Universidad de Alcalá

Estadística

Práctica 7

Intervalos de confianza para parámetros

Contenido

1. Introducción

Con esta práctica se utiliza R y RStudio para estimar (inferir) con un nivel de confianza determinado, el intervalo en el que se encuentra el valor de los parámetros de una población (media, varianza y proporción) a partir de una muestra de la misma.

En estadística hay que diferenciar entre conceptos “parámetros” y “estadísticos”. El término “parámetro” suele utilizarse en referencia a una población mientras que el término “estadístico” suele utilizarse en referencia a una muestra extraída de dicha población. En la siguiente tabla se muestra la nomenclatura utilizada habitualmente.

PARÁMETROS (POBLACIONALES)	**ESTADÍSTICOS (MUESTRALES)**
Media poblacional (μ)	Media muestral ($\bar{x}$)
Desviación estándar poblacional (σ)	Desviación estándar muestral (s)
Varianza poblacional (σ^2)	Varianza muestral (s^2)
Tamaño población (N)	Tamaño muestra (n)
Proporción poblacional (p)	Proporción muestral ($\hat{p}$)

Se puede calcular el intervalo en el que se puede asegurar que se encuentra el valor de un parámetro poblacional con un determinado nivel de confianza, obteniendo lo que se conoce como Intervalo de Confianza, con un límite inferior `Li` y un límite superior `Ls`:

Intervalo de confianza con nivel de confianza $1 - \alpha = (Li, Ls)$

El nivel de confianza $(1 - \alpha)$ es un valor de 0 a 1 y representa la probabilidad de acertar en la predicción, es decir la probabilidad de que el valor real del parámetro se encuentre entre los límites del intervalo. Se suele utilizar en % multiplicando por 100. En la expresión del nivel de confianza aparece un valor α (alfa) que se denomina nivel de error o nivel de significación, y representa lo contrario que la confianza, es decir, la probabilidad de no acertar en la predicción.

En la siguiente tabla se indican las fórmulas para calcular los parámetros poblacionales y su posible programación utilizando el lenguaje R.

Parámetro a calcular	Se conoce la varianza pobla-cional (σ^2)	Intervalo de confianza *(Li, Ls)*	Código R
Media poblacional (μ)	Si	$Li = \bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ $Ls = \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$	n=tamaño muestra z=qnorm(1-alfa/2,0,1) m=media muestral sigma=desviación poblacional Li=m-z*(sigma/sqrt(n)) Ls=m+z*(sigma/sqrt(n))
	No $n \geq 30$	$Li = \bar{x} - z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$ $Ls = \bar{x} + z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$	n=tamaño muestra z=qnorm(1-alfa/2,0,1) m=media muestral s=desviación muestral Li=m-z*(s/sqrt(n)) Ls=m+z*(s/sqrt(n))
	No $n < 30$ Distribución Normal	$Li = \bar{x} - t_{\frac{\alpha}{2},n-1} \frac{s}{\sqrt{n}}$ $Ls = \bar{x} + t_{\frac{\alpha}{2},n-1} \frac{s}{\sqrt{n}}$	n=tamaño muestra t=qt(1-alfa/2,n-1) m=media muestral s=desviación muestral Li=m-t*(s/sqrt(n)) Ls=m+t*(s/sqrt(n))
Varianza poblacional (σ^2)	No	$Li = \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},n-1}}$ $Ls = \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}$	n=tamaño muestra chi.li=qchisq(1-alfa/2,n-1) chi.ls=qchisq(alfa/2,n-1) s=desviación muestral Li=((n-1)*s^2)/chi.li Ls=((n-1)*s^2)/chi.ls
Proporción poblacional (p)	---	$L_i = \hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ $L_s = \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	n=tamaño muestra z=qnorm(1-alfa/2,0,1) p=proporción Li=p-z*sqrt(p*(1-p)/n) Li=p+z*sqrt(p*(1-p)/n)

Existen en R funciones predefinidas que se pueden utilizar para los cálculos de algunos intervalos sin tener que programar las fórmulas de la tabla anterior.

