Inferencia Estadística

Talller 1 Corte 2

Sebastián López Montenegro (97500) - Juan Pablo Silvestre (99127)

2024-09-24

Punto 1

Un estudio de una empresa de software monitorea el tiempo en minutos que tardan los empleados en resolver un problema técnico utilizando una nueva plataforma. El tiempo de resolución t sigue una distribución normal y depende de la complejidad del problema. Se propone que la complejidad del problema está relacionada con una variable x que mide el número de componentes afectados en el sistema, y que el tiempo de resolución se modela como:

$$t(x) = 12 + 3x + \epsilon$$

Donde ϵ es un término de error normalmente distribuido con media cero y desviación estándar desconocida. Se registraron los siguientes tiempos de resolución para problemas con distintos niveles de complejidad x:

x	**Tiempo (min)**
1	15
2	18
3	22
1	14
3	24
2	19
1	13
2	20
3	21

Con estos datos, estima un intervalo de confianza del 95 para la media del tiempo de resolución en problemas con 2 componentes afectados ($x = 2$)

Definición de parametros necesarios

Se requiere hacer el cálculo de el intervalo de confianza, sin embargo, en este caso contamos con que el número de elementos es menor a 25 y tampoco se conoce la desviación estandar, por lo que se debe hacer la aproximación del intervalo con t-student.

```
datos=c(18,19,20)
n=length(datos)
alpha=0.05
x.bar=mean(datos)
s=sd(datos)
```

Se definieron las variables: datos (conjunto de datos donde $x = 2$), n (tamaño de la muestra), $\alpha = 0.05$ (nivel de significancia del 5%), $\bar{x}$ (media muestral de los datos) y S (desviación estándar de los datos).

```
ME.t=qt(1-(alpha/2),n-1)*s/sqrt(n)
ic.t=c(x.bar-ME.t,x.bar+ME.t)
cat("Intervalo de confianza: [",ic.t,"] minutos con una confianza de",(1-alpha)*100,"%")
```

```
## Intervalo de confianza: [ 16.51586 21.48414 ] minutos con una confianza de 95 %
```

Luego de calcular el margen de error, se halla intervalo de confianza, esto siguiendo las siguientes formulas:

Fórmula Margen de Error

$$ME \rightarrow \frac{t_{(1-\alpha/2)}\hat{\sigma}}{\sqrt{n}}$$

Fórmula Intervalo de confianza (para μ)

$$\mu \rightarrow \left[\bar{x} - \frac{t_{(1-\alpha/2)}\hat{\sigma}}{\sqrt{n}}, \bar{x} + \frac{t_{(1-\alpha/2)}\hat{\sigma}}{\sqrt{n}}\right]$$

Con un nivel de confianza del 95% (con un nivel de significancia del 5%) hay evidencia muestral suficiente para determinar el IC para la media real/poblacional del tiempo que toma resolver problemas con complejidad 2, en minutos. Este intervalo será [16.5158 21.4841] Minutos.

Punto 2

Una compañía de telecomunicaciones realiza una encuesta para conocer la satisfacción de sus clientes con el servicio de Internet en diferentes ciudades. La encuesta se realizó en tres ciudades con las siguientes proporciones de clientes satisfechos:

- Ciudad A: 320 de 500 clientes satisfechos.
- Ciudad B: 410 de 600 clientes satisfechos.
- Ciudad C: 150 de 300 clientes satisfechos.

Se desea calcular el intervalo de confianza del 98% para la proporción de clientes satisfechos en cada una de las ciudades. Además, ¿es posible concluir que existe una diferencia estadísticamente significativa entre la satisfacción en la Ciudad A y la Ciudad C?

Función para calcular el intervalo de confianza (Proporción)

Como se requiere hacer el cálculo del intervalo de confianza para varias ciudades, primero se creará una función para usar con respecto a los datos de cada ciudad

```
calcularIntervaloP=function(n,c,alpha=0.02){
  pg=c/n
  cota=qnorm(1-(alpha/2))*sqrt(pg*(1-pg)/n)
  cotSup=pg+cota
  cotInf=pg-cota
  cat("Intervalo de confianza: [",cotInf,",",cotSup,"], con una confianza de",(1-alpha)*100,"%")
}
```

En esta función se calcula el valor de pg ($\hat{p}$) y luego se halla el margen de error para las cotas del intervalo teniendo en cuenta que para este caso el intervalo sigue la fórmula:

Fórmula Intervalo de confianza (Proporción)

$$P \rightarrow \left[\hat{p} - Z_{(1-\alpha/2)}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + Z_{(1-\alpha/2)}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

Cálculo intervalos de confianza

Ciudad A

```
n=500
CS=320
calcularIntervaloP(n,CS)
```

```
## Intervalo de confianza: [ 0.590062 , 0.689938 ], con una confianza de 98 %
```

