PRACTICA 6, SEMANA 10

Nelson de Jesus Magaña Godinez

2022-10-11

Contents

CONTRASTES DE HIPÓTESIS

Conceptos básicos.

Contraste de hipótesis. Un contraste de hipótesis (también conocido como test de hipótesis) es una técnica estadística que se utiliza para comprobar la validez de una afirmación en base a la información recogida en una muestra de observaciones. Es un proceso estadístico mediante el cual se investiga si una propiedad que se supone que cumple una población es compatible con lo observado en una muestra de dicha población. Es un procedimiento que permite elegir una hipótesis de trabajo de entre dos posibles y antagónicas.

Hipótesis Estadística. Todo contraste de hipótesis se basa en la formulación de dos hipótesis exhaustivas y mutuamente exclusivas:

1. Hipótesis nula (H_0)
2. Hipótesis alternativa (H_1)

Contrastes de hipótesis paramétricos.

El propósito de los contrastes de hipótesis es determinar si un valor propuesto (hipotético) para un parámetro u otra característica de la población debe aceptarse como plausible con base en la evidencia muestral.

Podemos considerar las siguientes etapas en la realización de un contraste:

1. El investigador formula una hipótesis sobre un parámetro poblacional, por ejemplo que toma un determinado valor
2. Selecciona una muestra de la población
3. Comprueba si los datos están o no de acuerdo con la hipótesis planteada, es decir, compara la observación con la teoría

- Si lo observado es incompatible con lo teórico entonces el investigador puede rechazar la hipótesis planteada y proponer una nueva teoría
- Si lo observado es compatible con lo teórico entonces el investigador puede continuar como si la hipótesis fuera cierta.

Contrastes de hipótesis para la media de una población normal.

El objetivo es probar uno de los siguientes contrastes de hipótesis con respecto de μ

$H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$

$H_0 : \mu \leq \mu_0$ $H_1 : \mu > \mu_0$

$H_0 : \mu \geq \mu_0$ $H_1 : \mu < \mu_0$

donde μ_0 es un valor conocido dado de antemano. Para ello se toma una mas concreta $x_1, x_2, ..., x_n$ cuya media valdrá:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Se distinguen dos situaciones: a) Varianza poblacional conocida y b) varianza poblacional desconocida.

Contrastes de hipótesis para la media de una población normal con Varianza conocida.

El caso en el que se desea resolver un contraste de hipótesis para la media de una variable continua y, además, se conoce el valor de la varianza de dicha variable en toda las poblaciones es el más sencillo de todos y, a la vez, el menos usual.

Supongamos una muestra aleatoria $X_1, X_2, ..., X_n$ de tamaño n de valores de una variable aleatoria que sigue una distribución normal de media μ desconocida, y de desviación típica σ conocida. Se plantea el siguiente contraste:

$H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$

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

que sigue una distribución normal de media 0 y desviación típica 1 cuando la hipótesis nula es cierta. A continuación se busca el cuantil $1 - \alpha/2$ de una distribución normal y se comparan ambos valores.

En el contraste de hipótesis bilateral, si el valor absoluto del estadístico de contraste es mayor que el cuantil, se rechazará la hipótesis nula. En caso contrario, no se rechazará.

En el contraste de hipótesis unilateral

- Con hipótesis alternativa del tipo <, el valor crítico $-z_1 - \alpha$ y la hipótesis nula se rechaza cuando $Z < -z_1 - \alpha$
- Con hipótesis alternativa del tipo >. el valor crítico $-z_1 > -\alpha$ y la hipótesis nula se rechaza cuando $Z > -z_1 - \alpha$

R no incluye una función específica para la resolución de contrastes de hipótesis de este tipo. Aun así, pueden resolverse de una forma muy sencilla como se muestra en el siguiente ejemplo.

Supuesto Práctico 1

Consideramos la base de datos empleados.xls que contiene una serie de variables medidas en los empleados de una empresa. A partir de dicha base de datos y suponiendo la normalidad de la variable altura, ¿puede concluirse con un 95% de confianza que la altura media de los empleados es de 185 cm, sabiendo que la varianza poblacional es 6?

