Práctica 7

Alejandro Cáceres
UPC - Statistics 2019/2020

Objetivo

- Intervalos de confianza

Intervalos de confianza

Generalmente hacemos un sólo experimento con n mediciones.

Por ejemplo el resultado de medir la impedancia de 10 altavoces de 4 Ohmios es:

```
exp1 <- c(3.889533, 4.800898, 4.262898, 4.361508,
5.617505, 3.423186, 3.575229, 6.961411, 3.833203,
3.560329)

barx <- mean(exp1)

> barx
[1] 4.42857
```

Intervalos de confianza

- tenemos **n** valores como resulatado de n mediciones $(x_1, ..., x_n)$, que provienen de las variables aleatorias $(X_i, ...X_n)$ que son n mediciones de la misma varibale aleatoria $X = X_i = .. = Xn$.
- tenemos **1** valor para la media muestral $\bar{x} = 1/n \sum x_i$, que proviene de **otra** variable aleatoria

$$\bar{X} = \frac{1}{n} \sum X_i$$

Intervalos de confianza

Recordemos que:

- hay 2 distribuciones una para X (distribución poblacional, para la que asumimos modelos de probabilidad; binomial, poisson, etc...) y otra para $\bar{X}$ (que intentamos averiguar desde los modelos para X: TCL, t, χ^2).
- no sabemos $\mu = E(X)$, nunca lo observamos directamente.
- $\bar{x}$ lo obetnemos de los datos y lo usamos para estimar μ, ya que $E(\bar{X}) = E(X)$.
- podemos tomar $\mu \sim \bar{x}$ $(\hat{\mu} = \bar{x})$, pero no sabemos **qué tan cerca** estamos de μ.

Intervalos de confianza

Busquemos los números f_{inf} y f_{sup} tal que

$$P(f_{inf} \leq \bar{X} - \mu \leq f_{sup}) = 0.95$$

O sea los valores que contienen al 95% de las diferencias entre $\bar{X}$ y μ.
Escribámos esta proabilididad como:

$$P(\bar{X} - f_{sup} \leq \mu \leq \bar{X} - f_{inf}) = 0.95$$

el intervalo aleatorio $(L, U) = (\bar{X} - f_{sup}, \bar{X} - f_{inf})$ tiene una probailidad del 95% de contener a μ

Intervalos de confianza

Volvamos a la probabilidad incial

$$P(f_{inf} \leq \bar{X} - \mu \leq f_{sup}) = 0.95$$

y dividamos por $\sigma_{\bar{X}}$

$$P(\frac{f_{inf}}{\sigma_{\bar{X}}} \leq \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} \leq \frac{f_{sup}}{\sigma_{\bar{X}}}) = 0.95$$

remplazemos $\sigma_{\bar{X}} = \sigma_X / \sqrt{n}$

$$P(\frac{f_{inf}}{\sigma_X/\sqrt{n}} \leq \frac{\bar{X} - \mu}{\sigma_X/\sqrt{n}} \leq \frac{f_{sup}}{\sigma_X/\sqrt{n}}) = 0.95$$

Intervalos de confianza

$$P(\frac{f_{inf}}{\sigma_X/\sqrt{n}} \leq \frac{\bar{X}-\mu}{\sigma_X/\sqrt{n}} \leq \frac{f_{sup}}{\sigma_X/\sqrt{n}}) = 0.95$$

Asumimos 1: que X se distribye normalmente $N(\mu,\sigma)$

Asumimos 2: que sabemos (nos dan) σ_X como información adicional. Por ejemplo el fabricante sabe que la desviación estándard de los amplificdores que produce es $\sigma_X = 1.4\Omega$

Entonces: $Z = \frac{\bar{X}-\mu}{\sigma_X/\sqrt{n}}$ es una variable aleatoria estandard $N(\mu = 0, \sigma = 1)$:

$$P(z_{0.025} \leq Z \leq z_{0.975}) = 0.95$$

Intervalos de confianza

$$P(z_{0.025} \leq Z \leq z_{0.975}) = 0.95$$

```
> z0.025 <- qnorm(0.025)
[1] -1.959964
> z0.975 <- qnorm(0.975)
[1] 1.959964
```

o sea:

$$f_{inf} = z_{0.025} * \sigma_X / \sqrt{n}$$
$$f_{sup} = z_{0.975} * \sigma_X / \sqrt{n}$$

Para nuestro experimento, tenemos n=10 y $\sigma_X = 1.4cm$ y $f_{inf} = -f_{sup}$

```
> fsup <- 1.959964*1.4/sqrt(10)
> fsup
[1] 0.8677131
```

Intervalos de confianza

Entonces para el intervalo aleatorio

$$(L, U) = (\bar{X} - f_{sup}, \bar{X} - f_{inf})$$

hemos hecho la observación

$$(l, u) = (\bar{x} - f_{sup}, \bar{x} + f_{sup})$$

```
> CI <- c(barx-fsup,barx+fsup)
> CI
[1]] 3.560857 5.296283
```

Intervalos de confianza

El intervalo $(l, u) = c(3.503469, 5.353670)$ es una observación de (L,U) para la cual tenemos una confianza del 95% de haber atrapado a μ.

