Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística

Primer Semestre 2018

Curso	:	**Probabilidad y Estadística**
Sigla	:	**EYP1113**
Pauta.	:	**I3**
Profesores	:	**Ricardo Aravena C. y Ricardo Olea O.**

Problema 1

En cierta universidad, la escuela de ingeniería que tiene una composición de 3.600 hombres y 1.200 mujeres. Para efecto de adherir a paros o marchas han decidido que se deben cumplir dos criterios:

i. Quorum superior al 40 % del padrón.

ii. Se aprueba la adhesión a un paro si al menos 3/4 de las mujeres lo aprueban.

Usted, interesado en "predecir" el comportamiento frente al próximo movimiento, realiza una rápida encuesta entre los alumnos y alumnas de esta escuela obteniendo

Resultados	Hombres	Mujeres
Num Casos	85	45
Vota	25	40
Aprueba	15	35

(a) **[1.5 Ptos]** ¿Se logra el quorum exigido? Indicar hipótesis, test a utilizar, y su conclusión para un nivel de significancia del 1 %.

(b) **[1.5 Ptos]** ¿Para qué nivel de significancia se puede afirmar que la adhesión al paro de las mujeres supera a la adhesión al paro de los hombres?

(c) **[1.5 Ptos]** ¿se aprueba o rechaza el paro? Indicar hipótesis, test a utilizar, especificar supuestos y su conclusión para un nivel de significancia del 5 %.

(d) **[1.5 Ptos]** Frente a críticas que el error de las estimaciones es muy alto, indique cual debe ser el tamaño muestral mínimo para que la estimación de la aprobación del paro tenga un error no mayor a un 5 % con un 90 % de confianza. Sea explícito en sus supuestos. Fundamente.

Solución

(a) Se pide realizar una prueba de hipótesis para r una proporción

$$\text{H}_0 : p = p_0 \quad \text{vs} \quad \text{H}_a : p > p_0, \qquad \textbf{[0.5 Ptos]}$$

con $p_0 = 0{,}4$.

Bajo H_0, el estadísticos de prueba $Z_0 = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0\,(1 - p_0)}{n}}} = 2{,}327373$, con $\hat{p} = \dfrac{65}{130} = 0{,}5$. **[0.5 Ptos]**

Alternativa 1: Dado que $Z_0 = 2{,}327373 > 2{,}326 = k_{0{,}99}$, existe suficiente evidencia al 1 % de significancia para rechazar H_0, es decir, se debería alcanzar el quorum. **[0.5 Ptos]**

Alternativa 2: Como valor-p $= P(Z > 2{,}327373) = 0{,}0099 < 0{,}01 = \alpha$, existe suficiente evidencia al 1 % de significancia para rechazar H_0, es decir, se debería alcanzar el quorum. **[0.5 Ptos]**

(b) Se pide realizar un test de comparación de proporciones

$$\mathrm{H}_0 : p_H = p_M \quad \text{vs} \quad \mathrm{H}_a : p_H < p_M \qquad \textbf{[0.5 Ptos]}$$

Bajo H_0, el estadísticos de prueba $Z_0 = \dfrac{\hat{p_H} - \hat{p_M}}{\sqrt{\hat{p}\,(1-\hat{p})\left(\dfrac{1}{n} + \dfrac{1}{m}\right)}} = -2{,}560111,$ **[0.5 Ptos]**

con $\hat{p} = \dfrac{15+35}{25+40} = 0{,}7692308.$

Se pide

$$\textbf{[0.5 Ptos]} \quad \text{valor-p} = \Phi(-2{,}560111)1 = 1 - \Phi(2{,}560111) \approx 1 - 0{,}9948 = 0{,}0052, \qquad \textbf{[0.5 Ptos]}$$

que sería el menor nivel de significancia para el cual se puede afirmar que adhesión al paro de las mujeres supera a la adhesión al paro de los hombres.

(c) Se pide un test de proporción solo mujeres, es decir, $n = 40$.

$$\mathrm{H}_0 : p = p_0 \quad \text{vs} \quad \mathrm{H}_a : p < p_0, \qquad \textbf{[0.5 Ptos]}$$

con $p_0 = 0{,}750$ y por que el al menos implica igualdad, razón por la cuál $\mathrm{H}_a : p < p_0$.

