OCW-UPM Estadística para Ingeniería Civil y Medioambiental

Autores: E. M. García del Toro, C. Hermoso, E. J. Huertas

LABORATORIO DE ESTADÍSTICA CON MATLAB

PRÁCTICA 5 - Inferencia estadística: estimación puntual y por intervalos de confianza. **PARTE A**

En esta práctica, vamos a trabajar los siguientes casos de *inferencia estadística*, de acuerdo con los contenidos del correspondiente tema de la asignatura. Los casos a estudio son los siguientes:

1) Estimación puntual:
   - *a*) Estimación de parámetros poblacionales en la distribución binomial
   - *b*) Estimación de parámetros poblacionales en la distribución normal

2) Estimación por intervalos de confianza:
   - *a*) Intervalo de confianza para la media de una población de varianza conocida
   - *b*) Intervalo de confianza para la media de una distribución normal de varianza desconocida
   - *c*) Intervalo de confianza para la varianza poblacional de una distribución normal
   - *d*) Intervalo de confianza para la proporción (distribución binomial)

Observación importante sobre notación: En lo que sigue, recuérdese que se denota por z_p, $t_{n;p}$, $\chi^2_{n;p}$, y $F_{n_1,n_2;p}$, respectivamente, el valor de la abcisa de una $N(0,1)$ normal estándar, t_n de Student, $\chi^2_{n;p}$ de Pearson, y $F_{n_1,n_2;p}$ de Fisher, que deja a su **derecha** -bajo la correspondiente función de densidad- un área de probabilidad p.

El cálculo de estas abcisas (también llamadas **valores críticos**) puede realizarse en MATLAB, respectivamente mediante los siguientes comandos:

```
norminv   %Inverse of the normal cumulative distribution function (cdf).
    X = norminv(P,MU,SIGMA) returns the inverse cdf for the normal
    distribution with mean MU and standard deviation SIGMA, evaluated at
    the values in P. Default values for MU and SIGMA are 0 and 1, respectively.

tinv      %Inverse of Student's T cumulative distribution function (cdf).
    X=tinv(P,V) returns the inverse of Student's T cdf with V degrees
    of freedom, at the values in P.

chi2inv   %Inverse of the chi-square cumulative distribution function (cdf).
    X = chi2inv(P,V)  returns the inverse of the chi-square cdf with V
    degrees of freedom at the values in P.

finv   %Inverse of the F cumulative distribution function.
    X=finv(P,V1,V2) returns the inverse of the F distribution
    function with V1 and V2 degrees of freedom, at the values in P.
```

También podremos hallar estos valores mediante la herramienta gráfica `disttool` que ya hemos visto en otras ocasiones, o en última instancia, si no disponemos de computador, mediante las tablas de las respectivas distribuciones.

Estimación puntual y por intervalo de confianza para la media de una población normal

Varianza poblacional σ^2 conocida

Se desea estimar el caudal medio μ (en metros cúbicos por segundo) de cierto producto químico que circula por la tubería de una fábrica. Se supone que dicho caudal sigue una distribución normal, de media μ y de varianza

