

PRACTICA 5, SEMANA 10

Nelson de Jesus Magaña Godinez

2022-10-10

Contents

Intervalo de confianza para la media en una población normal con varianza conocida.	2
Intervalo de confianza para la media en una población normal con varianza desconocida	4
Intervalo de confianza para la proporción.	5
Intervalo de confianza para el cociente de varianzas en dos poblaciones normales independientes.	6
a) Intervalo de confianza para la diferencia de medias en dos poblaciones normales independientes cuando las varianzas poblacionales son desconocidas pero supuestas iguales.	7
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.	8
Intervalo de confianza para la diferencia de medias en dos poblaciones normales relacionadas.	9
Intervalo de confianza para la diferencia de dos proporciones.	10
Ejercicios Guiados	11
Ejercicio Guiado1	11
Ejercicio Guiado 2	18
Ejercicio Guiado 3.	19
Ejercicio Guiado 4.	20
Ejercicios Propuestos	22
Ejercicio Propuesto 1	22
Ejercicio Propuesto 2	22
Ejercicio Propuesto 3	23
Ejercicio Propuesto 4	24

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.

Para estimar la media poblacional μ de una población Normal de media μ (desconocida) y de varianza σ^2 (conocida), $N(\mu, \sigma^2)$, se selecciona una muestra aleatoria X_1, X_2, \dots, X_n ; de tamaño n de valores de una variable aleatoria de esta población y se calcula su media muestral, como mejor estimador puntual de μ . La construcción del intervalo de confianza se hace tomando como base este estimador. Para calcular un intervalo de confianza para μ partimos de la variable aleatoria

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

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.

```
library(Rcmdr)
```

```
## Loading required package: splines
```

```
## Loading required package: RcmdrMisc
```

```
## Loading required package: car
```

```
## Loading required package: carData
```

```
## Loading required package: sandwich
```

```
## Loading required package: effects
```

```
## 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
```

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

```
empleados <-
  readXL("C:/Users/Usuario/Desktop/respaldo/Desktop/PAQUETE R/PRACTICAS_S10/empleados-1.xls",
    rownames=FALSE, header=TRUE, na="", sheet="Respuestas",
    stringsAsFactors=TRUE)
```

Guardamos el conjunto de datos activos

```
save("empleados",
  file="C:/Users/Usuario/Desktop/respaldo/Desktop/PAQUETE R/PRACTICAS_S10/Dataset.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 <- 7.5
```

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

```
n <- nrow(empleados)
n
```

```
## [1] 99
```

```
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] 176.4605
```

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

```
## [1] 177.5395
```

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.

Supongamos una muestra aleatoria X_1, X_2, \dots, X_n ; de tamaño n de valores de la variable aleatoria que sigue una distribución Normal de media μ y de varianza σ^2 , ambas desconocidas. Para calcular un intervalo de confianza, en este caso, partimos de la variable aleatoria

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

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.

En la parte superior izquierda se muestra una lista con todas las variables cuantitativas del archivo de datos que son susceptibles de ser contrastadas, de la cual debemos elegir exclusivamente una.

A continuación se muestran diversas Hipótesis alternativas así como el nivel de confianza. En este caso queremos obtener un intervalo de confianza, por lo que dejamos marcada la opción por defecto Media poblacional! = μ_0 y como Hipótesis nula: $\mu = 0.0$. Lo único que tenemos que cambiar es el nivel de confianza (el valor se introduce en tanto por uno). Nivel de confianza: .90.

```
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.

Dada una variable aleatoria X con distribución de probabilidad binomial de parámetros n y π , esto es, $X \rightarrow B(n, \pi)$; con π desconocido. El objetivo es determinar un intervalo de confianza para el parámetro π . Para ello, se extrae una muestra aleatoria X_1, X_2, \dots, X_n de tamaño n de dicha distribución. Sea p la proporción muestral. Entonces, se sabe que el estadístico

$$z = \frac{p - \pi}{\sqrt{\frac{p(1-p)}{n}}}$$

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 pestana 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

```
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
```

Analicemos las salidas que proporciona el programa Nos dice que en el estudio hay 87 hombre y 12 mujeres.

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

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

Consideramos dos variables aleatorias independientes X_1, X_2, \dots, X_n e Y_1, Y_2, \dots, Y_n con distribuciones normales de parámetros (μ_1, σ_1) y (μ_2, σ_2) , respectivamente, de las que vamos a tomar muestras aleatorias independientes de tamaños n_1 y n_2 , respectivamente.

Nuestro objetivo, en este caso, es obtener un intervalo de confianza para la diferencia de las medias de ambas distribuciones, es decir, para $\mu_1 - \mu_2$. Pero previo al cálculo de este intervalo, debemos determinar si las varianzas de ambas distribuciones o, equivalentemente, sus desviaciones típicas, σ_1 y σ_2 , aun siendo desconocidas, pueden asumirse iguales o no. El cálculo del intervalo de confianza se realiza de forma diferente dependiendo de si las varianzas (desviaciones típicas) pueden asumirse iguales o no.

En primer lugar determinemos el Intervalo de confianza para el cociente de varianzas

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

Para decidir si las varianzas de las dos distribuciones pueden asumirse iguales o no construiremos un intervalo de confianza para el cociente de ambos valores, esto es, para σ_1^2/σ_2^2 . En este caso, partimos de la variable aleatoria

$$F = \frac{s_1^2 \sigma_2^2}{s_2^2 \sigma_1^2}$$

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

La pestaña Datos muestra dos listas de variables.

- La lista de Grupos, muestra todas las variables cualitativas del fichero de datos. En esta lista tenemos que seleccionar cuál es la variable que nos va a dividir la muestra de observaciones en dos submuestras independientes. -En nuestro caso, al tratar con hombres y mujeres, seleccionamos el Sexo del empleado como variable de agrupación.
- La lista de Variable explicativa, muestra todas las variables cuantitativas del fichero de datos. Señalamos la variable principal sobre la cual se va a llevar a cabo el contraste (en nuestro caso, Peso).

Seleccionamos la pestaña Opciones

En la pestaña Opciones podemos personalizar el contraste. Como en este caso sólo nos interesa el intervalo de confianza, sólo hemos de mirar el nivel de confianza. Por defecto es 0.95, por lo que no cambiamos ninguna de las opciones de esta pestaña y mantenemos las opciones por defecto. Pulsamos Aceptar y se muestra la siguiente salida