Bajo H_0, el estadísticos de prueba $Z_0 = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0\,(1-p_0)}{n}}} = 1{,}825$, con $\hat{p} = \dfrac{35}{40} = 0{,}875.$ **[0.5 Ptos]**

Alternativa 1: Dado que $Z_0 = 1{,}825 > -1{,}645 = k_{0{,}05}$, no existe suficiente evidencia al 5 % de significancia para rechazar H_0, es decir, se aprueba el paro. **[0.5 Ptos]**

Alternativa 2: Como valor-p $= P(Z < 1{,}825) \approx 0{,}9664 > 0{,}05 = \alpha$, no existe suficiente evidencia al 5 % de significancia para rechazar H_0, es decir, se aprueba el paro. **[0.5 Ptos]**

(d) Se pide un n de mujeres que manifiestan ir a votar, para que la aprobación vistan en el punto (c) tenga menor error suponiendo que el quorum se cumple.

Alternativa 1: Criterio varianza máxima

$$n = \left(\frac{k_{0{,}95}}{2\cdot\omega}\right)^2 = \left(\frac{1{,}645}{2\cdot 0{,}05}\right)^2 = 270{,}6025 \to 271 \text{ mujeres dispuestas a votar en entrevistar} \qquad \textbf{[1.5 Ptos]}$$

Alternativa 2: Tomando el 87,5 % observado en (c) como base

$$n = \left(\frac{k_{0{,}95}\,\sqrt{\hat{p}\,(1-\hat{p})}}{\omega}\right)^2$$
$$= \left(\frac{1{,}645\cdot\sqrt{0{,}875\cdot(1-0{,}875)}}{0{,}05}\right)^2 = 118{,}3886 \to 119 \text{ mujeres dispuestas a votar en entrevistar} \qquad \textbf{[1.5 Ptos]}$$

+ 1 Punto Base

Problema 2

Esta semana, la ciudad de Santiago fue afectada por un graupel (entre granizo y copo de nieve) y por lo extraño del fenómeno, se aprovecho de tomar mediciones del diámetro de estas estructuras de cristales sólidos de hielo sobre un suave núcleo de origen nivoso en diversas zonas de la región metropolitana cuyo comportamiento empírico se puede ajustar por un modelo IGAU(β) definido por la siguiente función de densidad

$$f(x) = \frac{1}{x}\,\phi[\ln(x),\,\beta], \qquad \phi(z,\,\beta) = \frac{\sqrt{\beta}}{\exp(z/2)}\,\phi_{\text{Normal}}\left\{\sqrt{\beta}\left[\frac{\exp(z)+1}{\exp(z/2)}\right]\right\}$$

con valor esperado uno y varianza $1/\beta$.

(a) **[3.0 Ptos]** Obtenga el estimador máximo verosímil de β y su distribución asintótica.

(b) **[2.0 Ptos]** Obtenga el estimador de momento de β y su ECM aproximado de 1er orden.

(c) **[1.0 Ptos]** ¿Son igualmente de eficientes ambos estimadores, según los resultados obtenidos en (a) y (b)?

Solución

(a) Sean X_1, X_2,..., X_n una muestra aleatoria de diámetros de graupeles cuyo comportamiento será ajustado por una distribución IGAU(β).

La función de verosimilitud y su logarítmo natural están dadas por

$$L(\beta) = \left(\prod_{i=1}^{n} X_i\right)^{-1} \beta^{n/2} \left(\prod_{i=1}^{n} X_i\right)^{-1/2} (2\,\pi)^{-n/2}\,\exp\left[-\frac{\beta}{2}\sum_{i=1}^{n}\frac{(X_i+1)^2}{X_i}\right] \qquad \textbf{[0.6 Ptos]} \qquad (1)$$

$$\ln\,L(\beta) = -\frac{3}{2}\sum_{i=1}^{n}\ln(X_i) + \frac{n}{2}\,\ln(\beta) - \frac{n}{2}\,\ln(2\,\pi) - \frac{\beta}{2}\sum_{i=1}^{n}\frac{(X_i+1)^2}{X_i} \qquad \textbf{[0.6 Ptos]} \qquad (2)$$

