Intervalos de confianza

Caterine Melissa Guerrero España

2022-09-13

Introducción

el objetivo mediante la **intervalos de confianza** o **estimación confidencial** es la determinación de dos valores θ_1^* y θ_2^* que verifiquen $\theta_1^* \leq \theta_2^*$, tales que, al constituirse en intervalo θ_1^*, θ_2^* contengan, con una probabilidad prefijada, el verdadero valor del parámetro que deseamos estimar. De forma gráfica, un intervalo de confianza puede representarse del siguiente modo:

Supuesto Práctico 1

Con el fin de estudiar el número medio de flexiones continuadas que pueden realizar los alumnos, un profesor de educación física somete a 75 de ellos, elegidos aleatoriamente, a una prueba. El número de flexiones realizado por cada alumno, así como su sexo y si realizan o no deporte se muestran en el fichero Flexiones.txt.

Se sabe que el número de flexiones se distribuye según una Normal de varianza poblacional 7.5. ¿Determinar el intervalo de confianza a un nivel de confianza del 95% para el número medio de flexiones?

Solución En primer lugar debemos importar, en R, los datos que contienen el número de flexiones realizadas por cada alumno. Para ello, utilizamos la orden read.table

```
datos<-read.table("Flexiones.txt", header=TRUE)
datos
```

```
##    Flexiones Sexo Deporte
## 1         60    H       0
## 2         41    H       0
## 3         53    M       1
## 4         53    M       0
## 5         41    H       0
## 6         56    H       0
## 7         50    H       0
## 8         53    M       1
## 9         50    M       1
## 10        48    M       0
## 11        50    M       1
## 12        48    M       1
## 13        56    H       0
## 14        52    M       1
## 15        54    M       0
## 16        50    H       1
## 17        50    H       0
## 18        54    H       0
## 19        52    H       1
## 20        48    H       0
```

```
## 21          48    H        1
## 22          35    M        1
## 23          50    M        1
## 24          41    M        1
## 25          56    M        1
## 26          52    M        1
## 27          56    M        0
## 28          54    H        1
## 29          53    H        0
## 30          53    M        0
## 31          53    H        0
## 32          41    M        1
## 33          48    M        0
## 34          50    H        1
## 35          50    M        1
## 36          52    H        0
## 37          53    M        0
## 38          35    H        0
## 39          35    H        0
## 40          54    M        0
## 41          46    M        1
## 42          48    H        0
## 43          50    M        0
## 44          48    H        0
## 45          41    M        0
## 46          48    M        1
## 47          60    H        1
## 48          53    M        0
## 49          54    M        1
## 50          56    H        1
## 51          50    H        1
## 52          41    H        0
## 53          60    M        1
## 54          60    M        1
## 55          54    H        0
## 56          54    H        0
## 57          53    H        0
## 58          35    M        0
## 59          54    H        0
## 60          48    M        0
## 61          50    H        0
## 62          54    H        0
## 63          54    H        0
## 64          53    H        0
## 65          52    H        0
## 66          50    H        0
## 67          52    H        0
## 68          48    H        1
## 69          46    H        1
## 70          53    H        0
## 71          50    H        0
## 72          35    H        0
## 73          50    H        1
## 74          60    M        1
```

```
## 75         50    H       0
```

A continuación, introducimos en R los datos relativos al nivel de significación y la varianza poblacional de la variable que proporciona el enunciado.

```
alpha<-0.05
varianza<-0.75
```

Calculamos por separado cada uno de los elementos restantes que necesitamos para obtener el intervalo de confianza.

```
n<-nrow(datos)
media<-mean(datos$Flexiones)
cuantil<-qnorm(1-alpha/2)
```

Por último, calculamos los extremos inferior y superior del intervalo de acuerdo a la expresión que se vio anteriormente:

Por tanto:

```
lim_inf<-media-cuantil*sqrt(varianza)/sqrt(n)
lim_inf
```

```
## [1] 49.91067
```

```
lim_sup<-media+cuantil*sqrt(varianza)/sqrt(n)
lim_sup
```

```
## [1] 50.30266
```

Por lo que el intervalo de confianza que buscamos es (49.91067, 50.30266)

Intervalo de confianza para la media en una población normal con varianza desconocida

Supongamos, en este caso, que la varianza poblacional de la variable de interés es desconocida. Nuestro objetivo sigue siendo el cálculo de un intervalo de confianza para la media de dicha variable.

Supuesto Práctico 2

Considerando nuevamente el conjunto de datos que se ha presentado en el supuesto práctico 1, relativo a las flexiones de los alumnos. Calcular un intervalo de confianza a un nivel de confianza del 98% para el número medio de flexiones. Suponer en este caso que el número de flexiones se distribuye según una distribución Normal de varianza desconocida.

```
t.test(datos$Flexiones, conf.level=0.98)
```

```
##
##  One Sample t-test
##
## data:  datos$Flexiones
## t = 72.58, df = 74, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 98 percent confidence interval:
##  48.46512 51.74822
## sample estimates:
## mean of x
##  50.10667
```

De toda la información que devuelve la función t.test, sólo nos interesa la relativa al intervalo de confianza

98 percentconfidenceinterval: 48.46512 51.74822

Por lo tanto, el intervalo de confianza para el número medio de flexiones a un nivel de confianza del 98% es (48.46512, 51.74822).

