TECNOLÓGICO NACIONAL DE MÉXICO.

Instituto Tecnológico de Tuxtla Gutiérrez.

Subdirección Académica.

Área: Ciencias Básicas.

AÑO SABÁTICO

Elaboración de Estrategias didácticas (Objetivos Educacionales) Aprendizajes Basados en Problemas (ABP) para: "Estadística Inferencial I y II". Asignatura/Sector.

Autor: **Javier Alfaro Mendoza**

Periodo: del 28 de enero de 2018 al 28 de enero de 2019.

Elaboración de Estrategias didácticas (Objetivos Educacionales) Aprendizajes Basados en Problemas (ABP) para: "Estadística Inferencial I y II". Asignatura/Sector.

Éste libro es dedicado con
cariño a mis alumnos,
motivo de mi dedicación.

Agradecimientos.

El presente trabajo, que se concluye como un Libro elaborado durante el ejercicio del periodo Sabático, fue realizado con mucho entusiasmo y es producto de muchos años de mi quehacer docente, pero no hubiese sido posible si no se contara con el decidido apoyo de la Jefatura del Departamento de Ciencias Básicas y sobre todo del respaldo de la Academia de dicho departamento, por lo que manifiesto mi sincero agradecimiento a mis compañeros que apoyaron éste proyecto.

Introducción.

El presente trabajo, es el resultado de la experiencia y trabajos de varios años del estudio de la materia de Estadística que se imparte en el Instituto Tecnológico de Tuxtla Gutiérrez, y de los demás tecnológicos del Sistema Nacional Mexicano de enseñanza técnica.

El trabajo se divide en dos partes, tanto en el contenido como en el tiempo realizado. En la primera etapa, se hace un estudio del programa de Estadística Inferencial I, elaborado e el primer semestre del año, del cual ya se entregó el reporte correspondiente. La segunda etapa considera el programa de Estadística Inferencial II correspondiente al segundo semestre del año 2018, del cual el reporte final es la conclusión del año sabático y se presenta en un solo reporte final el total del trbajo que incluye a los dos programas de estadística Inferencial.

El título del trabajo desarrollado en el año sabático es "Elaboración *de Estrategias Didácticas (Objetivos Educacionales) Aprendizajes Basados en Problemas (ABP) para: Estadística Inferencial I y II"* con el uso de software estadístico Minitab y Excel

Los temas de éste reporte, están organizados de acuerdo al programa oficialmente autorizado con clave AEF1024 para Estadística Inferencial I y Clave AEF1025 para el de Estadística Inferencial II, que se indica en el Índice correspondiente como Anexo 2; es importante hacer la aclaración que se le ha incluido en un principio, parte de la Estadística Descriptiva, ya que es necesario, al menos, así se consideró pertinente, porque se recurre constantemente a éste tema, y el alumno debe adquirir practica y habilidad en el uso del software y utilizarlo para la resolución de problemas que le auxilien en los demás temas afines.

Es importante hacer la aclaración, que al principio del Primer periodo de trabajo, es decir, en la Estadística Inferencial I, son pocos los problemas resueltos por unidad, debido a que los cálculos no son laboriosos como en los siguientes, donde presentamos una mayor cantidad de problemas a resolver con software. Por otro lado, los problemas aquí presentados, incluyen una "hoja de trabajo" o de cálculo que contiene el enunciado y las preguntas que se pueden extraer de la información que cada carátula de software que utilice y además Minitab requiere de un orden de los datos para poder procesarlo. También, antes de presentar los resultados ya sea en Minitab o Excel, se presenta los COMANDOS dentro de un recuadro para que el alumno tenga presente y a mano los pasos a seguir para cada caso y le sirva para su consulta posterior.

Introducción.

Respecto al programa de Estadística Inferencial I, en la primera Unidad se manejan dos problemas, el primero de la forma de utilizar el software para obtener muestras representativas de la población y el segundo se aplica conceptos para entender la Teoría Central de Límites, resolviendo un problema. En la segunda Unidad, se refiere a la Teoría de estimación, se presentan tres problemas relacionados con los Intervalos de Confianza, una breve introducción explicando en Power Point de la forma de calcular analíticamente el tamaño de una muestra y una tabla preestablecida para éste fin que se encuentra en el anexo. En la Unidad Tres, de Pruebas de Hipótesis, se presentan ocho problemas con diversos panoramas prácticos de aplicación. La Unidad cuatro, con el tema de Bondad de ajuste, se presentan cuatro problemas de CHI-cuadrada y dos problemas de un enfoque gráfico de normalidad. Para la Unidad cinco, que se refiere a la Regresión Lineal Simple y Múltiple, se presentan dos problemas de Correlación como parte introductoria y cinco problemas de Regresión Lineal Simple así como cuatro problemas de Regresión Lineal Múltiple. Los problemas fueron seleccionados cuidadosamente que además de ser prácticos tienen un alto contenido didáctico. Hay también en algunos unidades, unos breves recordatorios teóricos, como en el caso de la Regresión, Multicolinealidad, la obtención práctica de tamaños de muestras, Tablas para muestras colocadas en el los anexos y una evaluación.

En cuanto al programa de Estadística Inferencial II, es importante hacer la aclaración, que la unidad I de este, es la misma de la unidad V del programa de Estadística Inferencial I, por lo que ya no se incluye para no repetir dicha unidad. La unidad II, se refiere al tema de Series de tiempo que para empezar se presenta un documento en Power Point para explicar algunos temas y dar ejemplos sobre datos de algunas empresas que son muy didácticas y se presentan tres problemas relacionados con lo mismo. En la unidad III se presentan cinco problemas sobre ANOVA de un factor, resueltos en Excel y Minitab. La unidad IV de Diseños de Bloques, se presenta una brevísima explicación referente a los diseños de los bloques que se pueden presentar o acomodar los problemas y por último la unidad V referente a ANOVA de dos o más factores en diferentes arreglos, con problemas didácticos y prácticos para el alumno.

Para concluir, es importante considerar, que con éste trabajo, se pretende que el alumno, además de aprender estadística, aprenda el uso de los medios tecnológicos de la información para resolver problemas laboriosos de estadística y, aprovechar que los software existentes, tienen una mayor cantidad de cálculos, que muchas veces no son estudiados en el pizarrón por el docente.

CONTENIDO

PROBLEMA 1. 100 DATOS ESTATURA

Excel. Estaturas.

A continuación se dan los valores de una muestra de 100 datos de estaturas en pulgadas de estudiantes de una Universidad y se desea conocer: la Media, Moda, Mediana, Rango, Desviación Estándar, Varianza, Histograma y el Polígono de frecuencias así como el Intervalo de confianza de la media al 95% de Nivel de Confianza.

X			
60	66	68	70
61	66	68	70
61	66	68	70
62	66	68	70
62	66	68	70
63	66	68	70
63	66	68	70
64	66	68	70
64	66	68	71
64	66	68	71
64	67	68	71
64	67	68	71
64	67	68	71
64	67	68	71
65	67	68	71
65	67	69	71
65	67	69	71
65	67	69	72
65	67	69	72
65	67	69	72
65	67	69	72
65	67	69	72
65	67	69	73
66	67	70	73
66	68	70	74

Columna1	
Media	67.54
Error típico	0.28653979
Mediana	68
Moda	68
Desviación estándar	2.86539789
Varianza de la muestra	8.21050505
Curtosis	-0.17052315
Coeficiente de asimetría	-0.18950358
Rango	14
Mínimo	60
Máximo	74
Suma	6754
Cuenta	100
Nivel de confianza (95.0%)	0.56855711
	Intervalo de confianza

COMANDOS:

Datos – análisis de datos – seleccionar Estadística Descriptiva – aceptar – rango de entrada – seleccionar todos los datos con el cursor – agrupados por columnas o filas (según como estén ordenados los datos) – resumen estadístico – nivel de confianza – (seleccionar el deseado) – rango de salida (marcar una celda vacía con el cursor donde se desea anotar los resultados – Aceptar.

NOTA: Para construir un Histograma y Polígono a gusto del investigador:

COMANDOS:

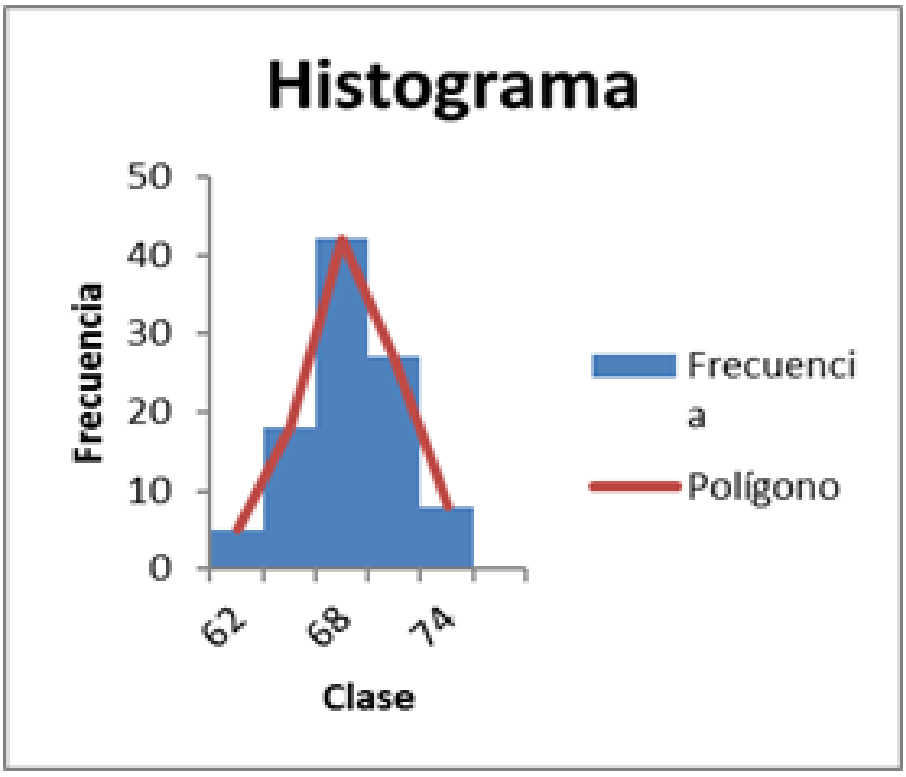
Datos – Análisis de datos – Histograma – Aceptar – Rango de entrada (valores de X) – Rango de clase – (valores de clase de salida) crear gráfico – Rango de salida con el cursor en una celda vacía – Aceptar – Seleccionar barra azul con clic derecho – Seleccionar datos – Agregar –Nombre de serie (polígono) – Valores de la serie (limpiar ícono) – Seleccionar frecuencias – Aceptar – Aceptar – Clic derecho – Seleccionar barra roja – Dar formato de serie – Poner en ceros – Enter – Clic derecho barra roja – Cambiar tipo de gráfica – Cuadro combinado – Listo.

Marcas de clase

Entrada	**Salida**
60	62
63	65
66	68
69	71
72	74

Clase	*Frecuencia*
62	5
65	18
68	42
71	27
74	8
y mayor...	0

Clase	*Frecuencia*
62	5
65	18
68	42
71	27
74	8
y mayor...	0



Problema 1.

Hoja de trabajo. Estaturas.

Minitab.

Una muestra de 100 estudiantes de una Universidad se obtuvo las estaturas en pulgadas como se señala en la columna C1 (X) y se desea determinar: La Media, Moda, Mediana, Rango, Desviación Estándar, Varianza, Intervalo de Confianza de la media a 95%, Histograma y Polígono de frecuencias.

X			
60	66	68	70
61	66	68	70
61	66	68	70
62	66	68	70
62	66	68	70
63	66	68	70
63	66	68	70
64	66	68	70
64	66	68	71
64	66	68	71
64	67	68	71
64	67	68	71
64	67	68	71
64	67	68	71
65	67	68	71
65	67	69	71
65	67	69	71
65	67	69	72
65	67	69	72
65	67	69	72
65	67	69	72
65	67	69	72
65	67	69	73
66	67	70	73
66	68	70	74

COMANDOS:
CARGAR HOJA DE TRABAJO:

Archivo – Abrir hoja de trabajo – Tipo de archivo (Excel) – marcar Escritorio – seleccionar archivo – Abrir – listo

COMANDOS DE CÁLCULOS:

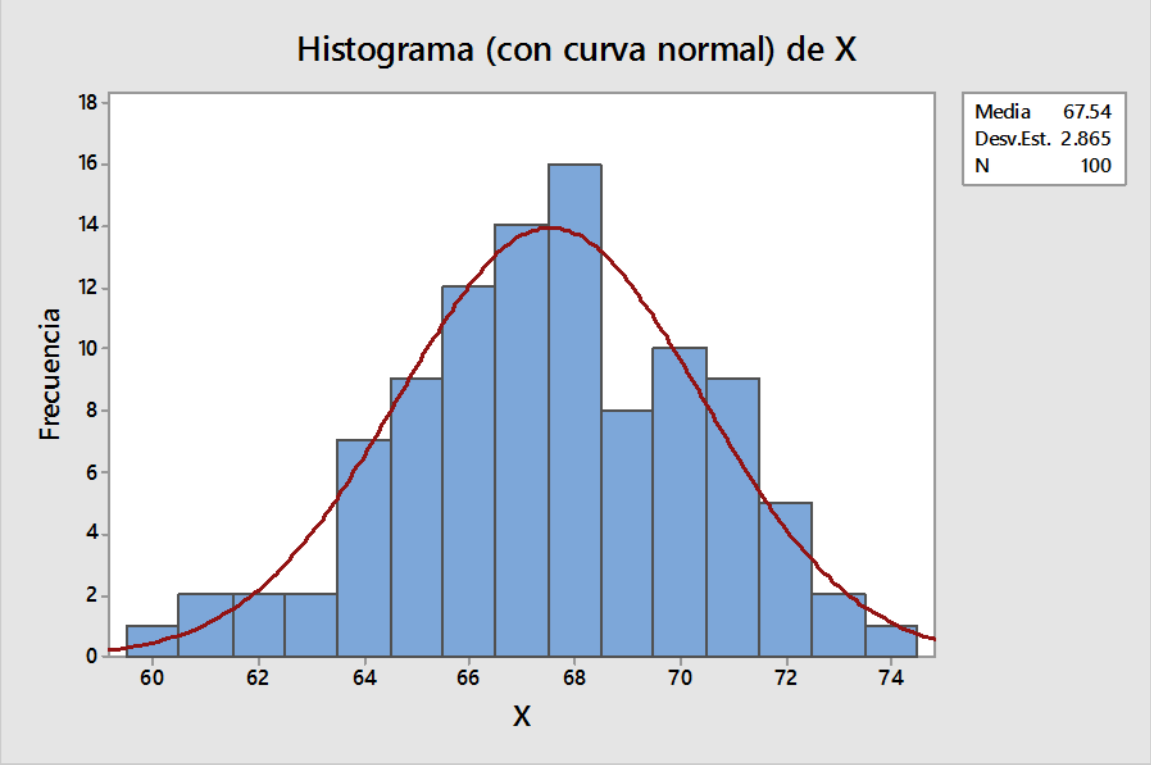
Estadísticas – Mostar estadísticos – Colocar cursos en cuadro de variables – Marcar C1 X – seleccionar – seleccionar gráficos – seleccionar estadísticas – marcar los cálculos deseados – aceptar – aceptar.

Estadísticos descriptivos: X

Variable	Conteo total	N	Media	Desv.Est.	Varianza	Mínimo	Q1	Mediana
X	100	100	67.540	2.865	8.211	60.000	66.000	68.000

Variable	Q3	Máximo	Rango
X	70.000	74.000	14.000

Histograma (con curva normal) de X



PROBLEMA 2. 100 CALIFICACIONES

Hoja de Trabajo. Calificaciones.

X			
42	64	74	78
48	65	75	79
50	65	75	79
51	65	75	79
52	67	75	79
53	67	75	79
54	68	75	80
54	68	75	81
55	69	75	82
56	69	75	70
57	69	75	71
58	69	75	72
59	69	76	72
60	70	76	82
60	70	76	82
60	70	76	83
60	70	77	84
61	73	77	85
62	73	77	85
62	73	77	85
63	73	77	85
63	73	78	86
63	73	78	93
64	74	78	95
64	74	78	97

Un grupo de 100 calificaciones de estadística, se obtuvieron los Siguientes Promedios (X) según tabla adjunta. (Columna A)

Determinar:

- La media, moda, mediana, desviación estándar y varianza.
- Construir el Histograma y el Polígono de Frecuencias
- Intervalo de Confianza?

Problema 2.

X			
42	64	73	78
48	65	73	78
50	65	74	78
51	65	74	78
52	67	74	78
53	67	75	79
54	68	75	79
54	68	75	79
55	69	75	79
56	69	75	79
57	69	75	80
58	69	75	81
59	69	75	82
60	70	75	82
60	70	75	82
60	70	75	83
60	70	76	84
61	70	76	85
62	71	76	85
62	72	76	85
63	72	77	85
63	73	77	86
63	73	77	93
64	73	77	95
64	73	77	97

Excel. Calificaciones.

Un grupo de 100 calificaciones de estadística, se obtuvieron los siguientes Promedios (X) según tabla adjunta. (Columna A):

Determinar:

- La media, moda, mediana, desviación estándar y varianza.
- Construir el Histograma y el Polígono de Frecuencia.
- Intervalo de Confianza?

Columna1	
	MARCA DE CLASES
Media	71.14
Error típico	1.02523907
Mediana	73
Moda	75
Desviación estándar	10.2523907
Varianza de la muestra	105.111515
Curtosis	0.2608531
Coeficiente de asimetría	-0.31879329
Rango	55
Mínimo	42
Máximo	97
Suma	7114
Cuenta	100
Nivel de confianza.(95.0%)	2.03429674

Entrada	Salida
42	49
50	57
58	65
66	73
74	81
82	89
90	97

Clase	*Frecuencia*
49	2
57	9
65	18
73	23
81	35
97	3
y mayor...	0

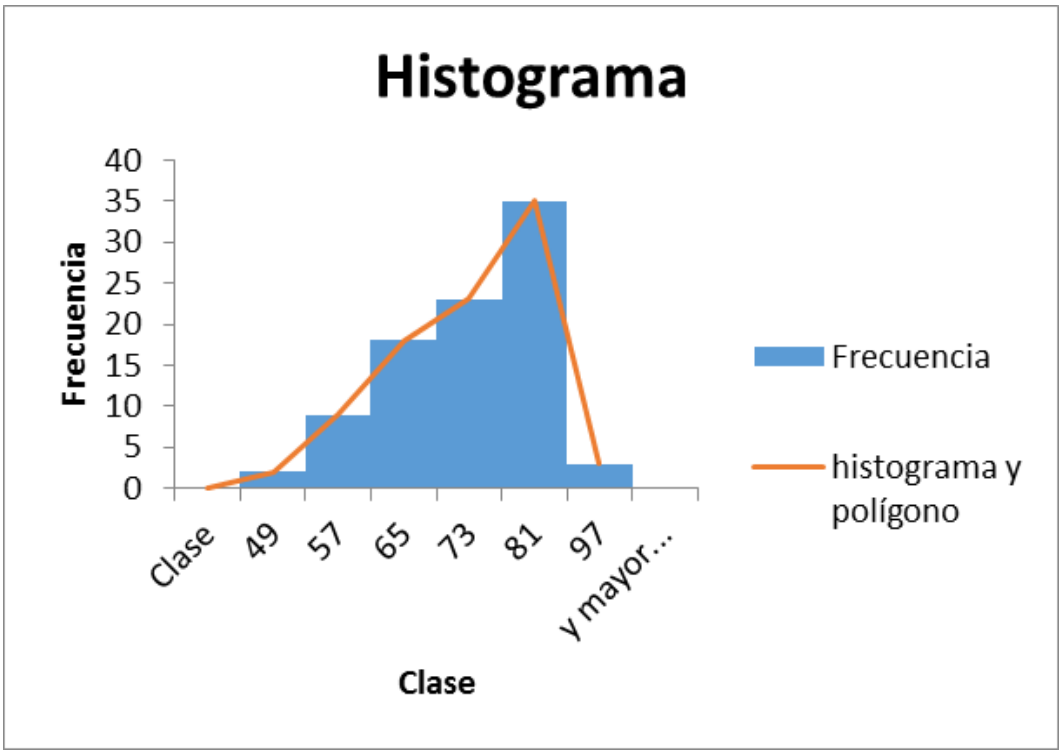
COMANDOS:

DATOS – análisis de datos – seleccionar Estadística Descriptiva – aceptar – rango de entrada – seleccionar todos los datos con el cursor – agrupados por columnas o filas (según como estén ordenados los datos) – resumen estadístico – nivel de confianza – (seleccionar el deseado) – rango de salida (marcar una celda vacía con el cursor donde se desea anotar los resultados – aceptar.

NOTA: Para construir un Histograma y Polígono a gusto del investigador:

COMANDOS:

Datos – análisis de datos – Histograma – Aceptar – Rango de entrada (valores de X) –rango de clase – (valores de clase de salida) crear gráfico – Rango de salida con el cursor en una celda vacía – aceptar – seleccionar barra azul con clic derecho – seleccionar datos – agregar –nombre de serie (polígono) – valores de la serie (limpiar ícono) – seleccionar frecuencias – aceptar – aceptar – clic derecho seleccionar barra roja – dar formato de serie – poner en ceros – enter – clic derecho barra roja – cambiar tipo de gráfica – cuadro combinado listo.



Problema 2.

Minitab. Calificaciones.

Un grupo de 100 calificaciones de estadística, se obtuvieron las siguientes Promedios (X) según tabla adjunta. (Columna A), Determinar:

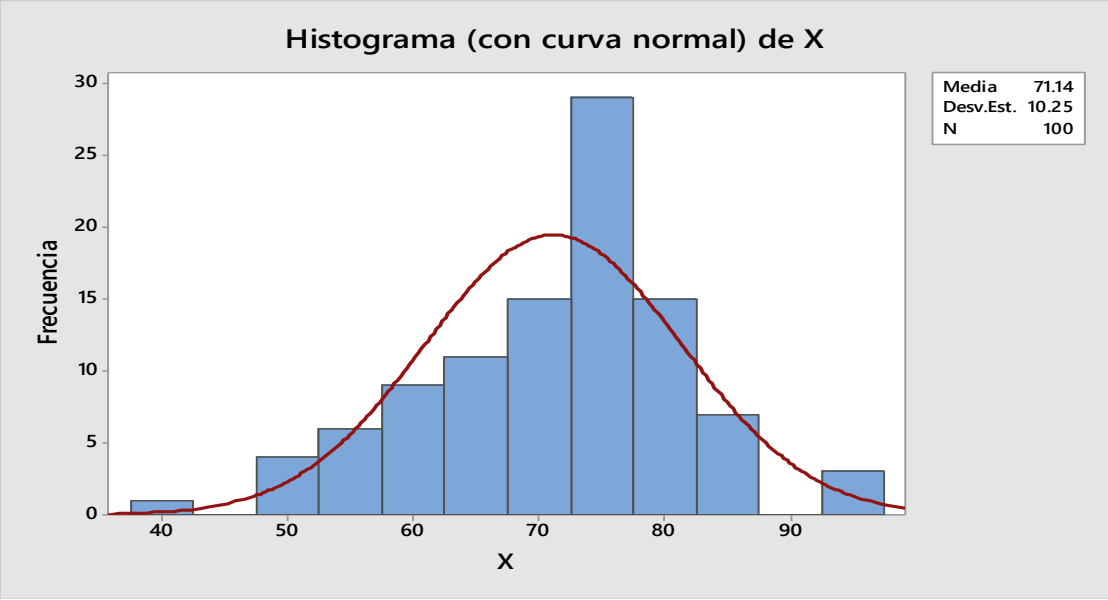
- La media, moda, mediana, desviación estándar y varianza.
- Construir el Histograma y el Polígono de Frecuencias
- Intervalo de Confianza?

COMANDOS:
CARGAR HOJA DE TRABAJO:
Archivo – Abrir hoja de trabajo – Tipo de archivo (Excel) – marcar Escritorio – seleccionar archivo – Abrir – listo.

COMANDOS DE CÁLCULOS:
Estadísticas – Mostar estadístico – Colocar cursos en cuadro de variables – Marcar C1 X – seleccionar – seleccionar gráficos – seleccionar estadísticas –marcar los cálculos deseados – aceptar – aceptar.

Estadísticos descriptivos: X

Variable	N	N*	Media	Error estándar de la media	Desv.Est.	Mínimo	Q1	Mediana	Q3	Máximo
X	100	0	71.14	1.03	10.25	42.00	64.00	73.00	77.75	97.00



PROBLEMA 3. PROMEDIOS DE MATEMÁTICAS

Excel. Promedios

La lista adjunta (A) son calificaciones de matemáticas de tres grupos: Determinar La Media, Moda, Mediana, Rango, Desviación Estándar, Varianza, Histograma y Polígono; así como el Intervalo de Confianza?

X			
53	71	78	94
57	71	78	95
59	72	78	95
60	72	78	96
60	73	79	97
60	73	79	
61	73	79	
61	73	80	
62	74	81	
62	74	82	
62	74	82	
62	75	83	
63	75	84	
63	75	85	
65	75	85	
65	75	85	
65	75	86	
66	75	87	
67	76	88	
67	76	88	
68	76	88	
68	76	89	
68	77	90	
69	77	93	
71	78	93	

Columna1	
Media	75.25
Error típico	1.1599187
Mediana	75
Moda	75
Desviación estándar	10.3746283
Varianza de la muestra	107.632911
Curtosis	-0.55539388
Coeficiente de asimetría	0.17120668
Rango	44
Mínimo	53
Máximo	97
Suma	6020
Cuenta	80
Nivel de confianza.(95.0%)	2.30876042

COMANDOS:

DATOS – análisis de datos – seleccionar Estadística Descriptiva – aceptar – rango de entrada – seleccionar todos los datos con el cursor – agrupados por columnas o filas (según como estén ordenados los datos) – resumen estadístico – nivel de confianza – (seleccionar el deseado) – rango de salida (marcar una celda vacía con el cursor donde se desea anotar los resultados – aceptar.

NOTA: Para construir un Histograma y Polígono a gusto del investigador:

COMANDOS:

Datos-análisis de datos – Histograma – Aceptar – Rango de entrada (valores de X) – rango de clase – (valores de clase de salida) crear gráfico – Rango de salida con el cursor en una celda vacía – aceptar – seleccionar barra azul con clic derecho – seleccionar datos – agregar – nombre de serie (polígono) – valores de la serie (limpiar ícono) – seleccionar frecuencias – aceptar – aceptar – clic derecho seleccionar barra roja – dar formato de serie -poner en ceros enter – clic derecho barra roja – cambiar tipo de gráfica – cuadro combinado – listo.

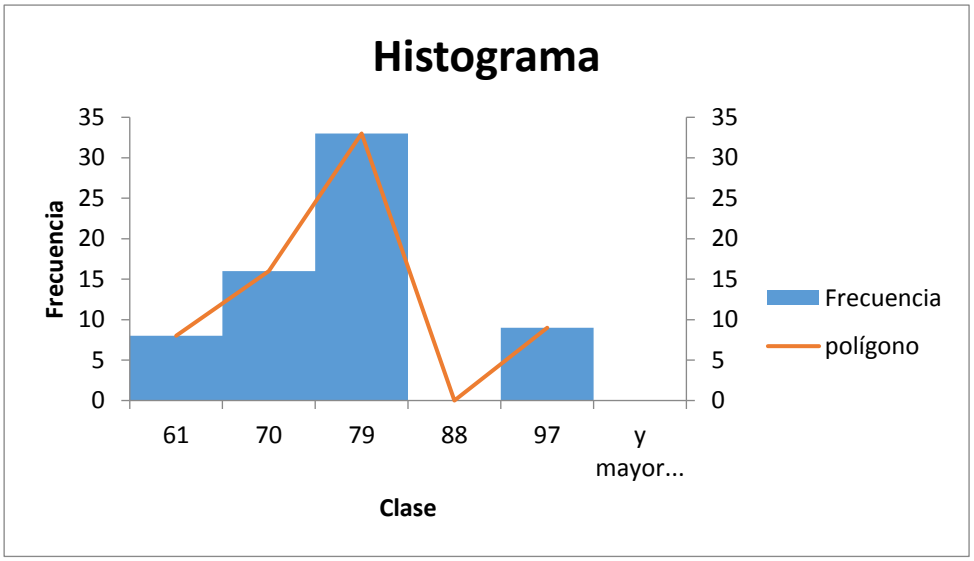
97-53=44+1=45/9=5

Maraca de clases

53	61
62	70
71	79
80	88
89	97

c=9

	Frecuencia
61	8
70	16
79	33
88	*Clase*
97	9
y mayor...	0



Problema 3.

Minitab. Promedios.

La lista adjunta (A) son calificaciones de matemáticas De tres grupos: Determinar La Media, Moda, Mediana, Rango, Desviación Están dar, Varianza, Histograma y Polígono; así como el Intervalo de Confianza.

COMANDOS:

CARGAR HOJA DE TRABAJO: Archivo – Abrir hoja de trabajo – Tipo de archivo (Excel) – marcar Escritorio –seleccionar archivo – Abrir – listo.

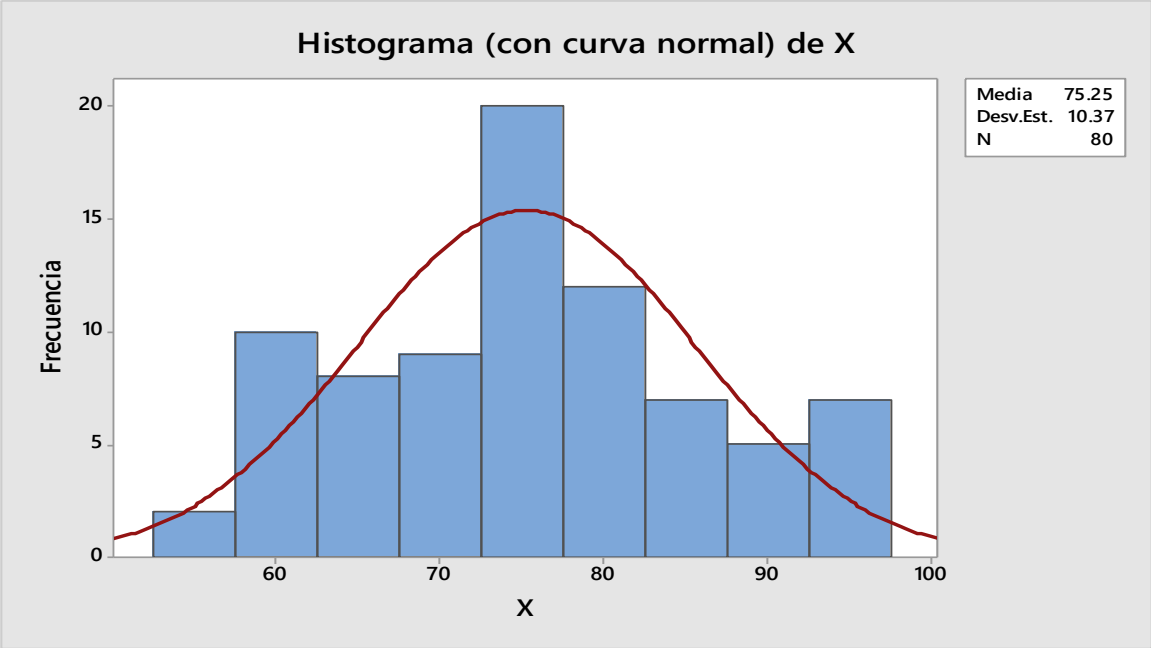
COMANDOS DE CÁLCULOS:

Estadísticas – Mostar estadísticos – Colocar cursos en cuadro de variables – Marcar C1 X – seleccionar – seleccionar gráficos – seleccionar estadísticas –marcar los cálculos deseados – aceptar – aceptar.

Estadísticos descriptivos: X

Variable	N	Media	Desv.Est.	Varianza	Mínimo	Q1	Mediana	Q3	Máximo
X	80	75.25	10.37	107.63	53.00	67.25	75.00	82.00	97.00

Variable	Modo	N para moda
X	75	7



PROBLEMA 4. HISTOGRAMAS DE FRECUENCIAS.

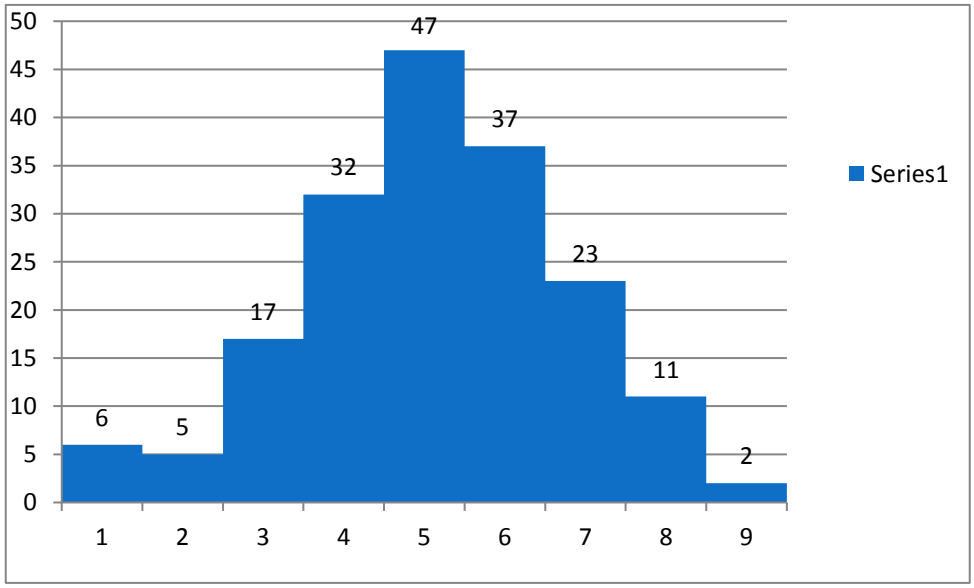
Problema 4. **Excel**.

En la columna A se tiene las ganancias de la ventas de 180 autos usados durante un mes en la columna B esta las frecuencias de cada ganancia, Construir un Histograma de Frecuencias para tener una visión más clara del comportamiento de las ventas de la empresa.

Ganancia Dlls.
400
800
1200
1600
2000
2400
2800
3200
3600

COMANDOS:

Marcar con el cursor las columnas de Frecuencias – seleccionar gráfico D (la primera) – Clic derecho sobre la gráfica – Dar formato sobre serie de datos (poner en ceros) – Enter



PROBLEMA 5. DIAGRAMA DE PUNTOS.

Problema 5. **Minitab.**

Tionesta	Sheffield
23	31
30	30
29	32
35	26
33	35
32	37
25	44
32	38
27	44
28	43
36	36
35	37
28	36
33	31
31	34
37	30
39	34
35	40
32	43
36	42
26	37
32	31
27	36
30	33

Los departamentos de servicios de Tionesta Ford Lincoln Mercury y Sheffield Motors Inc., Dos de las cutro distribuidoras de Aplplewood Auto Group, abrieron 24 días hábilies Del mes pasado. A continuación aparece el número de vehículos que recibieron servicio el mes pasado en ambas distribuidoras.

Tionesta Ford Lincoln Mercury					
Lunes	Martes	Miércoles	Jueves	Viernes	Sábado
23	33	27	28	39	26
30	32	28	33	35	32
29	25	36	31	32	27
35	32	35	37	36	30

Sheffield Motors Inc.					
Lunes	Martes	Miércoles	Jueves	Viernes	Sábado
31	35	44	36	34	37
30	37	43	31	40	31
32	44	36	34	43	36
26	38	37	30	42	33

Minitab proporciona un diagrama de puntos y permite calcular la media, la mediana, Los valores máximos y mínimos y la desviación estándar de la cantidad de autos que recibieron servicio en cada concesionaria durante los pasados 24 días hábiles.

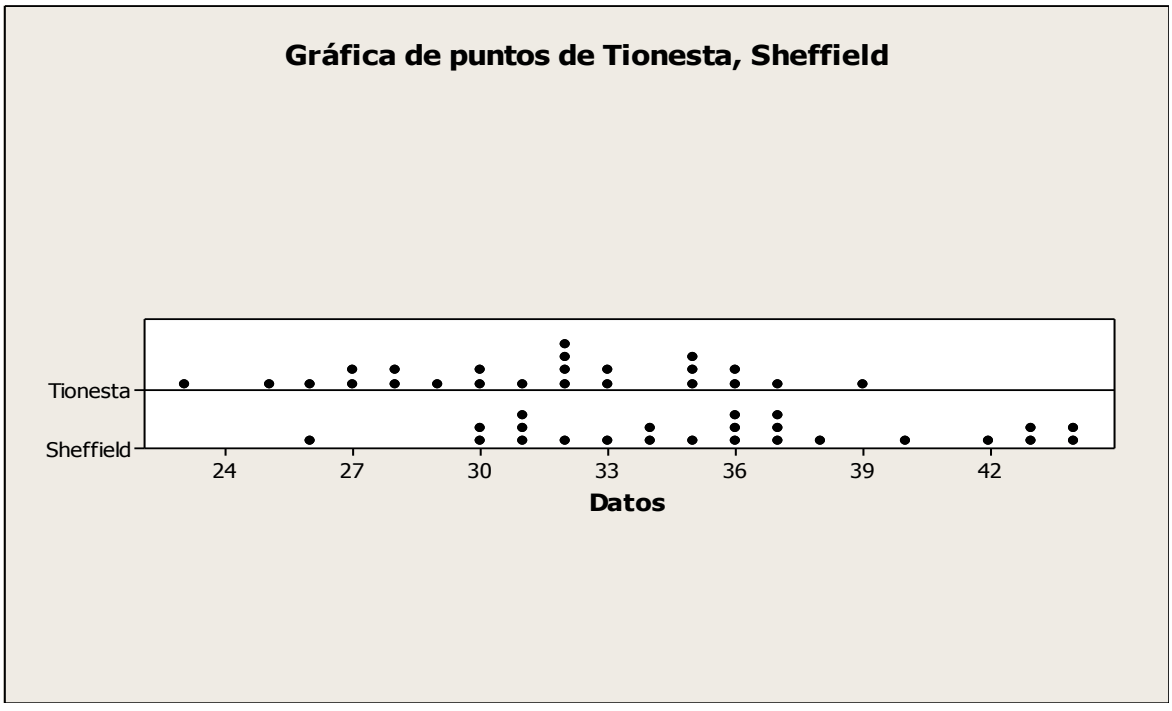
COMANDOS:

Cargar Hoja de Trabajo – Gráficas – Gráfica de puntos – seleccionar múltiples y simples – aceptar – variable de gráficas: seleccionar C1 y C2 – aceptar.

Estadísticos descriptivos: Tionesta, Sheffield

Variable	N	N*	Media	Error estándar de la media	Desv.Est.	Mínimo	Q1	Mediana	Q3
Tionesta	24	0	31.292	0.839	4.112	23.000	28.000	32.000	35.000
Sheffield	24	0	35.83	1.01	4.96	26.00	31.25	36.00	39.50

Variable	Máximo
Tionesta	39.000
Sheffield	44.00



ANÁLISIS:

° Al observar los esquemas de puntos, se puede ver que el número de autos que recibieron servicio en la distribuidora Sheffield están más dispersos, tienen una desviación estándar Sheffield de 4.96 autos por día, mayor que Tionesta de 4.112

° Tionesta dio servicio a menos automóviles en cualquier día, 23.

° Sheffield dio servicio a 26 autos su día más bajo.

° Tionesta dio servicio exactamente a 32 autos en cuatro días diferentes.

° Los números de autos que recibieron servicio se acumulan alrededor de 36 en el caso de Sheffield y 32 en el caso de Tionesta.

° A partir de la estadística descriptiva, es posible visualizar que Sheffield dio servicio a un promedio de 35.85 autos diarios y Tionesta un promedio de 31.292 autos al día en el mismo periodo.

PROBLEMA EVALUACIÓN

De las siguientes calificaciones de Química que aparecen en la columna A.

Determinar usando el Software de Excel y Minitab: la Media, Moda, Mediana, desviación Estándar, Varianza, Rango, Intervalo de Confianza y las gráficas Histograma y Polígono de frecuencias?.

X			
60	69	70	71
65	69	70	71
65	69	70	71
65	69	70	71
66	69	70	72
66	69	70	72
66	69	70	72
67	69	70	72
67	69	70	72
67	69	71	72
67	69	71	72
68	70	71	73
68	70	71	73
68	70	71	73
68	70	71	73
68	70	71	74
68	70	71	74
68	70	71	75
69	70	71	75
69	70	71	76

ESTADÍSTICA INFERENCIAL I

UNIDAD I.

DISTRIBUCIONES MUESTRALES.

PROBLEMA 1.1. MUESTREO

Excel.

En una pensión donde se dan alojamiento y desayuno. El negocio cuenta con 8 habitaciones. A continuación aparece el número de estas ocho habitaciones rentadas diariamente durante el mes de junio. Extraer una muestra de 5 noches de junio.

Junio	Habitaciones rentadas
1	0
2	2
3	3
4	2
5	3
6	4
7	2
8	3
9	4
10	7
11	3
12	4
13	4
14	4
15	7
16	0
17	5
18	3
19	6
20	2
21	3
22	2
23	3
24	6
25	0
26	4
27	1
28	1
29	3
30	3

muestra 1	muestra 2	muestra 3
4	0	4
3	0	4
3	4	0
5	6	4
3	1	7

COMANDOS:

Datos – análisis de datos – muestra – aceptar – rango de entrada (seleccionar todos los datos de la columna de habitaciones rentadas) – marcar aleatorio y señalar el número del tamaño de la muestra – colocar el cursor en rango de salida y señalar la celda donde se desea el resultado – aceptar.

PROBLEMA 1.2. DISTRIBUCIÓN MUESTRAL DE MEDIAS

Excel.

Demostración del Teorema central del límite: "Si todas las muestras de un tamaño en particular se seleccionan de cualquier población, la distribución muestral de la media se aproxima a una distribución normal. Esta aproximación mejora con muestras más grandes".

Tartus Industries cuenta con siete empleados de producción (a quienes se les considera la población). En la siguiente tabla se incluyen los ingresos por hora por cada uno de ellos.

Si graficamos los datos de la tabla tal como están, observamos que no se tiene una distribución normal (fig, 1).

Empleado	Ingreso/hora
Joe	7
Sam	7
Sue	8
Bob	8
Jan	7
Art	8
Ted	9

frecuencias	
7	3
8	3
9	1

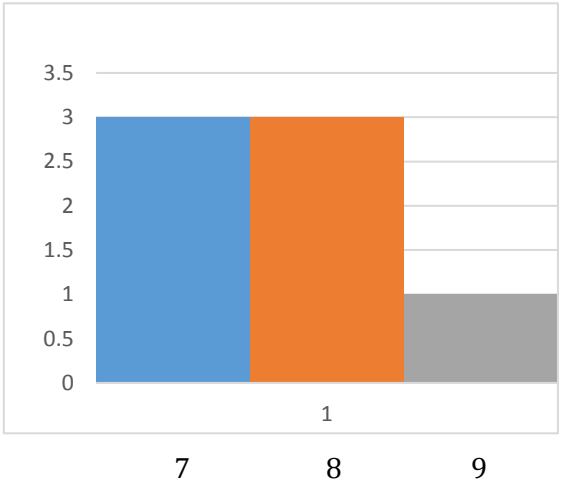


figura 1

Pero, si sacamos el número de combinaciones posibles en muestras de 2 en 2, se tienen 21 muestras posibles.

nCr = n!/r! (n-r)! = 7!/ 2!(7-2)! = 21

Si graficamos en un Histograma, observamos que la distribución muestral de la distribución muestral de medias tiene una mayor aproximación a la distribución normal.

1	7+7	7
2	7+8	7.5
3	7+8	7.5
4	7+7	7
5	7+8	7.5
6	7+9	8
7	7+8	7.5
8	7+8	7.5
9	7+7	7
10	7+8	7.5
11	7+9	8
12	8+8	8
13	8+7	7.5
14	8+8	8
15	8+9	8.5
16	8+7	7.5
17	8+8	8
18	8+9	8.5
19	7+8	7.5
20	7+9	8
21	8+9	8.5

Frecuencia	
7	3
7.5	9
8	6
8.5	3

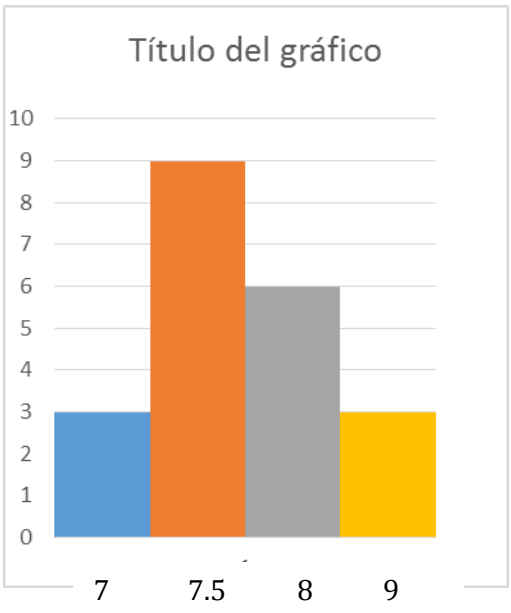


Fig.2

UNIDAD II.

ESTIMACIÓN.

PROBLEMA 2.1. INTERVALOS DE CONFIANZA, CLIENTES MALL

Excel.

El gerente de Inlet Square Mall, cerca de FT. Myers, Florida, desea estimar la cantidad media que gastan los clientes que visitan el centro comercial. Una muestra de 20 clientes revela las siguientes cantidades:

¿Cuál es la mejor estimación de la media poblacional?

Utilizar un intervalo de confianza de 95%

¿Concluiría de forma razonable que la media poblacional es de $50? ¿Y de $60?

Amount
48.16
42.22
46.82
51.45
23.78
41.86
54.86
37.92
52.64
48.59
50.82
46.94
61.83
61.69
49.17
61.46
51.35
52.68
58.84
43.88

48.16	
Media	49.4105263
Error típico	2.12315231
Mediana	50.82
Moda	#N/A
Desviación estándar	9.25460634
Varianza de la muestra	85.6477386
Curtosis	2.06145372
Coeficiente de asimetría	-0.99914822
Rango	38.05
Mínimo	23.78
Máximo	61.83
Suma	938.8
Cuenta	19
Nivel de confianza(95.0%)	**4.46057748**

48.16	
Nivel de confianza 95.0%)	4.46057748

COMANDOS: Datos – análisis de datos – estadística descriptiva – aceptar – rango de entrada – seleccionar toda la columna de datos numéricos – agrupados por: columna o fila según como estén ordenados los datos.

En este caso en columnas – marcar resumen estadístico – nivel de confianza deseado – rango de salida, marcar la celda donde se desea los cálculos – aceptar.

¿Cuál es la mejor estimación de la media poblacional? La media= 49.4105263 (Estimado puntual)

¿Cuál es el intervalo de confianza a 95%? : 53.87 y 44.95 (49.41+4.46=53.87 y 49.41-4.46=44,95)

¿Se puede concluir de forma razonable que los clientes gastan $50 dls.? 60 dls.?

Conclusiones:

Se puede concluir que sí, porque $50 dls. Esta dentro del intervalo de confianza pero no el de $60 dls. que cae fuera del intervalo.

También se puede interpretar que el 95% de los clientes del centro comercial gastan entre 53.87 a 44.95 dls. Y el 5 % restante están fuera de este rango o intervalo

También se puede afirmar que el 2.5% de los clientes gastan menos de 44.95 dls. por visita y otro 2.5% gasta más de 53.87 dls.

Minitab.

Primero se determina la media y la Desv. stand. y luego prueba t de una muestra.

El gerente de Inlet Square Mall, cerca de FT. Myers, Florida, desea estimar la cantidad que gastan los clientes que visitan el centro comercial. Una muestra de 20 clientes con las siguientes cantidades: (la señalada en la hoja de Excel)

¿Cuál es la mejor estimación de la media poblacional?

Utilizar un intervalo de confianza de 95%

¿Concluiría de forma razonable que la media poblacional es de $50? ¿Y de $60?

COMANDOS:

Estadísticas – estadística básica – t de una muestra – seleccionar datos –seleccionar una o más muestras, cada una de una columna – opciones determinar qué nivel de confianza se desea trabajar (en este caso 95%) – definir la hipótesis alternativa diferente respecto de la hipotética – aceptar – en el 2o. recuadro seleccionar Amount – aceptar.

Nota: en el caso de seleccionar datos resumidos, se tiene que calcular primero la media y la Desv. Std. que es información que no se tiene.

Amount	Media1	N1	Des. Std.
48.16	49.35	20	9.01
42.22			
46.82			
51.45			
23.78			
41.86			
54.86			
37.92			
52.64			
48.59			
50.82			
46.94			
61.83			
61.69			
49.17			
61.46			
51.35			
52.68			
58.84			
43.88			

T de una muestra: Amount

Variable	N	Media	Desv.Est.	Error estándar de la media	IC de 95%
Amount	20	49.35	9.01	2.02	(45.13, 53.57

El mejor estimado de la media es el estimado puntual = 49.35 dls.

El Intervalo de confianza esta entre 45.13 y 53.57 dls. de la media

$50 dls. sí es posible gasten los clientes puesto que está dentro del intervalo de confianza.

$60 dls. es probable que nó por estar fuera del intervalo de confianza.

Se puede asegurar con un nivel de confianza del 95% que los clientes gastan en promedio entre 45.13 y 53.57 dls. por visita al centro comercial.

PROBLEMA 2.2. INTERVALO DE CONFIANZA. GERENTES

Excel.

n	Media	Desv. Std	NC	IC
256	45420	2050	95%	

La American Management Association desea información acerca del ingreso medio de los gerentes de la industria del menudeo. Una muestra aleatoria de 256 gerentes revela una media muestral de $45,420.00, La desviación estándar de esta muestra es de 2050-

La asociación le gustaría responder a las siguientes preguntas:

1.- ¿Cuál es la media de la población?

2.- ¿Cuál es el conjunto de valores de la media razonable para media poblacional?

3.- ¿Cómo se debe de interpretar estos resultados?

COMANDOS: Fórmulas – f (insertar función) – seleccionar categoría (estadísticas) –seleccionar función (INTERVALO DE CONFIANZA NORMAL) – Aceptar – en alfa anotar el nivel de significancia deseado – en desv. Stad. y media notar datos – aceptar.

NOTA: En la parte inferior del recuadro ya está el resultado.

Conclusiones:

El mejor estimado de la media es $45,420.00 y es un estimado puntual, se puede asegurar con un nivel de confianza del 95% que el intervalo de confianza esta entre $45671 .00 y $45,169.00

Estos resultados se interpreta que el 95 % de los gerentes de la industria al menudeo ganan entre $45,671.00 y $45,169.00 y solo el 5% está fuera de este rango.