Esto también lo escribimos como:

$$\bar{x} = 4.4 \pm 0.8$$

O sea que al estimar μ con $\bar{x}$, no podemos estar muy seguros de la precisioón dada por el decimal .4. Los decimales siguientes carecen de confianza y los podemos descartar.

Intervalos de confianza

Los intervalos de confianza para

- un experimento de n muestras.
- de una distribución normal para X.
- del que sabemos la desviación estándard para X: σ_X.

se pueden calcular con la función
z.test

Intervalos de confianza

```
> install.packages("TeachingDemos") #solo se instala una vez
> library(TeachingDemos) #cada vez que queramos usar el paquete
> z.test(exp1,sd=1.4)


One Sample z-test

data:  exp1
z = 10.003, n = 10.00000, Std. Dev. = 1.40000, Std.
Dev. of the sample mean = 0.44272, p-value <
2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 3.560857 5.296283
sample estimates:
mean of exp1
     4.42857
```

Intervalos de confianza

Calculemos los intervalos de confianza para

- un experimento de n muestras
- una distribución para X normal (o desconocida si $n > 30$)
- pero **no** sabemos su desviación estándar $\sigma = \sqrt{V(X)}$

Intervalos de confianza

Recordemos que queremos:

$$P(\frac{f_{inf}}{\sigma_X/\sqrt{n}} \leq \frac{\bar{X}-\mu}{\sigma_X/\sqrt{n}} \leq \frac{f_{sup}}{\sigma_X/\sqrt{n}}) = 0.95$$

Asumimos 1: que X es una distribución normal $N(\mu, \sigma)$
Asumimos 2: podemos sustituir σ_x por su estimador $s = \frac{1}{n-1}\Sigma_i(x_i - \bar{x})^2$ que es calculado de los datos.
Entonces: $T = \frac{\bar{X}-\mu}{s/\sqrt{n}}$ es una variable de una distibución t-student con n-1 grados de libertad $t(n-1)$.

Intervalos de confianza

$$P(\frac{f_{inf}}{s/\sqrt{n}} \leq \frac{\bar{X}-\mu}{s/\sqrt{n}} \leq \frac{f_{sup}}{s/\sqrt{n}}) = 0.95$$

$$P(t_{0.025,n-1} \leq T \leq t_{0.975,n-1}) = 0.95$$

siendo T una variable aleatoria con una dostribución student-t con n-1 grados de libertad

```
> qt(0.025,df=9)
[1] -2.262157
> qt(0.975,df=9)
[1] 2.262157
```

$$f_{sup} = t_{0.975,n-1} s/\sqrt{n} = -f_{inf}$$

Intervalos de confianza

```
barx <- mean(exp1)
> barx
[1] 4.42857
s <-sd(exp1)
> s
[1] 1.112419
```

Entonces:

```
> fsup <- 2.262157*s/sqrt(10)
> fsup
[1] 0.7957762
> CI <- c(barx-fsup,barx+fsup)
> CI
[1] 3.632794 5.224346
```

Intervalos de confianza

En este caso se usa la fución **t.test**

```
> t.test(exp1, conf.level = 0.95)

One Sample t-test

data:  exp1
t = 12.589, df = 9, p-value = 5.114e-07
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 3.632794 5.224346
sample estimates:
mean of x
```

Notemos que el nivel de confianza se puede cambiar con el parámetro **conf.level**

Intervalos de confianza

Veamos el intervlo de confianza para la varianza.

Para un experimento queremos saber el intervalo de confiaza para su varianza muestral,

```
> s<-sd(exp1)
> s^2
[1] 1.237475
```

Recordemos que s es un valor de la variable aleatoria $S^2 = 1/(n-1)\Sigma_i(X_i - \mu)^2$ y que

$$S^2(n-1)/\sigma^2 \to \chi^2_{n-1}$$

Intervalos de confianza

Buscamos

$$P(\chi^2_{0.025,n-1} \leq S^2(n-1)/\sigma^2 \leq \chi^2_{0.975,n-1}) = 0.95$$

En términos del intervalo aleaorio (L,U) tal que $P(L \leq \sigma^2 \leq U) = 0.95$ tenemos

$$P(\frac{S^2(n-1)}{\chi^2_{0.975,n-1}} \leq \sigma^2 \leq \frac{S^2(n-1)}{\chi^2_{0.025,n-1}}) = 0.95$$

Intervalos de confianza

```
> chiInf<-qchisq(0.025, df=n-1)
> chiInf
[1] 2.700389
> chiSup<-qchisq(0.975, df=n-1)
> chiSup
[1] 19.02277
```

y para nuestros datos

```
> n<-10
> CI<-c(s^2*(n-1)/chiSup, s^2*(n-1)/chiInf)
> CI
[1] 0.5854708 4.1243219
```

Intervalos de confianza

En R se calcula con **sigma.test**

```
> sigma.test(exp1)

One sample Chi-squared test for variance

data:  exp1
X-squared = 11.137, df = 9, p-value = 0.5328
alternative hypothesis: true variance is not equal to 1
95 percent confidence interval:
 0.5854708 4.1243219
sample estimates:
var of exp1
   1.237475
```