
Primer Semestre 2014

Curso : Probabilidad y Estadística
Sigla : EYP1113
Examen
Profesor : Ricardo Aravena (Sec. 1 y 3) y Ana María Araneda (Sec. 2 y 4)
Ayudantes : Carlos Cayuman, Fabián Fuentealba, Alonso Molina, Genaro Olave.

- Se permite el uso de calculadora científica básica.
- No se permite usar apuntes, correctores y cualquier aparato de transmisión electrónica (por ejemplo celulares y aparatos con bluetooth y wifi).
- Alumnos que escriban sus soluciones con lápiz mina, o cualquier tipo de lápiz borrable, renuncian a su derecho a re-corrección.
- El alumno que sea sorprendido copiando o en otras actividades reñidas con las normas de comportamiento académico, será calificado con nota 1.0 (uno.cero) en la interrogación y su caso será informado a la Dirección de Docencia de la Escuela de Ingeniería.
- En su lugar de trabajo Ud. debe tener solo lápices, sus cuadernillos y calculadora.
- Recuerde poner su N° de lista en ambos cuadernillos.

Problema 1

El molibdeno es un metal de transición. Este metal puro es de color blanco plateado y muy duro; además, tiene uno de los puntos de fusión más altos entre todos los elementos. En pequeñas cantidades, el molibdeno se emplea en distintas aleaciones de acero para endurecerlo o hacerlo más resistente a la corrosión. Suponga que se obtiene una aleación de molibdeno y se desea determinar si ésta satisface ciertos requerimientos relacionados a la resistencia a tracción (en estas pruebas se mide la deformación o alargamiento de la probeta entre dos puntos fijos para diferentes cargas - el método anterior se conoce como la curva tensión-deformación). En lo que sigue, asuma que la resistencia a tracción sigue una distribución Normal.

- a) (1,5 puntos) Con un 95 % de confianza, determine el mínimo tamaño de muestra necesario para estimar la media de la resistencia a tracción con un error (o precisión) no mayor 2 GPa, cuando $\sigma = 6$ GPa.

Solución: La precisión corresponde a la mitad del ancho del intervalo de confianza de dos colas para la media. Luego, se quiere que

$$A = z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \leq 2. \quad [0,6]$$

En nuestro caso, $z_{0,975} = 1,96$ [0,3] y $\sigma = 6$. Despejando n , se obtiene $n > 34,57$ [0,3]. Es decir, el tamaño de muestra mínimo para tener la precisión requerida es de 35 observaciones. [0,3]

Suponga que, por costo disponible, se realizan 24 mediciones, obteniendo una resistencia a tracción promedio de 18 GPa, con una desviación estándar de 5 GPa.

- b) (1,5 puntos) ¿Es válido el supuesto realizado respecto a la varianza en a)? Use $\alpha = 10$ %.

Solución: Se requiere testear las hipótesis: $H_0 : \sigma = 6$, vs. $H_1 : \sigma \neq 6$ [0,3]. El estadístico del test es

$$X_0^2 = (n-1) \frac{S^2}{\sigma_0^2} = (24-1) \frac{5^2}{6^2} = 15,97. \quad [0,5]$$

Como $\alpha = 10$ %, se rechaza H_0 para valores del estadístico menores al cuantil 5 % o mayores al cuantil 95 % de la distribución Chi-Cuadrado con $n-1 = 23$ grados de libertad. En este caso $\chi_{23,0.05}^2 = 13,09$ [0,2] y $\chi_{23,0.95}^2 = 35,16$ [0,2]. Luego, con significancia 10 %, no podemos rechazar la hipótesis realizada en a). [0,3]

- c) (2 puntos) Usando el resultado en b), ¿existe evidencia suficiente, que permita afirmar que la media de la resistencia a tracción es inferior a 20 GPa? Use un nivel de significancia del 5 %.

Solución: Considerando el resultado en b), utilizaremos el valor $\sigma = 6$ como el verdadero valor de σ . Se requiere testear las hipótesis $H_0 : \mu = 20$, vs. $H_1 : \mu < 20$ (también se puede plantear $H_0 : \mu \geq 20$). [0,5] El estadístico del test es:

$$z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{18 - 20}{6/\sqrt{24}} = -1,63. \quad [0,5]$$

Alternativa 1: Rechazamos H_0 si $z_0 < z_{0,05} = -1,64$ [0,5] **por el cuantil**. Luego, con significancia del 5 %, no existe evidencia para afirmar que la media es menor a 20 GPa. [0,5]

Alternativa 2: Rechazamos H_0 si el valor-p es menor que $\alpha = 0,05$. En este caso, el valor-p corresponde a:

$$valor - p = P(Z < -1,63) = \Phi(-1,63) = 1 - \Phi(1,63) = 1 - 0,9484 = 0,0516. \quad [0,5].$$

Luego, con significancia del 5 %, no existe evidencia para afirmar que la media es menor a 20 GPa. [0,5]

- d) (1 punto) Suponga que en realidad la media de la resistencia a tracción es de 19 GPa. ¿Cuál es la probabilidad de cometer un error tipo II con la regla utilizada en c)?