Son las funciones predefinidas `t.test()` y `prop.test()`, y la función `z.test()` del paquete `BSDA`, que se pueden utilizar como se indica en la siguiente tabla.

Parámetro a calcular	**Código R**	**Ejemplo**

Media poblacional – muestra grande (μ)	`test= z.test(muestra, mu=media muestral, sigma.x=desviación típica muestral, conf.level = 0.95)`
	
	Intervalo:
	`test$conf.int`
		`> muestra=rnorm(50,5,1)`
		
		`> prueba=z.test(muestra, mu=5, sigma.x=1, conf.level=0.95)`
		
		`> prueba$conf.int`
		`[1] 4.713116 5.267477`

Media poblacional – muestra pequeña (μ)	`test=t.test(muestra, conf.level = 0.95)`
	
	Intervalo:
	`test$conf.int`
		`> muestra=c(5,5,4,6,4,5,5,3,3,6)`
		
		`> prueba=t.test(muestra, conf.level = 0.95)`
		
		`> prueba$conf.int`
		`[1] 3.831014 5.368986`

Proporción poblacional (p)	`test=prop.test(ng, n, conf.level = 0.95, correct=FALSE)`
	
	Intervalo:
	`test$conf.int`
	
	NOTA: `n` es el tamaño de la muestra.
	`ng` es el tamaño del grupo de datos de la muestra para el que se calcula la proporción, es decir la proporción muestral `p` sería `ng/n`.
		`> prueba=prop.test(70,100, conf.level = 0.95, correct=FALSE)`
		
		`> prueba$conf.int`
		`[1] 0.6041515 0.7810511`

2. Ejemplo: Intervalo de confianza para la media de calificaciones de estudiantes (muestra grande)

ENUNCIADO

Mediante una encuesta, se saben las calificaciones de acceso a la universidad de 74 estudiantes de la asignatura Estadística del Grado en Ingeniería en Sistemas de Información de la Universidad de Alcalá del curso 2021-22, de una población de 108 matriculados. Si se supone que las calificaciones tienen una distribución Normal, responder a las siguientes preguntas:

a) **Calcular los intervalos de confianza para la media poblacional con niveles de confianza del 90%, 95% y 99%, aplicando las fórmulas.**
b) **Calcular los intervalos de confianza para la media poblacional con niveles de confianza del 90%, 95% y 99%, utilizando la función `z.test` del paquete BSDA.**
c) **Dibujar una gráfica con los tres intervalos de confianza para compararlos visualmente.**

SOLUCIÓN

Primero hay que leer los datos de las notas, disponibles en el archivo encuesta.csv:

```
> encuesta = read.csv2("encuesta.csv")
> (nota=encuesta$NOTA)
 [1]  8.500  7.100  8.635  8.624  8.200  8.700  7.210  7.630  8.400  8.300
[11]  8.200  9.100  9.789 10.107  8.015  7.310  7.500  8.710  8.336  9.210
[21]  7.800 10.300  7.990  6.900  7.800 10.000  8.590  7.000  8.050 10.799
[31]  7.994  8.550  7.340  6.750  9.560  7.417  6.936  7.210  7.680 10.277
[41]  7.860 10.260  7.270  5.800  7.300  7.140  8.600  7.500  8.000  7.540
[51]  7.292  7.830  6.750  9.806  6.800  6.445  6.650  7.804 10.270  7.600
[61]  7.870  7.000  7.085  7.480  8.070  5.820  6.500  9.900  7.500  6.500
[71]  9.456  8.000  7.800  7.654
```

a) Calcular los intervalos de confianza para la media poblacional con niveles de confianza del 90%, 95% y 99%.

