**Estadística I**
**Grado en Matemáticas, UAM, 2018-2019**

## Hoja 6. Intervalos de confianza

Los ejercicios de los dos primeros bloques son, en su mayor parte, ejercicios de implementación numérica de los intervalos de confianza más habituales. Consúltese la hoja resumen colgada en la web de la asignatura. Pueden ser útiles las siguientes funciones de Excel:

- Para $\alpha \in (0, 1/2]$, el percentil $z_\alpha$ se calcula con `=inv.norm.estand(`$1-\alpha$`)`.
- Para $\alpha \in (0, 1/2]$, el percentil $t_{\{n;\alpha\}}$ se calcula con `=inv.t(`$1-\alpha$`;`$n$`)`.
- Para $\alpha \in (0, 1)$, el percentil $\chi^2_{\{n;\alpha\}}$ se calcula con `=inv.chicuad(1-`$\alpha$`;`$n$`)`.
- Para $\alpha \in (0, 1)$, el percentil $F_{\{n_1,n_2;\alpha\}}$ se calcula con = `inv.f(`$1-\alpha$`;`$n_1$`;`$n_2$`)`.

Para calcular la media y la cuasidesviación típica muestrales de unos datos (situados en un rango de la hoja de cálculo): `=promedio(rango)` y `=desvest(rango)`, respectivamente.

### Intervalos de confianza para una distribución

**1.** Se sabe que el peso de los recién nacidos sigue una distribución normal. En una muestra aleatoria de 100 de ellos se obtiene una media muestral de 3 kg y una cuasidesviación típica de 0.5 kg. Calcula un intervalo de confianza al 95 % para la media poblacional.

Solución (con 3 decimales): $(2.901, 3.099)$.

**2.** Observamos el peso en gramos de una muestra de 10 aspirinas, obteniendo: 1.19, 1.23, 1.18, 1.21, 1.27, 1.17, 1.15, 1.14, 1.19, 1.20. Suponiendo normalidad, halla un intervalo de confianza al 80 % para la varianza.

Solución (con 6 decimales): $(0.000886, 0.003121)$.

**3.** Se quiere estudiar la proporción $p$ de declaraciones de la renta que presentan algún defecto. En una muestra preliminar pequeña (muestra piloto) de tamaño 50 se han observado 22 declaraciones defectuosas. ¿Cuál es el tamaño muestral necesario para estimar $p$ cometiendo un error máximo del 1 % con una probabilidad de 99 %?

Solución: $n \geq 16\,349$ (alternativa más conservadora: $n \geq 16\,588$).

**4.** Antes de encargar un gran lote de pilas alcalinas a una fábrica, queremos estimar la proporción $p$ de pilas defectuosas que podemos esperar. Se prueban 300 pilas, y encontramos 42 defectuosas.

a) Estima la proporción de pilas defectuosas (nivel de confianza 90 %).
b) Si deseamos obtener una estimación (con el mismo nivel de confianza) cuyo error sea inferior a 0.01, ¿cuántas pilas tendríamos que probar?

Solución: a) $p \in (10.70\,\%, 17.30\,\%)$. b) $n \geq 3\,258$ (o, sin usar la estimación previa, $n \geq 6\,764$).

**5.** Se quiere determinar la cantidad de nitrógeno en 8 muestras diferentes de harina. Estas medidas se realizan, en cada una de las muestras, con dos métodos diferentes. Se obtienen los siguientes resultados:

| Muestra de harina | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Método 1 | 2.0 | 1.4 | 2.3 | 1.2 | 2.1 | 1.5 | 2.4 | 2.0 |
| Método 2 | 1.8 | 1.5 | 2.5 | 1.0 | 2.0 | 1.3 | 2.3 | 2.1 |

Halla un intervalo al 95 % para la diferencia de medias, especificando las hipótesis utilizadas.
Solución (con 3 decimales): $\mu_1 - \mu_2 \in (-0.084, 0.184)$.

## Intervalos de confianza para dos distribuciones

**6.** Tenemos las siguientes muestras de la cantidad de una cierta substancia en dos hortalizas:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Tomate : | 777 | 790 | 759 | 790 | 770 | 758 | 764 |
| Pepino : | 782 | 773 | 778 | 765 | 789 | 797 | 782 |

a) Halla un intervalo de confianza para la diferencia de medias (nivel 95 %), suponiendo normalidad y que las varianzas en las dos poblaciones de hortalizas son iguales.

b) Halla un intervalo de confianza para la diferencia de medias si las varianzas en las dos poblaciones son posiblemente distintas. Compara el intervalo con el obtenido en b).

Solución (con 3 decimales): a) $(-22.368, 5.797)$. b) $(-22.512, 5.940)$.