Solución: La probabilidad de cometer error Tipo II es:

$$P(\text{Error Tipo II}) = P(\text{no rechazar } H_0 | H_1 \text{ es cierto}). \quad [0,2]$$

En nuestro caso, el test en c) rechaza H_0 si

$$\frac{\bar{X} - 20}{6/\sqrt{24}} < z_{0,05} = -1,64,$$

y $\mu = 19$, luego,

$$\begin{aligned} P(\text{Error Tipo II}) &= P(\text{no rechazar } H_0 | \mu = 19) \\ &= P\left(\frac{\bar{X} - 20}{6/\sqrt{24}} \geq -1,64 \mid \mu = 19\right) \quad [0,3] \\ &= P\left(\frac{\bar{X} - 19}{6/\sqrt{24}} \geq -1,64 + \frac{20 - 19}{6/\sqrt{24}} \mid \mu = 19\right) \quad [0,4] \\ &= 1 - \Phi(-0,83) = \Phi(0,83) = 0,7967. \quad [0,1] \end{aligned}$$

[1,0] punto base

Problema 2

Para analizar el impacto del Mundial de Fútbol Brasil 2014 en la UC, se lleva a cabo un estudio mediante encuestas. Interesa analizar dos de las preguntas realizadas:

- i. ¿Está de acuerdo con la inversión realizada por la UC para contar con pantallas gigantes para ver los partidos del Mundial?
SI _____ NO _____.
- ii. ¿Ha concurrido usted a las carpas con pantalla gigante para ver, al menos, un partido?
SI _____ NO _____.

De los 400 estudiantes entrevistados, sólo 280 responden positivamente a la primera pregunta. Respecto a la segunda pregunta, de los 280 estudiantes que respondieron positivamente a la primera pregunta, 250 han concurrido al menos una vez a mirar un partido. En cambio, entre los que respondieron negativamente, sólo 22 han concurrido a ver al menos un partido. En base a estos resultados, responda:

- a) ¿Existe evidencia suficiente para afirmar que más de $2/3$ de los alumnos de la UC está de acuerdo con la inversión realizada para contar con pantallas gigantes para el Mundial? Sea explícito: hipótesis, regla de decisión, valor-p. Entregue su conclusión para un nivel de significancia del 5 %.

Solución: Sea p la proporción de alumnos en la UC que está de acuerdo con la inversión realizada para contar con pantallas gigantes. Se pide testear $H_0 : p = 2/3$, vs. $H_1 : p > 2/3$ (también se puede plantear $H_0 : p \leq 2/3$) [0,5]. El estadístico del test corresponde a:

$$\begin{aligned} z_0 &= \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \\ &= \frac{280/400 - 2/3}{\sqrt{\frac{2/3 \times (1-2/3)}{400}}} = 1,41. \end{aligned} \quad [0,5]$$

El valor-p es:

$$\begin{aligned} \text{valor} - p &= P(Z > 1,41) \\ &= 1 - \Phi(1,41) \\ &= 1 - 0,9207 = 0,0793. \end{aligned} \quad [0,5]$$

Para decidir:

Alternativa 1: Dado que $\text{valor} - p \geq 0,05 = \alpha$, con significancia 5 %, no existe evidencia para afirmar que la proporción de alumnos de la UC que está de acuerdo con la inversión es mayor que $2/3$. [0,5]

Alternativa 2: Se rechaza H_0 si $z_0 > z_{0,95} = 1,64$, lo cual en nuestro caso no se cumple. Luego, con significancia 5 %, no existe evidencia para afirmar que la proporción de alumnos de la UC que está de acuerdo con la inversión es mayor que $2/3$. [0,5]

- b) Entre los alumnos de la UC que no están de acuerdo con la inversión realizada, ¿es posible afirmar que menos de un 25 % concurre? (es decir, no está de acuerdo, pero disfruta). Utilice $\alpha = 0,05$.