No se conoce la varianza poblacional y es una muestra grande (74 >30), por lo que se aplican las fórmulas:

$$Li = \bar{x} - z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

$$Ls = \bar{x} + z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

Los comandos en R serían los siguientes:

```
> n=length(nota)
> m=mean(nota)
> s=sd(nota)
```

```
> #nivel de confianza del 90%
> alfa=1-0.9
> z=qnorm(1-alfa/2,0,1)
> (Li90=m-z*(s/sqrt(n)))
[1] 7.806304
> (Ls90=m+z*(s/sqrt(n)))
[1] 8.238858

> #nivel de confianza del 95%
> alfa=1-0.95
> z=qnorm(1-alfa/2,0,1)
> (Li95=m-z*(s/sqrt(n)))
[1] 7.764871
> (Ls95=m+z*(s/sqrt(n)))
[1] 8.280292

> #nivel de confianza del 99%
> alfa=1-0.99
> z=qnorm(1-alfa/2,0,1)
> (Li99=m-z*(s/sqrt(n)))
[1] 7.683892
> (Ls99=m+z*(s/sqrt(n)))
[1] 8.36127
```

b) Calcular los intervalos de confianza para la media poblacional con niveles de confianza del 90%, 95% y 99%, utilizando la función `z.test` del paquete BSDA.

Con `z.test()` se obtienen los mismos valores.

```
> install.packages("BSDA")
> library(BSDA)

> test90=z.test(nota,mu=mean(nota),sigma.x=sd(nota),conf.level=0.90)
> test90$conf.int
[1] 7.806304 8.238858
attr(,"conf.level")
[1] 0.9
>
> test95=z.test(nota,mu=mean(nota),sigma.x=sd(nota),conf.level=0.95)
> test95$conf.int
[1] 7.764871 8.280292
attr(,"conf.level")
[1] 0.95
>
> test99=z.test(nota,mu=mean(nota),sigma.x=sd(nota),conf.level=0.99)
> test99$conf.int
[1] 7.683892 8.361270
attr(,"conf.level")
[1] 0.99
```

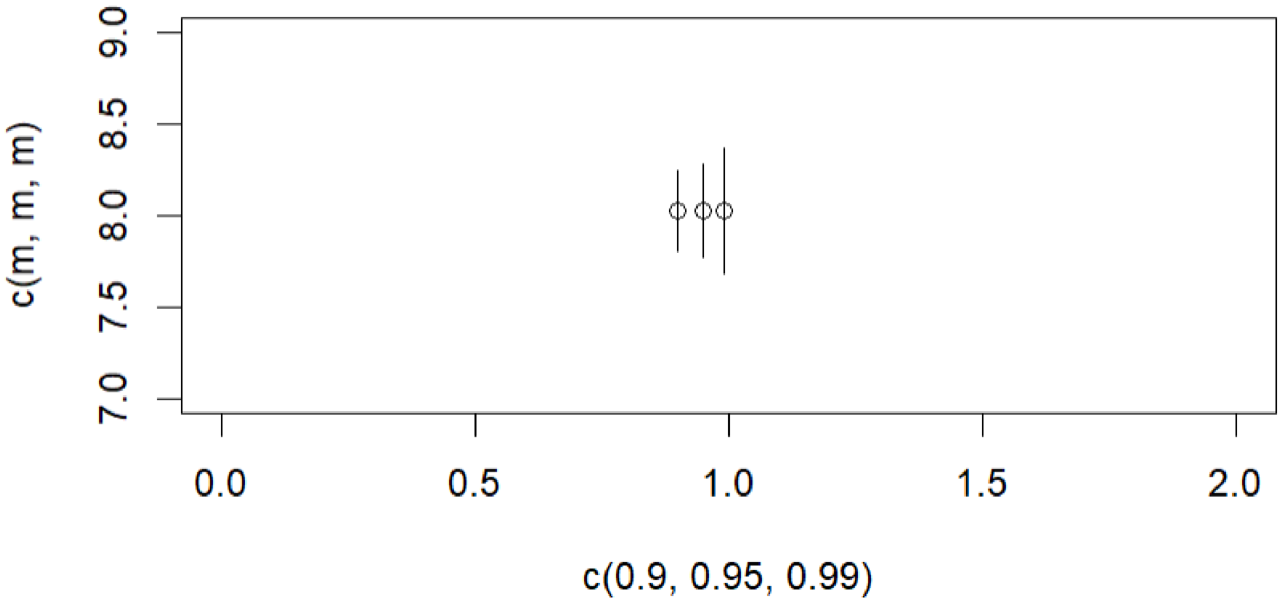
Podemos hacer una tabla para comparar los resultados, en la que se comprueba que cuanto mayor nivel de confianza menor precisión porque el intervalo es más amplio:

Nivel de confianza	Intervalo de confianza
90%	(7.806304, 8.238858)
95%	(7.764871, 8.280292)
99%	(7.683892, 8.361270)

c) Dibujar una gráfica con los tres intervalos de confianza para compararlos visualmente.

