

# PEC2 - MANUEL ROJAS GARCÍA

## Pregunta 1:

Los datos del archivo LasVegas.csv, obtenidas del repositorio UCI (<https://archivebeta.ics.uci.edu/ml/datasets>) contienen información sobre las reseñas redactadas en TripAdvisor por clientes de 21 hoteles de Las Vegas. Importe los datos al programa R y observe los nombres de las variables.

Importamos los datos y creamos la variable dataset:

```
dataset <- read.csv("C:/Users/Manuel/Desktop/UOC/SEMESTRE 3 (Sep 2023 - Feb 2024)/Estadística/PEC 2/LasVegas.csv")
head(dataset)
```

```
## User.country Nr..reviews Nr..hotel.reviews Helpful.votes Score Period.of.stay
## 1 USA 11 4 13 5 Dec-Feb
## 2 USA 119 21 75 3 Dec-Feb
## 3 USA 36 9 25 5 Mar-May
## 4 UK 14 7 14 4 Mar-May
## 5 Canada 5 5 2 4 Mar-May
## 6 Canada 31 8 27 3 Mar-May
## Traveler.type Pool Gym Tennis.court Spa Casino Free.internet
## 1 Friends NO YES NO NO YES YES
## 2 Business NO YES NO NO YES YES
## 3 Families NO YES NO NO YES YES
## 4 Friends NO YES NO NO YES YES
## 5 Solo NO YES NO NO YES YES
## 6 Couples NO YES NO NO YES YES
## Hotel.name Hotel.stars Nr..rooms User.continent
## 1 Circus Circus Hotel & Casino Las Vegas 3 3773 North America
## 2 Circus Circus Hotel & Casino Las Vegas 3 3773 North America
## 3 Circus Circus Hotel & Casino Las Vegas 3 3773 North America
## 4 Circus Circus Hotel & Casino Las Vegas 3 3773 Europe
## 5 Circus Circus Hotel & Casino Las Vegas 3 3773 North America
## 6 Circus Circus Hotel & Casino Las Vegas 3 3773 North America
## Member.years Review.month Review.weekday
## 1 9 January Thursday
## 2 3 January Friday
## 3 2 February Saturday
## 4 6 February Friday
## 5 7 March Tuesday
## 6 2 March Tuesday
```

- a) La variable Traveler.type indica el tipo de viajero clasificado en Business, Couples, Families, Friends, Solo (según si se han hospedado en el hotel por negocios, en pareja, en familia, con amigos o solos). La variable Hotel.stars indica el número de estrellas del hotel que pueden ser 3, 3.5, 4, 4.5 o 5. Elabore una tabla de contingencia entre las variables Traveler.type y Hotel.stars:

Creamos la variable tabla que contiene una tabla de contingencia entre los valores indicados. Para tener una visualización mejor usamos addmargins e incluimos los sumatorios.

<https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/addmargins>

```
Tabla <- table(dataset$Traveler.type, dataset$Hotel.stars)
Tablasconsum <- addmargins(Tabla)
head(Tablasconsum)
```

```
##
##           3 3,5   4 4,5   5 Sum
## Business  15  11  14   4  30  74
## Couples   36  25  58  10  85 214
## Families  22  25  17   4  42 110
## Friends   19   9  21   5  28  82
## Solo       4   2  10   1   7  24
## Sum       96  72 120  24 192 504
```

- b) Si escogemos un individuo de la base de datos al azar ¿cuál es la probabilidad que corresponda a un cliente que se ha hospedado con amigos y en un hotel de 5 estrellas?

Filtramos la tabla anterior en amigos y 5 estrellas. Contamos el total individuos y dividimos para obtener la probabilidad.

$$28/504 = 0.05555556$$

Realizamos la operación en R.

```
Hotel5 <- subset(dataset, Hotel.stars >= 5 & Traveler.type == "Friends")
Amigo5 <- nrow(Hotel5)
filas <- nrow(dataset)
resultado <- Amigo5/filas
head(resultado)
```

```
## [1] 0.05555556
```

- c) ¿Cuál es la probabilidad que un cliente que viaja por negocios se haya hospedado en un hotel de 3 estrellas?

Realizamos la misma operación pero solo seleccionando los sumatorios de “business. Realizamos la división para obtener el resultado.

$$15/74 = 0.2027$$

Realizamos la operación en R.

```
Hotel3 <- sum(dataset$Hotel.stars == 3 & dataset$Traveler.type == "Business")
Negocios <- sum(dataset$Traveler.type == "Business")
Resultado = Hotel3/Negocios
head(Resultado)
```

```
## [1] 0.2027027
```

- d) ¿Cuál es la probabilidad que un cliente que se ha hospedado en un hotel de 4 estrellas viaje en pareja?