Solución: Sea p la proporción de alumnos que concurre a ver al menos un partido, entre los alumnos que no están de acuerdo con la inversión realizada. Se pide testear $H_0 : p = 0,25$, vs. $H_1 : p < 0,25$ (también se puede plantear $H_0 : p \geq 0,25$) [0,5]. El estadístico del test corresponde a:

$$\begin{aligned} z_0 &= \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \\ &= \frac{22/120 - 0,25}{\sqrt{\frac{0,25 \times (1-0,25)}{120}}} = -1,69. \end{aligned} \quad [0,5]$$

Para decidir:

Alternativa 1: El valor-p es:

$$\begin{aligned} \text{valor} - p &= \Phi(-1,69) \\ &= 1 - \Phi(1,69) \\ &= 1 - 0,9545 = 0,0455. \quad [0,5] \end{aligned}$$

Dado que $\text{valor} - p < 0,05 = \alpha$, con significancia 5 %, podemos afirmar que la proporción de alumnos que concurre a ver al menos un partido, entre los alumnos que no están de acuerdo con la inversión realizada, es menor que un 25 %. [0,5]

Alternativa 2: Se rechaza H_0 si $z_0 < z_{0,05} = -1,64$ [0,5] **por cuantil**, lo cual en nuestro caso se cumple. Luego, con significancia 5 %, podemos afirmar que la proporción de alumnos que concurre a ver al menos un partido, entre los alumnos que no están de acuerdo con la inversión realizada, es menor que un 25 %. [0,5]

- c) Considere el test utilizado en b). Obtenga la potencia del test, si la verdadera proporción de alumnos que concurre a ver al menos un partido, dentro de los estudiantes que no están de acuerdo con la inversión realizada, es 20 %.

Solución: Si $p = 0,2$, la potencia del test corresponde a:

$$\begin{aligned} \mathcal{P}(0,2) &= P(\text{rechazar } H_0 \mid p = 0,2) \quad [0,3] \\ &= P\left(\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} < z_{0,05} \mid p = 0,2\right) \quad [0,4] \\ &= P\left(\hat{p} < -1,64\sqrt{\frac{0,25(1-0,25)}{120}} + 0,25 \mid p = 0,2\right) \\ &= P(\hat{p} < 0,185 \mid p = 0,2) \\ &= P\left(\frac{\hat{p} - 0,2}{\sqrt{\frac{0,2(1-0,2)}{120}}} < \frac{0,185 - 0,2}{\sqrt{\frac{0,2(1-0,2)}{120}}} \mid p = 0,2\right) \quad [1,0] \\ &= \Phi(-0,41) = 1 - \Phi(0,41) \\ &= 1 - 0,6591 = 0,3409. \quad [0,3] \end{aligned}$$

[1,0] punto base

Problema 3

Dado el riesgo asociado a las acciones transadas en la bolsa, se ha estudiando que las ganancias obtenidas por su compra y venta tienen valor esperado nulo. Con el objeto de conocer en mayor detalle la distribución de las ganancias asociadas a una acción determinada, se ha planteado un modelo con función de densidad

$$f_x(x) = \frac{1}{2\sigma} \exp \left\{ -\frac{|x|}{\sigma} \right\}, \quad x \in R, \sigma > 0.$$

Para estimar el parámetro σ asociado a esta acción, se dispone de una muestra aleatoria de n observaciones independientes de las ganancias generadas por su compra y venta.

- a) Encuentre el estimador de máxima verosimilitud de σ . No es necesario que muestre que el punto crítico es máximo.

Solución: La función de verosimilitud corresponde a:

$$\begin{aligned} L(\sigma, \mathbf{x}) &= \prod_{i=1}^n f_x(x_i) \\ &= \prod_{i=1}^n \frac{1}{2\sigma} \exp \left\{ -\frac{|x_i|}{\sigma} \right\} \\ &= \left(\frac{1}{2\sigma} \right)^n \exp \left\{ -\frac{\sum_{i=1}^n |x_i|}{\sigma} \right\}. \quad [0,5] \end{aligned}$$

Tomando logaritmo natural:

$$\log L(\sigma, \mathbf{x}) = -n \log(2\sigma) - \frac{\sum_{i=1}^n |x_i|}{\sigma}. \quad [0,5]$$

Para buscar el máximo, buscamos puntos críticos:

$$\frac{d \log L(\sigma, \mathbf{x})}{d\sigma} = -\frac{n}{\sigma} + \frac{\sum_{i=1}^n |x_i|}{\sigma^2}. \quad [0,5]$$

Assumiendo $\sigma \neq 0$ y despejando, se obtiene que el estimador de máxima verosimilitud de σ corresponde a:

$$\hat{\sigma} = \frac{\sum_{i=1}^n |X_i|}{n}. \quad [0,5]$$

- b) Encuentre la esperanza y la varianza del estimador de máxima verosimilitud encontrado en a). ¿Es éste insesgado?