Derivando parcialmente con respecto a β e igualando a cero se tiene que el estimador de máxima verosimilitud de β está dado por

$$\hat{\beta} = \frac{n}{\displaystyle\sum_{i=1}^{n}\frac{(X_i+1)^2}{X_i}} \qquad \textbf{[0.6 Ptos]}$$

Derivando (2) nuevamente con respecto a β se tiene

$$\textbf{[0.3 Ptos]} \qquad \frac{\partial^2}{\partial\beta^2}\,\ln\,L(\beta) = -\frac{n}{2\,\beta^2} \to I(\beta) = -\text{E}\left(-\frac{n}{2\,\beta^2}\right) = \frac{n}{2\,\beta^2} \qquad \textbf{[0.3 Ptos]}$$

Luego, su distribución asintótica está dada por

$$\hat{\beta} \overset{\text{aprox}}{\sim} \text{Normal}\left(\beta,\,\sqrt{\frac{2\,\beta^2}{n}}\right) \qquad \textbf{[0.6 Ptos]}$$

(b) Como el primer momento no depende de β procedemos a igualar el segundo momento teórico al empírico, es decir,

$$\text{E}(X^2) = \frac{1}{\beta} + 1 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \overline{X^2} \qquad \textbf{[0.5 Ptos]}$$

Despejando, el estimador de momentos de β está dado por

$$\widetilde{\beta} = \frac{1}{\overline{X^2} - 1} = \frac{n}{\left(\displaystyle\sum_{i=1}^{n} X_i^2\right) - n} \qquad \textbf{[0.5 Ptos]}$$

Su esperanza y varianza aproximada de 1er orden están dadas por

$$\mathrm{E}(\widetilde{\beta}) = \mathrm{E}[g(X_1^2, \ldots, X_n^2)] \approx \frac{n}{\left(\displaystyle\sum_{i=1}^{n} \mu_{X_i^2}\right) - n} = \frac{n}{\frac{n}{\beta} + n - n} = \beta \qquad \textbf{[0.5 Ptos]}$$

y como es insesgado,

$$\mathrm{ECM}(\widetilde{\beta}) = \mathrm{Var}(\widetilde{\beta}) = \mathrm{Var}[g(X_1^2, \ldots, X_n^2)] \overset{\text{ind}}{\approx} \sum_{i=1}^{n} \sigma_{X_i^2}^2 \left[\frac{-2\,n\,\sqrt{\mu_{X_i^2}}}{\left(\displaystyle\sum_{j=1}^{n} \mu_{X_j^2} - n\right)^2}\right]^2 \overset{\text{id}}{=} \frac{2\,\beta^2}{n}\,2\,\beta\,(\beta{+}1)\,\sigma_{X^2}^2 \qquad \textbf{[0.5 Ptos]}$$

con $\sigma_{X^2}^2 = \mathrm{Var}(X^2)$.

(c) Como $\sigma_{X^2}^2 = \mathrm{Var}(X^2) > \mathrm{Var}(X) = \frac{1}{\beta}$, entonces

$$\mathrm{Var}(\widetilde{\beta}) > \frac{2\,\beta^2}{n}\,2\,(\beta + 1) = \mathrm{Var}(\hat{\beta})\,2\,(\beta + 1) \qquad \textbf{[0.5 Ptos]}$$

y como ambos estimadores son insesgados, el más eficiente es el que tiene menor varianza y como $2\,(\beta + 1) > 1$, entonces $\hat{\beta}$ es más eficiente que $\widetilde{\beta}$. **[0.5 Ptos]**

+ 1 Punto Base

Problema 3

Frente al tema de género que se ha instalado en la discusión nacional, un investigador busca determinar si existen diferencias basales en el desempeño según género. Específicamente, su hipótesis es que frente a situaciones de stress (por ejemplo, desarrollo de una evaluación) las mujeres tienen un mayor control. Con el fin de verificar o refutar su hipótesis lleva a cabo un cuasi-experimento que consisten en someter a una situación estresantes a dos grupos - Hombres y Mujeres - seleccionados al azar dentro de los alumnos y alumnas del curso, registrando la presión arterial durante la evaluación. Los resultados son:

Resultados	Hombres	Mujeres	Ambos
Num Casos	15	17	32,00
Promedio	85	92	88,00
Mediana	80	85	84,00
Desv. Estándar	8	13	10,00
Promedio LN	-	-	4,48
Desv. Estándar LN	-	-	0,12

Asumiendo que la presión arterial diastólica se comporta de acuerdo a una distribución normal y teniendo claro que un alza de presión implica una pérdida de control frente a situaciones estresante,

(a) **[4.0 Ptos]** ¿Es válida la afirmación del experto para un nivel del 5 %? Debe plantear hipótesis, indicar test a utilizar, especificar y validar supuestos (cuando sea necesario) y sus conclusiones deben estar basada en el valor-p.

(b) **[2.0 Ptos]** Al revisar los datos conjuntos (hombres y mujeres juntos) se aprecia que el promedio no coincide con la mediana, por tanto, se propone que la presión arterial diastólica se puede asumir lognormal de parámetros λ y ζ, donde este último valor conocido e igual a 0,5 mmHg. Lleve a un test apropiado para determinar si existe evidencia que permita afirmar que la mediana de las presiones arterial diastólica es superior a 80 mmHg. Use $\alpha = 5\,\%$.

Solución

(a) Test de comparación de medias, muestras con distribución normal y varianzas desconocidas.

Primero se realiza un test de comparación de varianzas

$$\mathrm{H}_0 : \sigma_H = \sigma_M \quad \text{vs} \quad \mathrm{H}_a : \sigma_H \neq \sigma_M, \qquad \textbf{[0.5 Ptos]}$$

Bajo H_0, el estadísticos de prueba $F_0 = \dfrac{S_H^2}{S_M^2} \sim F(15-1,\, 17-1)$. **[0.5 Ptos]**

Como $F_0 = 0{,}3786982 > 0{,}3424658 = F_{0{,}025}(14,\, 16) \to$ valor-p $> 5\,\%$, entonces no existe suficiente evidencia para rechazar la igualdad de varianzas al 5 % de significancia. **[0.5 Ptos]**

Dado el resultado anterior, se realizará un test de comparación de medias con varianzas desconocidad, pero iguales.

$$\mathrm{H}_0 : \mu_H = \mu_M \quad \text{vs} \quad \mathrm{H}_a : \mu_H > \mu_M, \qquad \textbf{[0.5 Ptos]}$$

Bajo H_0, es estadístico de prueba

$$T_0 = \frac{\overline{X}_H - \overline{X}_M}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim \text{t-Student}(n + m - 2) \qquad \textbf{[0.5 Ptos]}$$

con $S_p^2 = \dfrac{(n-1)\,S_H^2 + (m-1)\,S_F^2}{n + m - 2} = 120.$ **[0.5 Ptos]**

Reemplazando $T_0 = -1{,}803859 \to 95\,\% <$ valor-p $< 97{,}5\,\%$. Por lo tanto no existe suficiente evidencia para rechazar H_0 y no se apoya la hipótesis que las mujeres tienen más control. **[1.0 Ptos]**

(b) Tenemos que la muestra conjunta distribuye Log-Normal(λ, $\zeta = 0{,}5$), donde la mediana es igual a e^{λ}.

El estimador máximo verosímil de λ es

$$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} \ln(X_i) \overset{\text{aprox}}{\sim} \text{Normal}\left(\lambda,\, \frac{\zeta}{\sqrt{n}}\right)$$

mientras que de la mediana e^{λ} es

$$e^{\hat{\lambda}} = \exp\left[\frac{1}{n}\sum_{i=1}^{n} \ln(X_i)\right] \overset{\text{aprox}}{\sim} \text{Normal}\left(e^{\lambda},\, \frac{\zeta\, e^{\lambda}}{\sqrt{n}}\right)$$

Alternativa 1: Hacer un test sobre la mediana

$$\text{H}_0 : e^{\lambda} = 80 \quad \text{vs} \quad \text{H}_a : e^{\lambda} > 80, \qquad \textbf{[0.5 Ptos]}$$