Abrimos R-Commander, para ello, en primer lugar nos situamos en R y mediante la instrucción:

Se abre la siguiente ventana que corresponde a R-Commander

```
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
```

Los datos de Excel, empleados.xls, los importamos a R-Commander mediante el Menú

Datos/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)
```

Pulsamos Editar conjunto de datos o Visualizar conjunto de datos y nos muestra el conjunto de datos

Los visualizamos y los guardamos (Datos/Conjunto de datos activo/Guardar el conjunto de datos activo) como fichero empleados.RData.

El contraste de hipótesis asociado a este ejercicio es

$H_0 : \mu = 185$ $H_1 : \mu \neq 185$ Expresión 8: Contraste de hipótesis del supuesto práctico 1

En R-Commander no existe ninguna función que nos ayude a resolver este contraste directamente, de manera que debemos calcular el valor del estadístico de contraste y ver si verifica o no la condición de rechazo.

En R-Script escribimos las sentencias

```
alpha<- 0.05
varianza <- 6
mu0<-185
n <- nrow(empleados)
media <- mean(empleados$Altura)
z0<-(media-mu0)/(sqrt(varianza)/sqrt(n))
z0
```

```
## [1] -32.49615
```

```
cuantil<-qnorm(1-alpha/2)
cuantil
```

```
## [1] 1.959964
```

Las señalizamos todas y pulsamos Ejecutar, se muestra la siguiente salida

```r
z0
```

```
## [1] -32.49615
```

```r
cuantil
```

```
## [1] 1.959964
```

Como $|Z_0| > Z_1 - \alpha/2$ rechazamos la hipótesis nula, por lo que la altura de los empleados es distinta de 185.

Contrastes de hipótesis para la media de una población normal con Varianza desconocida

Supongamos que la varianza poblacional de la variable de interés es desconocida. Nuestro objetivo sigue siendo la resolución del contraste de hipótesis para la media de dicha variable.

$H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$

$H_0 : \mu \leq \mu_0$ $H_1 : \mu > \mu_0$

$H_0 : \mu \geq \mu_0$ $H_1 : \mu < \mu_0$

Supongamos, de nuevo, una muestra aleatoria $X_1, X_2, ..., X_n$, de tamaño n de valores de la variable aleatoria que sigue una distribución normal de media μ y desviación típica σ, ambas desconocidas. Para resolver el contraste de hipótesis para μ en este caso partimos del estadístico de contraste

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

Supuesto Práctico 2

Considerando el conjunto de datos de empleados y asumiendo que la variable que mide la edad de los empleados sigue una distribución normal con varianza desconocida, contrastar con un nivel de significación del 10% si la edad media poblacional puede considerarse igual a 25 años frente a que esta edad es menor.

En primer lugar, planteamos el contraste de hipótesis asociado a este supuesto

$H_0 : \mu = 25$ $H_1 : \mu < 25$

Para calcular un contraste de hipótesis para la media de una población normal con varianza desconocida mediante R-Commander, elegimos la opción

Estadísticos/ Medias/ Test t para una muestra

- Elegimos una variable, que debe ser aquella cuya media estemos contrastando.
- Indicamos cuál es la hipótesis alternativa que vamos a contrastar.
- Especificamos el valor con el que estamos comparando la media.
- Especificamos el nivel de confianza, que se calcula como 1 - nivel de significación (en tanto por uno). Pulsamos Aceptar y se muestra la siguiente salida

```r
with(empleados, (t.test(Edad, alternative='less', mu=25,
  conf.level=.90)))
```