Solución: Para la esperanza necesitamos primero:

$$\begin{aligned} E(|X_i|) &= \int_{-\infty}^{\infty} |x| \frac{1}{2\sigma} \exp \left\{ -\frac{|x|}{\sigma} \right\} dx \quad [0,1] \\ &= \int_{-\infty}^0 (-x) \frac{1}{2\sigma} \exp \left\{ \frac{x}{\sigma} \right\} dx + \int_0^{\infty} x \frac{1}{2\sigma} \exp \left\{ -\frac{x}{\sigma} \right\} dx \\ &= \underbrace{\frac{1}{2} \int_0^{\infty} x \frac{1}{\sigma} \exp \left\{ -\frac{x}{\sigma} \right\} dx}_{E(\text{Exp}(1/\sigma))=\sigma} + \underbrace{\frac{1}{2} \int_0^{\infty} x \frac{1}{\sigma} \exp \left\{ -\frac{x}{\sigma} \right\} dx}_{E(\text{Exp}(1/\sigma))=\sigma} \\ &= \sigma. \quad [0,5] \end{aligned}$$

Luego,

$$E(\hat{\sigma}) = \frac{\sum_{i=1}^n E(|X_i|)}{n} = \sigma. \quad [0,2]$$

Luego, $\hat{\sigma}$ es insesgado para σ [0.2]. Para su varianza necesitamos:

$$\begin{aligned}
 E(|X_i|^2) &= \int_{-\infty}^{\infty} |x|^2 \frac{1}{2\sigma} \exp\left\{-\frac{|x|}{\sigma}\right\} dx & [0,1] \\
 &= \int_{-\infty}^0 (-x)^2 \frac{1}{2\sigma} \exp\left\{\frac{x}{\sigma}\right\} dx + \int_0^{\infty} x^2 \frac{1}{2\sigma} \exp\left\{-\frac{x}{\sigma}\right\} dx \\
 &= \frac{1}{2\sigma} \int_0^{\infty} \underbrace{x^2 \exp\left\{-\frac{x}{\sigma}\right\}}_{\text{Kernel Gama}(3,1/\sigma)} dx + \frac{1}{2\sigma} \int_0^{\infty} \underbrace{x^2 \exp\left\{-\frac{x}{\sigma}\right\}}_{\text{Kernel Gama}(3,1/\sigma)} dx \\
 &= \frac{1}{2\sigma} \frac{\Gamma(3)}{(1/\sigma)^3} + \frac{1}{2\sigma} \frac{\Gamma(3)}{(1/\sigma)^3} \\
 &= 2\sigma^2. & [0,5]
 \end{aligned}$$

Luego, la varianza de $\hat{\sigma}$ corresponde a:

$$\begin{aligned}
 \text{Var}(\hat{\sigma}) &= \frac{\text{Var}(|X_i|)}{n} & [0,2] \\
 &= \frac{E(|X_i|^2) - E^2(|X_i|)}{n} \\
 &= \frac{2\sigma^2 - \sigma^2}{n} = \frac{\sigma^2}{n}. & [0,2]
 \end{aligned}$$

- c) Encuentre la distribución asintótica del estimador de máxima verosimilitud encontrado en a). ¿Alcanza su varianza la cota de Cramer-Rao para cualquier valor de n ?

Solución: Para n grande, $\hat{\sigma}$ distribuye aproximadamente Normal [0.2], de media σ [0.2], y su varianza se acerca a la cota de Cramer-Rao. Para obtenerla,

$$\begin{aligned}
 I_n(\sigma) &= -E\left(\frac{d^2 \log L(\sigma, \mathbf{x})}{d\sigma^2}\right) \\
 &= -E\left(\frac{n}{\sigma^2} - 2 \frac{\sum_{i=1}^n |X_i|}{\sigma^3}\right) & [0,4] \\
 &= -\frac{n}{\sigma^2} + 2 \frac{\sum_{i=1}^n E(|X_i|)}{\sigma^3} & [0,4] \\
 &= -\frac{n}{\sigma^2} + 2 \frac{\sum_{i=1}^n \sigma}{\sigma^3} \\
 &= \frac{n}{\sigma^2}. & [0,4]
 \end{aligned}$$

