

INTERVALOS DE CONFIANZA

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

El intervalo de confianza para la media de una variable continua con el valor de la varianza de dicha variable conocida en toda la población es el intervalo menos usual.

Supuesto Práctico 1

El archivo empleados.xls nos informa de la edad, altura, peso, sexo y posesión de coche de 100 empleados de una empresa. Suponiendo la normalidad de la variable Altura, calcular el intervalo de confianza sobre la altura media poblacional a un 95% de confianza, sabiendo que la varianza poblacional es 6.

Solución

Vamos a trabajar con la base de datos empleados.xls que contiene las variables edad, altura, peso, sexo y posesión de coche, medidas en los empleados de una empresa.

Recordemos, que para poder utilizar R Commander hemos de instalar y cargar el paquete Rcmdr, que no viene incluido por defecto en R. Como dicho paquete ya está instalado, únicamente tenemos que cargarlo. Para ello, en el menú principal se selecciona Paquetes/Cargar paquete.. y en la ventana que se despliega, se selecciona el paquete Rcmdr y se pulsa Ok.

Una vez activado R Commander, pasamos a importar datos de Excel a R-Commander mediante el Menú: Datos/Importar datos/desde un archivo de Excel.

```
# incluya este fragmento de código literalmente para especificar las opciones
knitr::opts_chunk$set(comment=NA, prompt=TRUE, out.width=750, fig.height=8, fig.width=8)
library(Rcmdr)
```

Warning: package 'Rcmdr' was built under R version 4.1.3

Loading required package: splines

Loading required package: RcmdrMisc

Warning: package 'RcmdrMisc' was built under R version 4.1.3

Loading required package: car

Warning: package 'car' was built under R version 4.1.3

Loading required package: carData

Warning: package 'carData' was built under R version 4.1.3

Loading required package: sandwich

Warning: package 'sandwich' was built under R version 4.1.3

Loading required package: effects

Warning: package 'effects' was built under R version 4.1.3

lattice theme set by effectsTheme()
See ?effectsTheme for details.

La interfaz R-Commander sólo funciona en sesiones interactivas

Attaching package: 'Rcmdr'

The following object is masked from 'package:base':

errorCondition

```
library(car)
library(RcmdrMisc)
```

```
empleados <- readXL("C:/Users/user/OneDrive/Paquete R/PRACTICAS-S10/empleados-1.xls", rown
```

A continuación guardamos el archivo (Datos/Conjunto de datos activo/Guardar el conjunto de datos activo)

```
> save("empleados", file="C:/Users/user/OneDrive/Paquete R/PRACTICAS-S10/Empleados.RData")
```

como fichero Empleados.RData.

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

```
> 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(empleados)
> media<-mean(empleados$Altura)
> cuantil<-qnorm(1-alpha/2)
```

Los seleccionamos todos y pulsamos Ejecutar

Por último, calculamos los extremos inferior y superior del intervalo de acuerdo

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

```
[1] 176.8294
```

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

```
[1] 177.1706
```

Por lo que el intervalo de confianza que buscamos es (176.4605, 177.5395).

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 el conjunto de datos de empleados.xls y asumiendo que la variable que mide la altura de los empleados sigue una distribución Normal con varianza desconocida. Calcular un intervalo de confianza a un nivel de confianza del 90% para la altura media poblacional.

Solución

Accedemos al menú Test t para una muestra, seleccionando en el menú principal: Estadísticos/ Medias/ Test t para una muestra.

Test t para una muestra: Altura

```
> with(empleados, (t.test(Altura, alternative='two.sided', mu=0.0,  
+   conf.level=.90)))
```

One Sample t-test

```
data:  Altura  
t = 214.41, df = 98, p-value < 2.2e-16  
alternative hypothesis: true mean is not equal to 0  
90 percent confidence interval:  
 175.6292 178.3708  
sample estimates:  
mean of x  
 177
```

De toda la información que devuelve, sólo nos interesa la relativa al intervalo de confianza. El resto hace referencia a los contrastes de hipótesis.

El intervalo de confianza pedido es (175.6292, 178.3708).

Intervalo de confianza para la proporción

Para calcular un intervalo de confianza para la proporción de una población normal mediante R-Commander, seleccionamos en el menú principal: Estadísticos/Proporciones/ Test de proporciones para una muestra.

