

# EP04

Jaime Riquelme

2024-10-11

Como equipo N°5, se nos solicitó responder 4 preguntas de la siguiente lista, utilizando R y RStudio. Para ello identificaremos el tipo de prueba a utilizar para cada pregunta, y luego procederemos a realizar el análisis correspondiente junto a utilizar funciones de R para obtener los resultados necesarios. Finalmente, se presentarán las conclusiones obtenidas de cada análisis.

## Pregunta 1

1. Un hotel está implementando un nuevo protocolo de limpieza para mejorar la calidad de servicio y la satisfacción de los huéspedes. Para evaluar la efectividad de este protocolo, se seleccionaron aleatoriamente 25 habitaciones y se aplicó el protocolo de limpieza antiguo y nuevo de forma alternada. Después de la primera limpieza, se evaluó la calidad de limpieza de cada habitación. Luego, después de la segunda limpieza con el protocolo alternativo, se volvió a evaluar la calidad de limpieza. Los resultados se presentan a continuación: 4 habitaciones mostraron una calidad de limpieza baja en ambas ocasiones. 5 habitaciones con calidad de limpieza baja con el protocolo antiguo mostraron una mejora con el nuevo. 9 habitaciones mantuvieron una alta calidad de limpieza con ambos protocolos. 7 habitaciones mostraron una alta calidad de limpieza con el protocolo antiguo, pero una disminución con el protocolo nuevo. ¿Hay evidencia de que el nuevo protocolo de limpieza haya tenido un efecto significativo en la calidad de limpieza de las habitaciones del hotel?

## Respuesta pregunta 1

Para responder esta pregunta utilizaremos una prueba de McNemar, ya que se trata de un diseño de medidas repetidas y se busca determinar si hay una diferencia significativa entre las proporciones con respuestas dicotómicas. Verificamos las condiciones para realizar la prueba:

Las muestras de la prueba tienen la característica de ser apareadas, ya que se evalúa la misma habitación con dos protocolos de limpieza diferentes (25 habitaciones evaluadas en dos ocasiones). con dos posibles exitos, siendo esta de tipo dicotómica.

**Definimos las hipótesis nula y alternativa:** H0: No hay diferencia significativa en la calidad de limpieza de las habitaciones del hotel con el nuevo protocolo de limpieza.

HA: Existe una diferencia significativa en la calidad de limpieza de las habitaciones del hotel con el nuevo protocolo de limpieza.

*# Definimos los datos para la creatividad de la tabla*

```
baja_baja <- 4 #Calidad baja con ambos protocolos
baja_alta <- 5 #Calidad baja con el protocolo antiguo y mejora con el nuevo
alta_baja <- 7 #Calidad alta con el protocolo antiguo y disminución con el nuevo
```

```

alta_alta <- 9 #Calidad alta con ambos protocolos

# Creamos la tabla de contingencia (Antigua vs Nueva)

tabla <- matrix(c(baja_baja, baja_alta, alta_baja, alta_alta), nrow = 2, byrow = TRUE)

#Asginamos nombres a las filas y columnas de la tabla

rownames(tabla) <- c("Baja calidad A", "Alta calidad A")
colnames(tabla) <- c("Baja calidad N", "Alta calidad N")

#Mostramos la tabla de contingencia
print("Tabla de Contingencia (Antiguo vs Nuevo)")

## [1] "Tabla de Contingencia (Antiguo vs Nuevo)"

print(tabla)

##           Baja calidad N Alta calidad N
## Baja calidad A           4           5
## Alta calidad A           7           9

# Realizamos la prueba de McNemar con un nivel de significancia de 0.05

PruebaMcNemar <- mcnemar.test(tabla)

#Mostramos los resultados de la prueba
print("Resultados de la Prueba de McNemar")

## [1] "Resultados de la Prueba de McNemar"

print(PruebaMcNemar)

##
## McNemar's Chi-squared test with continuity correction
##
## data:  tabla
## McNemar's chi-squared = 0.083333, df = 1, p-value = 0.7728

```

## Conclusiones

Con un valor de p de 0.7728, no hay evidencia suficiente para rechazar la hipótesis nula, por lo que no se puede afirmar que el nuevo protocolo de limpieza haya tenido un efecto significativo en la calidad de limpieza de las habitaciones del hotel.

