Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística

Temporada Académica de Verano (enero 2025)

Curso : Probabilidad y Estadística
Sigla : EYP1113
Profesores : Hernán Robledo A. - Felipe Ossa M.

PAUTA INTERROGACIÓN 3

Pregunta 1 (6 puntos)

Sea $X_1, \ldots, X_n$ una muestra independiente e idénticamente distribuida que sigue una distribución cuya función de densidad está dada por

$$f(x) = \begin{cases} \dfrac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \exp\left(-\dfrac{x}{\beta}\right), & \text{si } x > 0, \\ 0, & \text{si } x \leq 0, \end{cases}$$

donde $\beta > 0$. Asuma que $\alpha = 2$ (conocido), que $\sum_{i=1}^n x_i = 23.5$, y que $n = 10$.

a) (3 ptos) Basado en los datos, calcule la estimación de máxima verosimilitud de β.

b) (3 ptos) Calcule la estimación de la varianza del estimador de máxima verosimilitud de β.

(**Ayuda**: $\mathsf{E}(X) = \alpha\beta$ y $\mathsf{Var}(X) = \alpha\beta^2$)

Solución:

a) Se debe obtener el estimador de máxima verosimilitud (EMV) para β.

$$L(\beta) = \prod_{i=1}^n f(x_i) = \prod_{i=1}^n \frac{1}{\Gamma(\alpha)\beta^\alpha} x_i^{\alpha-1} \exp\left(-\frac{x_i}{\beta}\right) = \frac{1}{\Gamma(\alpha)^n \beta^{n\alpha}} \left(\prod_{i=1}^n x_i\right)^{\alpha-1} \exp\left(-\frac{1}{\beta}\sum_{i=1}^n x_i\right)$$

$$\log L(\beta) = -n\log(\Gamma(\alpha)) - n\alpha\log(\beta) + (\alpha-1)\sum_{i=1}^n \log(x_i) - \frac{1}{\beta}\sum_{i=1}^n x_i$$

Derivamos respecto a β e igualamos a cero para obtener el EMV.

$$\frac{\partial}{\partial\beta}\log L(\beta) = \frac{-n\alpha}{\beta} + \frac{1}{\beta^2}\sum_{i=1}^n x_i = 0 \quad \Longrightarrow \quad \hat{\beta} = \frac{1}{\alpha n}\sum_{i=1}^n x_i = \frac{\bar{x}}{\alpha}$$

Como $\alpha = 2$ y $\bar{x} = 23.5/10 = 2.35$, la estimación de máxima verosimilitud de β es $\hat{\beta} = 2.35/2 = 1.175$.

b) Debemos calcular la varianza de $\hat{\beta}$.

$$\mathsf{Var}(\hat{\beta}) = \mathsf{Var}\left(\frac{\bar{x}}{\alpha}\right) = \frac{1}{\alpha^2 n^2}\mathsf{Var}\left(\sum_{i=1}^n x_i\right) = \frac{1}{\alpha^2 n^2}\sum_{i=1}^n \mathsf{Var}\left(x_i\right) = \frac{1}{\alpha^2 n^2}\sum_{i=1}^n \alpha\beta^2 = \frac{\beta^2}{\alpha n}.$$

Como alternativa, se acepta en este caso la varianza asintótica,

$$\text{Var}(\hat{\beta}) = \frac{1}{I_n(\beta)}$$

donde

$$\begin{aligned} I_n(\beta) &= -E\left[\frac{d^2}{d\beta^2}\ln L(\beta)\right] \\ &= -E\left[\frac{d}{d\beta}\frac{-n\alpha}{\beta} + \frac{1}{\beta^2}\sum_{i=1}^{n} X_i\right] \\ &= -E\left[\frac{n\alpha}{\beta^2} - \frac{2}{\beta^3}\sum_{i=1}^{n} X_i\right] \\ &= -\frac{n\alpha}{\beta^2} + \frac{2}{\beta^3}\sum_{i=1}^{n} E(X_i) \\ &= -\frac{n\alpha}{\beta^2} + \frac{2}{\beta^3}\cdot n\alpha\beta \\ &= \frac{n\alpha}{\beta^2} \end{aligned}$$

La estimación de la varianza del EMV se obtiene reemplazando $\hat{\beta}$ en la fórmula de la varianza. Así,

$$\text{Var}(\hat{\beta}) = \frac{1.175^2}{2\cdot 10} = 0.06903125$$

Asignación de Puntaje:

a)
Logro 1 [1 pto]: Plantea correctamente función de verosimilitud $L(\beta)$ o de log verosimilitud $\log L(\beta)$.
Logro 2 [1 pto]: por obtener correctamente la derivada de la log verosimilitud.
Logro 3 [1 pto]: por igualar a cero la derivada de la log verosimilitud [0.3 pto] y obtener correctamente el estimador $\hat{\beta}$ [0.5 pto] y su estimación $\hat{\beta} = 1.175$ [0.2 pto].

