Universida de Vigo

Departamento
de Estatística
e Investigación
Operativa

TERCEIRA PRÁCTICA EXCEL: INFERENCIA ESTATÍSTICA

Nesta práctica traballaremos sobre o ficheiro `alumnos.xls`. É recomendable almacenalo no escritorio e abrilo unha vez gardado.

Inferencia estatística.

1. **Intervalo de confianza para a media μ dunha poboacion normal:**

   a) Calcula un intervalo ó 90 % de confianza para a altura media supoñendo que a altura segue unha distribución normal $N(\mu, \sigma)$ con desviación estándar $\sigma = 8,58$.

   Procedemento: Da teoría sabemos que o intervalo ó nivel $1-\alpha$ é $\left(\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$.

   O centro do intervalo, $\bar{X}$, calculamos a partir dos datos e a amplitude do intervalo, $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$, calculámola usando a función de Excel *intervalo.confianza* tal e como se describe a continuación.

   Nunha folla en branco, calculamos o centro do intervalo, $\bar{X}$. Para iso, usaremos a función `promedio` aplicada sobre a variable *Altura*, tal e como se explicou na práctica anterior.

   Calculamos a amplitude, $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$. Situámonos na celda B3 e escribimos `=`. Seleccionamos *Fórmulas + Más funciones + Estadísticas + INTERVALO CONFIANZA*. En "Alfa" introducimos 0,1. En "Desviación estándar" introducimos 8,58. En "Tamaño" introducimos 102.

   A continuación calculamos o extremo inferior del intervalo $\left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$ na celda B6 e o superior $\left(\bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$ na celda B7.

   b) Calcula un intervalo ó 90 % de confianza para a altura media supoñendo que a altura segue unha distribución normal, sen facer ningún suposto sobre a desviación estándar. Compara este intervalo co calculado no apartado anterior.

   Procedemento: Da teoría sabemos que o intervalo ó nivel $1-\alpha$ é $\left(\bar{x} \pm t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}\right)$.

   O centro do intervalo, $\bar{X}$, xa o calculamos. A amplitude do intervalo, $t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}$, calculámolo como segue. Temos dúas alternativas:

   i) Usando unha das opcións de Excel. Situamonos na folla *Datos*. Accedemos ó menú *Datos+Análisis de Datos+Estadística Descriptiva*. En "Rango de entrada" introducimos os datos da variable *Altura*. En "Opciones de salida" seleccionamos "Rango de salida" e introducimos a celda `D1` da folla de datos onde está o exercicio anterior. Seleccionamos "Nivel de confianza para la media" e introducimos 90. A continuación calculamos o extremo inferior na celda `D6` e o superior na celda `D7`.

   ii) Introducindo a fórmula dos extremos do intervalo directamente. O valor $t_{n-1,\alpha/2}$ calculámolo mediante a función de Excel `=INV.T(probabilidad,grados de libertad)`. Esta función devolve o valor que nunha distribución t_{n-1} deixa á esquera a probabilidade que introducimos no seu argumento. Como o $t_{n-1,\alpha/2}$ defínese o valor nunha t_{n-1} deixa a dereita unha probailidade $\alpha/2$ (0.05 no noso caso), en "probabilidad" introducimos 0.95 e en grados de libertad 101.

   Para calcular a cuasidesviación típica S úsase a función `=DESVEST(...)` na que se introducen os valores da variable *altura*.

   A continuación calculamos o extremo inferior na celda `G6` e o superior na celda `G7`.
   Obviamente os valores calculados en i y ii son iguais.

   Este intervalo é maior (e polo tanto menos preciso) que o do caso normal. Esto é debido a que no anterior a varianza se asume coñecida.