Podemos dibujar primero los puntos de la media en cada intervalo, y luego una línea centrada en cada punto desde el límite inferior al límite superior de cada intervalo.

```
> plot (c(0.90,0.95,0.99),c(m,m,m), xlim=c(0,2),ylim=c(7,9))
> lines(c(0.90,0.90), c(Li90,Ls90))
> lines(c(0.95,0.95), c(Li95,Ls95))
> lines(c(0.99,0.99), c(Li99,Ls99))
```



3. Ejemplo: Intervalo de confianza para la media de calificaciones de estudiantes (muestra pequeña)

ENUNCIADO

En el ejemplo anterior, considerar sólo como muestra las calificaciones de los 19 estudiantes del grupo A2 y responder a las siguientes preguntas:

- **a) Calcular el intervalo de confianza para la media poblacional con nivel de confianza del 95% aplicando las fórmulas.**
- **b) Calcular el intervalo de confianza para la media poblacional con nivel de confianza del 95% utilizando la función `t.test`.**
- **c) Representar gráficamente los intervalos de todos los grupos (A1, A2, B1, B2) para un nivel de confianza del 95%, utilizando la función `ggplot()` del paquete `ggplot2`.**

SOLUCIÓN

a) Calcular el intervalo de confianza para la media poblacional con nivel de confianza del 95% aplicando las fórmulas.

En este caso, al ser una muestra pequeña (19<30), se debe suponer que el estimador de la media se adapta mejor a una función de probabilidad T-Student con n-1 (19-1) grados de libertad en lugar de una Normal.

No se conoce la varianza poblacional y es una muestra pequeña, por lo que se aplican las fórmulas:

$$Li = \bar{x} - t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}}$$

$$Ls = \bar{x} + t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}}$$

Los comandos en R serían los siguientes:

```
> encuesta.A2=encuesta[encuesta$GRUPO=="A2",]
> (nota.A2=encuesta.A2$NOTA)
 [1]  6.900  7.800 10.000  8.590  7.000  8.050 10.799  7.994  8.550  7.340
[11]  6.750  9.560  7.417  6.936  7.210  7.680 10.277  7.860 10.260

> n=length(nota.A2)
> m=mean(nota.A2)
> s=sd(nota.A2)

> #nivel de confianza del 95%
> alfa=1-0.95
> t=qt(1-alfa/2,n-1)
> (Li=m-t*(s/sqrt(n)))
[1] 7.635589
> (Ls=m+t*(s/sqrt(n)))
```

```
[1] 8.887884
```

b) Calcular el intervalo de confianza para la media poblacional con nivel de confianza del 95% utilizando la función `t.test`.

Con `t.test()` se obtienen los mismos valores.

```
> test=t.test(nota.A2, conf.level = 0.95)
> test$conf.int
[1] 7.635589 8.887884
attr(,"conf.level")
[1] 0.95
```

Podemos hacer una tabla para comparar los resultados:

Nivel de confianza	Tamaño muestra	Intervalo de confianza
95%	74	(7.764871, 8.280292)
95%	19	(7.635589, 8.887884)