Z de una muestra de estadística básica (por ser muestra grande)

Minitab.

La American Management Association desea información acerca de los ingresos medios de los gerentes de la industria del menudeo. Una muestra aleatoria de 256 gerentes revela una media muestral de $45420. La Desv. Std. de esta muestra es de $2050

La asociación le interesa contestar las siguientes preguntas:

1.- ¿Cuál es la media de la población (el mejor estimado)?
2.- ¿Cuál es el conjunto de valores razonable de la media poblacional?
3.- ¿Cómo se debe de interpretar estos resultados?

NOTA: Recordar que los salarios e ingresos tienen una distribución con sesgo positivo, pues unos cuantos individuos ganan unos pocos más que la mayoría, lo cual sesga la distrib., pero el Teorema Central del límite estipula que si se selecciona una muestra grande, la distribución de las medias muéstrales, tendera a seguir una distribución normal.

COMANDOS: intervalos de confianza: Estadística – estad. Básica – Z de una muestra – seleccionar datos resumidos – anotar los valores de la media, tamaño de la muestra y la Desv. Stad. que son datos – en opciones seleccionar el nivel de confianza deseado y la hipótesis alternativa deseado. Aceptar – aceptar.

Media	Desv. Std.	n
45420	2050	256

Z de una muestra

```
La desviación estándar supuesta = 2050

                Error
                estándar
                de la
 N    Media     media     IC de 95%
256   45420       128   (45169, 45671)
```

Conclusiones:

1.- El mejor estimado de la media es $45420.00 que es un estimado puntual.

2.- El intervalo de confianza a un nivel del 95% está entre $45169 y $45671

3.- Se interpreta que 95% de los gerentes de la industria del menudeo se encuentra entre este rango de salarios y solo el 5 % de ellos ganan menos de $45168 y más de $45671.

PROBLEMA 2.3. INTERVALO DE CONFIANZA: LLANTAS

Excel.

Un fabricante de llantas desea investigar la durabilidad de las llantas que fabrica. Toma una muestra de 10 llantas y las recorre 50,000 millas y mide el espesor de cuerda que le queda y tiene un espesor de 0.32" en promedio con una desv. std. De 0.09"

Utilizar un nivel de confianza de 95%

¿Sería razonable que el fabricante concluyera y garantizara sus llantas que después de 50,000 millas, la cantidad media de cuerda restante sea de 0.30 pulgs.?

n	media	Des. Std.	NC	nivel significancia	IC
10	0.32	0.09	95%	0.05	0.06438212

COMANDOS: Fórmulas – fx (insertar función) – seleccionar categoría (estadísticas) – seleccionar una función (INRERVALO DE CONFIANZA T) – aceptar en Alfa el nivel de significancia deseado (en este caso 0.05) – Desv. Std. (0-09 para este caso) – tamaño de la muestra (10 este caso) – aceptar

NOTA: en la parte inferior del recuadro ya nos da el resultado.

Conclusiones:

El intervalo de confianza a un nivel del 95% está entre : 0.256 y 0.384 pulgs.

Con un nivel de confianza del 95% se puede garantizar que las llantas después de recorrer 50,000 millas, les queda aún 0.30 pulg. de cuerda y solo el 5% de ellas probablemente no.

Otra forma de interpretarlo es que de 100 llantas fabricadas 95 de ellas se garanticen que terminaran después de 50,000 millas recorridas con 0.30 pulg. de cuerda y solo 5 de ellas no.

¿Porqué?

Porque 0.30 pulg. esta dentro del intervalo de confianza de más menos 0.064 pulg.

Minitab.

Un fabricante de llantas desea investigar la durabilidad de sus productos. Una muestra de 10 llantas que recorrieron 50,000 millas reveló una media muestral de 0.32 pulgs. de cuerda restante con una. Desv. Std. de 0.09 pulg..

Construir un intervalo de confianza de 95% de la media poblacional.

Con estos resultados, ¿puede el fabricante asegurar a sus clientes que las llantas que fabrica después de recorrer 50,000 millas aún le quedan 0.30 pulgs. de cuerda restante?

Media	Desv. Std	n	NC	NS
0.32	0.09	10	95.00%	0.05

T de una muestra

N	Media	Desv.Est.	Error estándar de la media	IC de 95%
10	0.3200	0.0900	0.0285	(0.2556, 0.3844)

Sí puede el fabricante garantizar su producto con un nivel de confianza del 95% que sus llantas después de recorrer 50,000 millas le quedan aún 0.30 pulgs. de espesor de cuerda restante, porque este valor se encuentra dentro del intervalo de 0.2556 a 0.3844 de pulg.

Solo el 5%de las llantas que fabrica, probablemente no lo cumplan o no lo garantizan.

(Una muestra es una decisión que se toma para que la estimación del parámetro poblacional sea bueno). Esta decisión está basada en tres variables:

1.- El margen de error.

2.-El Nivel de Confianza y

3.-La Variabilidad o Dispersión de la población.

(Desviación estándar)

Margen de error. (E)

Es el máximo error admisible determinado por el investigador, que se tolerará al estimar un parámetro poblacional.

Es la magnitud del error (E) que se suma y resta de la media muestral o proporción muestral para determinar los puntos extremos del intervalo de confianza. Existe una compensación entre el margen de error y el tamaño de la muestra.

Un margen de error pequeño requiere de una muestra más grande y de más tiempo y dinero para recolectarla.

Un margen de error más grande permitirá tener una muestra más pequeña y un intervalo de confianza más amplio.

Desviación Estándar de la población.

Si la población se encuentra muy dispersa, se requiere una muestra grande. Por el contrario, si se encuentra concentrada (homogénea), el tamaño de la muestra que se requiere será menor.

Para tal fin, es necesario utilizar un *estimador* de la desviación estándar de la población.

Sugerencias para determinar el estimador *variabilidad.*

1.-**Realizar un estudio piloto**.

Es uno de los más comunes: Se aplica a una ***pequeña*** muestra de la población a estudiar y se calcula la desviación estándar y se utiliza este valor como la desviación estándar de la población en la correspondiente ecuación para calcular el tamaño de la muestra.

2.- Utilizar un estudio comparativo.

Aplicar este enfoque cuando se encuentre disponible un estimador de la dispersión de *otro estudio* similar.

3.-Enfoque basado en el intervalo.

Se necesita conocer o contar con un cálculo de los valores máximos y mínimos de la población. Aplicando el concepto de una distribución normal de dividir toda el área de la dispersión en 6 partes o desviaciones estándares.

De esta manera, podemos estimar la desviación estándar de la población dividiendo el rango (valor máx –valor mín) entre 6.

CÁLCULO DEL TAMAÑO DE LA MUESTRA DE LA **MEDIA** POBLACIONAL (Variable de razón o cuantitativa).

Se puede expresar la interacción entre estos tres factores mediante la siguiente fórmula (notar que es la ecuación para determinar los extremos de los intervalos de confianza

$$E = Z\left(\frac{\sigma}{\sqrt{n}}\right) \qquad n = \left(\frac{z\sigma}{E}\right)^2$$

n = Tamaño de la muestra
Z = Valor del coeficiente del nivel de confianza deseado.
E =Tamaño del error máximo admisible
ϭ = Desviación estándar

TAMAÑO DE LA NUESTRA PARA LA **PROPORCIÓN.**

En el caso de la distribución binomial, el margen de error es:

$$E = Z\sqrt{\pi(1-\pi/n} \qquad \mathbf{n = \pi(1-\pi)(Z/E)^2}$$

n = tamaño de la muestra

Z = coeficiente del nivel de confianza deseado.

π = es la proporción de la población

E = es el máximo error tolerable.

En este caso, la desviación estándar de la población de una distribución normal está representada por $\pi/1-\pi$).

Para encontrar el valor de una proporción de la población, podemos hallar un estudio similar o conducir un estudio piloto. Si no se puede encontrar un valor confiable, entonces se debe usar un valor de π de 0.50, es decir, se sobre estima el tamaño de la muestra con este valor.

OTRO MÉTODO

Determinación de la muestra para la media poblacional (Variable de razón o cuantitativa). Analíticamente el tamaño de la muestra se determina:

$$n = \frac{(NS^2)}{\frac{NB^2}{Z^2} + S^2}$$

B es la precisión, pero como es una variable cuantitativa o de razón, B se da en términos desviación estándar S, es decir, D= B/S que se sustituye en la ecuación anterior y queda:

$$n = \frac{N}{\frac{ND^2}{Z^2} + 1}$$

Si se da, que B= 0.05, quiere decir que se da en la mitad de la desviación estándar. Es decir, que el estimado de la media, difiere en la mitad de la desviación estándar o menos

Lo práctico es que D no rebase 0.20, pero en tablas se da hasta 1.0.

USANDO LAS TABLAS: (ESTAS SE ENCUENTRAN EL ANEXO 1)

Ejemplo:

Si tenemos una población de egresados de N= 500, un nivel de confianza de 95%. Los egresados tuvieron una calificación entre 6 y 10, entonces 10-6=4 por lo tanto 4/6 = 0.66. Si escogemos una precisión de B = 0.1 entonces D=B/S = 0.1/0.66 = 0.15.

La precisión debe ser una fracción pequeña de la desviación estándar.

Si D=0.15; N.C. = 95% ; N = 500 de tablas n = 128 a N:C: 99% n = 186

Con la ecuación se tiene:

$$n = \frac{N}{\frac{ND^2}{Z^2} + 1}$$

$$n = \frac{500}{\frac{(500)(0.15)^2}{(1.96)^2} + 1}$$

n = 127.226

Para determinar el tamaño de una muestra para una proporción

$$n = \frac{N(1-p)}{\frac{(N-1)B^2}{Z} + p(1-p)}$$

N tamaño de la muestra.
B precisión
p proporción (estimado)
Z coeficiente del nivel de confianza deseado

p puede tener un valor máximo de 0.50 que da el valor más grande de "n", esto se hace cuando no se cuenta datos históricos o estudio piloto. Para cuando p>0.50 se toma (1-p).

B es la precisión, es una cantidad pequeña de la proporción de la variable cualitativa. Por ejemplo, si B=0.05, se dice que se desea hacer una estimación con precisión de 0.05 o menor de p

Usando tablas, con los siguientes datos se tiene un ejemplo:

N.C. = 80 %; N=200; B=0.05; y p=0.75

Usar p=(1-p) = (1-0.75)= 0.25 n= 77

N.C. = 90; N=200: p=0.75

Usar p=(1-p)= (1-0.75)=0.25 El tamaño de la muestra es de

n=101

Si se desea que B=0.07; N.C. = 90%; N= 257 y p= 0.22, entonces se busca en tablas con B=0.07; NC=90%; N=300 y p=0.25 y el tamaño de la muestra es de **n=78**.

Es decir, 257 se sube a 300 y p se sube a 0.25

Si se desea que B=0.07; N.C. = 90%; N= 257 y p= 0.22, entonces se busca en tablas con B=0.07; NC=90%; N=300 y p=0.25 y el tamaño de la muestra es de **n=78**.

Es decir, 257 se sube a 300 y p se sube a 0.25

$$n = \frac{N(1-p)}{\frac{(N-1)B^2}{Z} + p(1-p)}$$

UNIDAD III.

PRUEBA DE HIPÓTESIS.

PROBLEMA 3.1. PRUEBA DE HIPÓTESIS. COSTO QUEJAS.

Minitab.

Cost	
45	58
48	40
48	63
58	56
38	59
49	43
53	78
54	63
51	56
76	61
62	64
67	69
51	57

El departamento de quejas de una compañía de seguros en USA informa que el costo medio para tramitar una queja es de 60.00 dlls. Una comparación de la industria demostró que esta cantidad es mayor que las demás compañías de seguros, así que la compañía tomo medidas para reducir gastos.
Para evaluar el efecto de las medidas de reducción de gastos, el supervisor del Dpto. de quejas selecciono una muestra aleatoria de 26 quejas atendidas el mes pasado.

La información se muestra en la columna C1.

¿Es razonable concluir que el costo medio de atención de una queja ahora es menor que 60.00 dlls. con un nivel de significancia de 0.01?

COMANDOS:

Cargar hoja de trabajo – Estadística – estadística básica – t de una muestra seleccionar en recuadro superior una o más muestras en una columna – clic en recuadro intermedio y cargar costo – marcar en prueba de hipótesis – clic en media hipo tética Ho (60 en este caso) – opciones – seleccionar NC deseado – seleccionar hipótesis alterna H1 deseado(en este caso media <60) – aceptar – aceptar.

T de una muestra: cost

```
Prueba de μ = 60 vs. < 60

                                    Error
                                 estándar    Límite
                                    de la  superior
Variable   N  Media  Desv.Est.     media    de 95%      T      P
cost      26  56.42      10.04      1.97     59.79  -1.82  0.041
```

Para este problema se considera que la Ho=60 y la H1<60. de una cola. t crítica a 99% y 25 gl tiene un valor de -2.485. (de tablas)

Por lo tanto tcal=-1.82 es menor que tcrit=-2.485 por lo que cae en la zona de aceptación de la Ho; es decir se acepta la Ho.

Por lo tanto se concluye que las medidas para reducir costos no bajaron los costos por queja. Para corroborar esto, el valor de P=0.041>que el nivel de significancia =0.01 (recordar la regla de decisión que si P>n. significancia se acepta la Ho)

PROBLEMA 3.2. PRUEBA DE HIPÓTESIS. BARRAS

Minitab.

Length
42
39
42
45
43
40
39
41
40
42
43
42

La longitud media de una pequeña barra de contrapeso es de 43 mm. Al supervisor de producción le preocupa que hayan cambiado los ajustes de la máquina de producción de barras. Solicita una investigación al departamento de ingeniería que selecciona una muestra aleatoria de 12 barras y las mide.

Los resultados aparecen en la columna C1.

¿Es razonable concluir que cambio la longitud media de las barras?. Utilizar el nivel de significancia de 0.02

COMANDOS:
Cargar hoja de trabajo – estadística – estadística básica – t de una muestra – seleccionar en recuadro superior una o más muestras en una columna – clic en recuadro intermedio y cargar C1 (length) – clic en prueba de hipótesis – clic en media hipotética Ho (en este caso 43) – opciones – seleccionar NC deseado – seleccionar hipótesis alterna H1 deseada en este caso media diferente de 43 – aceptar – aceptar.

T de una muestra: length

```
Prueba de mu = 43 vs. no = 43
Ho=media ´poblacional = 43
H1: media pob.diferente de 43
t crítica a 98 (percentil 0.99)y 11 gl = 2.72

                                  Error
                                 estándar
                                  de la
Variable  N   Media  Desv.Est.    media      IC de 95%          T      P
length    12  41.500    1.784     0.515  (40.367, 42.633)    -2.91  0.014
```

Ho: si t calc.>tcrítica , la Ho se rechaza
H1: si t calc. <tcrítica se acepta Ho. Ho=43 H1 diferente de 43 dos colas
t crítica a 98% y 11 gl. =2.72 (de tablas)
P se interpreta como la prob. de que Ho sea cierta. Dado que p=0.014<que n. de significancia, se rechaza la Ho. Cae en la zona de rechazo.
Dado que tcal.=-2.91>tcritica=2.72 se rechaza la Ho y se acepta la H1

PROBLEMA 3.3. PRUEBA DE HIPÓTESIS 2 MUESTRAS PAREADAS

Excel. Prueba de Hipótesis. 2 muestras pareadas. Muestras dependientes.

La compañía Nickel Savings and Loan desea comparar las dos compañías que contrata para valuar las casas. Nickel Savins seleccionó una muestra de 10 propiedades y programa los avalúos de las dos empresas.

Los resultados en miles de dlls. Están en las columnas B y C.

Casa	**Schadek**	**Bowyer**
1	235	228
2	210	205
3	231	219
4	242	240
5	205	198
6	230	223
7	231	227
8	210	215
9	225	222
10	249	245

Con un nivel de significancia de 0.05 ¿se puede concluir que hay una diferencia significativa entre los avalúos medios de las casas de las dos empresas valuadoras?

Ho: $\mu_1 = \mu_2$ No hay diferencias entre los avalúos de las empresas valuadoras.

$\mu d = 0$

H_1: $\mu_1 \neq \mu_2$ Sí hay diferencias entre los avalúos de las empresas valuadoras.

dos colas

$\mu d \neq 0$

COMANDOS:

Datos – Análisis de datos – Seleccionar Prueba t para medias de dos muestras emparejadas – aceptar – Seleccionar rango para variable 1 (en este caso es Schadek) – Seleccionar datos para la variable 2 (en este caso es Bowyer) – Seleccionar diferencia hipotética entre las medias y marcar 0 (cero) – Seleccionar Alfa con 0.05 – Rango de salida – con el cursor seleccionar un espacio vacío donde estarán los cálculos – Aceptar – Listo

Prueba t para medias de dos muestras emparejadas

	Variable 1	*Variable 2*
Media	226.8	222.2
Varianza	208.844444	204.177778
Observaciones	10	10
Coeficiente de correlación de Pearson	0.95314381	
Diferencia hipotética de las medias	0	
Grados de libertad	9	
Estadístico t	3.30450068	
P(T<=t) una cola	0.00458195	
Valor crítico de t (una cola)	1.83311292	
P(T<=t) dos colas	0.0091639	
Valor crítico de t (dos colas)	2.26215716	

Conclusiones:

Dado que tcalc. = 3.3 > que tcrítica=2.26 se rechaza la Ho. es decir cae en la zona de rechazo por otro lado, P dos colas =0.009 < que el nivel de significancia de 0.05 , se rechaza la Ho Por lo que se acepta que sí hay diferencias entre los avalúos de las dos casas y Bowter tiene significativamente avalúos más bajos que Schadek.

Recordar la regla de decisión: *si tcal > tcrítica se rechaza la Ho.*

Si P < que nivel de significancia se rechaza la Ho.

Hay menor probabilidad de que la Ho sea cierta.

Minitab. Prueba de Hipótesis. 2 muestras pareadas.

PRUEBA DE HIPÓTESIS 2 MUESTRAS DEPENDIENTES

La compañía Nickel Savings and Loan desea comparar las dos compañías que contrata para valuar las casas. Nickel Savins selecciono una muestra de 10 propiedades y programa los avalúos de las dos empresas.

Los resultados en miles de dlls. Están en las columnas B y C. Con un nivel de significancia de 0.05 ¿se puede concluir que hay una diferencia significativa entre los avalúos medios de las casas de las dos empresas valuadoras?

Casa	**Schadek**	**Bowyer**
1	235	228
2	210	205
3	231	219
4	242	240
5	205	198
6	230	223
7	231	227
8	210	215
9	225	222
10	249	245

Al elaborar hoja de trabajo que sea en columnas. MINITAB NO ACEPTA FILAS.

COMANDOS:
PARA CARGAR HOJA DE TRABAJO:
En hoja de Excel elaborar la hoja de trabajo y guardar en escritorio. Abrir software minitab –Archivo – abrir hoja de trabajo – clic en escritorio – clic en nombre – clic seleccionar archivo – clic en tipo – clic en excel – Abrir – Listo, aparece cargada la hoja de trabajo.

COMANDOS PARA CÁLCULOS:
Clic estadística – Estadística básica-t pareada- clic en muestra 1 y cargar los datos (en este caso Schadek) con seleccionar – clic en muestra 2 y cargar los datos (en este caso Bowyer) selccionar – clic en opciones – clic nivel de confianza y anotar el deseado – clic diferencia hipotética y anotar 0.0 – clic hipótesis alterna – seleccionar la que se desea en este caso diferencia hipotética (H1) –Aceptar- Aceptar – Listo.

IC y Prueba T pareada: Schadek, Bowyer

```
T pareada para Schadek - Bowyer

                                    Error
                                    estándar
                                       de la
             N   Media  Desv.Est.     media
Schadek     10  226.80      14.45      4.57
Bowyer      10  222.20      14.29      4.52
Diferencia  10    4.60       4.40      1.39
IC de 95% para la diferencia media: (1.45, 7.75)
Prueba t de diferencia media = 0 (vs. ≠ 0):Valor T = 3.30 Valor p = 0.009
```

t crítica a 95% y 9 gl = 2.26

Conclusiones:

Dado que t calc. =3.3> que tcrítica= 2.26 se rechaza la Ho que dice que la diferencia entre ambas casas son iguales.

Por otro lado P=0.009 es < que el nivel de significancia de 0.05, se rechaza la Ho

Se concluye que sí existe una diferencia significativa entre los avalúos de las dos casas y que Bowyer tiene avalúos más bajos que schadek

PROBLEMA 3.4. PRUEBA DE HIPÓTESIS. 2 MUESTRAS. PERFUMES.

Minitab. PRUEBA DE PROPORCIONES DE DOS MUESTRAS.

La Compañia de perfumes Manelli desarrollo una fragancia nueva que planea comercializar con el nombre de Heavenly. Varios estudios de mercado indican que Heavely tiene buen potencial de mercado. El departamento de Manelli tiene interés si hay alguna diferencia entre las proporciones de mujeres jóvenes y mayores que comprarían el perfume si saliera al mercado. Hay dos poblaciones independientes una de mujeres jóvenes y otra de mayores. A cada una de las mujeres muestreadas, se le pedirá que huela el perfume e indique si le gusta el perfume como para comprarlo.

Ho: p1=p2 H1. p1 diferente de p2 nivel de significancia de 0.05

COMANDOS: CARGAR HOJA DE TRABAJO: Estadística – Estadística básica – 2 proporciones – Datos resumidos – No. de eventos: M1:19 – M2:62 – Ensayos: M1:100 – M2:200 – opciones: nivel de confianza: 95 – diferencia hipotética: 0 – Hipótesis alterna: diferente de Ho. – Método de prueba: utilice cálculo agrupado de la proporción – aceptar. aceptar.

Estadística básica dos proporciones

n1	n2	media 1	media 2	p1	p2
100	200	19	62	0.19	0.31

Muestra	X	N	Muestra p
1	19	100	0.190000
2	62	200	0.310000

```
Diferencia = p (1) - p (2)
Estimación de la diferencia: -0.12
IC de 95% para la diferencia: (-0.220102, -0.0198978)
Prueba para la diferencia = 0 vs. ≠ 0: Z = -2.21  Valor p = 0.027

Prueba exacta de Fisher: Valor p = 0.028
```

Z = 1.96 a 95% de nivel de confianza. (Porque n>30) Alfa: nivel de significancia = 0.05
Dado que Zcal=-2.21 es mayor que 1.96 se re rechaza la Ho. Es decir sí existe una diferencia significativa entre entre las mujeres jóvenes y las mujeres adultas por comprar la loción.

Por otro lado; el valor de p=0.027 es menor que el Nivel de significancia de 0.05, por lo que se rechaza la Ho. Corroborando la decisión de rechazar la Ho. Por lo que es muy probable que la loción en estudio sea más del gusto de mujeres adultas.

Este estudio puede servir para recomendar a que sector de las mujeres se puede dirigir la publicidad de un producto.

PROBLEMA 3.5. PRUEBA DE HIPÓTESIS. VIDES.

Minitab.

La Familia Daimon posee un viñedo grande a orillas del lago Erie, en Nueva York. Los viñedos deben fumigarse a enfermedades. Dos nuevos insecticidas acaban de salir al mercado: Pernod 5 yAction. (1 y 2 respectivamente). Para probar su eficacia, se seleccionaron 3 hileras y se fumigaron con Pernod 5 y otras 3 se fumigaron con Action.

Cuando las uvas maduraron, se revisaron 400 vides tratadas con Pernod 5 para saber si no estaban infectadas. De igual forma, se revisó 400 vides tratadas con Action. Los resultados son:

Insecticida	número	infectadas
Perdon 5	400	24
Action	400	40

Con un nivel de significancia de 0.05
¿Se puede concluir que existe una diferencia entre la proporción de vides Infectadas empleando Pernod 5 en comparación con las fumigadas con Actión?

COMANDOS: Cargar hoja de trabajo –_ Estadística – Estadística básica – 2 proporciones – Datos resumidos – No. de eventos M1:24-M2:40 – No. de ensayos:
M1: 400-M2:400_- Opciones: Nivel de confianza: 95%-Diferencia Ho. : 0 - H1: diferente de Ho. – Utilice cálculo por separado-aceptar – aceptar.

n1	n2	media 1	media 2	P1	P2
400	400	24	40	0.06	0.1

Prueba e IC para dos proporciones

```
Muestra   X    N  Muestra p
1        24  400   0.060000
2        40  400   0.100000


Diferencia = p (1) - p (2)
Estimación de la diferencia: -0.04
IC de 95% para la diferencia: (-0.0774963, -0.00250368)
Prueba para la diferencia = 0 vs. ≠ 0: Z = -2.09  Valor p = 0.037

Prueba exacta de Fisher: Valor p = 0.050
```

Z crítica a 95% =1.96

Conclusiones:
Dado que Z=1.96 es menor que Zcal=2.09 se rechaza la Ho. Cae en la zona de rechazo.
Dado que p=0.037 es menor que el nivel de significancia también se rechaza la Ho.
Por lo tanto se concluye que existe una diferencia significativa entre los dos insecticidas; y es probable que el Pernod 5es más efectivo que Action.

PROBLEMA 3.6. PRUEBA DE HIPÓTESIS. ENTRENAMIENTO.

Excel.

La publicidad que realiza Sylph Fitness Center afirma que, al terminar su entrenamiento, las personas bajaran de peso.
Una muestra aleatoria de 8 participantes recientes reveló los pesos siguientes antes y después de terminar el entrenamiento.
Con un nivel de significancia de 0.01. Se puede concluir que los participantes bajaran de peso?

Preguntas:

¿Cuál es el valor de t crítica?
¿Cuál es el valor de t calculada?
Hacer un análisis y dar su opinión basándose en la regla de decisión.

Nombre	Antes	Después	
Hunter	155	154	
Cashman	228	207	
Mervin	141	147	
Massa	162	157	
Creola	211	196	α= 0.01
Peterson	164	150	
Redding	184	170	
Poust	172	165	

Ho μd= 0 (no hay diferencia entre las medias del antes y después, la diferencia (d) es de cero)

H1 μd ≠0 (Que sí hay diferencias entre el antes y el después).

COMANDOS:

Datos-análisis de datos – prueba t para media de dos muestras emparejadas – aceptar – rango para la variable 1: marcar la columna "antes" – prueba para la variable 2: marcar la columna "después" – diferencia hipotética Ho = 0 - alfa=0.01 – Rango de salida: seleccionar la celda donde desea anotar los cálculos. Aceptar.

Prueba t para medias de dos muestras emparejadas

99%

	Variable 1	*Variable 2*
Media	177.125	168.25
Varianza	857.839286	485.642857
Observaciones	8	8
Coeficiente de correlación de Pearson	0.98110103	
Diferencia hipotética de las medias	0	
Grados de libertad	7	
Estadístico t	2.86100329	
P(T<=t) una cola	0.01215135	
Valor crítico de t (una cola)	2.99795157	
P(T<=t) dos colas	0.02430269	
Valor crítico de t (dos colas)	3.4994833	

Conclusiones:

El valor de t crítica de una cola a 99 % de N.C y de 7 gl es de 2.99

Para dos colas es de 3.499

T calculada es de 2.86 y de p=0.024

Dado que t calculada = 2.86 es menor que tcrítica =2.99 se acepta la Ho

Dado que p =0.024 es mayor que el N. de significancia = 0.01. Se acepta la Ho.

Por lo que se concluye al aceptar la Ho. que no existe diferencia entre el antes y el después del Entrenamiento físico.

Minitab.

La publicidad que realiza Sylph Fitness Center afirma que, al terminar su entrenamiento, las personas bajaran de peso. Una muestra aleatoria de 8 participantes recientes reveló los pesos siguientes antes y después de terminar el entrenamiento.

Con un nivel de significancia de 0.01. Se puede concluir que los participantes bajaran de peso?

Preguntas:

¿Cuál es el valor de t crítica?

¿Cuál es el valor de t calculada?

Hacer un análisis y dar su opinión basándose en la regla de decisión.

Nombre	Antes	Después	
Hunter	155	154	
Cashmn	228	207	
Mervin	141	147	
Massa	162	157	
Creola	211	196	α= 0.01
Petersn	164	150	
Redding	184	170	
Poust	172	165	

COMANDOS:

Cargar hoja de trabajo>Estadística>>Estadística básica>t pareada>checar que en el recuadro superior Éste en: cada muestra está en una columna>M1: cargar C1: antes>>m2: cargar C2 después>Opciones>NC: 95>Diferencia hipotética>>0.00>H1: diferencia (signo diferencia) diferencia hipotética>aceptar> aceptar. Listo.

Antes	Después
155	154
228	207
141	147
162	157
211	196
164	150
184	170
172	165

IC y Prueba T pareada: antes, después

```
T pareada para antes - después

                                   Error
                                estándar
                                   de la
            N  Media   Desv.Est.    media
Antes       8  177.1       29.3     10.4
Después     8  168.3       22.0      7.8
Diferencia  8   8.88       8.77     3.10


IC de 95% para la diferencia media: (1.54, 16.21)
Prueba t de diferencia media= 0 (vs. ≠ 0): Valor T= 2.86 Valor p= 0.024
```

La Ho se considera = 0 es decir que no hay diferencias entre el antes y el después. La H1 es que sí hay diferencias.

La t crítica de una cola es de 2.998 y de dos colas es 3.50 para el caso que sea. Con 7 g.l y 99 % de N. C.

Dado que Tcalculada es de 2.86 es menor que tcrítica, no se rechaza la Ho. Es decir, no existe diferencia significativa entre el antes y el después del programa, por lo que se considera que los estudiantes no bajaron de peso con el ejercicio.

PROBLEMA 3.7. PRUEBA DE HIPÓTESIS. PODADORAS.

Excel.
Poblaciones independientes.

welles	Atkins
2	3
4	7
9	5
3	8
2	4
	3

Considerar varianzas iguales.

Owens Lawn Care, Inc. Fabrica y ensambla podadoras de césped que envía a USA y a Canadá. Se han propuesto dos procedimientos distintos para el montaje del motor al chasis de la podadora.

La pregunta es: *¿existe una diferencia entre ellos con respecto al tiempo medio para montar los motores al chasis de la podadora?*, El primer procedimiento lo desarrolló el Sr Herb Welles, un antiguo empleado de Owens (designado como procedimiento 1), y el otro lo desarrollo el Sr William Atkins, Vicepresidente de ingeniería de Owens, (designado procedimiento 2).

Para evaluar los métodos, se decidió realizar un estudio de tiempos y movimientos. Se midió el tiempo de montaje en una muestra de cinco empleados según el método de Welles y seis con el método de Atkins.

Los resultados están en la columna A y B.

¿Hay alguna diferencia entre los tiempos medios de montaje?

Utilizar un Nivel de confianza del 95%

Ho: $\mu_1 = \mu_2$ que no hay diferencias entre las medias

H_1 : $\mu_1 \neq \mu_2$ que sí hay diferencias entre las medias

Estadístico de prueba es "t"

Regla de decisión: si Tcalc>tcrít se acepta la Ho Si p>α se acepta la Ho

COMANDOS:
Datos – análisis de datos – prueba t para dos muestras suponiendo varianzas iguales – aceptar – rango para la variable 1 (Welles) – rango para la variable 2 (Atkins) – diferencia hipotética Ho=0
alfa= 0.05 – rango de salida(seleccionar la celda donde desea anotar los cálculos – aceptar.

Prueba t para dos muestras suponiendo varianzas iguales

	Variable 1	*Variable 2*
Media	4	5
Varianza	8.5	4.4
Observaciones	5	6
Varianza agrupada	6.222222222	
Diferencia hipotética de las medias	0	
Grados de libertad	9	
Estadístico t	-0.662052112	
P(T<=t) una cola	0.262263176	
Valor crítico de t (una cola)	1.833112923	
P(T<=t) dos colas	0.524526351	
Valor crítico de t (dos colas)	2.262157158	

Conclusiones:

El valor de t crítica es de 2.26 a 95% de N.C. y 9 gl.

Se concluye que dado que t calculada es de -0.662 es menor que t crítica =2.26 por lo que se acepta la Ho.

También, dado que p=0.524 es mucho mayor que el nivel de significancia de 0.05, se acepta la Ho.

Por lo que se concluye que no existe diferencia significativa entre los tiempos medios de montaje de los motores al chasis de las podadoras en ambos procedimientos.

Minitab.

2 muestras. Suponiendo varianzas iguales. Ambas poblaciones son independientes.

welles	Atkins
2	3
4	7
9	5
3	8
2	4
3	

Owens Lawn Care, Inc. fabrica y ensambla podadoras de césped que envía a distribuidores instalados en Estados Unidos y Canadá. Se han propuesto dos procedimientos distintos para el montaje del motor al chasis de la podadora.

La pregunta es: *¿existe una diferencia entre ellos con respecto al tiempo medio para montar los motores al chasis de las podadoras?*

El primer procedimiento lo desarrolló el Sr. Herb Welles un antiguo empleado de El primer procedimiento lo desarrolló el Sr. Herb Welles un antiguo empleado de Owens (designado como procedimiento 1), y el otro lo desarrollo William Atkins, vicepresidente de ingeniería de Owens (designado como procedimiento 2). Para evaluar los dos métodos, se decidió realizar un estudio de tiempos y movimientos.

Se midió el tiempo de montaje en una muestra de 5 empleados según el método de Welles y 6 en el método de Atkins.
Los resultados, en minutos. Aparecen en la C1 y C2 respectivamente.

Utilizar un nivel de confianza del 95%

COMANDOS:

Cargar hoja de trabajo – Estadística – Estadística aplicada – t de dos muestras – cada muestra en su columna – M1: Welles – M2: Atkins – opciones – N.C:95% - Diferencia hipotética Ho = 0 – alternativa: diferente de la Ho – asumir varianzas iguales – aceptar – aceptar.

Prueba T e IC de dos muestras: welles, Atkins

```
T de dos muestras para welles vs. Atkins

                              Error
                            estándar
                               de la
        N  Media  Desv.Est.     media
welles  5   4.00       2.92       1.3
Atkins  6   5.00       2.10      0.86


Diferencia = μ (welles) - μ (Atkins)
Estimación de la diferencia:  -1.00
IC de 95% para la diferencia:  (-4.42, 2.42)
Prueba T de diferencia= 0 (vs. ≠): Valor T= -0.66  Valor p= 0.525 GL= 9
Ambos utilizan Desv.Est. agrupada = 2.4944


* NOTA * Se canceló el comando.
```

Con 9 g.l. (n1+n2)-2=9 y 95 % de N.C. la t crítica tiene un valor de 2.26
El nivel de significancia es del 5 %

Conclusiones:

Dado que tcalculada es de -0.66 < que tcrítica es de 2.26 no existe diferencia significativa Entre los dos procedimientos.

Es decir no hay diferencia entre el procedimiento 1 del procedimiento 2, es decir no se rechaza la Ho y dado que p=0.525 es mucho mayor que el nivel de significancia que es de 0.05, tampoco se rechaza la Ho.

PROBLEMA 3.8. PRUEBA DE HIPÓTESIS. TOALLAS.

Excel.

Con desviaciones estándar desiguales.

Toalla 1	Toalla 2
8	12
8	11
3	10
1	6
9	8
7	9
5	9
5	10
12	11
	9
	8
	10

El personal de un laboratorio de pruebas del consumidor evalúa la absorción de toallas de papel. Se desea compara un conjunto de toallas de una marca con un grupo similar de toallas de otra marca. De cada una de ellas se sumerge una pieza del papel en un tubo con un fluido, se deja que el papel escurra en una charola durante dos minutos y después se evalúa la cantidad de líquido que el papel absorbió.

Una muestra aleatoria de 9 toallas del papel de la primera marca absorbió las cantidades siguientes de líquido milímetros:

8 8 3 1 9 7 5 5 12

Una muestra aleatoria independiente de 12 toallas de la otra marca absorbió las cantidades siguientes de líquido en milímetro:

12 11 10 6 8 9 9 10 11 9 9 8 10

Utilizar un nivel de confianza del 90 %

COMANDOS:

Datos – análisis de datos – Prueba t para dos muestras suponiendo varianzas desiguales – aceptar – rango para la variable 1 (toallas 1) – rango para la variable 2 (toallas 2) diferencia hipotética Ho =0 – alfa = 0.10) – rango de salida seleccionar la celda donde desea anotar los cálculos aceptar.

diferencia hipotética= 0

Prueba t para dos muestras suponiendo varianzas desiguales

	Variable 1	*Variable 2*
Media	6.44444444	9.416666667
Varianza	11.0277778	2.628787879
Observaciones	9	12
Diferencia hipotética de las medias	0	
Grados de libertad	11	
Estadístico t	-2.47309843	
P(T<=t) una cola	0.01547536	
Valor crítico de t (una cola)	1.36343032	
P(T<=t) dos colas	0.03095072	
Valor crítico de t (dos colas)	1.79588481	

S1=3.32 t CRÍTICA A .90 Y 10 gl (dos colas)=1.81

S2=1.621 t CRÍTICA A .90 Y 11 gl (dos colas)=1.80

NOTA: en el cálculo de los grados de libertad se redondea hacia abajo a un entero para que nos dé un valor más alto de t crítica. La máquina redondea hacia arriba.

Conclusiones:

Dado que tcal.es = -2.47 es mucho mayor que t crítica=1.81, se rechaza la Ho. de igual manera p = 0.0309 < alfa=0.10 se rechaza la Ho.

Es decir, sí existe diferencia significativa entre la tasa de absorción de las toallas de papel por lo que se recomienda que la marca 2 retiene mayor cantidad de fluido que la marca 1.

Minitab.

Toalla 1	Toalla 2
8	12
8	11
3	10
1	6
9	8
7	9
5	9
5	10
12	11
	9
	8
	10

Estadístico de prueba de medias sin diferencia, varianzas desiguales.

El personal de un laboratorio de pruebas del consumidor evalúa la absorción de y toallas de papel.

Se desea comparar un conjunto de toallas de una marca con un grupo similar de toallas de otra marca. De cada una de ellas se sumerge una pieza de papel en un tubo con un fluido, se deja que el papel escurra en una charola durante dos minutos y después se evalúa la cantidad de líquido que el papel absorbió de la charola. Una muestra aleatoria de 9 toallas de papel de la primera marca absorbió las cantidades de líquido en mm (C1). Una muestra independiente y de manera aleatoria de 12 toallas de la segunda marca absorbió las cantidades medias de líquido en mm (C2).

Utilizando el nivel de significancia de 0.10, probar si existe una diferencia significativa entre las cantidades medias de líquidos que absorbieron los dos tipos de toallas.

Consideraciones:

Tienen una dist. Normal y no se conocen las Desv. Stad. Por lo que se usa el estadístico de prueba "t". La cantidad de absorción varia de la 1a. marca de 1 a 12 mm, mientras que la 2a. marca varía de 6 a 12 mm, es decir existe mayor variabilidad en la marca 1; por lo que se supone que las Desv. Std. no son iguales.

Ho: U1=U2 y H1: U1 diferente de U2

COMANDOS:

Cargar hoja de trabajo – estadística – estadística aplicada – t de dos muestras – cada una esta en su columna – muestra 1 (C1) – muestra 2 (C2) – opciones N.C. Diferencia hipotética Ho =0 – diferentes de Ho – aceptar – aceptar

Prueba T e IC de dos muestras: toalla 1, toalla 2

```
T de dos muestras para toalla 1 vs. toalla 2

                                 Error
                               estándar
                                  de la
           N  Media  Desv.Est.    media
toalla 1   9   6.44       3.32      1.1
toalla 2  12   9.42       1.62     0.47


Diferencia = μ (toalla 1) - μ (toalla 2)
Estimación de la diferencia:  -2.97
IC de 90% para la diferencia:  (-5.15, -0.79)
Prueba T de diferencia= 0 (vs. ≠): Valor T= -2.47  Valor p= 0.033 GL= 10
```

Se considera que las varianzas son desiguales.

Conclusiones:

Dado que t=-2.47 cae en la zona de rechazo de la Ho. Por lo que se rechaza ésta.

De igual manera p=0.033 es mucho menor que alfa=0.10 por lo que se concluye al rechazar la Ho que sí existe diferencia significativa en la capacidad de absorción de fluido entre las dos marcas de toallas, siendo la de mayor absorción la marca 2.

NOTA: en este software si redondea hacia abajo el valor de los grados de libertad calculados.

UNIDAD IV.

PRUEBAS DE BONDAD DE AJUSTE.

PROBLEMA 4.1 CHI- CUADRADA. ADAPTACIÓN

Minitab.

La Federal Correction Agency investiga: ¿un hombre liberado de una prisión Federal se adapta de manera diferente a la vida civil si regresa a su ciudad natal o si va a vivir a otra parte?. En otras palabras, ¿hay una relación entre la adaptación a la vida civil y el lugar de residencia después de salir de prisión?

Utilizar el nivel de significancia de 0.01

Se suprime la Hoja de trabajo.

Esta es una prueba de Chi-cuadrada

COMANDOS:

Abrir Minitab – cargar hoja de trabajo – Prueba Chi.cuadrada para asociación –Datos resumidos en una tabla de dos factores – columnas que contiene la tabla: seleccionar: sobresaliente, buena, regular e insatisfactorio – clic en estadística – prueba Chi-cuadrada – aceptar. Listo.

Residencia	Sobresaliente	Buena	Regular	Insatisfactorio
Cd natal	27	35	33	25
No cd natal	13	15	27	25

	Sobresaliente	Buena	Regular	Insatisfactorio	Total
1	27	35	33	25	120
	24.00	30.00	36.00	30.00	
	0.375	0.833	0.250	0.833	
2	13	15	27	25	80
	16.00	20.00	24.00	20.00	
	0.563	1.250	0.375	1.250	
Total	40	50	60	50	200

Chi-cuadrada = 5.729, GL = 3, Valor P = 0.126

Conclusiones:

Ho: No hay relación entre la adaptación a la vida civil y el lugar donde se radique el individuo después de salir de prisión.

H1: Hay relación entre la adaptación a la vida civil y el lugar donde se radique el individuo después de salir de prisión.

Regla de Decisión:
Rechazar Ho si el valor de Chi-cuadrada calculada es mayor que Chi-cuadrada crítica y si p<alfa=0.01 se rechaza la Ho.

Chi-cuadrada crítica a 3 gl y un nivel de significancia de 0.01 es de 11.345 (en tablas)

Dado que Chi-cuadrada = 5.729 es menor que Chi-cuadrada crítica = 11.345 no se rechaza la Ho, caen en la zona de aceptación de Ho.

Es decir, no hay evidencia de una relación entre la adaptación a la vida civil y el lugar de residencia del individuo al salir de prisión.

Por otro lado P = 0.126 es mayor que Alfa =0.01 por lo que también se acepta la Ho.

PROBLEMA 4.2. CHI- CUADRADA. JUGADORES.

Minitab.

Un científico social tomo una muestra de 140 personas y las clasificó de acuerdo con su nivel de ingreso, y si jugaron o no en la lotería nacional mes pasado. La información de la muestra aparece a continuación. ¿Es posible concluir que jugar a la lotería se relaciona con el nivel de ingreso?

Utilizar un nivel de significancia de 0.05

Se suprime la Hoja de Trabajo.

Participación	I. bajo	I. medio	I. alto
Jugaron	46	28	21
No jugaron	14	12	19

COMANDOS: Abrir Minitab – cargar hoja de trabajo – estadística – tablas – prueba Chi-cuadrada por asociación – datos resumidos en una tabla de dos factores – columnas que contiene la tabla – seleccionar:

Ingreso bajo, Ingreso medio, e Ingreso alto-clic en estadística- prueba Chi-cuadrada. Aceptar – aceptar.

Prueba Chi-cuadrada para asociación: Filas de la hoja de trab, Columnas de la hoja de t

```
Filas: Filas de la hoja de trabajo Columnas: Columnas de la hoja de trabajo


        I. bajpo  I. Medio  I. alto  Todo

1             46        28       21    95
           40.71     27.14    27.14

2             14        12       19    45
           19.29     12.86    12.86
```

```
Todo          60        40       40   140

Contenido de la celda:      Conteo
                            Conteo esperado


Chi-cuadrada de Pearson = 6.544, GL = 2, Valor p = 0.038
Chi-cuadrada de la tasa de verosimilitud  = 6.410, DF = 2, Valor p = 0.041
```

Considerar: Ho = No hay relación entre jugar y el ingreso.
H1= sí hay relación entre jugar y el engreso.
Chi-cuadrada crítica con dos gl y alfa = 0.05 tiene un valor de = 5.991

Conclusiones:

Dado que chi-cuadrada calculada = 6.544 es mayor que chi-cuadrada crítica = 5.991 se rechaza la Ho.
Por otro lado p=0.041 es menor que alfa=0.05 tambien se rechaza la Ho, por lo que se concluye que sí hay relación entre el nievel de ingreso y el jugar a la loteria, al ser rechazada la Ho y aceptada la H1

PROBLEMA 4.3. CHI-CUADRADA. KRUSKALS-WALLIS.NO PARAMÉTRICO.

Minitab.

El Hospital System of the Carolinas opera tres hospitales en el área de Great Charlote: St. Luke´s Memorial, en el lado poniente de la ciudad, Swedish Medical Center al Sur, y el Piedmon Hospital en el lado Este. El director de administración está preocupado acerca del tiempo de espera de los pacientes con lesiones de tipo deportivo, que no ponen en peligro la vida, y que llegan durante las tardes entre semana a los tres hospitales. Específicamente, ¿existe una diferencia en los tiempos de espera en los tres hospitales?

Para averiguarlo, el director seleccionó una muestra aleatoria de pacientes en los tres hospitales y determinó el tiempo, en minutos, en que se entra a un hospital y el momento en que termina el tratamiento. Los tiempos en minutos están en la columna A y B.

Score	Group
56	St. Luke´s
39	St. Luke´s
48	St. Luke´s
38	St. Luke´s
73	St. Luke´s
60	St. Luke´s
62	St. Luke´s
103	Swedish
87	Swedish
51	Swedish

```
I. bajpo  I. Medio  I. alto  Todo

1             46        28       21    95
           40.71     27.14    27.14

2             14        12       19    45
           19.29     12.86    12.86

Todo          60        40       40   140

Contenido de la celda:      Conteo
                            Conteo esperado


Chi-cuadrada de Pearson = 6.544, GL = 2, Valor p = 0.038
Chi-cuadrada de la tasa de verosimilitud  = 6.410, DF = 2, Valor p = 0.041
```

Considerar: Ho = No hay relación entre jugar y el ingreso.
H1= sí hay relación entre jugar y el ingreso.

Chi-cuadrada crítica con 2 gl y alfa = 0.05 tiene un valor de = 5.991

Conclusiones:
Dado que Chi-cuadrada calculada = 6.544 es mayor que Chi-cuadrada crítica = 5.991 se rechaza la Ho.

Por otro lado p=0.041 es menor que alfa=0.05 también se rechaza la Ho, por lo que se concluye que sí hay relación entre nivel de ingreso y el jugar a la lotería, al ser rechazada la Ho y aceptada la H1

COMANDOS: Abrir Minitab – cargar hoja de trabajo – estadística – No paramétricos – Kruskal_wallis – respuesta: Seleccionar C1 (Score) – C2 (Group) – aceptar. Listo

Prueba de Kruskal-Wallis: Score vs. Group

```
Prueba de Kruskal-Wallis en Score

                          Clasificación
Group        N  Mediana   del promedio      Z
Piedmont     6    51.50            8.8  -1.05
St. Luke´s   7    56.00            8.4  -1.38
Swedish      8    88.00           15.0   2.32
General     21                    11.0

H = 5.38  GL = 2  P = 0.068
H = 5.39  GL = 2  P = 0.067  (ajustados para los vínculos)
```

CONSIDERACIONES:

Kruskal-Wallis no requiere que la muestra provenga de una población normal (si no se usaría la prueba de ANOVA)

La prueba es de Chi-cuadrada pero se usa la técnica de Kruskal-Wallis con la sigla "H"

Ho= Las distribuciones de las poblaciones de los tiempos de espera es igual para los tres hospitales.

H1=No todas las distribuciones de los poblaciones son iguales.

Chi-cuadrada crítica con 2 gl y un nivel de significancia de 0.05 es de 5.99

Conclusiones:

Como el valor calculado de (H) es de 5.38 es menor que el valor de Chi-cuadrada = 5.99, no se rechaza la Ho.

PROBLEMA 4.4. CHI-CUADRADA. NO PARAMÉTRICO. BANCOS.

Minitab.

Score	Group
208	S. Englewood
307	S. Englewood
199	S. Englewood
142	S. Englewood
91	S. Englewood
296	S. Englewood
91	S. West Side
62	S. West Side
86	S. West Side
91	S. West Side
80	S. West Side
302	S. Great Northen
103	S. Great Northen
319	S. Great Northen
340	S. Great Northen
180	S. Great Northen
99	S. Sylvania
116	S. Sylvania
189	S. Sylvania
103	S. Sylvania
100	S. Sylvania
131	S. Sylvania

El gerente del banco regional Statewide Financial Bank tiene interés en el índice de movimientos de dinero de las cuentas de cheques personales en cuatro sucursales.

(El índice de movimientos es la velocidad a la que el dinero en una cuenta se deposita y se retira; por ejemplo: si una cuenta tiene un índice de 300 se dice que es una cuenta extremadamente activa, pero si solo tiene uno o dos cheques en índice puede ser de 30

Los índices de rotación de las muestras seleccionadas de las cuatro sucursales bancarias aparecen ya ordenadas en las columnas A y B

Con un nivel de significancia de = 0.01 y la prueba de Kruskal-Wallis, determinar si hay una diferencia significativa entre los índices de rotación de las cuentas de cheques personales de las cuatro sucursales.

Se suprime la Hoja de Trabajo

COMANDOS:

Abrir Minitab – cargar hoja de cálculo – Estadística – No paramétricos – Kruskas-Wallis – respuesta: seleccionar C1 y C2 – aceptar, listo.

Prueba de Kruskal-Wallis: Score vs. Group

```
Prueba de Kruskal-Wallis en Score

                              Clasificación
Group              N  Mediana   del promedio      Z
S. Englewood       6   203.50          14.8   1.47
S. Great Northen   5   302.00          17.1   2.19
S. Sylvania        6   109.50          10.4  -0.48
S. West Side       5    86.00           3.2  -3.25
General           22                   11.5

H = 13.64  GL = 3  P = 0.003
H = 13.67  GL = 3  P = 0.003  (ajustados para los vínculos)
```

Ho= Las distribuciones de las poblaciones son idénticas

H1=Las distribuciones de las poblaciones no son idénticas.

El valor de Chi-cuadrada crítica K-1=4-1=3 gl y 99% de N.C. es de 11.345

Dado que H=13.64 es mayor que Chi-cuadrada crítica =11.345 se rechaza la Ho al igual P=0.003 <alfa=0.01 se rechaza la Ho. Por lo que se concluye que los movimientos de las cuentas de cheques no son iguales en las 4 sucursales.

