Intervalos de confianza

Ejercicio 1

Como la varianza es conocida, un intervalo de confianza para μ al nivel 95 % es

$$\overline{X} \pm 1.96 \cdot \frac{5}{\sqrt{n}}$$

Para que el ancho del intervalo sea menor a 1 debemos tener

$$2 \cdot 1.96 \cdot \frac{5}{\sqrt{n}} \leq 1$$

de donde $n \geq 384.16$, es decir $n \geq 385$.

Ejercicio 2

1. El t-intervalo es

$$\overline{X} \pm 2.26 \cdot \frac{S}{\sqrt{n}} = 65 \pm 2.26 \cdot \frac{5.98}{\sqrt{10}} = 65 \pm 4.27$$

2. Es un procedimiento que calcular un intervalo que contiene al verdadero valor de μ el 95 % de las veces (si uno lo repite varias veces).
3. El z-intervalo es

$$\overline{X} \pm 1.96 \cdot \frac{\sigma}{\sqrt{n}} = 65 \pm 1.96 \cdot \frac{5}{\sqrt{10}} = 65 \pm 3.10$$

   Conocer la verdadera varianza permite achicar un poco el intervalo.

Ejercicio 3

1. El z-intervalo para p al nivel 95 % es

$$\overline{X} \pm 1.96 \cdot \frac{\sqrt{\overline{X}(1-\overline{X})}}{\sqrt{n}}$$

   El término $\sqrt{\overline{X}(1-\overline{X})}$ es a lo sumo $1/2$, de donde podemos acotar el error por

$$\frac{1.96}{2} \frac{1}{\sqrt{n}} = \frac{0.98}{\sqrt{n}}.$$

   Si queremos que el error sea menor a 0.01, basta con encuestar a

$$n \geq \left(\frac{0.98}{0.01}\right)^2 = 9604$$

   personas.

2. El extremo inferior del z-intervalo es
$$\overline{X} - 1.645 \cdot \frac{\sqrt{\overline{X}(1-\overline{X})}}{20}$$
y queremos que sea mayor que 0.5. La igualdad produce una ecuación para $\overline{X}$ que podemos resolver:
$$\overline{X} - 1.645 \cdot \frac{\sqrt{\overline{X}(1-\overline{X})}}{20} = 0.5$$
Reagrupando, elevando al cuadrado, y reagrupando nuevamente, se obtiene
$$\left(1+\frac{1.645^2}{400}\right)\overline{X}^2 - \left(1+\frac{1.645^2}{n}\right)\overline{X} + 1/4 = 0$$
$$\overline{X} = \frac{1+\frac{1.645^2}{400} \pm \sqrt{\left(1+\frac{1.645^2}{400}\right)^2 - \left(1+\frac{1.645^2}{400}\right)}}{2\left(1+\frac{1.645^2}{400}\right)}$$
Obtenemos las soluciones $\overline{X} = 0.548$ y $\overline{X} = 0.465$. La segunda debe descartarse pues no es la que buscamos: al elevar al cuadrado el término $-1.645 \cdot \frac{\sqrt{\overline{X}(1-\overline{X})}}{20}$ que aparece en la fórmula para el extremo inferior del intervalo, hemos perdido la información sobre su signo. La ecuación cuadrática contiene dos soluciones, una para un extremo inferior por encima de 0.5 (la 0.548) y otra para un extremo superior por encima de 0.5 (la 0.465). Como nosotros queremos que el intervalo entero esté por encima, tenemos $\overline{X} = 0.548$.

Ejercicio 4

Esperamos que $40 \cdot 0.05 = 2$ de ellos estén equivocados. La probabilidad de que una binomial de parámetros $n = 40$ y $p = 0.05$ sea mayor o igual a 10 es 0.000021. Sería muy sorprendente que eso pasara.

Ejercicio 5

La distribución de $Q = (n-1)S^2/\sigma^2$ es χ^2 con $n-1$ grados de libertad. Podemos encontrar un intervalo que contenga al valor de Q con probabilidad $\alpha = 0.95$, es decir
$$\mathbf{P}(a \leq Q \leq b) = 0.95$$
imponiendo las condiciones
$$\mathbf{P}(Q < a) = 0.05/2 = 0.025, \quad \mathbf{P}(Q > b) = 0.05/2 = 0.025.$$
En nuestro caso, $n-1 = 26$, y de la tabla sacamos que $a = 13.84$ y $b = 41.92$.

