PRUEBA DE HIPÓTESIS

Media poblacional con desviación estándar de la población (σ) conocido

En la clase anterior identificamos algunas pruebas que minimizan la probabilidad de cometer error tipo I y II. Una vez escogida la prueba, necesitamos decidir si rechazar o no la hipótesis nula. En esta clase aprenderemos de manera más formal cómo probar hipótesis sobre la media en la población cuando conocemos la desviación estándar de la población (σ). Para facilitar el cálculo asumiremos que la población de la que se extraen las muestras tiene una distribución normal. En caso de que no haya certeza de la normalidad, podemos ocupar el mismo método solamente si el tamaño de la muestra es suficientemente grande.

Prueba de una cola (50 min)

Supongamos que somos parte de un departamento del SERNAC y que estamos atendiendo una demanda de consumidores contra una cierta marca de café. Los envases de café y las propagandas dicen que los tarros traen al menos 3 gramos de café. Los consumidores han puesto esto en tela de juicio y nos piden investigar.

Nuestra hipótesis nula será probar si en efecto el tarro de café tiene al menos 3 gramos, mientras que la alternativa será que esto no ocurre y que tiene menos de 3 gramos. Como vimos en clases anteriores, estamos frente a una prueba de hipótesis de una cola:

$$H_0: \mu \geq 3$$

$$H_0: \mu < 3$$

Supondremos que:

- Conocemos la desviación estándar de la población: σ=0.18
- La muestra incluye 36 envases de café, con una distribución de pesos normal, sobre los cuales se calculó una media muestral $\bar{x}$ como estimación de la media poblacional μ
- El nivel de significancia o α, que nos indica la probabilidad de cometer el error tipo I, es igual a 0.01.

Si los datos que tomamos nos permiten rechazar la hipótesis nula, tendremos la evidencia estadística para inferir que el envase de café grande no tiene al menos 3 libras de café, y por lo tanto podremos tomar medidas en contra de la empresa y a favor de los clientes afectados.

Estadístico de pruebas

Dados los supuestos mencionados arriba, nuestro estadístico de prueba será el siguiente:

$$z = \frac{\bar{x} - H_0}{\sigma_{\bar{x}}}$$

En la fórmula anterior, z es el valor que calcularemos para determinar cuál es la probabilidad del valor $\bar{x}$, asumiendo su distribución normal (recuerden las clases de estimación), $\bar{x}$ es la media extraída de la muestra, H_0 es el valor de la hipótesis nula (3), y $\sigma_{\bar{x}}$ es el error estándar cuya forma es:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Donde σ es la desviación estándar de la población y n es el número de la muestra. Por lo tanto, podemos reemplazar los valores en la fórmula:

$$z = \frac{\bar{x} - 3}{0.03}$$

Para hacer la inferencia de la prueba de hipótesis, el estadístico de prueba z nos entregará un valor que debemos comparar con el valor crítico $\bar{z}$. Otra forma de proceder es ocupando el valor-p.

Valor crítico

El método del valor crítico depende directamente de nuestra elección del nivel de significancia o α, que nos indica la probabilidad de cometer el error tipo I. Para nuestro ejemplo ocuparemos los dos niveles de significancia ocupados con mayor frecuencia: α=0.01 y α= 0.05.

Dado nuestros supuestos, el valor $\bar{x}$ se distribuye de manera normal. Por lo tanto podemos usar la tabla de probabilidad de la distribución normal para extraer el valor $\bar{z}$ más cercano a los niveles de significancia. Para el caso de α=0.01 el valor crítico es $\bar{z} = -2.33$ y para α=0.05 el valor crítico es $\bar{z} = -1.64$. Estos valores críticos nos ayudarán a inferir si rechazamos o no la hipótesis nula.

Como nos encontramos en el lado izquierdo de la distribución, si el valor de prueba z es menor o igual al valor crítico $\bar{z}$, estamos en el área donde nuestro nivel de significancia sugiere rechazar la hipótesis nula. Es decir, la regla es rechazar H_0 si $z \leq \bar{z}$.

Supongamos que el valor encontrado en la muestra de $\bar{x} = 2.94$, por lo que el valor de z será:

$$z = \frac{\bar{x} - 3}{0.03} = \frac{2.94 - 3}{0.03} = -2.00$$

En este caso el valor de prueba z es menor que el valor crítico $\bar{z} = -1.64$ para α=0.05. Sin embargo, el valor de prueba z no es menor que el valor crítico $\bar{z} = -2.33$ para α=0.01. Por lo tanto solamente podemos rechazar la hipótesis nula e inferir que los envases de café no tienen en promedio el nivel que sus etiquetas dicen con α=0.05. Esto nos llevaría a tomar medidas contra dicha empresa con bastante certeza, pero no con el máximo nivel de confianza usualmente utilizado (99%).

Valor-p

Existe un método alternativo para probar hipótesis conocido como el valor-p. El cálculo del valor-p depende del valor de nuestro estadístico de prueba z. Supongamos ahora que el valor de $\bar{x}$ calculado en la muestra es $\bar{x} = 2.92$ (no 2.94 como en el ejemplo anterior). Entonces nuestro valor de z de prueba será -2.67:

$$z = \frac{\bar{x} - 3}{0.03} = \frac{2.92 - 3}{0.03} = -2.67$$

En la tabla de la distribución normal buscaremos la probabilidad asociada a nuestro valor de prueba z, que en este caso nos entrega un valor-p = 0.0038 (1-0.9962). Ahora nuestro criterio de decisión dependerá de la comparación entre el valor-p y el nivel de significancia que escogimos (0.01 o 0.05). En concreto, rechazaremos la hipótesis nula si el valor-p ≤ α.

En este caso el valor-p es menor que ambos niveles de significancia (p<0.05 y p<0.01). Por lo que tanto tenemos evidencia suficiente para rechazar la hipótesis nula y asumir que los tarros de café no contienen en promedio 3 libras de café.

STATA

Volvamos a ocupar la base de datos CASEN y hagamos un test t para los niveles de ingreso que aparecen ahí. Esta vez lo haremos intentando probar como hipótesis nula que el ingreso de personas menores de 30 años es igual a $500.000, contra la alternativa de que sus ingresos son superiores a eso. Para ello nuestro comando en Stata debe ser:

```
ttest ytotaj=500000 if edad<30

One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
  ytotaj |    1269    353411.3    13823.47     492434     326291.9   380530.7
------------------------------------------------------------------------------
mean = mean(ytotaj)                                          t = -10.6043
Ho: mean = 500000                           degrees of freedom =     1268

 Ha: mean < 500000          Ha: mean != 500000           Ha: mean > 500000
 Pr(T < t) = 0.0000       Pr(|T| > |t|) = 0.0000          Pr(T > t) = 1.0000
```

Ejercicio práctico

Prepare un breve informe y entréguelo al final de la clase.

a.- Ocupando las fórmulas revisadas en clases, calcule manualmente el valor del estadístico t y concluya cual es la inferencia que podemos hacer de este valor.

B.- Obtenga en Stata resultados de nueve pruebas de hipótesis, considerando que los ingresos son de 450000, 600000 y 700000, para los menores de 30 años, menores de 40 años y menores de 50 años.

c.- Obtenga en Stata resultados de una prueba de ingresos= 700,000 si los individuos están casados. Consideren como Ha que su salario es superior a 700,000.