Luego, para n grande, es posible aproximar la varianza de $\hat{\sigma}$ por

$$\frac{1}{I_n(\sigma)} = \frac{\sigma^2}{n}, \quad [0,2]$$

la que corresponde a la varianza exacta de $\hat{\sigma}$ (obtenida en b)) para cualquier valor de n . [0,2]

[1,0] punto base

Problema 4

Suponga que interesa identificar algunos factores que ayuden a explicar el crecimiento de la industria manufacturera de los países en vías de desarrollo. Para esto, se recolectó información de 48 países en vías de desarrollo, correspondiente a:

- Y : Crecimiento porcentual de la industria manufacturera,
- X_1 : Crecimiento agrícola porcentual,
- X_2 : Crecimiento porcentual de las exportaciones,
- X_3 : Tasa porcentual de inflación,

ajustándose los siete modelos de regresión posibles:

Modelo	n	k	SCE	SCR	SCT	R ²	r ²
$Y \sim X_1$	48	1	871.4817	80.46502	951.9467	0.08452681	0.06462522
$Y \sim X_2$	48	1	726.0273	225.91941	951.9467	0.23732360	0.22074368
$Y \sim X_3$	48	1	882.1841	69.76260	951.9467	0.07328415	0.05313815
$Y \sim X_1+X_2$	48	2	644.4629	307.48376	951.9467	0.32300524	0.29291659
$Y \sim X_1+X_3$	48	2	804.8916	147.05503	951.9467	0.15447822	0.11689948
$Y \sim X_2+X_3$	48	2	656.3386	295.60805	951.9467	0.31053005	0.27988694
$Y \sim X_1+X_2+X_3$	48	3	577.9669	373.97975	951.9467	0.39285788	0.35146183

Proponga un modelo de regresión múltiple adecuado, comenzando desde un modelo simple, hasta una propuesta más compleja, si es que corresponde. Justifique paso a paso.

Utilice la siguiente tabla de percentiles de la distribución F , y un nivel de significancia de 5 %.

qf(0.95, df1, df2)												
	df2 = 40	df2 = 41	df2 = 42	df2 = 43	df2 = 44	df2 = 45	df2 = 46	df2 = 47	df2 = 48	df2 = 49	df2 = 50	
df1 = 1	4.084746	4.078546	4.072654	4.067047	4.061706	4.056612	4.051749	4.047100	4.042652	4.038393	4.034310	
df1 = 2	3.231727	3.225684	3.219942	3.214480	3.209278	3.204317	3.199582	3.195056	3.190727	3.186582	3.182610	
df1 = 3	2.838745	2.832747	2.827049	2.821628	2.816466	2.811544	2.806845	2.802355	2.798061	2.793949	2.790008	
df1 = 4	2.605975	2.599969	2.594263	2.588836	2.583667	2.578739	2.574035	2.569540	2.565241	2.561124	2.557179	
df1 = 5	2.449466	2.443429	2.437693	2.432236	2.427040	2.422085	2.417356	2.412837	2.408514	2.404375	2.400409	
df1 = 6	2.335852	2.329771	2.323994	2.318498	2.313264	2.308273	2.303509	2.298956	2.294601	2.290432	2.286436	
df1 = 7	2.249024	2.242894	2.237070	2.231530	2.226253	2.221221	2.216417	2.211827	2.207436	2.203232	2.199202	
df1 = 8	2.180170	2.173989	2.168117	2.162530	2.157208	2.152133	2.147288	2.142658	2.138229	2.133988	2.129923	
df1 = 9	2.124029	2.117797	2.111875	2.106241	2.100873	2.095755	2.090868	2.086198	2.081730	2.077452	2.073351	
df1 = 10	2.077248	2.070965	2.064994	2.059313	2.053901	2.048739	2.043811	2.039101	2.034595	2.030279	2.026143	

Solución:

1. Según mayor r^2 y R^2 , el modelo seleccionado sería

$$Y = \beta_0 + \beta_2 \cdot X_2 + \varepsilon \quad [0.3 \text{ Ptos.}] \quad (1)$$

La significancia de este modelo la contrastamos con las siguientes hipótesis:

$$H_0 : \beta_2 = 0 \quad \text{vs} \quad H_1 : \beta_2 \neq 0 \quad [0.3 \text{ Ptos.}]$$

Se rechaza H_0 si $F > F_{1-\alpha}(1, n-2)$, [0.3 Ptos.] con

$$F = \frac{SCR/1}{SCE/(n-1-1)} \sim F(1, n-2) \quad [0.3 \text{ Ptos.}]$$

Reemplazando tenemos que

$$[0.3 \text{ Ptos.}] \quad F = \frac{225.91941/1}{726.0273/46} = 14.31391 > 4.051749 = F_{0.95}(1, 46) \quad [0.3 \text{ Ptos.}]$$

Por lo tanto, se concluye que el aporte de la variable X_2 es significativo. [0.2 Ptos.]