ENFOQUE GRÁFICO DE NORMALIDAD.

Enfoques gráficos y estadísticos para confirmar la normalidad de una población.

Una desventaja de la prueba de bondad de ajuste de la normalidad es que se compara una frecuencia de distribución de un conjunto de datos con un grupo esperado de frecuencias de distribución normal. Cuando se organizan los datos de distribución de frecuencias, se sabe que se pierde información con respecto a esos datos. Esto es, no se tienen los datos crudos. Existen varias pruebas en que se usan datos crudos en vez de utilizar datos agrupados en una distribución de frecuencias. Estas pruebas incluyen las pruebas de normalidad Kolmogorob-Smirinof, Lilliefors y Anderson – Darling. Para complementar estas pruebas estadísticas, se dispone de métodos gráficos para tener un acceso visual a la normalidad de una distribución.

Se utilizan valores "p" para evaluar la hipótesis de normalidad.

Nos enfocaremos en la prueba de normalidad Anderson – Darling, que se basa en dos pasos:

1. -Se crean dos distribuciones acumulativas. La primera es una distribución acumulativa de datos crudos y, la segunda, es una distribución acumulativa normal.

2. -Se comparan las dos distribuciones acumulativas para determinar la mayor diferencia numérica absoluta entre ambas. Utilizando una prueba estadística, si la diferencia es amplia, se rechaza la hipótesis nula de que los datos siguen una distribución normal.

Además, se puede graficar la distribución acumulada de los datos crudos y la distribución acumulativa normal. La gráfica de una distribución acumulativa normal es una línea recta. La gráfica de los datos crudos estará diseminada alrededor de la recta que representa la acumulativa normal. Mediante la gráfica, se puede observar que los datos están normalmente distribuidos si la diseminación está relativamente cerca de la línea recta que representa la distribución acumulativa normal.

PROBLEMA 4.5. NORMALIDAD. AUTOS.

ENFOQUE GRÁFICO Y ESTADÍSTICO PARA CONFIRMAR LA NORMALIDAD.

1387	1761	1115	2070	2063
1754	1915	1124	2454	2083
1817	2119	1532	1606	2856
1040	1766	1688	1680	2989
1273	2201	1822	1827	910
1529	996	1897	1915	1536
3082	2813	2445	2084	1957
1951	323	2886	2639	2240
2692	352	820	842	2695
1206	482	1266	1963	1325
1342	1144	1741	2059	2250
443	1485	1772	2338	2279
754	1509	1932	3043	2626
1621	1638	2350	1059	1501
870	1961	2422	1674	1752
1174	2127	2446	1807	2058
1412	2430	369	2056	2370
1809	1704	978	2236	2637
2415	1876	1238	2928	1426
1546	2010	1818	1269	2944
2148	2165	1824	1717	2147
2207	2231	1907	1797	1973
2252	2389	1938	1955	2502
1428	335	1940	2199	783
1889	963	2197	2482	1538
1166	1298	2646	2701	2339
1320	1410	1461	3210	2700
2265	1553	1731	377	2222
1323	1648	2230	1220	2597
1761	2071	2341	1401	2742
1919	2116	3292	2175	1837
2357	1500	1108	1118	2842
2866	1549	1295	2584	2434
732	2348	1344	2666	1640
1464	2498	1906	2991	1821
1626	294	1952	934	2487

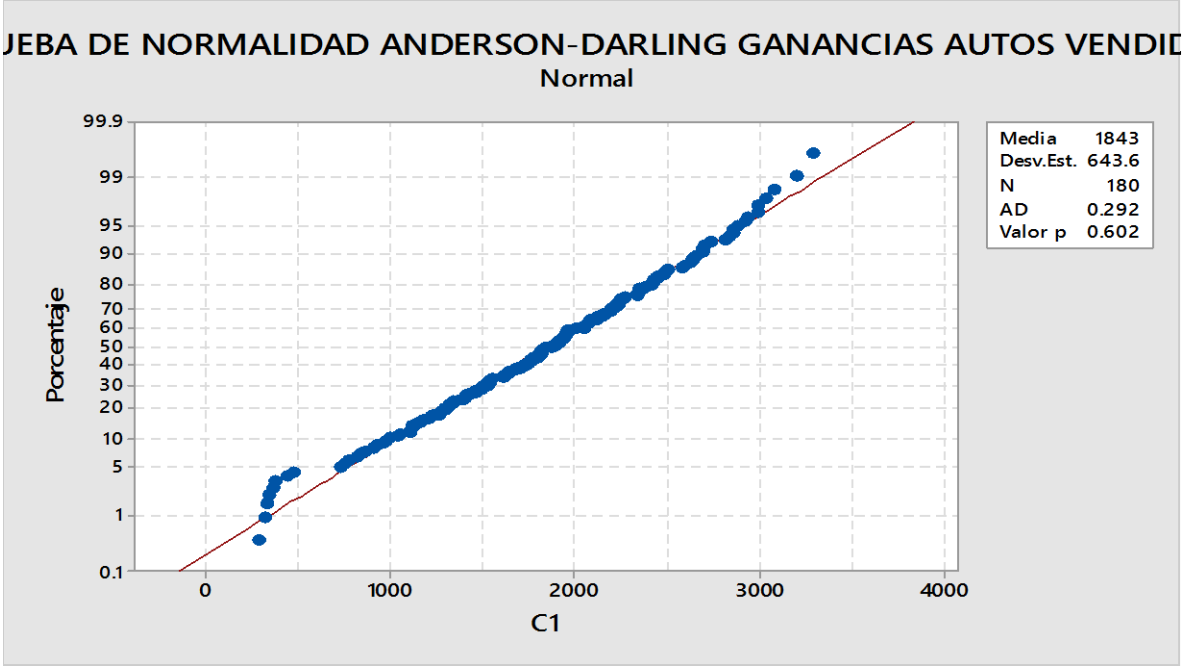
Minitab

En la columna A se tienen las ganancias de una Empresa que vende autos usados y desea saber si las ventas de cada mes tienen un comportamiento normal.

La muestra es las utilidades de 180 autos vendidos con diferentes precios durante un mes.

COMANDOS:

Abrir minitab – cargar hoja de trabajo – estadística – prueba de trabajo –prueba de normalidad – variable cargar C1 (columna de datos) – seleccionar: Anderson -Darling –ponerle título – aceptar.



Ho = la distribución es normal. Que la diferencia entre la distribución normal y los datos crudos no es significativa.

NOTAS: La línea roja es la distribución acumulativa normal, los puntos azules es la distribución acumulativa de datos crudos.

En el recuadro superior derecho hay 5 datos:

- Media: es la media de todos los datos
- Desv. Est. es la desviación estándar.
- N es el número de datos crudos.
- AD es el estadístico de prueba de Anderson - Darling,
- Valor p es el valor de cada estadístico de prueba que se utiliza para tomar una decisión con respecto de la Ho y se contrasta con el nivel de significancia (alfa). Como el valor de p=0.602 es mayor que alfa =0.05 por lo que no se rechaza la Ho.

PROBLEMA 4.6. NORMALIDAD MEDICAMENTOS.

ENFOQUES GRÁFICOS Y ESTADÍSTICOS PARA CONFIRMAR LA NORMALIDAD ANDERSON - DARLING.

Minitab.

9.2

8.7

8.9

8.6

8.8

8.5

8.7

9

Una máquina se calibra para llenar una pequeña botella con 9.0 gramos de medicamento. Una muestra de 8 botellas reveló las siguientes cantidades (en gramos) en cada botella. Se realizó una prueba de hipótesis con respecto a la media.

Para hacer la prueba, la suposición fue que los datos muestrales seguían una distribución normal. Los datos están en la columna A

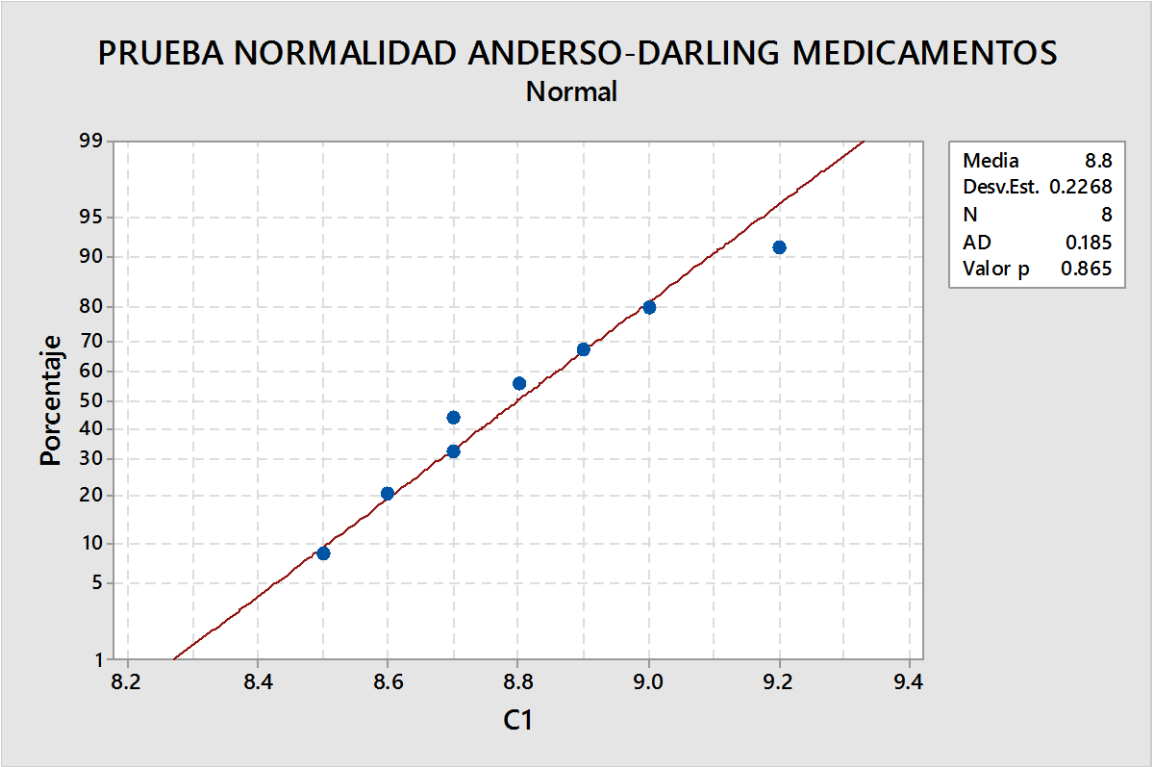
¿Es razonable suponer que los datos tienen un comportamiento normal?

Utilizar un nivel de significancia de 0.01

Explicar.

COMANDOS.

Abrir minitab – cargar hoja de trabajo – estadística – Prueba de normalidad – cargar datos (C1) – seleccionar Anderson – Darling – poner título a la gráfica – aceptar.



Ho = no hay diferencias entre las distribuciones, que son normales.

Las gráficas tienen una coincidencia del 86.5 % (p=0.865) que al contrastarse con el nivel de significancia que es de 5% (alfa=0.05) no se puede rechazar la Ho de que la distribución es normal.

UNIDAD V.

REGRESIÓN LINEAL SIMPLE Y MÚLTIPLE.

INTRODUCCIÓN A LA REGRESIÓN LINEAL.

BREVE APUNTE DE REGRESIÓN LINEAL.

REGRESIÓN LINEAL SIMPLE Y COMPUESTA

REGRESION LINEAL SIMPLE.

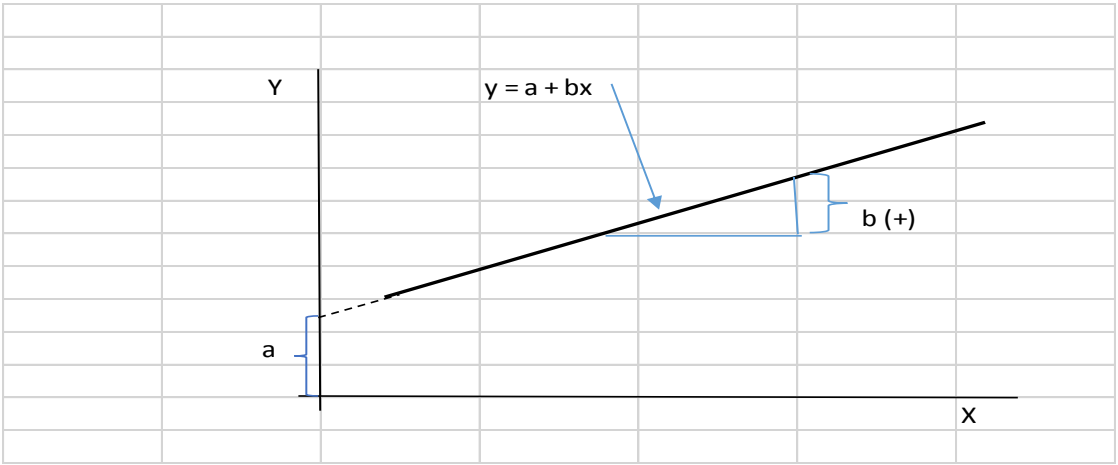
En la regresión lineal se trata de predecir los valores de una variable, cuando se conoce el valor de la otra. Por ejemplo, las calificaciones finales del curso se pueden predecir a partir de las puntuaciones según un test de capacidad mental. En este caso concreto las puntuaciones según el test de aptitud son las que predicen y constituyen los valores de **la variable independiente**. Lo que se predice, es decir, las calificaciones finales, es la **variable dependiente o función.** Mediante una gráfica o la ecuación de la regresión se puede inferir los valores de la variable dependiente a partir de los que adquiera la variable independiente. Es casi una norma emplear la letra X para la variable independiente y Y para la variable dependiente.

Recordemos los conocimientos sobre la recta. Donde su ecuación es:

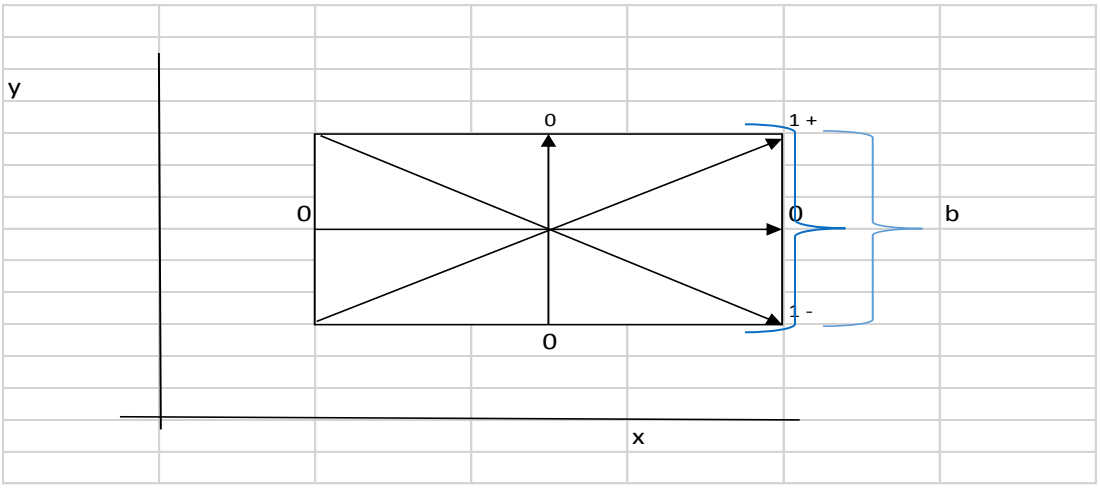
$$Y = a + b\,X$$

De donde Y como ya dijimos, es la variable dependiente de los valores que adquiera X que es la variable independiente. "a" es la ordenada al origen y "b" la pendiente de la recta.

Veamos una gráfica.



Recordemos que la pendiente puede tener diferente signo. Veamos:



Si la pendiente b tiene un valor de 1 (+) equivale a una correlación de Y con X del 100 %, es una correlación funcional. Por el contrario si b es (-1) también tiene una correlación del 100% pero es inversamente proporcional; es decir, en la medida que los valores de X aumentan, los valores de Y disminuyen. Cuando b tiene un valor de cero quiere decir que no existe correlación entre Y y X (la recta es totalmente horizontal o vertical).

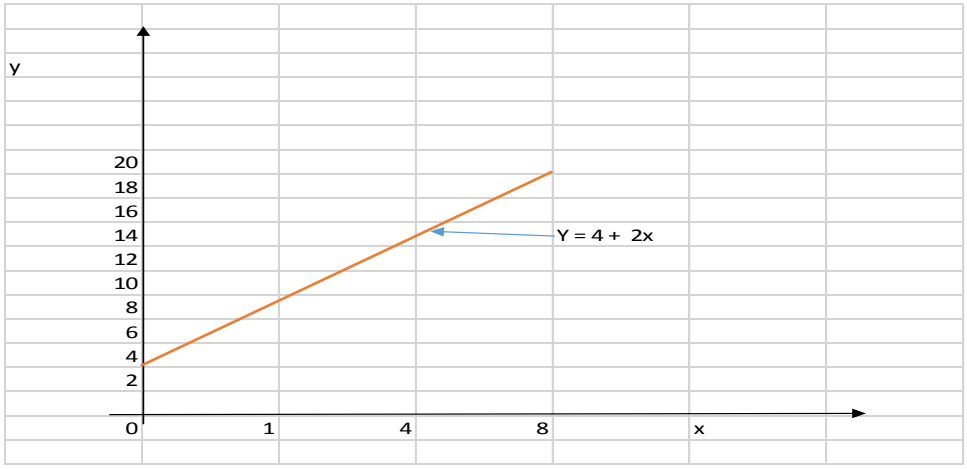
Por ejemplo, si la ecuación es $Y = 4 + 2X$ o sea

Si

X	Y
0	4
1	6
4	12
8	20

2 es valor de la pendiente y 4 es la ordenada al origen. Con estos datos se puede construir la recta. Esta es una correlación funcional.

Esto quiere decir que por cada unidad que aumente X, Y aumentara 2. (pendiente)

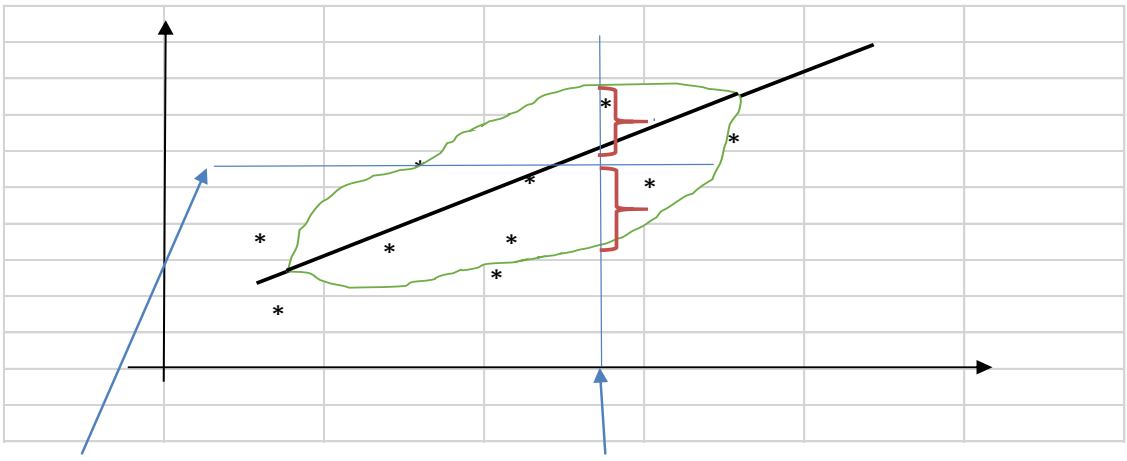


Ahora veamos el término **"correlación".** El concepto es la relación entre dos variables, en este caso, la variable Y, y la variable X. El Coeficiente de correlación llamado también de Pearson que lo designaremos como Υ, es una medida de la relación entre las variables X y Y.

Cuando la correlación es directa como el caso de la ecuación Y = 4 + 2 X, se dice que hay una relación funcional, pero no siempre es así, es decir, puede haber algunos puntos que no son coincidentes con una recta, están dispersos en el cuadrante, forman una "nube" en torno a ella; veamos dos condiciones e interpretación:

1.- Observar si la nube forma una línea en promedio para comprobar si hay o no correlación. En la medida que unos datos se alejan de una recta, el valor calculado del coeficiente de correlación de Pearson es más pequeño. La cuantía de Υ refleja la varianza (medida de dispersión) estimada mediante una recta sobre si los datos son esencialmente lineales o no. Una descripción de la regresión lineal es que la media de las X ($\dot{X}$), y la media de las Y ($\bar{Y}$) se encuentran en una recta. Es decir, las medias de las X para una misma media de Y. Sin embargo, si una relación es curvilínea, se utiliza un coeficiente de correlación llamado "eta", que veremos más adelante.

2.- HOMOCEDASTICIDAD: quiere decir si las desviaciones estándar (o la varianza) de los valores X y Y tienden a ser iguales. (Las distancias entre las llaves de la siguiente figura). Si son iguales existe homocedasticidad y sí hay correlación.



Veamos un ejercicio del cálculo del coeficiente de correlación de Pearson:

X	Y	x	y	Zx	Zy	ZxZy	x^2	y^2	xy	
20	12	* 7	* 2	1.61	0.54	0.8694	49	4	14	
18	16	5	6	1.15	1.62	1.8637	25	36	30	
16	10	3	0	0.69	0	0	9	0	0	
15	14	2	4	0.46	1.08	0.4968	4	16	8	
14	12	1	2	0.23	0.54	0.1242	1	4	2	
12	10	-1	0	-0.23	0	0	1	0	0	
12	9	-1	-1	-0.023	-0.27	0.0621	1	1	1	
10	8	-3	-2	-0.69	-0.54	0.3726	9	4	6	
8	7	-5	-3	-1.15	-0.81	0.9315	25	9	15	
5	2	-8	-8	-1.84	-2.16	3.9744	64	64	64	
130	100	σx=4.34	σy= 3.71			8.6947	188	138	140	
$\dot{X}$=13	$\bar{Y}$=10									

*** x = 20-13=7 * y = 12-10=2 (se codifica: a X se le resta 13 y a Y se le resta 10 para minimizar la cantidad de dígitos)**

$r = \frac{\sum ZxZy}{N} = \frac{8.6947}{10} = 0.87$ es decir, existe una correlación entre X y Y del 0.87 o sea 87%

La recta que aparece en el estudio de regresión suele escribirse en una forma análoga a la anterior:

$Y' = a + bX$ y se lee a Y´ como valor previsto de Y. El valor de Y´, no suele ser el mismo que el de Y, puesto que la puntuación que se predice a partir de dicha ecuación no es, en general, igual a la que se obtendría en realidad. Los valores de Y´ están, normalmente, más próximos a la media de Y ($\bar{Y}$) que los valores observados de Y. Por esta razón el fenómeno se conoce con el nombre de REGRESIÓN.

CÁLCULO DE LOS COEFICIENTES "a" Y "b"

La diferencia entre la puntuación obtenida de Y y la prevista de Y´, se llama *error de predicción.* **La recta de regresión o línea de ajuste óptimo, se suele definir como aquella recta respecto de la cuál, la suma de los cuadrados de los errores de predicción es mínima (método de los mínimos cuadrados).**

Para obtener los valores de **a** y de **b**, que hacen mínima la suma de los cuadrados de los errores de predicción, hay que derivar la expresión de la ecuación de la recta con Y´, respecto de **a** y de **b** e igualar a cero cada una de dichas derivadas.se obtiene un sistema de dos ecuaciones del que se deducen los citados coeficientes:

$$byx = \frac{\sum XY - [(\sum X)(\sum Y)/N]}{\sum X^2 - [(\sum X)^2/N]}$$

ayx = $\bar{Y}$ - byx (X) **Y´=a + byx(X)**

Que son los coeficientes de regresión de y sobre x. Lo que permite predecir los valores de y dado los de x.
Veamos un problema pero resuelto con Excel por lo laborioso del cálculo. En primer lugar veremos un problema resuelto con software de correlación con los comandos correspondientes y uno de regresión lineal simple. (Ir al software)

ERROR TÍPICO DE ESTIMACIÓN:

Como vemos, en general, las puntuaciones Y observadas no coinciden con las puntuaciones previstas de y (Y´). Existe un error en toda predicción y su cuantía se halla por medio de un estadístico conocido con el nombre de "Error típico de estimación" o error estándar de estimación, que para este caso, aparece en el cálculo de Excel en el primer recuadro del problema de alcohol en la sangra VS cervezas consumidas y que tiene un valor de 0.017726.

Si el coeficiente de correlación es grande (cercano a 1.00), el error típico de estimación es pequeño y recíprocamente. En el caso de una relación perfecta entre X y Y (correlación funcional), la Y observada es, exactamente, igual a la prevista Y´. Esto significa que todos los valores de Y ´pertenecen a la recta de regresión. En estas condiciones no existe desviación respecto de la recta para los diversos valores correspondientes a X, no se comete error alguno al hacer las estimaciones.

REGRESIÓN LINEAL MÚLTIPLE.

En la correlación y regresión lineal múltiple, emplean variables independientes adicionales denotadas por X1, X2,…… Xn que ayudan a explicar o predecir mejor la variable dependiente Y. Sin embargo, las variables independientes adicionales permiten hacer algunas consideraciones nuevas. El análisis de regresión múltiple sirve como técnica descriptiva o como técnica de inferencia.

La forma descriptiva general de una ecuación lineal múltiple es la siguiente:

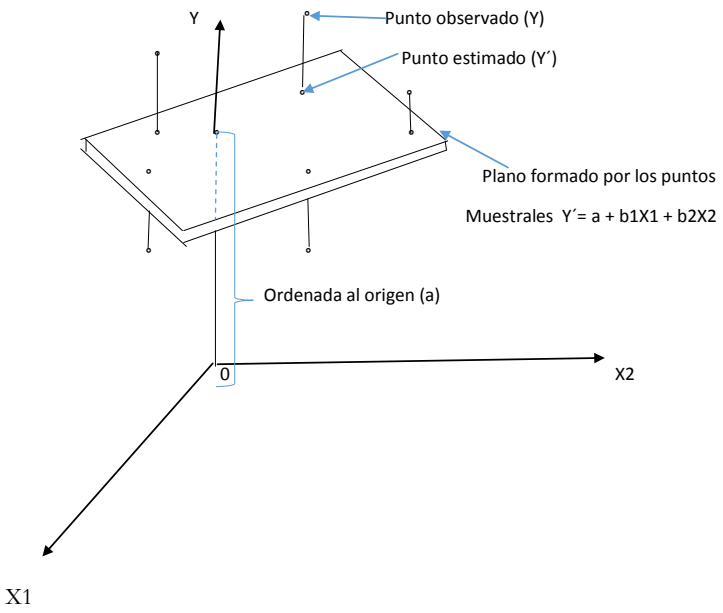
Y´= a + b1X1 + b2X2 + ----------+bkXk

De donde:

a es la intersección, b es la pendiente de cada variable independiente X, que es la cantidad en que Y cambia para cada X.

Cuando hay dos variables independientes la ecuación es: Y´= a + b1X1 + b2X2

Cuando hay dos variables independientes, esta relación se presenta de forma gráfica como un plataforma, como la siguiente figura:



Si un análisis de regresión múltiple incluye más de dos variables independientes, **no se puede emplear una gráfica para ilustrar el análisis, ya que las gráficas están limitadas a tres dimensiones.** Para ilustrar la interpretación de la intersección (a) y los dos coeficientes de regresión, supóngase que el rendimiento por galón de combustible de un vehículo tiene una relación directa con el octanaje de la gasolina (X1) y una inversa con el peso del automóvil (X2),. Suponga que la ecuación de regresión, calculada con software estadístico es:

Y´= 6.3 + 0.2 X1 - 0.001 X2 X1 es el octanaje de la gasolina y X2 es el peso del auto.

6.3 es la intersección (a) que es valor de Y´ cuando X1 y X2 tienen un valor de cero. Es importante tener en cuenta que en general, una ecuación de regresión no se utiliza fuera del rango de los valores muestrales.

El valor de b1 es 0.2 indica, que por cada aumento de 1 en el contenido de octanos en la gasolina el automóvil recorrería 2/10 de una milla por galón o sea 0,2 millas, sin importar el peso del auto. El valor b2 de - 0.001 revela que, por cada aumento de una libra en el peso del vehículo, el número de millas recorrido por galón **disminuye** en 0.001, sin importar el contenido de octanos de la gasolina.

Por ejemplo, un automóvil con gasolina de 92 octanos en el depósito de combustible y con un peso de 2000 libras recorrería un promedio de 22.7 millas por galón, calculado por:

$Y' = a + b1\ X1 + b2\ X2 = 6.3 + 0.2\ (92) - 0.001(2000) = 22.7$

Los valores de los coeficientes a y b se determinan mediante el método de los mínimos cuadrados.

Veamos un ejemplo de *regresión múltiple con tres variables independientes* resuelto en **Excel** y **Minitab**.

Para su identificación es el ejercicio de **Salsberry Realty** (problema 5.8), que vende casas usadas y desea saber los costos de calefacción de cada casa como información a sus potenciales clientes.

NOTA:

Para fines didácticos aprovecharemos lo siguiente: en la solución de este problema se toman tres variables independientes: X1 es la temperatura externa, X2 es el espesor del material aislante de la casa y, X3 es la edad del calentador. Cuando una variable se desea eliminar por no ser buen predictor, al eliminarlo se debe de iniciar el cálculo de nuevo sin esta variable eliminada. También al agregar una nueva variable a la ecuación, como en este caso, "garage" como X3, también hacer de nuevo los cálculos. En este caso se ha querido meter una variable **nominal** que no tiene un valor numérico, y **"tiene garaje con calefacción" es 1 y "no tiene garaje con calefacción" es cero, ya que Excel y Minitab no acepta letras, solo números. Es importante el análisis que se hace, sobre cómo interpretar cada uno de los coeficientes de la ecuación.**

De igual manera, en este problema (en la parte inferior) se explica cómo interpretar y definir el resumen, que es el primer recuadro del cálculo que hace el software de Excel.

CORRELACIÓN DE PEARSON

PROBLEMA 5.1. CORRELACIÓN DE PEARSON

Excel.

Determinar la correlación de Pearson entre la variable dependiente "Y" y la variable independiente "X" de los siguientes valores:

X	Y
20	12
18	16
16	10
15	14
14	12
12	10
12	9
10	8
8	7
5	2

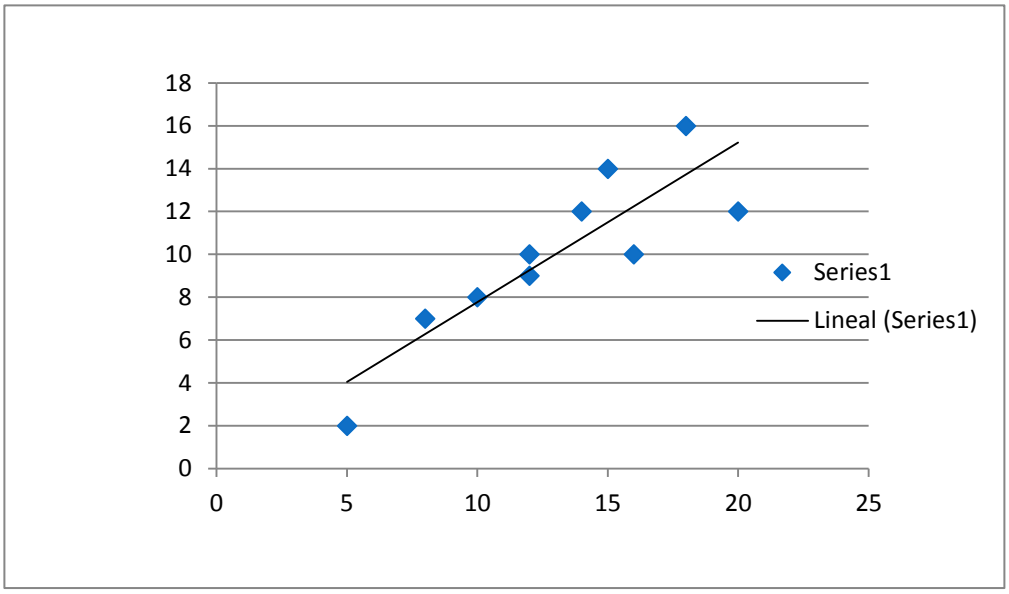
COMANDOS:

Datos – análisis de datos – coeficiente de correlación – rango de entrada (marcar los datos de los valores de X y de Y) – agrupados en columnas (para éste caso) – rango de salida (señalar con el cursor la celda donde uno desea anotar los cálculos) – aceptar. Listo

	Columna 1	*Columna 2*
Columna 1	1	
Columna 2	**0.86917969**	1

Existe una correlación entre las variables X y Y del 86.91%

Si graficamos se ve:

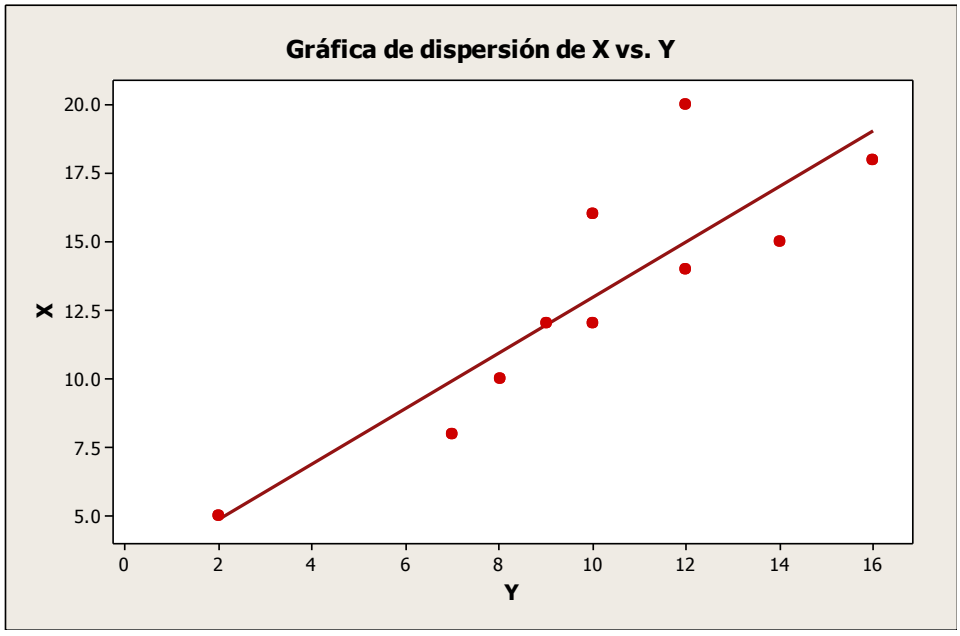


Minitab.

Determinar la correlación de Pearson que existe entre los valores de X y la variables Y.

COMANDOS:

Abrir Minitab – cargar hoja de trabajo – estadística – correlación –variable: seleccionar las columnas C1 y C2 – método: seleccionar correlación de Pearson. Aceptar. Listo



PROBLEMA 5.2. CORRELACIÓN. CASAS

X	Y
360	15
153	10
276	12
475	14
188	7
672	23
374	16
449	12
152	6
573	19
662	26

Determinar la correlación de Pearson de la variable independiente X : construcción de casas de interés social contra los datos de ventas de estufas que es la variable dependiente Y para hacer pronósticos a futuro de la empresa.

COMANDOS:

Datos – análisis de datos – coeficiente de correlación – rango de entrada (marcar los datos de los valores de X y de Y) – agrupados en columnas (para éste caso) – rango de salida (señalar con el cursor la celda donde uno desea anotar los cálculos) – aceptar. Listo

	Columna 1	*Columna 2*
Columna 1	1	
Columna 2	0.92368524	1

Minitab.

Determinar la correlación de Pearson de la variable independiente X : casas de interés social contra los datos de ventas de estufas que es la variable Y para hacer pronósticos a futuro de la empresa.

COMANDOS:

Abrir Minitab – cargar hoja de trabajo – estadística – correlación – variable: seleccionar las columnas C1 y C2 – método: seleccionar correlación de Pearson. Aceptar. Listo

Correlación: X, Y

```
Correlación de Pearson de X y Y = 0.924
Valor p = 0.000
```

REGRESIÓN LINEAL SIMPLE.

PROBLEMA 5.3. REGRESIÓN LINEAL SIMPLE. OXÍGENO

Excel.

X	Y
3	5
7	11
11	21
15	16
18	16
27	28
29	27
30	25
30	35
31	30
31	40
32	32
33	34
33	32
34	34
36	37
36	38
36	34
37	36
38	38
39	39
39	36
39	45
40	39
41	41
42	40
42	44
43	37
44	44
45	46
46	46
47	49
50	51

De los desechos residuales de una fábrica, se determinaron las reducciones de sólidos disueltos correspondientes a reducciones de oxígeno, para estos datos se puedan aplicar en otras muestras. Determinar la ecuación de regresión lineal simple para hacer estimaciones de DBO (degradación biológica de oxígeno) a partír de los sólidos disueltos (X) para estimar oxígeno degradado (Y) con los datos de las 33 muestras?

PREGUNTAS:

1.-Establecer las hipótesis
2.- ¿Qué es el coeficiente de correlación de Pearson y su valor en este problema?
3.- ¿Qué es el coeficiente de determinación y su valor en este problema?
4.-Determina la ecuación de regresión para este problema.
5.- ¿Cuál sería el valor estimado de Y si X es de 35, gráfica y analíticamente?
6.- ¿Cómo interpretas el valor del Error típico en este problema?
7.- ¿Cuál es valor crítico de "F" y qué tipo de prueba es?
8.-Cuál es el valor crítico de "t" y qué tipo de prueba es?
9.- ¿Cuáles son los estadísticos de prueba en la regresión lineal simple?
10.- ¿Cuáles son las reglas de decisión para cada estadístico de prueba?
11.- ¿A qué conclusión llegas de aceptar o rechazar Ho con los tres estadísticos de prueba.?. Explicar porqué

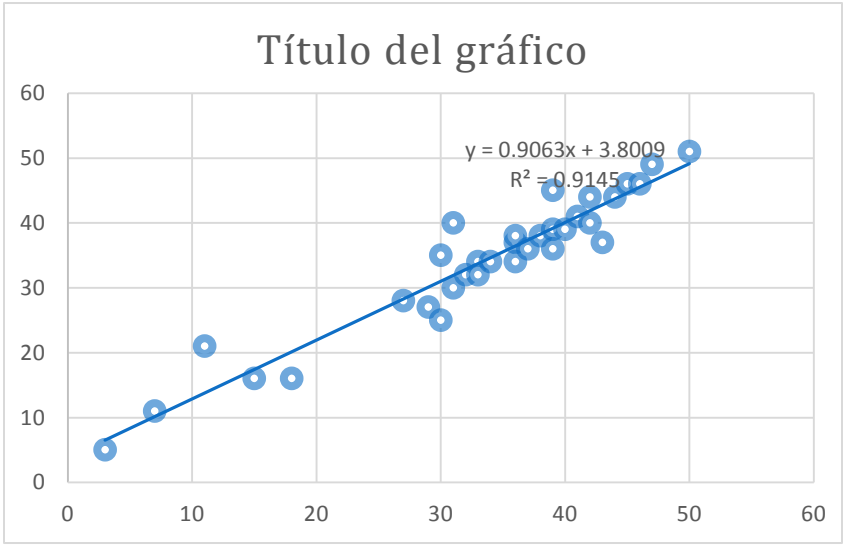
COMANDOS:

Datos-Análisis de datos-Regresión-Aceptar-Rango Y entrada (seleccionar columna de Y)- Rango X de entrada (seleccionar columna de X)-Rango salida (colocar el cursor en una celda vacía)- Aceptar.

	Resumen						
Estadísticas de la regresión							
Coeficiente (	0.95629268						
Coeficiente (	0.91449569						
R^2 ajustadc	0.91173749						
Error típico	3.20729814						
Observacion	33						
ANÁLISIS DE VARIANZA							
	Grados de liberta	*na de cuadrad*	*io de los cua*	*F*	*alor crítico de F*		
Regresión	1	3410.62555	3410.62555	331.554844	4.1377E-18		
Residuos	31	318.889602	10.2867613				
Total	32	3729.51515					
	Coeficientes	*Error típico*	*Estadístico t*	*Probabilidad*	*Inferior 95%*	*Superior 95%*	
Intercepción	3.80087869	1.75626931	2.1641776	0.03827121	0.21894382	7.38281356	
Variable X 1	0.90631431	0.04977384	18.2086475	4.1377E-18	0.8047999	1.00782873	

COMANDOS PARA LA GRÁFICA:

Marcar las columnas A y B (x y Y respectivamente) – insertar – seleccionar gráfico de dispersión-seleccionar la 1a. Gráfica de dispersión-con clic derecho tocar con el cursor uno de los puntos de la gráfica-agregar línea de tendencia – en opciones marcar "lineal" en automático – presentar ecuación en la gráfica – Presentar valor de R en el gráfico – enter. Listo



RESPUESTAS:

1.-Ho: b=0 no hay correlación entre la variable independiente X y la dependiente Y. H1: b‡ 0 sí hay correlación o diferente de 0.

2.-El Coeficiente de correlación de Pearson, es la correlación que existe entre la variable X (sólidos disueltos) y la variable Y (oxígeno degradado). Su valor es de: 0.95629 o sea 92.63 %.

3.-El coeficiente de determinación es la proporción de los valores de Y explicados a partir de los valores que la da X, y es de 0.9145 y se denomina con la letra R. La diferencia de 1 (1-0.9145) es el residuo y se define como los valores no explicados de Y que probablemente se deba a errores de muestreo o simplemente se desconoce.

R^2 es el coeficiente de determinación ajustado y se utiliza cuando hay muchas variables dependientes que pueden alterar los resultados, (este concepto se verá ampliamente en regresión lineal múltiple más adelante).En este caso tiene un valor de = 0.91173749.

4.-Y*= 3.0087869+ 0.90631431(X)

5.-Y*= 3.0087869+ 0.90631431(35) = 34.729

6.- Es el error estándar o típico y se interpreta como la desviación estándar que hay que sumar y restar a cada valor de Y estimado y en ese rango se encuentra el 68% de los datos, (teoría de limite central).

Para este problema es de 3.20729814

7.-F es la prueba de Fisher y es la prueba global. El valor crítico de F a .95 y 1 grado de libertad en numerador y 31 en denominador es de 4.16 (de tablas interpolando).

8.- t es la prueba individual y para este caso el valor crítico es para 0.975 y 31 g.l. es de =2.04

9.- Los estadísticos de prueba son : F de Fisher como prueba global; t de Studen como prueba individual y p para determinar la probabilidad de aceptar o rechazar la Ho.

10.-Las reglas de decisión para los estadísticos de prueba son:

Si Fcalculada es mayor que Fcrítica se rechaza ala Ho. De lo contrario se acepta.

Si t calculada es mayor que t crítica se rechaza la Ho. De lo contrario se acepta
Si p calculada es mayor que el nivel de significancia se acepta la Ho.

11.-Para F: se rechaza la Ho porque Fcalculada = 331.55 es mayor que la F crítica= 4.16

Para t : se rechaza la Ho. Porque tcalculada = 18.208 es mayor que tcrítica= 2.04

Para p en F: se rechaza la Ho porque Fcalculada es de 4.1377E-18 mucho más chica que el nivel de significancia de 0.05. por lo que se rechaza la Ho. En ambos casos es muy remota de suceder o de ser cierta la Ho.

NOTA: Con la gráfica también es posible hacer pronósticos de Y, pero no tienen la misma exactitud.Si prolongamos la recta hacia la línea de Y, en el punto donde la toca ahí es valor de "a" =3.80087869. La pendiente "b" tiene un valor de 0.90631431.

Minitab.

X	Y
3	5
7	11
11	21
15	16
18	16
27	28
29	27
30	25
30	35
31	30
31	40
32	32
33	34
33	32
34	34
36	37
36	38
36	34
37	36
38	38
39	39
39	36
39	45
40	39
41	41
42	40
42	44
43	37
44	44
45	46
46	46
47	49
50	51

De los desechos residuales de una fábrica, se determinaron las reducciones de sólidos disueltos correspondientes a reducciones de oxígeno, para estos datos se puedan aplicar en otras muestras.

Determinar la ecuación de regresión lineal simple para hacer estimaciones de DBO (degradación biológica de oxígeno) a partir de los sólidos disueltos (X) para estimar oxígeno degradado (Y) con los datos de las 33 muestras.

COMANDOS:

Cargar hoja de trabajo-Estadística- regresión-regresión-ajustar modelo de regresión.

Respuestas: C2-(columna de Y)-predicciones continuas: cargar C1 (columnas X)-opciones_ Nivel de confianza: 95-tipode intervalo: bilateral-aceptar- aceptar. Listo

NOTA:

Para no duplicar las anotaciones no se dan las Hojas de Trabajo. Ya que son iguales a la descripción del problema.

Análisis de regresión: Y vs. X

Análisis de Varianza

Fuente	GL	SC Ajust.	MC Ajust.	Valor F	Valor p
Regresión	1	3410.6	3410.63	331.55	0.000
X	1	3410.6	3410.63	331.55	0.000
Error	31	318.9	10.29		
Falta de ajuste	23	158.2	6.88	0.34	0.979
Error puro	8	160.7	20.08		

```
Total              32     3729.5


Resumen del modelo
                     R-cuad.  R-cuad.
      S  R-cuad.  (Ajustado)   (pred)
3.20730   91.45%      91.17%   90.19%


Coeficientes
                    EE del
Término      Coef   coef.  Valor T  Valor p   VIF
Constante    3.80    1.76     2.16    0.038
X          0.9063  0.0498    18.21    0.000  1.00


Ecuación de regresión
```

Y = 3.80 + 0.9063 X

F crítica a 0.95 1/31 gl= 4.16 (tablas). tcrítica a 0.975 31 gl = 2.04 Alfa=0.05
Ho =no hay correlación de X vs Y. H1 = Sí hay correlación entre X y Y

Tenemos los siguientes resultados para el análisis:

Coef. de determinación =91.45 %= R
R ajustado=91.17 %. Error típico= S =3.207 (se toma como Desv. Std.)
a=3.80 (ordenada al origen); b=0.9063 (pendiente); Ecuación: Y*=3.80+0.9063 (X); Fcalculada=331.55 (Fisher); p=0.00 (redondeada) t=18.21.

Con esta información concluimos: Que dado que Fcalculada=331.55>Fcrítica=4.16 se rechaza la Ho.(prueba global)

tcalculada=18,21 >tcrítica=2.04 se rechaza la Ho. (Prueba individual). Se tiene la ecuación para predicciones.

Conclusión:

Se puede asegurar que existe una muy buena correlación entre las variables independientes X y la variable dependientes Y.

Que los sólidos disueltos son valores aceptables como predictores del DBO de las muestras residuales.

Se tiene también la ecuación para hacer cualquier predicción y el factos Desv. Std o error típico. También podemos comprobar que los valores coinciden con los de Excel.

Trabajador	Cervezas consumidas X	Alcohol en sangre Y
1	6	0.1
2	7	0.09
3	7	0.09
4	4	0.1
5	5	0.1
6	3	0.07
7	3	0.1
8	6	0.12
9	6	0.09
10	3	0.07
11	3	0.05
12	7	0.08
13	1	0.04
14	4	0.07
15	2	0.06
16	7	0.12
17	2	0.05
18	1	0.02

Problema 5.4. **Excel**. Regresión L. simple. Alcohol.

Se desea tener un cuadro de referencia rápida de la cantidad de cervezas tomadas por los trabadores de una planta contra el contenido de alcohol en la sangre. Para tal fin se tomaron 18 muestras de sangre a igual número de trabajadores que ingerieron cervezas, los resultados estan en las columnas A, B y C

Determinar la ecuación que nos permita determinar cuanto alcohol hay en sangre si ha consumido X número de cervrzas?

Utilizar un nivel de confianza del 95%

Ho: b=0 Que no hay correlación entre X y Y pendiente = 0

H1: b ≠0 Que sí hay correlación entre X y Y pendiente diferente de 0

COMANDOS: Datos-Análisis de datos-Regresión-Aceptar-Rango Y entrada(seleccionar columna de Y)-Rango X de entrada (seleccionar columna de X)-seleccionar nivel de confianza deseado-rango de salida (seleccionar con el cursor la celda donde se desea tener los cálculos-aceptar.listo

Resumen

Estadísticas de la regresión	
Coeficiente	0.778795256
Coeficiente	0.60652205
R^2 ajustad	0.581929679
Error típico	0.017726699
Observacio	18

ANÁLISIS DE VARIANZA

	Grados de liberta	*a de cuadra*	*lio de los cua*	*F*	*alor crítico de F*
Regresión	1	0.00775	0.00775	24.6630156	0.00014011
Residuos	16	0.0050278	0.00031424		
Total	17	0.0127778			

	Coeficientes	*Error típico*	*Estadístico t*	*Probabilidad*	*Inferior 95%*	*Superior 95%*	*nferior 95.0%*
Intercepció	0.036141732	0.0095681	3.77730531	0.00164988	0.01585821	0.05642525	0.01585821
Variable X 1	0.009992842	0.0020122	4.96618723	0.00014011	0.00572722	0.01425846	0.00572722

Conclusion:

Y = 0.03614173 + 0.00999284 (X)

Y´=a+b(X)

Ecuación para estimar la cantidad de alcohol en sangre según la cantidad de cervezas consumidas,

t crítica para para 0.975 nc y 16 gl = 2.12 Fcrítica a 1 y 16gl y =4.49 (95%)

El estadístico de prueba es **F** de **Fisher.**

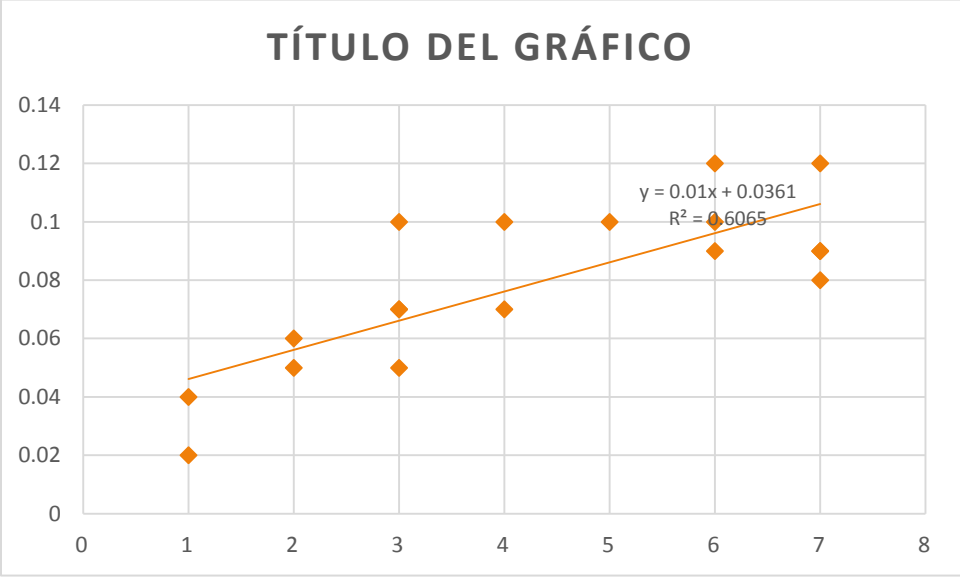
Dado que Fcalculada =24.66 es mayor que Fcrítica=4.49 se rechaza la Ho y se acepta la H1 es decir sí hay correlación entre la cantidad de cervezas consumidas (X) y la cantidad

Observar que el valor de p = 0.00014 en estadístico t es igual a p para Fisher por ser una sola variable independiente.

