Ejercicios Tema 3 SOLUCIONES - Intervalos de Confianza. Taller 2

Ricardo Alberich, Juan Gabriel Gomila y Arnau Mir

Curso completo de estadística inferencial con R y Python

Contenidos

1 Estimación por intervalos taller 2

1.1 Ejercicio 1

Una infección por un virus puede haber perjudicado a muchos ordenadores. Desde el Centro de Alerta Temprana (CAT) se quiere calcular la proporción de ordenadores infectados. El jefe del centro os pide que calculéis el tamaño de una muestra para que el intervalo de confianza de la proporción muestral de ordenadores infectados tenga amplitud de a lo sumo 0.01 con una probabilidad del 90%.

1.1.1 Solución

El el peor de los casos la proporción $p = 0.5$. La aproximación por el TLC del tamaño muestral que da una amplitud de la proporción en tanto por 1 del 0.01 es en este caso

Recordemos que la **Fórmula de Laplace**, En este caso la amplitud $A_0 = 0.01$

$$A_0 \geq 2 z_{1-\frac{\alpha}{2}} \sqrt{\frac{0.5^2}{n}} = \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \Rightarrow n \geq \left\lceil \frac{z_{1-\frac{\alpha}{2}}^2}{A_0^2} \right\rceil.$$

por lo tanto

```
A0=0.01
alpha=1-0.9
n=ceiling(qnorm(1-alpha/2)^2/A0^2)
n
```

```
## [1] 27056
```

Para asegurar este error, y en el peor de los casos $p = 0.5$, necesitamos una muestra de tamaño $n = 27056$.

1.2 Ejercicio 2

Se han medido los siguientes valores (en miles de personas) para la audiencia de un programa de televisión en distintos días (supuestos igualmente distribuidos e independientes):

$$521, 742, 593, 635, 788, 717, 606, 639, 666, 624.$$

Construir un intervalo de confianza del 90%, para la audiencia poblacional media y otro para la varianza poblacional, bajo la hipótesis de que la población de audiencias sigue una ley normal.

1.2.1 Solución

Nos dicen que la población es normal así que todo es más fácil, Cargamos los datos de la muestra

```
muestra=c(521, 742, 593, 635, 788, 717, 606, 639, 666, 624)
muestra
```

```
##  [1] 521 742 593 635 788 717 606 639 666 624
```

El intervalos para μ lo calcularemos con la función `t.test`

```
t.test(muestra,conf.level = 0.95)$conf.int
```

```
## [1] 597.1755 709.0245
## attr(,"conf.level")
## [1] 0.95
```

Con funciones básicas y utilizando la fórmula

$$\left(\overline{X}-t_{n-1,1-\frac{\alpha}{2}}\frac{\widetilde{S}_X}{\sqrt{n}},\overline{X}+t_{n-1,1-\frac{\alpha}{2}}\frac{\widetilde{S}_X}{\sqrt{n}}\right).$$

```
n=length(muestra)
n
```

```
## [1] 10
```

```
alpha=1-0.95
alpha
```

```
## [1] 0.05
```

```
media=mean(muestra)
media
```

```
## [1] 653.1
```

```
stilde=sd(muestra)
stilde
```

```
## [1] 78.17708
```

```
IC=c(media-qt(1-alpha/2,df=n-1)*stilde/sqrt(n),media+qt(1-alpha/2,df=n-1)*stilde/sqrt(n))
IC
```

```
## [1] 597.1755 709.0245
```

Lo calculamos con R

El intervalo para σ^2 es

$$\left(\frac{(n-1)\widetilde{S}_X^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}},\frac{(n-1)\widetilde{S}_X^2}{\chi^2_{n-1,\frac{\alpha}{2}}}\right),$$

Lo calculamos con R y el paquete `EnvStats`

```
#install.packages("EnvStats")
EnvStats::varTest(muestra,conf.level=0.95)$conf.int
```

```
##      LCL      UCL
##  2891.53 20369.25
## attr(,"conf.level")
## [1] 0.95
```

Si hacemos el cálculo con funciones básicas

```
n=length(muestra)
stilde=sd(muestra)
alpha=1-0.95
IC=c((n-1)*stilde^2/qchisq(1-alpha/2,df=n-1),(n-1)*stilde^2/qchisq(alpha/2,df=n-1))
IC
```

```
## [1]  2891.53 20369.25
```

obtenemos el mismo resultado:-).

1.3 Ejercicio 3

Supongamos que la empresa para la que trabajamos está en un proyecto de investigación, financiado con fondos de la Comunidad Europea, que pretende extender una nueva aplicación de las TIC. Una de las tareas del proyecto es realizar una encuesta de opinión sobre el grado de aceptación que tendría esta nueva tecnología en el mercado europeo. De entre todas las universidades y empresas participantes en el proyecto, es a tu empresa a la que le toca hacer el protocolo de la encuesta, llevarla a cabo y redactar esta parte del informe final. Como eres el último que llegó a la empresa y el resto de miembros del equipo no se acuerda de la estadística que vio en la carrera, te toca a ti cargar con la responsabilidad. Claro que el coste de la encuesta depende del número n de entrevistas que se realicen y el error de las proporciones de las contestaciones disminuye cuando n aumenta. Como no sabes cuánto dinero está dispuesto a gastar tu jefe, tabula los tamaños muestrales para los errores $\pm 5\%, \pm 3\%, \pm 2\%, \pm 1\%$, y para niveles de confianza 0.95 y 0.99, suponiendo el peor caso. Añade un comentario para que el equipo de dirección del proyecto, en el que hay componentes ignorantes en materia de encuestas, vea como quedarían redactado los datos técnicos de la encuesta, y pueda decidir el tamaño muestral leyendo tu informe.