```
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
```

sta 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.

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

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-1}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

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.

$$Z = \frac{(\bar{X} - \bar{X}_2)(\mu_1 - \mu_2)}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Supuesto Práctico 5

Sabiendo que las varianzas son iguales (Supuesto práctico 4), obtener un intervalo de confianza al 95% para la diferencia del peso medio entre hombres y mujeres. ¿Puede suponerse que el peso medio entre hombres y mujeres es igual?

Solución

Para obtener el Test t para muestras independientes con R-Commander. Seleccionamos en el menú principal: Estadísticos/Medias/ Test t para muestras independientes.

La pestaña Datos muestra dos listas de variables. Como ya se ha comentado con anterioridad, en la lista de la izquierda (Grupos) tenemos que escoger la variable a partir de la cual se formarán los dos grupos de observaciones, seleccionamos Sexo. En la de la derecha (Variable explicada) seleccionamos la variable cuya diferencia de medias en las poblaciones queremos estudiar, seleccionamos Peso.

Seleccionamos Opciones

La pestaña muestra todas las opciones del contraste que podemos modificar. En nuestro caso como sólo buscamos el intervalo de confianza, especificamos el nivel de confianza que se va a asumir al calcular el intervalo e indicamos si las varianzas de las dos poblaciones pueden suponerse iguales o no. En nuestro caso suponemos que las varianzas son iguales

Cuando hemos seleccionado las variables adecuadas y hemos marcado las opciones que nos interesan, pulsamos en Aceptar, de manera que en la consola se muestra la siguiente salida

```
t.test(Peso-Sexo, alternative='two.sided', conf.level=.95,
      var.equal=TRUE, data=empleados)

##
## Two Sample t-test
##
## data:  Peso by Sexo
## t = 3.0597, df = 97, p-value = 0.002865
## alternative hypothesis: true difference in means between group Hombre and group Mujer is not equal to 0
## 95 percent confidence interval:
##   4.167581 19.556557
## sample estimates:
## mean in group Hombre  mean in group Mujer
##           76.19540           64.33333
```

El intervalo de confianza que se incluye entre los resultados es:

Se puede afirmar que el intervalo de confianza a un 95% de confianza para la diferencia de las medias del peso medio para hombres y mujeres es (4.167581, 19.556557). Como el 0 no está dentro de este intervalo, no tenemos suficiente evidencia muestral para decir que el peso medio de hombres y mujeres sea el mismo.

Antes	Despues
72.2	73
73.5	74.5
70	74
71.5	74.5
76	75
80.5	82

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

Sean X_1, X_2, \dots, X_n e Y_1, Y_2, \dots, Y_n dos muestras aleatorias de tamaño n y relacionadas, de tal forma que la primera procede de una población $N(\mu_1, \sigma_1)$ y la segunda de una población $N(\mu_2, \sigma_2)$.

Antes de proporcionar el intervalo para la diferencia de medias de estas dos poblaciones, se hace necesario indicar qué se entiende por muestras relacionadas. Se dicen que dos muestras X_1, X_2, \dots, X_n e Y_1, Y_2, \dots, Y_n están relacionadas o apareadas cuando los datos de las muestras vienen por parejas, uno de cada una de ellas, de manera que cada individuo proporciona dos observaciones.

Para calcular un intervalo de confianza para la diferencia de medias en dos poblaciones normales relacionadas en R-Commander. Seleccionamos en el menú principal: Estadísticos/Medias/ Test t para datos relacionados.

Supuesto Práctico 6

Se desea evaluar la eficacia de un fármaco para la reducción del nivel de glucosa en pacientes. Para ello, se selecciona una muestra de 10 pacientes a los que se les mide su nivel de glucosa en sangre antes y después del suministro del medicamento. Los resultados aparecen recogidos en la siguiente tabla:

Calcular el intervalo de confianza para la diferencia de medias al 90%. ¿Son iguales los niveles medios de glucosa antes y después del medicamento?

```
supuesto6 <-
  read.table("C:/Users/Usuario/Desktop/respaldo/Desktop/PAQUETE R/PRACTICAS_S10/s10.txt",
    header=TRUE, stringsAsFactors=TRUE, sep=" ", na.strings="NA",
    dec=".", strip.white=TRUE)
```

Solución

En primer lugar, debemos crear un nuevo conjunto de datos con la información que nos proporciona la tabla. El conjunto de datos estará formado por dos variables con los niveles de glucosa antes y después de la aplicación del fármaco. Organizamos los datos en un fichero .txt (supuesto6.txt)

A continuación seleccionamos en el menú principal: Datos/Importar datos/desde archivo de texto

A continuación, accedemos al menú Test t para datos relacionados en R-Commander. Seleccionamos en el menú principal: Estadísticos/Medias/ Test t para datos relacionados

en muestra la siguiente pantalla

La pestaña Datos muestra dos listas de variables cada una de las cuales incluye todas las variables cuantitativas que son susceptibles de ser analizadas. Seleccionamos, en nuestro caso, Antes y Después

Es importante destacar que, a diferencia del caso de muestras independientes, cuando trabajamos con muestras pareadas no necesitamos una variable de agrupación, sino que debemos seleccionar las dos variables a analizar de forma separada.

En la pestaña Opciones personalizamos el contraste conforme al problema que estemos resolviendo. Como en este caso sólo nos interesa el intervalo de confianza, introducimos el valor del nivel de confianza dado en el enunciado.

Si hacemos clic en Aceptar, el programa nos devuelve la siguiente salida:

```
with(supuesto6, (t.test(Antes, Despues,
  alternative='two.sided', conf.level=.90, paired=TRUE)))

##
## Paired t-test
##
## data: Antes and Despues
## t = 0.52714, df = 9, p-value = 0.6108
## alternative hypothesis: true mean difference is not equal to 0
## 90 percent confidence interval:
## -9.290501 16.790501
## sample estimates:
## mean difference
## 3.75
```

En la segunda parte de los resultados se incluye el intervalo de confianza al 90% para la diferencia de las medias de la variable glucosa, que es (-9.290501, 16.790501). Este intervalo incluye el valor 0, lo que significa que el 0 es un valor posible para la diferencia entre las medias. Por ello, concluimos que puede asumirse que la diferencia entre dichas medias es 0, o dicho de otro modo, que ambos niveles medios de glucosa son iguales.

Intervalo de confianza para la diferencia de dos proporciones.

Vamos a construir un intervalo de confianza para la diferencia de proporciones $\pi_1 - \pi_2$. Para ello, consideramos dos muestras aleatorias X_1, X_2, \dots, X_{n_1} e Y_1, Y_2, \dots, Y_{n_2} de tamaño n_1 y n_2 , independientes entre sí, extraídas de poblaciones con distribuciones binomiales $B(1, \pi_1)$ y $B(1, \pi_2)$, respectivamente.

Construimos, a continuación, un intervalo de confianza para la diferencia de proporciones $\pi_1 - \pi_2$. Para ello, partimos de la variable aleatoria

$$Z = \frac{(p_1 - p_2)(\pi_1 - \pi_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

Supuesto Práctico 7

A partir del conjunto de datos empleados.xls, obtener un intervalo de confianza al 85% para la diferencia entre la proporción de empleados hombres y mujeres que tienen coche. ¿Pueden considerarse ambas proporciones iguales?

Solución

Dado que la hipótesis se ha planteado sobre la proporción de hombres y mujeres que tienen coche, es necesario recodificar la variable coche. Para ello seleccionamos Datos/Modificar variables del conjunto de datos activo/Recodificar variables

En esta pantalla:

- Se selecciona Coche en Variables a recodificar
- En Nuevo nombre o prefijo para variables múltiples recodificadas hemos puesto como nombre de la variable: coche_rec
- En la ventana Introducir directrices de recodificación: “No”=“2No” y “Sí”=“zSí”

Se pulsa Aceptar. Seleccionando Visualizar conjunto de datos activos se muestra el archivo de datos resultantes

Accedemos al menú Test de proporciones para dos muestras de R-Commander, seleccionando en el menú principal: Estadísticos/Proporciones/ Test de proporciones para dos muestras

Se muestra la siguiente pantalla

La pestaña Datos muestra dos listas con las variables cualitativas que incluye el conjunto de datos. De la primera lista seleccionamos la variable de agrupación (en nuestro caso es el Sexo, ya que distinguimos entre hombres y mujeres) y de la segunda, la variable de interés (que es, si el empleado tiene coche, coche_rec).

```
empleados <- within(empleados, {  
  Coche_rec <- Recode(Coche, '"No"="2No"; "Si"="1Si"; ;',  
    as.factor=TRUE, to.value="", interval=":", separator=";")  
})
```

Pulsamos la pestaña OpcionesEn la pestaña Opciones indicamos el nivel de confianza propuesto en el enunciado, 85% Se pulsa Aceptar y R-Commander muestra la siguiente salida

```
local({ .Table <- xtabs(~Sexo+Coche_rec, data=empleados)  
  cat("\nPercentage table:\n")  
  print(rowPercents(.Table))  
  prop.test(.Table, alternative='two.sided', conf.level=.85,  
    correct=FALSE)  
})
```

```
##  
## Percentage table:  
##      Coche_rec  
## Sexo      2No  Si Total Count  
##  Hombre 47.1 52.9   100     87  
##   Mujer 50.0 50.0   100     12  
  
##  
## 2-sample test for equality of proportions without continuity correction  
##  
## data: .Table  
## X-squared = 0.03492, df = 1, p-value = 0.8518  
## alternative hypothesis: two.sided  
## 85 percent confidence interval:  
##  -0.2503366  0.1928653  
## sample estimates:  
##      prop 1      prop 2  
## 0.4712644 0.5000000
```

El intervalo de confianza pedido al 85% de confianza es (-0.2729805, 0.3304517). El 0 está dentro de este intervalo, por lo que podemos concluir que las proporciones de hombre y mujeres que tienen coche coinciden.

Ejercicios Guiados

Ejercicio Guiado1

En la base de datos universidad.txt tenemos información sobre las variables, Coeficiente intelectual y Asistir a clase de estadística de dos grupos de alumnos, dependiendo del turno de clase en el que se encuentren. El

turno de mañana se define como A y el de tarde como B.

Se pide:

- Obtener un intervalo de confianza a un nivel del 99% para el cociente intelectual medio, sabiendo que la varianza poblacional es igual a 3
- Obtener un intervalo de confianza a un nivel del 95% para el cociente intelectual medio
- Obtener un intervalo de confianza a un nivel del 98% para la diferencia media de cociente intelectual entre el grupo A y B. ¿Puede suponerse que el cociente intelectual medio entre ambos grupos es igual?
- Obtener un intervalo de confianza a un nivel del 90% para la proporción de alumnos en el grupo A. Y un intervalo de confianza al 90% para la proporción de alumnos en el grupo B
- Obtener un intervalo de confianza a un nivel del 93% para la diferencia entre la proporción de alumnos en el grupo A y B que tienen clase de estadística.

Solucion.

En primer lugar debemos importar el archivo de datos universidad.txt.

Para ello, seleccionamos en el menú principal: Datos/Importar datos/desde archivo de texto

Se muestra la siguiente pantalla

donde, en Introducir el nombre del conjunto de datos escribimos universidad y pulsamos Aceptar. En el directorio seleccionamos el archivo universida.txt y pulsamos abrir.

R-Commander muestra el archivo universidad en Conjunto de datos

Pulsamos Visualizar conjunto de datos y se muestra