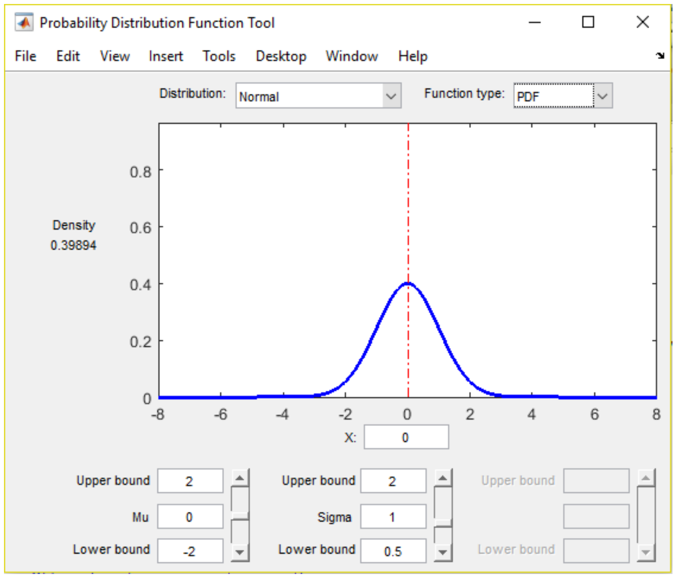

Figura 1: Ventana de la herramienta `disttool` en MATLAB

(poblacional) conocida $\sigma^2 = 144\ \left(m^3/s\right)^2$. Se llevan a cabo las siguientes 25 medidas (en m^3/s) de dicho caudal, en instantes elegidos al azar:

90.8, 81.5, 83.2, 79.7, 93.3, 65.0, 88.9, 90.9, 85.0, 72.5, 77.6, 101.4, 82.5, 93.6, 105.9, 71.1, 87.2, 75.0, 93.6, 61.8, 108.9, 88.7, 93.7, 80.6, 72.6

a) Hallar una estimación puntual del caudal medio μ de producto químico que circula por la tubería y de la varianza σ^2, mediante el estimador puntual que considere más adecuado en cada caso **(ejemplo de caso** $1b$**)**.

b) Se considera que la estimación puntual obtenida en el apartado anterior puede estar bastante alejada del verdadero valor poblacional del caudal medio μ que circula por la tubería. Para realizar entonces una inferencia más realista de este valor, se pide calcular tres Intervalos de Confianza distintos para μ, uno al 99 %, otro al 95 % y otro al 90 % de confianza (**ejemplos del caso** $2a$)).

c) Si forzamos al Intervalo de Confianza a tener un tamaño de 2 m^3/s, ¿qué nivel de confianza le corresponde a este intervalo?

Solución:

apartado a) Según se ha estudiado en las clases de teoría, el estimador más adecuado para la media poblacional desconocida μ es la media muestral

$$\bar{x} = \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right).$$

Utilizando MATLAB con la muestra proporcionada en el enunciado, introducimos los datos en cierto vector fila de longitud 25:

```
datos=[90.8, 81.5, 83.2, 79.7, 93.3, 65.0, 88.9, 90.9, 85.0, 72.5, 77.6, 101.4, 82.5, 93.6, 105.9,
71.1, 87.2, 75.0, 93.6, 61.8, 108.9, 88.7, 93.7, 80.6, 72.6]
```

y simplemente ejecutamos el comando `mean(datos)`, lo que nos da el valor de la media muestral $\bar{x} = 85\ m^3/s$.

Para estimar la varianza poblacional σ^2 el estimador puntual más adecuado es la cuasi-varianza muestral

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} (x_i - \bar{x})^2\right),$$

que en MATLAB se calcula mediante el comando `var(datos)` y cuyo valor para esta muestra es $s^2 = 139{,}04$.

apartado b)

En las clases teóricas vimos que la expresión para este Intervalo de Confianza es:

$$I = \left[\bar{x} \mp z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] = \left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\, \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right], \tag{1}$$

donde $z_{\alpha/2}$ es el valor de la abcisa en la distribución normal estándar N(0,1) que deja a su derecha un área de valor $\alpha/2$ (ver figura 1)

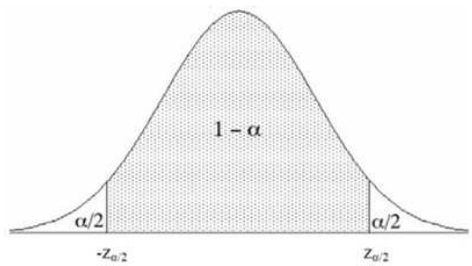

Figura 2: Valores críticos y gráfica de la pdf de la distribución Normal Estándar N(0,1)