```
##
##  One Sample t-test
##
## data:  Edad
## t = -10.718, df = 98, p-value < 2.2e-16
## alternative hypothesis: true mean is less than 25
## 90 percent confidence interval:
##      -Inf 21.05504
## sample estimates:
## mean of x
##  20.51515
```

Entre los resultados que proporciona R-Commander encontramos, por ejemplo, el valor del estadístico de contraste (t = -10.718). Pero en este caso, la resolución del contraste se hará basándonos en el p-valor (p-value = < 2.2e-16)

El p-valor es una probabilidad (oscila, por lo tanto, entre 0 y 1).

- Si el p-valor es mayor que el nivel de significación, no rechazamos la hipótesis
- Si el p-valor es menor que el nivel de significación, rechazamos la hipótesis nula en favor de la hipótesis alternativa En nuestro caso, el p-valor vale < 2.2e-16. El nivel de significación es $\alpha = 0{,}10$. Como 2.2e-16 < 0,10, rechazamos la hipótesis nula, por lo que podemosnconsiderar que la edad media de los empleados es menor de 25 años.

Contrastes de hipótesis para la diferencias de medias de dos poblaciones normales e independientes.

De un modo general, dos muestras se dice que son independientes cuando las observaciones de una de ellas no condicionan para nada a las observaciones de la otra, siendo dependientes en caso contrario. En realidad, el tipo de dependencia que se considera a estos efectos es muy especial: cada dato de una muestra tiene un homónimo en la otra, con el que está relacionada, de ahí el nombre alternativo de muestras apareadas. Por ejemplo, supongamos que se quiere estudiar el efecto de un medicamento, sobre la hipertensión, a un grupo de 20 individuos. El experimento se podría planificar de dos formas:

1. Aplicando el medicamento a 10 de estos individuos y dejando sin tratamiento al resto. Transcurrido un tiempo se miden las presiones sanguíneas de ambos grupos y se contrasta la hipótesis $H_0 : \mu 1 = \mu_2$ vs $H_1 : \mu 1 \neq \mu 2$ para evaluar si las medias son iguales o no. Como las muestras están formadas por individuos distintos sin relación entre sí, se dirá que son muestras independientes.
2. Aplicando el medicamento a los 20 individuos disponibles y anotando su presión sanguínea antes y después de la administración del mismo. En este caso los datos vienen dados por parejas, presión antes y después y tales datos están relacionados entre sí. Las muestras son apareadas. Consideramos ahora dos variables aleatorias independientes X_1 y X_2 con distribuciones Normales de parámetro (μ_1, σ_1) y $(\mu_2, \sigma 2)$ respectivamente, de las que vamos a tomar muestras aleatorias independientes de tamaños n_1 y n_2, respectivamente.

Nuestro objetivo, en este caso, es resolver un contraste de hipótesis para la diferencia de las medias de ambas distribuciones, es decir, para μ_1 y μ_2. Este contraste presentará alguna de las formas que se muestran a continuación

$H_0 : \mu_1 - \mu_2 = d_0$ $H_0 : \mu_1 - \mu_2 \neq d_0$

$H_0 : \mu_1 - \mu_2 \leq d_0$ $H_0 : \mu_1 - \mu_2 > d_0$

$H_0 : \mu_1 - \mu_2 \geq d_0$ $H_0 : \mu_1 - \mu_2 < d_0$

Supuesto Práctico 3

Continuando con los datos relativos a los empleados y asumiendo que el peso en hombres y el peso en mujeres se distribuyen según distribuciones normales con medias y varianzas desconocidas. Contrastar si el peso en ambas poblaciones puede considerarse igual con un nivel de confianza del 95%

Solución

En primer lugar vamos a realizar el contraste sobre la igualdad de varianzas

$H_0 : \sigma^2_{hombre} = \sigma^2_{mujer}$ $H_0 : \sigma^2_{hombre} \neq \sigma^2_{mujer}$

Resolvemos el contraste de hipótesis sobre las varianzas con R-Commander mediante el menú:

Estadísticos / Varianzas / Test F para dos varianzas

```
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
```

Submenú Datos: En el cuadro de la izquierda seleccionamos la variable que establece los grupos, Sexo, mientras que en el cuadro de la derecha elegimos la variable cuya varianza queremos contrastar en ambos grupos, Peso.

Pulsamos la pestaña Opciones

Submenú Opciones: Indicamos la forma de la hipótesis nula (en el contraste de varianzas será siempre bilateral) y el nivel de confianza en base al cual se calcularán los resultados.

Pulsamos Aceptar y se muestra la siguiente salida

```
Tapply(Peso ~ Sexo, var, na.action=na.omit, data=empleados)
```

```
##    Hombre     Mujer
## 166.99626  92.06061
```

```
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
```

La primera parte de los resultados incluye el valor del estadístico de contraste junto con los grados de libertad del numerador y del denominador. Justo a continuación se muestra el p-valor asociado al contraste. Teniendo en cuenta la significación que se ha fijado para este contraste (el 5%), rechazaremos la hipótesis nula a favor de la alternativa cuando el p-valor sea menor que 0.05. En caso contrario, no contaríamos con evidencia muestral para el rechazo de tal hipótesis nula.

En este Supuesto se verifica que p-valor $= 0.2752 > 0{,}05 = \alpha$, por lo que no rechazaríamos la hipótesis nula del contraste, lo que equivale a decir que la varianza de la variable peso puede suponerse igual en los dos grupos.

Una vez contrastada la igualdad de varianzas, resolveremos el contraste sobre la igualdad de medias.

$H_0 : \mu_{hombre} = \mu_{mujer}$ $H_1 : \mu_hombre \neq \mu_{mujer}$ Para ello, debemos acceder al menú

Estadísticos / Medias / Test t para muestras independientes

Submenú Datos: En el cuadro de la izquierda seleccionamos la variable que establece los grupos, Sexo, mientras que en el cuadro de la derecha elegimos la variable cuya media queremos contrastar en ambos grupos, Peso.

Submenú Opciones: Indicamos la forma de la hipótesis nula y el nivel de confianza en base al cual se calcularán los resultados. El test de igualdad de varianzas que hemos realizado anteriormente nos permite responder a la pregunta sobre si las varianzas pueden asumirse iguales o no. Pulsamos Aceptar

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 t
## 95 percent confidence interval:
##   4.167581 19.556557
## sample estimates:
## mean in group Hombre  mean in group Mujer
##             76.19540             64.33333
```