Para graficar:

COMANDOS PARA LA GRÁFICA:

Marcar las columnas A y B (x y Y respectivamente) – insertar – seleccionar gráfico de dispersión – seleccionar la 1a. Gráfica de dispersión – con clic derecho tocar con el cursor uno de los puntos de la gráfica – agregar línea de tendencia – en opciones marcar "lineal" en automático – presentar ecuación en la gráfica – Presentar valor de R en el gráfico – enter. Listo



PROBLEMA 5.4. REGRESIÓN LINEAL SIMPLE. ALCOHOL.

Minitab.

Se desea tener un cuadro de referencia rápida de la cantidad de cervezas tomadas por los trabadores de una planta contra el contenido de alcohol en la sangre. Para tal fin se tomaron 18 muestras de sangre a igual número de trabajadores que ingirieron cervezas, los resultados están en las columnas A, B y C.

Determinar la ecuación que nos permita determiner, ¿cuánto alcohol hay en sangre si ha consumido X número de cervezas?.

Utilizar un nivel de confianza del 95%

Trabajador	Cervezas (X)	Alcohol (Y)
1	6	0.10
2	7	0.09
3	7	0.09
4	4	0.10
5	5	0.10
6	3	0.07
7	3	0.10
8	6	0.12
9	6	0.09
10	3	0.07
11	3	0.05
12	7	0.08
13	1	0.04
14	4	0.07
15	2	0.06
16	7	0.12
17	2	0.05
18	1	0.02

COMANDOS:

Cargar hoja de trabajo – Estadística - regresión regresion – ajustar modelo de regression – Respuestas: cargarC3 (Y) – predicciones contínuas: cargar C2 (X) – opciones – seleccionar nivel de confianza deseado – tipo de intervalo: bilateral – aceptar – aceptar. listo.

Análisis de regresión: Alcohol (Y) vs. Cervezas (X)

Análisis de Varianza

Fuente	GL	SC Ajust.	MC Ajust.	Valor F	Valor p
Regresión	1	0.007750	0.007750	24.66	0.000
Cervezas (X)	1	0.007750	0.007750	24.66	0.000
Error	16	0.005028	0.000314		
Falta de ajuste	5	0.001686	0.000337	1.11	0.409
Error puro	11	0.003342	0.000304		
Total	17	0.012778			

Resumen del modelo

S	R-cuad.	R-cuad. (ajustado)	R-cuad. (pred)
0.0177267	60.65%	58.19%	49.47%

Coeficientes

Término	Coef	EE del coef.	Valor T	Valor p	VIF
Constante	0.03614	0.00957	3.78	0.002	
Cervezas (X)	0.00999	0.00201	4.97	0.000	1.00

Ecuación de regresión

Alcohol (Y) = 0.03614 + 0.00999 Cervezas (X)

Ho: b=0 que no hay correlación entre X y Y

H1b diferente de 0: que sí hay correlación entre (X) y (Y)

Fcrítica a 1/16 gl = 4.49, tcrítica a 0.975 y 16 gl = 2.12

Dado que Fcalculada=24.66>Fcrítica=4.49 se rechaza la Ho y se acepta la H.

Es decir si existe correlación entre las variables X y Y. Sí es confiable la relación entre las cervezas consumidas y el contenido de alcohol en sangre para hacer predicciones rápidas.

Por otro lado se tiene la ecuación Y*=0.03614 +0.0099 (X): Esta ecuación se puede interpretar también de la siguiente manera:

> Por cada cerveza que la persona consuma, se incrementa en 0.0099 grados de alcohol en sangre y 0.03614 es la ordenada al origen (eje Y).

PROBLEMA 5.5. REGRESIÓN LINEAL SIMPLE. IMPRESORAS.

Excel.

El gerente de una empresa dedicada a la venta de impresoras, desea incrementar las ventas y busca una variable independiente (X) que tenga correlación con las ventas de las impresoras (Y).

Toma una muestra de las ventas hechas en el mes de enero por 10 agentes y determina cuantas llamadas telefónicas realizo cada uno de ellos. Los datos están en las columnas A y B respectivamente.

1.-¿Puede el gerente considerar que existe correlación entre las llamadas y las ventas de impresoras para así incrementar las llamadas telefónicas. Explicar dando cifras?

2.-¿Determinar la ecuación de regresión para futuras estimaciones de ventas?

3.-¿Cuántas impresoras estimaría vender si un vendedor hiciera 50 llamadas telefónicas?

Usar un nivel de confianza del 95%.

vendedor nombre	llamadas X	ventas Y
Tom Keller	20	30
Jeff hall	40	60
Brian Virost	20	40
Greg Fish	30	60
Susan Welch	10	30
Carlos ramirez	10	40
Rich Niles	20	40
Mike Kiel	20	50
Mark Reynolds	20	30
Soni Jones	30	70

Ho= No hay correlación entre las llamadas telefónicas y las ventas. Es decir b=0
H1= sí, hay correlación entre las llamadas telefónicas y las ventas. b diferente de 0

COMANDOS: Datos – análisis de datos – regresión – aceptar – rango Y de entrada: seleccionar los datos de la columna Y – rango X de entrada: seleccionar datos de la columna X – NC:95% - rango de salida: seleccionar con el cursor una celda vacía donde desee colocar los cálculos – aceptar – listo.

Resumen

Estadísticas de la regresión	
Coeficiente de correlación múltiple	0.759014109
Coeficiente de determinación R^2	0.576102418
R^2 ajustado	0.52311522
Error típico	9.900823995
Observaciones	10

ANÁLISIS DE VARIANZA						
	Grados de libertad	*na de cuadrad*	*io de los cua*	*F*	*alor crítico de F*	
Regresión	1	1065.78947	1065.78947	10.8724832	0.01090193	
Residuos	8	784.210526	98.0263158			
Total	9	1850				
	Coeficientes	*Error típico*	*Estadístico t*	*Probabilidad*	*Inferior 95%*	*Superior 95%*
Intercepción	18.94736842	8.49881856	2.22941204	0.05634865	-0.65094232	38.5456792
Variable X 1	1.184210526	0.35914063	3.29734487	0.01090193	0.35603074	2.01239031

A48:G62F crítica a 1 y 8 gl con .05 N.S. es de 5.32

RESPUESTAS: 1.- Dado que Fcalculada es mayor que Fcrítica se rechaza la Ho, Es decir, se acepta qiue existe correlación entre las llamadas telefónicas (X) y las ventas de impresoras (Y).

Fcalculada = 10.87 contra Fcrítica = 5.32

2.-La ecuación de regresión es: $Y^*=18.94 + 1.18(X)$

3.- $Y^* = 18.94 + 1.18(50) = 77.94 = 78$ impresoras probablemente sean vendidas a un nivel de confianza del 95 %

La probabilidad de que la Ho sea cierta es de 0.0109 que se contrasta con con el nivel de significancia del 0.05, es decir es menor que alfa. Pocas posibilidades de que la Ho sea cierta.

La prueba individual de "t" se recomienda para la regresión lineal múltiple cuando hay varias variables independientes X

En la columna F donde dice: "valor crítico de F" es el valor de "p" = 0.0109

Minitab

Llamadas	Ventas
20	30
40	60
20	40
30	60
10	30
10	40
20	40
20	50
20	30
30	70

COMANDOS:

Cargar hoja de trabajo – Estadística – regresión regresión - ajustar modelo de regresión – Respuestas: cargarC2 (Y) – predicciones contínuas: cargar C1 (X) – opciones –Seleccionar nivel de confianza deseado – tipo de intervalo: bilateral –aceptar.aceptar. listo.

Análisis de regresión: Ventas vs. Llamadas

Análisis de Varianza

Fuente	GL	SC Ajust.	MC Ajust.	Valor F	Valor p
Regresión	1	1065.8	1065.79	10.87	0.011
Llamadas	1	1065.8	1065.79	10.87	0.011
Error	8	784.2	98.03		
Falta de ajuste	2	404.2	202.11	3.19	0.114
Error puro	6	380.0	63.33		
Total	9	1850.0			

Resumen del modelo

S	R-cuad.	R-cuad. (ajustado)	R-cuad. (pred)
9.90082	57.61%	52.31%	33.03%

Coeficientes

Término	Coef	EE del coef.	Valor T	Valor p	VIF
Constante	18.95	8.50	2.23	0.056	
Llamadas	1.184	0.359	3.30	0.011	1.00

Ecuación de regresión

Ventas = 18.95 + 1.184 Llamadas

Ho : No hay correlación entre las llamadas y las ventas, es decir b=0
H1 : sí hay correlación entre las llamadas telefónicas (X) y las ventas (Y).

El valor crítico de F con 1 y 8 gl. y 95% de NC = 5.32
Dado que el valor de Fcalculada es de 10.87 es más grande que Fcrítica = 5.32 se rechaza la Ho y se acepta la H1. Es decir sí hay correlación entre la las variables X = llamadas tel. contra las Y = ventas.

Por otro lado p=0.11 es mayor que el nivel de significancia de 0.05. También se rechaza la Ho.

PROBLEMA 5.6. REGRESIÓN LINEAL SIMPLE. VENTAS JANSEN AND FOOD.

Excel.

La empresa Jansen and Food que vende alimentos preparados, desea hacer pronósticos de ventas para el año 2018. En la columna A y columna B, se anotaron las ventas anuales en millones de dólares tenidos en un periodo de 6 años.

Determinar:

- ✓ La ecuación por el método de los mínimos cuadrados, para poder calcular las ventas probables para el año de 2018.
- ✓ Construir la gráfica.

Año	Ventas M$$	Código (año)
2012	7	1
2013	10	2
2014	9	3
2015	11	4
2016	13	5
2017	12	6

COMANDOS:

Datos – análisis de datos – regression – aceptar – rango de entrada Y: seleccionar datos de la columna B que son las ventas – rango de entrada X : seleccionar código de datos columna C – nivel de confianza de 95% o el deseado-rango de salida: seleccionar con el curso una celda donde se desea anotar los cálculos –aceptar – listo.

Resumen	
Estadísticas de la regresión	
Coeficiente de correlación múltiple	0.89076899
Coeficiente de determinación R^2	0.79346939
R^2 ajustado	0.74183673
Error típico	1.09761647
Observaciones	6

ANÁLISIS DE VARIANZA							
	Grados de liberta	*na de cuadra*	*dio de los cua*	*F*	*alor crítico de F*		
Regresión	1	18.5142857	18.5142857	15.3675889	0.01724548		
Residuos	4	4.81904762	1.2047619				
Total	5	23.3333333					
	Coeficientes	*Error típico*	*Estadístico t*	*Probabilidad*	*Inferior 95%*	*Superior 95%*	*nferior 95.0%*
Intercepción	6.73333333	1.02182532	6.58951506	0.00274686	3.89629143	9.57037524	3.89629143
Variable X 1	1.02857143	0.26238052	3.92015165	0.01724548	0.30008632	1.75705654	0.30008632

a= 6.7333 Ordenada al origen. Pendiente b= 1.02857

Ecuación para estimar ventas para el año 2018 (7 código)

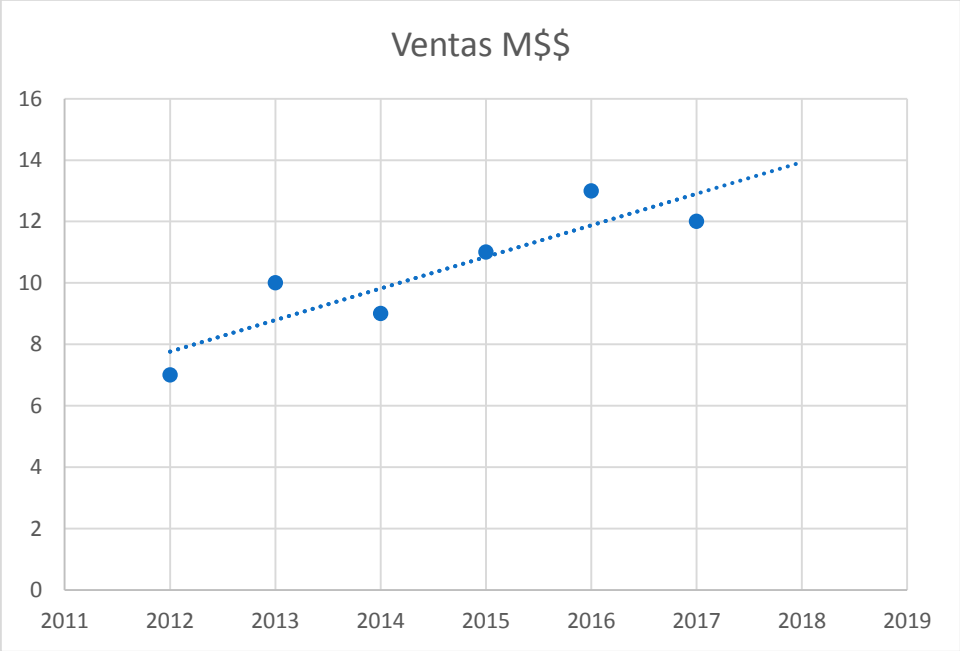
Y* = 6.733 +
1.02857(X)

Y* = 6.733 + 1.02857(7) = 13.93329 Dlls. Venta estimada para 2018

Con un más menos de 1.09 y ahí cae el 68 % de los casos

COMANDOS GRÁFICO:

Seleccionar columnas A y B – Insertar – seleccionar gráfico de nubes – la primera – con cursor clic derecho en un punto – seleccionar línea de tendencia – lineal – extrapolar: adelante: 1 – enter – listo.



Minitab.

La empresa Jansen and Food que que vende alimentos preparados, desea hacer pronósticos de ventas para el año 2018. En la columna A y columna B, se anotaron las ventas anuales en millones de dólares tenidos en un periodo de 6 años.

Determinar:

- ✓ La ecuación por el método de los mínimos cuadrados, para poder calcular las ventas probables para el año de 2018.
- ✓ Construir la gráfica.

Análisis de regresión: Ventas M$$ vs. Código (año)

Análisis de Varianza

Fuente	GL	SC Ajust.	MC Ajust.	Valor F	Valor p
Regresión	1	18.514	18.514	15.37	0.017
Código (año)	1	18.514	18.514	15.37	0.017
Error	4	4.819	1.205		
Total	5	23.333			

Resumen del modelo

S	R-cuad.	R-cuad. (ajustado)	R-cuad. (pred)
1.09762	79.35%	74.18%	45.60%

Coeficientes

Término	Coef	EE del coef.	Valor T	Valor p	VIF
Constante	6.73	1.02	6.59	0.003	
Código (año)	1.029	0.262	3.92	0.017	1.00

Ecuación de regresión

Ventas M$$ = 6.73 + 1.029 Código (año)

El pronóstico para el año 2018 (7) : Y*= 6.73 + 1.029(7) = 13.933 Mlls dlls.

PROBLEMA 5.7. REGRESIÓN LINEAL SIMPLE COEFICIENTES.

Excel.

Y	X
80	61
95	28
94	74
101	46
105	44
89	38
106	72
92	41
105	49
107	69
111	82
114	76
83	39
112	64
91	77
88	50
105	55
106	59
105	86
80	63
85	31
93	57
85	70
92	43
90	70
89	54
85	51
96	58
85	63
98	73
101	71
106	76
112	76
93	59
110	71

Se ha seleccionado a una muestra de 35 alumnos y se les practicó una evaluación de matemáticas para ver si existe correlación con el coeficiente de inteligencia de dichos alumnos.

Los resultados de la prueba de matemáticas están anotados en la columna B (X) y del Índice del coeficiente de inteligencia practicado en otra prueba a los mismos alumnos, se da en la columna A (Y).

La pregunta es : *¿existe correlación entre la variable independiente X (calificaciones del examen de matemáticas) con la variable dependiente Y (coeficiente de inteligencia)?*

Realizar una prueba de Hipótesis a un nivel de confianza del 95%

¿Qué tan confiable es esta prueba?

¿Cuál es la ecuación para que con los resultados de un examen de matemáticas se pueda medir el CI de manera rápida ?

Establecimiento de las Hipótesis:

Ho : No existe correlación entre una evaluación de matemáticas y el coeficiente de inteligencia de un alumno.

H1 : sí, existe correlación entre la evaluación de matemáticas y el coeficiente de inteligencia de los alumnos.

COMANDOS:

Datos – análisis de datos – regresión – aceptar – rango Y de entrada: seleccionar los datos de la columna Y (CI) – rango X de entrada: seleccionar los datos de la columna X (puntuación de matemáticas) – seleccionar un nivel de confianza de 95% - Rango de salida: con el cursor seleccionar una celda donde se desea anotar los resultados del cálculo – aceptar – listo.

Resumen	
Estadísticas de la regresión	
Coeficiente de correlación múltiple	0.42869016
Coeficiente de determinación R^2	0.18377526
R^2 ajustado	0.15904117
Error típico	9.27093481
Observaciones	35

ANÁLISIS DE VARIANZA							
	Grados de liberta	*a de cuadrad*	*io de los cua*	*F*	*alor crítico de F*		
Regresión	1	638.613765	638.613765	7.43004118	0.0101857		
Residuos	33	2836.35766	85.9502322				
Total	34	3474.97143					
	Coeficientes	*Error típico*	*Estadístico t*	*Probabilidad*	*Inferior 95%*	*Superior 95%*	*nferior 95.0%*
Intercepción	79.4158351	6.57749856	12.0738658	1.1837E-13	66.0338137	92.7978566	66.0338137
Variable X 1	0.29076611	0.10667145	2.72581019	0.0101857	0.07374142	0.5077908	0.07374142

Fcrítica a 1/33 gl y a NC de 95% tiene un valor de 4.14

La correlación entre las dos variables es de 42.86 % es baja, dado que Fcalculada es de 7.43 es mayor que Fcrítica = 4.14: se rechaza la Ho y se acepta la H1. Es decir sí existe correlación entre las variables pero se considera que es baja.

La probabilidad de que la Ho se cierto es de 0.0101 que al ser contrastada con el nivel de significancia (alfa) que es de 0.05. por lo que se rechaza la Ho.

COEFICIENTE DE DETERMINACION:
Proporción de la variación total de la variable dependiente "Y" que se explica o contabiliza, por la variación de la variable independiente "X". Que tiene un valor de: 18.37 % tambien baja.

Por lo tanto, al ser baja la correlación entre las dos variables (calificación de matemáticas y el CI) no es una prueba confiable.

La ecuación de regresión por éste método es : Y* = 79.4158+ 0.2907 (X).

Por otro lado podemos decir: que los valores no explicados o contabilizados o que no se conocen de la variable Y es de 1 -0.1837 = 0.8163; es decir es del orden del 81.63 %.

Hoja de Trabajo

Y	X
80	61
95	28
94	74
101	46
105	44
89	38
106	72
92	41
105	49
107	69
111	82
114	76
83	39
112	64
91	77
88	50
105	55
106	59
105	86
80	63
85	31
93	57
85	70
92	43
90	70
89	54
85	51
96	58
85	63
98	73
101	71
106	76
112	76
93	59
110	71

Minitab.

Se ha seleccionado a una muestra de 35 alumnos y se les practicó una evaluación de matemáticas para ver si existe correlación con el coeficiente de inteligencia de dichos alumnos. Los resultados de la prueba de matemáticas están anotados en la columna B (X) y del Índice del coeficiente de inteligencia practicado en otra prueba a los mismos alumnos, se da en la columna A (Y).

La pregunta es: *¿existe correlación entre la variable independiente X (calificaciones del examen de matemáticas) con la variable dependiente Y (coeficiente de inteligencia)?*

Realizar:

Una prueba de Hipótesis a un nivel de confianza del 95%

¿Qué tan confiable es esta prueba?

¿Cuál es la ecuación para que con los resultados de un examen de matemáticas se pueda medir el CI de manera rápida?

Establecimiento de las Hipótesis:

Ho : No existe correlación entre una evaluación de matemáticas y el coeficiente de inteligencia de un alumno.

H1 : sí, existe correlación entre la evaluación de matemáticas y el coeficiente de inteligencia de los alumnos.

COMANDOS:

Cargar hoja de trabajo – estadística – regresión – regresión - ajustar modelo de regresión respuestas: cargar columna C1 (Y) – predictores: cargar columna C2 (X) – opciones: 95% - aceptar – aceptar – listo.

Análisis de regresión: Y vs. X

Análisis de Varianza

Fuente	GL	SC Ajust.	MC Ajust.	Valor F	Valor p
Regresión	1	638.6	638.61	7.43	0.010
X	1	638.6	638.61	7.43	0.010
Error	33	2836.4	85.95		
Falta de ajuste	27	2651.7	98.21	3.19	0.076
Error puro	6	184.7	30.78		
Total	34	3475.0			

Resumen del modelo

S	R-cuad.	R-cuad. (ajustado)	R-cuad. (pred)
9.27093	18.38%	15.90%	10.00%

Coeficientes

Término	Coef	EE del coef.	Valor T	Valor p	VIF
Constante	79.42	6.58	12.07	0.000	
X	0.291	0.107	2.73	0.010	1.00

Ecuación de regresión

Y = 79.42 + 0.291 X

Ecuación de regresión para pronósticos rápidos

Dado que, el coeficiente de determinación (18,38 %) es tan bajo, no se recomienda la aplicación de este tipo de pronósticos.

La Ho se rechaza pues Fcalculada es de 7.43 es mayor que Fcrítica a 1/33 gl = 4.14. aunque se acepta la H1 es baja la correlación entre ambas variables.

REGRESIÓN LINEAL MULTIPLE

MULTICOLINEALIDAD EN LA REGRESIÓN LINEAL MÚLTIPLE

La Multicolinealidad existe cuando las variables independientes están correlacionadas. Las variables independientes correlacionadas dificultan las inferencias acerca de los coeficientes de regresión individuales y sus efectos individuales sobre la variable dependiente.

En la práctica, es casi imposible seleccionar variables que carezcan por completo de alguna relación. En otras palabras, es casi imposible crear un conjunto de variables independientes que no estén correlacionadas hasta cierto punto. Sin embargo, la comprensión general del punto de multicolinealidad es importante. Pero hay que destacar que la multicolinealidad no afecta la capacidad de una ecuación de regresión múltiple para predecir la variable dependiente. No obstante, cuando se tenga interés en evaluar la relación entre cada variable independiente y la variable dependiente. La multicolinealidad puede presentar resultados inesperados.

Por ejemplo, si se usan dos promedios de calificaciones de preparatoria (calificaciones de matemáticas y calificaciones de aptitudes verbales) como variables independientes y la calificación de admisión a la univesidad (variable dependiente) y las variables independientes con multicolinealidad alta, se esperaría que las dos variables independientes estén positivamente correlacionadas con la variable dependiente. Sin embargo una de las variables independientes puede tener un signo negativo inesperado e inexplicable. En esencia, estas dos variables independientes son redundantes cuando se trata de explicar la misma variación de la variable dependiente.
Otra razón para evitar variables independientes correlacionadas es que puedan generar resultados erróneos en las pruebas de hipótesis de las variables independientes individuales. Esto se debe a la inestabilidad del error estándar de estimación.

A continuación, se dan algunas pistas que indican problemas con la multicolinealidad:

1.- Una variable independiente conocida como factor de predicción importante resulta con un coeficiente de regresión que no es significativo.

2.-Un coeficiente de regresión que debería tener un signo positivo resulta negativo, o lo contrario.

3.- Cuando se agrega o elimina una variable independiente, hay un cambio drástico de los valores de los coeficientes de regresión restante.

En la evaluación de la ecuación de regresión múltiple, una aproximación para reducir los efectos de la multicolinealidad es seleccionar con cuidado las variables independientes incluidas en la ecuación. Una regla general es que, si la correlación entre dos variables independientes se encuentra entre -0.70 y 0.70, es probable que no haya problema al emplear las dos variables independientes. Otra prueba más precisa es utilizar el Factor de inflación de la varianza (VIF) y se determina con:

$$VIF = 1 / 1-R^2$$

R^2 es el coeficiente de determinación de la variable independiente seleccionada sirve como variable dependiente y las variables independientes restantes, como variables independientes.

Un VIF mayor que 10 se considera insatisfactorio, e indica que la variable independiente se debe de eliminar del análisis. Veamos un ejemplo del VIF utilizando el problema 1. Regresión Lineal Múltiple. Calefacción.Veamos un problema de aplicación del VIF:

Problema 1. Regresión L. Múltiple. Calefacción: Donde se relaciona el costo de calefacción con las variables independientes: temperatura externa, cantidad de aislamiento y antigüedad del calentador.

Cost (Y)	Temp (X1)	Insul (X2)	Age (X3)
250	35	3	6
360	29	4	10
165	36	7	3
43	60	6	9
92	65	5	6
200	30	5	5
355	10	6	7
290	7	10	10
230	21	9	11
120	55	2	5
73	54	12	4
205	48	5	1
400	20	5	15
320	39	4	7
72	60	8	6
272	20	5	8
94	58	7	3
190	40	8	11
235	27	9	8
139	30	7	5

¿Parece que hay un problema de multicolinealidad?

Encontrar e interpretar el VIF de cada una de las variables independientes.

Empleando Minitab con una **prueba de correlación** de la variable dependiente y las variables independientes.

Una parte de la pantalla es la siguiente:

Correlación: Cost (Y), Temp (X1), Insul (X2), Age (X3)

	Cost (Y)	Temp (X1)	Insul (X2)
Temp (X1)	-0.812		
Insul (X2)	-0.257	-0.103	
Age (X3)	0.537	-0.486	0.064

Contenido de la celda: Correlación de Pearson

Los datos de azul indica la correlación entre las variables independientes. Ninguna de las correlaciones entre ellas sobre pasa -0.70 y 0.70, por lo que no se sospecha problemas de multicolinealidad.

La correlación mayor entre las variables independientes es -0.486 entre antigüedad y temperatura.

Para confirmar esta conclusión calcule el VIF de cada una de las variables independientes.
Primero se considera la variable independiente, temperatura. Emplee Minitab para determinar el coeficiente de determinación múltiple con la temperatura como variable dependiente, y la cantidad de aislamiento y antigüedad del calentador como variables independientes. La captura de la pantalla es la siguiente:

Resumen del modelo

S	R-cuad.	R-cuad. (ajustado)	(pred)
16.0311	24.14%	15.22%	2.72%

Coeficientes

Término	Coef	EE del coef.	Valor T	Valor p	VIF
Constante	58.0	12.3	4.70	0.000	
Insul (X2)	-0.51	1.49	-0.34	0.737	1.00
Age (X3)	-2.51	1.10	-2.27	0.036	1.00

Ecuación de regresión

Temp (X1) = 58.0 - 0.51 Insul (X2) - 2.51 Age (X3)

El coeficiente de determinación es 0.2414 (24.14%)

VIF = 1 / 1 - 0.2114 = 1.32 que es menor que 10, **lo que indica que la variable independiente temperatura,** no está muy correlacionada con las demás variables independientes.

De igual manera se procede con la variable independiente aislamiento como *variable dependiente* y la temperatura y edad del calentador como variables independientes. Se determina el coeficiente de determinación R_2^2 que se sustituye en la ecuación del VIF.

Por fortuna Minitab genera los valores del VIF de cada una de las variables independientes los cuales se reportan en la columna derecha con el encabezado "VIF" de la captura de la pantalla

de Minitab. Los valores para este problema de alrededor de 1.0, de aquí que se concluya que no hay problema de multicolinealidad en este ejemplo.

Resumen del modelo

S	R-cuad.	R-cuad. (ajustado)	R-cuad. (pred)
51.0486	80.42%	76.75%	68.72%

Coeficientes

Término	Coef	EE del coef.	Valor T	Valor p	VIF
Constante	427.2	59.6	7.17	0.000	
Temp (X1)	-4.583	0.772	-5.93	0.000	1.32
Insul (X2)	-14.83	4.75	-3.12	0.007	1.01
Age (X3)	6.10	4.01	1.52	0.148	1.31

Cost (Y)	Temp (X1)	Insul (X2)	Age (X3)
250	35	3	6
360	29	4	10
165	36	7	3
43	60	6	9
92	65	5	6
200	30	5	5
355	10	6	7
290	7	10	10
230	21	9	11
120	55	2	5
73	54	12	4
205	48	5	1
400	20	5	15
320	39	4	7
72	60	8	6
272	20	5	8
94	58	7	3
190	40	8	11
235	27	9	8
139	30	7	5

PROBLEMA 5.8. REGRESIÓN LINEAL MÚLTIPLE. CALEFACCIÓN.

Hoja de Trabajo.

Cost (Y)	Temp (X1)	Insul (X2)	Age (X3)
250	35	3	6
360	29	4	10
165	36	7	3
43	60	6	9
92	65	5	6
200	30	5	5
355	10	6	7
290	7	10	10
230	21	9	11
120	55	2	5
73	54	12	4
205	48	5	1
400	20	5	15
320	39	4	7
72	60	8	6
272	20	5	8
94	58	7	3
190	40	8	11
235	27	9	8
139	30	7	5

Salsberry Realty vende casas en la costa este de USA. Una de las preguntas más frecuentes de los potenciales compradores es: si compramos ésta casa: ¿Cuánto gastaremos en calefacción en invierno?

El departamento de investigación de Salsberry se le pidió considerar algunas directrices respecto de los costos de calefacción de casas unifamiliares. Se considera que tres variables se relacionan con dichos costos:

1.-Temperatura externa media diaria

2.-número de pulgadas del aislamiento del ático,

3.-los años de uso del calentador.

Para el estudio, el Dpto. de investigación selecciono una muestra de 20 casas aleatoriamente vendidas recientemente. Determino el costo de calefacción del mes de enero pasado, así como la temperatura externa en enero de la región, el número de pulgadas del aislamiento del ático y los años de uso del calentador.

La información se reporta en la tabla adjunta en las columnas A, B, C, y D.

1.- ¿Cuál es la ecuación de regresión para estimar el costo de calefacción de una casa?

2.- ¿Cuánto se estima que costará la calefacción de una casa que se encuentra en Filadelfia si la temperatura externa promedio es de 30°F, si el ático tiene un espesor de 5 pulgadas de aislamiento y el calentador tiene 10 años de servicio?

Usar un nivel de confianza del 95%.

Establecimiento de las hipótesis:

Ho : que no existe correlación entre las variables independientes Temperatura externa (X1), espesor del asilamiento (X2) y la edad del calentador (X3) con la variable dependiente Costo de calefacción (Y).

Es decir, que b= 0 (pendiente). También podemos decir que b1=b2=b3=0

H1: Que sí existe correlación entre las variables independientes con la dependiente; es decir b≠ 0

También podemos decir que b1≠0;b2≠0 y b3≠0

Lo correcto es usar las letras griegas: β1, β2 y β3 Porque lo se trata es determinar las pendientes poblacionales y no las pendientes muéstrales (b1, b2 y b3).

Regla de decisión: si Fcalculada es mayor que Fcrítica se rechaza la Ho. Cae en la zona de rechazo de la Ho.

Si p es mayor que alfa (nivel de significancia) se acepta la Ho. Cae en la zona de aceptación de la Ho.

COMANDOS:

Datos-análisis de datos – regresión – aceptar – rango Y de entrada: Seleccionar la columna de Y costo de calefacción – Rango X de entrada: seleccionar todos los datos de las tres columnas X1, X2 y X3 – nivel de confianza: 95% - con el cursor seleccionar la celda donde se anotaran los cálculos – aceptar – listo.

Resumen

Estadísticas de la regresión		
Coeficiente de correlación múltiple	0.8967553	Correlación de Pearson
Coeficiente de determinación R^2	0.8041701	Coeficiente de determinación
R^2 ajustado	0.767452	Coeficiente de determinación ajustado.
Error típico	51.048554	Variación de valor estimado Y*
Observaciones	20	

ANÁLISIS DE VARIANZA

	Grados de libertad	*Suma de cuadrados*	*Promedio de los cuadrados*	*F*	*Valor crítico de F*
Regresión	3	171220.5	57073.491	21.90118203	6.56178E-06
Residuos	16	41695.28	2605.9548		
Total	19	212915.8			

	Coeficientes	*Error típico*	*Estadístico t*	*Probabilidad*	*Inferior 95%*	*Superior 95%*	*Inferior 95.0%*
Intercepción	427.1938	59.60143	7.1675094	2.23764E-06	300.8444175	553.543189	300.844417
Variable X 1	-4.582663	0.772319	-5.933637	2.10035E-05	-6.21990652	-2.94541874	-6.21990652
Variable X 2	-14.83086	4.754412	-3.119389	0.006605963	-24.9097665	-4.7519589	-24.9097665
Variable X 3	6.1010321	4.01212	1.5206504	0.147862484	-2.40428274	14.6063469	-2.40428274

1.-Ecuación de regresión: Y* = a + X1(b1) +X2(b2)+ X3(b3)

Y* = 427.19 -4.58(X1) - 14.83 (X3) + 6.10 (X3)

Como se interpreta esta ecuación:

a = 427.19 ordenada al origen (lugar donde la recta cruza el eje Y), costo base.

b1= - 4.58 es la pendiente de la variable 1 (temperatura Externa) y quiere decir que por cada °F que aumente la temperatura externa, el costo bajará 4,58 dlls (signo negativo la pendiente)

b2 = -14.83 es la pendiente de la variable 2 (aislamiento del ático en pulg) y quiere decir que por cada pulgada que aumente el espesor del aislamiento, el costo de calefacción disminuirá 14.83 dlls (signo negativo la pendiente)

b3 = 6.1 es la pendiente de la variable X3 edad del calentador en uso en años, quiere decir que por cada año de uso que aumente el calentador, aumentará en 6.10 dls de costo del calefacción (signo positivo de la pendiente)

2.-El costo de calefacción para la casa de Filadelfia es de Y* = 427.19-4.58(30°F)-14.83(5")+6.10(10 años)

Y* = 276.64 Dlls, Costo de calefaccion estimado de la casa de Filadelfia.

Prueba de hipótesis: Fcrítica a 3/16 gl y 95% NC = 3.24 (de tablas)

tcrítica a 0.975 y 16 gl = 2.12

Prueba global de "F":

Dado que F calculada es de 21.90 es mucho mayor que Fcrítica = 3.24 se rechaza la Ho y se acepta H1

Es decir, sí existe correlación entre las variables independientes X!, X2 y X3 con la variable dependiente Y.

De igual manera p calculada = 6.56E-06 es mucho menor que Alfa (nivel de significancia) de 0.05 por lo que también se rechaza la Ho y se acepta la H1.

Prueba individual "t":

Para X1 temp. Externa: tcalculada = -5.93 es mayor que tcrítica=2.12; se rechaza la Ho.

Para X2 aislamiento del ático: tcalculada= -3.11 > que tcrítica = 2.12; se rechaza la Ho.

Para X3 edad del calentador: tcalculada= 1.52 < que t crítica =2.12; se acepta la Ho.

Con respecto a p: pX1 =2.1E-05 < alfa = 0.05 , se rechaza la Ho

pX2 = 0.0066 < alfa = 0.05, se rechaza la Ho.

pX3 = 0.1478 > alfa = 0.05 se acepta la Ho.

Esto quiere decir que, las variables independientes X1 y X2 al rechazar la Ho y aceptar la H1 si existe correlación con la variable dependiente Y* Costo de calefacción, que son buenas predictoras del costo, pero no así con X3 (edad del calentador) que se acepta la Ho, es decir no existe buena correlación y por lo tanto esta variable se desecha como buena predictora del costo de calefacción.

La prueba individual "t"; nos va a servir para conocer cuál de las variables independientes son o no buenas predictoras para la variable dependiente Y*, pues F es una prueba global.

NOTA: Para continuar con el problema, hay que repetir todo el cálculo para encontrar la nueva ecuación de regresión, pero ahora con dos variables independientes (Temp y espesor del aislamiento del ático) y se elimina la edad del calentador por ser mala predictora.

Al repetir el procedimiento con solo dos variables independientes, observar que los coeficientes de la ecuación cambian.

Vamos a repetir el proceso, pero ahora solo con dos variables independientes (X1 y X2)

Resumen

Estadísticas de la regresión

Coeficiente de correlación múltiple	0.8808337
Coeficiente de determinación R^2	0.775868
R^2 ajustado	0.7494996
Error típico	52.982366
Observaciones	20

ANÁLISIS DE VARIANZA

	Grados de libertad	*Suma de cuadrados*	*Promedio de los cuadrados*	*F*	*Valor crítico de F*
Regresión	2	165194.5	82597.261	29.42408379	3.01497E-06
Residuos	17	47721.23	2807.1311		
Total	19	212915.8			

	Coeficientes	*Error típico*	*Estadístico t*	*Probabilidad*	*Inferior 95%*	*Superior 95%*	*Inferior 95.0%*
Intercepción	490.28593	44.40984	11.040028	3.56342E-09	396.5893463	583.982507	396.589346
Variable X 1	-5.149884	0.701887	-7.337201	1.16062E-06	-6.63073548	-3.66903245	-6.63073548
Variable X 2	-14.71815	4.933918	-2.983055	0.00835087	-25.1278063	-4.30849064	-25.1278063

Observemos los cambios en los coeficientes y en los componentes de la ecuación.

$$Y^* = 490.2859 - 5.1498(X1) - 14.718(X2)$$

Uso de una variable cualitativa:

Ahora vamos a agregar una nueva columna de datos (X4) que denominaremos Garage. Es una variable de escala Nominal porque diremos si una casa tiene o no calefacción en el Garage.

Cost (Y)	Temp (X1)	Insul (X2)	**Garage(X4)**
250	35	3	**0**
360	29	4	**1**
165	36	7	**0**
43	60	6	**0**
92	65	5	**0**
200	30	5	**0**
355	10	6	**1**
290	7	10	**1**
230	21	9	**0**
120	55	2	**0**
73	54	12	**0**
205	48	5	**1**
400	20	5	**1**
320	39	4	**1**
72	60	8	**0**
272	20	5	**1**
94	58	7	**0**
190	40	8	**1**
235	27	9	**0**
139	30	7	**0**

Codificaremos la palabra "no" y "sí"

No tiene calefacción, código = 0

Sí tiene calefacción, código =1

Repetimos el procedimiento:

Resumen

Estadísticas de la regresión	
Coeficiente de correlación múltiple	0.9326512
Coeficiente de determinación R^2	0.8698383
R^2 ajustado	0.845433
Error típico	41.618416
Observaciones	20

ANÁLISIS DE VARIANZA

	Grados de libertad	*Suma de cuadrados*	*Promedio de los cuadrados*	*F*	*Valor crítico de F*
Regresión	3	185202.3	61734.09	35.6413338	2.58644E-07
Residuos	16	27713.48	1732.0926		
Total	19	212915.8			

	Coeficientes	*Error típico*	*Estadístico t*	*Probabilidad*	*Inferior 95%*	*Superior 95%*	*Inferior 95.0%*
Intercepción	393.66568	45.00128	8.7478765	1.70718E-07	298.2672182	489.064143	298.267218
Variable X 1	-3.962847	0.652657	-6.071864	1.61754E-05	-5.34641919	-2.57927523	-5.34641919
Variable X 2	-11.33395	4.001531	-2.832404	0.01201021	-19.8168214	-2.85108608	-19.8168214
Variable X 3	77.432105	22.78282	3.3987056	0.003670177	29.13467978	125.729529	29.1346798

Ecuación de regresión con nueva variable Garage:

$$Y^* = 393.665 - 3.9628\ (X1) - 11.33(X2) + 77.43(X3)$$

Ahora observemos que tanto para la prueba global como para la prueba individual, todas las variables independientes son buenas predictoras de Y*, porque en todas es rechazada la Ho y aceptada la H1.

Fcalculada=35.64>Fcrítica = 3.24 se rechaza la Ho. P=2.58E-07<alfa=0.05 se rechaza la Ho.

tX1 calculada = -6.07>tcrítica =2.12 se rechaza la Ho.

tX2 calculada= -2.83>tcrítica =2.12 se rechaza la Ho y

tX4calculada= 3.398 tcrítica =-2.12 se rechaza la Ho.

pX1=1.61E-05<alfa=0.05 se rechaza la Ho

pX2=0.0120<alfa=0.05 se rechaza la Ho.

pX4 = 0.00367 < alfa=0.05 se rechaza la Ho.

Ahora demos un repaso a la definición de los coeficientes que se encuentran en el resumen, Para el primer problema donde tomamos a tres variables independientes (temp X1, Aislante del ático X2 y Edad del calentador en años X3):

0.8967 es coeficiente de correlación de Pearson y se define como la relación que existe entre las variables independientes X1, X2 y X3 y la variable dependiente Y.

0.80417 es el coeficiente de determinación múltiple y se define como el porcentaje de variación de la variable dependiente Y, explicada o determinada por el conjunto de variables independientes X1, X2 y X3 (para este caso). La diferencia (1-0.80417) = 0.1958 es la parte no explicada de la variable Y, que se puede deber a error de muestreo, a faltante de otras variables independientes no tomadas en cuenta o simplemente se desconoce.

0.76745 Es el coeficiente de determinación ajustado y se utiliza cuando existen muchas variables independientes (más de 4) que por el solo hecho de ser muchas pueden dar un valor muy alto de F, pues en el cálculo de éste estadístico de prueba, el término SSE está en el denominador de la fórmula y por lo tanto incrementa dicho valor de F.

51.0485 es el valor del error típico que se puede tomar como desviación estándar, y es el valor que se incrementa y disminuye del valor estimado de Y*, es el intervalo (± una vez la desviación estándar) donde caen 2/3 partes de todos los valores estimados de Y (tomado de la Teoría de limite central).

Minitab.

Eliminamos la escritura de la Hoja de cálculo.

COMANDOS:

Cargar hoja de trabajo – estadística – regresión – regresión – ajustar modelo de regresión – respuestas: cargar columna de Y (C1) – Predictores: cargar columnas de X1, X2 y X3 (C2, C3 y C4) – en opciones determinar el N.C. deseado – aceptar – aceptar. listo

análisis de regresión: Cost (Y) vs. Temp (X1), Insul (X2), Age (X3)

Análisis de Varianza

Fuente	GL	SC Ajust.	MC Ajust.	Valor F	Valor p
Regresión	3	171220	57073	21.90	0.000
Temp (X1)	1	91751	91751	35.21	0.000
Insul (X2)	1	25357	25357	9.73	0.007
Age (X3)	1	6026	6026	2.31	0.148
Error	16	41695	2606		
Total	19	212916			

Resumen del modelo

S	R-cuad.	R-cuad. (ajustado)	R-cuad. (pred)
51.0486	80.42%	76.75%	68.72%

Coeficientes

Término	Coef	EE del coef.	Valor T	Valor p	VIF
Constante	427.2	59.6	7.17	0.000	
Temp (X1)	-4.583	0.772	-5.93	0.000	1.32
Insul (X2)	-14.83	4.75	-3.12	0.007	1.01
Age (X3)	6.10	4.01	1.52	0.148	1.31

Ecuación de regresión

Cost (Y) = 427.2 - 4.583 Temp (X1) - 14.83 Insul (X2) + 6.10 Age (X3)

Los resultados son los mismos que con Excel:

Se tiene la ecuación de regresión: Y* = 427.2 -4.583(X1) - 14.83(X2) + 6.10 (X3)

Fcalculada= 21.90 >Fcrítica =3.24 se rechaza la Ho. p = 0.00 < alfa= 0.05 se rechaza la Ho.

tX1 = -5.93 y tX2 = -3.12 las dos > t crítica = 2.12 se rechaza la Ho

tX3 = 1.52 es < que tcrítica = 2.12 se acepta la Ho. no es buena predictora la edad del calentador

pX1 = 0.00 < que alfa = 0.05 se rechaza la Ho. Sí es buena predictora.

pX2 = 0.007 < que alfa = 0.05 se rechaza la Ho. Sí es buena predictora.

pX3 = 0.148 >que alfa=0.05 se acepta la Ho. Es mala predictora

En conclusión: la temp. Externa y el espesor del aislamiento son buenas predictoras del valor de Y (consto de calefacción).

No así X3 edad del calentador que es aceptada la Ho no es buen predictor del costo de calefacción Y.

PROBLEMA 5.9. REGRESIÓN LINEAL MÚLTIPLE. IMPUESTOS.

Excel.

Mes	X1	X2	X3	Y
Enero	45	16	71	29
Febrero	42	14	70	24
Marzo	44	15	72	27
Abril	45	13	71	25
Mayo	43	13	75	26
Junio	46	14	74	28
Julio	44	16	76	30
Agosto	45	16	69	28
Septiembre	44	15	74	28
Octubre	43	15	73	27

El Servicio Interno de Contribuciones (IRS: Internal Revennue Service) está tratando de estimar la cantidad mensual de impuestos no pagados descubierto por su departamento de auditorías. En el pasado, el IRS estimaba esta cantidad con base en el número esperado de horas de trabajo de auditorías de campo. En los últimos años, sin embargo, las horas de trabajo de auditorías de campo se han convertido en un pronosticador errático de los impuestos no pagados reales.

Como resultado, la dependencia está buscando otro factor para mejorar la ecuación de estimación. El Dpto. de auditorías tiene un registro del número de horas que usa sus computadoras para detectar impuestos no pagados.

La pregunta es: ¿podría combinar esta información con los datos referente a las horas de trabajo de auditoria de campo y obtener una ecuación más precisa para los impuestos no pagados descubiertos cada mes?

A continuación se presentan esos datos de los últimos 10 meses.

X1= Hrs. De trabajo de auditorías de campo (2 ceros omitidos)

X2= Hrs. En computadoras (2 ceros omitidos)

Y= Impuestos reales no pagados descubiertos (mlls de dls)

Posteriormente se agrega una tercera columna (X3), es una recompensa ofrecida a informantes de contribuyentes que no pagan impuestos: X3: Recompensa a informantes (tres ceros omitidos)

Nivel de Confianza = 95%

Regla de decisión: si Fcalculada es mayor que Fcrítica se rechaza la Ho. Cae en la zona de rechazo de la Ho.

Si p es mayor que alfa (nivel de significancia) se acepta la Ho. Cae en la zona de aceptación de la Ho.

COMANDOS:

Datos – análisis de datos – regresión – aceptar – rango y de entrada: cargar la columna de e (y) – rango x de entrada: cargar los datos de la columna x1 (auditoría de campo) y x2 (horas de computadora) – nivel de confianza deseado – rango de salida: con el cursor seleccionar una celda vacía donde se desea se anoten los cálculos.

Ho: $\beta 1 = \beta 2 = \beta 3 = 0$.que las pendientes es de 0, que no hay correlación o que son malas predictoras de Y.

H1 : $\beta 1 \neq 0$, $\beta 2 \neq 0$,$\beta 3 \neq 0$ que las pendientes son diferentes de cero, que sí hay correlación o que son buenas predictoras de Y.

Primero haremos el análisis con X1 y X2:

Resumen

Estadísticas de la regresión	
Coeficiente de correlación múltiple	0.85377
Coeficiente de determinación R^2	0.72892
R^2 ajustado	0.65147
Error típico	1.07064
Observaciones	10

ANÁLISIS DE VARIANZA

	Grados de libertad	*Suma de cuadrados*	*Promedio de los cuadrados*	*F*	*Valor crítico de F*
Regresión	2	21.57613	10.78806	9.41147107	0.01037106
Residuos	7	8.023873	1.146268		
Total	9	29.6			

	Coeficientes	*Error típico*	*Estadístico t*	*Probabilidad*	*Inferior 95%*	*Superior 95%*	*Inferior 95.0%*
Intercepción	-13.82	13.3233	-1.037253	0.33411481	45.3242267	17.6849694	-45.3242267
Variable X 1	0.56366	0.303274	1.858586	0.10543015	-0.1534683	1.28078925	-0.1534683
Variable X 2	1.09947	0.313139	3.511123	0.0098445	0.35901339	1.8399256	0.35901339

Ecuación de regresión: Y* = -13.82 + 0.563 (X1) + 1.09947(X2).

Estadístico de prueba (global) F:

Fcrítica a 0.92 con 2/7 gl = 4.74

tcrítica a 0.975 con 7 gl = 2.36

Dado que F calculada = 9.41 >que Fcrítica = 4.74 se rechaza la Ho.

p=0.01037 < que alfa = 0.05 se rechaza la Ho.

Estadístico de prueva individual t:

tX1calculada = 1.85 <tcrítica = 2.36, se acepta la Ho.

PX1 =0.105>alfa=0.05 se acepta la Ho.

tX2 calculada = 3.51 >tcrítica =2.36 se rechaza la Ho.

pX2 = 0.0098 <alfa= 0.05 se rechaza la Ho.

Por lo anterior se confirma que las horas de auditoría de campo no son buenas predictoras de los impuestos reales no pagados por los contribuyentes.

Ahora vamos a agregar una tercera columna de información que es X3 (pago a informantes) y procedemos a calcular de nuevo sin la X1 (horas de auditorías de campo):

Mes	X2	X3	Y
Enero	16	**71**	29
Febrero	14	**70**	24
Marzo	15	**72**	27
Abril	13	**71**	25
Mayo	13	**75**	26
Junio	14	**74**	28
Julio	16	**76**	30
Agosto	16	**69**	28
Septiembre	15	**74**	28
Octubre	15	**73**	27

Resumen

Estadísticas de la regresión	
Coeficiente de correlación múltiple	0.91303
Coeficiente de determinación R^2	0.83363
R^2 ajustado	0.7861
Error típico	0.83875
Observaciones	10

ANÁLISIS DE VARIANZA

	Grados de libertad	*Suma de cuadrados*	*Promedio de los cuadrados*	*F*	*Valor crítico de F*
Regresión	2	24.67549	12.33774	17.5376058	0.00187824
Residuos	7	4.924515	0.703502		
Total	9	29.6			

	Coeficientes	*Error típico*	*Estadístico t*	*Probabilidad*	*Inferior 95%*	*Superior 95%*	*Inferior 95.0%*
Intercepción	-20.133	9.996407	-2.014023	0.08386653	43.7707439	3.50474822	43.7707439
Variable X 1	1.28756	0.242474	5.310102	0.00111113	0.71420242	1.86092339	0.71420242
Variable X 2	0.3918	0.123689	3.167653	0.0157593	0.09932595	0.68428297	0.09932595

Y* = -20.133 + 1.287 (X2) + 0.3918 (X3)

Fcalculada = 17.53 > Fcrítica = 4.74 se rechaza la Ho; p=0.00187<alfa0.05 tambien se rechaza la Ho

tX2 calculada = 5.31 > tcrítica =2.36 se rechaza la Ho; pX2 calculada = 0.0011<alfa =0.05 se rechaza la Ho.

tX3 cal. = 3.167 > tcrítica = 2.36 se rechaza la Ho; pX3 cal. = 0.0157 < alfa =0.05 se rechaza la Ho.