- Una agencia de marketing desea determinar si hay una diferencia significativa en la efectividad de tres estrategias publicitarias utilizadas para promocionar un nuevo producto. Para ello, se ha recopilado información de personas que fueron expuestas a las tres estrategias publicitarias, registrando si mostraron una aceptación (A) o rechazo (R) a cada una de ellas. ¿Qué puede concluir la agencia de marketing sobre la efectividad de las estrategias publicitarias para promover el nuevo producto? Indicación: obtenga la muestra de 50 personas a partir del archivo “EP04 Datos.csv” que se encuentra en el directorio compartido, usando la semilla 347. Considere un nivel de significación  $\alpha=0,05$

## Respuesta pregunta 2

```
#Cargamos las librerias necesarias  
library(dplyr)
```

Para responder esta pregunta, procederemos a verificar las condiciones de nuestra prueba junto a la tabla de contingencia, para definir correctamente el metodo a utilizar.

```
##  
## Attaching package: 'dplyr'  
  
## The following objects are masked from 'package:stats':  
##  
##     filter, lag  
  
## The following objects are masked from 'package:base':  
##  
##     intersect, setdiff, setequal, union
```

```
library(RVAideMemoire)
```

```
## *** Package RVAideMemoire v 0.9-83-7 ***
```

```
library(rcompanion)  
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --  
## v forcats   1.0.0      v readr     2.1.5  
## v ggplot2   3.5.1      v stringr  1.5.1  
## v lubridate 1.9.3      v tibble   3.2.1  
## v purrr     1.0.2      v tidyr    1.3.1
```

```
## -- Conflicts ----- tidyverse_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag()     masks stats::lag()  
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
# Cargamos los datos del archivo "EP04 Datos.csv"
```

```
datoscsv <- read.csv2("EP04 Datos.csv")
```

```
#Obtenemos una muestra de 50 personas a partir de los datos proporcionados
```

```
Muestra_50 <- sample_n(datoscsv, 50)
```

```
#Mostramos los datos iniciales  
head(Muestra_50)
```

```
##      id estrategia_1 estrategia_2 estrategia_3
## 1  291             R             A             R
## 2 1446             A             A             A
## 3  241             R             R             R
## 4  227             R             A             R
## 5  664             R             A             A
## 6   85             A             R             A
```

```
# Definimos la semilla para la generación de la muestra
```

```
set.seed(347)
```

```
# Verificamos las condiciones para realizar la prueba
```

```
# Para realizar esta prueba, utilizaremos la prueba de Q de Cochran, ya que
```

```
# 1. La variable dependiente es dicotómica (aceptación o rechazo)
```

```
#2. se trata medidas repetidas por cada persona, ya que se evalúan las tres estrategias publicitarias.
```

```
#3. Las muestras son relacionadas, ya que se evalúa la misma persona en las tres estrategias publicitarias.
```

```
#4. el tamaño de la muestra es mayor a 40.
```

```
#Dado que cumplimos las condiciones para realizar la prueba de W de Cochran, procedemos a realizar el análisis
```

```
#Llevamos la matriz de datos a formato largo
```

```
Muestra_50_Largo <- Muestra_50 %>% pivot_longer(c("estrategia_1", "estrategia_2", "estrategia_3"), names_to = "Estrategias", values_to = "Aceptacion_Rechazo")
```

```
Muestra_50_Largo[["id"]] <- factor(Muestra_50_Largo[["id"]])
```

```
Muestra_50_Largo[["Estrategias"]] <- factor(Muestra_50_Largo[["Estrategias"]])
```

```
#Mostramos los datos en formato largo
```

```
print("Datos en formato largo")
```

```
## [1] "Datos en formato largo"
```

```
print(Muestra_50_Largo)
```

```
## # A tibble: 150 x 3
##   id   Estrategias Aceptacion_Rechazo
##   <fct> <fct>         <chr>
## 1 291   estrategia_1 R
## 2 291   estrategia_2 A
## 3 291   estrategia_3 R
## 4 1446  estrategia_1 A
## 5 1446  estrategia_2 A
## 6 1446  estrategia_3 A
## 7 241   estrategia_1 R
## 8 241   estrategia_2 R
## 9 241   estrategia_3 R
## 10 227  estrategia_1 R
## # i 140 more rows
```

```

# Realizamos la prueba Q de Cochran con un nivel de significancia de 0.05

Prueba_Cochran <- cochrان.qtest(Aceptacion_Rechazo ~ Estrategias | id, data = Muestra_50_Largo, alpha = 0.05)

#Mostramos los resultados de la prueba
print("Resultados de la Prueba de Q de Cochran")

## [1] "Resultados de la Prueba de Q de Cochran"

print(Prueba_Cochran)

##
## Cochran's Q test
##
## data: Aceptacion_Rechazo by Estrategias, block = id
## Q = 1.8049, df = 2, p-value = 0.4056
## alternative hypothesis: true difference in probabilities is not equal to 0
## sample estimates:
## proba in group estrategia_1 proba in group estrategia_2
##                0.52                0.58
## proba in group estrategia_3
##                0.44

#Falta concluir si se debe usar una prueba pos-hoc o no para este caso dependiendo del valor de p obtenido