b)
Logro 4 [1.0 pto]: por aplicar correctamente propiedades de varianza que los términos constantes salen al cuadrado.
Logro 5 [1.5 pto]: por aplicar correctamente propiedades de varianza que varianza de suma de variables independientes es la suma de las varianzas.
Logro 6 [0.5 pto]: por obtener la fórmula de la varianza $\text{Var}(\hat{\beta}) = \frac{\beta^2}{\alpha n}$ [0.3 pto] y su estimación $\text{Var}(\hat{\beta}) = 0.06901325$ [0.2 pto].

Pregunta 2 (6 puntos)

Sea $X_1, \ldots, X_n$ una muestra independiente e idénticamente distribuida de una distribución Binomial Negativa con parámetro $r = 5$ y parámetro de probabilidad p desconocido. En un experimento se observó que $\bar{x} = 12.6$ para $n = 25$ observaciones. Se sabe que el estimador de máxima verosimilitud para p y su varianza asintótica están dadas por:

$$\hat{p} = \frac{r}{\bar{x}} = \frac{r \cdot n}{\sum_{i=1}^{n} x_i}, \quad \text{Var}(\hat{p}) = \frac{1}{I_n(p)} = \frac{p^2(1-p)}{r \cdot n}.$$

A partir de la muestra, determine un intervalo de confianza asintótico para p al 99 % de confianza.

Solución:

Por propiedad de los estimadores de máxima verosimilitud, se tiene que

$$\frac{\hat{p} - p}{\sqrt{\text{Var}(\hat{p})}} \overset{aprox.}{\sim} \text{Normal}(0, 1).$$

Así, podemos obtener el intervalo al $(1-\alpha)100\,\%$ de confianza como

$$\begin{aligned} < p >_{99\,\%} &= \left(\hat{p} \pm k_{1-\alpha/2}\sqrt{\text{Var}(\hat{p})}\right) \\ &= \left(\hat{p} \pm k_{1-\alpha/2}\sqrt{\frac{\hat{p}^2(1-\hat{p})}{r \cdot n}}\right) \end{aligned}$$

Una confianza del 99 % implica que $\alpha = 1\,\%$. Así, $k_{1-0.01/2} = k_{0.995} = \texttt{qnorm(0.995)} = 2.575$. Finalmente se obtiene el intervalo de confianza reemplazando todos los valores.

$$< p >_{99\,\%} = (0.3258,\ 0.4678)$$

Asignación de Puntaje:

Alternativa 1:
Logro 1 [4 pto]: por plantear que $(\hat{p} - p)/\sqrt{\text{Var}(\hat{p})} \sim \text{Normal}(0, 1)$ [1 ptos] y utilizar este resultado para derivar la fórmula del intervalo de confianza para p [3 ptos].
Logro 2 [1.5 pto]: por obtener $\hat{p}$, $\text{Var}(\hat{p})$, y $k_{1-\alpha/2}$.

- [0.5 ptos] por $\hat{p} = 0.3968$
- [0.5 ptos] por $\text{Var}(\hat{p}) = 0.02756$
- [0.5 ptos] por $k_{1-\alpha/2} = k_{0.995} = 2.575829$.

Logro 3 [0.5 ptos]: por obtener que $< p >_{99\,\%} = (0.3258,\ 0.4678)$.

Alternativa 2 (sólo cambia Logro 1):
Logro 1 [4.0 pto]: por plantear fórmula de un intervalo de confianza: $p \in \left(\hat{p} \pm k_{1-\alpha/2}\sqrt{\text{Var}(\hat{p})}\right)$
Logro 2 [1.5 pto]: por obtener $\hat{p}$, $\text{Var}(\hat{p})$, y $k_{1-\alpha/2}$.

- [0.5 ptos] por $\hat{p} = 0.3968$
- [0.5 ptos] por $\text{Var}(\hat{p}) = 0.02756$
- [0.5 ptos] por $k_{1-\alpha/2} = k_{0.995} = 2.575829$.

Logro 3 [0.5 ptos]: por obtener que $< p >_{99\,\%} = (0.3258,\ 0.4678)$.