Ciudad B

```
n=600
CS=410
calcularIntervaloP(n,CS)
```

```
## Intervalo de confianza: [ 0.6391543 , 0.7275124 ], con una confianza de 98 %
```

Ciudad C

```
n=300
CS=150
calcularIntervaloP(n,CS)
```

```
## Intervalo de confianza: [ 0.4328441 , 0.5671559 ], con una confianza de 98 %
```

Conclusión Para una confianza de 98% se concluyó que:

Para responder a la pregunta se debe observar qué la ciudad A toma un intervalo [0.59 , 0.68] y la ciudad C toma el intervalo [0.43,0.56], por lo cual se puede suponer que existe una diferencia notoria ya que están separados, prueba de esto es que si se realiza la intersección de ambos intervalos esta será nula ó vacía, esto tiene sentido incluso si se mira que en las proporciones de las ciudades, la ciudad A tiene más de la mitad de clientes satisfechos, mientras que en la C son sólo la mitad de clientes satifechos. En el mejor de los casos para que la direfencia sea la minima, sería que en la ciudad A tuviese una satisfacción de 0.59 y la ciudad C una satisfacción de 0.56, teniendo entonces solo un 3% de diferencia, aproximadamente.

Punto 3

Datos

Se creará un vector con los datos de las pruebas de resistencia realizadas.

```
pruebas=c(78.4, 80.2, 79.0, 77.5, 81.6, 79.8, 82.0, 80.5, 79.2, 78.1,
77.9, 80.3, 81.1, 79.5, 80.9, 77.8, 78.6, 80.7, 81.4, 82.1,
79.3, 78.9, 79.4, 77.6, 80.0, 81.8, 79.6, 80.4, 78.0, 82.3)
```

```
n=length(pruebas)
alpha=0.1
desv.est=sd(pruebas)

q1=qchisq(1-(alpha/2),n-1)
q2=qchisq(alpha/2,n-1)
a=(n-1)*(desv.est**2)/q1
b=(n-1)*(desv.est**2)/q2
cat("Intervalo de confianza: [",a,",",b,"], con una confianza de",(1-alpha)*100,"%")
```

```
## Intervalo de confianza: [ 1.4289 , 3.433951 ], con una confianza de 90 %
```

Fórmula Intervalo de Confianza (Varianza)

$$\sigma^2 \to \left[\frac{(n-1)\hat{\sigma}^2}{X^2_{(\alpha/2,n-1)}}, \frac{(n-1)\hat{\sigma}^2}{X^2_{(1-\alpha/2,n-1)}}\right]$$

Con una confianza del 90% la varianza real de la resistencia de los envases estará entre [1.42 , 3.43] megapascales cuadrados. Se puede identificar que en el mejor de los casos tiene una ligera variabilidad, sin embargo, se hará énfasis en el peor de los casos. La vairanza en el peor de los casos es 3.43 lo que su pone una varianza más significativa, si esta variabilidad está dentro de los límites para los que se requiera usar el material se podría usar estas aleaciones, en caso contrario, si se requiere mas uniformidad en los materiales se debe estar preocupado por mejorar en calidad para que la resistencia sea más homogénea.

Punto 4

Se desea estimar el ingreso promedio de los habitantes de una región para desarrollar un programa de apoyo económico. A partir de un estudio preliminar, se estima que la desviación estándar del ingreso es de 4,500 dolares. El gobierno quiere que el intervalo de confianza del 95 % tenga un error máximo de 500 dolares. ¿Cuál debe ser el tamaño mínimo de la muestra para realizar esta estimación?

Datos

```r
S = 4500
alpha = 0.05
ME.ast = 500

n.ast = ceiling((qnorm(1 - (alpha / 2)) * S / ME.ast)^2)
cat("El tamaño mínimo de la muestra para realizar la estimación
    (error máximo de 500 dólares) es de:", n.ast, "\n")
```

```
## El tamaño mínimo de la muestra para realizar la estimación
##     (error máximo de 500 dólares) es de: 312
```

Fórmula después de despejar para n (μ)

$$n = \left(\frac{z_{(1-\alpha/2)}\sigma}{ME}\right)^2$$

Si luego de recolectar los datos se obtiene una muestra con los siguientes ingresos (en miles de dólares), calcula el intervalo de confianza del 95 % para la media:

```r
data = c(38.5, 42.1, 40.3, 37.8, 45.6, 41.2, 39.7, 43.5, 40.0,
         39.8, 38.9, 44.1, 40.6, 37.9, 41.7, 42.5, 39.1, 40.9, 41.4, 43.0)
media = mean(data)
n = length(data)

S_data=sd(data)

ME = qnorm(1 - (alpha / 2)) * S_data / sqrt(n)

IC = c(media - ME, media + ME)
```

```r
cat("El intervalo de confianza es: [", IC, "] (En miles de
dólares)")
```

```
## El intervalo de confianza es: [ 40.0086 41.8514 ] (En miles de
## dólares)
```