```
library(Rcmdr)
```

```
universidad <-  
read.table("C:/Users/Usuario/Desktop/respaldo/Desktop/PAQUETE R/PRACTICAS_S10/universidad.txt",  
header=TRUE, stringsAsFactors=TRUE, sep=" ", na.strings="NA",  
dec=".", strip.white=TRUE)
```

- Obtener un intervalo de confianza a un nivel del 99% para el cociente intelectual medio, sabiendo que la varianza poblacional es igual a 3

En este caso nos encontramos ante un intervalo de confianza sobre la media de una población normal con varianza conocida, por lo que hay que calcularlo mediante código

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

```
alpha <- 0.01  
varianza <- 3
```

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

```
n <- nrow(universidad)  
media <- mean(universidad$C.I.)  
cuantil <- qnorm(1-alpha/2)
```

Por último, calculamos los extremos inferior y superior del intervalo

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

```
## [1] 99.20477
```

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

```
## [1] 103.1952
```

Por lo que el intervalo de confianza que buscamos es (99.20477, 103.1952).

Nota: Recordar que para que se ejecute una instrucción hay que pulsar Ejecutar

b) Obtener un intervalo de confianza a un nivel del 95% para el cociente intelectual medio

En este caso nos encontramos ante un intervalo de confianza sobre la media de una población normal con varianza desconocida. Para realizarlo con R-Commander, seleccionamos en el menú principal: Estadísticos/Medias/ Test t para una muestra

En la parte superior izquierda aparece una lista con todas las variables cuantitativas del archivo de datos que son susceptibles de ser contrastadas, de la cual debemos elegir exclusivamente una. Elegimos C.I

El resto de opciones las dejamos por defecto, ya que el enunciado nos pide el intervalo de confianza al 95%.

Pulsamos Aceptar y se muestra la siguiente salida

```
with(universidad, (t.test(C.I., alternative='two.sided',
mu=0.0, conf.level=.95)))
```

```
##
## One Sample t-test
##
## data: C.I.
## t = 79.024, df = 4, p-value = 1.537e-07
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## 97.64442 104.75558
## sample estimates:
## mean of x
## 101.2
```

Por lo que el intervalo de confianza es (97.64442, 104.75558).

c) Obtener un intervalo de confianza a un nivel del 98% para la diferencia media de cociente intelectual entre el grupo A y B. ¿Puede suponerse que el cociente intelectual medio entre ambos grupos es igual?

En este caso nos encontramos ante un intervalo de confianza para la diferencia de medias en dos poblaciones normales independientes.

En primer lugar tenemos que saber si las varianzas de ambas distribuciones son iguales. Para ello, seleccionamos en R-Commander: Estadísticos/ Varianzas/ Test F para dos varianzas. Se muestra la siguiente pantalla

La pestaña Datos muestra dos listas de variables. La lista de la izquierda (Grupos) incluye todas las variables cualitativas del fichero de datos. En esta lista tenemos que seleccionar cuál es la variable que nos va a dividir la muestra de observaciones en dos submuestras independientes. En nuestro caso, Grupo. En la lista de la derecha se incluyen las variables cuantitativas del fichero de datos. Aquí tenemos que señalar la variable principal sobre la cual se va a llevar a cabo el contraste (en nuestro caso, C.I.). Seleccionamos Opciones

En la pestaña Opciones podemos personalizar el contraste. Como en este caso sólo nos interesa el intervalo de confianza, sólo hemos de mirar el nivel de confianza. Por defecto es 0.95, por lo que lo cambiamos a 0.98. Pulsamos Aceptar y se muestra la siguiente salida.

```
Tapply(C.I. ~ Grupo, var, na.action=na.omit,
      data=universidad) # variances by group
```

```
##           A           B
## 12.33333  8.00000
```

```
var.test(C.I. ~ Grupo, alternative='two.sided',
         conf.level=.98, data=universidad)
```

```
##
## F test to compare two variances
##
## data:  C.I. by Grupo
## F = 1.5417, num df = 2, denom df = 1, p-value = 0.9897
## alternative hypothesis: true ratio of variances is not equal to 1
## 98 percent confidence interval:
##  3.083642e-04 1.518580e+02
## sample estimates:
## ratio of variances
##           1.541667
```

El intervalo de confianza para el cociente de las varianzas, (3.083642e-04, 1.518580e+02). Dicho intervalo incluye al 1 entre sus posibles valores. Esto implica que a un nivel de confianza del 98% 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 de las varianzas de ambas distribuciones, procedemos a calcular el intervalo de confianza para la diferencia de las medias cuando las varianzas poblacionales son iguales. Para ello, seleccionamos en R-Commander: Estadísticos/Medias/Test t para muestras independientes

La pestaña Datos muestra dos listas de variables. Como ya se ha comentado con anterioridad, en la lista de la izquierda (Grupos) tenemos que escoger la variable a partir de la cual se formarán los dos grupos de observaciones, en nuestro caso Grupo. En la de la derecha (Variable explicada) seleccionamos la variable cuya diferencia de medias en las poblaciones queremos estudiar, en nuestro caso C.I.. Pulsamos Opciones

La pestaña Opciones muestra todas las opciones del contraste que podemos modificar. En nuestro caso como sólo buscamos el intervalo de confianza, especificamos el nivel de confianza que se va a asumir al calcular el intervalo (98%) e indicamos si las varianzas de las dos poblaciones pueden suponerse iguales o no, en nuestro caso Sí.

Una vez que hemos seleccionado las variables adecuadas y hemos marcado las opciones que nos interesan, pulsamos en Aceptar, de manera que en la consola aparece una salida como esta:

```
t.test(C.I.~Grupo, alternative='two.sided', conf.level=.98,
      var.equal=TRUE, data=universidad)
```

```
##
## Two Sample t-test
##
## data: C.I. by Grupo
## t = 0.11066, df = 3, p-value = 0.9189
## alternative hypothesis: true difference in means between group A and group B is not equal to 0
## 98 percent confidence interval:
## -13.34472 14.01139
## sample estimates:
## mean in group A mean in group B
## 101.3333 101.0000
```

El intervalo de confianza a un 98% de confianza para la diferencia de las medias del cociente intelectual entre el grupo A y el B es (-13.34472, 14.01139). Como el 0 está dentro de este intervalo, tenemos suficiente evidencia muestral para decir que el cociente intelectual medio del grupo A y B son iguales.

- d) Obtener un intervalo de confianza a un nivel del 90% para la proporción de alumnos en el grupo A. Y un intervalo de confianza al 90% para la proporción de alumnos en el grupo B

En este caso nos encontramos ante un intervalo de confianza para la proporción.

Obtener un intervalo de confianza a un nivel del 90% para la proporción de alumnos en el grupo A. Dado que la hipótesis que se ha planteado se ha hecho sobre el grupo A, no es necesario hacer ninguna recodificación de la variable. Seleccionamos: Estadísticos/Proporciones/Test de proporciones para una muestra

La pestaña Variable 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 la variable Grupo.

En la pestaña Opciones modificamos el valor del nivel de confianza a un 90%.

Cuando presionamos el botón Aceptar, obtenemos los resultados del análisis, que son los que se muestran a continuación:

```
local({
  .Table <- xtabs(~ Grupo , data= universidad )
  cat("\nFrequency counts (test is for first level):\n")
  print(.Table)
  prop.test(rbind(.Table), alternative='two.sided', p=.5,
    conf.level=.90, correct=FALSE)
})
```

```
##
## Frequency counts (test is for first level):
## Grupo
## A B
## 3 2
```

```
## Warning in prop.test(rbind(.Table), alternative = "two.sided", p = 0.5, : Chi-
## squared approximation may be incorrect
```

```
##
## 1-sample proportions test without continuity correction
##
## data:  rbind(.Table), null probability 0.5
## X-squared = 0.2, df = 1, p-value = 0.6547
## alternative hypothesis: true p is not equal to 0.5
## 90 percent confidence interval:
##  0.2724832 0.8572935
## sample estimates:
##      p
## 0.6
```

Por lo que el intervalo de confianza, a un nivel de confianza del 90% para la proporción de estudiantes en el grupo A es (0.2724832, 0.8572935).

Obtener un intervalo de confianza a un nivel del 90% para la proporción de alumnos en el grupo B. Dado que la hipótesis que se ha planteado se ha hecho sobre el grupo B es necesario hacer una recodificación de la variable. Para ello seleccionamos Datos/Modificar variables del conjunto de datos activo/Recodificar variables

Se muestra la siguiente pantalla

En la casilla: Nuevo nombre o prefijo para la variable múltiple recodificadas ponemos Grupo_rec.

```
universidad <- within(universidad, {
  Grupo_rec <- Recode(Grupo, "A"="2A"; "B"="1B",
    as.factor=TRUE, to.value="=", interval=":", separator=";")
})
```

Una forma de recodificar la variable es ponerle un número delante del carácter. De este modo la primera categoría de la variable es aquella con el número más bajo, en este caso la B.

Una vez recodificada la variable, pasamos a calcular el intervalo de confianza. En este caso nos encontramos ante un intervalo de confianza para la proporción. Seleccionamos: Estadísticos/Proporciones/Test de proporciones para una muestra. Se muestra la siguiente pantalla

La pestaña Variable 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 la variable Grupo_rec.

En la pestaña Opciones modificamos el valor del nivel de confianza a un 90%.

Cuando presionamos el botón Aceptar, obtenemos los resultados del análisis, que son los que se muestran a continuación:

```
local({
  .Table <- xtabs(~ Grupo_rec , data= universidad )
  cat("\nFrequency counts (test is for first level):\n")
  print(.Table)
  prop.test(rbind(.Table), alternative='two.sided', p=.5,
    conf.level=.90, correct=FALSE)
})
```

```
##
## Frequency counts (test is for first level):
## Grupo_rec
## 1B 2A
##  2  3
```



```
## Warning in prop.test(rbind(.Table), alternative = "two.sided", p = 0.5, : Chi-
## squared approximation may be incorrect
```

```
##
## 1-sample proportions test without continuity correction
##
## data:  rbind(.Table), null probability 0.5
## X-squared = 0.2, df = 1, p-value = 0.6547
## alternative hypothesis: true p is not equal to 0.5
## 90 percent confidence interval:
##  0.1427065 0.7275168
## sample estimates:
##      p
## 0.4
```

Por lo que el intervalo de confianza, a un nivel de confianza del 90% para la proporción de estudiantes en el grupo B es (0.1427065, 0.7275168).

- e) Obtener un intervalo de confianza a un nivel del 93% para la diferencia entre la proporción de alumnos en el grupo A y B que no tienen clase de estadística.

Dado que la hipótesis que se ha planteado se ha hecho sobre los alumnos que no tienen clase de estadística no es necesario hacer una recodificación de la variable Estadística. En este caso nos encontramos ante un intervalo de confianza para la diferencia de dos proporciones. Seleccionamos: Estadísticos/Proporciones/Test de proporciones para dos muestras. Se muestra la siguiente pantalla

La pestaña Datos muestra dos listas con las variables cualitativas que incluye el conjunto de datos. De la primera lista seleccionamos la variable de agrupación (en nuestro caso es el Grupo) y de la segunda, la variable de interés (Estadística). Pulsamos Opciones

Ya en la segunda pestaña (Opciones) indicamos el nivel de confianza propuesto en el enunciado, 93%. Pulsamos Aceptar y se muestra la siguiente salida

