PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
FACULTAD DE MATEMÁTICAS
Departamento de Estadística
EYP1113 – Probabilidades y Estadística
Segundo semestre 2024

Ayudantes: Michael Ramón: maramon@uc.cl
Juan Merino: jcamilo.merino@uc.cl
Diego Pérez: dtperez1@uc.cl
Christoph Bürger: christoph.buerger@uc.cl
Oscar Guerrero: oscar.guerrero@uc.cl

Ayudantía 10

Test de Hipótesis

Problema 1 *Test de Hipótesis I*

En TAV es común que los y las estudiantes estudien en grupos. Usted busca comprobar que más del 40 % de los estudiantes lo haces. Por otra parte, se dice respecto a los estudiantes *que estudian en forma solitaria*, que estos obtienen una nota media inferior a 5,0.

Con objetivo de comprobar o refutar sus apreciaciones (hipótesis), aplica una encuesta a los $n = 113$ alumnos del curso TAV 2024, obteniendo los siguientes datos muestrales:

Característica	Respuesta
Número de encuestados	113
¿Estudia solo?, Si	58
De los que estudian solos:	
Nota promedio	4.76
Desv. estándar	0.7

Lleve a cabo las dos pruebas de hipótesis correspondientes, use un nivel de significancia $\alpha = 1\,\%$ cada uno. Sea explícito:

- Señale la hipótesis a utilizar.
- Señale el estadístico de prueba a usar.
- Determine el criterio para concluir (valor-p o valor crítico).
- Finalice con una conclusión en el contexto del problema.

Suponga que la nota de los alumnos que estudian solo distribuye Normal.

Solución:

Hipótesis 1: Más del 40 % de los estudiantes estudian en grupo.

Debido a que se está tratando de porcentajes y no se menciona que los estudiantes que estudian en grupo están dentro de la Normalidad, entonces se tiene una prueba de proporciones. Sea p la verdadera proporción de estudiantes que estudian en grupo. Las hipótesis a contrastar son:

$$H_0 : p = 0{,}4 \qquad \text{vs} \qquad H_a : p > 0{,}4$$

El estadístico de prueba, bajo H_0, es:

$$Z_0 = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0\,(1 - p_0)}{n}}} \overset{\cdot}{\sim} \text{Normal}(0, 1)$$

, donde:

$$\hat{p} = \frac{\#\text{ Alumnos que estudian en grupo}}{\#\text{ Alumnos en total}} = \frac{113 - 58}{113} = 0{,}487$$
$$n = 113$$
$$p_0 = 0{,}4$$

Evaluando el estadístico de prueba se tiene:

$$Z_0 = 1{,}888$$

- *Criterio de valor-p*: El valor-p de Z_0 es:

  $$\text{valor-p} = P(Z > 1{,}888) = 1 - \Phi(1{,}89) = 0{,}0294$$

  Debido a que valor-p $> \alpha$, entonces no se rechaza H_0, o bien no existe evidencia estadística suficiente para afirmar que más que el 40 % de los estudiantes estudia en grupos, con 1 % de significancia.

- *Criterio de valor crítico*: El valor crítico debido a la hipótesis alternativa es:

  $$k_{1-\alpha} = k_{0{,}99} = 2{,}33$$

  Debido a que $Z_0 < k_{0{,}99}$, no se rechaza H_0, o bien no existe evidencia estadística suficiente para afirmar que más que el 40 % de los estudiantes estudia en grupos, con 1 % de significancia.

Alternativa: Se pudo haber realizado el test de hipótesis si se utilizaba la hipótesis de menos del 60 % de los estudiantes estudia en solitario y el desarrollo es consistente.

Hipótesis 2: Los estudiantes que estudian en forma solitaria tienen notas promedio inferior a 5,0.

Se sabe que las notas de los alumnos que estudian en solitario distribuyen Normal, no se tienen los valores poblacionales (μ, σ), sino los datos muestrales ($\overline{X}_n$, S_n). Las hipótesis a contrastar son:

$$H_0 : \mu = 5 \qquad \text{vs} \qquad H_a : \mu < 5$$

, debido a que no se conoce la desviación estándar poblacional σ, entonces, el estadístico de prueba, bajo H_0, es:

$$T_0 = \frac{\overline{X}_n - \mu_0}{S/\sqrt{n}} \sim \text{t-Student}(n-1)$$

, donde:

$$\overline{X}_{58} = 4{,}76 \qquad S_{58} = 0{,}7 \qquad n = 58$$

Debido a que $n > 30$, es posible aproximar la distribución t–Student a la Normal estándar.

