UNIVERSIDAD ICESI — Facultad de Ingeniería, Departamento de Matemáticas y Estadística

EXAMEN FINAL DE INFERENCIA ESTADÍSTICA – PERÍODO 2011 - 1

Cali, mayo 17 de 2011.

1. En una empresa, la gerencia comercial prueba un proceso nuevo sobre venta de artículos y se requiere evaluarlo, para saber si cumple con la meta de vender como mínimo 155 artículos en promedio por semana. Se toma una muestra aleatoria de artículos vendidos por semana durante 10 semanas, obteniéndose los siguientes valores (considere que el número de artículos vendidos por semana, se distribuye aproximadamente normal):

145	150	153	148	141	152	146	154	139	148

   a. Construya un intervalo con un nivel de confianza del 98% y concluya si se cumple con el requerimiento de la gerencia.
   b. La gerencia considera que las ventas promedio de artículos por semana son consistentes si la desviación es inferior a 6 artículos, con base en un intervalo del 95% concluya si la gerencia puede considerar las ventas consistentes.
   c. Si el proceso actual ha presentado un promedio semanal de ventas de artículos de 140 con una desviación de 10 artículos, al obtenerse una muestra de 36 semanas ¿Cuál es la probabilidad de que la media de dicha muestra difiera de la media de la población en más de 5 artículos?
   d. El gerente comercial solicita obtener una muestra actualizada del proceso actual de ventas de artículos y sugiere una confiabilidad del 95% y un error permisible como máximo de 3 artículos y como variabilidad la desviación de 10 artículos por semana. ¿Cuántas semanas se deben evaluar, sabiendo que sólo se tiene información de 52 semanas?

(20%)

2. Para una autopista entre los pueblos A y B, una muestra de 300 residentes adultos del pueblo A reveló que 63 estaban a favor de incrementar el límite de velocidad para la autopista de 80 a 120 km/h, en tanto que una muestra de 180 residentes del pueblo B arrojó que 65 estaban a favor del incremento. ¿Indican los datos que en el pueblo B prefieren más una velocidad mayor en la autopista que en el pueblo A? plantee y resuelva la hipótesis correspondiente, utilice un $\alpha=0{,}05$.

(15%)

3. Una reconocida tienda comercial de la ciudad de Cali quiere determinar si la ubicación de la tienda tiene incidencia en el volumen de ventas de la compañía. Se tomaron los siguientes datos DE VENTAS (en millones de pesos) durante un período de cuatro trimestres (use α = 0,05):

	Ubicación		
Trimestre	Sur	Centro	Norte
1	580	490	480
2	560	480	340
3	510	450	300
4	450	430	310

   a. Indique el nombre del diseño, defina los factores y la variable de respuesta.
   b. Plantee y valide las hipótesis correspondientes al diseño. Use la notación de hipótesis estadística.

ANÁLISIS DE VARIANZA

Fuente de variación	*Sumas de cuadrados*	*Grados de libertad*	*Cuadrados medios*	*F calc*	*valor-p*	*F crítica*
Filas	24833.3	3	8277.8	5.9	0.032	4.757
Columnas	57316.7	2	28658.3	20.4	0.002	5.143
Error	8416.7	6	1402.8			
Total	90566.7	11				

   c. Use un procedimiento estadístico para determinar "la mejor" ubicación de la tienda en la ciudad de Cali.

(20%)

4. En el artículo "Marijuana use in College" (Youth and Society: 323-334) Cada uno de los 445 estudiantes universitarios fue clasificado según su frecuencia de consumo de marihuana y si los padres consumían alcohol y drogas psicoactivas.

		Nivel estándar de consumo De marihuana		
		Nunca	Ocasional	Regular
Padres Consumen Alcohol y drogas	Ninguno	141	54	40
	Uno	68	44	51
	Ambos	17	11	19

¿La información sugiere que la adicción de padres está relacionada con la adicción de sus hijos estudiantes en la población de la que se extrajo la muestra? Realice un procedimiento de la inferencia estadística para justificar su respuesta. Utilice un $\alpha=0,05$

(15%)

5. Una empresa aseguradora de productos agrícolas investiga cómo la cantidad de lluvia caída afecta las ganancias de los agricultores que siembran un determinado cultivo. Con el propósito de construir un modelo de regresión lineal simple que permita relacionar las dos variables, toma información sobre la cantidad de lluvia (en mm) caída en diferentes épocas y las ganancias obtenidos en millones de pesos por hectárea por los agricultores en dicha temporada, los resultados obtenidos:

Lluvia	260	420	125	500	110	280	130	360	220	130	140	150
Ganancias	2.1	1.8	3.1	1.2	4.5	2.1	3.2	1.9	2.5	2.9	2.8	2.8

a. Estime la ecuación de regresión e interprete los coeficientes del modelo
b. Compruebe e interprete la hipótesis de linealidad del modelo
c. Calcule el coeficiente de determinación del modelo e interprételo
d. Compruebe los supuestos del modelo de regresión (plantee las hipótesis respectivas para cada supuesto)
e. Si el IDEAM pronostica que en la próxima temporada de lluvias caerán 375 mm, estime las ganancias que obtendrán los agricultores. Construya un intervalo de predicción individual del 95% para las ganancias estimadas.