2. Tomando como referencia el modelo en propuesto en (1), entonces por mayor incremento en r^2 y R^2 , el siguiente modelo seleccionado sería

$$Y = \beta_0 + \beta_1 \cdot X_1 + \beta_2 X_2 + \varepsilon \quad [0.3 \text{ Ptos.}] \quad (2)$$

El aporte de la variable ingresada al modelo propuesto en (1) la contrastamos con las siguientes hipótesis:

$$H_0 : \beta_1 = 0 \quad \text{vs} \quad H_1 : \beta_1 \neq 0 \quad [0.3 \text{ Ptos.}]$$

Se rechaza H_0 si $F > F_{1-\alpha}(1, n-3)$, [0.3 Ptos.] con

$$F = \frac{(SCR_{(1)} - SCE)/1}{SCE/(n-2-1)} \sim F(1, n-3) \quad [0.3 \text{ Ptos.}]$$

Reemplazando tenemos que

$$[0.3 \text{ Ptos.}] \quad F = \frac{(726.0273 - 644.4629)/1}{644.4629/45} = 5.695282 > 4.056612 = F_{0.95}(1, 45) \quad [0.3 \text{ Ptos.}]$$

Por lo tanto, se concluye que el aporte de la variable X_1 , en presencia de X_2 es significativo. [0.2 Ptos.]

3. Tomando como referencia el modelo en propuesto en (2), interesa ver si el aporte de la variable X_3 es significativo en presencia de X_1 y X_2 . El modelo final sería:

$$Y = \beta_0 + \beta_1 \cdot X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon \quad [0.3 \text{ Ptos.}] \quad (3)$$

Se observa un incremento importante en los coeficientes de determinación r^2 y R^2 , para ver si esa ganancia en calidad de ajuste es significativa, la contrastamos con las siguientes hipótesis:

$$H_0 : \beta_3 = 0 \quad \text{vs} \quad H_1 : \beta_3 \neq 0 \quad [0.3 \text{ Ptos.}]$$

Se rechaza H_0 si $F > F_{1-\alpha}(1, n-4)$, [0.3 Ptos.] con

$$F = \frac{(SCR_{(1)} - SCE)/1}{SCE/(n-3-1)} \sim F(1, n-4) \quad [0.3 \text{ Ptos.}]$$

Reemplazando tenemos que

$$[0.3 \text{ Ptos.}] \quad F = \frac{(644.4629 - 577.9669)/1}{577.9669/44} = 5.062269 > 4.061706 = F_{0.95}(1, 44) \quad [0.3 \text{ Ptos.}]$$

Por lo tanto, se concluye que el aporte de la variable X_3 , en presencia de X_1 y X_2 es significativo, y el modelo final sería el propuesto en (3) [0.2 Ptos.]