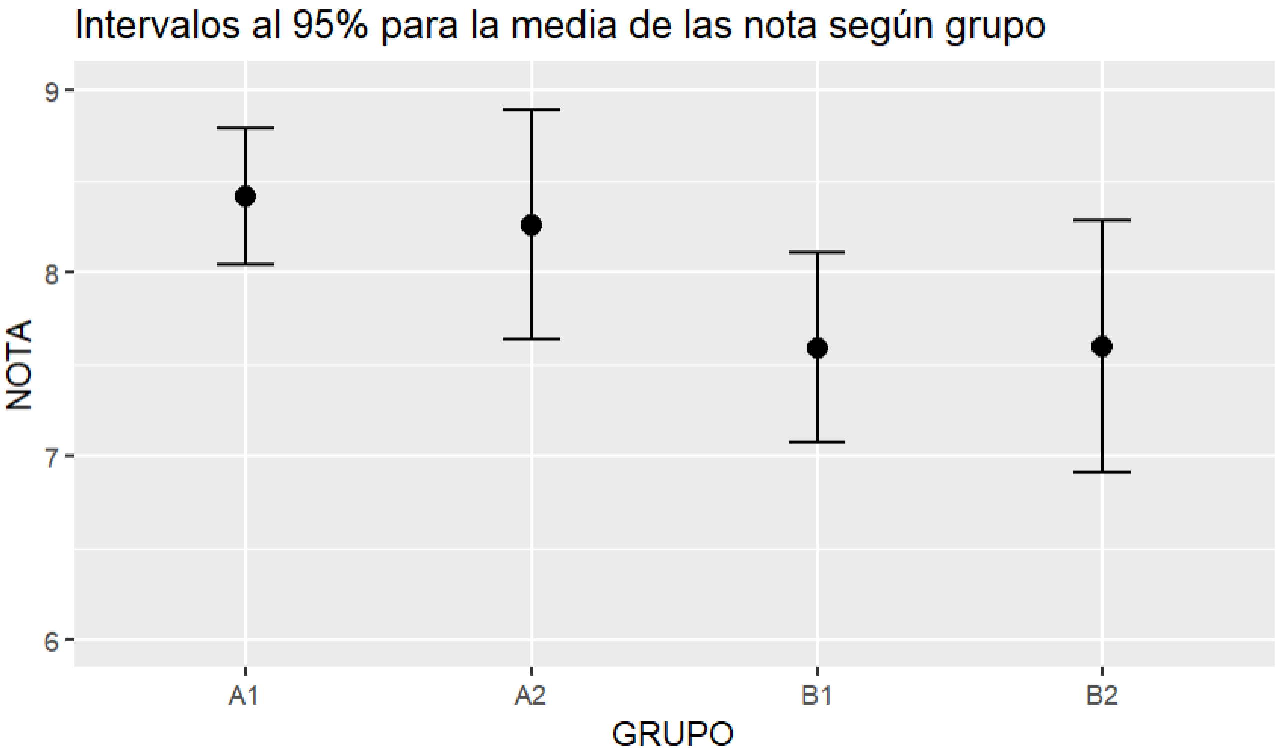
c) Representar gráficamente los intervalos de todos los grupos (A1, A2, B1, B2) para un nivel de confianza del 95%, utilizando la función `ggplot()` del paquete `ggplot2`.

Se podrían representar gráficamente los intervalos de todos los grupos (A1, A2, B1, B2) al 95% de confianza con el siguiente código[1]:

```
> install.packages("ggplot2")
> library(ggplot2)  #Para utilizar la función ggplot()
> install.packages("Rmisc")
> library(Rmisc)    #Para utilizar la función summarySE()

> (resumen.encuesta=summarySE(encuesta, measurevar="NOTA", groupvars="GRUPO"))
  GRUPO  N     NOTA        sd        se        ci
1    A1 23 8.420261 0.8623132 0.1798047 0.3728922
2    A2 19 8.261737 1.2991030 0.2980347 0.6261476
3    B1 19 7.593000 1.0718574 0.2459010 0.5166187
4    B2 13 7.597308 1.1327137 0.3141582 0.6844920

> g.puntos=ggplot(resumen.encuesta, aes(x=GRUPO, y=NOTA))+ geom_point(size = 3) +
ylim(6, 9)
> g.puntos

> g.intervalos = g.puntos +  geom_errorbar(aes(ymin=NOTA-ci, ymax=NOTA+ci), width
= 0.2) +
```

[1] Código adaptado de https://rstudio-pubs-static.s3.amazonaws.com/80947_18e337f559fd4ac594506e39901d445a.html

```
+   ggtitle("Intervalo al 95% para la media de las notas según grupo")
> g.intervalos
```