Observemos que la ecuación tiene un cambio importante y las condiciones son otras.

En todos los casos la Ho es rechazada y aceptada la H1.

Por lo que las horas de computadora X2 y el pago a informantes X3 son dos variables independientes que se pueden considerar buenas predictoras de Y impuestos no pagados por los contribuyentes.

Si tenemos la ecuación de ésta última prueba que es: Y* = -20.133 + 1.287(X2) + 0.3918(X3); vamos a interpretar los coeficientes de dicha ecuación:

a = -20.133 es la ordenada al origen y es la base del cálculo, es decir, si X2 y X3 tienen pendiente cero, el valor de Y* = -20.133, o se a que se determinaría la cantidad de $20´133,000 de impuestos no pagados.

Por cada 100 horas de computadora que se aumente (no olvidar que a los datos se le omitieron dos ceros o dígitos a X2), se incrementaran en 1´287,560 dlls de impuestos no pagados (se le omitieron seis dígitos o ceros a Y, puesto que estan dados en millones de dólares).

De igual manera, por cada mil dólares que se le aumente al pago a informantes (se omitieron tres ceros o dígitos a X3), se incrementaran 391,800 dlls re recaudación de impuestos no pagados (porque se le omitieron seis dígitos o ceros x Y).

Minitab.

COMANDOS:

Cargar hoja de trabajo – regresión – regresión – ajustar modelo de regresión – respuestas cargar columna de Y – predictores cargar columnas de X1 y X2 – opciones: 95% - NC:95 – aceptar – aceptar – listo.

Análisis de regresión: Y vs. X2, X3

```
Análisis de Varianza

Fuente     GL  SC Ajust.  MC Ajust.  Valor F  Valor p
Regresión   2     24.675    12.3377    17.54    0.002
  X2        1     19.837    19.8368    28.20    0.001
  X3        1      7.059     7.0590    10.03    0.016
Error       7      4.925     0.7035
Total       9     29.600


Resumen del modelo

                      R-cuad.  R-cuad.
       S  R-cuad.  (ajustado)   (pred)
0.838750   83.36%      78.61%   68.13%


Coeficientes

                 EE del
Término     Coef  coef.  Valor T  Valor p   VIF
Constante  -20.1   10.0    -2.01    0.084
```

```
X2          1.288    0.242      5.31     0.001   1.01
X3          0.392    0.124      3.17     0.016   1.01


Ecuación de regresión

Y = -20.1 + 1.288 X2 + 0.392 X3
```

Observemos que es la misma ecuación de regresión y los resultados de la prueba de hipótesis son los mismos, por lo tanto las conclusiones son también las mismas.

PROBLEMA 5.10. REGRESIÓN LINEAL MÚLTIPLE. CERDITOS:

Hoja de Cálculo.

X1	X2	Y
39	8	7
52	6	6
49	7	8
46	12	10
61	9	9
35	6	5
25	7	3
55	4	4

Un productor de alimentos balanceados para cerdos desea determinar qué relación existe entre la edad de un cerdo cuando empieza a recibir un complement alimenticio de reciente creación, el peso inicial del animal y el aumento de peso en un periodo de una semana con el complemento alimenticio.

X1= peso inicial en libras

X2= Edad inicial en semanas

Y= Aumento de peso en libras.

¿Cuánto podemos esperar que un cerdo aumente de peso en una semana con el complement alimenticio si tiene 9 semanas y pesa 48 Lbs.

Emplear un nivel de significancia de 0.05

PREGUNTAS:

1.- Determinar las hipótesis nula y alternativa

2.- ¿Cuál es la variable dependiente y cuales las independientes?

3.- ¿Cuál es la ecuación de regresión para este caso?

4.- Determinar el valor crítico de F y de t

5.- ¿Cómo interpreta el valor de P para F?

6.- ¿Cómo interpreta el valor del error típico?

7.- ¿A qué conclusión llega respecto de las pruebas de F y t?

8.- ¿Cómo interpreta los valores de p para "t" ?

COMANDOS:

Cargar hoja de trabajo – regresión – regresión – ajustar modelo de regresión – respuestas cargar columna de Y – predictores cargar columnas de X1 y X2 – opciones:95% - NC:95 – aceptar – aceptar – listo.

Análisis de regresión: Y vs. X1, X2

Análisis de Varianza

Fuente	GL	SC Ajust.	MC Ajust.	Valor F	Valor p
Regresión	2	37.009	18.5046	18.54	0.005
X1	1	10.520	10.5202	10.54	0.023
X2	1	25.929	25.9294	25.98	0.004
Error	5	4.991	0.9981		
Total	7	42.000			

Resumen del modelo

S	R-cuad.	R-cuad. (ajustado)	R-cuad. (pred)
0.999073	88.12%	83.36%	65.83%

Coeficientes

Término	Coef	EE del coef.	Valor T	Valor p	VIF
Constante	-4.19	1.89	-2.22	0.077	
X1	0.1048	0.0323	3.25	0.023	1.00
X2	0.807	0.158	5.10	0.004	1.00

Ecuación de regresión

Y = -4.19 + 0.1048 X1 + 0.807 X2

Y*= -4.19+0.1048(48)+0.807(9) = 8.09 lbs. + - 0.9981 y ahí caen el 68% de los pronósticos a 95% de N:C: X1= peso en lbs. de los cerdos; X2 = edad en semanas de los cerditos

1.- Ho: que no hay incremento significativo con el complemento alimenticio
H1: que sí hay incremento con el complemento alimenticio.

2.-La variable dependiente es Y peso final de los cerditos.
las variables independientes son : peso en lbs al inicio (X1) y edad en semanas al inicio (X2)

3.- Y* = -4.19 + 0.1040(X1) + 0.807(X2)

4.-Fcrítica a 95%, 2 y 5 gl = 5.79; crítica a 5 gl y 0.975 = 2.57; Alfa = 0.05
Fcal=18.54 >Fcrítica= 5.79 Se rechaza la Ho. tcalculada X1 =3.25 >tcrítica=2.57 Se rechaza la Ho

5.-p es la probabilidad de aceptar o rechazar la H. para éste caso P= 0.005< alfa=0.05 se rechaza la Ho.

6.-El error típico tiene un valor de 0.9981 y es el más-menos como intervalo donde caen el 68%b de los casos estimados

7.-Dado que Fcalculada = 18.54>que Fcrítica=5.79 Se rechaza la Ho. Para p = 0.005 <alfa=0.05 se rechaza la Ho.

F es la prueba global.

La prueba individual para tX1 = 3.25>tcrítica =2.57 se rechaza la Ho. igual para p=0.023< alfa=0.05 se rechaza la Ho.
para tX2 =5.10>tcrítica = 2.57 se rechaza la Ho. Igual para p=0.004< alfa =0.05 se rechaza la Ho.

8. p es la probabilidad de que la Ho de tX1 (0.023) sea cierta. Igualmente para la probabilidad de que la Ho de tX2(0.004) sea cierta. Pero en ambas la probabilidad de X1 y X2 de t es menor que alfa= 0.05 (5%) por lo que se rechaza la Ho por ser menor el área de la curva, es decir cae en la zona de rechazo de la Ho.

X1	X2	Y	Problema 5.10. **Excel**. Regresión Lineal Múltiple. Cerditos:
39	8	7	Un productor de alimentos balanceados para cerdos
52	6	6	desea determinar que relación existe entre la edad de
49	7	8	un cerdo cuando empieza a recibir un complemento
46	12	10	alimenticio de reciente creación, el peso inicial del
61	9	9	animal y el aumento de peso en un periodo de una sema-
35	6	5	na con el complemento alimenticio.
25	7	3	X1= peso inicial en libras
55	4	4	X2= Edad inicial en semanas
			Y= Aumento de peso en libras.

¿Cuánto podemos esperar que un cerdo aumente de peso en iuna semana con el complemento alimenticio si tenía 9 semanas y pesaba 48 libras?
Emplear un nivel de significancia de 0.05

PREGUNTAS:
1.- Determinar las hipótesis nula y alternativa?
2.-Cuál es la variable dependiente y cuales las independientes?
3.-Cuál es la ecuación de regresión para este caso?
4.- Determinar el valor crítico de F y de t ?
5.-Como interpreta elvalor de P para F?
6.-Como interpreta el valor del error típico?
7.-A qué conclusión llega respecto de las pruebas de F y t ?
8.-Como interpreta los valores de p para "t" ?

COMANDOS: Datos-Análisis de datos-Regresión-Aceptar-rango de entrada (seleccionar columna de Y) - Rango X entrada (seleccionar columnas X1 y X2) Rango de salida (colocar cursor en celda vacia) - Aceptar.

Resumen

Estadísticas de la regresión	
Coeficiente d	0.93870818
Coeficiente d	0.88117304
R^2 ajustado	0.83364226
Error típico	0.99907279
Observacione	8

	Grados de libertad	Suma de cuadrados	Promedio de los cuadrados	F	Valor crítico de F
Regresión	2	37.0092678	18.5046339	18.5389971	0.00486729
Residuos	5	4.99073219	0.99814644		
Total	7	42			

	Coeficientes	Error típico	Estadístico t	Probabilidad	Inferior 95%	Superior 95%	Inferior 95.0%
Intercepción	-4.1917094	1.88811919	2.22004491	0.0771239	-9.04527431	0.6618555	-9.04527431
Variable X 1	0.10483433	0.03229147	3.24650208	0.02278415	0.02182646	0.18784219	0.02182646
Variable X 2	0.80650253	0.15823657	5.09681501	0.00378031	0.39974248	1.21326259	0.39974248

RESPUESTAS:

1.-Ho: μ=0 no hay correlación, sí la hay, se debe al azar. También se puede expresar que el complemento alimenticio no incrementa el peso de los cerditos más que el alimento tradicional.

H1: μ1‡ 0 sí hay correlación entre X1 y X2 con Y, es decir entre el peso inicial, la edad en semanas con el peso final de los cerditos. También se puede decir a un nivel de significancia del 95% que el complemento alimenticio sí incrementa significativamente el peso de los cerditos.

2.-Las variables independientes son: Peso inicial en libras (X1) y Edad inicial en semanas (X2), y la variable dependiente es el aumento de peso en libras (Y*)

3.-Y*= -4,1917094+0.10483433(X1)+0.80650253(X2)

4.-Valor crítico de F a N.C.=95% y 2 y 5 grados de libertad=5.79. para "t" a 0.975 de N.C. y n-(k+1)=8-(2+1)=5 es de 2.57

5.- p es la probabilidad de que la Ho se cierta o no. se contrasta con α=0.05 Nivel de significancia p=0.004867 < que alfa =0.05. por lo tanto se rechaza la Ho.

6.-Se interpreta como el error de predicción y a cada valor de Y* se le ±0.99907279 y ahí cae el 68% de los valores de Y (teoría de limites central)

7.-F es la prueba global y la Ho se rechaza ya que Fcalculada=18.53 > que Fcrítica=5.79 t es la prueba individual y para X1 se rechaza la Ho porque tcalc=3.2465 >tcrít=2.57 para X2 se rechaza la Ho porque tcalc.=5.0968>tcrít=2.57

8.. La regla de decisión es que si p>que α se acepta la Ho. De lo contrario se rechaza y en este caso para F es de 0.004867 es menor que α = 0,05, así que es muy baja la probabilidad de aceptar la Ho.

Para las pruebas individuales donde p de t de X1=0.022784 y t2 de X2=0.00378 son menores que α=0.05 por lo que se rechaza la Ho.

Al rechazar la Ho tanto en F como para t y para los respectivos valores de p, es aceptada la H1 y se concluye que el alimento incrementa los pesos finales en una semana de los cerditos; cabe aclarar que esta prueba solo nos indica el incremento de los cerditos que habría que compararlo con un lote de cerditos alimentados con el alimento tradicional sin el complemento para que el granjero decida si la diferencia en peso le es benéfico económicamente.

Con respecto a la pregunta del problema, se tiene:

$Y^* = -4.19+0.1048(X1)+0.8065(X2)$

$Y^* = -4.19+0.1048\ (48)+0.8065\ (9) = 8.09$

El puerco en cuestión ganaría 8.09 lbs. $\pm$ 0.999 lbs . Intervalo donde estarán el 68% de los casos.

PROBLEMA 5.11. REGRESIÓN LINEAL MÚLTIPLE. ARRESTOS POLICIACOS.

Hoja de Trabajo.

Y	X1	X2	X3
390.6	68	81.6	4.3
504.3	94	75.1	3.9
628.4	125	97.3	5.6
745.6	175	123.5	8.7
585.2	113	118.4	11.4
450.3	82	65.4	9.6
327.8	46	61.6	12.4
260.5	32	54.3	18.3
477.5	89	97.4	4.6
389.8	67	82.4	6.7
312.4	47	56.4	8.4
367.5	59	71.3	7.6
474.4	61	67.4	9.8
494.6	87	96.3	11.3
487.5	92	86.4	4.7

En las estaciones de policias de USA están interesados en predecir los números de arrestos esperados que deben procesar cada mes para programar mejor a los empleados de oficina. En los datos históricos, el número promedio de arrestos (Y) cada mes tiene influencia el número de oficiales de la fuerza policiaca (X1), la población de la ciudad en miles (X2) y el porcentaje de personas desempleadas en la ciudad (X3). Los datos de estos factores en15 ciudades se presentan en la tabla adjunta.

1.-Determinar la ecuación de regresión?

2.-Qué porcentaje de la variación total en el número de arrestos (Y) explica esta ecuación?

3.-El Departamento de policia de Chapelboro deseapronosticar el número de arrestos mensuales, si esta población tiene 75,000 habitantes, 82 elementos en su fuerza policiaca y un porcentaje de desempleo de 10.5 % . ¿Cuántos arrestos se pronostica para cada mes?

Utilizar un nivel de confianza del 95%

Minitab.

COMANDOS:

Cargar hoja de trabajo – regresión – regresión – ajustar modelo de regresión – respuestas cargar columna de Y – predictores cargar columnas de X1, X2 y X3 – opciones:95% -NC:95 – aceptar – aceptar – listo.

Análisis de regresión: Y vs. X1, X2, X3

Análisis de Varianza

Fuente	GL	SC Ajust.	MC Ajust.	Valor F	Valor p
Regresión	3	215137	71712.4	63.87	0.000
X1	1	43202	43201.7	38.48	0.000
X2	1	86	85.6	0.08	0.788
X3	1	0	0.0	0.00	0.997

```
Error      11      12350     1122.7
Total      14     227487
```

Resumen del modelo

```
                    R-cuad.  R-cuad.
      S  R-cuad.  (ajustado)   (pred)
33.5070   94.57%      93.09%   88.80%
```

Coeficientes

```
                 EE del
Término     Coef   coef.  Valor T  Valor p   VIF
Constante  166.2    49.3     3.37    0.006
X1         3.313   0.534     6.20    0.000  4.62
X2         0.246   0.891     0.28    0.788  4.40
X3          0.01    2.46     0.00    0.997  1.14
```

Ecuación de regresión

Y = 166.2 + 3.313 X1 + 0.246 X2 + 0.01 X3

Ajustes y diagnósticos para observaciones poco comunes

```
                          Resid
Obs      Y  Ajuste  Resid   est.
 13  474.4   385.0   89.4   2.82  R
```

Residuo grande R

RESPUESTAS:

1. - Ecuación de regresión Y*= 166.2 + 3.313(X1) + 0.246(X2) + 0.01(X3).
2.-Es el coeficiente de regresión = 94.57 %
3. - Y* = 166.2 + 3.313(82) + 0.246 (75) + 0.01 (10.5) = 456.421 se esperan 456 arrestos para Chapelboro por mes.

Y	X1	X2	X3	
390.6	68	81.6	4.3	Problema 5.11. **Excel**.Regresión L. Múltiple. Arrestos policia-
504.3	94	75.1	3.9	cos. En las estaciones de policias de USA estan inte-
628.4	125	97.3	5.6	resados en predecir los números de arrestos espe-
745.6	175	123.5	8.7	rados que deben procesar cada mes para programar
585.2	113	118.4	11.4	mejor a los empleados de oficina. En los datos his-
450.3	82	65.4	9.6	tóricos, el número promedio de arrestos (Y) cada
327.8	46	61.6	12.4	mes tiene influencia el número de oficiales de la
260.5	32	54.3	18.3	fuerza policiaca (X1), la población de la ciudad en
477.5	89	97.4	4.6	miles (X2) y el porcentaje de personas desemplea-
389.8	67	82.4	6.7	das en la ciudad (X3). Los datos de estos factores en
312.4	47	56.4	8.4	15 ciudades se presentan en la tabla adjunta.
367.5	59	71.3	7.6	1.-Determinar la ecuación de regresión?
474.4	61	67.4	9.8	2.-Qué porcentaje de la variación total en el núme-
494.6	87	96.3	11.3	ro de arrestos (Y) explica esta ecuación?
487.5	92	86.4	4.7	3.-El Departamento de policia de Chapelboro desea
				pronosticar el número de arrestos mensuales, si
				esta población tiene 75,000 habitantes, 82 elemen-
				tos en su fuerza policiaca y un porcentaje de de --
				sempleo de 10.5 % . Cuantos arrestos se pronostica
				para cada mes?
				Utilizar un nivel de confianza del 95%

COMANDOS: Datos- análisis de datos-regresión- aceptar-rango de entrada Y (seleccionar datos de columna Y) -Rango Xde entrada(seleccionar catos de columnas X1, X2 y X3)-rango de salida (colocar el ciursor en celda vacia deseada).aceptar.

Resumen

Estadísticas de la regresión	
Coeficiente (	**0.97247703**
Coeficiente (	**0.94571157**
R^2 ajustado	0.93090564
Error típico	33.5070177
Observacion	15

ANÁLISIS DE VARIANZA

	Grados de liberta	*na de cuadra*	*io de los cua*	*F*	*alor crítico de F*
Regresión	3	215137.273	71712.424	**63.873815**	3.0286E-07
Residuos	11	12349.9226	1122.7202		
Total	14	227487.196			

	Coeficientes	*Error típico*	*Estadístico t*	*Probabilidad*	*Inferior 95%*	*Superior 95%*	*nferior 95.0%*
Intercepción	**166.23397**	49.2689747	3.3740091	0.00620803	57.7936879	274.674252	57.7936879
Variable X 1	**3.31270601**	0.5340333	**6.2031825**	**6.686E-05**	2.13730665	4.48810538	2.13730665
Variable X 2	**0.24595389**	0.8906065	**0.2761645**	**0.78754136**	-1.7142578	2.20616558	-1.7142578
Variable X 3	**0.0107561**	2.45797238	**0.004376**	**0.99658682**	-5.39920462	5.42071682	-5.39920462
RESPUESTAS:							
1.- Ecuación	Y* = a +X1b1+X2b2+X3b3		Y* = 166.23+3.31(X1) +0.246(X2) + 0.0107(X3)				
2.- Coeficien	**0.94571157**	Es el coeficiente de determoinación.			94.57%		
3.-	Y* = 166.23+3.31(82) +0.246(75) + 0.0107(10.5) = 456.46						

Se pronostican 456 arrestos para Chapelboro.

ESTADÍSTICA INFERENCIAL II

UNIDAD I.

Esta Unidad es la misma que la Unidad V de Estadística Inferencial I.

UNIDAD II.

SERIES DE TIEMPO

SERIES DE TIEMPOS

Una serie de tiempo es un grupo de datos registrados durante un periodo de tiempo (semanal, trimestral o anual).

Un **análisis de la historia**, que es una serie de tiempo, es útil para que la administración tome decisiones hoy y planee con base en una predicción, o proyección de largo plazo. (Suponiendo que los patrones pasados continuaran en el futuro).

* Las proyecciones de largo plazo son esenciales a fin de dar tiempo suficiente para que los departamentos de compras, manufacturas, ventas, finanzas y otros de la compañía elaboren planes para construir nuevas plantas, solicitar financiamiento, desarrollar productos nuevos y métodos de ensamble innovadores.

*La competencia por el dinero de los consumidores, la presión para obtener utilidades para los accionistas, el deseo de obtener mayor participación en el mercado y las ambiciones de los ejecutivos son algunas **fuerzas de motivación de negocios.**

¨Una proyección no es más que una declaración de los objetivos de la administración¨.

Se trata pues, de proyectar eventos futuros.

COMPONENTES DE UNA SERIE DE TIEMPO:

Consta de 4 componentes:

- Tendencia
- Variación cíclica
- Variación estacional
- Variación irregular.

TENDENCIA SECULAR

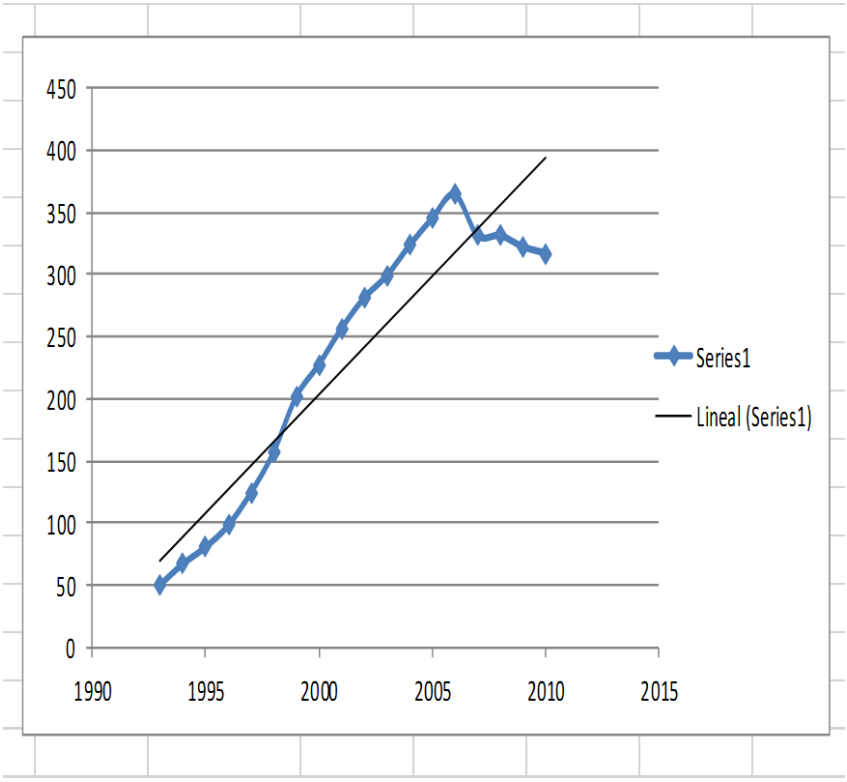
Dirección uniforme de una serie de tiempo de largo plazo.

PROBLEMA 2.1: VARIACIONES.

Ejemplo: Home Depot se fundó en 1978, y es el minorista más grande de los Estados Unidos en artículos para mejorar el hogar.

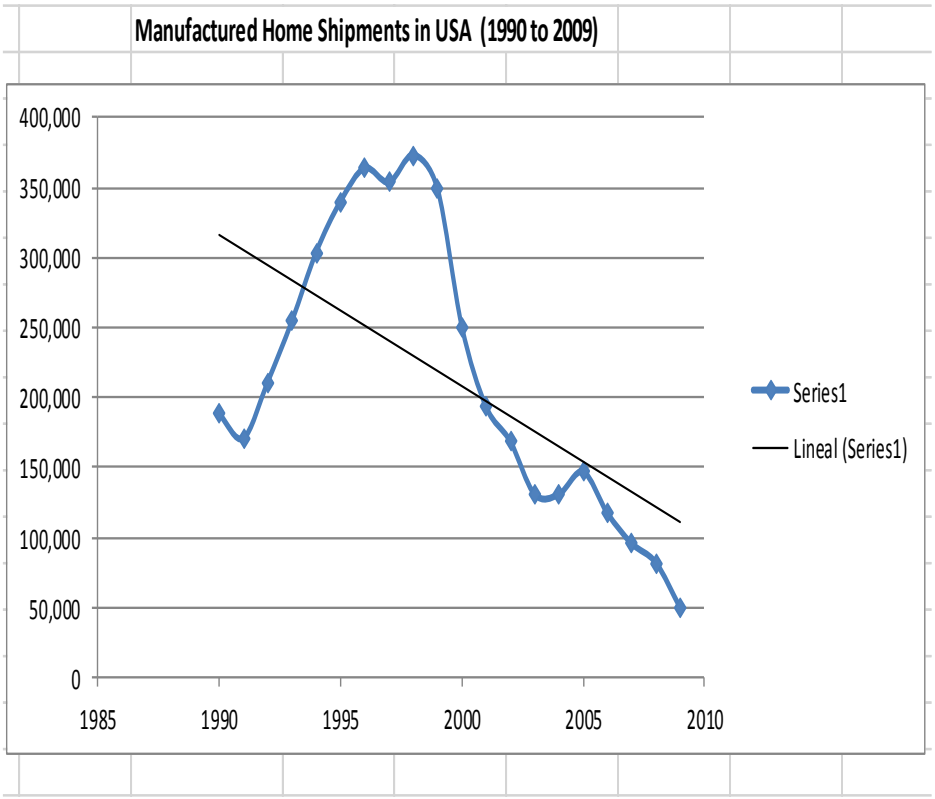
En la siguiente gráfica se muestra el Número de empleados y se puede observar el crecimiento, que aumento drásticamente en los primeros 15 años. De 50,000 empleados en 1993 a 364,000 para 2006. Desde entonces, el número de empleados a disminuido a 317,000 para 2010.

	YEAR	associates (miles)
1	1993	50.6
2	1994	67.3
3	1995	80.8
4	1996	98.1
5	1997	124.4
6	1998	156.7
7	1999	201.4
8	2000	227.3
9	2001	256.3
10	2002	280.9
11	2003	298.8
12	2004	323.1
13	2005	344.8
14	2006	364
15	2007	331
16	2008	331
17	2009	322
18	2010	317



El número de casa prefabricadas construidas en Estados Unidos presento un aumento uniforme de 1990 a 1996, luego permaneció casi igual hasta 1999, cuando el número empeso a declinar. En 2002 el número era menor que en 1990 y continuó declinando hasta 2009.

	year	shipments
1	1990	188,172
2	1991	170,713
3	1992	210,787
4	1993	254,276
5	1994	303,392
6	1995	339,601
7	1996	363,411
8	1997	353,377
9	1998	372,843
10	1999	348,671
11	2000	250,550
12	2001	193,229
13	2002	168,491
14	2003	130,815
15	2004	130,748
16	2005	146,800
17	2006	117,300
18	2007	95,700
19	2008	81,900
20	2009	49,800

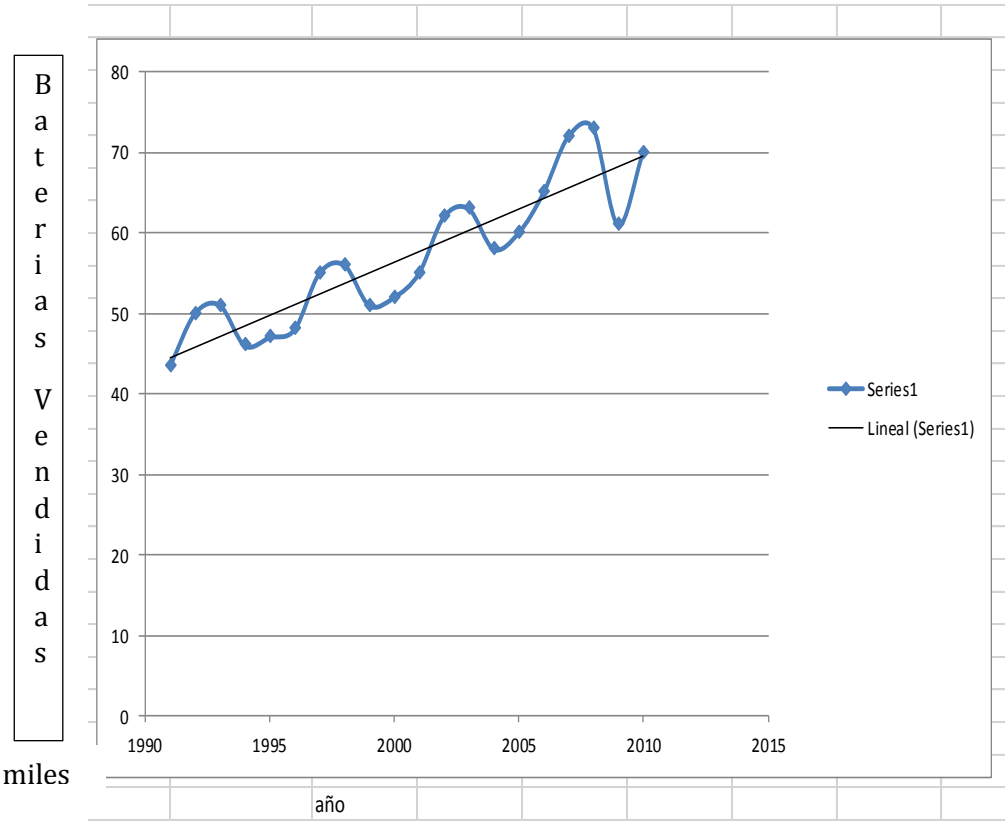


VARIACIÓN CÍCLICA:

Aumento y reducción de una serie de tiempo durante periodos mayores de un año. (un cliclo de negocios habitual consiste en un periodo de **prosperidad**, seguidos por periodos de **recesión, depresión** y luego **recuperación**) Hay tendencias mayores de un año y menores de la secular.

En la siguiente tabla se presentan las unidades anuales de baterías que vendió National Battery Retailers, Inc. Desde 1991 hasta 2010. Resalta el ciclo natural del negocio. Los periodos son de recuperación, seguidos por prosperidad, luego recesión y, por último el ciclo ascendiente con recuperación

año	baterias ventas (miles)
1991	43.5
1992	50
1993	51
1994	46
1995	47
1996	48
1997	55
1998	56
1999	51
2000	52
2001	55
2002	62
2003	63
2004	58
2005	60
2006	65
2007	72
2008	73
2009	61
2010	70



VARIACIÓN ESTACIONAL:

Patrones de cambio en una serie de tiempo en un año. (estos patrones tienden a repetirse cada año) Muchos negocios tratan de equilibrar los efectos estacionales y se dedican a otras actividades de **compensació**n estacional.

Este tipo de serie de tiempo se utiliza por ejemplo con las ventas de autos, los embarques de botellas de cola, la construcción residencial en donde la proyección se hace con un año de anticipación o dos, por mes; lo que es esencial para lograr una programación adecuada.

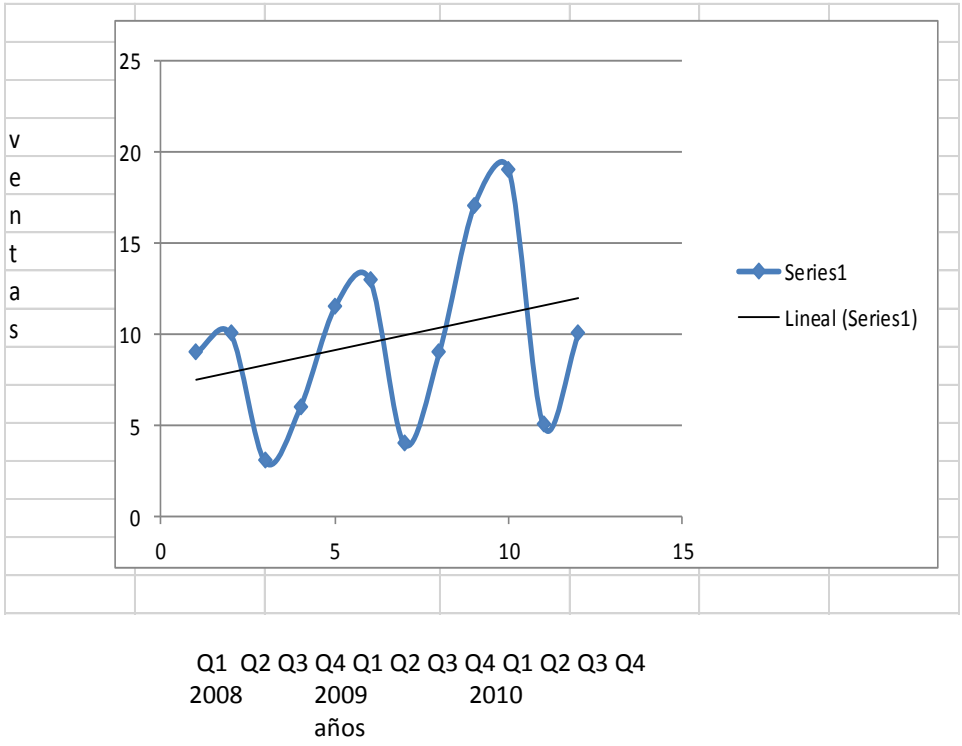
Ejercicio de variación estacional: En el siguiente cuadro aparecen las ventas trimestrales, en millones de dlls de Hercher Sporting Goods, Inc. Dicha compañía de artículos deportivos del área de Chicago se especializa en la venta de equipo y artículos deportivos a preparatorias, universidades y ligas juveniles.

También tiene varias tiendas de descuentos en algunos de los centros de comerciales más grandes. Para este negocio existe un patrón estacional distintivo.

La mayoria de sus ventas es en el primero y segundo trimestre del año, cuando las escuelas y organizaciones compran equipo para la próxima temporada.

Durante el verano se mantiene ocupada con la venta de equipo

Año	trimestres	ventas mlls. Dlls.
2008	Q1	9
	Q2	10
	Q3	3
	Q4	6
2009	Q1	11.5
	Q2	13
	Q3	4
	Q4	9
2010	Q1	17
	Q2	19
	Q3	5
	Q4	10



VARIACION IRREGULAR:

Las variaciones irregulares son impredecibles, pero es posible en algunos casos identificarlas: por ejemplo, el efecto de una huelga importante o de una guerra en la economía no se pueden predecir. A estas también se les llama con frecuencias fluctuaciones azarosas, porque incluso, algunas ni siquiera se pueden identificar. Por supuesto, no es posible proyectar a futuro.

PROMEDIO MÓVIL:

Un promedio móvil es útil para **suavizar** una serie de tiempo y apreciar su tendencia. Además es el método básico para medir la fluctuación estacional. ES DECIR EL PROMEDIO MÓVIL SOLO SUAVIZA LAS FLUCTUACIONES DE LOS DATOS. Mientras que el método de los mínimos cuadrados, expresa la tendencia en términos de una ecuación matemática:

$$Y^* = a+bX$$

Para aplicar el promedio móvil a una serie de tiempo, los datos deben seguir una tendencia muy lineal y tener un patrón rítmico definido de las fluctuaciones (que se repita, por ejemplo, cada tres años).

Para aplicar el Promedio Móvil en una serie de tiempo:

1.- los datos deben seguir una tendencia lineal.

2.-Tener un patrón rítmico definido de las fluctuaciones. (Que se repita).

Una serie de tiempo, con el método del promedio móvil utilizando el software estadístico de Excel y Minitab, nos da la tendencia de manera gráfica.

Mientras que por otro lado, el método de los mínimos cuadrados expresa la tendencia en términos de una ecuación matemática:

$$Y^* = a + b(X)$$

PROMEDIO MÓVIL PONDERADO:

En un promedio móvil se utiliza la misma ponderación para cada observación. En otras palabras, en este caso, cada valor de datos tiene una ponderación de un tercio. De manera similar, en el caso de un promedio móvil de 5 años, cada valor de datos tiene una ponderación de un quinto.

En la mayoría de las aplicaciones se emplea el valor de la ponderación uniformizado como una proyección a futuro.

Pero también es posible tener unos datos con **diferente ponderación** para cada valor, por ejemplo: a la primera observación o más reciente se le da la ponderación mayor, la cual disminuye con valores de datos más antiguos. (VER EJEMPLO DE CEDAR FAIR).

PROBLEMA 2.2. SERIES DE TIEMPO. CEDAR FAIR

Excel.

Cedar Fair opera siete parques de diversiones y cinco parques acuáticos independientes. Su Asistencia combinada (en miles) durante los últimos 17 años aparece en la siguiente tabla. Un socio le pide estudiar la tendencia de la asistencia.

Calcular un promedio móvil de 3 años y un promedio móvil ponderado de 3 años con Ponderaciones de 0.2, 0.3 y 0.5 para los años sucesivos.

año	código	asistencia (miles)	promedio móvil 3 años	Prom. Móvil ponderado 0.2,0.3,0.5	
1993	1	5761			
1994	2	6148	6230.6667	6388.1	0
1995	3	6783	6792	6987	
1996	4	7445	7211	7292.6	
1997	5	7405	8766.6667	9435.5	
1998	6	11450	10026.333	10528	
1999	7	11224	11459	11508.7	
2000	8	11703	11605.667	11700.7	
2001	9	11890	11991	12097.6	
2002	10	12380	12150.333	12182.5	
2003	11	12181	12372.667	12408.8	
2004	12	12557	12479.333	12553.3	
2005	13	12700	14852.333	15971.4	
2006	14	19300	18033.333	19380	
2007	15	22100	21373.333	21850	
2008	16	22720	21985.333	21804	
2009	17	21136			
2010	18	22192.49			
2010	18	21844.72			

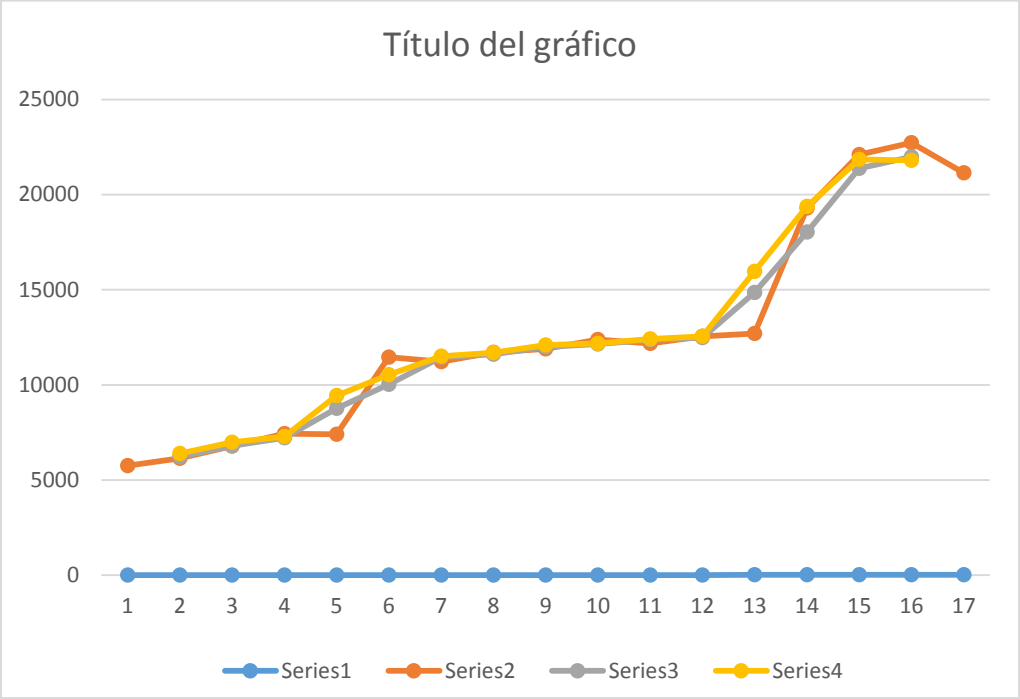
Vamos a proceder a realizar el promedio móvil de 3 años:

COMANDOS:

Se coloca el cursor en la 2a celda de las siguientes columnas de datos porque son 3 celdas a analizar (D12) y se seleccionan los 3 primeros datos de la columna anterior como lo señala la formula (C11, C12, C13/3) y se divide en 3 porque son 3 años. Se marca la celda y se arrastra hacia abajo hasta la penúltima celda para dejar un espacio vacío como se hizo arriba y aparecerán todos los cálculos.

Para el promedio móvil ponderado se procede:
Se coloca el cursor en la celda (E12) y se marcan las celdas como lo indica la fórmula, Marcar con el cursor la celda y arrastrarla hacia abajo hasta la penúltima celda y quedan los cálculos en la columna.

Para graficar se marca con el cursor las 3 columnas (B, C, D y E) y seleccionar la gráfica deseada que se encuentra en la parte superior de la pantalla.



NOTA: En la mayoría de los casos, es común usar una ponderación con el fin de dar un mayor peso a los datos más recientes, disminuyendo los de mayor antigüedad

Conclusiones:

Al observar la gráfica con cuidado, se encuentra que la tendencia de la Asistencia es ascendente de manera uniforme, con 360,000 visitantes más cada año.
Sin embargo, hay un "salto" de casi tres millones por año entre 1997 y 1998. Es probable que esto refleje que Cedar Fair adquirió Knott´s Berry Farm en 1997, lo que generó un incremento repentino en la asistencia. Ocurrió un auge similar en 2006 con la compra de King´s Island, cerca de Cincinnati. El promedio móvil ponderado sigue los datos de manera más cercana que el promedio móvil, lo que refleja la influencia adicional que recibe el periodo más reciente. En otras palabras, el método ponderado, con forme se da la ponderación mayor al periodo más reciente, no será tan uniforme. Sin embargo, quizá sea más preciso como herramienta de proyección.

Por otro lado, si queremos tener una ecuación, que nos permita hacer un cálculo analítico de una proyección más lineal, es conveniente utilizar un software de regresión lineal como se vio en la unidad V del programa de estadística inferencial I.

Minitab.

Hoja de trabajo.

Cedar Fair opera siete parques de diversiones y cinco parques acuáticos independientes. Su asistencia combinada (en miles) durante los últimos 17 años aparece en la siguiente tabla. Un socio le pide estudiar la tendencia de la asistencia.

Calcular un promedio móvil de 3 años y un promedio móvil ponderado de 3 años con ponderaciones de 0.2, 0.3 y 0.5 para los años sucesivos.

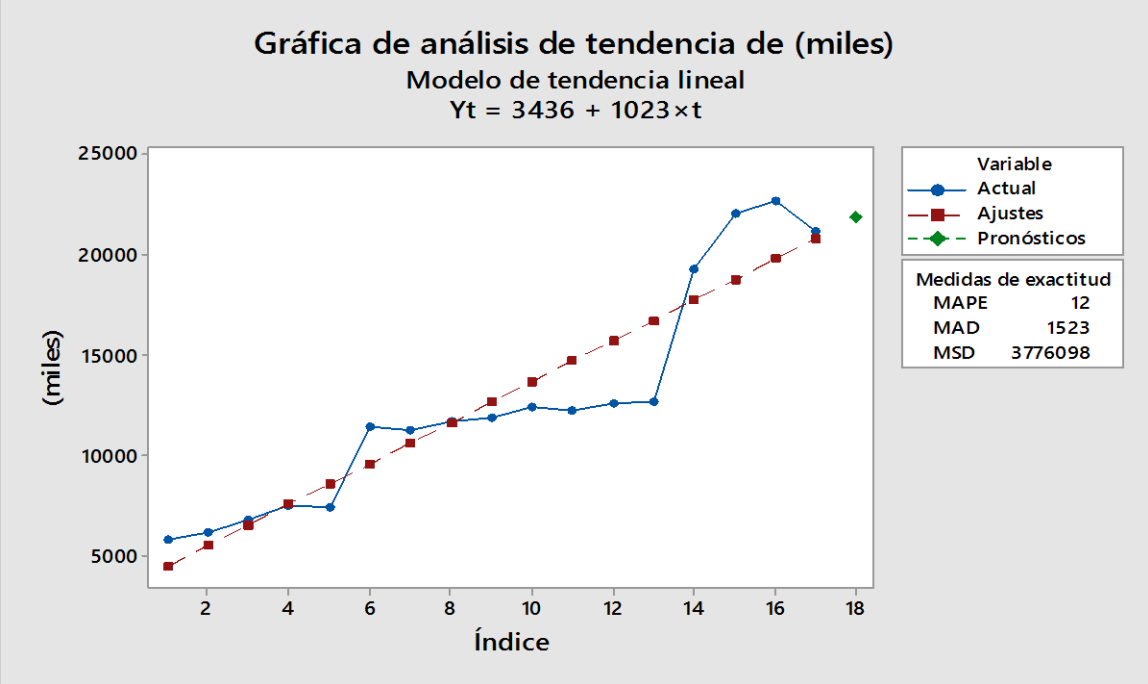
año	código	(miles)	móvil 3 años	0.2,0.3,0.5
1993	1	5761	*	*
1994	2	6148	6230.66667	6388.1
1995	3	6783	6792	6987
1996	4	7445	7211	7292.6
1997	5	7405	8766.66667	9435.5
1998	6	11450	10026.3333	10528
1999	7	11224	11459	11508.7
2000	8	11703	11605.6667	11700.7
2001	9	11890	11991	12097.6
2002	10	12380	12150.3333	12182.5
2003	11	12181	12372.6667	12408.8
2004	12	12557	12479.3333	12553.3
2005	13	12700	14852.3333	15971.4
2006	14	19300	18033.3333	19380
2007	15	22100	21373.3333	21850
2008	16	22720	21985.3333	21804
2009	17	21136		

COMANDOS:

Estadística – Series de tiempo – Análisis de tendencia – Variables: miles – Tipo de modelo: lineal – Seleccionar pronósticos: 1 – Aceptar. Listo

Análisis de tendencia para (miles)

```
Datos                          (miles)
Longitud                         17
Número de valores faltantes      0


Ecuación de tendencia ajustada

Yt = 3436 + 1023×t


Medidas de exactitud

MAPE       12
MAD      1523
MSD   3776098


Pronósticos

Período   Pronóstico
18          21844.7
```



Conclusión:

Minitab nos da los datos de el pronóstico como en este caso es de 21844.7 que es el mismo valor que nos da el pronóstico de la gráfica en el punto verde. Si tocamos el punto verde con el cursor, la gráfica nos da el mismo resultado de 21 844.7.

Tanto la gráfica cómo el cálculo de Minitab nos dan la ecuación lineal del problema para hacer pronósticos.

PROBLEMA 2.3 PROMEDIO MOVIL DE 3 Y 5 AÑOS

Excel.

AÑO	y=ventas	PRM.MV3A	PRM.MV5A
1991	5		
1992	6	6.33333333	
1993	8	8	6.8
1994	10	7.66666667	6.4
1995	5	6	6.6
1996	3	5	7
1997	7	6.66666667	7.4
1998	10	9.66666667	8.6
1999	12	11	9.8
2000	11	10.6666667	11
2001	9	11	12
2002	13	12.3333333	13.2
2003	15	15.3333333	14
2004	18	16	14.4
2005	15	14.6666667	14.6
2006	11	13.3333333	15
2007	14	14	15.8
2008	17	17.6666667	
2009	22		

Una empresa realiza un análisis de sus ventas y desea hacer una proyección a futuro y desea Emplear un promedio móvil a 3 y a 5 años para tal fin.

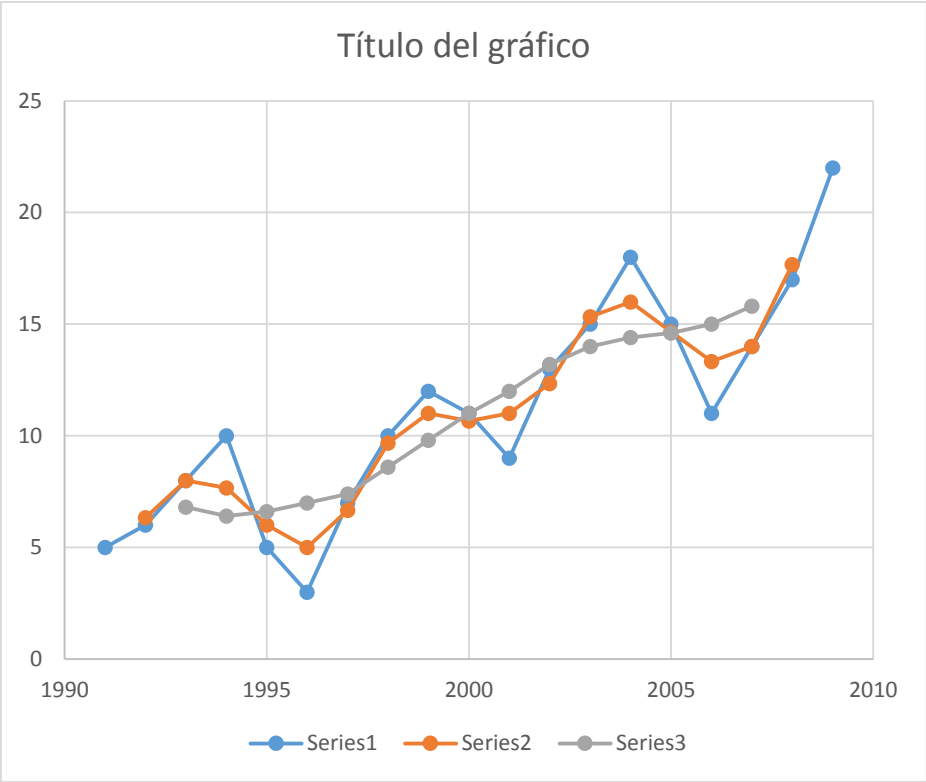
El periodo a analizar es de 1991 a 2009 del comportamiento de sus ventas en millones de Dlls. Como datos se tiene únicamente las columnas A y B.

Las columnas C y D se determinan con las fórmulas de Excel como se indica colocando el cursor en las Celdas C8 para el promedio móvil de 3 años y en la celda D9 como se indica y aplicar las fórmulas Señaladas, arrastrando el cursor hacia abajo se tienen todos los cálculos respectivos, terminando en la celda correspondiente, dejando una celda vacía para el de 3 años y dos celdas para la de 5 años.

Se procede a graficar los resultados y hacer el análisis al observar las gráficas.

Para la gráfica:

COMANDOS: marcar las columnas A, B, C y D , y seleccionar en la parte de arriba la gráfica de "Dispersión" y escoger la gráfica deseada.



Aquí se puede observar que el promedio móvil de 5 años, "amortigua" más las curvas o los "picos" y la tendencia es más clara o más obvia que la del promedio móvil de 3 años.

