

Tarea N°2

Daniel Felipe Villa

2022-03-30

Caso de Estudio N°3:

Habiendo hecho esto, procederemos con el Intervalo de Confianza (IC)

Contexto: Suponga que se desea realizar un estudio de muestreo en un municipio A del departamento de Antioquia para estimar la **proporción de votantes** registrados con intención de voto por un candidato X . Para ello:

- Se dispone del listado de todos los habitantes mayores de edad que conforman el municipio, $N = 10000$ Habitantes.
- Se opta por seleccionar una MAS sin remplazo de dicha población, $n = 500$.
- Se encuentra que en dicha muestra sólo 350 personas estaban inscritas para votar y de éstas 150 estaban a favor del candidato X .

¿Cómo estimar la proporción de votantes registrados con intención de voto por el candidato X ?

Solución

extraemos los datos del enunciado anterior

N : 10.000 habitantes (*de la población total*)

n : 500 hab. tamaño de la muestra

N_k : Desconocido

a_k : 150 (*voto a favor del candidato X*)

n_k : 350

```
ak <- 150
nk <- 350
N <- 10000
n <- 500
```

$$\hat{p}_k = \frac{a_k}{n_k} = \frac{150}{350} = 0.4285714 \approx 0.43$$

```
# "p" gorro:
pg <- round(ak/nk,2)
```

I.C.

Habiendo hecho esto, procederemos con el Intervalo de Confianza (IC)

Como N_k es desconocido utilizare:

$$\hat{Var}[\hat{p}_k] = \frac{N-n}{N} * \frac{\hat{p}_k(1-\hat{p}_k)}{n_k-1}$$

(1)

```
# Varianza estimada:
var_p <- (N-n)/N * pg*(1-pg)/(nk-1)
print("# Varianza estimada:")
```

```
## [1] "# Varianza estimada:"
```

```
print(var_p)
```

```
## [1] 0.0006671777
```

un intervalo al 95% esta dado por:

$$\hat{p}_k \pm t_{1-\alpha/2, n_k-1} \sqrt{\hat{V}(\hat{p}_k)}$$

```
## [1] "# IC para P_k"
```

```
## [1] "#           Limite Inf ,   Limite Sup"
```

```
## [1] "( 0.379198383195855 , 0.480801616804145 )"
```

Con esto vemos un IC al 95% de la proporción de votantes que tienen intención de voto por el candidato X.

Caso de Estudio N°4:

Contexto: Una multinacional de abrir nuevos puestos de trabajo en un barrio de Medellín.

para ello necesitas estimar de las personas que **NO trabajan**, el tiempo en (meses) que el jefe de hogar ha completado sin trabajar.

La empresa cuenta con lista de todos los hogares del barrio bajo estudio, conformado por 1000 hogares. se decide:

- seleccionar una muestra piloto 10 hogares y se entrevista al jefe de hogar

los datos que se obtuvieron fueron los siguientes:

Hogar	1	2	3	4	5	6	7	8	9	10
Trabaja (1: SI 0: NO)	1	0	1	0	0	1	0	1	0	0
Tiempo (en meses)		3		12	5		7		8	2

¿Como estimar la **proporción** de hogares donde el jefe de hogar *NO* trabaja?

¿Como estimar el **tiempo promedio** en meses, de los jefes de hogar que *NO* han tenido trabajo?

Solución:

Primera Pregunta:

Extraemos los datos del enunciado:

```
#Valores necesarios:
```

```
nk <- 6
```

```
N <- 1000
```

```
n <- 10
```

```
# proporción estimada
```

```
pg <- nk/n
```

```
pg
```

```
## [1] 0.6
```

```
#Tiempo en meses sin trabajar
```

```
tm <- c(3,12,5,7,8,2)
```

Para hallar IC, primero hallaremos el $\hat{V}(\hat{p})$

Nota: es la misma formula anterior (1)

```
# Varianza estimada:
```

```
var_p <- (N-n)/N * pg*(1-pg)/(n-1)
```

```
print("# Varianza estimada:")
```

```
## [1] "# Varianza estimada:"
```

```
print(var_p)
```

```
## [1] 0.0264
```

IC por notación calcularemos el β (error de estimación) *(esto con el fin de que no sea tan repetitivo todo)*

$$\beta = t_{\alpha/2, n-1} \sqrt{\frac{\hat{p}(1-\hat{p})}{n-1} \frac{N-n}{N}}$$

```
# error de estimación:
b <- qt(0.025,n, lower.tail = F)*sqrt(var_p)
b
```

```
## [1] 0.3620297
```

```
#Calculando el IC
print("#      Limite Inf ,   Limite Sup")
```

```
## [1] "#      Limite Inf ,   Limite Sup"
```

```
paste("(", pg-b, ",", pg+b, ")")
```

```
## [1] "( 0.237970287912041 , 0.96202971208796 )"
```

Según el IC decimos con un nivel de confianza del 95% en el barrio muestreado de Medellín los jefes de hogar que no trabajan estan dentro del 24% y el 96%; esto es preocupante ya que deja mucho que desear acerca de esta población, haciéndonos preguntar...¿cómo generan ingresos para el hogar?

Segunda pregunta

Estimar la media:

$$\hat{\mu} = \bar{y}_k = \frac{1}{n} \sum y_{ki}$$

```
#Media estimada
yk <- mean(tm)
yk
```

```
## [1] 6.166667
```

Ahora estimamos la varianza de los datos de interés:

$$S_k^2 = \frac{1}{n-1} \sum (y_{ki} - \bar{y})^2$$

```
# varianza estimada
sk <- var(tm)
sk
```

```
## [1] 13.36667
```

Por ultimo calculo la varianza de la media estimada:

$$\hat{Var}(\bar{y}) = \frac{N-n}{N} \frac{S_k^2}{n_k}$$

Se utilizo una función creada para futuros trabajos tenerla presente y solamente llamarla cuando sea necesario:

```
#Funcion de varianza estimada de la media:
varmu <- function(N,n,sk,nk){
  "Ya que no se conoce Nk..."
  (N-n)/N * (sk)/nk
}
```

```
## [1] "# Calculando la varianza estimada de la media con Nk desconocida: "
```

```
## [1] 2.2055
```

IC Un IC al 95% para \bar{y} (promedio de meses sin trabajar) estan dados por:

$$\bar{y} \pm t_{1-\alpha/2, nk-1} \sqrt{\hat{V}(\bar{y})}$$

```
LI <- yk - qt(0.025,nk-1,lower.tail = F)*sqrt(varyk)
```

```
LU <- yk + qt(0.025,nk-1,lower.tail = F)*sqrt(varyk)
```

```
print("# IC para y_k")
```

```
## [1] "# IC para y_k"
```

```
print("#      Limite Inf ,      Limite Sup")
```

```
## [1] "#      Limite Inf ,      Limite Sup"
```

```
paste("(", LI, ", ", LU, ")")
```

```
## [1] "( 2.34911463133005 , 9.98421870200329 )"
```

Podemos afirmar con un $\alpha = 0.05$ que los tiempos promedios en los cuales no han tenido ningún tipo de trabajo los responsables del hogar entre 3 a 10 meses; unido con lo anterior me deja pensando como hacen para subsistir estas familias...¿dependerán completamente del ingreso solidario del gobierno?