En la primera parte de los resultados se muestra el valor del estadístico de prueba, así como los grados de libertad correspondientes. Como es habitual al realizar un contraste de hipótesis también aparece el p-valor, que para este ejemplo toma el valor 0.002865. Como este p-valor es menor que 0.05 (recordemos que estamos considerando una significación del 5%) rechazamos la hipótesis nula a favor de la alternativa, es decir, el peso medio de hombres y mujeres no puede considerarse igual.

Contrastes de hipótesis para la diferencia de medias de dos poblaciones normales relacionadas.

Hasta ahora, hemos considerado que las muestras que se extraen de las dos poblaciones son independientes. En algunas situaciones puede que esto no sea así. Supongamos, por ejemplo, que queremos medir el efecto de un fármaco en un grupo de personas. Para ello, lo lógico es realizar las mediciones oportunas antes y después del suministro del fármaco sobre el mismo grupo de pacientes. Es por ello por lo que las muestras de las dos poblaciones no serán, en ningún caso, independientes. En estos casos, se dice que las muestras son pareadas o están relacionadas y la resolución de contrastes de hipótesis conlleva el uso de técnicas estadísticas distintas a las que se usan cuando las muestras son independientes. Se dice que dos muestras $X_1, X_2, ..., X_n$ e $Y_1, Y_2, ..., Y_n$ 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.

Sean $X_1, X_2, ..., X_n$ e $Y_1, Y_2, ..., Y_n$ 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 $N(\mu_2, \sigma_2)$.

El contraste que debemos resolver será alguno de los siguientes:

$H_0 : \mu_1 - \mu_2 = d_0$ $H_0 : \mu_1 - \mu_2 \neq d_0$

$H_0 : \mu_1 - \mu_2 \leq d_0$ $H_0 : \mu_1 - \mu_2 > d_0$

$H_0 : \mu_1 - \mu_2 \geq d_0$ $H_0 : \mu_1 - \mu_2 < d_0$

Supuesto Práctico 4

Se desea evaluar la eficacia de un fármaco para la reducción del nivel de glucosa en los empleados de una fábrica. Para ello, se selecciona una muestra de 10 empleados 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:

Puede suponerse, a un nivel de confianza del 90% que el medicamento es eficaz en el sentido de que su ingesta implica una reducción en el nivel medio de glucosa en sangre?

En este caso, nos piden resolver el siguiente contraste de hipótesis:

$H_0 : \mu_{Glu-An} = \mu_{Glu-Des}$ $H_1 : \mu_{Glu-An} < \mu_{Glu-Des}$

el cual equivale, claramente, a este otro

$H_0 : \mu_{Glu-An} - \mu_{Glu-Des} = 0$ $H_0 : \mu_{Glu-An} - \mu_{Glu-Des} < 0$

En primer lugar, debemos crear un nuevo conjunto de datos Datos / Nuevo conjunto de datos con la información que nos proporciona la tabla 1. El conjunto de datos estará formado por dos variables con los niveles de glucosa antes y después de la aplicación del fármaco.

```
p4 <-
  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)
```

En R-Commander, las opciones para resolver un contraste de hipótesis sobre las medias de una variable cuantitativa cuando las muestras están relacionadas se encuentran en Estadísticos / Medias / Test t para datos relacionados.

En la primera pestaña que aparece, Datos, encontramos 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 convenga (Antes y Después en nuestro caso).

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 segunda pestaña, Opciones, podemos personalizar el contraste conforme al problema que estemos resolviendo.