Realizamos la misma operación anterior pero el operador de 4 estrellas. Realizamos la división para obtener el resultado.

$$58/214 = 0.2710$$

Realizamos la operación en R

```
Hotel4 <- sum(dataset$Hotel.stars == 4 & dataset$Traveler.type == "Couples")
Parejas <- sum(dataset$Traveler.type == "Couples")
Resultado = Hotel4/Parejas
head(Resultado)
```

```
## [1] 0.271028
```

## Pregunta 2:

- a) Si escogemos un cliente de la base de datos al azar. ¿Cuál es la probabilidad que se hospede en un hotel de 5 estrellas?

Hacemos un sumatorio de todos los clientes de 5 estrellas y lo dividimos por el total para poder obtener la probabilidad de hospedarse en dichas condiciones.

$$192/504 = 0.3809524$$

```
Cliente5 <- subset(dataset, Hotel.stars == 5)
Totales5 <- nrow(Cliente5)
Totaldataset <- nrow(dataset)
resultado <- Totales5/Totaldataset
resultado
```

```
## [1] 0.3809524
```

- b) Escogemos al azar 10 clientes de la base de datos, con reposición. Consideramos la variable que nos indica el número de clientes, entre los 10, que se hospedan en un hotel de 5 estrellas.

- i.) ¿Qué distribución sigue esta variable? ¿De qué parámetros?

La distribución que sigue esta variable es Binomial, que tiene una distribución  $B(n,p)$  donde  $n$  será el número de clientes 10 y  $p$  la probabilidad de éxito de la probabilidad de hotel de 5 estrellas, es decir,  $192/504 = 0,3809$ .

$B(10, 0,3809)$

- ii.) ¿Cuál es la probabilidad que exactamente 3 de los 10 se hospeden en un hotel de 5 estrellas?

<https://r-coder.com/distribucion-binomial-r/>

Realizamos los cálculos con la fórmula.

$$P(Y = 3) = (10!/3!7!) * 0,3809^3 * (1-0.3809)^7$$

Utilizamos la bibliografía superior para realizar los cálculos en R

```
dbinom(x = 3, size = 10, prob = 0.3809)
```

```
## [1] 0.2311744
```

- c) Supongamos ahora que sabemos que entre los que se hospedan en un hotel de 5 estrellas el 25% usa el parking del hotel. Sabemos también que un 30% de los clientes de la base de datos que han usado el parking de su hotel se han hospedado en un hotel de 5 estrellas. ¿Cuál es la probabilidad que un cliente use el parking del hotel?

<https://rpubs.com/maarnuro/594459>

En este caso tenemos que resolver con el teorema de bayes. Se realiza un árbol de probabilidades, donde el 25% de los clientes de 5 estrellas usan parking y un 30% del total también.

$$0.25 \cdot 0.30 = 0.075$$

```
0.25*0.30
```

```
## [1] 0.075
```

En segundo lugar, el árbol de decisión tiene un 75% de no usar parking y un 30% de si usar el parking aún no siendo de 5 estrellas.

$$0.75 \cdot 0.30 = 0.225$$

```
0.75*0.3
```

```
## [1] 0.225
```

Aplicamos la fórmula

$$0.25 / (0.075 + 0.225) = 0.833$$

```
0.25/(0.075+0.225)
```

```
## [1] 0.8333333
```

## Problema 3

Supongamos ahora que sabemos que la edad de los clientes de uno de estos hoteles se distribuye siguiendo una distribución normal de media 52 y desviación típica 11. Escogemos un cliente al azar: a) ¿Cuál es la probabilidad que tenga más de 60 años?

<https://r-coder.com/distribucion-normal-r/>

```
pnorm(60, mean = 52, sd = 11, log = FALSE, lower.tail = FALSE)
```

```
## [1] 0.2335295
```

- b) ¿Cuál es la probabilidad que tenga menos de 40 años?

```
pnorm(40, mean = 52, sd = 11, log = FALSE, lower.tail = TRUE)
```

```
## [1] 0.1376564
```

- c) Encuentre una edad de forma que el 75% de los clientes de este hotel sean menores que ese valor y el 25% de los clientes sean mayores.

Seguimos usando la web superior para encontrar el cuartil con qnorm

```
qnorm(0.75, mean = 52, sd = 11, lower.tail = TRUE, log.p = FALSE)
```

```
## [1] 59.41939
```