Departamento de Ingeniería Civil Industrial
Estadística para la Economía y Gestión - IN3401
Prof.: Constanza Contreras Piña
Aux's: Paz Montaño - Rubén Ortega - Camila J. Pulgar
Carolina Salgado - Florencia Signorini

INGENIERIA INDUSTRIAL
UNIVERSIDAD DE CHILE

Auxiliar Pre Control

Preparación Control 1

Parte 1: Comentes.

1. Obtener un p-valor de 0.01 siempre es mejor que obtener un p-valor de 0.05 porque el efecto encontrado es mucho mayor.
2. Cuando el resultado de un test de hipótesis nos da que NO es estadísticamente significativo aceptamos la hipótesis nula.
3. Si aumenta el poder estadístico, entonces es más probable que se rechace la hipótesis nula.
4. Teniendo una muestra grande, ésta va a distribuir normal gracias al Teorema Central del Límite.
5. Imagine que usted se ha planteado un test de hipótesis para determinar si existe una diferencia significativa en las notas obtenidas por los alumnos varones y por las alumnas del curso IN3401, más especificamente usted desea saber si con un 95% la media de notas de los hombres (μ_H) es mayor a la de las mujeres (μ_M). Dado que usted cuenta con una cantidad lo suficientemente grande, deberá plantearse un test de hipótesis considerando el valor del estadístico $z_\alpha = 1.96$

Solución

1. Obtener un p-valor de 0.01 siempre es mejor que obtener un p-valor de 0.05 porque el efecto encontrado es mucho mayor

   Falso , el p - valor representa la probabilidad de error de aceptar la H_A como cierta cuando en realidad no lo es , suponiendo que la hipótesis nula H0 es cierta, asociado al error tipo 1.

   A un p-valor de 0.01, son más sorprendentes los resultados bajo la H_0 que a un valor de 0.05 pero no nos dice nada sobre la magnitud del erecto

   - El valor de p no representa la probabilidad de que la hipótesis nula sea cierta: como hemos dicho, partimos del supuesto de que la hipótesis nula es cierta y es bajo ese supuesto en el que calculamos el valor de p.
   - Un p $<$ 0,05 significa que la hipótesis nula es falsa y una p $>$ 0,05 que la hipótesis nula es verdadera: siempre nos movemos en el terreno de la probabilidad.

     Un p $<$ 0,05 quiere simplemente decir que es poco probable que la H_0 sea cierta, luego la rechazamos para abrazar la alternativa, pero siempre tenemos cierta probabilidad de cometer lo que se denomina un error de tipo 1: rechazar la hipótesis nula cuando en realidad es verdadera.

Por otra parte, el valor de p > 0,05 no afirma que la H0 sea verdadera, ya que puede ocurrir que la diferencia sea real y el estudio no tenga potencia para detectarla. Estaremos ante el error de tipo 2: no rechazar la hipótesis de nulidad (y afirmar que no existe el efecto) cuando en realidad sí que existe en la población (pensad, por ejemplo, que el tamaño muestra no sea el suficiente). Así como podemos rechazar H0, nunca podemos afirmar lo contrario: H0 solo es falsable, nunca podemos afirmar que sea cierta.

2. Cuando el resultado de un test de hipótesis nos da que NO es estadísticamente significativo que aceptamos la hipótesis nula.

   Si el resultado de un test de hipótesis nos da que No es estadísticamente significativo , ya sea por que nuestro estadístico experimental está dentro de la región de aceptación o el p-valor obtenido es mayor al nivel de significancia con el cual se contrasta, se concluye que no se puede rechazar la hipótesis nula. Es decir, no tenemos pruebas suficientes para concluir que la hipótesis nula sea falsa, afirmando esto con (1-α)*100% de probabilidad de no equivocarnos.

3. Si aumenta el poder estadístico, entonces es más probable que se rechace la hipótesis nula.

   Verdadero dado que el poder estadístico es la probabilidad de que la hipótesis alternativa sea aceptada cuando la hipótesis alternativa es verdadera (es decir, la probabilidad de no cometer un error del tipo II), si aumentamos dicho poder disminuimos el error tipo 2.

   Que es lo mismo que aumente la probabilidad de rechazar la nula dado que es falsa.

4. Teniendo una muestra grande, ésta va a distribuir normal gracias al Teorema Central del Límite.

   Falso, el Teorema Central del Límite nos dice que, sin importar la distribución de la muestra aleatoria, si tiene esperanza y varianza finita, para un numero suficientemente grande (n>30), **EL PROMEDIO** va a converger en distribución a una normal de media cero y varianza σ^2

   Recordando que:

$$\frac{\bar{X}_n - \mu}{SE(\bar{X}_n)} \xrightarrow{d} \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{d} \mathbb{N}(0,1)$$

Parte 2: Mentalidad Tiburón 2.0.