Así, debemos indicar qué forma tiene la hipótesis alternativa: bilateral (hipótesis alternativa de la forma $\neq$), de diferencia negativa (hipótesis alternativa de la forma $<$) o de diferencia positiva (hipótesis alternativa de la forma $>$). También podemos modificar el nivel de confianza de acuerdo al enunciado del problema.

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

```
with(p4, (t.test(Antes, Despues, alternative='less',
  conf.level=.90, paired=TRUE)))
```

```
##
##  Paired t-test
##
## data:  Antes and Despues
## t = 0.52714, df = 9, p-value = 0.6946
## alternative hypothesis: true mean difference is less than 0
## 90 percent confidence interval:
##      -Inf 13.58867
## sample estimates:
## mean difference
##            3.75
```

La primera parte de la salida incluye el valor del estadístico de contraste junto con los grados de libertad correspondientes, así como el p-valor asociado. En este caso, dicho p-valor es 0.6946. Dada la significación que se ha fijado para el contraste (10%), rechazaremos la hipótesis nula siempre y cuando el p-valor sea inferior a 0.10. En cualquier otro caso, no contaremos con evidencia para rechazar esta hipótesis nula. Por lo tanto deducimos que los niveles de glucosa son iguales antes y después del tratamiento.

Contrastes de hipótesis para el parámetro p de una distribución Binomial.

Supongamos que X es una variable aleatoria con distribución de probabilidad binomial con parámetro n y pi, $X \to B(n, \pi)$, de la que se extrae una muestra aleatoria $X_1, X_2, ..., X_n$ de tamaño n. Sea p la proporción poblacional. Se desea contrastar si el parámetro π puede ser igual a un valor π_0, es decir se desea resolver uno de los siguientes contrastes

Contraste bilateral $H_0 : \pi = \pi_0$ $H_1 : \pi \neq \pi_0$

Contraste unilateral

$H_0 : \pi \geq \pi_0$ $H_1 : \pi < \pi_0$

$H_0 : \pi \leq \pi_0$ $H_o : \pi > \pi_0$

El contraste de hipótesis para el parámetro p (proporción de éxitos) de una distribución Binomial se basa en la distribución del estadístico muestral π para un tamaño muestral n suficientemente grande.

Denotando por $\hat{p}$ la proporción de éxitos de la muestra de una distribución Binomial, se verifica que

$$Z = \frac{\hat{p} - \pi_0}{\sqrt{\frac{\hat{p} - (1 - \hat{p})}{n}}}$$

sigue una distribución normal de media 0 y desviación típica 1 bajo la hipótesis nula.

a) Para la hipótesis alternativa $H_1 : \pi \neq \pi_0$ la correspondiente región de no rechazo es $(-z_{\alpha/2}, z_{\alpha/2})$

b) Para la hipótesis alternativa $H_1 : \pi > \pi_0$ la correspondiente región de no rechazo es $(-\infty, z_\alpha)$
c) Para la hipótesis alternativa $H_1 : \pi < \pi_0$ la correspondiente región de no rechazo es $(-z_\alpha, \infty)$.

Supuesto práctico 5 A partir del conjunto de datos de empleados.xls, una empresa de estudios sociales quiere contrastar si la proporción de hombres es superior al 50% con un nivel de confianza al 95%.

En este caso nos piden resolver el siguiente contraste

$H0 : \pi_{nombres} = 0.5$ $H_1 : \pi_{hombres} > 0.5$

Respuesta

Recordemos que estamos trabajando con la base de datos empleados.xls que contiene una serie de variables medidas en los empleados de una empresa.

En primer lugar, importamos los datos de Excel a R-Commander mediante el Menú

Datos/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)
```

Pulsamos Aceptar.

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 primera pestaña del cuadro de diálogo, submenú Datos, que aparece, encontramos una lista con todas las variables cualitativas que pueden utilizarse en este tipo de contrastes, de entre las cuales tenemos que elegir una. Elegimos Sexo

En la pestaña Opciones tenemos que indicar si la hipótesis alternativa de nuestro contraste es del tipo "distinto de" (!=), "menor que" (<) o "mayor que" (>). También introducimos el valor de la proporción que queremos contrastar, que debe coincidir con el valor empleado en la hipótesis nula, y el nivel de confianza. Por último, podemos elegir entre tres tipos de pruebas diferentes, aunque nosotros siempre seleccionaremos la opción por defecto, que se corresponde con la Aproximación normal.