Pregunta 3 (6 puntos)

Enero es el mes en que la ciudad de Santiago registra las temperaturas más altas del año. Además, debido al calentamiento global, las personas perciben que cada enero es más caluroso que los anteriores. Para evaluar esta percepción, se midió la temperatura (en grados Celsius), en distintos puntos de la ciudad al mediodía y en distintos días de enero, obteniendo en total $n = 280$ registros. Con la muestra obtenida se calcularon las siguientes estadísticas:

Promedio	Mediana	Desviación Estándar	Coef. de Asimetría	Coef. de Variación
31.05	30.75	1.41	-0.09	0.045

Asuma que las mediciones de temperatura siguen una distribución normal. ¿Hay evidencia estadística para concluir que el promedio de temperaturas de este mes supera los 30.84 grados (el promedio de la década de los 80)? Use un nivel de significancia del 1 % e indique claramente las hipótesis, el estadístico de prueba, el criterio de decisión (valor-p o valor crítico) y la conclusión en el contexto del problema.

Ayuda: recuerde que en la hipótesis alternativa se indica aquello que se busca evaluar.

Solución:

Asumiendo las mediciones de temperatura siguen una distribución normal con media μ y desviación σ, se pide evaluar la hipótesis

$$H_0 : \mu = 30.84 \quad vs \quad H_a : \mu > 30.84$$

Para evaluar la hipótesis, se utiliza el estadístico de prueba

$$T_0 = \frac{\bar{x} - \mu_0}{S/\sqrt{n}} \sim t_{n-1},$$

donde $\bar{x} = 31.05$, $S = 1.41$, y $n = 280$. Se utiliza el estadístico con distribución t-student ya que se desconoce la varianza de la población, y porque la cantidad S es estimada a través de la muestra registrada.
Al calcular T_0 se tiene

$$T_0 = \frac{31.05 - 30.84}{1.41/\sqrt{280}} = 2.492179$$

Para evaluar la hipótesis nula se puede considerar cualquiera de los siguientes casos:

- A una significancia de $\alpha = 1\,\%$, se rechaza H_0 si $T_0 > t_{(1-\alpha,n-1)} = t_{(0.99,279)} = 2.339788$. Como $T_0 = 2.492179$, existe evidencia para rechazar H_0, por lo que $\mu > 30.84$, y se puede concluir que el promedio de temperaturas de este mes supera los 30.84 grados.
- Se calcula el valor-p como valor-p $= \mathrm{P}(T_{n-1} > T_0) = \mathrm{P}(T_{279} > 2.492179) = 1 - \texttt{pt(2.492179,279)} = 0.0066$. Como el valor-p es menor a la significancia de $1\,\% = 0.01$, entonces existe evidencia para rechazar H_0, por lo que $\mu > 30.84$, y se puede concluir que el promedio de temperaturas de este mes supera los 30.84 grados.

De manera alternativa, el estudiante puede utilizar la tabla de cuantiles t-student del formulario, pero notar que no se dispone de la t-student con 279 grados de libertad, por lo que se revisa la tabla de cuantiles normales. En ese caso, se tiene que se rechaza H_0 si $T_0 > z_{0.99} = 2.326$, o si nota que el valor-p es igual a $\mathrm{P}(Z > 2.492179) = 0.0063$, rechazando H_0 por ser menor a $\alpha = 1\,\%$.

Asignación de Puntaje:

Logro 1 [1 pto]: por plantear correctamente la hipótesis de interés. $H_0 : \mu = 30.84$ vs $H_a : \mu > 30.84$ o $H_0 : \mu < 30.84$ vs $H_a : \mu > 30.84$
Logro 2 [1.5 ptos]: por plantear o utilizar el estadístico de prueba correcto. $T_0 = (\bar{x} - \mu_0)/(S/\sqrt{n}) \sim t_{n-1}$.
Logro 3 [1 pto]: por obtener que $T_0 = 2.492179$.
Logro 4 [1.5 ptos]: por calcular valor-p $= 0.0066$ o por obtener $t_{(0.99,279} = 2.339788$.
Logro 5 [1 pto]: por tomar la decisión de acuerdo al contexto del problema en base a los resultados. A un $\alpha = 0.01$, se rechaza H_0. Descontar [0.5 pto] si no da conclusión en base a contexto del problema.

Observación:

- Si el estudiante plantea incorrectamente el estadístico de prueba T_0, y calcula todo el resto correctamente, no asignar el **Logro 2** pero sí asignar los Logros 3, 4, y 5 si corresponden.

