Prueba Bonus SOL114

En el marco de las elecciones presidenciales de segunda vuelta en Chile, un centro de estudios sociales realiza una encuesta a nivel nacional para analizar el comportamiento electoral y ciertas características socioeconómicas de la población. El estudio se aplicó a individuos mayores de 18 años, residentes en todas las regiones, con muestreo aleatorio estratificado. Se recolectaron, entre otras, las siguientes cuatro variables:

- Sexo: Hombre (H) / Mujer (M)
- Intención de voto (segunda vuelta): Kast / Jara
- Ingreso mensual (en miles de pesos chilenos)
- Años de escolaridad

Se levantó una muestra grande (n = 3.000), pero a continuación se muestran 7 registros como ejemplo ilustrativo:

ID	Sexo	Voto	Ingreso (mil)	Escol. (años)
1	H	Kast	980	12
2	M	Jara	650	14
3	H	Kast	1.200	16
4	M	Jara	720	13
5	M	Kast	820	11
6	H	Kast	1.050	15
7	M	Jara	580	10

La siguiente table muestra los principales estadísticos descriptivos de la muestra:

Variable	Media	Desv. Est.	Proporción
Ingreso (mil pesos)	834	270	—
Escolaridad (años)	13.2	2.6	—
Sexo (Hombre)	—	—	0.47
Voto Kast	—	—	0.52
Voto Jara	—	—	0.48

Además, la siguiente tabla muestra la matriz de varianza–covarianza entre las variables Ingreso y Escolaridad. Una matriz de varianza–covarianza resume cuánta variación tiene cada variable por sí sola (las varianzas, ubicadas en la diagonal) y cómo se relacionan entre sí (las covarianzas, ubicadas fuera de la diagonal).

$$\Sigma = \begin{pmatrix} Var(\text{Ingreso}) & Cov(\text{Ingreso, Escolaridad}) \\ Cov(\text{Ingreso, Escolaridad}) & Var(\text{Escolaridad}) \end{pmatrix} = \begin{pmatrix} 72900 & 312.4 \\ 312.4 & 6.76 \end{pmatrix}$$

donde:

- Var(Ingreso): varianza de los ingresos individuales.
- Var(Escolaridad): varianza de los años de escolaridad.
- Cov(Ingreso, Escolaridad): covarianza entre ambas variables.

Importante:

- Todos los estimadores presentados son insesgados para sus respectivos parámetros teóricos poblacionales.
- Para todos los procedimientos inferenciales se trabajará con un nivel de confianza del 99% ($\alpha = 0.01$).

Preguntas

1. (20%) Estime los intervalos de confianza al 99% para:

a) La proporción de intención de voto por Kast

b) La proporción de intención de voto por Jara

Interprete ambos resultados en el contexto electoral.

Para construir los intervalos de confianza al 99% para proporciones utilizo la fórmula:

$$IC_{99\%} = \hat{p} \pm z_{0.005} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

donde $z_{0.005} = 2.576$ para un nivel de confianza del 99%.

Para Kast tenemos $\hat{p}_{Kast} = 0.52$ y $n = 3000$. Calculando el error estándar:

$$EE_{Kast} = \sqrt{\frac{0.52 \cdot 0.48}{3000}} = \sqrt{\frac{0.2496}{3000}} = \sqrt{0.0000832} = 0.00912$$

El margen de error es:

$$ME_{Kast} = 2.576 \cdot 0.00912 = 0.0235$$

Por lo tanto el intervalo de confianza para Kast es:

$$IC_{99\%}(Kast) = [0.52 - 0.0235, 0.52 + 0.0235] = [0.4965, 0.5435]$$

Para Jara tenemos $\hat{p}_{Jara} = 0.48$ y $n = 3000$. El error estándar es:

$$EE_{Jara} = \sqrt{\frac{0.48 \cdot 0.52}{3000}} = \sqrt{\frac{0.2496}{3000}} = 0.00912$$

El margen de error es:

$$ME_{Jara} = 2.576 \cdot 0.00912 = 0.0235$$

Por lo tanto el intervalo de confianza para Jara es:

$$IC_{99\%}(Jara) = [0.48 - 0.0235, 0.48 + 0.0235] = [0.4565, 0.5035]$$

Interpretación: Con un 99% de confianza, la verdadera proporción poblacional de votantes por Kast se encuentra entre 49.65% y 54.35%, mientras que para Jara se encuentra entre 45.65% y 50.35%. Ambos intervalos se traslapan considerablemente, lo que indica que la elección está técnicamente empatada dentro del margen de error estadístico. Ninguno de los candidatos puede declararse claramente ganador basándose únicamente en esta encuesta, ya que los intervalos incluyen valores donde cualquiera de los dos podría obtener la mayoría.