```

3. En otro planeta se realiza un estudio sobre la preferencia de hábitat de dos especies de alienígenas. Después de observar a una muestra de 17 alienígenas de la especie EA14012-A y 10 de la especie EA14013-B durante meses, se ha determinado que 4 alienígenas de la primera y 8 de la segunda prefieren hábitats subterráneos, mientras los demás prefieren hábitats acuáticos. ¿Existe relación entre las especies alienígenas y elegir hábitats subterráneos o hábitats acuáticos?

## Respuesta pregunta 3

```

# Definimos los parametros conocidos

EA14012_A <- 17 #Cantidad de alienigenas de la especie EA14012-A
EA14013_B <- 10 #Cantidad de alienigenas de la especie EA14013-B

Subterraneo_A <- 4 #Cantidad de alienigenas de la especie EA14012-A que prefieren hábitats subterráneos
Subterraneo_B <- 8 #Cantidad de alienigenas de la especie EA14013-B que prefieren hábitats subterráneos

Acuatico_A <- EA14012_A - Subterraneo_A #Cantidad de alienigenas de la especie EA14012-A que prefieren hábitats acuáticos
Acuatico_B <- EA14013_B - Subterraneo_B #Cantidad de alienigenas de la especie EA14013-B que prefieren hábitats acuáticos

#Mostramos la cantidad de alienigenas de cada especie que prefieren hábitats subterráneos y acuáticos por especie

print("Cantidad de alienigenas de la especie EA14012-A que prefieren hábitats subterráneos")

```

Para responder esta pregunta, procederemos a verificar las condiciones de nuestra prueba junto a la tabla de contingencia, para definir correctamente el metodo a utilizar.

```
## [1] "Cantidad de alienigenas de la especie EA14012-A que prefieren hábitats subterráneos"
```

```
print(Subterraneo_A)
```

```
## [1] 4
```

```
print("Cantidad de alienigenas de la especie EA14013-B que prefieren hábitats subterráneos")
```

```
## [1] "Cantidad de alienigenas de la especie EA14013-B que prefieren hábitats subterráneos"
```

```
print(Subterraneo_B)
```

```
## [1] 8
```

```
print("Cantidad de alienigenas de la especie EA14012-A que prefieren hábitats acuáticos")
```

```
## [1] "Cantidad de alienigenas de la especie EA14012-A que prefieren hábitats acuáticos"
```

```
print(Acuatico_A)
```

```
## [1] 13
```

```
print("Cantidad de alienigenas de la especie EA14013-B que prefieren hábitats acuáticos")
```

```
## [1] "Cantidad de alienigenas de la especie EA14013-B que prefieren hábitats acuáticos"
```

```
print(Acuatico_B)
```

```
## [1] 2
```

```
# Como nos encontramos a dos variables dicotómicas y además tenemos observaciones menores a 5, utilizar
```

```
# Definimos las hipótesis nula y alternativa
```

```
#H0: Las variables son independientes, es decir, no existe relación entre las especies alienígenas y la
```

```
#HA: Las variables no son independientes, es decir, existe relación entre las especies alienígenas y la
```

```
# Creamos la tabla de contingencia utilizando "rep"
```

```
alienigenas <- c(rep("EA14012-A", EA14012_A), rep("EA14013-B", EA14013_B))
```

```
habitat <- c(rep("Subterráneo", Subterraneo_A), rep("Acuático", Acuatico_A), rep("Subterráneo", Subterraneo_B), rep("Acuático", Acuatico_B))
```

```
datos1 <- data.frame(alienigenas, habitat)
```

```
tabla1 <- xtabs(~.,datos1)
```

```
print("Tabla de Contingencia")
```

```
## [1] "Tabla de Contingencia"
```

```
print(tabla1)
```

```
##           habitat
## alienigenas Acuático Subterráneo
##   EA14012-A      13      4
##   EA14013-B       2      8
```

```
# Realizamos la prueba exacta de Fisher con un nivel de significancia de 0.05
```

```
Prueba_fisher <- fisher.test(tabla1)
print("Resultados de la Prueba de Fisher")
```

```
## [1] "Resultados de la Prueba de Fisher"
```

```
print(Prueba_fisher)
```

```
##
## Fisher's Exact Test for Count Data
##
## data:  tabla1
## p-value = 0.006916
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##   1.499202 156.126293
## sample estimates:
## odds ratio
##   11.52041
```

```
print("Valor P value Obtenido")
```

```
## [1] "Valor P value Obtenido"
```

```
print(Prueba_fisher$p.value)
```

```
## [1] 0.00691584
```