La empresa ConstanzaContreras, líder en inversiones, le ha solicitado a los alumnos destacados del IN3401 (aquellos que fueron a la Auxilir #4) realizar un análisis estadístico sobre los datos entregados acerca de la tasa de interés fijada por los bancos al momento de otorgar un crédito. Esto con el objetivo de comprender de forma básica algunas dimensiones que analizan las entidades financieras al momento de otorgar una tasa.

Sabiendo que usted es una persona con Mentalidad Tiburón, se le solicita:

1. Analizar las variables con las que cuenta, evalué la necesidad de normalizar/estandarizar los datos o transformar las variables (por ejemplo a dummy o aplicar logaritmo). Justifique.

2. Seleccionar 3 variables que considere que pueden explicar el comportamiento de la variable *Interest_Rate*. Explique el por qué de su elección y que comportamiento espera de ellas.

3. Realizar dos gráficos en los cuales contraste dos variables del dataset que usted crea puedan estar correlacionadas o poseer un patrón interesante entre ellas. Explique su elección y comente los gráficos.

4. Genere un gráfico Boxplot para las variables Home.Ownership y Loan.Length, para evaluar el impacto (visualmente) que tiene en la tasa de interés fijada ubicación estante en las ventas.

5. Calcular la probabilidad de que un cliente solicite un monto mayor a 25.000 USD. Calcule la probabilidad de que un cliente sin mentalidad tiburón solicite un monto de crédito menor a 5.000 USD.

6. ¿Es razonable asumir que el monto solicitado de un crédito sigue una distribución Normal?

7. Finalmente, la auxiliar Pazi en sus ganas de ayudar, le asegura que no existe una diferencia en la varianza asociado a los periodos de préstamo de 36 y 60 meses, que le podría decir a su auxiliar al respecto.

   - Opción 1: Esta en lo correcto, por lo que es la mejor auxiliar del mundo.
   - Opción 2: Se ha observado que: $F > F_\alpha$ por lo cual rechazamos la hipótesis nula de igualdad de varianza asociado a los periodos de préstamo de 36 y 60 meses. Dado que la auxiliar se ha equivocado, debe abandonar el equipo docente inmediatamente

Solución

La solución se encuentra en tu corazón.

Ajajja ve el markdown en material docente !

Resumen

4. Poisson:

$$X \sim \text{poisson}(\lambda, k), \mathbb{P}(X = k) = \frac{\exp^{-\lambda} \lambda^k}{k!}$$

$$\mathbb{E}[X] = \lambda$$

$$\text{Var}[X] = \lambda$$

Distribuciones continuas:

1. Uniforme: $X \sim U(a, b), \mathbb{P}(X) = \frac{1}{b-a}(\text{ para } a \leq x \leq b).\mathbb{E}[X] = \frac{a+b}{2}, \text{Var}[X] = \frac{(b-a)^2}{12}$.
2. Exponencial: $X \sim \exp(\lambda), \mathbb{P}(X = k) = \lambda \exp^{-\lambda k} .\mathbb{E}[X] = \frac{1}{\lambda}, \text{Var}[X] = \frac{1}{\lambda^2}$.
3. Normal: $X \sim N(\mu, \sigma), \mathbb{P}(X = k) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \frac{-(k-\mu)^2}{2\sigma^2} \cdot \mathbb{E}[X] = \mu, \text{Var}[X] = \sigma^2$
4. Normal Estándar:
$$X \sim N\left(\mu, \sigma^2\right)$$
$$\rightarrow Z = \left(\frac{X-\mu}{\sigma}\right) \sim N(0, 1)$$
5. Chi-Cuadrado: Dado Z distribuye normal estándar.
$$X = Z^2 \sim X_1^2$$
6. T-Student: Dado Z distribuye normal estándar $y \times$ distribuye Chi-Cuadrado.
$$t = \frac{Z}{\sqrt{\frac{x}{n}}} \sim t_n$$

Teorema: Si $X \sim N\left(\mu, \sigma^2\right)$, entonces $aX + b \sim N\left(a\mu + b, a^2\sigma^2\right)$.

Población y Muestra:

- Población: Es un conjunto de individuos que poseen algunas características comunes observables en un contexto y lugar específico.
- Parámetro: Dato real que caracteriza a la población.
- Muestra Aleatoria: Corresponde a un subconjunto representativo de la población.
- Estimador: Es un estadístico que trata de estimar un parámetro.

Resumen

1. Estimador de la media:
$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

2. Estimador de la varianza
$$\hat{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \hat{\mu})^2$$

Propiedades:

- Insesgadez: Sea T(x) un estimador del parámetro θ. T(x) es insesgado si:
$$E[T(X)] = \theta$$

- Consistente: Un estimador X_n de θ es consistente si $X_n \xrightarrow{P} \theta$ con θ un escalar(no una variable aleatoria)

Teorema: Si x_n es un estimador insesgado de θ y $Var(X_n) \longrightarrow 0, cuandon {\longrightarrow} \infty$ entonces X_n es un estimador consistente.