Se observa también que tiene una tendencia lineal, por lo que también se puede utilizar el método De los mínimos cuadrados, es decir, la ecuación de la recta con un software de regresión lineal Simple y obtener una ecuación analítica para las proyecciones, ya que éste método, es únicamente gráfico pero es muy útil por lo obvio que resulta.

Minitab.

AÑO	y=ventas
1991	5
1992	6
1993	8
1994	10
1995	5
1996	3
1997	7
1998	10
1999	12
2000	11
2001	9
2002	13
2003	15
2004	18
2005	15
2006	11
2007	14
2008	17
2009	22

Una empresa realiza un análisis de sus ventas y desea hacer una proyección a futuro y desea emplear un promedio móvil de 3 y 5 años para tal fin.

El periodo a analizar es de 1991 a 2009 del comportamiento de sus ventas en millones de Dlls.

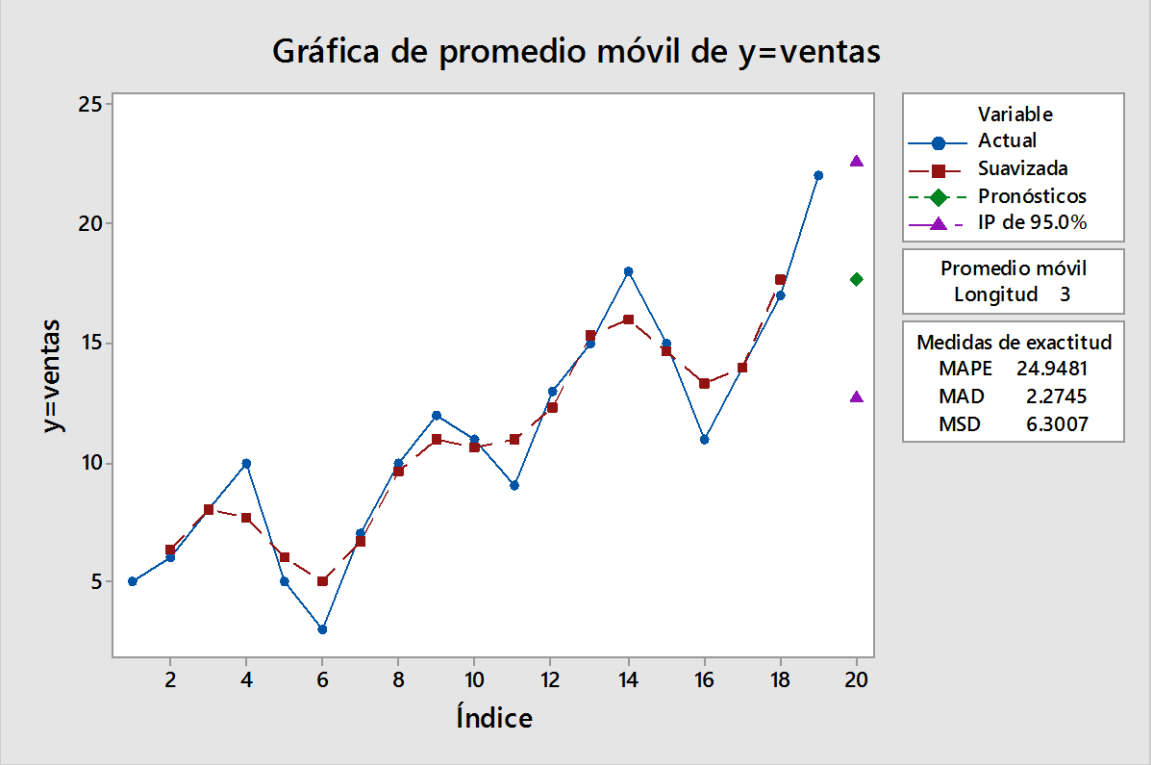
Se procede a graficar los resultados y hacer el análisis al observar las gráficas.

COMANDOS:

Cargar hoja de trabajo. – Estadística – Series de tiempo Promedio móvil – Variable: Y=ventas – longitud de MA: 3 – Marcar centrar promedio móviles – marcar generar pronósticos: No de pronósticos: 1 – Gráficas: seleccionar: suavizado vs real – Aceptar. Listo.

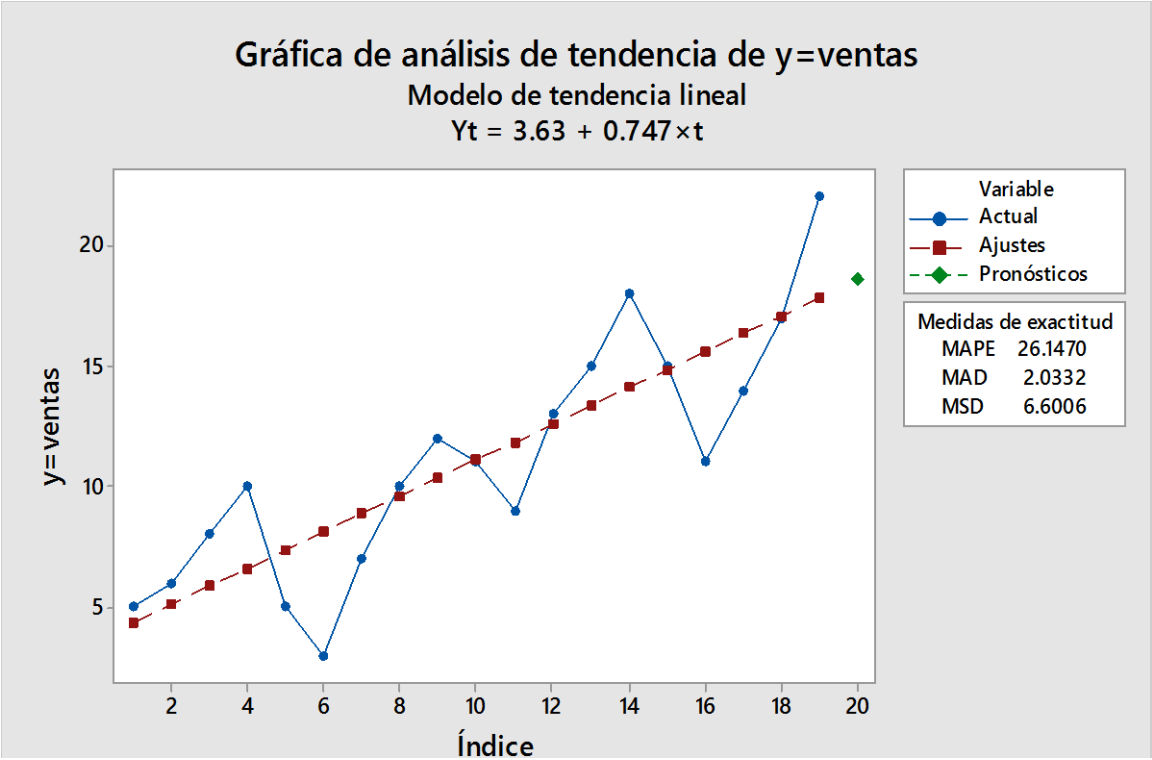
Promedio móvil de y=ventas

```
Datos                          y=ventas
Longitud                       19
Número de valores faltantes    0


Promedio móvil
Longitud  3


Medidas de exactitud
MAPE  24.9481
MAD    2.2745
MSD    6.3007

Pronósticos

Período  Pronóstico  Inferior  Superior
20          17.6667   12.7469   22.5864
```



Análisis de tendencia para y=ventas

```
Datos                          y=ventas
Longitud                       19
Número de valores faltantes    0


Ecuación de tendencia ajustada

Yt = 3.63 + 0.747×t



Medidas de exactitud

MAPE  26.1470
MAD    2.0332
MSD    6.6006


Pronósticos

Período  Pronóstico
20          18.5789
```



Primero se determina el promedio móvil de 3 años: se observa en la zona de cálculos los resultados y se tiene una gráfica siguiendo los comandos indicados.

En la gráfica se observa cómo se amortiguan los datos con el promedio móvil de 3 años, y se tiene un pronóstico como se pide en la construcción de la gráfica y los cálculos éste pronóstico en la gráfica se da con un punto verde y si se coloca el indicador de cursor, nos da el valor numérico del pronóstico como también en la parte del área de los cálculos de Minitab. Se procede de igual manera con el promedio móvil de 5 años.

La gráfica de Análisis de tendencias nos gráfica los datos y nos da la recta, así cómo la ecuación de la recta para hacer los pronósticos calculados con dicha ecuación.

Por ejemplo en Este caso el punto verde nos da un valor de 18.5789 y nos da la ecuación Yt = 3.63+ 0.747+t para realizar el cálculo.

LOS COMANDOS DE LA GRÁFICA DE ANÁLISIS DE TENDENDENCIAS SON:

Estadística – Series de tiempo – Análisis de tendencias – Variables: Y=ventas – Tipo de modelo: Lineal – Seleccionar pronósticos: 1 – aceptar. Listo

PROBLEMA 2.4. VARIACIÓN ESTACIONAL. TOYS INTERNATIONAL.

Excel.

En la siguiente tabla aparecen las ventas trimestrales de Toys International de 2004 a 2009. Las ventas se reportan en millones de Dlls. Determinar un Índice estacional trimestral con el método de la razón con el promedio móvil.

Año	Estación	codificación	ventas		P.M.4E	PMEC	IE	
2004	invierno	**1**	6.7	invierno				10/8.475=1.18
	primavera	2	4.6	primavera	8.5			
	verano	3	10	verano	8.45	8.475	1.179941	**1.179941**
	otoño	4	12.7	otoño	8.45	8.45	1.50295858	**1.50295858**
2005	invierno	**5**	6.5	invierno	8.4	8.425	0.77151335	**0.77151335**
	primavera	6	4.6	primavera	8.625	8.5125	0.54038179	**0.54038179**
	verano	7	9.8	verano	8.725	8.675	1.129683	**1.129683**
	otoño	8	13.6	otoño	8.825	8.775	1.54985755	**1.54985755**
2006	invierno	**9**	6.9	invierno	8.975	8.9	0.7752809	**0.7752809**
	primavera	10	5	primavera	9.1	9.0375	0.55325035	**0.55325035**
	verano	11	10.4	verano	9.125	9.1125	1.14128944	**1.14128944**
	otoño	12	14.1	otoño	9.25	9.1875	1.53469388	**1.53469388**
2007	invierno	**13**	7	invierno	9.35	9.3	0.75268817	**0.75268817**
	primavera	14	5.5	primavera	9.575	9.4625	0.58124174	**0.58124174**
	verano	15	10.8	verano	9.6	9.5875	1.12646675	**1.126466754**
	otoño	16	15	otoño	9.65	9.625	1.55844156	**1.55844156**
2008	invierno	**17**	7.1	invierno	9.725	9.6875	0.73290323	**0.73290323**
	primavera	18	5.7	primavera	9.6	9.6625	0.58990944	**0.58990944**
	verano	19	11.1	verano	9.825	9.7125	1.14285714	**1.14285714**
	otoño	20	14.5	otoño	9.95	9.8875	1.4664981	**1.4664981**
2009	invierno	**21**	8	invierno	10.025	9.9875	0.80100125	**0.80100125**
	primavera	22	6.2	primavera	10.125	10.075	0.61538462	**0.61538462**
	verano	23	11.4	verano				
	otoño	24	14.9	otoño				

Explicación del cuadro:

Columna A son los años (datos)
Columna B son las estaciones (datos)
Columna C se codifica las estaciones del 1 al 24

Columna F promedio móvil de 4 estaciones. (Se calcula como se indica en la fórmula)
Columna G Promedio móvil estacional centrado.
Columna H Índice específico.
Columna I

Se suman todos los índices de los inviernos, luego de primavera, luego veranos y de otoños. Se ajusta a 4 y se determina el índice estacional específico ajustado a 4 como se indica en el cuadro.

año	invierno	primavera	verano	otoño		
2004			1.18	1.503		gráfica de
2005	0.772	0.54	1.13	1.55		índice
2006	0.775	0.553	1.141	1.535		estacional
2007	0.753	0.581	1.126	1.558		específico
2008	0.733	0.59	1.143	1.466		
2009	0.801	0.615				
total	3.834	2.879	5.72	7.612		
media	0.767	0.576	1.144	1.522	4.009	4/4.009=0.9976903
ajustado	0.765	0.575	1.141	1.519	4	
índice	76.5	57.5	114.1	151.9		
	-23.5	-42.5	14.1	51.9		
	INVIERNO	PRIMAVERA	VERANO	OTOÑO		

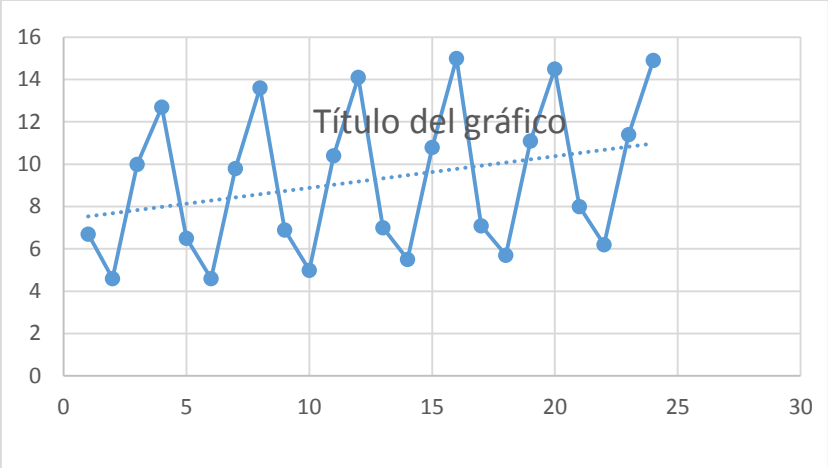
Índice trimestral

habitual					FACTOR DE
invierno	primavera	verano	otoño		CORRECCIÓN
0.76667738	0.57603359	1.14404747	1.52249	4.00924837	0.9977
0.76490884	0.57470482	1.14140843	1.51898	4	

INTERPRETACIÓN DEL ÍNDICE:

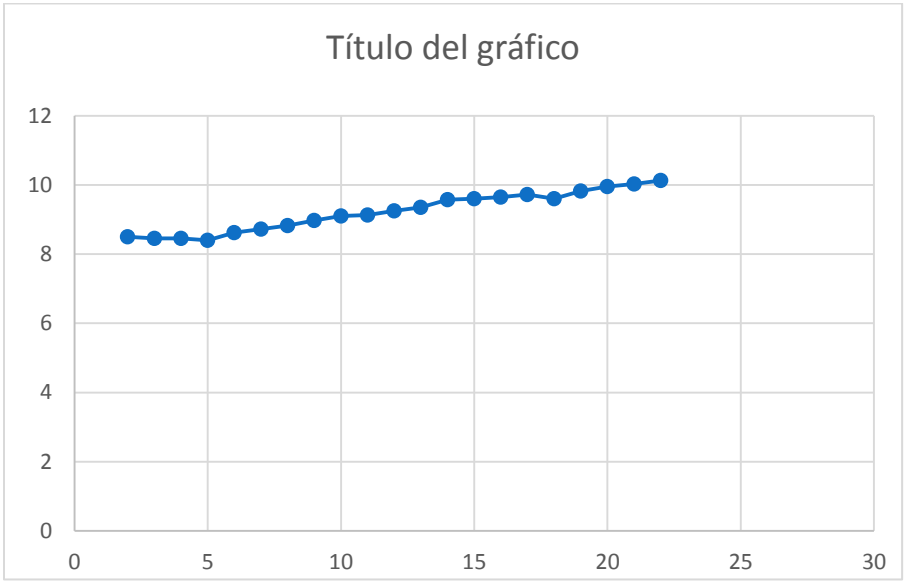
Se interpreta que de los datos del pasado, que en invierno se vende 23.5% por abajo del trimestre habitual:
En primavera se vende 42.5 % por abajo del trimestre habitual, En VERANO se vende 14.14% más del trimestre habitual y En otoño se vende 51.89 % más del trimestre habitual.

Si graficamos los datos de las ventas por cada estación nos quedaría:



COMANDOS:

Para la gráfica: Seleccionamos los datos de las estaciones codificadas (columnas C y D)-insertar-seleccionar gráficas de "Dispersión " y seleccionar la deseada, listo. Con el Mouse clic izquierdo tocamos un punto de la gráfica y seleccionamos la tendencia lineal y listo, nos da la recta en puntos. Si graficamos los datos codificados (columna C) con Promedio 4 estaciones (PM.4E) nos queda una gráfica que es casi una recta. Construimos la gráfica con los mismos comandos:



UNIDAD III.

DISEÑO DE EXPERIMENTOS 1 FACTOR

PROBLEMA 3.1, 1 FACTOR. COMPARACIÓN DE FERTILIZANTES

Excel.

La siguiente tabla muestra la producción por acre, en celemines, de cierta variedad de trigo cultivado en un tipo particular de suelo tratado con los fertilizantes A, B y C.

Calcular:

a). La producción media de los diferentes tratamientos.

b). La gran media de todos los tratamientos.

c). La variación total.

d). La variación entre tratamientos.

e). La variación dentro de los tratamientos y

f). Realizar el análisis del software correspondiente a los cálculos

1 acre=4047 m2 es decir 40áreas y 47 centiáreas, 1 hectárea=2.471 acres

1 celemin= 4,625 lts.

A	48	49	50	49
B	47	49	48	48
C	49	51	50	50

Ho: no hay diferencia entre cultivos

H1: si hay diferencia entre cultivos

a) al 0.05 de nivel de significancia

b) al nivel de 0.01 de significancia

COMANDOS:

Datos – Análisis de varianza de un factor – Aceptar – Rango de entrada – Marcar todos los dato numéricos de la tabla – Clic en filas porque así están ordenados los datos – clic en Alfa 0.05 o 0.01 según el caso – Rango de salida –colocar el cursor – Listo.

Análisis de varianza de un factor.

RESUMEN

Grupos	*Cuenta*	*Suma*	*Promedio*	*Varianza*
Fila 1	4	196	49	0.666667
Fila 2	4	192	48	0.666667
Fila 3	4	200	50	0.666667

ANÁLISIS DE VARIANZA

Origen de las variaciones	*Suma de cuadrados*	*Grados de libertad*	*Promedio de los cuadrados*	*F*	*Probabilidad*	*Valor crítico para F*
Entre grupos	8	2	4	6	0.0220854	4.2564947
Dentro de los grupos	6	9	0.6666667			
Total	14	11				

a).-196/4 =49 b).-192/4 = 48 c).- 200/4 =50

b).- 196+192+200/12 = 49

c).- 0.666667 La misma dentro y entre tratamientos

f):

F calculado es =6>F crítica 4.25 (de tablas) se rechaza la Ho al nivel de significancia de 0.05

F calculado=6<Fcrítico=8.02 (de tablas) no se rechaza Ho al nivel de significancia de 0.01

p = 0.022, se interpreta como el valor máximo al que podría aceptarse la Ho.

O dicho de otra forma: es la probabilidad de que la Ho sea cierta, además si contrastamos 0.022 con el valor del nivel de significancia es menor que 0.05, por lo que también se rechaza la Ho. Tiene menos probabilidad de ser cierta la Ho.

Sin embargo en el caso de α= 0.01 no se rechaza la Ho ya que F calculada = 6 < que F crítica =8.02 De la misma manera p=0.022 > α= 0.01, es decir es mayor la probabilidad de ser cierta la Ho.

Podemos concluir con este análisis que a un nivel de significancia de 0,05 sí existe una diferencia significativa entre cultivos con el uso de éstos fertilizantes siendo el de mejor resultado el Fertilizante "C".

No así, a un nivel de significancia de 0.01 al aceptar la Ho de que no hay diferencias. Es decir, no existe diferencias significativas entre los cultivos con los usos de estos tres fertilizantes.

Minitab.

La siguiente tabla muestra la producción por acre, en celemines, de cierta variedad de trigo cultivado en un tipo particular de suelo tratado con los fertilizantes A, B y C. Calcular:

A	B	C
48	47	49
49	49	51
50	48	50
49	48	50

a). La producción media de los diferentes tratamientos.

b). La gran media de todos los tratamientos.

c). La variación total.

d). La variación entre tratamientos.

e). La variación dentro de los tratamientos

f). Realizar el análisis Del software correspondiente a los cálculos

1 acre=4047 m2 es decir 40áreas y 47 centiáreas, 1 hectárea=2.471 acres

1 celemin= 4,625 lts.

a) al 0.05 de nivel de significancia
b) al nivel de 0.01 de significancia

COMANDOS:

Cargar hoja de trabajo – Estadística – anova – un solo factor – checar que el recuadro diga: los datos de respuestas están en una columna separada para cada nivel de factor - clic en respuestas: cargar las tres columnas una por una en el recuadro – aceptar – listo.

Seleccionar gráficas por intervalos.

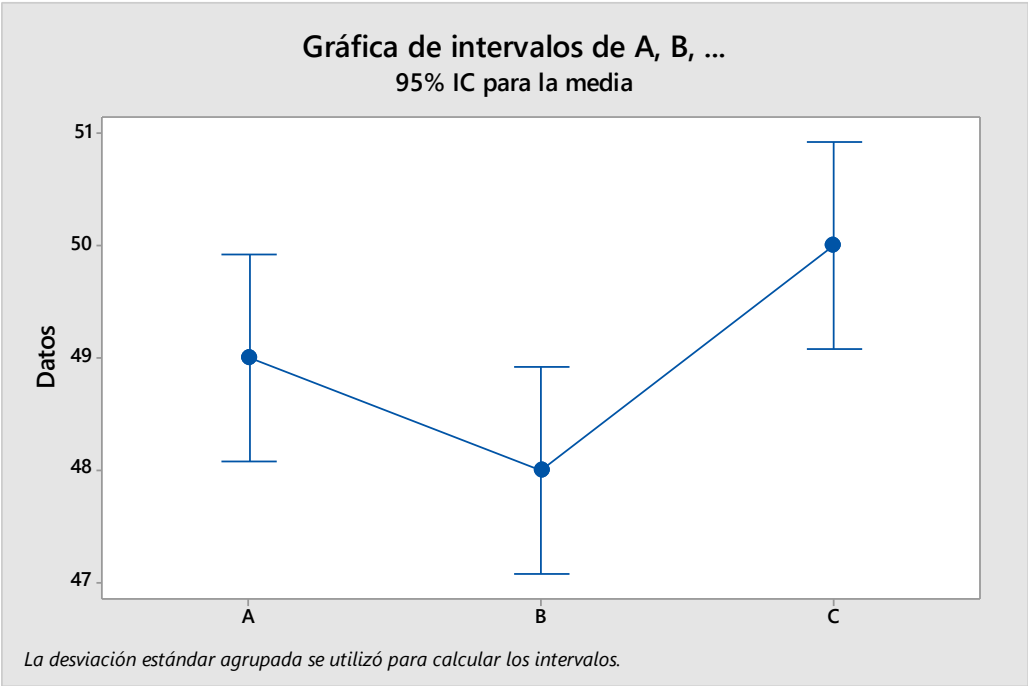
ANOVA unidireccional: A, B, C

```
Método

Hipótesis nula          Todas las medias son iguales
Hipótesis alterna       Por lo menos una media es diferente
Nivel de significancia  α = 0.05
```

```
Se presupuso igualdad de varianzas para el análisis.

Información del factor

Factor  Niveles  Valores
Factor        3  A, B, C


Análisis de Varianza

Fuente  GL  SC Ajust.  MC Ajust.  Valor F  Valor p
Factor   2      8.000     4.0000     6.00    0.022
Error    9      6.000     0.6667
Total   11     14.000

Resumen del modelo

                      R-cuad.  R-cuad.
       S  R-cuad.  (ajustado)   (pred)
0.816497   57.14%      47.62%   23.81%

Medias

Factor  N   Media  Desv.Est.      IC de 95%
A       4  49.000      0.816  (48.076, 49.924)
B       4  48.000      0.816  (47.076, 48.924)
C       4  50.000      0.816  (49.076, 50.924)

Desv.Est. agrupada = 0.816497
```

Ho: No existe diferencia significativa entre los cultivos.

H1: sí existe diferencias significativas entre los cultivos.



Conclusión:

A un nivel de significancia de 0,05 se rechaza la Ho ya que Fcrítica =4.25 (de tablas) es menor que Fcalculada = 6. De igual manera para la prueba de p =0.022 < 0.05 se rechaza la Ho. Es decir, sí hay diferencias significativas entre cultivos.

De igual manera para la prueba de p 0.022>que 0.01. Se acepta la Ho. Es decir No hay Diferencias significativas entre los cultivos con la aplicación de éstos tres fertilizantes.

LO PODEMOS OBSERVAR EN LA GRÁFICA ADJUNTA QUE LAS BARRAS VERTICALES NO SE TRASLAPAN. LO QUE INDICA QUE EXISTE UNA DIFERENCIA SIGNIFICATIVA ENTRE CULTIVOS CON EL USO (para 0.05) DE ESTOS TRES FERTILIZANTES.

De estos tres fertilizantes, Partiendo de la base que la Ho es e que no existe diferencias significativas y la H1 de que si existe diferencias significativas entre cultivos con el uso de estos tres fertilizantes.

PROBLEMA 3.2, 1 FACTOR. PRODUCTIVIDAD DE 5 MÁQUINAS.

Un productor de pernos desea comparar la productividad de cinco de sus máquinas (A, B, C, D, E). En la tabla se presentan la producción de cada una de ellas durante cinco días de la semana. Usando un nivel de significancia de 0.05. ¿Se puede considerar que hay diferencias significativas en la productividad entre las cinco máquinas?

A	B	C	D	E
68	72	60	48	64
72	53	82	61	65
77	63	64	57	70
42	53	75	64	68
53	48	72	50	53

A=C1
B=C2
C=C3
D=C4
E=C5

Ho = No existes diferencia significativa entre la producción de las 5 máquinas

H1 = Sí existes diferencia significativa entre la producción de las 5 máquinas

COMANDOS:

Datos-Análisis de datos-análisis de varianza de un factor-aceptar Rango de entrada-seleccionar los valores de la tablas de datos-clic en columnas porque así están ordenados los datos-seleccionar el nivel de significancia deseado-clic en rango de salida-seleccionar celda donde se desea queden los cálculos-listo.

Análisis de varianza de un factor

RESUMEN

Grupos	*Cuenta*	*Suma*	*Promedio*	*Varianza*
Columna 1	5	312	62.4	210.3
Columna 2	5	289	57.8	92.7
Columna 3	5	353	70.6	76.8
Columna 4	5	280	56	47.5
Columna 5	5	320	64	43.5

ANÁLISIS DE VARIANZA

Origen de las variaciones	*Suma de cuadrados*	*Grados de libertad*	*Promedio de los cuadrados*	*F*	*Probabilidad*	*Valor crítico para F*
Entre grupos	658.16	4	164.54	1.74745115	0.17921408	2.8660814
Dentro de los grupos	1883.2	20	94.16			

Total	2541.36	24

F crítica = 2.86 a 0.05 de nivel de significancia.

Dado que F calculada = 1.74 <F crítica = 2.86 se acepta la Ho Dado que p=0.179 >0.05 se acepta también la Ho, Es decir, no existe una diferencia significativa entre la producción de las 5 máquinas.

Si se trata de calificar o de adquirir a cualquiera de estas 5 máquinas, con ésta prueba de hipótesis nos podemos auxiliar para determinar qué máquina tiene mejor productividad.

Minitab.

A	B	C	D	E
68	72	60	48	64
72	53	82	61	65
77	63	64	57	70
42	53	75	64	68
53	48	72	50	53

Se trata de comparar la producción de 5 máquinas de diferentes marcas para seleccionar la de mayor producción.

COMANDOS:

Cargar hoja de trabajo – Estadística – Anova – Un solo factor – clic en Datos de respuestas están en una columna separadas por cada nivel de factor –Respuestas – cargar las columnas una por una (c1 a C5) seleccionar gráficas por intervalos. Aceptar. Listo

ANOVA unidireccional: A, B, C, D, E

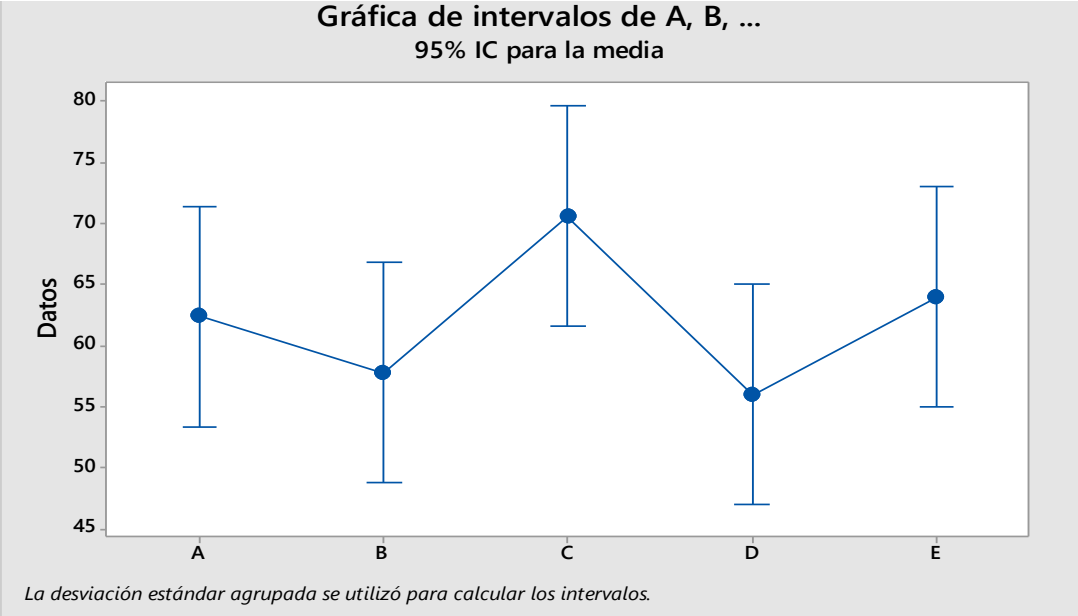
```
Método

Hipótesis nula          Todas las medias son iguales
Hipótesis alterna       Por lo menos una media es diferente
Nivel de significancia  α = 0.05

Se presupuso igualdad de varianzas para el análisis.

Información del factor

Factor  Niveles  Valores
Factor        5  A, B, C, D, E


Análisis de Varianza

Fuente  GL  SC Ajust.  MC Ajust.  Valor F  Valor p
Factor   4      658.2     164.54     1.75    0.179
Error   20     1883.2      94.16
Total   24     2541.4
```

Resumen del modelo

S	R-cuad.	R-cuad. (ajustado)	R-cuad. (pred)
9.70361	25.90%	11.08%	0.00%

Medias

Factor	N	Media	Desv.Est.	IC de 95%
A	5	62.40	14.50	(53.35, 71.45)
B	5	57.80	9.63	(48.75, 66.85)
C	5	70.60	8.76	(61.55, 79.65)
D	5	56.00	6.89	(46.95, 65.05)
E	5	64.00	6.60	(54.95, 73.05)



Conclusión:

El valor de Fcrítica a 0.05 es de 2.86
Dado que Fcalculada= 1.74 < Fcrítica=2.86 se acepta la Ho
De igual manera p=0.179 > 0.05 (nivelde significancia) se acepta la Ho. Es decir, hay mayor probabilidad de aceptar la Ho.

Por lo que se concluye que no existe diferencia significativa en la producción de las cinco máquinas Como se puede ver en la gráfica deintervalos, ya que todas se traslapan.

Los resultados en Minitab coinciden con los de Excel.

PROBLEMA 3.3, 1 FACTOR. 4 LÍNEAS AÉREAS.

Excel.

Northern	WTA	Pocono	Branson
94	75	70	68
90	68	73	70
85	77	76	72
80	83	78	65
	88	80	74
		68	65
		65	

Se selecciona el nivel de significancia de 0.05

4 empresas aéreas están preocupadas por el servicio que ofrecen y contrataron a Brunner Marketing Research,Inc., para encuestar a sus pasajeros sobre la adquisición de boletos, abordaje, servicios durante el vuelo, manejo del equipaje, comunicación del piloto, etc.

Hicieron 25 preguntas con diversas respuestas posibles: excelente con una calificación de 4 puntos, bueno 3, regular 2 y deficiente 1. Estas respuestas se sumaron, de modo que la calificación final fue una indicación de la satisfacción con el vuelo. Entre mayor la calificación, mayor es el nivel de satisfacción con el vuelo. La calificación mayor es de 100 puntos. Brunner selecciono y estudió al azar pasajeros de las 4 líneas que se presentan en la tabla siguiente.

La pregunta es ¿hay alguna diferencia entre los niveles de satisfacción medios con respecto a las 4 aerolíneas?

COMANDOS:

Datos-Análisis de datos-análisis de varianza de un factor-aceptar Rango de entrada-seleccionar los valores de la tablas de datos-clic en columnas porque así están ordenados los datos-seleccionar el nivel de significancia deseado-clic en rango de salida-seleccionar celda donde se desea queden los cálculos-listo.

Análisis de varianza de un factor

RESUMEN

Grupos	*Cuenta*	*Suma*	*Promedio*	*Varianza*
Columna 1	4	349	87.25	36.9167
Columna 2	5	391	78.2	58.7
Columna 3	7	510	72.8571429	30.1429
Columna 4	6	414	69	13.6

ANÁLISIS DE VARIANZA

Origen de las variaciones	*Suma de cuadrados*	*Grados de libertad*	*Promedio de los cuadrados*	*F*	*Probabilidad*	*Valor crítico para F*
Entre grupos	890.6837662	3	296.894589	8.99064	0.00074277	3.15990759
Dentro de los grupos	594.4071429	18	33.022619			
Total	1485.090909	21				

Una vez hecho los cálculos con el software, el alumno debe contestar las siguientes preguntas:

PREGUNTAS:

1.- ¿Cuál es el estadístico de prueba?

2.- ¿Cuál es la hipótesis nula y cuál la alternativa?

3.- ¿Cuál es nivel crítico de F?

4.- ¿Cuál es la regla de decisión?

5.- ¿A qué decisión se llega? (dar cifras)

6.- ¿Cuál es la regla de decisión de "p"?

7.- ¿A que conclusión llega respecto "p" y cómo interpreta "p"?

8.- ¿Cuál es la línea aérea que fue mejor calificada por los usuarios?

RESPUESTAS:

1.- El estadístico de prueba es F

2.-Ho: No existe diferencia significativa en las 4 Líneas Aéreas. Si las hay se debe a la casualidad o a error de muestreo.

H1: Sí existe existe diferencias significativa entre las líneas aéreas.

3.-Fcrítico = 3.15 a un nivel de confianza del 95%

4.-Si F calculada es mayor que F crítica se rechaza la Ho: (F calculada>F crítica)

5.- Dado que F calculada = 8.99>F crítica=3.1599, se rechaza la Ho.

6.- Si p es mayor que 5% se acepta la Ho de lo contrario se rechaza

7.- Dado que p =0.000742<que 0.05, se rechaza la Ho porque la probabilidad de que la Ho sea cierta es muy baja.

p es la probabilidad de aceptar o rechazar la Ho en comparación que el nivel de significancia.

8.-Northern ya que obtuvo el mayor promedio en la puntuación.

Minitab.

Un solo factor desapilado

Northern	WTA	Pocono	Branson
94	75	70	68
90	68	73	70
85	77	76	72
80	83	78	65
	88	80	74
		68	65
		65	

COMANDOS:

Cargar hoja de trabajo – Estadística – Anova – UN solo factor – los datos de respuestas están en una columna separada para cada nivel de factor – Respuestas – Cargar una por una las columnas – en gráficas marcar gráfica por intervalo – aceptar – aceptar – listo.

4 empresas aéreas están preocupadas por el servicio que ofrecen y contrataron a Brunner Marketing Research, Inc., para encuestar a sus pasajeros sobre la adquisición de boletos, abordaje, servicios durante el vuelo, manejo del equipaje, comunicación del piloto, etc.

Hicieron 25 preguntas con diversas respuestas posibles: excelente con una calificación de 4 puntos, bueno 3, regular 2 y deficiente 1. Estas respuestas se sumaron, de modo que la calificación final fue una indicación de la satisfacción con el vuelo.

Entre mayor la calificación, mayor es el nivel de satisfacción con el vuelo. La calificación mayor es de 100 puntos. Brunner selecciono y estudió al azar pasajeros de las 4 líneas que se presentan en la tabla siguiente. La pregunta es:

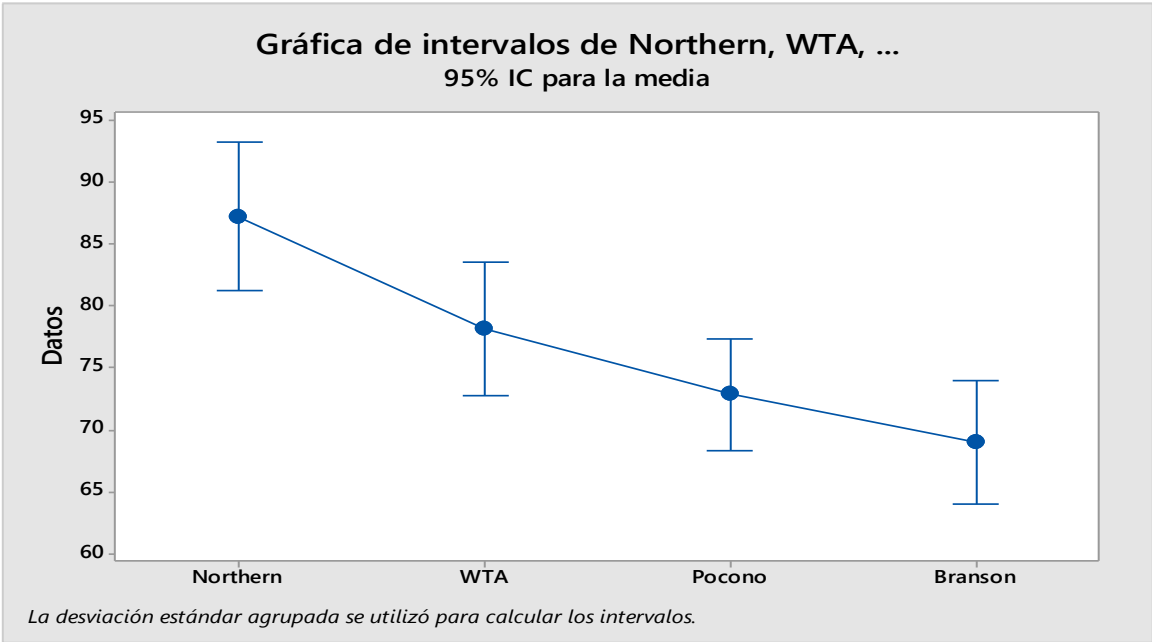
¿Hay alguna diferencia entre los niveles de satisfacción medios con respecto a las 4 aerolíneas?

Se selecciona el nivel de significancia de 0.05

ANOVA unidireccional: Northern, WTA, Pocono, Branson

```
Método

Hipótesis nula          Todas las medias son iguales
Hipótesis alterna       Por lo menos una media es diferente
Nivel de significancia  α = 0.05

Se presupuso igualdad de varianzas para el análisis.


Información del factor

Factor  Niveles  Valores
Factor        4  Northern, WTA, Pocono, Branson


Análisis de Varianza

Fuente  GL  SC Ajust.  MC Ajust.  Valor F  Valor p
Factor   3      890.7     296.89     8.99    0.001
```

```
Error    18      594.4      33.02
Total    21     1485.1


Resumen del modelo
                      R-cuad.  R-cuad.
      S  R-cuad.  (ajustado)   (pred)
5.74653   59.98%      53.30%   38.87%

Medias

Factor     N  Media  Desv.Est.     IC de 95%
Northern   4  87.25       6.08  (81.21, 93.29)
WTA        5  78.20       7.66  (72.80, 83.60)
Pocono     7  72.86       5.49  (68.29, 77.42)
Branson    6  69.00       3.69  (64.07, 73.93)


Desv.Est. agrupada = 5.74653
```



La desviación estándar agrupada se utilizó para calcular los intervalos.

Conclusión:

El valor de Fcrítica es de 3.16 y al comparase con la Fcalculada (8.99).

Se concluye que existe una diferencia significativa entre la Fcrítica y la Fcalculada por lo que se rechaza la Ho. Ya que la Fcalculada cae en la zona de rechazo de la Ho.
De igual manera el valor (redondeado) de p =0.001 es un valor muy chico respecto al valor del Nivel de significancia de 0.05; por lo que es improbable que la Ho sea cierta.

PROBLEMA 3.4, 1 FACTOR. PRUEBA PINTURAS.

Excel.

Pittsburgh Paints desea probar un aditivo formulado para aumentar la VIDA de las pinturas empleadas en condiciones calurosas y áridas del sureste de los Estados Unidos. Se pintó la parte superior de una pieza de madera con la pintura normal, y en la parte inferior se usó la pintura con aditivo. Se siguió el mismo procedimiento con un total de 10 piezas.

Después se sometió a cada pieza a una luz brillante. Los datos, el número de horas que duró la pintura de cada pieza antes de desvanecerse más allá de un cierto punto, son:

Utilizar un Nivel de confianza del 95%.

El alumno debera determinar en una prueba de un Factor, si existe diferencia significativa entre la pintura normal y la pintura con el Aditivo formulado, Es decir, se probara si el aditivo le da mayor tiempo de vida a las pinturas aplicadas en la madera.

	A	B	C	D	E	F	G	H	I	J
sin aditivo	325	313	320	340	318	312	319	330	333	319
con aditivo	323	313	326	343	310	320	313	340	330	315

Ho: $\mu_1 = \mu_2$ No hay diferencia entre medias de la pintura normal y la que tiene el aditivo.

H_1: $\mu_1 \neq \mu_2$ Sí hay diferencias entre medias de la pintura normal y la que tiene el aditivo.

COMANDOS:

Datos>Análisis de datos>seleccionar análisis de varianza de un factor>aceptar>rango de entrada> marcar todos los valores del cuadro de datos >agrupado por filas>alfa 0.05>>rango de salida>colocar el cursor en un espacio vacío> aceptar. Listo.

Análisis de varianza de un factor

RESUMEN

Grupos	*Cuenta*	*Suma*	*Promedio*	*Varianza*
Fila 1	10	3229	322.9	80.989
Fila 2	10	3233	323.3	132.01

ANÁLISIS DE VARIANZA

Origen de las variaciones	*Suma de cuadrados*	*Grados de libertad*	*Promedio de los cuadrados*	*F*	*Probabilidad*	*Valor crítico para F*
Entre grupos	0.8	1	0.8	0.0075	0.93	**4.41**
Dentro de los grupos	1917	18	106.5			
Total	1918	19				

Conclusión:

Dado los resultados de la prueba de Hipótesis de varianza de un factor se puede asegurar con un nivel de confianza del 95 % de que la Ho NO SE RECHAZA porque el Valor de F=4.41 que es el valor crítico es mayor que el de F calculada =0.0075.

También podemos decir que F calculada cae en la Zona de aceptación de la Ho;

Y como la Ho dice que "no hay diferencia entre la pintura normal y la pintura con el aditivo" se acepta la Ho.

Por otro lado, el valor de p=0.93, que quiere decir que existe una probabilidad del 93 % de que la Ho sea cierta y solo un 7% de probabilidad de rechazar la Ho.

Minitab.

Pittsburgh Paints desea probar un aditivo formulado para aumentar la VIDA de las pinturas empleadas en condiciones calurosas y áridas del sureste de los Estados Unidos.

Se pintó la parte superior de una pieza de madera con la pintura normal, y en la parte inferior se usó la pintura con aditivo. Se siguió el mismo procedimiento con un total de 10 piezas.

Después se sometió a cada pieza a una luz brillante. Los datos, el número de horas que duró la pintura de cada pieza antes de desvanecerse más allá de un cierto punto, son:

SIN	CON	SIN2	CON2
325	323	325	323
313	313	313	313
320	326	320	326
340	343	340	343
318	310	318	310
312	320	312	320
319	313	319	313
330	340	330	340
333	330	333	330
319	315	319	315

Utilizar Nivel de confianza del 95%.

El alumno debera determinar en una prueba de un Factor, si existe diferencia significativa entre la pintura normal y la pintura con el Aditivo formulado, Es decir, se probara si el aditivo le da mayor tiempo de vida a las pinturas aplicadas en la madera.

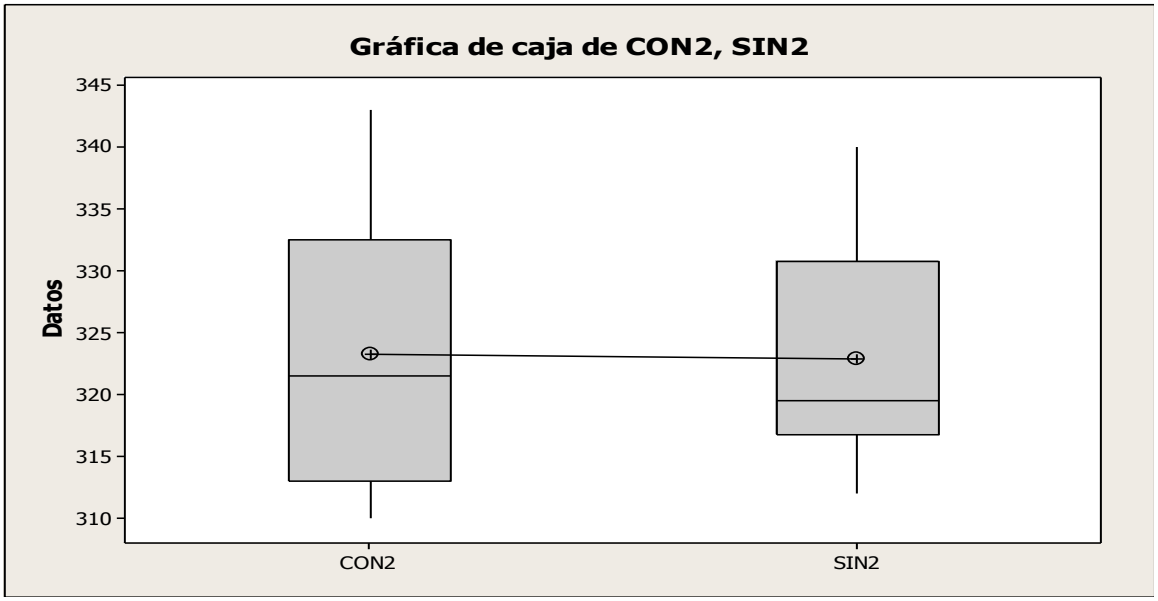
COMANDOS:

Cargar hoja de trabajo – Estadística – Anova – Un solo factor – clic en los datos de respuestas están en una sola columna separadas para cada nivel de factor – clic en respuestas – cargar las columnas una por una – seleccionar gráficas por intervalo – aceptar

```
Fuente  GL    SC   CM     F      P
Factor   1     1    1  0.01  0.932
Error   18  1917  106
Total   19  1918
S = 10.32   R-cuad. = 0.04%   R-cuad.(ajustado) = 0.00%
```

es mucho mayor que el 5% del valor de significancia.

```
                                ICs de 95% individuales para la media
                                basados en Desv.Est. agrupada
Nivel   N   Media  Desv.Est.     +---------+---------+---------+---------
CON2   10  323.30      11.49       (-----------------*----------------)
SIN2   10  322.90       9.00      (-----------------*----------------)
                                 +---------+---------+---------+---------
                               316.0     320.0     324.0     328.0


Desv.Est. agrupada = 10.32
```



Prueba T e IC de dos muestras: SIN, CON

```
T de dos muestras para SIN vs. CON

                              Error
                            estándar
                               de la
      N   Media  Desv.Est.     media
SIN  10  322.90       9.00       2.8
CON  10   323.3       11.5       3.6


Diferencia = μ (SIN) - μ (CON)
Estimación de la diferencia:  -0.40
IC de 95% para la diferencia:  (-10.14, 9.34)
Prueba T de diferencia = 0 (vs. ≠): Valor T = -0.09  Valor p = 0.932  GL = 17
```

Conclusión:

Fcritica a 95% es de 4.41. Como Fcalculada =0.01 < que Fcrítica=4.41 se acepta la Ho es decir no hay diferencia significativa entre la duración de las pinturas con y sin aditivo. POr otra parte p=0.932 es > Alfa de 0.05
Es decir, existe una probabilidad del 93.2% de que la Ho sea cierta es mucho mayor que el 5% del valor de significancia.

Como se puede observar las gráficas se traslapan, eso quiere decir, Que no existen diferencias significativas entre ellas porque se tiene en ellas la desviación Estándar de la media.
Lo mismo para la gráfica de caja.

Los valores de los cálculos que no coinciden con Excel se deben al redondeo.

PROBLEMA 3.5, 1 FACTOR. COMPARACIÓN DE LLANTAS.

Excel.

Una empresa desea probar 4 tipos de llantas: A, B, C, y D. Las vidas medias de las llantas determinadas por su rodaje, están dadas en miles de millas. Donde cada tipo es probado en 6 autos similares asignados aleatoriamente a las llantas.

Determinar si existe una diferencia significativa entre las llantas a los niveles de significancia de 0.05 y 0.01.

Determinar si existe una diferencia significativa en la vida útil de las 4 marcas diferentes de llantas.

A	33	38	36	40	31	35
B	32	40	42	38	30	34
C	31	37	35	33	34	30
D	29	34	32	30	33	31

Ho. No existe diferencia significativa en la vida útil de las llantas
H1: Sí existe diferencia significativa en la vida útil de las llantas.

COMANDOS:

Datos>Análisis de datos>seleccionar análisis de varianza de un factor>aceptar>rango de entrada> marcar todos los valores del cuadro de datos >agrupado por filas>alfa 0.05>>rango de salida>colocar el cursor en un espacio vacío> aceptar. Listo.

Análisis de varianza de un factor

RESUMEN

Grupos	*Cuenta*	*Suma*	*Promedio*	*Varianza*
Fila 1	6	213	35.5	10.7
Fila 2	6	216	36	22.4
Fila 3	6	200	33.33333	6.6667
Fila 4	6	189	31.5	3.5

F crítica 3,20 =3,10

ANÁLISIS DE VARIANZA

Origen de las variaciones	*Suma de cuadrados*	*Grados de libertad*	*Promedio de los cuadrados*	*F*	*Probabilidad*	*Valor crítico para F*
Entre grupos	77.5	3	25.83333	2.3883	0.09919	**3.1**
Dentro de los grupos	216.33	20	10.81667			
Total	293.83	23				

Conclusiones:

Dado que F calculada = 2.38 < que F crítica = 3.1 (de tablas) se acepta la Ho, Es decir, No existe diferencias significativas entre las marcas de las llantas en la vida útil. Por otro lado, El valor de p=0.099 es mayor que α =0.05 por lo que la probabilidad de que la Ho sea cierta es del 9.9 % mayor que el nivel de significancia del 5%

Ahora vamos a hacer el análisis considerando un nivel de significancia del 1 %

Análisis de varianza de un factor

RESUMEN

Grupos	*Cuenta*	*Suma*	*Promedio*	*Varianza*
Fila 1	6	213	35.5	10.7
Fila 2	6	216	36	22.4
Fila 3	6	200	33.33333	6.6667
Fila 4	6	189	31.5	3.5

ANÁLISIS DE VARIANZA

Origen de las variaciones	*Suma de cuadrados*	*Grados de libertad*	*Promedio de los cuadrados*	*F*	*Probabilidad*	*Valor crítico para F*
Entre grupos	77.5	3	25.83333	2.3883	0.09919	**4.94**
Dentro de los grupos	216.33	20	10.81667			
Total	293.83	23				

Se llega a las mismas conclusiones, pues F calculada =2.38<F crítica =4.94

La probabilidad de que la Ho sea cierta es la misma. No hay diferencias significativa entre la vida útil de las llantas.