**7.** Para un grupo de 41 alumnos de un instituto, se observaron las calificaciones obtenidas en las PAU y se obtuvo una cuasi-varianza muestral de 5.75. Para otro grupo de 25 estudiantes procedentes de otro instituto, se constató que la cuasi-varianza era 5.35. Suponiendo que la distribución de la nota de las PAU para cada uno de los centros es normal, halla el intervalo de confianza para el cociente de las varianzas al nivel de confianza del 90 %.
Solución (con 3 decimales): $(0.568, 1.927)$.

**8.** Una proporción $p_1$ de los móviles del modelo $A$ salen defectuosos de fábrica. Para el modelo $B$, esa proporción es $p_2$.

Se analizan 170 móviles del modelo $A$, y se detecta que 5 de ellos son defectuosos. En una muestra de tamaño 223 de móviles modelo $B$ se han encontrado 18 defectuosos. Calcula los intervalos de confianza para la diferencia $p_2 - p_1$ para $\alpha = 5\,\%$, $\alpha = 1\,\%$ y $\alpha = 0.1\,\%$.
Solución: Para $\alpha = 5\,\%$, IC $= (0.745\,\%, 9.516\,\%)$; para $\alpha = 1\,\%$, IC $= (-0.633\,\%, 10.894\,\%)$; y para $\alpha = 0.1\,\%$, IC $= (-2.232\,\%, 12.493\,\%)$.

## Ejercicios de corte teórico

**9.** La variable $X$ sigue una distribución $\mathcal{N}(\mu_0, \sigma^2)$. Aquí, $\mu_0$ es conocida.

Para estimar $\sigma^2$ (con muestras de tamaño $n$), usamos el estadístico

$$T_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_0)^2.$$

Analiza la distribución de este estimador y construye un intervalo de confianza $1-\alpha$ para $\sigma^2$.

**10.** Sea $X \sim \textsc{unif}[0,a]$. Consideramos el estimador $M_n = \text{máx}(X_1,\dots,X_n)$ del parámetro $a$.

a) Comprueba que, para todo $0 \le t \le 1$, $\mathbf{P}_a(M_n/a \le t) = t^n$.

b) Para $\alpha \in (0,1)$, calcula explícitamente los valores $c_1$ y $c_2$ para los que

$$\mathbf{P}_a(M_n/a \le c_1) = \mathbf{P}_a(M_n/a \ge c_2) = \alpha/2.$$

c) Obtenemos una muestra de $X$ de tamaño $n$. La estimación del parámetro (el máximo de esos números) es $\widehat{m}$. Calcula el intervalo de confianza $1-\alpha$ para $a$.

**11.** Sea $X = \delta + Y$, donde $Y \sim \textsc{exp}(1)$. Para estimar el parámetro $\delta > 0$, consideramos el estimador $T_n = \text{mín}(X_1,\dots,X_n)$.

a) Comprueba que

$$F_{T_n}(t) = \begin{cases} 0 & \text{si } t < \delta; \\ 1 - e^{-n(t-\delta)} & \text{si } t \ge \delta. \end{cases}$$

b) Para $\alpha \in (0,1)$, calcula explícitamente los valores $c_1$ y $c_2$ para los que

$$\mathbf{P}_\delta(T_n \le c_1) = \mathbf{P}_\delta(T_n \ge c_2) = \alpha/2.$$

c) Obtenemos una muestra de $X$ de tamaño $n$. La estimación del parámetro con $T_n$ es el número $\widehat{t}$. Calcula el intervalo de confianza $1-\alpha$ para $\delta$.

**12.** Sea $X \sim \textsc{exp}(\lambda)$, con $\lambda > 0$. Interesa estimar el parámetro $\theta = 1/\lambda$. Suponiendo que el tamaño $n$ de la muestra es grande, y usando la media muestral como estimador de $\theta$, halla un intervalo (aproximado) de confianza $1-\alpha$ para $\theta$.

**13.** La variable aleatoria $X$ sigue una distribución de Rayleigh, $X \sim \textsc{ray}(\theta)$, con $\theta > 0$. El estadístico estimador de $\theta$

$$T_n = \sqrt{\frac{1}{2}\overline{X^2}}$$

para muestras de tamaño $n$ de $X$ cumple el siguiente resultado de normalidad asintótica:

$$\sqrt{n}\,(T_n - \theta) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}(0, \theta^2/4).$$

Obtenemos una muestra de $X$ de tamaño $n$ (grande). La estimación del parámetro $\theta$ con el estimador $T_n$ es el número $\widehat{\theta}$. Calcula un intervalo (aproximado) de confianza $1-\alpha$ para $\theta$.