Probabilitat i Estadística 1

Problemes Tema 7. Intervalos de confianza

1. In order to ensure efficient usage of a server, it is necessary to estimate the mean number of concurrent users (that at a given moment can be assumed to be normally distributed). According to records, the average number of concurrent users at 100 randomly selected times is 37.7, with a standard deviation $\sigma = 9.2$. Construct a 90% confidence interval for the expectation of the number of concurrent users.

   Resultat: $\mathrm{IC}_{1-\alpha}(\mu) \equiv [36.18673, 39.21327]$

   Resolució:
   Definimos la variable aleatoria X como el número de usuarios conectados simultáneamente en un instante dado, que supondremos normal, siguiendo la indicación del enunciado. Sean μ y σ^2 su esperanza y su varianza. Por el enunciado sabemos que $\sigma^2 = 9.2^2$.
   Aplicaremos la fórmula para construir un intervalo de confianza $(1-\alpha)$ para la esperanza de una normal con varianza conocida:

   $$\mathrm{IC}_{1-\alpha}(\mu) \equiv \bar{x}_n \mp z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

   En este caso se ha tomado una muestra de tamaño $n = 100$ para la cual la media muestral ha sido $\bar{x}_n = 37.7$. Se pide que la confianza sea 90%, es decir, $1-\alpha = 0.9 \Leftrightarrow \alpha = 0.1$. Para este valor de α buscamos el valor de $z_{\alpha/2}$ con R:

   ```
   alpha <- 0.1
   qnorm(1 - alpha/2)
   1.644854
   ```

   Así,

   $$\begin{aligned}\mathrm{IC}_{1-\alpha}(\mu) &\equiv 37.7 \mp 1.644854\frac{9.2}{\sqrt{100}}\\ &\equiv 37.7 \mp 1.513266\\ &\equiv [36.18673, 39.21327].\end{aligned}$$

2. Installation of a certain hardware takes random time with a standard deviation of 5 minutes and can be assumed to be normally distributed.

   (a) A computer technician installs this hardware on 64 different computers, with the average installation time of 42 minutes. Compute a 95% confidence interval for the population mean installation time.

   (b) Suppose that the population mean installation time is 40 minutes. A technician installs the hardware on your PC. What is the probability that the installation time will be within the interval computed in the previous point?

   Resultat: (a) $\mathrm{IC}_{1-\alpha}(\mu) \equiv [40.775, 43.225]$. (b) 0.1789471

   Resolució:

   (a) Definimos la variable aleatoria X como el tiempo (en minutos) de instalación de ese hardware, que supondremos normal, siguiendo la indicación del enunciado. Sean μ y σ^2 su esperanza y su varianza. Por el enunciado sabemos que $\sigma^2 = 5^2$.
   Aplicaremos la fórmula para construir un intervalo de confianza $(1-\alpha)$ para la esperanza de una normal con varianza conocida:

   $$\mathrm{IC}_{1-\alpha}(\mu) \equiv \bar{x}_n \mp z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

   En este caso se ha tomado una muestra de tamaño $n = 64$ para la cual la media muestral ha sido $\bar{x}_n = 42$ minutos. Se pide que la confianza sea 95%, es decir, $1-\alpha = 0.95 \Leftrightarrow \alpha = 0.05$. Para este valor de α sabemos que $z_{\alpha/2} = 1.96$. Así,

   $$\begin{aligned}\mathrm{IC}_{1-\alpha}(\mu) &\equiv 42 \mp 1.96\frac{5}{\sqrt{64}}\\ &\equiv 42 \mp 1.225\\ &\equiv [40.775, 43.225].\end{aligned}$$

(b) Sea X_{n+1} el tiempo de instalación del harware en este nuevo equipo. Si $\mu = 40$, $X_{n+1} \sim N(40, 5^2)$, luego

$$\Pr(X_{n+1} \in [40.775, 43.225]) = \Pr(40.775 \leq X_{n+1} \leq 43.225)$$
$$= \Pr\left(\frac{40.775 - 40}{5} \leq \frac{X_{n+1} - 40}{5} \leq \frac{43.225 - 40}{5}\right) = \Pr(0.155 \leq Z \leq 0.645)$$
$$= \texttt{pnorm(0.645)} - \texttt{pnorm(0.155)} = 0.7405364 - 0.5615893 = 0.1789471.$$

Observar que la forma de construir el $\mathrm{IC}_{1-\alpha}(\mu)$ garantiza una probabilidad de cobertura $(1-\alpha)$ para $\mu = \mathbb{E}(X_{n+1})$, pero no hace ninguna referencia a la probabilidad con la que X_{n+1} ha de pertenecer a los intervalos que proporciona.