Los cálculos son los mismos, lo único que cambia es el valor crítico de F a 99% de nivel de confianza o el nivel de significancia ahora es de 1%. Y F crítica es de 4.94

Minitab.

A	B	C	D
33	32	31	29
38	40	37	34
36	42	35	32
40	38	33	30
31	30	34	33
35	34	30	31

Una empresa desea probar 4 tipos de llantas: A,B,C,y D. Las vidas medias de las llantas determinadas por su rodaje, están dadas en miles de millas. Donde cada tipo es probado en 6 autos similares asignados aleatoriamente a las llantas.

Determinar si existe una diferencia significativa entre las llantas a los niveles de significancia de 0.05 y 0.01.

Determinar si existe una diferencia significativa en la vida útil de las 4 marcas diferentes de llantas.

Ho. No existe diferencia significativa en la vida útil de las llantas

H1: Sí existe diferencia significativa en la vida útil de las llantas.

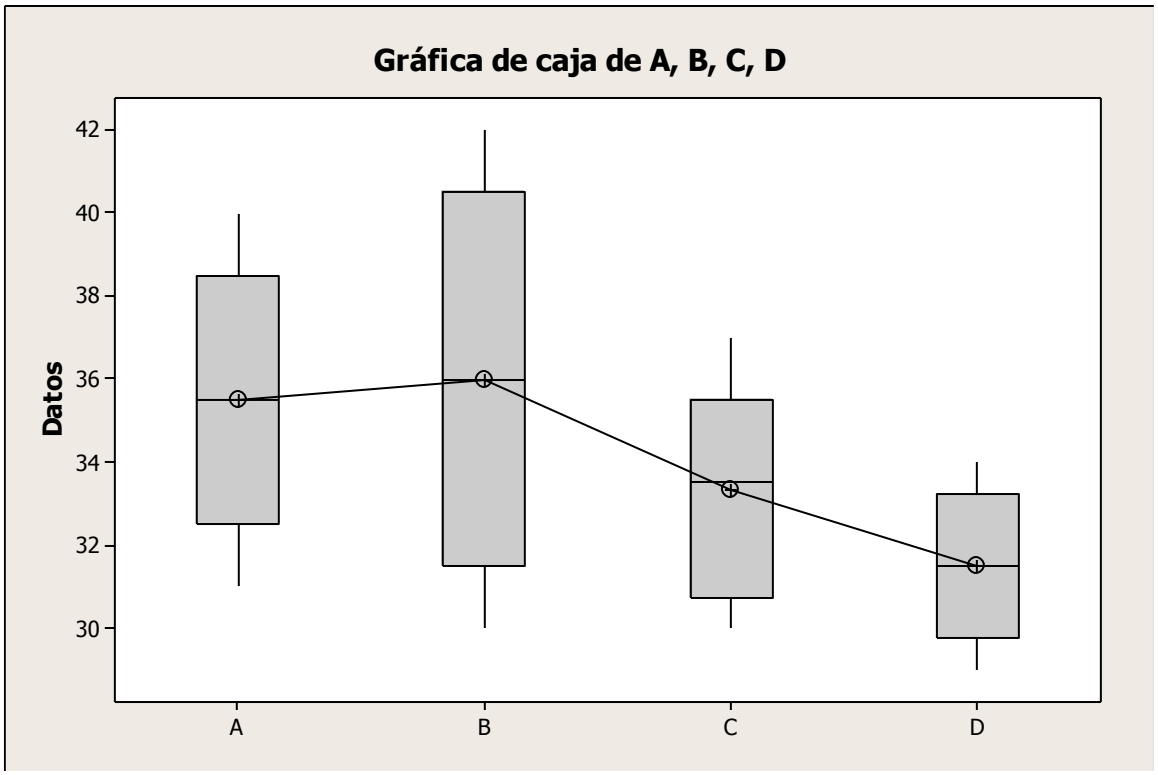
COMANDOS:

Cargar hoja de trabajo – Estadística – Anova – Un solo Factor – Clic en los datos de respuestas están en una columna separada para cada nivel de factor – clic en datos de respuestas – Cargar las columnas una por una – Seleccionar gráfica por intervalos – Aceptar – Listo.

ANOVA unidireccional: A, B, C, D

```
Fuente  GL     SC    CM     F      P
Factor   3   77.5  25.8  2.39  0.099
Error   20  216.3  10.8
Total   23  293.8

S = 3.289   R-cuad. = 26.38%   R-cuad.(ajustado) = 15.33%
```

```
                                ICs de 95% individuales para la media
                                basados en Desv.Est. agrupada
Nivel  N   Media  Desv.Est.  ----+---------+---------+---------+-----
A      6  35.500      3.271              (--------*---------)
B      6  36.000      4.733                (--------*--------)
C      6  33.333      2.582       (--------*--------)
D      6  31.500      1.871  (--------*--------)
                            ----+---------+---------+---------+-----
                                  30.0      33.0      36.0      39.0

Desv.Est. agrupada = 3.289
```

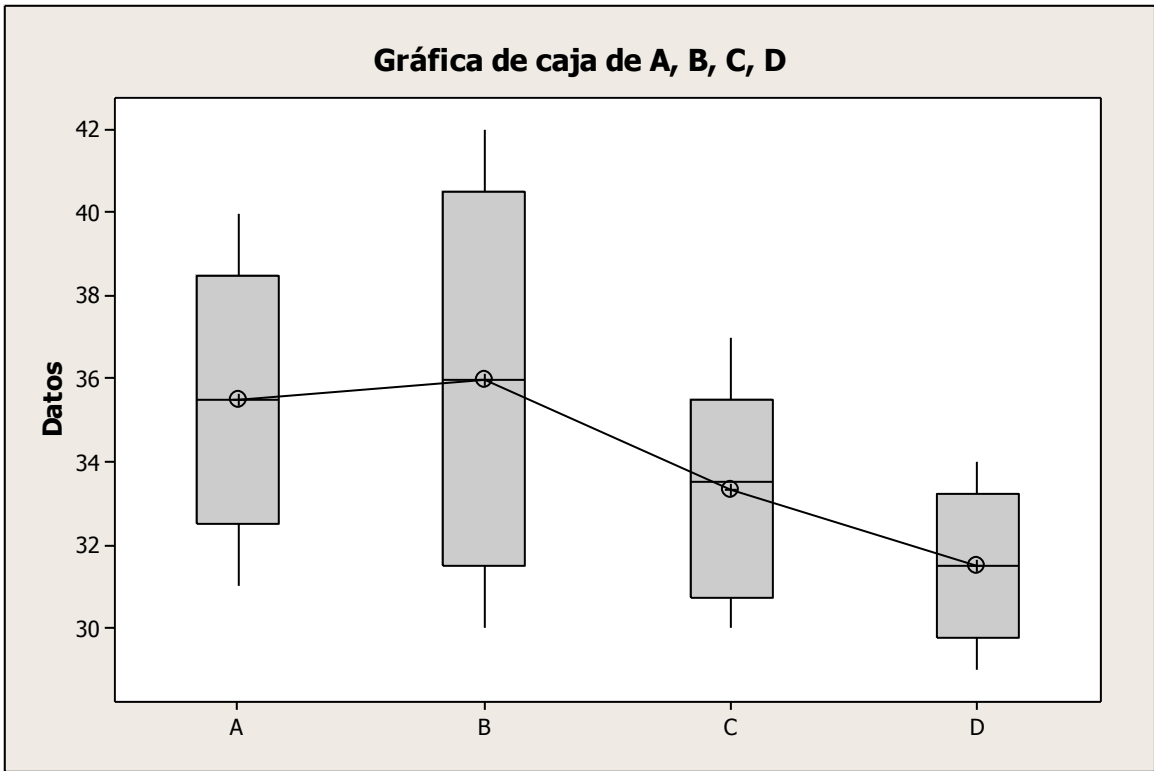


ANOVA unidireccional: A, B, C, D

Fuente	GL	SC	CM	F	P
Factor	3	77.5	25.8	2.39	0.099
Error	20	216.3	10.8		
Total	23	293.8			

S = 3.289 R-cuad. = 26.38% R-cuad.(ajustado) = 15.33%

```
                                ICs de 99% individuales para la media
                                basados en Desv.Est. agrupada
Nivel  N   Media  Desv.Est.   -+---------+---------+---------+--------
A      6  35.500      3.271              (---------*----------)
B      6  36.000      4.733               (----------*----------)
C      6  33.333      2.582         (----------*----------)
D      6  31.500      1.871    (----------*----------)
                               -+---------+---------+---------+--------
                              28.0      31.5      35.0      38.5

Desv.Est. agrupada = 3.289
```



Conclusión:

Fcrít, .95, 3,20=3.10. Por lo que se llega a las mismas conclusiones que en el caso del cálculo que se hizo con Excel ya que coinciden dichos cálculos con Minitab.

Respecto a la gráfica de puntos se puede observar que ninguna de ellas se traslapan por lo que se puede asegurar a un nivel de confianza del 95% que no hay diferencia en ninguna de las marcas de llantas como para considerarlas que sea significativa. Esta comparación se hace en función de la desviación Estándar de la media de cada una de ellas.

.

UNIDAD IV.

DISEÑO DE BLOQUES

DISEÑO EXPERIMENTAL CON BLOQUES AL AZAR

La técnica de análisis de varianza (ANOVA) se emplea después de que se han obtenido los resultados de un experimento. Sin embargo, para obtener la mayor cantidad de información posible, el diseño de un experimento debe planearse cuidadosamente de antemano; Esto se conoce como ***el diseño del experimento.***

Los siguientes son diseños importantes de diseños de experimentos:

D	A	C	C
B	D	B	A
D	C	B	D
A	B	C	A

FIGURA 1

Aleatorización completa

I	C	B	A	D
II	A	B	D	C
III	B	C	D	A
IV	A	D	C	B

FIG. 2

Bloques aleatorizados

D	B	C	A	
B	D	A	C	
C	A	D	B	
A	C	B	D	

FIG. 3

Cuadrado latino

B γ	A β	D δ	C α
A δ	B α	C γ	D β
D α	C δ	B β	A γ
C β	D γ	A α	B δ

FIG. 4

Cuadrado greco-latino

A, B, C, D : tratamientos I, II, III, IV : bloques.

1.- **Aleatorización completa.** Supóngase que se tiene un experimento agrícola como el ejemplo de la figura 1. Para diseñar un experimento como éste, se puede dividir la tierra en 4 X 4 = 16 parcelas, indicadas con cuadrados, aunque físicamente es posible utilizar cualquier forma y asignar cada tratamiento (por ejemplo, fertilizantes indicado por A, B, C y D) a 4 bloques elegidos al azar. El propósito de la aleatorización es eliminar las fuentes de error como la fertilidad del suelo.

2.-**_Bloques aleatorizados.** Cuando es necesario tener un conjunto completo de tratamientos para cada bloque, como en el ejemplo 2, los tratamientos A, B, C y D se introducen en orden aleatorio dentro de cada bloque: I,II,III y I(es decir, los renglones de la fig. 2) ; Por ésta razón, los bloques se conocen como bloques aleatorizados. Este tipo de diseño se utiliza cuando se desea controlar *una fuente de error o variabilidad*; a saber, **la diferencia de los bloques**.

3.- **Cuadrados latinos.** Para algunos propósitos se desea controlar *dos fuentes de error o variabilidad* al mismo tiempo, como **la diferencia en renglones y la diferencia en columnas**. Por ejemplo, en el experimento del ejemplo 1 los errores en distintos renglones y columnas podrían deberse a cambios en la fertilidad, en diferentes partes de la tierra. En tal caso es deseable

que cada tratamiento ocurra una vez en cada renglón y una vez en cada columna, como en la figura 3. Esta disposición se denomina *cuadrado latino,* debido a que se utilizan las letras latinas A, B, C y D.

4.- **Cuadrados greco-latinos.** Si se necesita controlar *tres fuentes de error o variabilidad,* se utiliza un *cuadrado greco-latino,* como se muestra en la fig. 4. Tal cuadrado consiste esencialmente de dos cuadrados latinos superpuestos con letra latinas A, B, C, y D usadas en un cuadrado y las letras griegas (alfa) α, (beta) β, (gamma) γ y (delta) δ usadas en el otro cuadrado. El requisito adicional que tiene que cubrirse es que cada letra latina debe emplearse una y solo una vez con cada letra griega; Cuando este requisito se cubre, se dice que el cuadrado es *ortogonal.*

Veamos ahora un ejemplo resuelto en Excel en donde se hace un ANOVA con dos vías o dos factores para comparar si hay diferencias en las medias de las rutas y diferencias en las medias de los conductores. También se utiliza el término **variable de bloqueo** a la segunda variable que para este caso es la de conductoras para ver el efecto de los valores de la F calculada, que es estadístico de prueba.

Para resolver estos problemas en minitab es muy importante construir, por separado, la hoja de trabajo llamada también de cálculo, es muy importante para que este software realice los cálculos correctamente, por lo deben observar con detenimiento el "arreglo" que debe tener, para cada caso. Solo para Minitab.

PROBLEMA 4.1, 2 FACTORES. MAÍZ VS PARCELAS.

Excel.

Hoja de trabajo.

Se siembran 4 tipos de maíz en cinco bloques. Cada bloque está dividido en cuatro parcelas, las cuales se asignan aleatoriamente a los cuatro tipos de maíz.

Determinar al nivel de significancia de 0.05 si las producciones en celemines por acre, como se muestra en la siguiente tabla, varían de manera significativa por diferencias en a) El suelo (es decir, los cinco bloques) y b) el tipo de maíz.

	Tipo de maíz			
	I	II	III	IV
Bloque A	12	15	10	14
Bloque B	15	19	12	11
Bloque C	14	18	15	12
Bloque D	11	16	12	16
Bloque E	16	17	11	14

COMANDOS: >datos>Análisis de datos>Análisis de varianza de dos factores con una sola muestra por grupo>Rango de entrada>cargar datos del cuadro (solo números) >alfa: 0.05>Rango de salida. >aceptar. Listo

Hipótesis:

Ho No existe diferencias significativas entre el Maíz

H1 sí existe diferencias significativas entre el Maíz.

Ho No existe diferencias significativas entre Bloques

H1 Sí existe diferencias significativas entre Bloques

Análisis de varianza de dos factores con una sola muestra por grupo

RESUMEN	*Cuenta*	*Suma*	*Promedio*	*Varianza*
Fila 1	4	51	12.75	4.91666667
Fila 2	4	57	14.25	12.9166667
Fila 3	4	59	14.75	6.25
Fila 4	4	55	13.75	6.91666667
Fila 5	4	58	14.5	7
Columna 1	5	68	13.6	4.3
Columna 2	5	85	17	2.5
Columna 3	5	60	12	3.5
Columna 4	5	67	13.4	3.8

ANÁLISIS DE VARIANZA

Origen de las variaciones	*Suma de cuadrados*	*Grados de libertad*	*Promedio de los cuadrados*	*F*	*Probabilidad*	*Valor crítico para F*
Filas	10	4	2.5	0.64655172	0.63989267	3.25916673
Columnas	67.6	3	22.5333333	5.82758621	0.01075118	3.49029482
Error	46.4	12	3.86666667			
Total	124	19				

Conclusión:

Sí existe diferencias entre los diversos tipos de Maíz pues Fcalculada =5.82> que Fcrítica = 3.49; es decir cae en la zona de rechazo de la Ho.

Respecto a "p" se concluye tambien que p=0.0107< que Nivel de significancia de 0.05 por lo que es baja la probabilidad de que la Ho sea cierta

Respecto a los Bloques o suelos No existe diferencias significativas toda vez que es menor que el valor de F; es decir cae en la zona de aceptación de la Ho. De igual manera p=0.6465 > que el nivel de significancia de 0.05. Es decir; es alta la probabilidad de que la Ho sea cierta.

Minitab.

hoja de trabajo.

Maíz	bloque	producción
1	1	12
1	2	15
1	3	14
1	4	11
1	5	16
2	1	15
2	2	19
2	3	18
2	4	16
2	5	17
3	1	10
3	2	12
3	3	15
3	4	12
3	5	11
4	1	14
4	2	11
4	3	12
4	4	16
4	5	14

Se siembran 4 tipos de maíz en cinco bloques. Cada bloque está dividido en cuatro las cuales se asignan aleatoriamente a los cuatro tipos de maíz.

Determinar al de significancia de 0.05 si las producciones en celemines por acre, como se muestra en la siguiente tabla, varían de manera significativa por diferencias en a) El suelo (bloques) y b) el tipo de maíz.

- Bloque A
- Bloque B
- Bloque C
- Bloque D
- Bloque E

Tipo de maíz			
I	II	III	IV
12	15	10	14
15	19	12	11
14	18	15	12
11	16	12	16
16	17	11	14

COMANDOS:

Cargar hoja de trabajo – Estadística – ANOVA – Modelo lineal gral. – Ajustar modelo lineal gral. – Respuesta: cargar producción C3 – Factores: cargar maíz y bloques C1 y C2 respectivamente. – Aceptar – listo.

Modelo lineal general: producción vs. Maíz, bloque

Método

Codificación de factores (-1, 0, +1)

Información del factor

Factor	Tipo	Niveles	Valores
Maíz	Fijo	4	1, 2, 3, 4
Bloque	Fijo	5	1, 2, 3, 4, 5

Análisis de Varianza

Fuente	GL	SC Ajust.	MC Ajust.	Valor F	Valor p
Maíz	3	67.60	22.533	5.83	0.011
B loque	4	10.00	2.500	0.65	0.640
Error	12	46.40	3.867		
Total	19	124.00			

Resumen del modelo

S	R-cuad.	R-cuad. (ajustado)	R-cuad. (pred)
1.96638	62.58%	40.75%	0.00%

Coeficientes

Término	Coef	EE del coef.	Valor T	Valor p	VIF
Constante	14.000	0.440	31.84	0.000	
Maíz					
1	-0.400	0.762	-0.53	0.609	1.50
2	3.000	0.762	3.94	0.002	1.50
3	-2.000	0.762	-2.63	0.022	1.50
bloque					
1	-1.250	0.879	-1.42	0.181	1.60
2	0.250	0.879	0.28	0.781	1.60
3	0.750	0.879	0.85	0.410	1.60
4	-0.250	0.879	-0.28	0.781	1.60

Ecuación de regresión

Producción = 14.000 - 0.400 Maíz_1 + 3.000 Maíz_2 - 2.000 Maíz_3
- 0.600 Maíz_4
- 1.250 bloque_1 + 0.250 bloque_2 + 0.750 bloque_3
- 0.250 bloque_4
+ 0.500 bloque_5

Conclusión:

Fcrítica a 95% = 3.49 para columnas o maíz

Fcrítica a 95% = 3.25 para bloques

Para el Maíz se rechaza la Ho. Es decir existen diferencias significativas entre el Maíz dado que Fcalculada es mayor que Fcrítica.

Respecto a los bloques dado que Fcalculada es menor que 1 se acepta la Ho.
Es decir no existen diferencias significativas entre Bloques.

Respecto a los valores de p para el maíz es de 0.011< que Alfa de 0.05 la probabilidad de aceptar la Ho.

PROBLEMA 4.2, 2 FACTORES. SHAMPOO.

Excel.

Una marca de Shampoo vende tres tipos de shampoo: Para cabello seco, normal y graso.

En la tabla siguiente se presenta las ventas, en millones de dólares, de los últimos 5 meses. Con un nivel de significancia de 0.05, comprobar si las ventas medias difieren entre los tres tipos de Shampoo y si difieren según el mes.

	Ventas (millones de dólares)		
Mes	Seco	Normal	Graso
Junio	7	9	12
Julio	11	12	14
Agosto	13	11	8
Septiembre	8	9	7
Octubre	9	10	13

COMANDOS:

Datos – análisis de datos-Análisis de varianza dedos factores con una sola muestra por grupo – Aceptar – Rango de entrada: seleccionar todos los datos del problema – Alfa: 0.05- Rango de salida: Seleccionar un espacio vacío – Aceptar. Listo.

Análisis de varianza de dos factores con una sola muestra por grupo

RESUMEN	*Cuenta*	*Suma*	*Promedio*	*Varianza*
Fila 1	3	28	9.33333333	6.33333333
Fila 2	3	37	12.3333333	2.33333333
Fila 3	3	32	10.6666667	6.33333333
Fila 4	3	24	8	1
Fila 5	3	32	10.6666667	4.33333333
Columna 1	5	48	9.6	5.8
Columna 2	5	51	10.2	1.7
Columna 3	5	54	10.8	9.7

ANÁLISIS DE VARIANZA

Origen de las variaciones	*Suma de cuadrados*	*Grados de libertad*	*Promedio de los cuadrados*	*F*	*Probabilidad*	*Valor crítico para F*
Filas	31.733333	4	7.93333333	1.71223022	**0.23969376**	**3.83785335**
Columnas	3.6	2	1.8	0.38848921	**0.69020774**	**4.45897011**
Error	37.0666667	8	4.63333333			
Total	72.4	14				

Ho: Las Medias de ventas de los Shampoos son iguales, no difieren entre sí.
H1: Las Medias de las ventas de los Shampoos SÍ difieren entre sí.

Ho: Las Medias entre los meses no varían entre sí, son iguales.
H1: Las Medias entre los meses SÍ varían entre sí.

REGLA DE DECISIÓN:

Para los tipos de Shampoo: "Si Fcalculada es >que Fcrítica = 4.4589 se rechaza la Ho."

Para los meses: "Si Fcalculada es > Fcrítica 3.8378 se rechaza la Ho."

Conclusión:

Para los tipos de Shampoo: la Fcalculada en los dos casos es menor que la Fcrítica, se acepta la Ho. No se puede rechazar. Así que, en resumen, No se puede rechazar la Ho es decir, no existe diferencia entre los tipos de Shampoo ni en los meses.

Minitab.

	Ventas (millones de dólares)		
Mes	Seco	Normal	Graso
Junio	7	9	12
Julio	11	12	14
Agosto	13	11	8
Septiembre	8	9	7
Octubre	9	10	13

Una marca de Shampoo vende tres tipos de shampoo: Para cabello seco, normal y g.

En la tabla siguiente se presenta las ventas, en millones de dólares, de los últimos meses. Con un nivel de significancia de 0.05, comprobar si las ventas medias difieren entre los tres tipos de Sampoo y si difieren según el mes.

Hoja trabajo.

MES	TIPOS	VENTAS
1	1	7
1	2	9
1	3	12
2	1	11
2	2	12
2	3	14
3	1	13
3	2	11
3	3	8
4	1	8
4	2	9
4	3	7
5	1	9
5	2	10
5	3	13

COMANDOS:

Cargar hoja de trabajo – Estadística – ANOVA – Modelo lineal general – alinear modelo lineal general –Respuestas: Ventas – Factores: mes y ventas –aceptar. Listo

Minitab. Modelo lineal general: VENTAS vs. MES, TIPOS

```
Método

Codificación de factores  (-1, 0, +1)


Información del factor

Factor  Tipo  Niveles  Valores
MES     Fijo        5  1, 2, 3, 4, 5
TIPOS   Fijo        3  1, 2, 3

Análisis de Varianza

Fuente   GL  SC Ajust.  MC Ajust.  Valor F  Valor p
```

```
  MES      4     31.733      7.933     1.71    0.240
  TIPOS    2      3.600      1.800     0.39    0.690
Error      8     37.067      4.633
Total     14     72.400


Resumen del modelo

                    R-cuad.  R-cuad.
      S  R-cuad.  (ajustado)   (pred)
2.15252   48.80%      10.41%    0.00%


Coeficientes

                   EE del
Término      Coef   coef.  Valor T  Valor p   VIF
Constante  10.200   0.556    18.35    0.000
MES
  1         -0.87    1.11    -0.78    0.458  1.60
  2          2.13    1.11     1.92    0.091  1.60
  3          0.47    1.11     0.42    0.686  1.60
  4         -2.20    1.11    -1.98    0.083  1.60
TIPOS
  1        -0.600   0.786    -0.76    0.467  1.33
  2         0.000   0.786     0.00    1.000  1.33


Ecuación de regresión

VENTAS = 10.200 - 0.87 MES_1 + 2.13 MES_2 + 0.47 MES_3 - 2.20 MES_4 + 0.47 MES_5
         - 0.600 TIPOS_1 + 0.000 TIPOS_2 + 0.600 TIPOS_3


Ajustes y diagnósticos para observaciones poco comunes

                           Resid
Obs  VENTAS  Ajuste  Resid   est.
  9    8.00   11.27  -3.27  -2.08  R

Residuo grande R
```

Conclusión:

Regla de decisión: si Fcalculada > que Fcríticase rechaza la Ho.

Dado que en los dos casos la Fcalculada es menor que la Fcrítica, NO se rechaza la Ho. Es decir no existen diferencias significativas en los dos casos (las ventas en los tipos de shampoo y en los meses).

PROBLEMA 4.3. CUADRADOS LATINOS FERTILIZANTES

Excel.

Hoja de trabajo

PARCELAS

2	3	4
C21	D25	B11
B12	A15	C19
A20	C23	D24
D21	B10	A17

Un granjero desea probar los efectos de 4 fertilizantes diferentes (A,B,C, y D) en la producción de trigo y utiliza un arreglo cuadrado latino.

Demostrar si hay diferencias en la producción tanto en los fertilizantes como en las diferentes parcelas usando un nivel de significancia de 0.05 y 0.01. Los datos se presentan en la siguiente tabla.

COMANDOS:

Para hoja de trabajo: abrir software de Minitab, clic en archivo – abrir hoja de trabajo – clic en escritorio si ahí está el archivo hoja de trabajo – clic en t tipo y seleccionar si el archivo está en Excel - clic en nombre, seleccionar – clic en abrir – y listo. Aparece cargada la sección inferior de Minitab.

Para el cálculo:

Clic en estadística_ Clic ANOVA – clic en modelo lineal general – clic en ajustar modelo lineal general – clic en respuestas - Seleccionar yield – factores – seleccionar (una por una) rows, columns, treatment – clic en aceptar – listo.

Minitab.

Un granjero desea probar los efectos de 4 fertilizantes diferentes (A,B, C, y D) en la producción de trigo y utiliza un arreglo cuadrado latino. Demostrar si hay diferencias en la producción tanto en los fertilizantes cómo en las diferentes parcelas usando un nivel de significancia de 0.05 y 0.01.

row	rows	columns	treatment	yield
1	1	1	1	18
2	1	2	3	21
3	1	3	4	25
4	1	4	2	11
5	2	1	4	22
6	2	2	2	12
7	2	3	1	15
8	2	4	3	19
9	3	1	2	15
10	3	2	1	20
11	3	3	3	23
12	3	4	4	24
13	4	1	3	22
14	4	2	4	21
15	4	3	2	10
16	4	4	1	17

COMANDOS:

Para hoja de trabajo: abrir software de Minitab, clic en archivo – abrir hoja de trabajo – clic en escritorio si ahí está el archivo hoja de trabajo – clic en t tipo y seleccionar si el archivo está en Excel – clic en nombre, seleccionar – clic en abrir – y listo. Aparece cargada la sección inferior de Minitab.

Para el cálculo:

Clic en estadística – Clic ANOVA – clic en modelo lineal general – clic en ajustar modelo lineal general – clic en respuestas – Seleccionar yield – factores – leccionar (juntas) rows, columns, treatment - clic en aceptar -listo.

NOTA: muy importante el arreglo de los datos.

Modelo lineal general: yield vs. rows, columns, treatment

```
Factor     Tipo  Niveles  Valores
rows       fijo        4  1, 2, 3, 4
columns    fijo        4  1, 2, 3, 4
treatment  fijo        4  1, 2, 3, 4


Análisis de varianza para yield, utilizando SC ajustada para pruebas

Fuente      GL  SC Sec.  SC Ajust.  CM Ajust.      F      P  FCRIT.
rows         3   29.188     29.188      9.729   4.92  0.047  4.76
columns      3    4.688      4.687      1.562   0.79  0.542  4.76
treatment    3  284.188    284.188     94.729  47.86  0.000  4.76
Error        6   11.875     11.875      1.979
Total       15  329.938
```

NO DA LAS MEDIAS

```
S = 1.40683   R-cuad. = 96.40%   R-cuad.(ajustado) = 91.00%


Observaciones inusuales de yield

                         EE de            Residuo
Obs    yield   Ajuste  ajuste  Residuo  estándar
  3  25.0000  23.1250  1.1122   1.8750      2.18 R

R denota una observación con un residuo estandarizado grande.
```

Conclusión:

Recordar que el valor "p" es el valor mínimo para un nivel de significancia preestablecido en el que se puede rechazar la Ho de igualdad de medias para cierto factor. Para un N.S.=0.05 en los renglones rechazamos Ho. Pero no en las columnas para N.S.= 0.01 tanto en renglones como en columnas se acepta Ho no hay diferencias

```
gl:N-1=4-1=3; a,b y c=N-1=4-1=3 para rows,columns, y tratment. N2-
1-(a+b+c)=16-1-(3+3+3)=6.  Total: n-1=16-1=15
```

DETERMINAR EL VALOR CRÍTICO DE F EN TABLAS PARA COMPARAR Y TOMAR LA DECISIÓN

Fcrítica a 99% =9.78 con 3 y 6 gl y para 95% Fcrítica = 4.76

En conclusión podemos asegurar con un nivel de confianza que entre fertilizantes sí existe una diferencia significativa en el rendimiento en la producción pero no así en las parcelas.
Al nivel de confianza del 99% se puede asegurar que no existe diferencia significativa entre los fertilizantes y entre las parcelas.

En éste arreglo estamos haciendo una prueba de Hipótesis entre los diversos tipos de fertilizantes y a la vez los diversos tipos de parcelas, es decir estamos comparando fertilizantes y parcelas a la vez.

PROBLEMA 4.4. ARREGLO GRECO-LATINO. GASOLINAS.

Hoja de trabajo

Excel

Interesa determinar si existe una diferencia significativa en el rendimiento en millas / galón entre las gasolinas A, B, C, Y D. Diseñar un experimento que utilice 4 conductores, 4 autos y 4 carreteras.

Dado que se trata del mismo número (4) de tipos de gasolina, conductores, autos y carreteras; se puede utilizar un cuadrado-grecolatino. Suponer que los distintos autos están representados por los renglones y los distintos conductores por las columnas, Como se muestra en la tabla.

Ahora se asignan aleatoriamente los distintos tipos de gasolina A, B, C, y D a los renglones y columnas, cubriendo el requisito de que cada letra aparezca solo una vez en cada renglón y en cada columna. Entonces, cada conductor tendrá una oportunidad de conducir cada auto y de usar cada tipo de gasolina y ningún auto será conducido dos veces con el mismo tipo de gasolina.

Luego se asignan aleatoriamente las 4 carreteras, denotadas por: α, β, γ y δ. Cubriendo el mismo requisito impuesto en los cuadrados latinos. Así, cada conductor tendrá también la misma oportunidad de recorrer cada una de las carreteras.

row	car	driver	gasoline	road	MPG
1	1	1	2	3	19
2	1	2	1	2	16
3	1	3	4	4	16
4	1	4	3	1	14
5	2	1	1	4	15
6	2	2	2	1	18
7	2	3	3	3	11
8	2	4	4	2	15
9	3	1	4	1	14
10	3	2	3	4	11
11	3	3	2	2	21
12	3	4	1	3	16
13	4	1	3	2	16
14	4	2	4	3	16
15	4	3	1	1	15
16	4	4	2	4	23

		CONDUCTOR			
	1	2	3	4	
Auto 1	**Bγ 19**	**Aβ 16**	**Dδ 16**	**Cα 14**	**62 A**
Auto 2	**Aδ 15**	**Bα 18**	**Cγ 11**	**Dβ 15**	**81 B**
Auto3	**Dα 14**	**Cδ 11**	**Bβ 21**	**Aγ 16**	**52 C**
Auto4	**Cβ 16**	**Dγ 16**	**Aα 15**	**Bδ 23**	**61 D**

A,B,C Y D SON LOS TIPOS DE GASOLINA A=1, B=2, C=3 Y D=4

CARRETERAS: α, β, γ, y δ 1,2,3,y 4 respectivamente

Minitab.

Interesa determinar si existe una diferencia significativa en el rendimiento de millas/galón, en las gasolinas (A, B, C, Y D). Diseñar un experimento que utilice 4 conductores, 4 autos y 4 carreters.

row	car	driver	gasoline	road	MPG
1	1	1	2	3	19
2	1	2	1	2	16
3	1	3	4	4	16
4	1	4	3	1	14
5	2	1	1	4	15
6	2	2	2	1	18
7	2	3	3	3	11
8	2	4	4	2	15
9	3	1	4	1	14
10	3	2	3	4	11
11	3	3	2	2	21
12	3	4	1	3	16
13	4	1	3	2	16
14	4	2	4	3	16
15	4	3	1	1	15
16	4	4	2	4	23

COMANDOS:

Los mismos que para los problemas anteriores de operadores contra máquinas y fertilizantes y parcelas

NOTA: Es muy importante el arreglo de los datos en la Hoja de trabajo.

Para mayor descripción del problema acudir a la HOJA DE TRABAJO RESPECTIVA YA QUE CUENTA CON MAYOR INFORMACIÓN DEL PROBLEMA.

Se trata de determinar que gasolina nos da el mayor rendimiento tomando en cuenta la influencia de error que implican los autos, los conductores y las carreteras; es por eso que se utiliza un Cuadrado Greco-Latino para eliminar esas tres fuentes de error.

Modelo lineal general: MPG vs. car, driver, gasoline, road

```
Factor    Tipo  Niveles  Valores
car       fijo        4  1, 2, 3, 4
driver    fijo        4  1, 2, 3, 4
gasoline  fijo        4  1, 2, 3, 4
road      fijo        4  1, 2, 3, 4


Análisis de varianza para MPG, utilizando SC ajustada para prueba

Fuente    GL  SC Sec.  SC Ajust.  CM Ajust.      F      P
car        3   16.500     16.500      5.500   2.75  0.214
driver     3    6.500      6.500      2.167   1.08  0.475
gasoline   3  111.500    111.500     37.167  18.58  0.019
road       3    7.500      7.500      2.500   1.25  0.429
Error      3    6.000      6.000      2.000
Total     15  148.000

S = 1.41421   R-cuad. = 95.95%   R-cuad.(ajustado) = 79.73%


Ho:   A=B=C=D QUE DAN EL MISMO RENDIMIENTO

H1:   que las gasolinas son diferentes resultados en rendimientos
```

CONCLUSIONES:

Para valores críticos de F: F a 0.95, 3 y3 F=9.28; F= 0.99 3, 3,=29.5. Por lo tanto, se puede rechazar la Ho de que las gasolinas son iguales al N.S. de 0.05 pero no al n.s. de 0.01.

CALCULO DE gl: total.N2-1=4x4(-1)=16. car,driver,gasoline yroad:N-1 = 4-1 =3
del Error: N2-1-4(N-1)=16-1-4(4-1) = 3

NOTA: para este caso: N = 4 N2 = N al cuadrado

SE CONCLUYE QUE A UN NIVEL DE CONFIANZA DEL 95% SE PUEDE ASEGURAR EUR ENTRE GASOLINAS SI HAY DIFERENCIAS SIGNIFICATIVAS EN CUANTO AL RENDIMIENTO. PERO NO ASÍ AL NIVEL DE CONFIANZA DEL 99%.

UNIDAD V.

DISEÑOS FACTORIALES

INTERACCIÓN

En Minitab se maneja el término de "**Interacción**" entre dos factores y considero importante explicar el concepto. La Interacción tiene lugar si la combinación de dos factores ejerce algún efecto sobre la variable de estudio, además de hacerlo en cada factor por sí misma. A la variable de estudio se le llama variable de "Respuesta".

Para aclarar más el concepto de Interacción, se tiene el siguiente ejemplo ilustrativo: Dieta y ejercicio hará bajar de peso a una persona. Solo la dieta bajara de peso a una persona. Solo el ejercicio lo bajara de peso; pero, si se combina dieta con ejercicio se encontrara un mayor efecto en bajar de peso.

Podemos decir que dieta y ejercicio hay interacción entre ellas.

PROBLEMA 5.1. CULTIVOS VS FERTILIZANTES.

Excel.

	cultivo 1	cultivo 2	cultivo 3	cultivo 4
fretilizante A	4.5	6.4	7.2	6.7
fertilizante B	8.8	7.8	9.6	7
fertilizante C	5.9	6.8	5.7	5.2

La siguiente tabla muestra la producción por acres, de cuatro cultivos sembrados en lotes tratados con tres tipos de fertilizantes.

Determinar si existe diferencia en la producción por acre: a) debido a los fertilizantes y b) por cultivos

Utilizar un nivel de confianza del 95%.

Para fertilizantes:

Ho: $\mu A=\mu B=\mu C$ y no hay diferencias significativas entre la producción debida a los fertilizantes

H_1: $\mu A \ddagger \mu B \ddagger \mu C$ y sí hay diferencias significativas entre la producción debida a los fertilizantes.

Para los cultvos:

Ho: μ1=μ2=μ3=μ4 y no hay diferencias significativas entre la producción debido a los cultivos

H1: μ1‡μ2‡μ3‡μ4 y sí hay diferencias significativas entre la producción debido a los cultivos

α= 0.01 y 0.05

Estadístico de prueba "F"

Regla de decisión: si Fcalc.>Fcrít. Se rechaza la Ho de lo contrario se acepta.

COMANDOS: Para dos factores

Datos – Análisis de datos – Análisis de datos de dos factores con una sola muestra por grupo – clic con Alfa = 0.05 -clic en rango de entrada y seleccionar los datos del problema – clic – en rango de salida y con el cursor seleccionar un espacio vacío – Aceptar. Listo

Análisis de varianza de dos factores con una sola muestra por grupo

RESUMEN	*Cuenta*	*Suma*	*Promedio*	*Varianza*
Fila 1	4	24.8	6.2	1.3933333
Fila 2	4	33.2	8.3	1.2933333
Fila 3	4	23.6	5.9	0.4466666
Columna 1	3	19.2	6.4	4.81
Columna 2	3	21	7	0.52
Columna 3	3	22.5	7.5	3.87
Columna 4	3	18.9	6.3	0.93

ANÁLISIS DE VARIANZA

Origen de las variaciones	*Suma de cuadrados*	*Grados de libertad*	*Promedio de los cuadrados*	*F*	*Probabilidad*	*Valor crítico para F*
Filas	13.68	2	6.84	6.23708207	**0.034258**	**5.14325285**
Columnas	2.82	3	0.94	0.85714286	**0.512185**	**4.75706266**
Error	6.58	6	1.096666667			
Total	23.08	11				

Para cultivos:

Hipótesis:

Ho no existe diferencias significativas entre Cultivos.

H1 Sí existe diferencias significativas entre cultivos.

Para Fertilizantes:

Hipótesis:

Ho No existe diferencias entre la producción de los fertilizantes.

H1 Sí existe diferencias significativas entre la producción de los fertilizantes.

Cultivos son las columnas: Dado que Fcrítica = 4.75 > Fcalculada = 0.857 se acepta la Ho pues cae en la Zona de aceptación de la Ho, Es decir NO existe diferencias entre los cultivos en su producción.

Respecto a p= 51.2% > que el nivel de significancia de 0.05; por lo que también es alta la probabilidad de que la Ho sea cierta.

Fertilizantes son las Filas: Dado que Fcrítica = 5.14 < que Fcalculada = 6.23 se rechaza la Ho pues cae en la Zona de rechazo de la Ho. Es decir; sí existen diferencias en la producción de los fertilizantes.

Respecto a p = 3.42 % < que 5% que es el valor del nivel de significancia de la Ho. por ser baja la probabilidad de que la Ho sea cierta se rechaza ésta.

Minitab.

La siguiente tabla muestra la producción por acre, de cuatro cultivos en lotes y tratados con tres tipos diferentes de fertilizantes.

Determinar si existe diferencia significativa en la producción por acre, a) debido a los fertilizantes y b) por los cultivos.

Cargar hoja de trabajo.

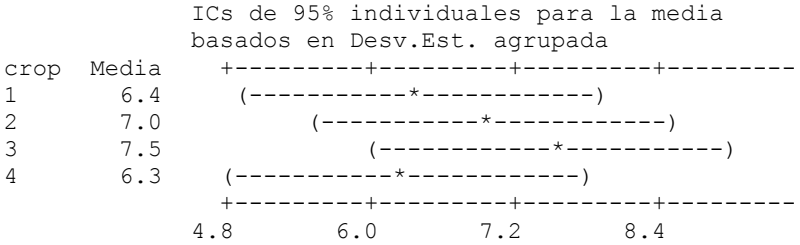
crop	fertilizer	yield
1	1	4.5
1	2	8.8
1	3	5.9
2	1	6.4
2	2	7.8
2	3	6.8
3	1	7.2
3	2	9.6
3	3	5.7
4	1	6.7
4	2	7
4	3	5.2

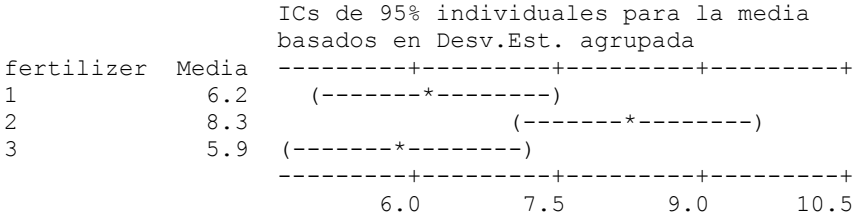
COMANDOS:

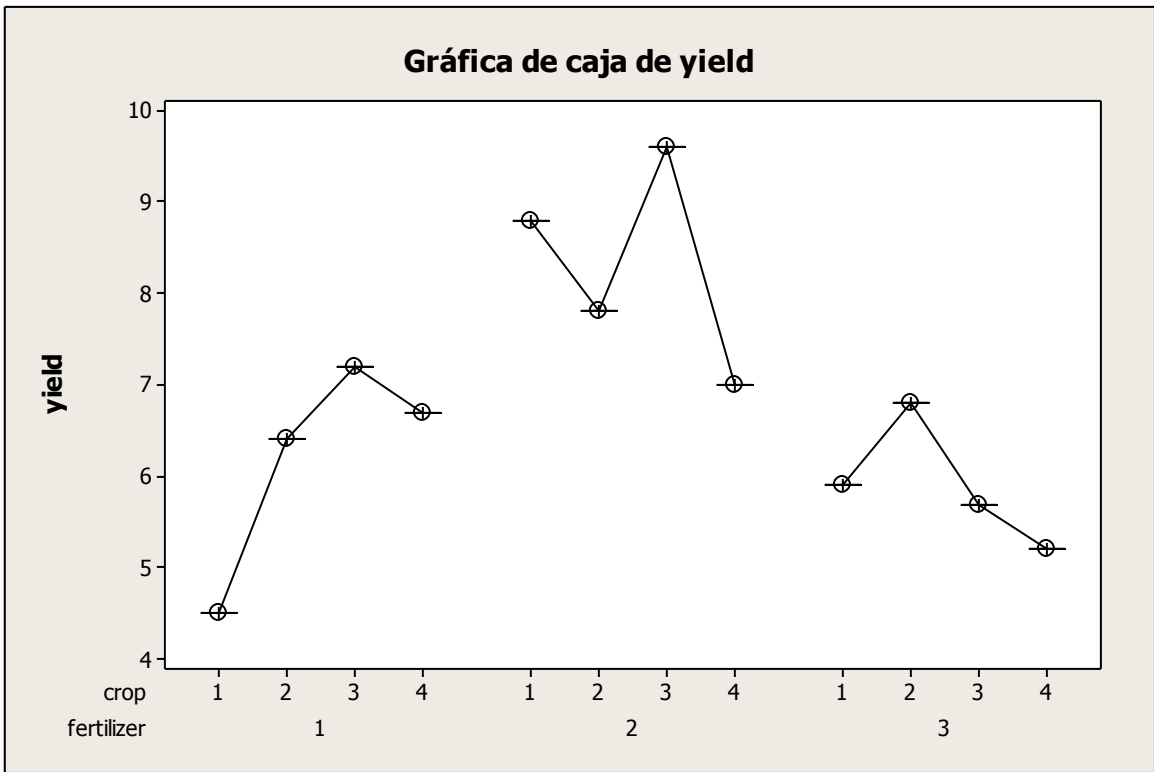
Estadística – ANOVA – modelo lineal gral. – ajustar modelo lineal gral. – Respuestas: Yield- Factores: crop fertilizer – aceptar. Listo

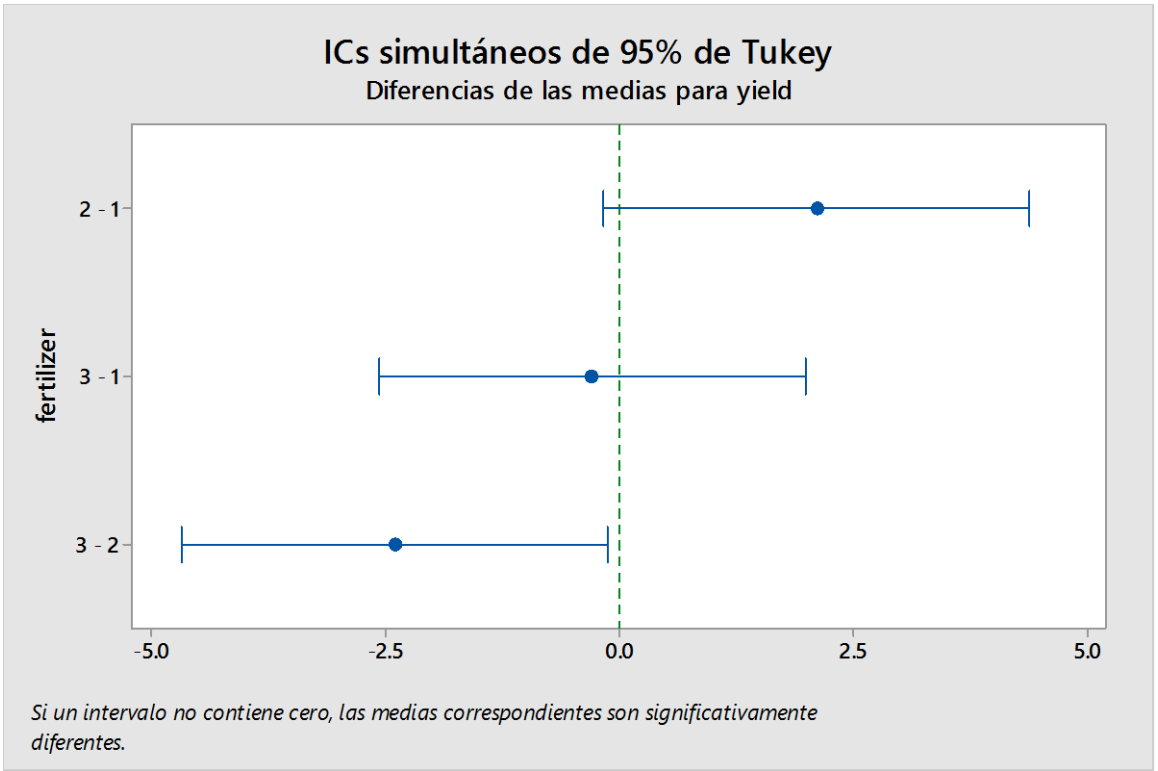
ANOVA de dos factores: yield vs. crop, fertilizer

```
eFuente      GL     SC       CM     F      P
crop          3   2.82  0.94000  0.86  0.512
fertilizer    2  13.68  6.84000  6.24  0.034
Error         6   6.58  1.09667
Total        11  23.08

S = 1.047   R-cuad. = 71.49%   R-cuad.(ajustado) = 47.73%
```






Gráfica de caja de yield
yield
10
9
8
7
6
5
4
crop
1 2 3 4 1 2 3 4 1 2 3 4
fertilizer
1
2
3


ICs simultáneos de 95% de Tukey
Diferencias de las medias para yield
fertilizer
2 - 1
3 - 1
3 - 2
-5.0
-2.5
0.0
2.5
5.0
Si un intervalo no contiene cero, las medias correspondientes son significativamente diferentes.

Conclusión:

F.05,2,6 =5.14 : Existe una diferencia significativa en la producción debida a los fertilizantes.

Porque Fcalculada=6.24>Fcrítica =5.14.

Como F<1 (o.86) se concluye que no hay diferencia significativa en la producción por los cultivos.
p=0.512 es el nivel mínimo de significancia con el que se podría rechazar una diferencia en la producción media de los cultivos. Es alta la probabilidad de aceptar la Ho.

PROBLEMA 5.2 2 FACTORES CON RÉPLICA

Excel.
Hoja de trabajo.

Un fabricante desea determinar la efectividad de 4 tipos de máquinas (A,B,C,y D) en la producción de tuercas. Para logra esto, se tiene el número de tuercas defectuosas producidas por cada máquina en cada día de cierta semana, en dos turnos de trabajo. Con un nivel de significancia de 0.05; a) existe diferencias entre las máquinas, b) entre los turnos.

Probar con "p" si existe diferencias entre máquinas y entre turnos. Probar.