Pregunta 4

Considere dos variables aleatorias contínuas X e Y, independientes, donde $E(X) = \mu_X$, $E(Y) = \mu_Y$, y Var$(X) =$ Var$(Y) = 4$ (conocidas). Los parámetros μ_X y μ_Y son desconocidos.

Se desea contrastar las siguientes hipótesis

$$H_0 \ : \ \mu_X = \mu_Y$$
$$H_a \ : \ \mu_X < \mu_Y$$

a través de un test utilizando muestras $X_1, \ldots, X_n$ y $Y_1, \ldots, Y_n$ independientes, de X e Y respectivamente, ambas de tamaño $n = 32$. Para evaluar las hipótesis, se propone una regla (región crítica) que consiste en rechazar la hipótesis nula H_0, si

$$\overline{X} - \overline{Y} < -0.5$$

donde $\overline{X}$ e $\overline{Y}$ son los promedios de las muestras respectivas.

a) (3 puntos) Encuentre la significancia *aproximada* del test, es decir, la probabilidad *aproximada* de cometer un Error Tipo I.

b) (3 puntos) Asumiendo el caso que, en realidad, $\mu_X = 4$ y $\mu_Y = 4.8$, calcule la probabilidad *aproximada* de cometer un Error Tipo II.

(**Ayuda I:** Aproximar las probabilidades utilizando el Teorema del Límite Central)

(**Ayuda II:** $P($Error Tipo II$) = P($No Rechazar $H_0 \mid H_a$ es verdadera$)$)

Solución:

a) La probabilidad solicitada corresponde a

$$P(\overline{X} - \overline{Y} < -0.5)$$

donde $\mu_X - \mu_Y = 0$. Gracias al Teorema del Límite Central

$$\overline{X} - \overline{Y} \overset{\text{aprox.}}{\sim} \text{Normal}\left(\mu_X - \mu_Y, \sqrt{\frac{4}{n} + \frac{4}{n}}\right)$$

Bajo la hipótesis nula, se tiene que $\overline{X} - \overline{Y} \overset{\text{aprox.}}{\sim}$ Normal$(0; 0.5)$. Por lo tanto,

$$P(\text{Error Tipo I}) = P(\overline{X} - \overline{Y} < -0.5) \approx \Phi\left(\frac{-0.5 - 0}{0.5}\right) = \Phi(-1) = 1 - 0.8413 = 0.1587$$

b) La probabilidad solicitada corresponde a

$$P(\overline{X} - \overline{Y} \geq -0.5)$$

donde $\mu_X - \mu_Y = -0.8$. Utilizando nuevamente el teorema del límite central,

$$\overline{X} - \overline{Y} \overset{\text{aprox.}}{\sim} \text{Normal}(-0.8; 0.5)$$

Por lo tanto,

$$\begin{aligned} P(\text{Error Tipo II} \mid \mu_X - \mu_Y = -0.8) &= P(\overline{X} - \overline{Y} \geq -0.5) \\ &\approx 1 - \Phi\left(\frac{-0.5 - (-0.8)}{0.5}\right) \\ &= 1 - \Phi(0.6) \\ &= 1 - 0.7257 \\ &= 0.2743 \end{aligned}$$

Asignación de Puntaje:

a)
Logro 1 [0.5 pto]: por plantear la probabilidad solicitada $P(\overline{X} - \overline{Y} < -0.5)$.
Logro 2 [2 pto]: por indicar por Teorema del Límite Central que $\overline{X} - \overline{Y}$ sigue una distribución normal [0.5 pto] con media cero bajo H_0 [1 pto] y con desviación estándar $\sqrt{8/n} = 0.5$ o varianza $8/n = 0.25$ [0.5 pto].
Logro 3 [0.5 pto]: por indicar cómo calcular una probabilidad bajo una distribución normal [0.3 pto] y obtener que $P(\text{Error Tipo I}) = 0.1587$.

b)
Logro 4 [0.5 pto]: por plantear la probabilidad solicitada $P(\overline{X} - \overline{Y} \geq -0.5)$.
Logro 5 [2 pto]: por indicar por Teorema del Límite Central que $\overline{X} - \overline{Y}$ sigue una distribución normal [0.5 pto] con media -0.8 bajo H_a [1 pto] y con desviación estándar $\sqrt{8/n} = 0.5$ o varianza $8/n = 0.25$ [0.5 pto].
Logro 6 [0.5 pto]: por indicar cómo calcular una probabilidad bajo una distribución normal [0.3 pto] y obtener que $P(\text{Error Tipo II}) = 0.2743$ [0.2 pto].