4. Ejemplo: Intervalo de confianza para las proporciones de estudiantes que tienen móvil Android

ENUNCIADO

Mediante una encuesta, se sabe que 50 de 74 estudiantes de la asignatura Estadística del Grado en Ingeniería en Sistemas de Información de la Universidad de Alcalá del curso 2021-22 tienen teléfonos móviles con sistema operativo Android. La población es de 108 estudiantes matriculados. Responder a las siguientes preguntas:

a) **Calcular los intervalos de confianza para la proporción poblacional de estudiantes que tienen móvil Android con un nivel de confianza del 90%, 95% y 99% usando las fórmulas.**
b) **Inferir el mínimo y máximo número de estudiantes de la población que tienen móvil Android con una confianza del 95%.**

SOLUCIÓN

a) Calcular los intervalos de confianza para la proporción poblacional de estudiantes que tienen móvil Android con un nivel de confianza del 90%, 95% y 99% usando las fórmulas.

```
> so=encuesta$SO
> n=length(so)
> so.android=so[so=="Android"]
> ng=length(so.android)
> (p=ng/n)
[1] 0.6756757

> # 90%
> alfa=1-0.90
> z=qnorm(1-alfa/2,0,1)
> Li90=p-z*sqrt(p*(1-p)/n)
> Ls90=p+z*sqrt(p*(1-p)/n)
> c(Li90,Ls90)
[1] 0.5861659 0.7651854

> # 95%
> alfa=1-0.95
> z=qnorm(1-alfa/2,0,1)
> Li95=p-z*sqrt(p*(1-p)/n)
> Ls95=p+z*sqrt(p*(1-p)/n)
> c(Li95,Ls95)
[1] 0.5690182 0.7823331

> # 99%
> alfa=1-0.99
> z=qnorm(1-alfa/2,0,1)
> Li99=p-z*sqrt(p*(1-p)/n)
> Ls99=p+z*sqrt(p*(1-p)/n)
> c(Li99,Ls99)
[1] 0.5355040 0.8158473
```

b) Inferir el mínimo y máximo número de estudiantes de la población que tienen móvil con Android con una confianza del 95%.

Como la población es de 108 estudiantes matriculados, sólo hay que multiplicar este valor por los límites del intervalo de confianza. Pero hay que redondear por defecto el valor mínimo (usando la función `floor`) y por exceso el máximo (usando la función `ceiling`).

```
> floor(108*Li95)
[1] 61
> ceiling(108*Ls95)
[1] 85
```

5. Ejercicios propuestos

1) Mediante una encuesta, se sabe el tiempo en minutos del viaje a la Escuela Politécnica de 74 estudiantes de la asignatura Estadística del Grado en Ingeniería en Sistemas de Información de la Universidad de Alcalá del curso 2021-22, de una población de 108 matriculados. Si se supone que tiene una distribución Normal, responder a las siguientes preguntas:
   a. Calcular los intervalos de confianza para la media poblacional con niveles de confianza del 90%, 95% y 99%, aplicando las fórmulas.
   b. Dibujar una gráfica con los tres intervalos de confianza para compararlos visualmente.

2) En caso del ejercicio anterior, considerar sólo como muestra los tiempos de viaje de los 19 estudiantes del grupo A2 y responder a las siguientes preguntas:
   a. Calcular el intervalo de confianza para la media poblacional con nivel de confianza del 95% aplicando las fórmulas.

3) Mediante una encuesta, se sabe que 26 estudiantes de una muestra de 74 de la asignatura Estadística del Grado en Ingeniería en Sistemas de Información de la Universidad de Alcalá del curso 2021-22 tienen una línea telefónica contratada con las compañías Jazzte/Orange/Simyo. La población es de 108 estudiantes matriculados. Responder a las siguientes preguntas:
   a. Calcular los intervalos de confianza para la proporción poblacional de estudiantes que tienen una línea telefónica contratada con las compañías Jazzte/Orange/Simyo con un nivel de confianza del 90%, 95% y 99%. NOTA: Utilizar las fórmulas.
   b. Estimar el mínimo y máximo número de estudiantes de la población que tienen una línea telefónica contratada con las compañías Jazzte/Orange/Simyo móvil con una confianza del 95%.