3. Salaries of entry-level computer engineers have Normal distribution with unknown mean and variance. Three randomly selected computer engineers have salaries (in \$1000s):

$$30, 50, 70$$

Construct a 90% confidence interval for the average salary of an entry-level computer engineer.

Resultat: $\mathrm{IC}_{1-\alpha}(\mu) \equiv [16.28291, 83.71709]$

Resolució:
Definimos la variable aleatoria X como el salario (en miles de dólares) de un ingeniero informático que acaba de entrar en su primer empleo. X es normal esperanza μ y varianza σ^2 desconocidas. Aplicaremos la fórmula para construir un intervalo de confianza $(1-\alpha)$ para la esperanza de una normal con varianza desconocida:

$$\mathrm{IC}_{1-\alpha}(\mu) \equiv \bar{x}_n \mp t_{n-1,\alpha/2}\frac{S}{\sqrt{n}},$$

donde S es la raíz cuadrada de la varianza muestral. En este caso se ha tomado una muestra de tamaño $n = 3$ con observaciones $x_1 = 30, x_2 = 50, x_3 = 70$. Para estos datos la media muestral vale $\bar{x}_n = 50$ y

$$S^2 = \frac{1}{3-1}\left((30-50)^2 + (50-50)^2 + (70-50)^2\right) = 400.$$

Se pide que la confianza sea 90%, es decir, $1 - \alpha = 0.9 \Leftrightarrow \alpha = 0.1$. Para este valor de α buscamos el valor de $t_{n-1,\alpha/2}$ con R:

```
alpha <- 0.1
n <- 3
qt(1 - alpha/2, df=n-1)
2.919986
```

Así,

$$\mathrm{IC}_{1-\alpha}(\mu) \equiv 50 \mp 2.919986\frac{\sqrt{400}}{\sqrt{3}}$$
$$\equiv 50 \mp 33.71709$$
$$\equiv [16.28291, 83.71709].$$

4. We have to accept or reject a large shipment of items. For quality control purposes, we collect a sample of 200 items and find 24 defective items in it. Construct a 96% confidence interval for the proportion of defective items in the whole shipment.

Resultat: $\mathrm{IC}_{1-\alpha}(p) \equiv [9.7\%, 14.3\%]$

Resolució:
Sea $X \sim B(n, p)$. Usamos proporción muestral $\hat{p} = X/n$ como estimador de p. Sabemos que un intervalo de confianza aproximada $(1-\alpha)$ para p viene dado por la expresión

$$\mathrm{IC}_{1-\alpha}(p) \equiv \hat{p} \mp z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},$$

Se pide que la confianza sea 96%, es decir, $1-\alpha = 0.96 \Leftrightarrow \alpha = 0.04$. Para este valor de α buscamos el valor de $z_{\alpha/2}$ con R:

```
alpha <- 0.04
qnorm(1 - alpha/2)
2.053749
```

Se ha observado el valor $x = 24$ de X. Así la proporción muestral observada es

$$\hat{p} = \frac{x}{n} = \frac{24}{200} = 0.12,$$

y el intervalo será

$$\begin{aligned} \mathrm{IC}_{1-\alpha}(p) &\equiv 0.12 \mp 2.053749\sqrt{\frac{0.12 \cdot 0.88}{200}} \\ &\equiv 0.12 \mp 0.2297825 \\ &\equiv [0.09702175, 0.1429783] \\ &\equiv [9.7\%, 14.3\%]\,. \end{aligned}$$

5. An electronic parts factory produces resistors. Statistical analysis of the output suggests that resistances follow an approximately Normal distribution with a standard deviation of 0.2 ohms. A sample of 52 resistors has the average resistance of 0.62 ohms.

   (a) Based on these data, construct a 95% confidence interval for the population mean resistance.

   (b) If the actual population mean resistance is exactly 0.6 ohms, what is the probability that an average of 52 resistances is 0.62 ohms or higher?