Bajo H_0, es estadístico de prueba

$$Z_0 = \frac{e^{\hat{\lambda}} - 80}{\zeta \cdot 80/\sqrt{n}} \sim \text{Normal}(0,\, 1) \qquad \textbf{[1.0 Ptos]}$$

Dado que $Z_0 = 1{,}164559 < 1{,}645 = k_{0{,}95}$, no existe suficiente evidencia al 5 % de significancia para rechazar H_0 y no se apoya la hipótesis que la mediana supera los 80 mmHg. **[0.5 Ptos]**

Alternativamente, dado que $Z_0 = 1{,}164559 \to$ valor-p $= 0{,}1230244 > 0{,}05 = \alpha$, no existe suficiente evidencia al 5 % de significancia para rechazar H_0 y no se apoya la hipótesis que la mediana supera los 80 mmHg. **[0.5 Ptos]**

Alternativa 2: Hacer un test sobre λ

$$\text{H}_0 : \lambda = \ln(80) \quad \text{vs} \quad \text{H}_a : \lambda > \ln(80), \qquad \textbf{[0.5 Ptos]}$$

Bajo H_0, es estadístico de prueba

$$Z_0 = \frac{\hat{\lambda} - \ln(80)}{\zeta/\sqrt{n}} \sim \text{Normal}(0,\, 1) \qquad \textbf{[1.0 Ptos]}$$

Dado que $Z_0 = 1{,}108442 < 1{,}645 = k_{0{,}95}$, no existe suficiente evidencia al 5 % de significancia para rechazar H_0 y no se apoya la hipótesis que la mediana supera los 80 mmHg. **[0.5 Ptos]**

Alternativamente, dado que $Z_0 = 1{,}108442 \to$ valor-p $= 0{,}1334995 > 0{,}05 = \alpha$, no existe suficiente evidencia al 5 % de significancia para rechazar H_0 y no se apoya la hipótesis que la mediana supera los 80 mmHg. **[0.5 Ptos]**

+ 1 Punto Base

Problema 4

Sean $Y1_1$, Y_2 e Y_3 variables aleatorias independientes con $\text{E}(Y_i) = 0$ y $\text{Var}(Y_i) = 1$, con $i = 1$, 2 y 3. Defina las siguientes variables aleatorias:

$$X_1 = \frac{Y_1}{\sqrt{1-\phi^2}}, \quad X_2 = \phi\, X_1 + Y_2, \quad X_3 = \phi\, X_2 + Y_3$$

donde $|\phi| < 1$.

(a) **[2.0 Ptos]** Determine el valor esperado de X_i, para $i = 1$, 2 y 3.

(b) **[4.0 Ptos]** Calcule $\text{Corr}(X_i,\, X_j)$, para todo $i,\, j \in \{1, 2, 3\}$.

Solución

(a) Del enunciado se tiene que X_1, X_2 y X_3 dependen de las Y's, y aplicando operador esperanza se tiene que

$$\textbf{[0.3 Ptos]} \quad \text{E}(X_1) = \text{E}\left(\frac{1}{\sqrt{1-\phi^2}}\, Y_1\right) = \frac{1}{\sqrt{1-\phi^2}}\, \text{E}(Y_1) = 0 \quad \textbf{[0.3 Ptos]}$$

$$\textbf{[0.3 Ptos]} \quad \text{E}(X_2) = \text{E}\left(\frac{\phi}{\sqrt{1-\phi^2}}\, Y_1 + Y_2\right) = \frac{\phi}{\sqrt{1-\phi^2}}\, \text{E}(Y_1) + \text{E}(Y_2) = 0 \quad \textbf{[0.4 Ptos]}$$

$$\textbf{[0.3 Ptos]} \quad \text{E}(X_3) = \text{E}\left(\frac{\phi^2}{\sqrt{1-\phi^2}}\, Y_1 + \phi\, Y_2 + Y_3\right) = \frac{\phi^2}{\sqrt{1-\phi^2}}\, \text{E}(Y_1) + \phi\, \text{E}(Y_2) + \text{E}(Y_3) = 0 \quad \textbf{[0.4 Ptos]}$$