Información adicional para comprobar supuestos e hipótesis del modelo

Nota: Para todas las pruebas utilice un $\alpha = 0.05$.

ANÁLISIS DE VARIANZA DE LA REGRESION

Fuente de variación	*Grados de libertad*	*Suma de cuadrados*	*Cuadrados medios*	*Razón F*	*Valor p*
Regresión			5.97811		0.0003
Residuos					
Total		7.9825			

D+	0.2200
D-	-0.1493
R^2 de regresión auxiliar	0.098
Durbin-Watson	1.9751
Error estándar para β_0	0.2747
Error estándar para β_1	0.0010

(30%)

FÓRMULAS PARA INFERENCIA

Distribuciones muestrales:

Del promedio:

$$Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

De la proporción:

$$Z = \frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$$

De la varianza:

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

Intervalos de Confianza:

Para el promedio:

$\overline{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ Muestras grandes.

$\overline{X} \pm t_{n-1;\alpha/2} \frac{S}{\sqrt{n}}$ Muestras pequeñas.

Para la proporción:

$$p \pm Z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}$$

Para la varianza:

$$\left(\frac{(n-1)S^2}{\chi^2_{n-1;\alpha/2}} ; \frac{(n-1)S^2}{\chi^2_{n-1;1-\alpha/2}} \right)$$

Tamaños de muestra:

Para el promedio:

$$n = \frac{Z^2_{\alpha/2}\sigma^2}{e^2}$$

Para la proporción:

$$n = \frac{Z^2_{\alpha/2}\pi(1-\pi)}{e^2}$$

Factor de corrección por población finita:

$$fcp = \sqrt[2]{\frac{N-n}{N-1}}$$

Diferencia de promedios:

Con varianzas desconocidas iguales:

$$t=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}};\quad S_p^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$

Con varianzas desconocidas diferentes:

$$t=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}};\quad v=\frac{\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)^2}{\frac{(S_1^2/n_1)^2}{n_1-1}+\frac{(S_2^2/n_2)^2}{n_2-1}}$$ grados de libertad del valor crítico

Diferencia de proporciones:

$$Z=\frac{(p_1-p_2)-(\pi_1-\pi_2)}{\sqrt{\frac{p_c(1-p_c)}{n_1}+\frac{p_c(1-p_c)}{n_2}}};\quad p_c=\frac{X_1+X_2}{n_1+n_2}$$

Intervalos de confianza post-anova

$$(\bar{X}_1-\bar{X}_2)\pm t_{n-k;\alpha/2}\sqrt{CME\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

Intervalo de predicción para $\hat{Y}$

$$\hat{Y}\pm t_{n-2;\alpha/2}S_{YX}\sqrt{1+h_i}$$

$$h_i=\frac{1}{n}+\frac{(X_i-\bar{X})^2}{\sum(X_i-\bar{X})^2}$$

$$S_{YX}=\sqrt{\frac{SCE}{n-2}}$$

Prueba Chi cuadrado

$$\chi^2_{calc}=\sum\frac{(f_o-f_e)^2}{f_e}$$

UNIVERSIDAD ICESI
Facultad de Ingeniería
Departamento de Matemáticas y Estadística

Diferencia de promedios:

Con varianzas desconocidas iguales:

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}; \quad S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

Con varianzas desconocidas diferentes:

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}; \quad v = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{(S_1^2/n_1)^2}{n_1 - 1} + \frac{(S_2^2/n_2)^2}{n_2 - 1}}$$ grados de libertad del valor crítico

Diferencia de proporciones:

$$Z = \frac{(p_1 - p_2) - (\pi_1 - \pi_2)}{\sqrt{\frac{p_c(1 - p_c)}{n_1} + \frac{p_c(1 - p_c)}{n_2}}}; \quad p_c = \frac{X_1 + X_2}{n_1 + n_2}$$

Intervalos de confianza post-anova

$$(\bar{X}_1 - \bar{X}_2) \pm t_{n-k;\alpha/2}\sqrt{CME\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Intervalo de predicción para $\hat{Y}$

$$\hat{Y} \pm t_{n-2;\alpha/2} S_{YX}\sqrt{1 + h_i}$$

$$h_i = \frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum (X_i - \bar{X})^2}$$

$$S_{YX} = \sqrt{\frac{SCE}{n-2}}$$

Prueba Chi cuadrado

$$\chi^2_{calc} = \sum \frac{(f_o - f_e)^2}{f_e}$$