Pregunta 5 (6 puntos)

En plena temporada de cosecha de duraznos, una productora agrícola registró, durante $n = 12$ días, la cantidad de toneladas cosechadas de duraznos por día trabajado (Y_i), como también la temperatura media diaria (X_i) en °C. Se modelará la cosecha media diaria en función de la temperatura a través de un modelo de regresión lineal, el cual establece que

$$E(Y \mid X = x) = \beta_0 + \beta_1 \cdot x$$

donde β_0 y β_1 son parámetros desconocidos.

El resumen de los datos fue el siguiente,

$$n = 12\ , \qquad \overline{x} = 26.78\ , \qquad \overline{y} = 2.34\ , \qquad \sum_{i=1}^{n} x_i^2 = 8653.56$$

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 = 45.4\ , \qquad \sum_{i=1}^{n} (y_i - \bar{y})^2 = 2.77\ ,$$

y la recta de regresión estimada de los datos derivó en la ecuación

$$y = \hat{\beta}_0 + \hat{\beta}_1 \cdot x = 0.6022 + 0.065 \cdot x$$

A partir de estos datos, ¿existe evidencia suficiente para afirmar que la pendiente de regresión β_1 es distinta de cero? Indique hipótesis a contrastar, valor del estadístico, valor crítico (o alternativamente valor-p) y la conclusión en base al contexto. Utilice una significancia de $\alpha = 5\,\%$.

(**Ayuda:** además del formulario, apoyarse en la fórmula auxiliar para $s^2_{Y|x}$,

$$s^2_{Y|x} = \frac{1}{n-2} \left[\sum_{i=1}^{n} (y_i - \bar{y})^2 - (\hat{\beta}_1)^2 \sum_{i=1}^{n} (x_i - \bar{x})^2 \right]$$

)

Solución:

Se debe contrastar las hipótesis

$$H_0 \ : \ \beta_1 = 0 \qquad\qquad H_a \ : \ \beta_1 \neq 0$$

con lo cual se utiliza el estadístico asociado,

$$T_1 = \frac{\hat{\beta}_1}{\left(\frac{s_{Y|x}}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}} \right)}$$

que bajo la hipótesis nula tiene distribución t-Student(10). Entonces se rechazaría H_0 si $|T_1| > t_{0.975,10} = 2.228$. De los datos obtenemos

$$s^2_{Y|x} = \frac{1}{10} \left[2.77 - (0.065)^2 \cdot 45.4 \right] = 0.2578$$

Por lo tanto,

$$T_1 = \frac{0.065}{\sqrt{\frac{0.2578}{45.4}}} = 0.8626\ ,$$

y vemos que $|0.8626| \not> 2.228$.

Alternativamente, puede calcularse el valor-p de T_1 como

$$\text{valor-p} = 2 \cdot (1 - P(T > 0.8626)) = 2(1 - \texttt{pt(0.8626,10)}) = 0.4086.$$

o bien concluir que valor-p $> 20\,\%$ viendo la tabla en formulario. Claramente vemos que el valor-p obtenido es mayor a α.

Entonces con un 5 % de significancia, no existe suficiente evidencia estadística para afirmar que la pendiente de regresión es significativamente distinta de cero.

Asignación de Puntaje:

Logro 1 [0.5 ptos]: por calcular $s^2_{Y|x} = 0.2578$.
Logro 2 [0.5 ptos]: por plantear correctamente la hipótesis de interés. $H_0 : \beta_1 = 0$ vs $H_a : \beta_1 \neq 0$.
Logro 3 [1.5 ptos]: por plantear o utilizar el estadístico de prueba $T_1 = (\hat{\beta}_1 - \beta_1)/s_{\hat{\beta}_1} \sim t_{n-2}$. Si no muestra indicios de que reemplaza $\beta_1 = 0$ en T_1 descontar [0.2 ptos].
Logro 4 [1 pto]: por obtener que $T_1 = 0.8626$.
Logro 5 [1.5 ptos]: por calcular valor-p $= 0.4085679$ o por obtener $t_{0.975,10} = 2.228$.
Logro 6 [1 pto]: por tomar la decisión de acuerdo al contexto del problema en base a los resultados. A un $\alpha = 0.05$, no se rechaza H_0. Descontar [0.5 pto] si no da conclusión en base a contexto del problema.

Observaciones:

- Si $s^2_{Y|x}$ es calculado incorrectamente sólo descontar **Logro 1** y asignar todo el resto de logros si corresponde.
- Si el estudiante plantea incorrectamente el estadístico de prueba T_1, y calcula todo el resto correctamente, no asignar el **Logro 3** pero sí asignar los Logros 4, 5, y 6 si corresponden.