```
local({
  .Table <- xtabs(~ Sexo , data= empleados )
  cat("\nFrequency counts (test is for first level):\n")
  print(.Table)
  prop.test(rbind(.Table), alternative='greater', 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 = 2.39e-14
## alternative hypothesis: true p is greater than 0.5
## 95 percent confidence interval:
##  0.8145345 1.0000000
## 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 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". 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.

Con respecto a la salida, el programa nos recuerda que estamos realizando un contraste para una proporción en una población y nos indica el valor de la proporción que se está contrastando (0.5). Justo debajo, se incluye el valor del estadístico de contraste junto con los grados de libertad correspondientes. También aparece el p-valor, el cual nos servirá para resolver el contraste. Para este ejemplo, se tiene un p-valor de 2.39e-14. Como este p-valor es mucho más pequeño que 0.05, que es la significación que se ha prefijado para el contraste, rechazamos la hipótesis nula a favor de la alternativa. Por ello, se puede asumir que la proporción de hombres es mayor a 0.5.

Contrastes de hipótesis para la diferencia de proporciones

Consideremos dos muestras aleatorias $X_1, X_2, ..., X_{n_1}$ e $Y_1, Y_2, ..., Y_{n_2}$ de tamaños n_1 y n_2 independientes entre sí, extraídas de poblaciones con distribuciones binomiales $B(n_1, \pi_1)$ y $B(n_2, \pi_2)$, respectivamente. Pretendemos resolver alguno de los siguientes contrastes de hipótesis:

$H_0 : \pi_1 - \pi_2 = \delta_0$ $H_1 : \pi_1 - \pi_2 \neq \delta_0$

$H_0 : \pi_1 - \pi_2 \geq \delta_0$ $H_1 : \pi_1 - \pi_2 < \delta_0$

$H_0 : \pi_1 - \pi_2 \leq \delta_0$ $H_1 : \pi_1 - \pi_2 > \delta_0$

Para ello, partimos del estadístico de contraste

$$Z = \frac{(\hat{p_1} - \hat{p_2}) - \delta_0}{\sqrt{\frac{\hat{p_1}(1-\hat{p_1})}{n_1} + \frac{\hat{p_2}(1-\hat{p_2})}{n_2}}}$$

con $\hat{p_1}$ y $\hat{p_2}$ las proporciones de individuos que presentan la característica de interés en la primera y la segunda muestra, respectivamente. Este estadístico de contraste sigue una distribución normal de media 0 y desviación típica 1 cuando la hipótesis nula del contraste en cuestión es cierta.

Supuesto práctico 6

A partir del conjunto de datos relativo a los empleados, contrastar al 85% si la diferencia de proporciones entre los hombres y mujeres que no tienen coche es la misma.

Respuesta

El contraste que vamos a resolver es

$H_0 : \pi_{hombrenocoche} = \pi_{mujernocoche}$ $H_1 : \pi_{hombrenocoche} \neq \pi_{mujernocoche}$

donde $\pi_{hombrenocoche}$ y $\pi_{mujernocoche}$ representan la proporciones de hombres y mujeres que no tienen coche, respectivamente.

Para realizar un contraste de hipótesis para la diferencia de dos proporciones

Estadísticos/ Proporciones/ Test de proporciones para dos muestras

```
library(abind, pos=17)
local({  .Table <- xtabs(~Sexo+Coche, data=empleados)
  cat("\nPercentage table:\n")
  print(rowPercents(.Table))
  prop.test(.Table, alternative='two.sided', conf.level=.85,
  correct=FALSE)
})
```

```
##
## Percentage table:
##         Coche
## Sexo       No   Sí 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
```

La salida comienza con una tabla de doble entrada que recoge la información de los porcentajes de cada categoría en cada uno de los dos grupos.

A continuación, encontramos el valor del estadístico de contraste, junto con los grados de libertad correspondientes. También aparece el p-valor, que para este ejemplo es 0.8518. Dado que este p-valor no es inferior al nivel de significación, esto es $0.8518 > 0.15$, no podemos rechazar la hipótesis nula del contraste. Por tanto, concluiremos que la proporción de hombres y mujeres que no tienen coche es la misma.

Contrastes de hipótesis no paramétricos.

En la sesión anterior hemos estudiado contrastes de hipótesis acerca de parámetros poblacionales, tales como la media y la varianza, de ahí el nombre de contrastes paramétricos. En estadística paramétrica se trabaja bajo el supuesto de que las poblaciones poseen distribuciones conocidas, donde cada función de distribución teórica depende de uno o más parámetros poblacionales. Sin embargo, en muchas situaciones, es imposible especificar la forma de la distribución poblacional. El proceso de obtener conclusiones directamente de las observaciones muestrales, sin formar los supuestos con respecto a la forma matemática de la distribución poblacional se llama teoría no paramétrica.