```
local({ .Table <- xtabs(~Grupo+Estadistica,
  data=universidad)
  cat("\nPercentage table:\n")
  print(rowPercents(.Table))
  prop.test(.Table, alternative='two.sided', conf.level=.93,
  correct=FALSE)
})
```

```
##
## Percentage table:
##      Estadistica
## Grupo  No  Si Total Count
##      A 33.3 66.7   100     3
##      B 50.0 50.0   100     2
```

```
## Warning in prop.test(.Table, alternative = "two.sided", conf.level = 0.93, :
## Chi-squared approximation may be incorrect
```

```
##
## 2-sample test for equality of proportions without continuity correction
```

```
##
## data: .Table
## X-squared = 0.13889, df = 1, p-value = 0.7094
## alternative hypothesis: two.sided
## 93 percent confidence interval:
## -0.9750999 0.6417665
## sample estimates:
## prop 1 prop 2
## 0.3333333 0.5000000
```

Por último, el programa devuelve el intervalo de confianza al 93% de confianza, (-0.9750999, 0.6417665). El 0 está dentro de este intervalo, por lo que podemos concluir que las proporciones de alumnos en ambos grupos que no tienen clase de estadística coinciden.

Ejercicio Guiado 2

En un hospital se elige una muestra de pacientes y se les mide la tasa cardíaca por la mañana (TCM) y a última hora de la tarde (TCT).

studiar mediante un intervalo de confianza al 99% si, por término medio, la tasa cardíaca es igual por la mañana y a última hora de la tarde.

Solución:

En primer lugar debemos importar el archivo de datos tcardiaca.txt.

Para ello, seleccionamos en el menú principal: Datos/Importar datos/desde archivo de texto

```
cardiaca <-
  read.table("C:/Users/Usuario/Desktop/respaldo/Desktop/PAQUETE R/PRACTICAS_S10/tcardiaca.txt",
    header=TRUE, stringsAsFactors=TRUE, sep=" ", na.strings="NA",
    dec=".", strip.white=TRUE)
```

En este caso nos encontramos ante un intervalo de confianza para la diferencia medias en dos poblaciones normales relacionadas. Seleccionamos: Estadísticos/Medias/Test t para datos relacionados. Se muestra la siguiente pantalla

La pestaña Datos muestra dos listas de variables cada una de las cuales incluye todas las variables cuantitativas que son susceptibles de ser analizadas. Seleccionamos en cada lista la variable que nos interese (TCM y TCT en nuestro caso).

Pulsamos Opciones

En la pestaña Opciones podemos personalizar el contraste. Como en este caso sólo nos interesa el intervalo de confianza, introducimos el valor del nivel de confianza dado en el enunciado (99%). Si hacemos clic en Aceptar, el programa nos devuelve la siguiente salida:

```
with(cardiaca, (t.test(TCM, TCT, alternative='two.sided',
  conf.level=.99, paired=TRUE)))
```

```
##
## Paired t-test
##
## data: TCM and TCT
## t = -1.6236, df = 4, p-value = 0.1798
```

```
## alternative hypothesis: true mean difference is not equal to 0
## 99 percent confidence interval:
##  -18.411312   8.811312
## sample estimates:
## mean difference
##           -4.8
```

En la segunda parte de los resultados se incluye el intervalo de confianza al 99% para la diferencia de las medias de la variable tasa cardíaca, que es (-18.411312, 8.811312). Este intervalo incluye el valor 0, lo que significa que el 0 es un valor posible para la diferencia entre las medias. Por ello, concluimos que puede asumirse que la diferencia entre dichas medias es 0, o dicho de otro modo, que ambas tasas cardíacas son iguales.

Ejercicio Guiado 3.

Una determinada empresa quiere saber si su nuevo producto tendrá más aceptación en la población de hombres o entre las mujeres. Para ello, considera una muestra aleatoria de 38 hombres y 62 mujeres, observando que sólo a 15 hombres y 33 mujeres les había gustado su producto. Construir un intervalo de confianza al 99% de confianza para la diferencia de proporciones de hombres y mujeres a los que les gusta el producto. ¿Puede suponerse que el producto gusta por igual en hombres y mujeres?

Solución:

En primer lugar debemos importar el archivo de datos empresa.xls

Para ello, seleccionamos en el menú principal: Datos/Importar datos/desde un archivo de Excel

```
empresa <-
  readXL("C:/Users/Usuario/Desktop/respaldo/Desktop/PAQUETE R/PRACTICAS_S10/empresa2.xls",
    rownames=FALSE, header=TRUE, na="", sheet="Respuestas",
    stringsAsFactors=TRUE)
```

Dado que la hipótesis que se ha planteado se ha hecho la diferencia entre mujeres y hombres a los que les gusta el producto, es necesario hacer una recodificación de la variable Producto. Para ello seleccionamos Datos/Modificar variables del conjunto de datos activo/Recodificar variables

En la casilla: Nuevo nombre o prefijo para la variable múltiple recodificadas ponemos Producto_rec.

```
empresa <- within(empresa, {
  Producto_rec <- Recode(Producto, "Si"="1Si"; "No"="2No",
    as.factor=TRUE, to.value="", interval=":", separator=";")
})
```

Una forma de recodificar la variable es ponerle un número delante del carácter. De este modo la primera categoría de la variable es aquella con el número más bajo, en este caso Sí

Se pulsa Aceptar y se muestra el siguiente conjunto de datos

A continuación realizamos un intervalo de confianza para la diferencia de dos proporciones. Seleccionamos: Estadísticos/Proporciones/Test de proporciones para dos muestras. Se muestra la siguiente pantalla

La pestaña Datos muestra dos listas con las variables cualitativas que incluye el conjunto de datos. De la primera lista seleccionamos la variable de agrupación (en nuestro caso es el Sexo) y de la segunda, la variable de interés (Producto_rec).

Pulsamos Opciones

En la pestaña Opciones podemos personalizar el contraste. Como en este caso sólo nos interesa el intervalo de confianza, introducimos el valor del nivel de confianza dado en el enunciado (99%). Si hacemos clic en Aceptar, el programa nos devuelve la siguiente salida:

```
local({ .Table <- xtabs(~Sexo+Producto_rec, data=empresa)
  cat("\nPercentage table:\n")
  print(rowPercents(.Table))
  prop.test(.Table, alternative='two.sided', conf.level=.99,
    correct=FALSE)
})

##
## Percentage table:
##      Producto_rec
## Sexo      2No    Sí Total Count
##  Hombre 60.5 39.5   100     38
##   Mujer 46.8 53.2   100     62

##
## 2-sample test for equality of proportions without continuity correction
##
## data: .Table
## X-squared = 1.7851, df = 1, p-value = 0.1815
## alternative hypothesis: two.sided
## 99 percent confidence interval:
##  -0.1239329  0.3989753
## sample estimates:
##    prop 1    prop 2
## 0.6052632 0.4677419
```

Según los resultados, el intervalo de confianza para la diferencia de proporciones que buscamos es (-0.1239329 0.3989753), el cual incluye al 0, por lo que se puede afirmar que el producto gusta por igual entre hombres y mujeres.

Ejercicio Guiado 4.

En una experiencia genética se extraen 20 moscas de una caja experimental y se mide la longitud del ala de cada una. Se obtuvieron los siguientes valores:

93, 90, 97, 90, 93, 91, 96, 94, 91, 91, 88, 93, 95, 91, 89, 92, 87, 88, 90, 86

Suponiendo que la longitud del ala sigue una distribución Normal. Construir un intervalo de confianza al 99% de confianza para

- La media μ
- La varianza σ^2

Solución:

En primer lugar debemos importar el archivo de datos moscas.txt

```
moscas <-
  read.table("C:/Users/Usuario/Desktop/respaldo/Desktop/PAQUETE R/PRACTICAS_S10/moscas.txt",
    header=TRUE, stringsAsFactors=TRUE, sep=" ", na.strings="NA",
    dec=".", strip.white=TRUE)
```

Hay que obtener un intervalo de confianza cuando la varianza poblacional es desconocida

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

```
with(moscas, (t.test(Longitud, alternative='two.sided',
  mu=0.0, conf.level=.99)))
```

```
##
## One Sample t-test
##
## data: Longitud
## t = 139.01, df = 19, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 99 percent confidence interval:
## 89.37195 93.12805
## sample estimates:
## mean of x
## 91.25
```

El intervalo de confianza para la longitud media de las alas al 99% de confianza, es (89.37195, 93.12805)

b) Construir un intervalo de confianza para la varianza σ^2 al 99% de confianza.

El intervalo de confianza que debemos calcular es

$$\left(\frac{(n-1)\sigma^2}{X_{1-\alpha/2}^2}, \frac{(n-1)\sigma^2}{X_{\alpha/2,n-1}^2} \right)$$

Expresión 28: Expresión del Intervalo de confianza para la varianza (media conocida)

Nota: R-Commander no incluye una función específica para el cálculo de intervalos de confianza en este tipo de situaciones. Por lo tanto calcularemos el intervalo de la siguiente forma

```
n <- length(moscas$Longitud)
varianza <- var(moscas$Longitud)
L1 <- (n-1) * varianza/qchisq(1-alpha/2,n-1)
L2 <- (n-1) * varianza/qchisq(alpha/2,n-1)
IC <- c(L1,L2)
IC
```

```
## [1] 4.244179 23.926166
```

El intervalo pedido es: (4.244179, 23.926166)

Ejercicios Propuestos

Ejercicio Propuesto 1

En la tabla siguiente se muestran los salarios mensuales en euros de 10 trabajadores de Madrid y Barcelona. Se pide:

- Obtener un intervalo de confianza a un nivel del 89% para el salario medio entre ambas ciudades
- Obtener un intervalo de confianza a un nivel del 90% para la diferencia media de salarios entre ambas ciudades. ¿Se pueden considerar iguales?
- Obtener un intervalo de confianza a un nivel del 90% para la proporción de trabajadores en Barcelona.

Solucion:

En primer lugar debemos importar el archivo de datos salario.xls

Para ello, seleccionamos en el menú principal: Datos/Importar datos/desde un archivo de Excel.

```
library(Rcmdr)
```

```
salario <- readXL("C:/Users/Usuario/Desktop/respaldo/Desktop/PAQUETE R/PRACTICAS_S10/salario.xlsx", rownames=FALSE, header=TRUE, na="", sheet="Salario")
```

- Obtener un intervalo de confianza a un nivel del 89% para el salario medio entre ambas ciudades

En este caso nos encontramos ante un intervalo de confianza sobre la media de una población normal con varianza desconocida. Para resolverlo, seleccionamos Estadísticos/Medias/Test t para una muestra

Ejercicio Propuesto 2

Para comprobar si un determinado fertilizante puede mejorar la producción de manzanas, se selecciona una muestra aleatoria simple de 10 árboles. En la tabla siguiente se muestra el peso (en Kgr) de manzanas por árbol recogidas antes y después del tratamiento

Obtener un intervalo de confianza al 98% para la diferencia de los pesos medios producida antes y después del tratamiento.

Solución:

En primer lugar se importa el archivo de datos manzanas.xls,

El intervalo de confianza al 98% para la diferencia de los pesos medios producida antes y después del tratamiento

```
manzanas <- readXL("C:/Users/Usuario/Downloads/manzanas.xlsx", rownames=FALSE, header=TRUE, na="", sheet="Manzanas")
```

```
with(manzanas, (t.test(Antes, Después, alternative='two.sided', conf.level=.98, paired=TRUE)))
```