   Resultat: (a) $\mathrm{IC}_{95\%}(\mu) \equiv [0.56563938, 0.67436062]$. (b) 0.2354208

   Resolució:
   Sea X =“Resistencia, en ohmios, de una de las piezas producidas en la fábrica”:

   $$X \sim N(\mu, \sigma^2 = 0.2^2).$$

   Se toma una muestra de X de tamaño $n = 52$ para la cual la media muestral vale $\bar{x}_n = 0.62$ ohmios.
   (a)

   $$\begin{aligned} \mathrm{IC}_{95\%}(\mu) &\equiv \bar{x}_n \mp z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \\ &\equiv 0.62 \mp 1.96\frac{0.2}{\sqrt{52}} \\ &\equiv 0.62 \mp 0.05436062 \\ &\equiv [0.56563938, 0.67436062]. \end{aligned}$$

   (b) Si $\mu = 0.6$, $\sigma^2 = 0.2^2$, $n = 56$, $X_1, \ldots, X_n$ i.i.d. $N(\mu, \sigma 2)$, entonces $\bar{X}_n = (1/n)\sum_{i=1}^n X_i \sim N(\mu, \sigma^2/n)$. Por lo tanto, si $x_0 = 0.62$,

   $$\begin{aligned} \Pr(\bar{X}_n \geq x_0) &= \Pr\left(\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \geq \frac{\sqrt{n}(x_0 - \mu)}{\sigma}\right) \\ &\quad (\text{sea } Z \sim N(0,1)) \\ &= \Pr\left(Z \geq \frac{\sqrt{n}(x_0 - \mu)}{\sigma}\right) \\ &= \Pr\left(Z \geq \frac{\sqrt{52}(0.62 - 0.60)}{0.2}\right) \\ &= \Pr(Z \geq 0.7211103) \\ &= 1 - \texttt{pnorm}(0.7211103) \\ &= 0.2354208. \end{aligned}$$

6. A sample of 250 items from lot A contains 10 defective items, and a sample of 300 items from lot B is found to contain 18 defective items. Construct a 98% confidence interval for the difference of proportions of defective items.

Resultat: $\text{IC}_{1-\alpha}(p_A - p_B) \equiv [-0.063, 0.023]$.

Resolució:
Sea $X_A \sim B(n_A = 250, p_A)$, de la cual se ha observado el valor $x_A = 10$. Así la estimación de p_A es

$$\hat{p}_{n_A} = \frac{x_A}{n_A} = \frac{10}{250} = 0.04.$$

Por otra parte, $X_B \sim B(n_B = 330, p_B)$ y se ha observado el valor $x_B = 18$. Así la estimación de p_B es

$$\hat{p}_{n_B} = \frac{x_B}{n_B} = \frac{18}{300} = 0.06.$$

Nos piden un intervalo de confianza 98%. Eso implica $\alpha = 0.02$ y entonces

$$z_{\alpha/2} = z_{0.01} = \texttt{qnorm(1 - 0.01)} = 2.326348.$$

Usaremos la fórmula del intervalo de confianza para la diferencia de proporciones:

$$\begin{aligned}
\text{IC}_{1-\alpha}(p_A - p_B) &\equiv (\hat{p}_A - \hat{p}_B) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_A(1-\hat{p}_A)}{n_A} + \frac{\hat{p}_B(1-\hat{p}_B)}{n_B}} \\
&\equiv (0.04 - 0.06) \pm 2.326348\sqrt{\frac{0.04 \cdot 0.96}{250} + \frac{0.06 \cdot 0.94}{300}} \\
&\equiv -0.02 \pm 2.326348\sqrt{0.0001536 + 0.000188} \\
&\equiv -0.02 \pm 2.326348 \cdot 0.01848242 \\
&\equiv -0.02 \pm 0.043 \\
&\equiv [-0.063, 0.023].
\end{aligned}$$

7. A news agency publishes results of a recent poll. It reports that candidate A leads candidate B by 10% because 45% of the poll participants supported Ms. A whereas only 35% supported Mr. B. What margin of error should be reported for each of the listed estimates, 10%, 35%, and 45%? Notice that 900 people participated in the poll, and the reported margins of error typically correspond to 95% confidence intervals.

Resultat: $\text{IC}_{1-\alpha}(p_A) \equiv [0.417, 0.483]$, $\text{IC}_{1-\alpha}(p_B) \equiv [0.319, 0.381]$, $\text{IC}_{1-\alpha}(p_A - p_B) \equiv [0.042, 0.158]$

Resolució:
Definimos las variables aleatorias

$$\begin{aligned}
X_A &= \text{“Número de encuestados que dicen que votarán A”} \\
X_B &= \text{“Número de encuestados que dicen que votarán B”} \\
X_O &= \text{“Número de encuestados que dicen que votarán a otros candidatos”}
\end{aligned}$$

