

Supletorio de Examen Final de Inferencia Estadística – período 081
 Cali, Mayo 24 de 2008.

1. Una empresa que se dedica a hacer encuestas se queja de que sus agentes realizan en promedio 53 encuestas por semana. Se ha introducido una forma más moderna de realizar las encuestas y la empresa quiere evaluar su efectividad. Los números de encuestas, realizados en una semana por una muestra aleatoria de 15 agentes que utilizaron la forma moderna son:

53	57	50	55	58	54	60	52	59	62	60	60	51	59	56
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

- Con un nivel de significancia del 5%, ¿puede concluirse que la cantidad media de entrevistas realizadas por los agentes es superior a 53 por semana?
- Con base en la información anterior, construir un intervalo de confianza del 95% para estimar la verdadera media poblacional.

10%

2. Se tomaron muestras independientes de 1500 hombres y de 1700 mujeres para comparar la cantidad de labores domésticas hechas por hombres y por mujeres en matrimonios con doble aportación económica. El estudio indicó que 990 de los hombres sentían que la división de tareas domésticas era justa, y que 1037 de las mujeres sentían que la división de tareas era justa. (American Journal of Sociology, 1994).

- ¿Es mayor la proporción de hombres que sentían que la división del trabajo doméstico era justa, que la proporción correspondiente de mujeres? Respalde su conclusión con una prueba estadística al nivel de significancia del 5%.
- Halle el valor-p de la prueba estadística.

10%

3. El gerente de una compañía de programas para computadora desea analizar, de acuerdo al tipo de industria, la cantidad de horas que los ejecutivos de alto nivel pasan frente a sus computadoras. Se obtiene una muestra de 5 ejecutivos de cada una de 3 industrias.

- ¿Al nivel de significancia del 5%, puede concluirse que existe una diferencia por industria en el número promedio de horas por semana que los ejecutivos dedican a trabajar en sus computadoras?
- De ser así, identifique cuál o cuáles grupos son diferentes.

Los datos muestrales obtenidos son los siguientes:

Banca	12	10	10	12	10
Comercio	8	8	6	8	10
Seguros	10	8	6	8	10

20%

4. Una gran corporación está interesada en determinar si existe relación entre el tiempo de recorrido hasta llegar al trabajo y el nivel de estrés de los empleados en sus labores diarias de ensamblaje. Para ello se tomó una muestra con los siguientes resultados:

Tiempo de Recorrido	Nivel de Estrés		
	Alto	Moderado	Bajo
Menos de 15 min	9	5	18
De 15 a 45 min	17	8	28
Más de 45 min	18	6	7

- Plantee las hipótesis correspondientes.
- Pruebe la hipótesis y entregue su conclusión al respecto con un nivel de significancia del 5% para la prueba estadística correspondiente.

20%

5. Al gerente de una empresa inmobiliaria de una ciudad, le interesa saber si existe algún tipo de relación lineal entre el costo de renta mensual de los apartamentos que alquila y el tamaño del inmueble definido por su extensión. Tomó una muestra de 25 apartamentos con su renta y su extensión correspondiente. (renta en miles de pesos y tamaño en pies cuadrados)

Renta	950	1600	1200	1500	950	1700	1650	935	875	1150	1400	1650	2300	1800	1400	1450	1100	1700	1200	1150	1600	1650	1200	800	1750
Tamaño	850	1450	1085	1232	718	1485	1136	726	700	956	1100	1285	1985	1369	1175	1225	1245	1259	1150	896	1361	1040	755	1000	1200

Con los datos procesados se obtuvo el siguiente resumen con un nivel de significancia del 5%:

Estadísticas de la regresión

Coefficiente de correlación	0.850060796
Coefficiente de determinación R ²	0.722603356
Error estándar	194.5953946
Observaciones	25

TABLA DE ANÁLISIS DE VARIANZA

Fuente de Variación	Grados de Libertad	Suma de Cuadrados	Cuadrados Medios	F _{calc}
Regresión				
Residuos			37867,3676	
Total		3139726		

TABLA DE ESTIMACIÓN DE COEFICIENTES

	Coefficientes	Error típico	Estadístico t	valor-p
Intercepción	177,12082	161,004277	1,1001001	0,28266985
X		0,13760841	7,74039821	7,5183E-08

Regresión Auxiliar

Estadísticas de la regresión

Coefficiente de correlación múltiple	0,0597562
Coefficiente de determinación R ²	0,003570803
R ² ajustado	-0,039752205
Error típico	55544,74384
Observaciones	25