Dado que la confianza es $(1-\alpha)$, y queremos tres intervalos $I_{0,99}$, $I_{0,95}$ e $I_{0,90}$, los respectivos valores de α son 0,01, 0,05, y 0,1. Estos valores de α nos sirven para calcular $z_{\alpha/2}$ en cada uno de los tres casos. Utilizando la herramienta `disttool`, tenemos que (esto puede hacerse también con la tabla de la N(0,1))

Para $\alpha = 0{,}01$, $\alpha/2 = 0{,}005$ y $z_{\alpha/2} = z_{0,005} = 2{,}5758$.
Para $\alpha = 0{,}05$, $\alpha/2 = 0{,}025$ y $z_{\alpha/2} = z_{0,025} = 1{,}96$.
Para $\alpha = 0{,}1$, $\alpha/2 = 0{,}05$ y $z_{\alpha/2} = z_{0,05} = 1{,}6449$.

Trabajo del alumno: Obtener de nuevo los valores críticos anteriores $z_{0,005}$, $z_{0,025}$, y $z_{0,05}$, pero esta vez utilizando el comando `norminv`. Obviamente, el resultado debe ser el mismo.

Los tres intervalos pedidos son entonces:

$$\begin{aligned}
I_{0,99} &= \left[\bar{x} - z_{0,005}\frac{\sigma}{\sqrt{n}},\, \bar{x} + z_{0,005}\frac{\sigma}{\sqrt{n}}\right] = \left[85 - 2{,}5758\frac{\sqrt{144}}{\sqrt{25}},\, 85 + 2{,}5758\frac{\sqrt{144}}{\sqrt{25}}\right] = [78{,}82\,,\,91{,}18]\ \ m^3/s,\\
I_{0,95} &= \left[\bar{x} - z_{0,025}\frac{\sigma}{\sqrt{n}},\, \bar{x} + z_{0,025}\frac{\sigma}{\sqrt{n}}\right] = \left[85 - 1{,}96\frac{\sqrt{144}}{\sqrt{25}},\, 85 + 1{,}96\frac{\sqrt{144}}{\sqrt{25}}\right] = [80{,}30\,,\,89{,}7]\ \ m^3/s,\\
I_{0,90} &= \left[\bar{x} - z_{0,05}\frac{\sigma}{\sqrt{n}},\, \bar{x} + z_{0,05}\frac{\sigma}{\sqrt{n}}\right] = \left[85 - 1{,}6449\frac{\sqrt{144}}{\sqrt{25}},\, 85 + 1{,}6449\frac{\sqrt{144}}{\sqrt{25}}\right] = [81{,}05\,,\,88{,}95]\ \ m^3/s,
\end{aligned}$$

cuyos tamaños respectivos son 12.36 m^3/s, 9,4 m^3/s y 7,9 m^3/s.
Todas las probabilidades que se calculan en esta práctica, pueden calcularse también con esta herramienta.

apartado c)
En la fórmula (1) vemos que, en esta caso, el ancho del intervalo es siempre

$$L = 2 \cdot z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Si obligamos a que L tenga tamaño fijo, como conocemos σ y n, podemos despejar $z_{\alpha/2}$ y con esto averiguar α. La confianza asociada a este intervalo de longitud fija será simplemente $(1-\alpha)$. En nuestro caso:

$$\begin{aligned}
2 &= 2 \cdot z_{\alpha/2}\frac{\sqrt{144}}{\sqrt{25}} \to z_{\alpha/2} = \frac{\sqrt{25}}{\sqrt{144}} = 0{,}4167 \underset{MATLAB}{\to}\\
& \quad z_{\alpha/2} \text{ deja a su izquierda un área de } 0{,}66155, \text{ por lo que}\\
\alpha/2 &= 1 - 0{,}66155 = 0{,}33845 \to \alpha = 2 \cdot 0{,}33845 = 0{,}677
\end{aligned}$$

que es un valor de α enorme, y que corresponde a un valor de confianza realmente muy bajo, de tan solo $(1-\alpha) = 1 - 0{,}677 = 0{,}323 \to 32{,}3\,\%$ de confianza.