Tenemos entonces que
$$13.84 \leq Q \leq 41.92$$
con probabilidad 0.95. De aquí podemos despejar σ^2, para obtener que
$$\frac{26S^2}{41.92} \leq \sigma^2 \leq \frac{26S^2}{13.84}$$
con probabilidad 0.95. El intervalo es entonces $\sigma \in [4.62, 8.03]$.

Ejercicio 6

Tenemos $n = 45$ datos normales, con $\overline{X} = 5$ y $S = 4$.

1. El promedio "estandarizado"
$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$
tiene distribución t con 44 grados de libertad. Podemos hallar c para que
$$\mathbf{P}(-c \leq T \leq c) = 0.8$$
Resulta $c = 1.30$. Despejando μ de la definición de T tenemos
$$\overline{X} - (1.30)\frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + (1.30)\frac{S}{\sqrt{n}}$$
con probabilidad 0.8. Luego el intervalo es 5 ± 0.78.

2. La distribución de $Q = (n-1)S^2/\sigma^2$ es χ^2 con $n-1$ grados de libertad. Podemos encontrar un intervalo que contenga al valor de Q con probabilidad $\alpha = 0.8$, es decir
$$\mathbf{P}(a \leq Q \leq b) = 0.8$$
imponiendo las condiciones
$$\mathbf{P}(Q < a) = 0.1, \quad \mathbf{P}(Q > b) = 0.1.$$
En nuestro caso, $n-1 = 44$, y de la tabla sacamos que $a = 32.49$ y $b = 56.37$.

   De aquí podemos despejar σ^2, para obtener que
$$\frac{44S^2}{56.37} \leq \sigma^2 \leq \frac{44S^2}{32.49}$$
con probabilidad 0.95. El intervalo es entonces $\sigma^2 \in [12.49, 21.67]$.

Ejercicio 7

Tenemos $n = 25$ datos normales, con $\overline{X} = 48.4$ y $S = 6.1$.

El promedio "estandarizado"
$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$
tiene distribución t con 24 grados de libertad. Podemos hallar c para que
$$\mathbf{P}(-c \leq T \leq c) = 0.95$$
Resulta $c = 2.06$. Despejando μ de la definición de T tenemos
$$\overline{X} - (2.06)\frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + (2.06)\frac{S}{\sqrt{n}}$$
con probabilidad 0.95. Luego el intervalo es 48.4 ± 2.51.

Ejercicio 8

Llamemos p a la verdadera proporción de personas en el distrito que votan por el partido del político. Haremos un z-intervalo pues el tamaño $n = 100$ de la muestra es grande y podemos usar el TCL.

El intervalo es $\overline{X} \pm (1.96)\sqrt{\overline{X}(1-\overline{X})}/\sqrt{n} = 0.5 \pm 0.098$. Como el extremo superior del intervalo es $0.598 < 0.7$, podemos afirmar que es poco probable (5 % de probabilidad) que el 70 % de las personas del distrito voten por el partido.

Ejercicio 9

La estimación por intervalo es de la forma $\overline{X} \pm (1.96)\sigma/\sqrt{n}$. Queremos entonces que

$$(1.96)\frac{\sigma}{\sqrt{n}} \leq 10$$

de donde $n \geq 15.36$, es decir $n \geq 16$.

Ejercicio 10

Lo único que varía son los valores crítico z_α.

En el caso de $\alpha_1 = \alpha_2 = 0.025$ tenemos que

$$z_{\alpha_1} = -1.96, \quad z_{\alpha_2} = 1.96.$$

En el caso de $\alpha_1 = 0.01$ y $\alpha_2 = 0.04$ tenemos que

$$z_{\alpha_1} = -2.33, \quad z_{\alpha_2} = 1.75.$$

La longitud del intervalo depende de la diferencia $z_{\alpha_2} - z_{\alpha_1}$, que vale 3.92 en el primer caso, y 4.08 en el segundo. La ventaja de tomar el intervalo simétrico es que es un intervalo más chico, y por lo tanto ganamos en precisión de estimación (manteniendo el mismo nivel de confianza).