En esta sesión vamos a realizar procedimientos que no exigen ningún supuesto, o muy pocos acerca de la familia de distribuciones a la que pertenece la población, y cuyas observaciones pueden ser cualitativas o bien se refieren a alguna característica ordenable. En estos casos, cuando no se dispone de información acerca de qué distribución de probabilidad sigue la variable a nivel poblacional, se pueden utilizar técnicas estadísticas no paramétricas para el planteamiento y resolución de contrastes de hipótesis no paramétricos. .Estas técnicas se basan exclusivamente en la información que se recoge en la muestra para resolver los contrastes.

Así, uno de los objetivos de esta sesión es el estudio de contrates de hipótesis para determinar si una población tiene una distribución teórica específica. La técnica que nos introduce a estudiar esas cuestiones se llama Contraste de la Chi-cuadrado para la Bondad de Ajuste. Una variación de este contraste se emplea para resolver los Contrastes de Independencia. Tales contrastes pueden utilizarse para determinar si dos características (por ejemplo preferencia política e ingresos) están relacionadas o son independientes. Y, por último estudiaremos otra variación del contraste de la bondad de ajuste llamado Contraste de Homogeneidad. Tal contraste se utiliza para estudiar si diferentes poblaciones, son similares (u homogéneas) con respecto a alguna característica. Por ejemplo, queremos saber si las proporciones de votantes que favorecen al candidato A, al candidato B o los que se abstuvieron son las mismas en dos ciudades.

El procedimiento Prueba de la Chi-cuadrado.

Hemos agrupado los procedimientos en los que el denominador común a todos ellos es que su tratamiento estadístico se aborda mediante la distribución Chi-cuadrado. El procedimiento Prueba de Chi-cuadrado tabula una variable en categorías y calcula un estadístico de Chi-cuadrado. Esta prueba compara las frecuencias observadas y esperadas en cada categoría para contrastar si todas las categorías contienen la misma proporción de valores o si cada categoría contiene una proporción de valores especificada por el usuario.

##Contraste de hipótesis no paramétrico para la independencia de los valores de una variable cualitativa.

Supongamos que se dispone de información sobre una variable cualitativa, X, y se quiere comprobar si todas las categorías de la variable aparecen por igual. Es decir, se pretende comprobar si las categorías de la variable son independientes o no. El contraste de hipótesis que se debe resolver es el siguiente:

H_0 : Las categorías de la variable X aparecen igual

H_1 : Las categorías de la variable X no aparecen igual

Para resolver este contraste en R-Commander se utiliza la opción Test Chi-Cuadrado de bondad de ajunte (solo para una variable), que encontramos en Estadísticos /Resúmenes/Distribución de frecuencias.

Supuesto Práctico 7

La directora de un hospital quiere comprobar si los ingresos en el hospital se producen en la misma proporción durante todos los días de la semana. Para ello, se anota el número de ingresos durante una semana cualquiera.

Contrastar, a un nivel de significación del 5%, si la hipótesis de la directora del hospital puede suponerse cierta. ¿Puede asumirse que las proporciones de ingresos de lunes a domingo son (0.15, 0.15, 0.15, 0.15, 0.20, 0.15, 0.05)?

Solución En primer lugar, que tenemos que abrir R-Commander, para ello, nos situamos en R y escribimos la siguiente instrucción:

library(Rcmdr)

Una vez situados en R-Commander, introducimos los datos. Para ello, creamos un fichero de texto como se muestra en la Imagen

```
sp7 <-
  read.table("C:/Users/Usuario/Desktop/respaldo/Desktop/PAQUETE R/PRACTICAS_S10/supuesto7.txt",
   header=TRUE, stringsAsFactors=TRUE, sep="", na.strings="NA",
   dec=".", strip.white=TRUE)
```

Para transformar la tabla de frecuencias en un conjunto de datos (data.frame) con el que R-Commander pueda trabajar hay que escribir las siguientes instrucciones en la ventana R Script

```
P<-rep(sp7$Dias,sp7$frecuencias)
Dias<-data.frame(P)
```

Seleccionamos ambas instrucciones y pulsamos Ejecutar

El contraste que se debe resolver es:

H_0 : Los ingresos en el hospital se producen en la misma proporción todos los días de la semana