<u>Varianza poblacional σ^2 desconocida</u>

d) En el mismo ejercicio anterior, hallar de nuevo los tres intervalos de confianza, pero en el caso en que la varianza poblacional σ^2 sea desconocida **(ejemplo del caso** $2b$**))**.

Solución:
apartado d) Si, como suele ocurrir en casos reales, no se conoce la varianza poblacional σ^2, habrá que estimarla con la cuasi-varianza muestral de las 25 observaciones. En este caso, vimos en las clases de teoría que este Intervalo de Confianza viene dado por:

$$I = \left[\bar{x} \mp t_{n-1;\alpha/2}\frac{s}{\sqrt{n}}\right] = \left[\bar{x} - t_{n-1;\alpha/2}\frac{s}{\sqrt{n}},\ \bar{x} + t_{n-1;\alpha/2}\frac{s}{\sqrt{n}}\right], \tag{2}$$

donde $t_{n-1;\alpha/2}$ es el valor de la abcisa en la distribución t de Student de $n-1$ grados de libertad, que deja a su derecha un área de valor $\alpha/2$ (ver figura 3)

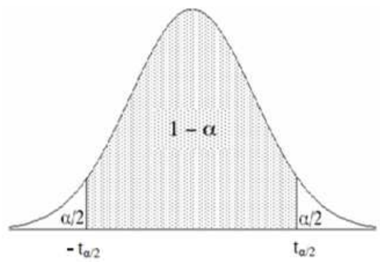

Figura 3: Valores críticos y gráfica de la pdf de la distribución t de Student

Si la estimación puntual obtenida fue $s^2 = 139{,}04$, utilizando por ejemplo el comando `tinv` de MATLAB para calcular $t_{24;0{,}005} =$ `tinv(1-0.005,24)`$= 2{,}7969$, $t_{24;0{,}025} =$ `tinv(1-0.025,24)`$= 2{,}0639$, $t_{24;0{,}05} =$ `tinv(1-0.05,24)`$=$ 1,7109, tendremos los siguientes Intervalos de Confianza:

$$\begin{aligned}
\tilde{I}_{0,99} &= \left[\bar{x} - t_{24;0,005}\frac{s}{\sqrt{n}},\ \bar{x} + t_{24;0,005}\frac{s}{\sqrt{n}}\right] = \left[85 - 2{,}7969\frac{\sqrt{139{,}04}}{\sqrt{25}},\ 85 + 2{,}7969\frac{\sqrt{139{,}04}}{\sqrt{25}}\right] = [78{,}40\,,\ 91{,}60]\ m^3/s,\\
\tilde{I}_{0,95} &= \left[\bar{x} - t_{24;0,025}\frac{s}{\sqrt{n}},\ \bar{x} + t_{24;0,025}\frac{s}{\sqrt{n}}\right] = \left[85 - 2{,}0639\frac{\sqrt{139{,}04}}{\sqrt{25}},\ 85 + 2{,}0639\frac{\sqrt{139{,}04}}{\sqrt{25}}\right] = [80{,}13\,,\ 89{,}87]\ m^3/s,\\
\tilde{I}_{0,90} &= \left[\bar{x} - t_{24;0,05}\frac{s}{\sqrt{n}},\ \bar{x} + t_{24;0,05}\frac{s}{\sqrt{n}}\right] = \left[85 - 1{,}7109\frac{\sqrt{139{,}04}}{\sqrt{25}},\ 85 + 1{,}7109\frac{\sqrt{139{,}04}}{\sqrt{25}}\right] = [80{,}97\,,\ 89{,}03]\ m^3/s,
\end{aligned}$$

cuyos tamaños respectivos son 13,2 m^3/s, 9,74 m^3/s y 8,06 m^3/s. Fijémonos que los resultados son parecidos a los del apartado **b)**.

Trabajo del alumno: Obtener de nuevo los valores críticos anteriores $t_{24;0{,}005}$, $t_{24;0{,}025}$, y $t_{24;0{,}05}$, pero esta vez utilizando la herramienta `disttool`. Obviamente, el resultado debe ser el mismo.