Teoría Asintótica:

1. ***Convergencia en Probabilidad:*** Una secuencia de una v.a X_n converge en probabilidad a una v.a X si para cualquier ε ¿0 se cumple que:
$$\lim_{n\to\infty} P\left(|X_n - X| < \varepsilon\right) = 1$$
$$X_n \xrightarrow{p} X$$

2. ***Ley de los grandes números:*** La ley débil de los grandes números establece que si $X_1, X_2, X_3, \ldots$ es una sucesión infinita de variables aleatorias independientes que tienen el mismo valor esperado μ ¡ ∞ y varianza $\sigma^2 < \infty$, entonces el promedio:
$$\bar{X}_n = \left(X_1 + \cdots + X_n\right)/n, \quad \textit{Es un estimador consistente de } \mu$$

Resumen

3. ***Convergencia en Distribución:*** Se dice que una sucesión $X_1, X_2, \ldots$ de variables aleatorias reales converge en distribución, o converge débilmente, a una variable aleatoria X si

$$\lim_{n\to\infty} F_n(x) = F(x)$$

Para todo punto $x \in \mathbb{R}$ en el que F es continua, donde F_n y F denotan las funciones de distribución acumulada de las variables aleatorias X_n y X, respectivamente.

$$X_n \xrightarrow{d} X$$

4. ***Teorema Central del Límite:*** Se dice que una sucesión $X_1, X_2, \ldots$ de variables aleatorias reales con $E(X_i) = \mu < \infty$ y varianza $\sigma^2 < \infty$ cumple que s

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu) \xrightarrow{d} \mathbb{N}(0, \sigma^2)$$

Para una muestra lo suficientemente grande.

En palabras humanas: No importa la distribución de la muestra aleatoria, si tiene esperanza y varianza finita, para un numero suficientemente grande (n>30), **EL PROMEDIO** va a converger en distribución a una normal de media cero y varianza σ^2

Si estandarizamos obtenemos que:

$$\frac{\bar{X}_n - \mu}{SE(\bar{X}_n)} \xrightarrow{d} \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{d} \mathbb{N}(0, 1)$$

Def Intervalos de Confianza (IC) : *Rango de valores que con cierta probabilidad contiene al parámetro de interés.*

$$\begin{gathered} IC = C(\bar{X}) = \bar{X} \pm z_\alpha \cdot SE(\bar{X}) \\ P(\mu \in C(X)) = 1 - \alpha \\ = P\left(\bar{X} - z_\alpha SE(\bar{X}) \leq \mu \leq \bar{X} + z_\alpha SE(\bar{X})\right) \\ = P\left(-z_\alpha \leq \frac{\bar{X}-\mu}{SE(\bar{X})} \leq z_\alpha\right) \end{gathered}$$

No se dice: "La probabilidad de que μ este en el intervalo es de 95%", por favor no lo diga !!!

	Poblacional	**Muestral**
Media	μ	$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$
Varianza	$\sigma^2 = \frac{\sum_1^N \left(x_i - \bar{X}\right)^2}{N}$	$\hat{s}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \hat{\mu}\right)^2$
SE o Error Típico	σ	$\hat{s}$

Table 1: Medidas de disperción poblacional y muestral.

Parámetro Poblacional	**Tamaño de la muestra**	**Estimador -**	$E(\hat{\theta})$	$V(\hat{\theta})$
μ	n	$\hat{\mu}$	$E(\mu)$	$\frac{\sigma^2}{n}$
Proporcion P	n	$\hat{p} = \frac{X}{n}$	P	$\frac{\mathrm{p}q}{\mathrm{n}}$
Y = $\mu_1 - \mu_2$	n_1, n_2	$\hat{Y} = \hat{\mu}_1 - \hat{\mu}_2$	$\mu_1 - \mu_2$	$\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$
Z = $P_1 - P_2$	n_1, n_2	$\hat{Z} = \hat{p}_1 - \hat{p}_2$	$P_1 - P_2$	$\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n2}$

Table 2: Estimación de Parámetros

Condiciones	**Estadistico**
X Normal σ Conocido	$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$
X Normal σ Desconocido	$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1,\frac{\alpha}{2}}$
X Distribución Cualquiera σ Conocido n $\geq$ 30	$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \approx N(0,1)$
X Distribución Cualquiera σ Desconocido n $\geq$ 30	$\frac{\bar{X}-\mu}{S/\sqrt{n}} \approx N(0,1)$

Table 3: Distribución del estadístico.