## Conclusiones

Con un valor P de 0.00691584 obtenido, rechazamos la hipótesis nula, en favor de la hipótesis alternativa, por lo que podemos concluir que existe una relación significativa entre las especies alienígenas y la preferencia de hábitats subterráneos o acuáticos.

4. Se llevó a cabo una investigación entre dos grupos demográficos: residentes urbanos y suburbanos. Seleccionaron aleatoriamente 35 residentes urbanos y 30 residentes suburbanos, y se les preguntó sobre su preferencia de transporte para desplazarse en la ciudad entre tres opciones comunes, obteniéndose que de los primeros 15 usan el Metro, 14 usan bicicletas y 8 su vehículo particular. mientras que entre los segundos 12 prefieren el Metro, 7 las bicicletas y 11 el vehículo particular. ¿Son iguales las distribuciones de las preferencias de transporte entre los dos grupos demográficos?

## Respuesta pregunta 4

```
# Definimos los parametros conocidos

Urbanos <- 35 #Cantidad de residentes urbanos
Suburbanos <- 30 #Cantidad de residentes suburbanos

Metro_U <- 15 #Cantidad de residentes urbanos que prefieren el Metro
Bicicletas_U <- 14 #Cantidad de residentes urbanos que prefieren las bicicletas
Vehiculo_U <- 8 #Cantidad de residentes urbanos que prefieren su vehículo particular

Metro_S <- 12 #Cantidad de residentes suburbanos que prefieren el Metro
Bicicletas_S <- 7 #Cantidad de residentes suburbanos que prefieren las bicicletas
Vehiculo_S <- 11 #Cantidad de residentes suburbanos que prefieren su vehículo particular

#Definicion de prueba a realizar: Como nos encontramos con dos variables Categoricals y una de ellas es
#Por otro lado verificamos que las observaciones son Independientes y que las frecuencias esperadas sean
#Para realizar esta prueba utilizaremos la prueba chi-cuadrado de homogeneidad, ya que se nos pide comp

# Definimos las hipótesis nula y alternativa

#H0: los residentes urbanos y suburbanos tienen la misma preferencia de transporte.
#HA: los residentes urbanos y suburbanos no tienen la misma preferencia de transporte.

# Creamos la tabla de contingencia

Urbanos <- c(Metro_U, Bicicletas_U, Vehiculo_U)
Suburbanos <- c(Metro_S, Bicicletas_S, Vehiculo_S)

tabla2 <- as.table(rbind(Urbanos, Suburbanos))

dimnames(tabla2) <- list(c("Urbanos", "Suburbanos"), c("Metro", "Bicicletas", "Vehiculo"))

print(tabla2)
```

Para responder esta pregunta, procederemos a verificar las condiciones de nuestra prueba junto a la tabla de contingencia, para definir correctamente el metodo a utilizar.

```
##           Metro Bicicletas Vehiculo
## Urbanos      15         14         8
## Suburbanos   12          7        11
```

```
#Realizamos la prueba de Chi-cuadrado de homogeneidad con un nivel de significancia de 0.05

Prueba_chi <- chisq.test(tabla2)
print("Resultados de la Prueba de Chi-cuadrado de homogeneidad")
```

```
## [1] "Resultados de la Prueba de Chi-cuadrado de homogeneidad"
```



```
print(Prueba_chi)
```

```
##  
## Pearson's Chi-squared test  
##  
## data:  tabla2  
## X-squared = 2.4356, df = 2, p-value = 0.2959
```

```
print("Valor P value Obtenido")
```

```
## [1] "Valor P value Obtenido"
```

```
print(Prueba_chi$p.value)
```

```
## [1] 0.2958814
```

### Conclusiones

Con un valor de P de 0.2958814 obtenido, no hay evidencia suficiente para rechazar la hipótesis nula, por lo que podemos concluir que las distribuciones de las preferencias de transporte entre los dos grupos demográficos son iguales.