Resolvamos el mismo ejemplo pero cuando el enunciado nos muestra la media (50.11), la cuasidesviación típica (5.98) y el tamaño muestral (75). Pero no proporciona los datos.

Solución Al no disponer de un vector numérico con los valores de la variable de interés, para resolver el ejemplo recurriremos a utilizar la expresión del intervalo de confianza para este caso

```
alpha<-0.2
n=75
cuasi<-5.98
media<-50.11
```

```
cuantil<-qt((1-alpha/2),8,lower.tail = T)
lim_inf<-media-cuantil*cuasi/sqrt(n)
lim_inf
```

```
## [1] 49.14548
```

```
lim_sup<-media+cuantil*cuasi/sqrt(n)
lim_sup
```

```
## [1] 51.07452
```

Por lo tanto, hay un 98% de confianza de que el intervalo [49.14548 , 51.07452 contenga el número medio de flexiones

Supuesto Práctico 3

A partir del conjunto de datos relativo al número de flexiones y el sexo de los alumnos, obtener un intervalo de confianza al 95% para la proporción de alumnos en la población. Del mismo modo, calcular un intervalo de confianza al 90% para la proporción de alumnas.

Solución Comenzaremos con el cálculo del intervalo de confianza para los chicos. Para realizar la llamada a la función prop.test necesitamos conocer, además del nivel de confianza, que viene indicado en el enunciado, el número de alumnos varones y el número total de estudiantes en la muestra. Para ello utilizamos la función, de R, table.

```
table(datos$Sexo)
```

```
##
##  H  M
## 43 32
```

e los 75 estudiantes que conforman la muestra, 43 son chicos. Por lo que la llamada a prop.test sería la siguiente:

```
prop.test(43,75)
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  43 out of 75, null probability 0.5
## X-squared = 1.3333, df = 1, p-value = 0.2482
## alternative hypothesis: true p is not equal to 0.5
```

```
## 95 percent confidence interval:
##  0.4539882 0.6851173
## sample estimates:
##         p
## 0.5733333
```

Como el nivel de confianza para este intervalo es el 95% no ha sido necesario incluir el argumento conf.level en la llamada a prop.test, puesto que este es el nivel de confianza por defecto.

De nuevo, los resultados de la función incluyen mucha más información aparte de la relativa al intervalo de confianza. Por ahora, nos centraremos únicamente en esta última.

Por lo que el intervalo de confianza, a un nivel de confianza del 95% para la proporción de alumnos varones en la población es (0.4539882, 0.6851173)

Repitamos el procedimiento para obtener ahora un intervalo de confianza para la proporción de chicas. En este caso, tenemos que tener en cuenta que el número de chicas en la muestra era 32.

```
prop.test(23,75,conf.level = 0.90)
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  23 out of 75, null probability 0.5
## X-squared = 10.453, df = 1, p-value = 0.001224
## alternative hypothesis: true p is not equal to 0.5
## 90 percent confidence interval:
##  0.2211717 0.4066284
## sample estimates:
##         p
## 0.3066667
```

Centrándonos en la parte de la salida que incluye el intervalo de confianza

90 percentconfidenceinterval: 0.2211717 0.4066284

Podemos concluir que el intervalo de confianza, considerando un nivel de confianza del 90%, para la proporción de chicas en la población es (0.2211717, 0.4066284).

Resolvamos este ejercicio utilizando la expresión del intervalo de confianza para la proporción

Solución Intervalo de confianza, a un nivel de confianza del 95% para la proporción de alumnos varones en la población

```
alpha<-0.02
n=75
p=32/75
cuantil<-qnorm(1-alpha/2)
lim_inf<-p-cuantil*sqrt(p*(1-p))/sqrt(n)
lim_inf
```

```
## [1] 0.2938074
```

```
lim_sup<-p+cuantil*sqrt(p*(1-p))/sqrt(n)
lim_sup
```

```
## [1] 0.559526
```

Por lo que el intervalo de confianza, a un nivel de confianza del 90%, para la proporción de chicas en la población es (0.3147318, 0.5386015).

Intervalo de confianza para la diferencia de medias en dos poblaciones normales independientes

falta texo

Intervalo de confianza para el cociente de varianzas en dos poblaciones normales independientes

falta texto

Supuesto Práctico 4

Continuando con los datos relativos a las flexiones realizadas por un grupo de estudiantes y asumiendo que el número de flexiones que realizan los chicos y las que realizan las chicas se distribuyen según sendas distribuciones normales con medias y varianzas desconocidas. Calcular un intervalo de confianza a un nivel de confianza del 95% para el cociente de varianzas en ambas poblaciones. ¿Puede asumirse que ambas varianzas son iguales?