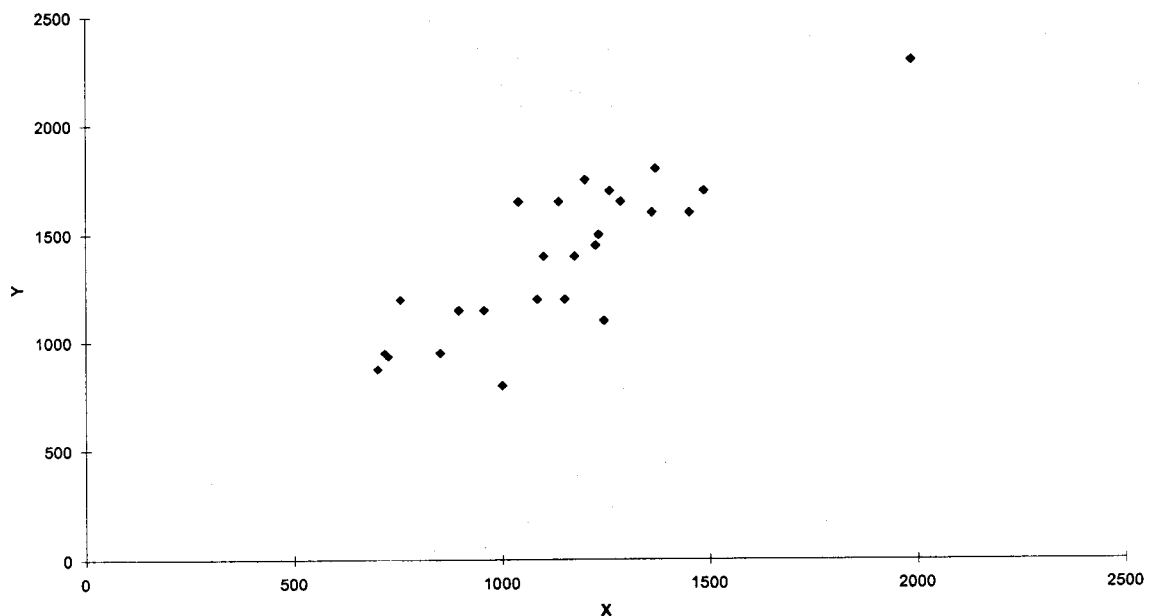
Estadístico de Durbin-Watson: 2,4383

Estadístico de Kolmogorov-Smirnov: 0,1413

Intervalo de Confianza del 95% para Y: (1153,01 – 1331,52) para un X=1000

Intervalo de Predicción del 95% para Y: (829,94 – 1654,59) para un X=1000

Scatter Diagram



- Complete la información de las tablas de análisis de varianza y de estimación de coeficientes.
- Escriba el modelo de regresión lineal simple e interprete el significado de la pendiente del modelo. Interprete el resultado del coeficiente de determinación
- Efectúe la prueba inherente a la pendiente del modelo con una significancia del 5%
- Para un valor de X de 1000, ¿cuál sería el intervalo para el valor de un solo apartamento Y que usted utilizaría? Escriba el intervalo.

e. Plantee y resuelva todas las hipótesis correspondientes a los supuestos de modelo de regresión encontrado, con base en los resultados relacionados arriba.

30%

6. Un grupo de investigación de mercados se interesa en probar la afirmación de un fabricante de automóviles, de que un nuevo modelo económico recorre cuando menos 25 Km por galón. Encuentre los valores del promedio para los cuales, la afirmación del fabricante se debe rechazar, sabiendo que se tomó una muestra de 30 automóviles y un $\alpha = 0.02$, Suponga además que $\sigma = 3$ Km por galón. ¿Cuál es el β si el rendimiento real es de 23 km por galón?

10%

Algunas fórmulas de Interés

Tabla de análisis de varianza

Fuente de Variación	Grados de libertad	Sumas de Cuadrados	Cuadrados Medios	F Calculada
Entre grupos	c-1	SSA	MSA	F
Dentro de grupos	n-c	SSW	MSW	
TOTAL	n-1	SST		

Diferencia de proporciones

$$Z \cong \frac{(p_{s_1} - p_{s_2}) - (p_1 - p_2)}{\sqrt{p(1-p) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

Chi cuadrado

$$X^2 = \sum_{\text{todas las celdas}} \frac{(f_o - f_e)^2}{f_e}$$

Rango crítico de Tukey-Kramer

$$RC = Q_U \sqrt{\frac{MSW}{2} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$$

Intervalo de Confianza para el promedio con σ desconocido

$$\bar{X} \pm t_{n-1} \frac{S_x}{\sqrt{n}}$$