Evaluando el estadístico de prueba se obtiene:

$$T_0 = -2{,}61$$

- *Criterio de valor-p*: El valor-p de T_0 aproximado es:

  $$\text{valor-p} = P(Z < -2{,}61) = 1 - \Phi(2{,}61) = 0{,}0045$$

  entonces, como valor-p $< \alpha$, se rechaza H_0 o bien existe evidencia suficiente para afirmar que, de los estudiantes que estudian en solitario, la nota media es menor a 5, con un 1 % de significancia.

- *Criterio de valor crítico*: El valor crítico debido a la hipótesis alternativa es

$$t_\alpha(n-1) = t_{0{,}01}(57) = -t_{0{,}99}(57) \approx -k_{0{,}99} = -2{,}33$$

, como $T_0 < t_{0{,}01}(57)$, se rechaza H_0 o bien existe evidencia suficiente para afirmar que, de los estudiantes que estudian en solitario, la nota media es menor a 5, con un 1 % de significancia.

Problema 2 *Test de hipótesis II*

Un estudio respecto a las horas de estudio diario en TAV, según el género, es llevado a cabo. Parte de los resultados muestrales se presentan en la siguiente tabla (asuma que los tiempos son independientes y se comportan de acuerdo con una distribución Normal):

Característica	Masculino	Femenino	Total
Número de casos	16	14	30
Tiempo: (en min)			
Promedio	46	52	49
Desv. Estándar	13	8	11

a) Considerando la muestra completa, ¿existe evidencia que permita afirmar que la variabilidad (en términos de la desviación estándar) de las horas de estudio es inferior a 15 min? Use $\alpha = 5\,\%$.

b) ¿Existe evidencia que permita afirmar qu el género Femenino dedica un mayor tiempo medio al estudio que sus congéneres Masculinos? Use $\alpha = 10\,\%$ y asuma varianzas iguales.

Solución:

Solución a)

Las hipótesis a contrastar son:

$$H_0 : \sigma = 15 \qquad \text{vs} \qquad H_a : \sigma < 15$$

Debido a que los parámetros poblacionales de la distribución Normal no se conocen (i.e. μ desconocida), el estadístico de prueba a utilizar, bajo H_0, es:

$$C_0 = \frac{(n-1)\,S^2}{\sigma_0^2} \sim \chi^2(n-1)$$

, donde:

$$S = 11 \qquad \sigma_0 = 15 \qquad n = 30$$

Evaluando el estadístico de prueba se tiene:

$$C_0 = 15{,}6$$

- *Criterio de valor-p*: El valor-p de C_0 es (utilizando la tabla chi-cuadrado):

$$\begin{gathered} c_{0{,}005}(29) < C_0 < c_{0{,}025}(29) \\ 13{,}121 < 15{,}6 < 16{,}047 \\ P(C < 13{,}121) < P(C < 15{,}6) < P(C < 16{,}047) \\ 0{,}5\,\% < \text{valor-p} < 2{,}5\,\% \end{gathered}$$

  , como valor-p $< \alpha$, se concluye que se rechaza H_0, es decir, existe evidencia suficiente para afirmar que la desviación estándar de tiempo dedicado al estudio es menor a 15 minutos, con un 5 % de significancia.

- *Criterio de valor crítico*: El valor crítico en base a H_0 es:

$$c_\alpha(n-1) = c_{0{,}05}(29) = 17{,}708$$

  , como $C_0 < c_{0{,}05}(29)$, se concluye que se rechaza H_0, es decir, existe evidencia suficiente para afirmar que la desviación estándar de tiempo dedicado al estudio es menor a 15 minutos, con un 5 % de significancia.