1.3.1 Solución

El tamaño de la muestra en el pero de los casos $p = 0.5$ es

$$A_0 \geq 2 z_{1-\frac{\alpha}{2}} \sqrt{\frac{0.5^2}{n}} = \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \Rightarrow n \geq \left\lceil \frac{z^2_{1-\frac{\alpha}{2}}}{A_0^2} \right\rceil .$$

Donde $A_0 es la amplitud del intervalo que como sabemos es dos veces el error $A_0 = 2 \cdot \mathrm{error}$.

Definimos una función que calcule n en función del error

```
sample_size=function(error,p=0.5,alpha=1-0.95){
  A0=2*error
n=ceiling(qnorm(1-alpha/2)^2/A0^2)
n
}
```

Aplicamos la función

```
sample_size(c(0.05,0.03,0.02,0.01))
```

```
## [1]  385 1068 2401 9604
```

Podemos presentar los resultados en una bonita tabla para el nivel de confianza 0.95

```
tabla_n=data.frame(error=c(0.05,0.03,0.02,0.01),
                   n=sample_size(c(0.05,0.03,0.02,0.01),alpha=1-0.95), conf.level=0.95)
tabla_n
```

```
##   error    n conf.level
## 1  0.05  385       0.95
## 2  0.03 1068       0.95
## 3  0.02 2401       0.95
## 4  0.01 9604       0.95
```

y otra bonita tabla para el nivel de confianza 0.99

```
tabla_n=data.frame(error=c(0.05,0.03,0.02,0.01),
                   n=sample_size(c(0.05,0.03,0.02,0.01),alpha=1-0.99), conf.level=0.99)
tabla_n
```

```
##   error     n conf.level
## 1  0.05   664       0.99
## 2  0.03  1844       0.99
## 3  0.02  4147       0.99
## 4  0.01 16588       0.99
```

1.4 Ejercicio 4

El número de reservas semanales de billetes de cierto vuelo de una compañía aérea sigue una distribución aproximadamente normal. Se toma un muestra aleatoria de 81 observaciones de números de reservas de este vuelo: el número medio de reservas muestral resulta ser 112, mientras que la desviación típica muestral es 36. Además de estos 81 vuelos, 30 llegaron a su destino con un retraso de más de 15 minutos.

1. Calcular un intervalo de confianza del 95% para el número medio poblacional de reservas en este vuelo.
2. Calcular un intervalo de confianza de 95% para la varianza poblacional de las reservas.
3. Calcular un intervalo de confianza del 95% para la proporción poblacional de vuelos que llegan con un retraso de más de 15 minutos.
4. Calcular el tamaño muestral que asegura un intervalo de confianza de amplitud 0.1 para la proporción de vuelos que llegan con un retraso de más de 15 minutos al nivel de confianza 95%.

1.4.1 Solución

1. $(104.16, 119.84)$.
2. $(972.343, 1814.08)$.
3. $(0.265, 0.475)$.
4. $n = 385$.

Se pueden obtener con el siguiente código

```
n=81
xmedia=112
s=36
alpha=1-0.95
x=30# retraso
phat=30/81# proporción muestral retraso
A0=0.1
#apartado1
IC1=c(xmedia-qnorm(1-alpha/2)*s/sqrt(n),xmedia+qnorm(1-alpha/2)*s/sqrt(n))
IC1
```

```
## [1] 104.1601 119.8399
```

```r
#apartado2
IC2=c((n-1)*s^2/qchisq(1-alpha/2,n-1),(n-1)*s^2/qchisq(alpha/2,n-1))
IC2
```

```
## [1]  972.3473 1814.0725
```

```r
#apartado3
IC3=c(phat-qnorm(1-alpha/2)*sqrt(phat*(1-phat)/n),phat+qnorm(1-alpha/2)*sqrt(phat*(1-phat)/n))
IC3
```

```
## [1] 0.2652066 0.4755342
```

```r
#apartado4
n=ceiling(qnorm(1-alpha/2)^2/A0^2)
n
```

```
## [1] 385
```

1.5 Ejercicio 5

Una empresa cervecera sabe que las cantidades de cerveza que contienen sus latas sigue una distribución normal con desviación típica poblacional 0.03 litros.

1. Se extrae una muestra aleatoria de 25 latas y, a partir de la misma, un experto en estadística construye un intervalo de confianza para la media poblacional del contenido en litros de las latas que discurre entre 0.32 y 0.34 ¿Cuál es el nivel de confianza de este intervalo?
2. Un gerente de esta empresa exige un intervalo de confianza del 99% que tenga una amplitud máxima de 0.03 litros a cada lado de la media muestral ¿Cuántas observaciones son necesarias, como mínimo, para alcanzar este objetivo?

1.5.1 Solución

1. 90.44%
2. $n = 7$ hemos consideramos una "amplitud"/"error" a cada lado de la media como 0.03. Luego la amlitud total esintervalo es de $A0 = 2 * 0.03 = 0.06$.

Se pueden obtener con el siguiente código

```r
#apartado 1
sigma=0.03
n=25
IC=c(0.32,0.34)
A0=(IC[2]-IC[1])# amplitud A0=2*cuantil*sigma/sqrt(n);
#cuantil=qnorm(1-alpha/2)
A0
```

```
## [1] 0.02
```

```r
cuantil=A0*sqrt(n)/(2*sigma)
cuantil
```

```
## [1] 1.666667
```

```r
nivel_confianza=1-2*(1-pnorm(cuantil))
nivel_confianza
```

```
## [1] 0.9044193
```

```r
round(nivel_confianza*100,2)
```

```
## [1] 90.44
```

```r
#apartado2
A0=0.03*2
alpha=1-0.99
n=ceiling((2*qnorm(1-alpha/2)*sigma/A0)^2)
n
```

```
## [1] 7
```