Supuesto Práctico 3

A partir del conjunto de datos de empleados.xls, obtener un intervalo de confianza al 95% para la proporción de empleados varones en la población.

Solución

Accedemos al menú Test de proporciones para una muestra de R-Commander, seleccionando en el menú principal: Estadísticos/Proporciones/ Test de proporciones para una muestra.

En la pestaña Datos del cuadro de diálogo, se muestra una lista con todas las variables cualitativas que pueden utilizarse en este tipo de contrastes, de entre las cuales tenemos que elegir una. En este caso elegimos Sexo.

Una vez elegido Sexo, pulsamos la pestaña Opciones

En la pestaña Opciones dejamos los valores que vienen por defecto, ya que el nivel de confianza es el pedido por el enunciado, 0.95. Pulsamos Aceptar y se muestra la siguiente salida

```
> # Test de proporciones para una muestra: Sexo
> local({
+   .Table <- xtabs(~ Sexo , data= empleados )
+   cat("\nFrequency counts (test is for first level):\n")
+   print(.Table)
+   prop.test(rbind(.Table), alternative='two.sided', p=.5, conf.level=.95,
+     correct=FALSE)
+ })
```

Frequency counts (test is for first level):

Sexo

Hombre	Mujer
87	12

1-sample proportions test without continuity correction

```

data:  rbind(.Table), null probability 0.5
X-squared = 56.818, df = 1, p-value = 4.78e-14
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.7999934 0.9292845
sample estimates:
      p
0.8787879

```

En primer lugar, se muestra una tabla con las frecuencias absolutas de cada categoría de la variable cualitativa. Es muy importante tener en cuenta que R-Commander realiza el contraste de hipótesis, y por lo tanto el intervalo, para la primera categoría de la variable.

Para R-Commander la primera categoría de una variable es la que primero aparece siguiendo el orden alfabético, en caso de que las categorías vengan dadas por cadenas de caracteres, o aquella con el número más bajo, en caso de que las categorías se identifiquen mediante un código numérico.

En este ejemplo, las dos posibles opciones para la variable Sexo son “Hombre” y “Mujer”, por lo que la primera de las categorías para R-Commander es “Hombre”. Dado que la hipótesis que se ha planteado se ha hecho sobre los hombres no es necesario hacer ninguna modificación.

Si, por el contrario, la hipótesis del problema se hubiera planteado sobre las mujeres, deberíamos hacer una recodificación previa de la variable para situar la categoría “Mujer” como la primera.

En la segunda parte se muestran los resultados del contraste de hipótesis, que analizaremos en la práctica sobre Contrastes de Hipótesis.

Y por último, muestra el resultado del intervalo de confianza pedido

Por lo que el intervalo de confianza, a un nivel de confianza del 95% para la proporción de empleados varones en la población es (0.7999934, 0.9292845).

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

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

Supuesto Práctico 4

Continuando con los datos del archivo empleados.xls y asumiendo que el peso en hombres y el peso en mujeres se distribuyen según 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

Accedemos a Test F para dos varianzas, de R-Commander, seleccionando en el menú principal: Estadísticos/ Varianzas/ Test F para dos varianzas

```
> # Test F de dos varianzas: Peso ~ Sexo  
> Tapply(Peso ~ Sexo, var, na.action=na.omit, data=empleados)
```

Hombre	Mujer
166.99626	92.06061

```
> # variances by group  
> var.test(Peso ~ Sexo, alternative='two.sided', conf.level=.95,  
+ data=empleados)
```

F test to compare two variances

```
data:  Peso by Sexo  
F = 1.814, num df = 86, denom df = 11, p-value = 0.2752  
alternative hypothesis: true ratio of variances is not equal to 1  
95 percent confidence interval:  
 0.6112226 3.8937784  
sample estimates:  
ratio of variances  
 1.813982
```

Esta salida, muestra que el intervalo de confianza para el cociente de las varianzas es (0.6112226, 3.8937784).

La interpretación del intervalo de confianza puede servirnos para concluir acerca de la igualdad de las varianzas. En este ejemplo, dicho intervalo es (0.6112226, 3.8937784) que, como podemos comprobar incluye al 1 entre sus posibles valores. Esto implica que a un nivel de confianza del 95% se puede suponer que el cociente entre las dos varianzas puede tomar el valor 1 o, lo que es lo mismo, que las dos varianzas son iguales.

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.