Solución En primer lugar determinamos el intervalo de confianza para el cociente de varianzas, para ello utilizamos la función var.test. Lo primero que tenemos que hacer para aplicar la función var.test es separar en dos variables los datos relativos a las flexiones realizadas por los chicos y por las chicas.

```
Flexiones.chicos<-datos$Flexiones[datos$Sexo=="H"]
Flexiones.chicas<-datos$Flexiones[datos$Sexo=="M"]
Flexiones.chicas
```

```
##  [1] 53 53 53 50 48 50 48 52 54 35 50 41 56 52 56 53 41 48 50 53 54 46 50 41 48
## [26] 53 54 60 60 35 48 60
```

```
Flexiones.chicos
```

```
##  [1] 60 41 41 56 50 56 50 50 54 52 48 48 54 53 53 50 52 35 35 48 48 60 56 50 41
## [26] 54 54 53 54 50 54 54 53 52 50 52 48 46 53 50 35 50 50
```

A continuación, utilizamos la función var.test tal y como se indica a continuación:

```
var.test(Flexiones.chicos,Flexiones.chicas)
```

```
##
##  F test to compare two variances
##
## data:  Flexiones.chicos and Flexiones.chicas
## F = 0.87506, num df = 42, denom df = 31, p-value = 0.679
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.4415454 1.6765483
## sample estimates:
## ratio of variances
##          0.8750585
```

Analizando la información relativa al intervalo de confianza que se incluye en la salida de var.test, podemos afirmar que el intervalo de confianza a un nivel de confianza del 95% para el cociente de las varianzas de las dos distribuciones es (0.4415, 1.6765). Este intervalo de confianza contiene al valor 1, por lo que se puede suponer que las varianzas de las dos distribuciones son idénticas.

Una vez se ha determinado la igualdad (o desigualdad) de las varianzas de ambas distribuciones, procedemos a calcular el intervalo de confianza para la diferencia de las medias propiamente dicho.

a) Intervalo de confianza para la diferencia de medias en dos poblaciones normales independientes cuando las varianzas poblacionales son desconocidas pero supuestas iguales

Si la varianzas poblacionales son desconocidas pero supuestas iguales, se parte de la variable aleatoria

b) Intervalo de confianza para la diferencia de medias en dos poblaciones normales independientes cuando las varianzas poblacionales son desconocidas, distintas y tamaños muestrales grandes

Si las varianzas de las poblaciones son desconocidas y, además, distintas y tamaños muestrales grandes, se sigue un procedimiento similar al que acabamos de describir en el caso de igualdad de varianzas para la obtención del intervalo de confianza, partiendo de una variable aleatoria.

#Supuesto Práctico 5 En vista de los resultados obtenidos en el supuesto práctico 4, y suponiendo que el número de flexiones que realizan los alumnos y las alumnas se distribuyen de acuerdo a variables normales de medias y varianzas desconocidas, obtener un intervalo de confianza al 95% para la diferencia del número medio de flexiones entre chicos y chicas. ¿Puede suponerse que el número medio de flexiones que realizan los chicos y las chicas es igual?

Solución Dado que en el supuesto práctico 4 se concluyó la igualdad de las varianzas del número de flexiones que hacen chicos y chicas, debemos establecer a TRUE el valor del parámetro var.equal cuando realicemos la llamada a la función t.test.

```
t.test(Flexiones.chicos,Flexiones.chicas, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  Flexiones.chicos and Flexiones.chicas
## t = -0.06154, df = 73, p-value = 0.9511
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.887271  2.714306
## sample estimates:
## mean of x mean of y
##  50.06977  50.15625
```

Atendiendo la información sobre el intervalo de confianza que se incluye entre los resultados

95 percentconfidenceinterval: -2.887271 2.714306

Se puede afirmar que el intervalo de confianza a un 95% de confianza para la diferencia de las medias del número de flexiones que hacen chicos y chicas es (-2.8872, 2.7143). Como el 0 está dentro de este intervalo, se puede decir que ambas medias son idénticas.

Intervalo de confianza para la diferencia de medias en dos poblaciones normales relacionadas

Supuesto Práctico 6

Para estudiar los efectos de un programa de control de peso, el profesor de educación física selecciona aleatoriamente a 6 alumnos y se les toma nota de sus pesos antes y después de pasar por el programa.

falta tabla

Construir un intervalo de confianza a un 95% de confianza para la diferencia de medias de los pesos antes y después de seguir el programa.

Solución Como puede observarse, los datos vienen por parejas: peso antes y después, dos datos por individuo. Parece lógico que los datos se encuentren relacionados entre sí.

En primer lugar, vamos a introducir los datos en R. Para ello definimos dos vectores

```
Antes<-c(72.0, 73.5, 70.0, 71.5, 76.0, 80.5)
Despues<-c(73.0, 74.5, 74.0, 74.5, 75.0, 82.0)
```

A partir de estos datos, vamos a aplicar la función t.test, para obtener el intervalo de confianza que buscamos.

```
t.test(Antes,Despues,paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  Antes and Despues
## t = -2.2238, df = 5, p-value = 0.07676
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.4135884  0.2469217
## sample estimates:
## mean of the differences
##               -1.583333
```