(b) Por la independencia entre Y_1, Y_2 e Y_3, se tiene que las covarianzas

$$\text{Cov}(X_1,\, X_1) = \text{Cov}\left(\frac{1}{\sqrt{1-\phi^2}}\, Y_1,\, \frac{1}{\sqrt{1-\phi^2}}\, Y_1\right) = \frac{1}{1-\phi^2} \quad \textbf{[0.4 Ptos]}$$

$$\text{Cov}(X_1,\, X_2) = \text{Cov}\left(\frac{1}{\sqrt{1-\phi^2}}\, Y_1,\, \frac{\phi}{\sqrt{1-\phi^2}}\, Y_1 + Y_2\right) = \frac{\phi}{1-\phi^2} \quad \textbf{[0.4 Ptos]}$$

$$\text{Cov}(X_1,\, X_3) = \text{Cov}\left(\frac{1}{\sqrt{1-\phi^2}}\, Y_1,\, \frac{\phi^2}{\sqrt{1-\phi^2}}\, Y_1 + \phi\, Y_2 + Y_3\right) = \frac{\phi^2}{1-\phi^2} \quad \textbf{[0.4 Ptos]}$$

$$\text{Cov}(X_2,\, X_2) = \text{Cov}\left(\frac{\phi}{\sqrt{1-\phi^2}}\, Y_1 + Y_2,\, \frac{\phi}{\sqrt{1-\phi^2}}\, Y_1 + Y_2\right) = \frac{\phi^2}{1-\phi^2} + 1 \quad \textbf{[0.4 Ptos]}$$

$$\text{Cov}(X_2,\, X_3) = \text{Cov}\left(\frac{\phi}{\sqrt{1-\phi^2}}\, Y_1 + Y_2,\, \frac{\phi^2}{\sqrt{1-\phi^2}}\, Y_1 + \phi\, Y_2 + Y_3\right) = \frac{\phi^3}{1-\phi^2} + \phi \quad \textbf{[0.4 Ptos]}$$

$$\text{Cov}(X_3,\, X_3) = \text{Cov}\left(\frac{\phi^2}{\sqrt{1-\phi^2}}\, Y_1 + \phi\, Y_2 + Y_3,\, \frac{\phi^2}{\sqrt{1-\phi^2}}\, Y_1 + \phi\, Y_2 + Y_3\right) = \frac{\phi^4}{1-\phi^2} + \phi^2 + 1 \quad \textbf{[0.4 Ptos]}$$

Mientras que

$$\text{Corr}(X_1,\, X_1) = 1 \quad \textbf{[0.2 Ptos]}$$

$$\text{Corr}(X_1,\, X_2) = \frac{\frac{\phi}{1-\phi^2}}{\sqrt{\frac{1}{1-\phi^2}\cdot\left(\frac{\phi^2}{1-\phi^2}+1\right)}} = \phi\sqrt{\frac{\frac{1}{1-\phi^2}}{\frac{\phi^2}{1-\phi^2}+1}} \quad \textbf{[0.3 Ptos]}$$

$$\text{Corr}(X_1,\, X_3) = \frac{\frac{\phi^2}{1-\phi^2}}{\sqrt{\frac{1}{1-\phi^2}\cdot\left(\frac{\phi^4}{1-\phi^2}+\phi^2+1\right)}} = \phi^2\sqrt{\frac{\frac{1}{1-\phi^2}}{\frac{\phi^4}{1-\phi^2}+\phi^2+1}} \quad \textbf{[0.3 Ptos]}$$

$$\text{Corr}(X_2,\, X_2) = 1 \quad \textbf{[0.2 Ptos]}$$

$$\text{Corr}(X_2,\, X_3) = \frac{\frac{\phi^3}{1-\phi^2}+\phi}{\sqrt{\left(\frac{\phi^2}{1-\phi^2}+1\right)\cdot\left(\frac{\phi^4}{1-\phi^2}+\phi^2+1\right)}} = \phi\sqrt{\frac{\frac{\phi^2}{1-\phi^2}+1}{\frac{\phi^4}{1-\phi^2}+\phi^2+1}} \quad \textbf{[0.4 Ptos]}$$

$$\text{Corr}(X_3,\, X_3) = 1 \quad \textbf{[0.2 Ptos]}$$

+ 1 Punto Base