+ 1 Punto Base

Tiempo: 2 Horas

Distribuciones

Distribución	Densidad de Probabilidad	Θ_X	Parámetros	Esperanza y Varianza
Binomial	$\binom{n}{x} p^x (1-p)^{n-x}$	$x = 0, \dots, n$	n, p	$\mu_X = n p$ $\sigma_X^2 = n p (1-p)$ $M(t) = [p e^t + (1-p)]^n, \quad t \in \mathbb{R}$
Geométrica	$p (1-p)^{x-1}$	$x = 1, 2, \dots$	p	$\mu_X = 1/p$ $\sigma_X^2 = (1-p)/p^2$ $M(t) = p e^t / [1 - (1-p) e^t], \quad t < -\ln(1-p)$
Binomial-Negativa	$\binom{x-1}{r-1} p^r (1-p)^{x-r}$	$x = r, r+1, \dots$	r, p	$\mu_X = r/p$ $\sigma_X^2 = r(1-p)/p^2$ $M(t) = \{p e^t / [1 - (1-p) e^t]\}^r, \quad t < -\ln(1-p)$
Poisson	$\frac{(\nu t)^x e^{-\nu t}}{x!}$	$x = 0, 1, \dots$	ν	$\mu_X = \nu t$ $\sigma_X^2 = \nu t$ $M(t) = \exp[\lambda (e^t - 1)], \quad t \in \mathbb{R}$
Exponencial	$\nu e^{-\nu x}$	$x \geq 0$	ν	$\mu_X = 1/\nu$ $\sigma_X^2 = 1/\nu^2$ $M(t) = \nu / (\nu - t), \quad t < \nu$
Gamma	$\frac{\nu^k}{\Gamma(k)} x^{k-1} e^{-\nu x}$	$x \geq 0$	k, ν	$\mu_X = k/\nu$ $\sigma_X^2 = k/\nu^2$ $M(t) = [\nu / (\nu - t)]^k, \quad t < \nu$
Normal	$\frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$	$-\infty < x < \infty$	μ, σ	$\mu_X = \mu$ $\sigma_X^2 = \sigma^2$ $M(t) = \exp(\mu t + \sigma^2 t^2 / 2), \quad t \in \mathbb{R}$
Log-Normal	$\frac{1}{\sqrt{2\pi}(\zeta x)} \exp\left[-\frac{1}{2}\left(\frac{\ln x - \lambda}{\zeta}\right)^2\right]$	$x \geq 0$	λ, ζ	$\mu_X = \exp\left(\lambda + \frac{1}{2}\zeta^2\right)$ $\sigma_X^2 = \mu_X^2 (e^{\zeta^2} - 1)$ $E(X^r) = e^{r\lambda} M_Z(r\zeta), \text{ con } Z \sim \text{Normal}(0,1)$
Uniforme	$\frac{1}{(b-a)}$	$a \leq x \leq b$	a, b	$\mu_X = (a+b)/2$ $\sigma_X^2 = (b-a)^2/12$ $M(t) = [e^{tb} - e^{ta}] / [t(b-a)], \quad t \in \mathbb{R}$
Beta	$\frac{1}{B(q, r)} \frac{(x-a)^{q-1} (b-x)^{r-1}}{(b-a)^{q+r-1}}$	$a \leq x \leq b$	q, r	$\mu_X = a + \frac{q}{q+r} (b-a)$ $\sigma_X^2 = \frac{q r (b-a)^2}{(q+r)^2 (q+r+1)}$
Hipergeométrica	$\frac{\binom{m}{x} \binom{N-m}{n-x}}{\binom{N}{n}}$	$\max\{0, n+m-N\} \leq x \leq \min\{n, m\}$	N, m, n	$\mu_X = n \frac{m}{N}$ $\sigma_X^2 = \left(\frac{N-n}{N-1}\right) n \frac{m}{N} \left(1 - \frac{m}{N}\right)$

Formulario

- Sea X_1, \dots, X_n una muestra aleatoria independiente e idénticamente distribuida con función de probabilidad p_X o de densidad f_X , determinada por un parámetro θ . Si $\hat{\theta}$ es el estimador máximo verosímil del parámetro θ , entonces:
 - $E(\hat{\theta}) \rightarrow \theta$, cuando $n \rightarrow \infty$.
 - $Var(\hat{\theta}) = \frac{1}{I_n(\hat{\theta})}$, con $I_n(\theta) = -E\left[\frac{\partial^2}{\partial \theta^2} \ln L(\theta)\right]$.
 - $\hat{\theta} \sim \text{Normal}\left(\theta, \sqrt{\frac{1}{I_n(\hat{\theta})}}\right)$, cuando $n \rightarrow \infty$.
 - El estimador máximo verosímil de $g(\theta)$ es $g(\hat{\theta})$, cuya varianza está dada por: $Var[g(\hat{\theta})] = \frac{[g'(\theta)]^2}{I_n(\theta)}$.

- Sean X_1, \dots, X_n variables aleatorias independientes e idénticamente distribuidas $\text{Normal}(\mu, \sigma)$, entonces

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim \text{Normal}(0, 1), \quad \frac{\bar{X}_n - \mu}{s/\sqrt{n}} \sim \text{t-student}(n-1), \quad \frac{s^2(n-1)}{\sigma^2} \sim \chi_{n-1}^2$$

$$\text{con } s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

Regresión Lineal Múltiple

Para el siguiente modelo de regresión

$$y_i = \beta_0 + \beta_1 \cdot x_{i1} + \dots + \beta_k \cdot x_{ik} + \varepsilon_i$$

con $E(\varepsilon_i) = 0$, $\text{Var}(\varepsilon_i) = \sigma^2$ y $E(\varepsilon_i \cdot \varepsilon_j) = 0$ para $i \neq j$, se define el modelo de regresión múltiple estimado como

$$\hat{y}_i = b_0 + b_1 \cdot x_{i1} + \dots + b_k \cdot x_{ik}$$

Descomposición de cuadrados

$$SCT = SCR + SCE$$

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Coefficiente de determinación

$$R^2 = \frac{SCR}{SCT} = 1 - \frac{SCE}{SCT} = 1 - \frac{(n-k-1)}{(n-1)} \frac{s_{Y|x}^2}{s_Y^2}$$