2. (24%) A partir de la siguiente tabla de contingencia entre sexo e intención de voto, donde cada celda reporta los recuentos observados y, entre paréntesis, los recuentos esperados bajo independencia, aplique el test Chi-cuadrado de independencia para evaluar si existe asociación entre ambas variables.

Sexo / Voto	Kast	Jara	Total
Hombre	735 (733)	675 (677)	1410
Mujer	825 (827)	765 (763)	1590
Total	1560	1440	3000

Para realizar el test Chi-cuadrado de independencia planteo las hipótesis:

H_0: Sexo e intención de voto son independientes

H_1: Sexo e intención de voto están asociados

El estadístico Chi-cuadrado se calcula como:

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Calculando para cada celda:

Para Hombre-Kast: $\frac{(735-733)^2}{733} = \frac{4}{733} = 0.00546$

Para Hombre-Jara: $\frac{(675-677)^2}{677} = \frac{4}{677} = 0.00591$

Para Mujer-Kast: $\frac{(825-827)^2}{827} = \frac{4}{827} = 0.00484$

Para Mujer-Jara: $\frac{(765-763)^2}{763} = \frac{4}{763} = 0.00524$

Sumando todos los términos:

$$\chi^2_{obs} = 0.00546 + 0.00591 + 0.00484 + 0.00524 = 0.02145$$

Los grados de libertad son:

$$gl = (filas - 1)(columnas - 1) = (2-1)(2-1) = 1$$

Bajo la hipótesis nula de independencia, el estadístico χ^2_{obs} sigue una distribución Chi-cuadrado con 1 grado de libertad. El valor crítico para $\alpha = 0.01$ con 1 grado de libertad es $\chi^2_{crtico} = 6.635$.

Como $\chi^2_{obs} = 0.02145 < 6.635$, no rechazo la hipótesis nula. El valor p asociado a este estadístico es extremadamente alto (cercano a 0.88), lo que indica que si las variables fueran independientes, obtener un estadístico como el observado o más extremo sería muy probable.

Conclusión: No existe evidencia estadística significativa al 99% de confianza para afirmar que el sexo y la intención de voto están asociados. Las diferencias observadas entre hombres y mujeres en sus preferencias electorales son tan pequeñas que pueden atribuirse al azar del muestreo. El sexo no parece ser un factor relevante en la elección entre Kast y Jara en esta muestra.

3. (20%) Use el test de diferencia de proporciones para evaluar si la proporción de votantes por Kast es significativamente distinta entre hombres y mujeres.

Primero calculo las proporciones muestrales de cada grupo:

Para hombres: $\hat{p}_H = \frac{735}{1410} = 0.5213$

Para mujeres: $\hat{p}_M = \frac{825}{1590} = 0.5189$

Planteo las hipótesis:

H_0: $p_H = p_M$ (las proporciones son iguales)

H_1: $p_H \neq p_M$ (las proporciones son diferentes)

El error estándar de la diferencia se calcula directamente como:

$$EE = \sqrt{\frac{\hat{p}_H(1-\hat{p}_H)}{n_H} + \frac{\hat{p}_M(1-\hat{p}_M)}{n_M}}$$

$$EE = \sqrt{\frac{0.5213 \cdot 0.4787}{1410} + \frac{0.5189 \cdot 0.4811}{1590}}$$

$$EE = \sqrt{\frac{0.2496}{1410} + \frac{0.2496}{1590}}$$

$$EE = \sqrt{0.0001770 + 0.0001570} = \sqrt{0.0003340} = 0.01828$$

El estadístico de prueba es:

$$z = \frac{\widehat{p}_H - \widehat{p}_M}{EE} = \frac{0.5213 - 0.5189}{0.01828} = \frac{0.0024}{0.01828} = 0.1313$$

Bajo la hipótesis nula de igualdad de proporciones, el estadístico z sigue una distribución Normal estándar $N(0,1)$. Para un test bilateral con $\alpha = 0.01$, el valor crítico es $z_{crtico} = \pm 2.576$.