		1 PRIMER TURNO					2 **SEGUNDO TURNO**			
	máquina	lunes	martes	miercoles	jueve	viernes	**lunes**	**martes**	**mierc.**	**jueves**
1	A	6	4	5	5	4	**5**	**7**	**4**	**6**
2	B	10	8	7	7	9	**7**	**9**	**12**	**8**
3	C	7	5	6	5	9	**9**	**7**	**5**	**4**
4	D	8	4	6	5	5	**5**	**7**	**9**	**7**

ARREGLO DE DATOS

row	machine	shift	defects
1	1	1	6
2	1	1	4
3	1	1	5
4	1	1	5
5	1	1	4
6	1	2	5
7	1	2	7
8	1	2	4
9	1	2	6
10	1	2	8
11	2	1	10
12	2	1	8
13	2	1	7
14	2	1	7
15	2	1	9
16	2	2	7
17	2	2	9
18	2	2	12
19	2	2	8
20	2	2	8
21	3	1	7
22	3	1	5
23	3	1	6
24	3	1	5
25	3	1	9
26	3	2	9
27	3	2	7
28	3	2	5
29	3	2	4
30	3	2	6
31	4	1	8
32	4	1	4
33	4	1	6
34	4	1	5
35	4	1	5
36	4	2	5
37	4	2	7
38	4	2	9
39	4	2	7
40	4	2	10

Minitab.

row	machine	shift	defects
1	1	1	6
2	1	1	4
3	1	1	5
4	1	1	5
5	1	1	4
6	1	2	5
7	1	2	7
8	1	2	4
9	1	2	6
10	1	2	8
11	2	1	10
12	2	1	8
13	2	1	7
14	2	1	7
15	2	1	9
16	2	2	7
17	2	2	9
18	2	2	12
19	2	2	8
20	2	2	8
21	3	1	7
22	3	1	5
23	3	1	6
24	3	1	5
25	3	1	9
26	3	2	9
27	3	2	7
28	3	2	5
29	3	2	4
30	3	2	6
31	4	1	8
32	4	1	4
33	4	1	6
34	4	1	5
35	4	1	5
36	4	2	5
37	4	2	7
38	4	2	9
39	4	2	7
40	4	2	10

Un fabricante desea determinar la efectividad de 4 tipos de máquinas (A, B, C y D) en la producción de tuercas. Para lograr esto, se obtiene el número de tuercas defectuosas producidas por cada máquina en cada día de cierta semana, en dos turnos de trabajo; los resultados se muestran en la tabla de datos de manera ordenada como se indica.

Realizar un análisis de varianza para determinar, al nivel de significancia del 0.05 si hay diferencias: a) entre las máquinas, y b) entre los turnos.

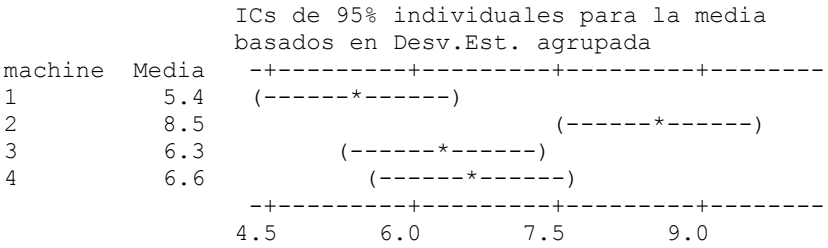
Utilizar Minitab para el análisis de varianza y probar si hay diferencias entre las máquinas y entre los turnos con el valor de "p" para probar.

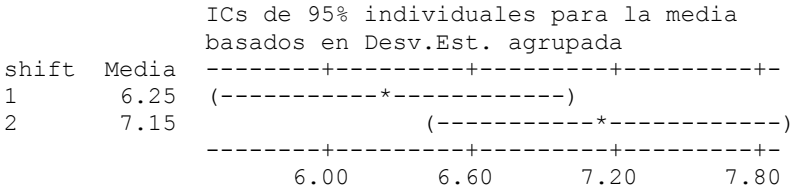
COMANDOS:

Cargar hoja de trabajo – Estadística – ANOVA – Modelo lineal Gral. – Ajustar Modelo lineal gral. –Respuestas:defects – Factores:machine chift – Aceptar. Listo.

ANOVA de dos factores: defects vs. machine, shift

```
Fuente       GL     SC       CM     F      P
machine       3   51.0  17.0000  6.42  0.002
shift         1    8.1   8.1000  3.06  0.090
Interacción   3    6.5   2.1667  0.82  0.494
Error        32   84.8   2.6500
Total        39  150.4

S = 1.628   R-cuad. = 43.62%   R-cuad.(ajustado) = 31.28%
```





gl por renglones 4-1=3; por columnas Dos turnos) =2-1=1; para encontrar los gl debido a la interacción se observa que hay 8 entradas en la tabla de sumas por máq,días,turno:8-1=7; puesto que 3 de estos 7 gl son consecuencia de los renglones y 1 de las columnas, los gl sobrantes 7-(3+1)=3 son por la interacción de la tabla original n=40: n-1=39 en total. y en la residual o error :39-7=32

F0.95,3,32,=2.90 F(6.42)>Fc(2.90) Existe una diferencia significativa entre las máquinas

El valor de P 0.002 indica que es muy pequeño para aceptar la Ho < que 0.05: se rechaza la Ho en máquinas.

F0.95,1,32,=4.15 F(3.06)<Fc4.15) No existe diferencia en los defectos debido a los turnos.

El valor de P en turnos es 0.09 > que 0.05 por lo que se acepta la Ho No hay diferencias en turnos.

shift:turno el valor de P 0.494 de interacción entre turnos y máquinas no hay diferencia significativa F <1 se acepta la Ho.

PROBLEMA 5.3. 2 FACTORES. RUTA DE AUTOBUSES.

Excel.

Una compañía en USA desea ampliar sus rutas de autobuses y se llevaron a cabo pruebas por 4 rutas para determinar los tiempos de recorridos y determinar las diferencias entre ellas. Se comparan Rutas y Choferes (De dos vías).

Driver	US-6	West End	Hickory St	Ruta 59
Deans	18	17	21	22
Snaverly	16	23	23	22
Ormson	21	21	26	22
Zollaco	23	22	29	25
Filbeck	25	24	28	28

Para iniciar, se realiza una prueba de hipótesis de un factor o una vía, es decir, solo se considera las 4 rutas.

Para ver el efecto de la variable de bloqueo y se observa un cambio en el valor de "F". Se tratas de un análisis de varianza de dos vías, donde se toma en cuenta el efecto de la otra vía o factor sobre la otra.

Ho: las medias de bloques o rutas son iguales. $\mu_1=\mu_2=\mu_3=\mu_4$

H1: las medias de los bloques o rutas son diferentes$\mu_1\ddagger\mu_2\ddagger\mu_3\mu\ddagger\mu_4$

Ho: las medias de los tratamientos o choferes son iguales. $\mu_1=\mu_2=\mu_3=\mu_4=\mu_5$

H1: las medias de los tratamientos o choferes son diferentes. $\mu_1\ddagger\mu_2\ddagger\mu_3\mu\ddagger\mu_4\ddagger\mu_5$

COMANDOS:

Primero cargar la hoja de Excel con los datos de la tabla y el enunciado del problema. Resolver el problema primero como de un factor para ver el efecto de la variable de bloqueo que en este caso son conductores: Datos – Análisis de varianza de un factor – Rango de entrada – Marcar todos los datos del cuadro – clic en agrupados por columnas-clic en alfa = 0.05 -clic en rango de salida y con el cursor seleccionar un espacio vacío – listo.

Análisis de varianza de un factor

RESUMEN

Grupos	*Cuenta*	*Suma*	*Promedio*	*Varianza*
Columna 1	5	103	20.6	13.3
Columna 2	5	107	21.4	7.3
Columna 3	5	127	25.4	11.3
Columna 4	5	119	23.8	7.2

ANÁLISIS DE VARIANZA

Origen de las variaciones	*Suma de cuadrados*	*Grados de libertad*	*Promedio de los cuadrados*	*F*	*Probabilidad*	*Valor crítico para F*
Entre grupos	72.8	3	24.2666667	2.48252344	0.09810502	3.23887152
Dentro de los grupos	156.4	16	9.775			
Total	229.2	19				

Para dos factores:

Datos – Análisis de datos – Análisis de datos de dos factores con una sola muestra por grupo – clic en alfa con 0.05 – Clic en rango de salida y con el curso seleccionar un espacio vacío – Aceptar – listo.

Análisis de varianza de dos factores con una sola muestra por grupo

RESUMEN	*Cuenta*	*Suma*	*Promedio*	*Varianza*
Fila 1	4	78	19.5	5.66666667
Fila 2	4	84	21	11.3333333
Fila 3	4	90	22.5	5.66666667
Fila 4	4	99	24.75	9.58333333
Fila 5	4	105	26.25	4.25
Columna 1	5	103	20.6	13.3
Columna 2	5	107	21.4	7.3
Columna 3	5	127	25.4	11.3
Columna 4	5	119	23.8	7.2

ANÁLISIS DE VARIANZA

Origen de las variaciones	*Suma de cuadrados*	*Grados de libertad*	*Promedio de los cuadrados*	*F*	*Probabilidad*	*Valor crítico para F*
Filas	119.7	4	29.925	9.78474114	0.00093357	3.25916673
Columnas	72.8	3	24.2666667	7.9346049	0.00350791	3.49029482
Error	SSE 36.7	12	3.05833333			
Total	229.2	19				

Observar como el efecto de los conductores influye aumentando el valor de "F" de 2.48 a 9.784. A esta variable de los "conductores" se le llama VARIABLE DE BLOQUEO, porque es las rutas lo que analizamos y conductores es la segunda variable incluida

DEFINICION:

Variable de bloqueo es una segunda variable de tratamiento que, cuando se incluye en el análisis de ANOVA, tendrá el efecto de reducir el término SSE dando como consecuencia el incremento de F calculada.

Conclusión:

Se concluye que los tratamientos como los bloques, es decir las rutas y los conductores difieren significativamente.

En ambos casos se rechaza la Ho y se acepta la H1 porque en ambos casos la Fcalculada > que la Fcrítica, es decir, para las filas o sea conductores Fcalculada=9.78>Fcrítica =3.259 y para las columnas po sea Rutas Fcalculada=7.93 > que Fcrítica=3.49

Por lo que se concluye que Si existen diferencias significativas tanto en conductores como en las rutas.

PROBLEMA 5.4 ARTICULOS PRODUCIDOS.

Excel.

Los artículos fabricados por una empresa son producidos por 3 operadores que utilizan, cada uno, una máquina.

El fabricante busca determinar si existe alguna diferencia: a) Entre los operadores y b) entre las máquinas. Se lleva a cabo un experimento para determinar el número de artículos diarios producidos por cada operador manejando cada máquina.

Usar un nivel de significancia del 0.05.

MÁQUINAS	OPERADOR 1	OPERADOR 2	OPERADOR 3
A	23	27	24
B	34	30	28
C	28	25	27

COMANDOS: PARA DOS FACTORES

DATOS – ANÁLISIS DE DATOS – ANÁLISIS DE DATOS DE DOS FACTORES CON UNA SOLA MUESTRA POR GRUPO – CLIC EN RANGO DE ENTRADA Y SELECCIONAR TODOS LOS DATOS DEL PROBLEMA.

CLIC EN ALFA CON 0.05- CLIC EN RANGO DE SALIDA Y CON EL CURSOR SELECCIONAR UN ESPACIO VACÍO – ACEPTAR –LISTO.

Análisis de varianza de dos factores con una sola muestra por grupo

RESUMEN	*Cuenta*	*Suma*	*Promedio*	*Varianza*
Fila 1	3	74	24.6666667	4.33333333
Fila 2	3	92	30.6666667	9.33333333
Fila 3	3	80	26.6666667	2.33333333
Columna 1	3	85	28.3333333	30.3333333

Columna 2	3	82	27.3333333	6.33333333
Columna 3	3	79	26.3333333	4.33333333

ANÁLISIS DE VARIANZA

Origen de las variaciones	*Suma de cuadrados*	*Grados de libertad*	*Promedio de los cuadrados*	*F*	*Probabilidad*	*Valor crítico para F*
Filas	56	2	28	4.30769231	**0.1005354**	**6.94427191**
Columnas	6	2	3	0.46153846	**0.66015625**	**6.94427191**
Error	26	4	6.5			
Total	88	8				

Hipótesis:

Ho. No existe una diferencia significativa entre los operadores de las máquinas.
H1. Sí existe diferencias significativas entre el rendimiento de los operadores.

Ho No existe diferencias significativas entre el rendimiento de las máquinas.
H1 Sí existe diferencias significativas en el rendimiento de las máquinas.

Cconclusiones:

Respecto a los Operadores se concluye que, dado que Fcrítica = 6.94>que Fcalculada = 4.30 se acepta la Ho de que no hay diferencias entre los Operadores en cuanto a su rendimiento.

En cuanto al rendimiento de las máquinas tambien se acepta la Ho de que no hay diferencias entre las máquinas porque la Fcrítica = 6.94 es mucho mayor que la Fcalculada= 0.461.

Ambas caen en lazona de aceptación de la Ho.

De igual manera se concluye respecto a la probabilidad de que la Ho sea cierta, ya que para los operadores p=0.1005 > 0.05 (nivel de significancia) es decir, existe una probabilidad del 10.05 % de que la Ho sea cierta. Respecto a las máquinas p=0.6601 > 0.05 (nivel de significancia) es decir, existe la probabilidad del 66.01 % de que la Ho sea cierta.

Minitab.

MÁQUINAS	OPERADOR 1	OPERADOR 2	OPERADOR 3
A	23	27	24
B	34	30	28
C	28	25	27

Los artículos fabricados por una empresa son producidos por 3 operadores que utilizan, cada uno, una máquina. El fabricante busca determinar si existe alguna diferencia:

a) Entre los operadores

b) entre las máquinas.

Se lleva a cabo un experimento para determinar el número de artículos diarios producidos por cada operador manejando cada máquina. Usar un nivel de significancia del 0.05.

Hoja de trabajo.

operador	máquinas	producción
1	1	23
1	2	34
1	3	28
2	1	27
2	2	30
2	3	25
3	1	24
3	2	28
3	3	27

COMANDOS:

Para hoja de trabajo: abrir software de Minitab, clic en archivo – abrir hoja de trabajo – clic en escritorio si ahí está el archivo hoja de trabajo – clic en t tipo y seleccionar si el archivo está en Excel – clic en nombre, seleccionar – clic en abrir – y listo. Aparece cargada la sección inferior de Minitab.

Para el cálculo:

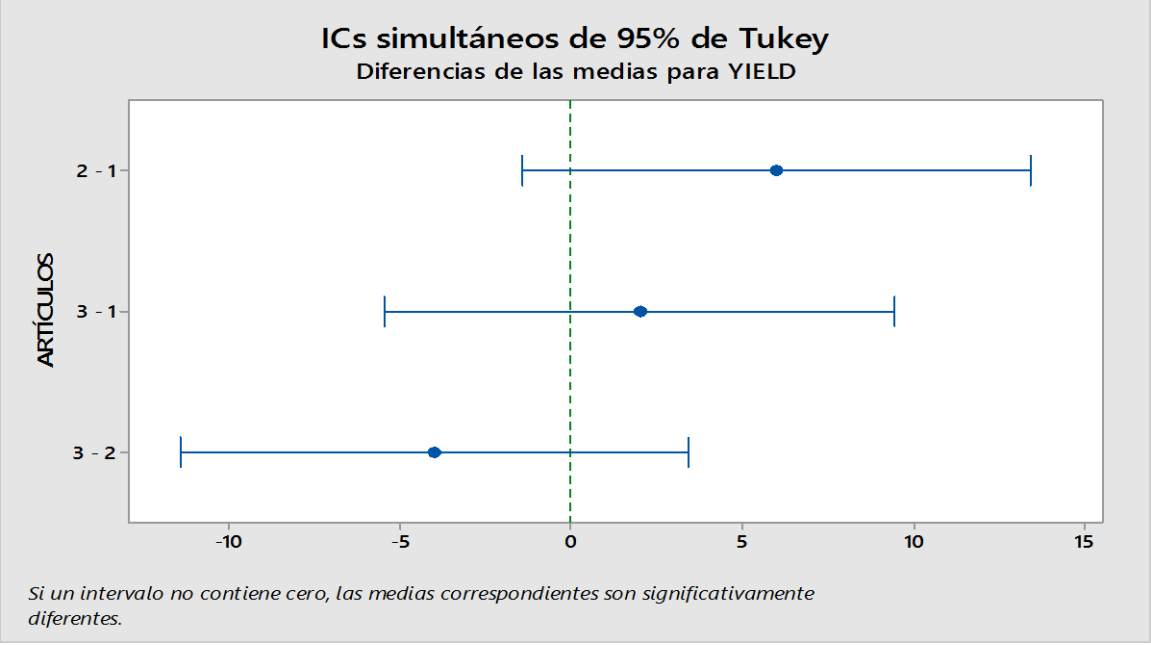
Clic en estadística – Clic ANOVA – clic en modelo lineal general - clic en ajustar modelo lineal genetral - clic en respuestas - Seleccionar yield – Factores; seleccionar (todas juntas) row – group – artículos – clic en aceptar. Listo.

NOTA: muy importante el arreglo de los datos.

gráficas: modelo lineal gral. – otra vez – comparaciones-respuesta:producción – tipo de comparaciones:en parejas – método:Fisher o Tukey – elegir términos para comparación: doble clic de c/u de las variables – aceptar. Aceptar

ANOVA de dos factores: YIELD vs. GROUP, ARTÍCULOS

```
Fuente      GL  SC     CM      F       P
GROUP        2   6    3.0   0.46   0.660
ARTÍCULOS    2  56   28.0   4.31   0.101      fCRÍT,0.95 2,4=6.94
Error        4  26    6.5
Total        8  88

S = 2.550    R-cuad. = 70.45%    R-cuad.(ajustada)


                ICs de 95% individuales para la media

                basados en Desv.Est. agrupada
GROUP    Media  ------+---------+---------+---------+---
1      28.3333          (------------*--------  -----)
2      27.3333      (-------------*-------------)
3      26.3333  (-------------*------------)
                ------+---------+---------+-         --------+---
                   24.0      27.0      30.0      33.0


                  ICs de 95% individuales para la media
                  basados en Desv.Est. agrupada
ARTÍCULOS    Media  ---------+---------+---------+---------+
1          24.6667  (----------*---------)
2          30.6667                (----------*---------)
3          26.6667       (----------*---------)
                  ---------+---------+---------+---------+
                        24.0      28.0      32.0      36.0
```



Conclusiones:

No existe diferencia significativa entre la producción

Entre la producción tanto de operadores cómo entre máquinas dado que la Fcalculada es menor que la Fcrítica en ambos casos, es decir, entre operadores y entre máquinas.

De igual manera p es mucho mayor que 0.05 (nivel de significancia) en ambos casos.

También podemos observar en la gráfica de intervalos que no se traslapan por lo que no hay diferencias significativas tanto en operadores cómo en máquinas.

Bibliografía

Murray R. Spiegel y Larry J. Stephens. Estadística (Schaum). Tercera Edición y Cuarta. Mc Graw Hill.

Douglas A. LInd, William G. Marchal y Samuel A Wathen. Estadística Aplicada a los Negocios y la Economía. 15 Edición. Mc Graw Hill.

Richard I Levin. Astadística Para Administración y Economía.7a. Edición revisada. Editorial Pearson.

Walohole. Probabilidad y Estadística. Editorial Mc Graw Hill.

Montgomery. Douglas C. Diseño y Análisis de Experimento. Grupo Editorial.

ANEXOS

ANEXO 1.

Tabla 1. Tamaño de la muestra para el intervalo de confianza de una orció

CONFIANZA=80% y N=200											
	VALOR INICIAL PARA LA PROPORCIÓN										
B											
	p=.01	p=.05	p=.10	p=.15	p=.20	p=.25	p=.30	p=.35	p=.40	p=.45	p=.50
0.01	90	160	177	183	186	188	190	190	191	191	191
0.02	34	100	131	145	154	159	163	165	167	168	168
0.03	17	61	91	108	119	127	132	136	138	139	140
0.04	10	40	64	80	91	99	105	109	111	113	113
0.05	7	28	46	60	70	77	82	86	89	90	91
0.06	5	20	35	46	54	61	66	69	72	73	73
0.07	4	15	27	36	43	49	53	56	58	59	60
0.08	3	12	21	29	35	39	43	46	48	49	49
0.09	2	10	17	24	29	33	36	38	40	41	41
0.10	2	8	14	20	24	27	30	32	34	34	35
0.15	1	4	7	9	12	13	15	16	17	17	17
0.20	1	2	4	6	7	8	9	9	10	10	10

CONFIANZA=80% y N=100											
	VALOR INICIAL PARA LA PROPORCIÓN										
B											
	p=.01	p=.05	p=.10	p=.15	p=.20	p=.25	p=.30	p=.35	p=.40	p=.45	p=.50
0.01	63	89	94	96	97	97	98	98	98	98	98
0.02	30	67	79	85	87	89	90	91	91	92	92
0.03	16	47	63	71	75	78	80	81	82	83	83
0.04	10	34	49	57	63	67	69	71	72	72	73
0.05	7	24	38	46	52	56	59	61	62	63	63
0.06	5	18	30	38	43	47	50	52	53	54	54
0.08	3	11	19	25	30	33	36	38	39	40	40
0.09	2	9	16	21	25	28	31	32	33	34	34
0.10	2	8	13	18	21	24	26	28	29	30	30
0.15	1	4	7	9	11	13	14	15	16	16	16
0.20	1	2	4	6	7	8	9	9	10	10	10

CONFIANZA=80% y N=75											
VALOR INICIAL PARA LA PROPORCIÓN											
B											
	p=.01	p=.05	p=.10	p=.15	p=.20	p=.25	p=.30	p=.35	p=.40	p=.45	p=.50
0.01	52	69	72	73	73	74	74	74	74	74	74
0.02	27	55	63	66	68	69	70	70	70	70	70
0.03	15	41	52	57	60	62	63	64	65	65	65
0.04	10	30	42	48	52	55	56	57	58	59	59
0.05	7	23	34	40	45	47	49	51	52	52	52
0.06	5	17	27	34	38	41	43	44	45	46	46
0.07	4	14	22	28	32	35	37	39	40	40	40
0.08	3	11	18	24	27	30	32	34	35	35	35
0.09	2	9	15	20	23	26	28	29	30	31	31
0.10	2	8	13	17	20	23	24	26	27	27	27
0.15	1	4	7	9	11	12	13	14	15	15	15
0.20	1	2	4	5	7	8	8	9	9	10	10

CONFIANZA=80% y N=50											
VALOR INICIAL PARA LA PROPORCIÓN											
B											
	p=.01	p=.05	p=.10	p=.15	p=.20	p=.25	p=.30	p=.35	p=.40	p=.45	p=.50
0.01	39	48	49	49	50	50	50	50	50	50	50
0.02	23	40	45	46	47	48	48	48	48	48	48
0.03	14	32	39	42	43	44	45	45	45	46	46
0.04	9	25	33	37	39	40	41	42	42	42	42
0.05	6	20	28	32	35	36	37	38	39	39	39
0.06	5	16	23	28	30	32	34	34	35	35	35
0.07	4	13	20	24	27	29	30	31	32	32	32
0.08	3	10	17	21	23	25	27	28	28	29	29
0.09	2	9	14	18	20	22	24	25	25	26	26
0.10	2	7	12	15	18	20	21	22	23	23	23
0.15	1	4	6	8	10	11	12	13	14	14	14
0.20	1	2	4	5	6	7	8	9	9	9	9

CONFIANZA=80% y N=40											
VALOR INICIAL PARA LA PROPORCIÓN											
B											
	p=.01	p=.05	p=.10	p=.15	p=.20	p=.25	p=.30	p=.35	p=.40	p=.45	p=.50
0.01	33	39	39	40	40	40	40	40	40	40	40
0.02	21	34	37	38	38	39	39	39	39	39	39
0.03	13	28	33	35	36	36	37	37	37	37	37
0.04	9	23	29	31	33	34	34	35	35	35	35
0.05	6	18	25	28	30	31	32	32	33	33	33
0.06	5	15	21	24	27	28	29	30	30	30	30
0.07	4	12	18	21	24	25	26	27	27	28	28
0.08	3	10	15	19	21	23	24	24	25	25	25
0.09	2	8	13	16	19	20	21	22	23	23	23
0.10	2	7	11	14	17	18	19	20	21	21	21
0.15	1	4	6	8	10	11	12	12	13	13	13
0.20	1	2	4	5	6	7	8	8	9	9	9

CONFIANZA=80% y N=30											
VALOR INICIAL PARA LA PROPORCIÓN											
B											
	p=.01	p=.05	p=.10	p=.15	p=.20	p=.25	p=.30	p=.35	p=.40	p=.45	p=.50
0.01	26	29	30	30	30	30	30	30	30	30	30
0.02	18	27	28	29	29	29	30	30	30	30	30
0.03	12	23	26	27	28	28	28	29	29	29	29
0.04	8	19	23	25	26	27	27	27	27	27	27
0.05	6	16	21	23	24	25	25	26	26	26	26
0.06	5	13	18	21	22	23	24	24	24	24	24
0.07	4	11	16	18	20	21	22	22	23	23	23

CONFIANZA=80% y N=30											
VALOR INICIAL PARA LA PROPORCIÓN											
B											
	p=.01	p=.05	p=.10	p=.15	p=.20	p=.25	p=.30	p=.35	p=.40	p=.45	p=.50
0.01	26	29	30	30	30	30	30	30	30	30	30
0.02	18	27	28	29	29	29	30	30	30	30	30
0.03	12	23	26	27	28	28	28	29	29	29	29
0.04	8	19	23	25	26	27	27	27	27	27	27
0.05	6	16	21	23	24	25	25	26	26	26	26
0.06	5	13	18	21	22	23	24	24	24	24	24
0.07	4	11	16	18	20	21	22	22	23	23	23
0.08	3	9	14	16	18	19	20	21	21	21	21
0.09	2	8	12	15	16	18	18	19	19	20	20
0.10	2	7	11	13	15	16	17	17	18	18	18
0.15	1	4	6	8	9	10	11	11	12	12	12
0.20	1	2	4	5	6	7	7	8	8	8	8

Tabla 1. Tamaño de la muestra para el intervalo de una confianza de una proporción.

CONFIANZA=90% y N=10,000											
VALOR INICIAL PARA LA PROPORCIÓN											
B											
	p=.01	p=.05	p=.10	p=.15	p=.20	p=.25	p=.30	p=.35	p=.40	p=.45	p=.50
0.01	261	1140	1959	2566	3022	3367	3624	3811	3938	4012	4036
0.02	67	312	574	795	977	1126	1245	1334	1397	1435	1447
0.03	30	141	264	370	460	534	594	641	674	693	700
0.04	17	80	150	212	264	308	344	371	391	402	406
0.05	11	52	97	137	171	199	223	241	254	261	264
0.06	8	36	68	95	119	139	156	169	178	183	185
0.07	6	27	50	70	88	103	115	125	131	135	137
0.08	5	21	38	54	68	79	89	96	101	104	105
0.09	4	16	30	43	54	63	70	76	80	83	83
0.10	3	13	25	35	44	51	57	62	65	67	68
0.15	2	6	11	16	20	23	26	28	29	30	30
0.20	1	4	7	9	11	13	15	16	17	17	17

CONFIANZA=90% y N=5,000											
VALOR INICIAL PARA LA PROPORCIÓN											
B											
	p=.01	p=.05	p=.10	p=.15	p=.20	p=.25	p=.30	p=.35	p=.40	p=.45	p=.50
0.01	255	1023	1638	2042	2321	2519	2661	2760	2826	2864	2876
0.02	67	302	543	736	890	1012	1107	1177	1226	1255	1264
0.03	30	139	257	357	439	507	561	602	631	648	654
0.04	17	80	148	207	257	299	332	358	376	387	390
0.05	11	51	96	135	168	196	218	235	247	255	257
0.06	8	36	67	95	118	138	154	166	175	180	182
0.07	6	27	50	70	87	102	114	123	130	134	135
0.08	5	21	38	54	67	79	88	95	100	103	104
0.09	4	16	30	43	53	62	70	75	79	82	83
0.10	3	13	25	35	43	51	57	61	65	67	67
0.15	2	6	11	16	20	23	26	28	29	30	30
0.20	1	4	7	9	11	13	15	16	17	17	17

CONFIANZA=90% y N=1,000											
VALOR INICIAL PARA LA PROPORCIÓN											
B											
	p=.01	p=.05	p=.10	p=.15	p=.20	p=.25	p=.30	p=.35	p=.40	p=.45	p=.50
0.01	212	563	710	776	813	836	851	861	867	871	872
0.02	63	244	379	464	521	560	588	607	620	627	629
0.03	29	126	214	278	326	361	388	407	420	427	430
0.04	17	75	133	178	214	241	263	279	289	296	298
0.05	11	49	89	122	148	169	186	198	207	212	214
0.06	8	35	64	88	108	124	137	147	153	157	159
0.07	6	26	48	66	82	94	105	112	118	121	122
0.08	5	20	37	52	64	74	82	88	93	95	96
0.09	4	16	30	41	51	60	66	71	75	77	78
0.10	3	13	24	34	42	49	54	59	62	63	64
0.15	2	6	11	16	19	23	25	27	29	29	30
0.20	1	4	7	9	11	13	15	16	16	17	17

CONFIANZA=90% y N=300											
VALOR INICIAL PARA LA PROPORCIÓN											
B											
	p=.01	p=.05	p=.10	p=.15	p=.20	p=.25	p=.30	p=.35	p=.40	p=.45	p=.50
0.01	142	244	268	277	281	284	286	287	287	288	288
0.02	55	156	202	223	236	243	248	252	254	255	255
0.03	28	97	143	169	186	197	204	209	213	215	215
0.04	16	64	102	126	143	155	163	169	173	175	176
0.05	11	45	74	95	111	122	130	136	140	142	143
0.06	8	33	56	73	87	97	104	110	113	116	116
0.07	6	25	43	58	69	78	84	89	93	95	95
0.08	5	19	34	46	56	63	69	74	77	78	79
0.09	4	16	28	38	46	52	58	61	64	65	66
0.15	2	6	11	15	19	22	25	26	28	29	29
0.20	1	4	7	9	11	13	14	15	16	17	17

CONFIANZA=90% y N=400											
VALOR INICIAL PARA LA PROPORCIÓN											
B											
	p=.01	p=.05	p=.10	p=.15	p=.20	p=.25	p=.30	p=.35	p=.40	p=.45	p=.50
0.01	161	306	344	359	367	371	374	376	377	378	378
0.02	58	179	242	274	293	305	313	318	322	324	324
0.03	28	106	162	197	219	235	246	253	258	261	262
0.04	17	68	111	141	162	178	189	197	202	205	206
0.05	11	46	79	103	122	135	146	153	158	161	162
0.06	8	33	58	78	93	105	114	121	125	128	129
0.07	6	25	45	61	73	83	91	96	100	103	103
0.08	5	20	35	48	58	67	73	78	82	84	84
0.09	4	16	29	39	48	55	60	65	67	69	70
0.10	3	13	24	32	40	46	50	54	56	58	58
0.15	2	6	11	15	19	22	24	26	27	28	29
0.20	1	4	7	9	11	13	14	15	16	17	17

CONFIANZA=90% y N=200											
VALOR INICIAL PARA LA PROPORCIÓN											
B											
	p=.01	p=.05	p=.10	p=.15	p=.20	p=.25	p=.30	p=.35	p=.40	p=.45	p=.50
0.01	115	174	185	190	192	193	194	194	195	195	195
0.02	51	124	151	163	169	173	176	178	179	179	179
0.03	27	84	116	132	142	148	153	155	157	158	159
0.04	16	58	87	105	116	123	129	132	135	136	136
0.05	11	42	66	82	94	101	107	111	114	115	116
0.06	8	31	51	66	76	83	89	93	96	97	98
0.07	6	24	40	53	62	69	74	78	80	82	82
0.08	5	19	33	43	51	57	62	66	68	69	70
0.09	4	15	27	36	43	48	53	56	58	59	60
0.10	3	13	22	30	36	41	45	48	50	51	51
0.15	2	6	11	15	18	21	23	25	26	27	27
0.20	1	4	6	9	11	12	14	15	16	16	16

CONFIANZA=90% y N=100											
VALOR INICIAL PARA LA PROPORCIÓN											
B											
	p=.01	p=.05	p=.10	p=.15	p=.20	p=.25	p=.30	p=.35	p=.40	p=.45	p=.50
0.01	74	93	97	98	98	99	99	99	99	99	99
0.02	41	77	87	90	92	93	94	94	95	95	95
0.03	24	60	74	80	83	86	87	88	88	89	89
0.04	15	45	61	69	74	77	79	80	81	81	82
0.05	10	35	50	59	64	68	70	72	73	74	74
0.06	7	27	41	50	55	59	62	64	65	66	66
0.07	6	21	34	42	48	52	54	56	58	58	59
0.08	5	17	28	36	41	45	48	50	51	52	52
0.09	4	14	24	31	36	39	42	44	45	46	46
010	3	12	20	26	31	34	37	39	40	41	41
0.15	2	6	10	14	17	19	21	22	23	24	24
0.20	1	4	6	9	10	12	13	14	15	15	15

CONFIANZA=90% y N=75											
VALOR INICIAL PARA LA PROPORCIÓN											
B											
	p=.01	p=.05	p=.10	p=.15	p=.20	p=.25	p=.30	p=.35	p=.40	p=.45	p=.50
0.01	59	71	73	74	74	74	75	75	75	75	75
0.02	36	61	67	70	71	71	72	72	72	72	72
0.03	22	50	59	63	66	67	68	68	69	69	69
0.04	14	40	51	56	59	61	63	63	64	64	64
0.05	10	31	43	49	53	55	57	58	59	59	59
0.06	7	25	36	43	47	50	52	53	54	54	54
0.07	6	20	31	37	41	44	46	48	49	49	49
0.08	5	17	26	32	36	39	41	43	44	44	4
0.09	4	14	22	28	32	35	37	39	40	40	40
0.10	3	12	19	24	28	31	33	35	36	36	36
0.15	2	6	10	13	16	18	20	21	22	22	22
0.20	1	4	6	8	10	11	13	13	14	14	14

CONFIANZA=90% y N=50											
VALOR INICIAL PARA LA PROPORCIÓN											
B											
	p=.01	p=.05	p=.10	p=.15	p=.20	p=.25	p=.30	p=.35	p=.40	p=.45	p=.50
0.01	43	49	50	50	50	50	50	50	50	50	50
0.02	29	44	47	48	48	49	49	49	49	49	49
0.03	19	38	43	45	46	47	47	47	47	47	47
0.04	13	32	38	41	43	44	44	45	45	45	45
0.05	9	26	34	37	39	41	42	42	43	43	43
0.06	7	22	29	34	36	38	39	39	40	40	40
0.07	6	18	26	30	33	34	36	36	37	37	37
0.08	4	15	22	27	29	31	33	34	34	35	35
0.09	4	13	20	24	27	29	30	31	32	32	32
0.10	3	11	17	21	24	26	27	28	29	29	29
0.15	2	6	10	12	15	16	18	18	19	19	20
0.20	1	4	6	8	10	11	12	12	13	13	13

CONFIANZA=90% y N=40											
VALOR INICIAL PARA LA PROPORCIÓN											
B											
	p=.01	p=.05	p=.10	p=.15	p=.20	p=.25	p=.30	p=.35	p=.40	p=.45	p=.50
0.01	35	39	40	40	40	40	40	40	40	40	40
0.02	26	36	38	39	39	39	39	40	40	40	40
0.03	18	32	35	37	38	38	38	38	38	39	39
0.04	13	27	32	34	35	36	37	37	37	37	37
0.05	9	23	29	32	33	34	35	35	35	35	35
0.06	7	20	26	29	31	32	33	33	33	34	34
0.07	5	17	23	26	28	30	30	31	31	32	32
0.08	4	14	20	24	26	27	28	29	29	30	30
0.09	4	12	18	21	24	25	26	27	27	28	28
0.10	3	10	16	19	22	23	24	25	25	26	26
0.15	2	6	9	12	14	15	16	17	18	18	18
0.20	1	4	6	8	9	10	11	12	12	13	13

CONFIANZA=90% y N=30											
VALOR INICIAL PARA LA PROPORCIÓN											
B											
	p=.01	p=.05	p=.10	p=.15	p=.20	p=.25	p=.30	p=.35	p=.40	p=.45	p=.50
0.01	28	30	30	30	30	30	30	30	30	30	30
0.02	21	28	29	30	30	30	30	30	30	30	30
0.03	16	25	28	28	29	29	29	29	29	29	29
0.04	11	23	26	27	28	28	28	28	28	29	29
0.05	9	20	24	25	26	27	27	27	27	28	28
0.06	7	17	21	24	25	25	26	26	26	26	26
0.07	5	15	19	22	23	24	24	25	25	25	25
0.08	4	13	18	20	21	22	23	24	24	24	24
0.09	4	11	16	18	20	21	22	22	23	23	23
0.10	3	10	14	17	18	20	20	21	21	21	21
0.15	2	5	9	11	12	14	14	15	15	16	16
0.20	1	3	6	7	9	10	10	11	11	11	12

Tabla 2. Tamaño de muestra para intervalo de confianza de la media en una variable medida en escala de razón

CONFIANZA=80%											
	TAMAÑO DE LA POBLACION (N)										
D											
	10000	5000	1000	500	400	300	200	100	75	50	30
0.01	6218	3834	943	486	391	295	198	100	75	50	30
0.02	2913	2256	805	446	365	280	191	98	74	50	30
0.03	1545	1338	647	393	329	258	181	95	73	49	30
0.04	932	853	507	337	288	233	168	92	70	48	30
0.05	617	582	397	285	249	206	154	87	68	47	29
0.06	437	419	314	239	214	182	140	83	65	46	29
0.07	325	315	252	201	183	159	126	78	62	44	28
0.08	251	245	205	170	157	139	113	72	59	42	27
0.09	199	195	169	145	135	122	101	67	55	41	27
0.10	162	160	142	124	117	107	91	63	52	39	26
0.15	73	72	69	64	62	59	54	43	38	30	22
0.20	41	41	40	38	38	37	35	30	27	23	18
0.25	27	27	26	25	25	25	24	21	20	18	15
0.30	19	19	18	18	18	18	17	16	15	14	12
0.35	14	14	14	14	13	13	13	12	12	11	10
0.40	11	11	11	11	11	10	10	10	10	9	8
0.45	9	9	9	8	8	8	8	8	8	7	7
0.50	7	7	7	7	7	7	7	7	7	6	6
0.60	5	5	5	5	5	5	5	5	5	5	4
0.70	4	4	4	4	4	4	4	4	4	4	4
0.80	3	3	3	3	3	3	3	3	3	3	3
0.90	3	3	3	3	3	3	3	2	2	2	2
1.00	2	2	2	2	2	2	2	2	2	2	2

CONFIANZA=90%

TAMAÑO DE LA POBLACION (N)											
D											
	10000	5000	1000	500	400	300	200	100	75	50	30
0.01	7302	4221	965	491	395	297	199	100	75	50	30
0.02	4036	2876	872	466	378	288	195	99	75	50	30
0.03	2312	1878	751	429	354	273	188	97	74	50	30
0.04	1447	1264	629	386	324	255	179	95	72	49	30
0.05	977	890	520	343	293	235	169	92	71	48	30
0.06	700	654	430	301	262	215	158	89	69	47	29
0.07	524	498	356	263	232	195	147	85	67	46	29
0.08	406	390	298	230	206	176	136	81	64	45	29
0.09	324	314	251	201	183	159	126	77	62	44	28
0.10	264	257	213	176	162	143	116	74	59	43	28
0.15	119	118	108	97	93	86	76	55	47	36	25
0.20	68	67	64	60	58	56	51	41	36	29	21
0.25	44	43	42	40	40	38	36	31	28	24	18
0.30	30	30	30	29	28	28	27	24	22	19	16
0.35	23	22	22	22	21	21	20	19	18	16	13
0.40	17	17	17	17	17	17	16	15	14	13	11
0.45	14	14	14	14	13	13	13	12	12	11	10
0.50	11	11	11	11	11	11	11	10	10	9	8
0.60	8	8	8	8	8	8	8	7	7	7	7
0.70	6	6	6	6	6	6	6	6	6	5	5
0.80	5	5	5	5	5	5	5	5	5	4	4
0.90	4	4	4	4	4	4	4	4	4	4	4
1.00	3	3	3	3	3	3	3	3	3	3	3

CONFIANZA=95%											
TAMAÑO DE LA POBLACION (N)											
D											
	10000	5000	1000	500	400	300	200	100	75	50	30
0.01	7935	4425	975	494	396	298	199	100	75	50	30
0.02	4900	3289	906	476	385	291	196	99	75	50	30
0.03	2992	2303	811	448	366	281	192	98	74	50	30
0.04	1937	1623	706	414	343	267	185	97	73	49	30
0.05	1332	1176	606	378	318	251	177	94	72	49	30
0.06	965	880	517	341	291	235	169	92	71	48	30
0.07	728	678	440	306	265	217	160	89	69	48	29
0.08	567	536	376	273	241	201	151	86	67	47	29
0.09	453	434	322	244	217	184	141	83	65	46	29
0.10	370	357	278	218	196	169	132	80	63	45	28
0.15	168	166	146	128	120	109	93	64	53	39	26
0.20	96	95	88	81	78	73	65	49	43	33	23
0.25	62	61	58	55	54	52	48	39	34	28	21
0.30	43	43	41	40	39	38	36	30	28	24	18
0.35	32	32	31	30	30	29	28	24	23	20	1
0.40	24	24	24	23	23	23	22	20	19	17	14
0.45	19	19	19	19	19	18	18	16	16	14	12
0.50	16	16	16	15	15	15	15	14	13	12	11
0.60	11	11	11	11	11	11	11	10	10	9	8
0.70	8	8	8	8	8	8	8	8	8	7	7
0.80	6	6	6	6	6	6	6	6	6	6	6
0.90	5	5	5	5	5	5	5	5	5	5	5
1.00	4	4	4	4	4	4	4	4	4	4	4

CONFIANZA=99%											
TAMAÑO DE LA POBLACION (N)											
D											
	10000	5000	1000	500	400	300	200	100	75	50	30
0.01	8691	4650	986	497	398	299	200	100	75	50	30
0.02	6240	843	944	486	391	295	198	100	75	50	30
0.03	4244	2980	881	469	380	289	195	99	75	50	30
0.04	2932	2267	806	447	365	280	191	98	74	50	30
0.05	2098	1734	727	421	348	270	186	97	73	50	30
0.06	1557	1347	649	394	329	259	181	95	73	49	30
0.07	1193	1066	576	366	309	246	175	94	72	49	30
0.08	940	859	510	338	289	233	168	92	70	48	30
0.09	758	704	451	311	269	220	161	90	69	48	29
0.10	623	586	399	286	250	207	154	87	68	47	29
0.15	287	279	228	186	170	149	120	75	60	43	28
0.20	164	161	143	125	118	107	91	63	52	39	26
0.25	106	104	96	88	84	79	70	52	44	34	24
0.30	74	73	69	65	63	60	54	43	38	30	22
0.35	54	54	52	49	48	46	43	36	32	27	20
0.40	42	42	40	39	38	37	35	30	27	23	18
0.45	33	33	32	31	31	30	29	25	23	20	16
0.50	27	27	26	26	25	25	24	21	20	18	15
0.60	19	19	19	18	18	18	17	16	15	14	12
0.70	14	14	14	14	14	13	13	12	12	11	10
0.80	11	11	11	11	11	11	10	10	10	9	8
0.90	9	9	9	9	9	8	8	8	8	8	7
1.00	7	7	7	7	7	7	7	7	7	6	6

CONFIANZA=99.5%											
TAMAÑO DE LA POBLACION (N)											
D											
	10000	5000	1000	500	400	300	200	100	75	50	30
0.01	8874	4702	988	497	398	299	200	100	75	50	30
0.02	6633	3988	952	488	393	296	198	100	75	50	30
0.03	4669	3183	898	473	383	291	196	99	75	50	30
0.04	3300	2482	832	454	370	283	193	99	74	50	30
0.05	2397	1934	760	432	355	274	189	97	74	50	30
0.06	1796	1523	687	408	339	264	184	96	73	49	30
0.07	1386	1217	617	382	321	253	178	95	72	49	30
0.08	1097	988	552	356	302	242	173	93	71	49	30
0.09	887	815	494	331	284	230	166	91	70	48	30
0.10	731	681	441	306	266	218	160	89	69	48	29
0.15	339	328	260	206	187	162	128	78	62	44	28
0.20	194	190	165	142	132	119	100	67	55	40	27
0.25	125	123	112	101	96	89	78	56	48	36	25
0.30	87	87	81	75	72	68	61	47	41	32	23
0.35	64	64	61	57	56	53	49	40	35	29	21
0.40	50	49	47	45	44	43	40	33	30	25	19
0.45	39	39	38	37	36	35	33	29	26	22	17
0.50	32	32	31	30	30	29	28	24	23	20	16
0.60	22	22	22	21	21	21	20	18	17	16	13
0.70	17	17	16	16	16	16	15	14	14	13	11
0.80	13	13	13	13	12	12	12	11	11	10	9
0.90	10	10	10	10	10	10	10	9	9	9	8
1.00	8	8	8	8	8	8	8	8	8	7	7

ANEXO 2.

PROGRAMAS:

PROGRAMA DE LA ASIGNATURA: **Estadística Inferencial I**
CLABE : AEF1024
CARRERA: Ingeniería Industrial y Gestión Empresarial

Caracterización de la asignatura.
Fundamentación.
La materia de Estadística Inferencial I:
Se plantea como una asignatura básica de la Carrera de Ingeniería en Logística e Industrial y común a la mayor parte de las Ingenierías.
Proporciona los elementos básicos para hacer análisis a partir del estadístico de la muestra y conceptos de la estimación estadística.
Permite establecer inferencias sobre una población, conclusiones a partir de la información que arrojan las pruebas de hipótesis.
A partir de las pruebas de bondad de ajuste, se establece el nivel de aplicabilidad de los conceptos del análisis estadístico.

UNIDAD	TEMAS	SUB-TEMAS
1	Distribuciones Fundamentales para el Muestreo	1.1 Introducción a la Estadística Inferencial 1.2 Muestreo: Introducción al muestreo y tipos de muestreo 1.3 Teorema del límite central 1.4 Distribuciones fundamentales para el muestreo 1.4.1 Distribución muestral de la media 1.4.2 Distribución muestral de ladiferencia de medias 1.4.3 Distribución muestral de la proporción 1.4.4 Distribución muestral de la diferencia de proporciones 1.4.5 Distribución t-student 1.4.6 Distribución muestral de la varianza 1.4.7 Distribución muestral de la relación de varianzas
2	Estimación	2.1 Introducción 2.2 Características de un estimador 2.3 Estimación puntual 2.4 Estimación por intervalos

		2.4.1 Intervalo de confianza para la media 2.4.2 Intervalo de confianza para la diferencia de medias 2.4.3 Intervalos de confianza para la proporción 2.4.4 Intervalos de confianza para la diferencia de proporciones 2.4.5 Intervalos de confianza para la varianza 2.4.6 Intervalos de confianza para la relación de varianzas
		2.5 Determinación del tamaño de muestra 2.5.1 Basado en la media de la Población 2.5.2 Basado en la proporción de la Población 2.5.3 Basado en la diferencia entre las medias de la Población
3	Pruebas de hipótesis	3.1 Introducción 3.2 Confiabilidad y significancia 3.3 Errores tipo I y tipo II 3.4 Potencia de la prueba 3.5 Formulación de Hipótesis estadísticas 3.6 Prueba de hipótesis para la media 3.7 Prueba de hipótesis para la diferencia de medias 3.8 Prueba de hipótesis para la proporción 3.9 Prueba de hipótesis para la diferencia de proporciones 3.10 Prueba de hipótesis para la varianza 3.11 Prueba de hipótesis para la relación de varianzas. 3.12 Uso de software estadístico
4	Pruebas de bondad de ajuste y pruebas no paramétricas	4.1 Bondad de ajuste 4.1.1 Análisis Ji-Cuadrada 4.1.2 Prueba de independencia 4.1.3 Prueba de la bondad del ajuste 4.1.4 Tablas de contingencia 4.1.5 Uso del software estadístico. 4.2 Pruebas no paramétricas 4.2.1 Escala de medición 4.2.2 Métodos estadísticos contra no paramétricos 4.2.3 Prueba de Kolmogorov – Smirnov 4.2.4 Prueba de Anderson – Darling

		4.2.5 Prueba de Ryan – Joiner 4.2.6 Prueba de Shappiro – Wilk. 4.2.7 Aplicaciones del paquete computacional
5	Regresión lineal simple y múltiple	5.1 Regresión Lineal simple 5.1.1 Prueba de hipótesis en la regresión lineal simple 5.1.2 Calidad del ajuste en regresión lineal simple 5.1.3 Estimación y predicción por intervalo en regresión lineal simple 5.1.4 Uso de software estadístico 5.2 Regresión lineal múltiple 5.2.2 Pruebas de hipótesis en regresión lineal múltiple 5.2.3 Intervalos de confianza y predicción en regresión múltiple 5.2.4 Uso de un software estadístico 5.3 Regresión no lineal

PROGRAMA DE LA ASIGNATURA: **Estadística Inferencial II.**
CLABE: AEF1025 Y GEG0908
CARRERAS: Ingeniería Industrial y Gestión Empresarial

Esta asignatura aporta al perfil del Ing. Industrial y en Gestión Empresarial la capacidad de realizar análisis de regresión simple y múltiple, análisis de serie de tiempo y diseño de experimentos en los diferentes ámbitos del quehacer empresarial. Se ha hecho una mención especial en el desarrollo de experimentos aplicados a la industria que permitirán mejorar la calidad de los productos y procesos. Muy importante será el poder identificar los diferentes factores que podrían resultar relevantes en el desarrollo de nuevos productos y de nuevas tecnologías; así como la importancia que tiene el análisis de regresión en identificar las variables explicitarías para estimar las variables dependientes.

UNIDAD	TEMA	SUBTEMAS
1	Regresión lineal múltiple	1.1 Regresión lineal múltiple. 1.1.1 Pruebas de hipótesis en regresión lineal múltiple. 1.1.2 Intervalos de confianza y predicción en regresión múltiple. 1.1.3 Uso de un software estadístico 1.2 Regresión no lineal.
2	Series de tiempo	2.1 Modelo clásico de series de tiempo. 2.2 Análisis de fluctuaciones. 2.3 Análisis de tendencia. 2.4 Análisis de variaciones cíclicas. 2.5 Medición de variaciones estacionales e irregulares. 2.6 Aplicación de ajustes estacionales. 2.7 Pronósticos basados en factores de 2.8 Tendencia y estacionales.
3	Diseño de experimentos de un factor	3.1 Familia de diseños para comparar tratamientos. 3.2 El modelo de efectos fijos. 3.3 Diseño completamente aleatorio y ANOVA. 3.4 Comparaciones o pruebas de rangos múltiples. 3.5 Verificación de los supuestos del modelo.
4	Diseño de bloques.	4.1 Diseños en bloques completos al azar. 4.2 Diseño en cuadrado latino. 4.3 Diseño en cuadrado grecolatino. 4.4 Uso de un software estadístico.
5	Diseños factoriales.	5.1 Diseños factoriales con dos factores. 5.2 Diseños factoriales con tres factores. 5.3 Diseño factorial general. 5.4 Modelos de efectos aleatorios. 5.5 Uso de un software estadístico.