Coefficiente de determinación ajustado

$$r^2 = 1 - \frac{(n-1)}{(n-k-1)} \frac{SCE}{SCT} = 1 - \frac{s_{Y|x}^2}{s_Y^2}$$

Distribuciones

$$T_{b_j} = \frac{b_j - \beta_j}{s_{b_j}} \sim \text{t-Student}(n-k-1) \quad F = \frac{SCR/k}{SCE/(n-k-1)} \sim F(k, n-k-1)$$

$$F = \frac{(SCE_{(m)} - SCE)/m}{SCE/(n-k-1)} \sim F(m, n-k-1)$$

con $SCE_{(m)}$ igual a la suma de errores al cuadrado del modelo sin m de los regresores que aparecen en el modelo completo.

Tablas de Percentiles p

Distribución Normal Estándar k_p											Distribución t-student $t_p(\nu)$				
k_p	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	ν	$t_{0.90}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359	1	3.078	6.314	12.706	31.821
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753	2	1.886	2.920	4.303	6.965
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141	3	1.638	2.353	3.182	4.541
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517	4	1.533	2.132	2.776	3.747
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879	5	1.476	2.015	2.571	3.365
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224	6	1.440	1.943	2.447	3.143
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549	7	1.415	1.895	2.365	2.998
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852	8	1.397	1.860	2.306	2.896
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133	9	1.383	1.833	2.262	2.821
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389	10	1.372	1.812	2.228	2.764
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621	11	1.363	1.796	2.201	2.718
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830	12	1.356	1.782	2.179	2.681
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015	13	1.350	1.771	2.160	2.650
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177	14	1.345	1.761	2.145	2.624
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319	15	1.341	1.753	2.131	2.602
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441	16	1.337	1.746	2.120	2.583
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545	17	1.333	1.740	2.110	2.567
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633	18	1.330	1.734	2.101	2.552
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706	19	1.328	1.729	2.093	2.539
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767	20	1.325	1.725	2.086	2.528
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817	21	1.323	1.721	2.080	2.518
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857	22	1.321	1.717	2.074	2.508
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890	23	1.319	1.714	2.069	2.500
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916	24	1.318	1.711	2.064	2.492
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936	25	1.316	1.708	2.060	2.485
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952	26	1.315	1.706	2.056	2.479
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964	27	1.314	1.703	2.052	2.473
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974	28	1.313	1.701	2.048	2.467
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981	29	1.311	1.699	2.045	2.462
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986	30	1.310	1.697	2.042	2.457
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990	∞	1.282	1.645	1.960	2.326

Distribución Chi-Cuadrado $c_p(\nu)$								
ν	$c_{0.025}$	$c_{0.05}$	$c_{0.10}$	$c_{0.90}$	$c_{0.95}$	$c_{0.975}$	$c_{0.99}$	$c_{0.995}$
1	0.00	0.00	0.02	2.71	3.84	5.02	6.63	7.88
2	0.05	0.10	0.21	4.61	5.99	7.38	9.21	10.60
3	0.22	0.35	0.58	6.25	7.81	9.35	11.34	12.84
4	0.48	0.71	1.06	7.78	9.49	11.14	13.28	14.86
5	0.83	1.15	1.61	9.24	11.07	12.83	15.09	16.75
6	1.24	1.64	2.20	10.64	12.59	14.45	16.81	18.55
7	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.95
9	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	3.82	4.57	5.58	17.28	19.68	21.92	24.72	26.76
12	4.40	5.23	6.30	18.55	21.03	23.34	26.22	28.30
13	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	6.26	7.26	8.55	22.31	25.00	27.49	30.58	32.80
16	6.91	7.96	9.31	23.54	26.30	28.85	32.00	34.27
17	7.56	8.67	10.09	24.77	27.59	30.19	33.41	35.72
18	8.23	9.39	10.86	25.99	28.87	31.53	34.81	37.16
19	8.91	10.12	11.65	27.20	30.14	32.85	36.19	38.58
20	9.59	10.85	12.44	28.41	31.41	34.17	37.57	40.00
21	10.28	11.59	13.24	29.62	32.67	35.48	38.93	41.40
22	10.98	12.34	14.04	30.81	33.92	36.78	40.29	42.80
23	11.69	13.09	14.85	32.01	35.17	38.08	41.64	44.18
24	12.40	13.85	15.66	33.20	36.42	39.36	42.98	45.56
25	13.12	14.61	16.47	34.38	37.65	40.65	44.31	46.93