Observar que la distribución conjunta de estas tres variables es multinomial:

$$(X_A, X_B, X_O) \sim M_3(n = 900; p_A, p_B, 1 - p_A - p_B).$$

Por las propiedades de la multinomial sabemos que

$$X_A \sim B(n, p_A),\ X_B \sim B(n, p_B),\ \text{Cov}(X_A, X_B) = -np_Ap_B.$$

Así, si definimos $Y = X_A - X_B$, tenemos que

$$\mathbb{E}(Y) = \mathbb{E}(X_A - X_B) = np_A - np_B,\ \text{Var}(Y) = \text{Var}(X_A - X_B)$$

$$= \text{Var}(A) + \text{Var}(B) - 2\text{Cov}(X_A, X_B) = np_A(1-p_A) + np_B(1-p_B) + 2np_Ap_B.$$

Observar que $Y = \sum_{i=1}^{n} Y_i$, donde Y_i son v.a.i.i.d. definidas como

$$Y_i = \begin{cases} 1 & \text{si el encuestado } i \text{ dice que votará A,} \\ -1 & \text{si el encuestado } i \text{ dice que votará B,} \\ 0 & \text{si el encuestado } i \text{ dice que votará otro candidato.} \end{cases}$$

Observar además que $\mathbb{E}(Y_i) = p_A - p_B$ y que $\mathbb{E}(Y_i^2) = p_A + p_B$, de donde se sigue que

$$\mathrm{Var}(Y_i) = p_A + p_B - (p_A - p_B)^2 = p_A(1 - p_A) + p_B(1 - p_B) + 2p_Ap_B,$$

lo que prueba, de forma alternativa, la expresión anterior para la varianza de Y. Observar que

$$\hat{p}_A - \hat{p}_B = \frac{X_A}{n} - \frac{X_B}{n} = \frac{Y}{n} = \bar{Y}_n \underset{\text{si } n \text{ grande}}{\approx} N\left(p_A - p_B, \frac{1}{n}\left(p_A(1 - p_A) + p_B(1 - p_B) + 2p_Ap_B\right)\right).$$

Ahora podemos aplicar la fórmula genérica

$$\mathrm{IC}_{1-\alpha}(\theta) \approxeq \hat{\theta} \mp z_{\alpha/2}\,\mathrm{se}(\hat{\theta})$$

para intervalos de confianza aproximados basados en estimadores asintóticamente normales. Así, trabajando con $1 - \alpha = 0.95$,

$$\begin{aligned} \mathrm{IC}_{1-\alpha}(p_A) &\approxeq \hat{p}_A \mp z_{\alpha/2}\sqrt{\frac{\hat{p}_A(1 - \hat{p}_A)}{n}} \\ &\equiv 0.45 \mp 1.96\sqrt{\frac{0.45 \cdot 0.55}{900}} \\ &\equiv 0.45 \mp 0.03250292 \\ &\equiv [0.4174971, 0.4825029], \end{aligned}$$

$$\begin{aligned} \mathrm{IC}_{1-\alpha}(p_B) &\approxeq \hat{p}_B \mp z_{\alpha/2}\sqrt{\frac{\hat{p}_B(1 - \hat{p}_B)}{n}} \\ &\equiv 0.35 \mp 1.96\sqrt{\frac{0.35 \cdot 0.65}{900}} \\ &\equiv 0.35 \mp 0.03116201 \\ &\equiv [0.318838, 0.381162], \end{aligned}$$

$$\begin{aligned} \mathrm{IC}_{1-\alpha}(p_A - p_B) &\equiv (\hat{p}_A - \hat{p}_B) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_A(1 - \hat{p}_A) + \hat{p}_B(1 - \hat{p}_B) + 2\hat{p}_A\hat{p}_B}{n}} \\ &\equiv (0.45 - 0.35) \pm 1.96\sqrt{\frac{0.45 \cdot 0.55 + 0.35 \cdot 0.65 + 2 \cdot 0.45 \cdot 0.35}{900}} \\ &\equiv 0.10 \pm 0.058 \\ &\equiv [0.042, 0.158] \\ &\equiv [4.2\%, 15.8\%]. \end{aligned}$$

8. If $[A, B]$ is a $(1 - \alpha)100\%$ confidence interval for the population variance (with $a \geq 0$), prove that $[\sqrt{A}, \sqrt{B}]$ is a $(1 - \alpha)100\%$ confidence interval for the population standard deviation.

   Resolució:
   Sea σ^2 la varianza poblacional y $\sigma = \sqrt{\sigma^2}$ la desviación típica poblacional. Teniendo en cuenta que la función $h(x) = x^2$ es monótona creciente en $[0, \infty)$ y que $\sigma \in [0, \infty)$, se tiene que

   $$\Pr(\sqrt{A} \leq \sigma \leq \sqrt{B}) = \Pr(A \leq \sigma^2 \leq B) = 1 - \alpha,$$

   lo que concluye la demostración.