Solución b)

Las hipótesis a contrastar son:

$$H_0 : \mu_F = \mu_M \qquad \text{vs} \qquad H_a : \mu_F > \mu_M$$

Este es un test de comparación de medias con varianzas poblacionales desconocidas pero iguales, por lo que el estadístico de prueba es:

$$T_0 = \frac{\overline{F} - \overline{M}}{S_p \sqrt{\dfrac{1}{n_F} + \dfrac{1}{n_M}}} \sim \text{t-Student}(n + m - 2)$$

, donde:

$$\overline{F} = 52 \qquad \overline{M} = 46 \qquad n_M = 16 \qquad n_F = 14 \qquad S_p = \sqrt{\frac{S_M^2\,(n_M - 1) + S_F^{(n_F-1)}}{n_F + n_M - 2}} = 10{,}966$$

Evaluando en el estadístico de prueba se tiene:

$$T_0 = 1{,}495$$

- *Criterio de valor-p*: El valor-p de T_0, en base a la tabla t-Student, es:

$$\begin{aligned} t_{0,90}(28) < T_0 < t_{0,95}(28) \\ 1{,}313 < 1{,}495 < 1{,}701 \\ P(T < 1{,}313) < P(T < 1{,}495) < P(T < 1{,}701) \\ 0{,}90 < P(T < 1{,}495) < 0{,}95 \\ 5\,\% < \text{valor-p} < 10\,\% \end{aligned}$$

  , como valor-p $< \alpha$, se rechaza H_0, se puede afirmar, con un 10 % de significancia, que el género femenino dedica en promedio más tiempo al estudio que el género masculino en este curso.

- *Criterio de valor crítico*: El valor crítico en base a H_a es:

$$t_{1-\alpha}(n_F + n_M - 2) = t_{0,90}(28) = 1{,}313$$

  , como $T_0 < t_{0,90}(28)$, se rechaza H_0, se puede afirmar, con un 10 % de significancia, que el género femenino dedica en promedio más tiempo al estudio que el género masculino en este curso.

Problema 3 *Estimadores*

Suponga que X es una variable aleatoria discreta con función de probabilidad dada por:

$$p_X(x) = \begin{cases} \theta^2 & x = -1 \\ 2\,\theta\,(1-\theta) & x = 0 \\ (1-\theta)^2 & x = +1 \end{cases}$$

con $\theta \in (0, 1)$.

a) A partir de la muestra aleatoria $x_1 = +1$, $x_2 = +1$, $x_3 = 0$, obtenga y evalúe el estimador de momentos de θ.

b) A partir de la muestra aleatoria $x_1 = -1$, $x_2 = +1\, x_3 = -1$, obtenga y evalúe el estimador máximo verosímil de θ.

Solución:

Solución a)

La esperanza de X es:

$$\mathrm{E}(X) = \sum_{i=1}^{3} x_i\, p_X(x_i) = (-1)(\theta^2) + (0)[2\,\theta\,(1-\theta)] + (1)(1-\theta)^2 = 1 - 2\,\theta$$

El método de estimación de momentos establece que:

$$\begin{aligned} \mathrm{E}(X) &= \overline{X} \\ 1 - 2\,\theta &= \overline{X} \\ \hat{\theta} &= \frac{1 - \overline{X}}{2} \end{aligned}$$

En base a la muestra aleatoria, se tiene que:

$$\overline{X} = \frac{1}{3} \sum_{i=1}^{3} x_i = \frac{1}{3}\,(+1 + 1 + 0) = \frac{2}{3}$$

, entonces:

$$\hat{\theta} = \frac{1}{6}$$

Solución b)

En base a la muestra aleatoria, la función de verosimilitud es:

$$\begin{aligned} L(\theta) &= p_X(-1)\, p_X(+1)\, p_X(-1) \\ &= \theta^2\,(1-\theta)^2\,\theta^2 \end{aligned}$$

La función de log–verosimilitud es:

$$\ln(L) = 4\,\ln(\theta) + 2\,\ln(1-\theta)$$

Derivando respecto a θ e igualando a cero:

$$\begin{aligned} \frac{\mathrm{d}\ \ln(L)}{\mathrm{d}\,\theta} &= 0 \\ \frac{4}{\theta} - \frac{2}{1-\theta} &= 0 \\ \tilde{\theta} &= \frac{2}{3} \end{aligned}$$