Fórmula Intervalo de Confianza (μ)

$$\mu \to \left[\bar{x} - \frac{t_{(1-\alpha/2)}\hat{\sigma}}{\sqrt{n}}, \bar{x} + \frac{t_{(1-\alpha/2)}\hat{\sigma}}{\sqrt{n}}\right]$$

Dado el nivel de confianza del 95% (nivel de significancia del 5%), el ingreso promedio poblacional/real de la región encuestada estará entre [40.0086 41.8514] miles de dólares.

Punto 5

Un instituto de investigación está realizando un estudio para determinar la proporción de personas que apoyan una nueva ley de protección ambiental. En un estudio piloto, se entrevistó a 150 personas y 75 de ellas se mostraron a favor de la ley.

El instituto desea ahora realizar una encuesta más grande con un margen de error del 2% y un nivel de confianza del 99%. ¿Cuál debería ser el tamaño mínimo de la muestra para la nueva encuesta?

Datos

```r
alpha = 0.01
p.gorro = 75 / 150
n = 150
me.ast = 0.02

n.ast = ceiling( p.gorro * (1 - p.gorro) * (qnorm(1 - alpha / 2) / me.ast) ^ 2 )

cat("El tamaño minimo de la muestra debe ser:", n.ast)
```

```
## El tamaño minimo de la muestra debe ser: 4147
```

Fórmula después de despejar para n(Proporción)

$$n = \hat{p}(1 - \hat{p})\left(\frac{z_{(1-\alpha/2)}}{ME}\right)^2$$

Una vez que se ha recolectado una muestra de 1200 personas, de las cuales 720 apoyan la ley, calcula el intervalo de confianza del 99 % para la proporción de apoyo a la ley.

```r
alpha = 0.01
p.gorro = 720 / 1200
n = 1200

ME = qnorm(1 - alpha / 2) * sqrt((p.gorro * (1 - p.gorro)) / n)
IC = c(p.gorro - ME, p.gorro + ME)

print(ME)
```

```
## [1] 0.03642773
```

```r
print(IC)
```

```
## [1] 0.5635723 0.6364277
```

Fórmula Intervalo de Confianza (Proporción)

$$P \rightarrow \left[\hat{p} - Z_{(1-\alpha/2)}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + Z_{(1-\alpha/2)}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

Dado el nivel de confianza del 99% (nivel significancia del 1%), la proporción real/poblacional de las personas que apoyan la ley es [56.35 63.64] %. Esto significa que hay evidencia suficiente para afirmar que más del 50% de las personas apoyan la ley.

Punto 6

Una empresa farmacéutica está probando la efectividad de un nuevo medicamento para reducir el colesterol en personas mayores de 50 años. Se seleccionaron 50 personas que tomaron el medicamento durante 6 meses, y se registraron las siguientes reducciones en los niveles de colesterol (en mg/dL):

Datos

```r
data = c(
18, 25, 22, 19, 30, 27, 20, 28, 23, 21, 24, 26, 31, 19, 29, 21, 22, 27, 25, 20,
23, 28, 24, 19, 30, 26, 27, 22, 21, 29, 18, 31, 25, 20, 23, 19, 27, 22, 24, 28,
26, 20, 23, 21, 29, 24, 19, 30, 22, 28)
```

Asumiendo que las reducciones siguen una distribución normal, calcula el intervalo de confianza del 95 % para la media de la reducción de colesterol.

```r
#Intervalo de Confianza por aproximacion normal

alpha = 0.05
n = length(data)
x.bar = mean(data)
S = sd(data)

#margen de error
me.n = qnorm(1 - (alpha / 2)) * S / sqrt(n)

IC.n = c(x.bar - me.n, x.bar + me.n)

print(IC.n)
```

```
## [1] 23.04008 25.15992
```

Fórmula Intervalo de Confianza (μ)

$$\mu \rightarrow \left[\bar{x} - \frac{t_{(1-\alpha/2)}\hat{\sigma}}{\sqrt{n}}, \bar{x} + \frac{t_{(1-\alpha/2)}\hat{\sigma}}{\sqrt{n}}\right]$$

Dado el nivel de confianza del 95% (nivel de significancia del 5%), la media/real de la reducción de colesterol en personas mayores de 50 años que tomen el medicamento durante 6 meses, estará entre [23.04008 25.15992] mg/dL, esta información es utilizada para concluir que la empresa farmacéutica logró que el medicamento sea efectivo.