Nota. Calquera intervalo de confianza estudiado se pode calcular escribindo a fórmula tanto do extremo superior como do inferior directamente nunha celda de Excel. Para esto, son de utilidade as seguintes funcións:

- **Promedio.** $\bar{X}$, `=promedio(rango de celdas)`.
- **Desviación típica.** S, `=desvest(rango de celdas)`.
- **Varianza.** S^2, `=var(rango de celdas)`.
- **Cuantil α da normal.** z_α, `=inv.norm(1-alpha; media; desviación estándar)`
- **Cuantil α da T de student.** $t_{n-1,\alpha}$, `=inv.t(1-alpha; grados de libertad)`
- **Cuantil α da χ^2.** $\chi^2_{n-1,\alpha}$, `=inv.chicuad(1-alpha; grados de libertad)`
- **Cuantil α da F de Snedecor.** $F_{n_1-1,n_2-1,\alpha}$,
  `=inv.f(1-alpha; grados de libertad 1; grados de libertad 2)`

2. **Contraste de hipótese para o cociente de varianzas en poboacións normais.**

   ¿Podemos considerar que a varianza da estatura é a mesma para os homes que para as mujeres? ($\alpha = 0.01$). Realizaremos o contraste $H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$ fronte a $H_1 : \frac{\sigma_1^2}{\sigma_2^2} \neq 1$.

   Procedemento: Descríbense dúas alternativas.

   a) Primero usaremos un dos menús de Excel. Situamonos na folla *Datos*. Seleccionamos *Datos+Análisis de Datos+Prueba F* para varianzas de dúas mostras. En “Rango para la variable 1:” introducimos as alturas das mulleres en “Rango para la variable 2:” a altura dos homes. En “Alfa” introducimos 0.005 (como é un test bilateral debemos introducir $\alpha/2$ en vez de α). En “Opciones de salida” seleccionamos “En una hoja nueva” e aceptamos.

      O criterio para saber cando rechazamos H_0 sería o seguinte:

      - Se “F” é menor que 1 entonces rexeitamos H_0 cando “F” é menor que “Valor crítico para F (una cola)”.
      - Se “F” é maior ou igual ca 1 entonces rexeitamos H_0 cando “F” é maior que “Valor crítico para F (una cola)”.

      No noso caso, “F” vale 0.8139 que é menor que 1. Como “Valor crítico para F (una cola)” vale 0.4785 aceptamos H_0. Podemos supoñer iguais as varianzas.

   b) Agora repetirémolo a través dunha das funcións de Excel.

      Situamonos nunha celda en blanco da folla de datos. Escribimos `=PRUEBA.F(Matriz1; Matriz2)` en `Matriz1` introducimos as alturas das mulleres e en `Matriz2` as alturas dos homes.

      Rexeitaremos H_0 cando o valor de esta función sexa menor que α. Como neste caso 0.4661 é maior que 0.01 aceptamos H_0. Podemos supoñer que as varianzas son iguais.

3. **Contraste de hipótese para a diferenza de medias en poboacións normais.**

   a) ¿Existen evidencias de que a altura media dos homes é distinta da das mulleres?

      Trátase dun contraste para a diferencia de medias: $H_0 : \mu_1 - \mu_2 = 0$ fronte a $H_1 : \mu_1 - \mu_2 \neq 0$ onde μ_1 fai referencia ás mulleres e μ_2 ós homes.

      Procedemento: Situámonos na folla de datos. Podemos asumir que a varianza de altura en mulleres e homes coinciden.

      Seleccionamos *Datos+Análisis de Datos+Prueba t para dos muestras suponiendo varianzas iguales*. En “Rango para la variable 1:” introducimos as alturas das mulleres e en “Rango para la variable 2:” as alturas dos homes. En “Alfa” introducimos 0.01. En “Opciones de salida” seleccionamos “En una hoja nueva” e aceptamos.
      O criterio xeral que se segue é que se o valor absoluto do “Estadístico t” é maior que “Valor crítico de t (dos colas)” entón rexeitamos H_0. No noso caso, “Estadístico t” = ?9.70 e “Valor crítico de t (dos colas)”= 2.62 polo que rexeitamos H_0. Existen evidencias de que a altura media nos homes é distinta da altura nas mulleres.