H_1 : Los ingresos en el hospital no se producen en la misma proporción todos los días de la semana

Para resolver este contraste seleccionamos en el menú: Estadísticos/ Resúmenes/Distribución de frecuencias

```
local({
  .Table <- with(sp7, table(Dias))
  cat("\ncounts:\n")
  print(.Table)
  cat("\npercentages:\n")
  print(round(100*.Table/sum(.Table), 2))
  .Probs <- c(0.142857142857143,0.142857142857143,
  0.142857142857143,0.142857142857143,0.142857142857143,
  0.142857142857143,0.142857142857143)
  chisq.test(.Table, p=.Probs)
})
```

```
##
## counts:
## Dias
##   Domingo    Jueves     Lunes    Martes Miercoles    Sabado   Viernes
##         1         1         1         1         1         1         1
##
## percentages:
## Dias
##   Domingo    Jueves     Lunes    Martes Miercoles    Sabado   Viernes
##     14.29     14.29     14.29     14.29     14.29     14.29     14.29
```

```
## Warning in chisq.test(.Table, p = .Probs): Chi-squared approximation may be
## incorrect
```

```
##
##  Chi-squared test for given probabilities
##
## data:  .Table
## X-squared = 5.522e-30, df = 6, p-value = 1
```

El estadístico de contraste, que sigue una distribución chi-cuadrado, toma el valor 5.522e-30. Los grados de libertad de la distribución chi-cuadrado para este ejemplo son 6. El p-valor asociado al contraste es mayor que 0.05 por lo que, considerando un nivel de significación del 5%, no se rechaza la hipótesis nula. Es decir, se concluye que los ingresos hospitalarios se producen en la misma proporción todos los días de la semana.

Para comprobar si podemos asumir que las proporciones de ingresos correspondientes a cada día de la semana (de Lunes a Domingo) son (0.15, 0.15, 0.15, 0.15, 0.20, 0.15, 0.05), seguimos los mismos pasos, pero teniendo en cuenta que, ahora, tenemos que introducir los valores de las nuevas proporciones consideradas y que R-Commander ordena los días de la semana alfabéticamente y no en el orden en el que nos los da el enunciado del problema.

Supuesto Práctico 8

Lanzamos un dado 720 veces y obtenemos los resultados que se muestran en la tabla.

Contrastar la hipótesis de que el dado está bien construido.

Solución

En primer lugar vamos a introducir los datos en R-Commander. Para ello, creamos un fichero de texto como el que aparece en la Imagen

```
Dataset <-
  read.table("C:/Users/Usuario/Desktop/respaldo/Desktop/PAQUETE R/PRACTICAS_S10/SUPUST8.txt",
   header=TRUE, stringsAsFactors=TRUE, sep="", na.strings="NA",
   dec=".", strip.white=TRUE)
```

Que el dado esté bien construido equivale a decir que todos sus valores aparecen en la misma proporción. Por tanto, el contraste de hipótesis que se debe resolver es el siguiente:

H_0 : Los valores del dado aparecen en la misma proporción

H_1 : Los valores del dado no aparecen en la misma proporción

Para ello, en el menú seleccionamos: Estadísticos/ Resúmenes/Distribución de frecuencias y se muestra la siguiente ventana en la que tenemos que seleccionar la variable en estudio (P), clickar en Test Chi-Cuadrado de bondad de ajuste.. y pulsar Aceptar.

```
Dataset <- within(Dataset, {
  Resultado <- as.factor(Resultado)
})
```

```
P<-rep(Dataset$Resultado, Dataset$Frecuencias)
Resultados<-data.frame(P)
```

```
local({
  .Table <- with(Resultados, table(P))
  cat("\ncounts:\n")
  print(.Table)
  cat("\npercentages:\n")
  print(round(100*.Table/sum(.Table), 2))
  .Probs <- c(0.166666666666667,0.166666666666667,
  0.166666666666667,0.166666666666667,0.166666666666667,
  0.166666666666667)
  chisq.test(.Table, p=.Probs)
})
```

```
##
## counts:
## P
##   1   2   3   4   5   6
## 116 120 115 120 125 124
##
## percentages:
```

```
## P
##     1     2     3     4     5     6
## 16.11 16.67 15.97 16.67 17.36 17.22
```

```
##
##  Chi-squared test for given probabilities
##
## data:  .Table
## X-squared = 0.68333, df = 5, p-value = 0.9839
```