Como $|z_{obs}| = 0.1313 < 2.576$, no rechazo la hipótesis nula. El valor p asociado a este estadístico es aproximadamente 0.90, lo que significa que si las proporciones poblacionales fueran iguales, obtener una diferencia como la observada o más extrema sería muy probable.

Conclusión: No hay evidencia estadística significativa al 99% de confianza para afirmar que la proporción de votantes por Kast difiere entre hombres y mujeres. La diferencia observada de apenas 0.24 puntos porcentuales es tan pequeña que puede explicarse fácilmente por variación muestral aleatoria. Este resultado es consistente con el test Chi-cuadrado realizado anteriormente.

4. (20%) Usando la matriz de varianza-covarianza y estadísticos provistos, calcule el coeficiente de correlación de Pearson entre ingreso y años de escolaridad.

- Explique qué mide.

- Interprete su valor, signo e intensidad.

El coeficiente de correlación de Pearson se calcula como:

$$r = \frac{Cov(X,Y)}{\sigma_X \cdot \sigma_Y}$$

De la matriz de varianza-covarianza tenemos:

$Var(Ingreso) = 72900$, por lo tanto $\sigma_{Ingreso} = \sqrt{72900} = 270$

$Var(Escolaridad) = 6.76$, por lo tanto $\sigma_{Escolaridad} = \sqrt{6.76} = 2.6$

$Cov(Ingreso, Escolaridad) = 312.4$

Calculando la correlación:

$$r = \frac{312.4}{270 \cdot 2.6} = \frac{312.4}{702} = 0.445$$

El coeficiente de correlación de Pearson mide la fuerza y dirección de la relación lineal entre dos variables cuantitativas. Toma valores entre -1 y 1, donde valores cercanos a 1 indican una relación lineal positiva fuerte, valores cercanos a -1 indican una relación lineal negativa fuerte, y valores cercanos a 0 indican ausencia de relación lineal.

En este caso, $r = 0.445$ indica una correlación positiva moderada entre ingreso y años de escolaridad. El signo positivo significa que a mayor escolaridad tiende a haber mayores ingresos. La magnitud de 0.445 sugiere que la relación es moderada, es decir, existe una tendencia clara pero no perfecta. Aproximadamente el 19.8% de la variabilidad en los ingresos puede explicarse por la escolaridad ($r^2 = 0.198$), lo que implica que aunque la educación es un factor relevante para explicar los ingresos, existen muchos otros factores que también influyen en esta variable.

5. (16%) Si X es una variable aleatoria e $Y = \frac{1}{2} \cdot X$, ¿cuál es la correlación de Pearson entre X e Y?

Explique el razonamiento y el significado del resultado.

Para encontrar la correlación entre X e $Y = \frac{1}{2}X$ utilizo la fórmula:

$$r_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \cdot \sigma_Y}$$

Primero calculo la covarianza. Por propiedades de la covarianza:

$$Cov(X,Y) = Cov\left(X, \frac{1}{2}X\right) = \frac{1}{2}Cov(X,X) = \frac{1}{2}Var(X) = \frac{1}{2}\sigma_X^2$$

Ahora calculo la desviación estándar de Y. Por propiedades de la varianza:

$$Var(Y) = Var\left(\frac{1}{2}X\right) = \left(\frac{1}{2}\right)^2 Var(X) = \frac{1}{4}\sigma_X^2$$

Por lo tanto:

$$\sigma_Y = \sqrt{\frac{1}{4}\sigma_X^2} = \frac{1}{2}\sigma_X$$

Sustituyendo en la fórmula de correlación:

$$r_{X,Y} = \frac{\frac{1}{2}\sigma_X^2}{\sigma_X \cdot \frac{1}{2}\sigma_X} = \frac{\frac{1}{2}\sigma_X^2}{\frac{1}{2}\sigma_X^2} = 1$$

El resultado es $r_{X,Y} = 1$, lo que indica una correlación perfecta positiva.

Este resultado tiene sentido porque Y es una transformación lineal positiva de X. Cuando una variable es una función lineal de otra con pendiente positiva, la correlación siempre es 1, independientemente de la constante multiplicativa o aditiva. El coeficiente de correlación es invariante ante cambios de escala y localización, por lo que solo captura la dirección y fuerza de la relación lineal, no la magnitud de los cambios. En este caso, cada vez que X aumenta, Y aumenta proporcionalmente, manteniendo una relación lineal perfecta, aunque Y sea exactamente la mitad de X.