```
##
## Paired t-test
##
## data:  Antes and Después
```

```
## t = -2.1262, df = 9, p-value = 0.0624
## alternative hypothesis: true mean difference is not equal to 0
## 98 percent confidence interval:
## -1.0006005  0.1406005
## sample estimates:
## mean difference
## -0.43
```

En la segunda parte de los resultados se incluye el intervalo de confianza al 98% para la diferencia de las medias los pesos, que es (-1.0006005, 0.1406005). Este intervalo incluye el valor 0, lo que significa que el 0 es un valor posible para la diferencia entre las medias. Por ello, concluimos que puede asumirse que la diferencia entre dichas medias es 0, o dicho de otro modo, que ambos pesos son iguales.

Ejercicio Propuesto 3

En una muestra aleatoria de 150 personas con pelo oscuro se encontró que 90 de ellas tenían los ojos azules. Construir un intervalo de confianza al 95% para la proporción de individuos que teniendo pelo oscuro en la población posee ojos azules. ¿Son compatibles estos resultados con la suposición de que dicha proporción vale 2/3).

Solución:

En primer lugar debemos importar el archivo de datos propuesto3.xls

```
p3 <- readXL("C:/Users/Usuario/Downloads/propuesto3-1.xlsx", rownames=FALSE, header=TRUE, na="", sheet=
```

Recodificación de la variable

```
p3 <- within(p3, {
  azules_rec <- Recode(Azules, "Si"="1Si"; "No"="2No", as.factor=TRUE, to.value="", interval=":", sep=
})
```

Dado que la hipótesis que se ha planteado se ha hecho sobre tener los ojos azules es necesario hacer una recodificación de la variable. Para ello seleccionamos Datos/Modificar variables del conjunto de datos activo/Recodificar variables

```
local({
  .Table <- xtabs(~ azules_rec , data= p3 )
  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):
## azules_rec
## 2No  Si
## 60  90

##
## 1-sample proportions test without continuity correction
##
```

```
## data:  rbind(.Table), null probability 0.5
## X-squared = 6, df = 1, p-value = 0.01431
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.3250432 0.4799508
## sample estimates:
##      p
## 0.4
```

El intervalo de confianza para la proporción de individuos que teniendo pelo oscuro en la población posee ojos azules, a un nivel del 95%, es (0.3250432, 0.4799508). Este resultado si es compatible con la suposición de que dicha proporción vale $2/3$, ya que $2/3$ no pertenece al intervalo.

Ejercicio Propuesto 4

En una piscifactoría se desea comparar el porcentaje de peces adultos que miden menos de 20 cm con los que miden más de 40 cm. Para ello, se toma una muestra de 200 peces observando que 40 de ellos miden menos de 20 cm y una muestra de 200 peces de los que 57 miden más de 40 cm. Halla un intervalo de confianza para:

- La diferencia de proporciones de peces adultos que miden más de 40 cm con los que miden menos de 20 cm al nivel de confianza del 0.95
- La diferencia de proporciones de peces adultos que miden menos de 20 cm con los que miden más de 40 cm al nivel de confianza del 0.95

Solución:

En primer lugar debemos importar el archivo de datos `propuesto4.xls`

```
p4 <- readXL("C:/Users/Usuario/Downloads/propuesto4-1.xlsx", rownames=FALSE, header=TRUE, na="", sheet=
```

- La diferencia de proporciones de peces adultos que miden más de 40 cm con los que miden menos de 20 cm al nivel de confianza del 0.95

recodificación de la variable

```
p4 <- within(p4, {
  resultados_rec <- Recode(resultado, "Si"="1Si"; "No"="2No", as.factor=TRUE, to.value="", interval=
})
```

Nos piden un intervalo de confianza para la diferencia de proporciones de peces adultos que miden más de 40 cm con los que miden menos de 20 cm. Seleccionamos: Estadísticos/Proporciones/Test de proporciones para dos muestras.

```
library(abind, pos=17)
local({ .Table <- xtabs(~medida+resultados_rec, data=p4)
  cat("\nPercentage table:\n")
  print(rowPercents(.Table))
  prop.test(.Table, alternative='two.sided', conf.level=.95, correct=FALSE)
})
```



```
##
## Percentage table:
##           resultados_rec
## medida      2No    Sí Total Count
##   masde40    71.5  28.5   100   200
##   menosde20  80.0  20.0   100   200

##
## 2-sample test for equality of proportions without continuity correction
##
## data: .Table
## X-squared = 3.9332, df = 1, p-value = 0.04734
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.168589075 -0.001410925
## sample estimates:
## prop 1 prop 2
##  0.715  0.800
```

El intervalo pedido es: (-0.168589075, -0.001410925)

- b) La diferencia de proporciones de peces adultos que miden menos de 20 cm con los que miden más de 40 cm al nivel de confianza del 0.95

El intervalo pedido es de peces adultos que miden menos de 20 cm con los que miden más de 40 cm. Tenemos que recodificar la variable medida. Para ello, en la ventana de Recodificar Variables

```
p4 <- within(p4, {
  medida_rec <- Recode(medida, '"menosde20"="1menosde20"; "masde40"="2masde40";', as.factor=TRUE, to.va
})
```

```
local({ .Table <- xtabs(~medida_rec+resultados_rec, data=p4)
  cat("\nPercentage table:\n")
  print(rowPercents(.Table))
  prop.test(.Table, alternative='two.sided', conf.level=.95, correct=FALSE)
})
```

```
##
## Percentage table:
##           resultados_rec
## medida_rec      2No    Sí Total Count
##   1menosde20  80.0  20.0   100   200
##   2masde40    71.5  28.5   100   200

##
## 2-sample test for equality of proportions without continuity correction
##
## data: .Table
## X-squared = 3.9332, df = 1, p-value = 0.04734
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.001410925 0.168589075
```

```
## sample estimates:  
## prop 1 prop 2  
## 0.800 0.715
```

ntervalo de confianza para la diferencia de proporciones entre los peces adultos que miden menos de 20 cm y los que miden más de 40 cm al 95% (0.001410925, 0.168589075).