

Tarea: Repaso de Gráficos en Estadística

Jesus Mudarra Luján

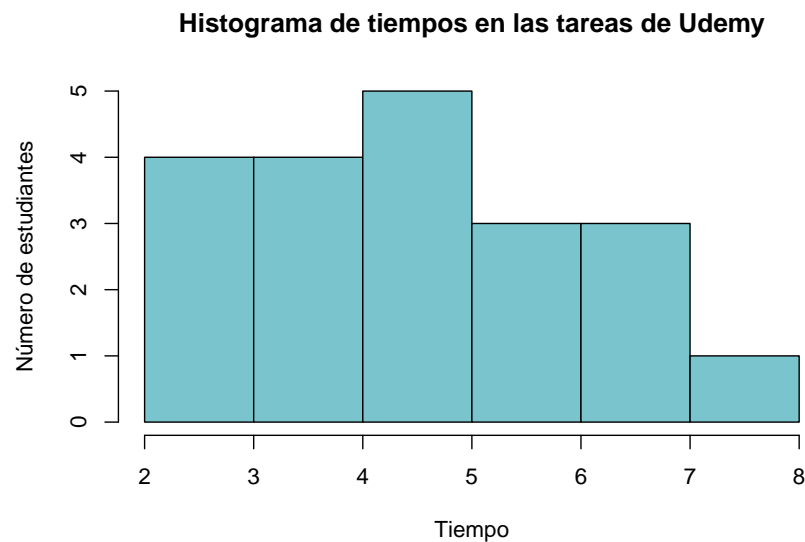
2023-10-04

Pregunta 1

Describe gráficamente el tiempo (en minutos) que estos 20 estudiantes tardaron en finalizar su tarea en Udemy utilizando el gráfico adecuado:

```
datos <- c(6.5, 5.8, 4.5, 6.2, 4.8, 7.3, 4.6, 3.9, 4.4, 5.5,  
          5.2, 6.7, 3.0, 2.4, 5.0, 3.6, 2.9, 4.0, 2.8, 3.6)
```

```
hist(datos, main = "Histograma de tiempos en las tareas de Udemy", col = "cadetblue3",  
      xlab = "Tiempo", ylab = "Número de estudiantes")
```



La mayoría de los alumnos tardaron en finalizar la tarea entre 4 y 5 minutos.

Pregunta 2

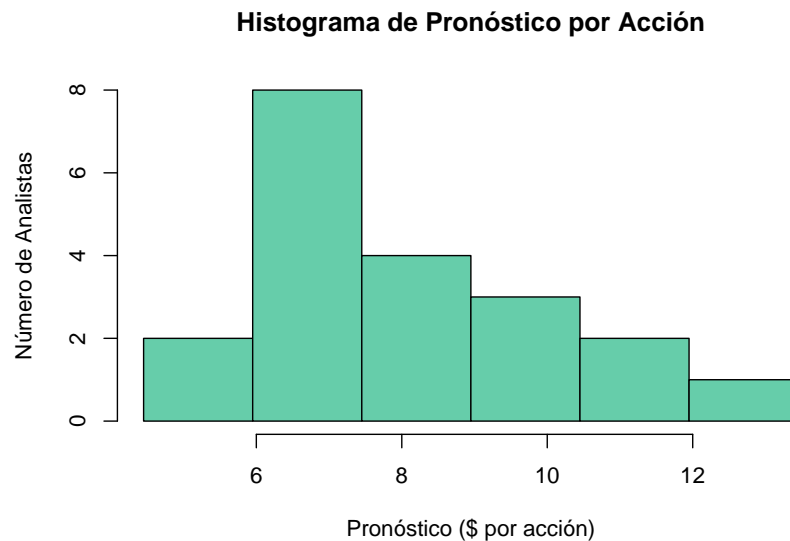
Se ha solicitado a una muestra de 20 analistas financieros que nos den unos pronósticos de ganancias del valor de la acción de nuestra empresa para el próximo año. Los resultados son los que siguen en la siguiente tabla:

Pronóstico (\$ por acción)	Número de Analistas
4.45 < 5.95	2
5.95 < 7.45	8
7.45 < 8.95	4
8.95 < 10.45	3
10.45 < 11.95	2
11.95 < 13.45	1

a) Representa gráficamente el histograma.

```
datos <- c(rep(5,2), rep(6,8), rep(8,4), rep(9,3), rep(11,2), 12)

hist(datos, breaks = c(4.45, 5.95, 7.45, 8.95, 10.45, 11.95, 13.45),
      main = "Histograma de Pronóstico por Acción", col = "aquamarine3",
      xlab = "Pronóstico ($ por acción)", ylab = "Número de Analistas")
```



b) Determina las frecuencias relativas.

	Frecuencia	Frecuencia relativa (%)
[4.45, 5.95)	2	10
[5.95, 7.45)	8	40
[7.45, 8.95)	4	20
[8.95, 10.45)	3	15
[10.45, 11.95)	2	10
[11.95, 13.45]	1	5

c) Determina las frecuencias acumuladas.

	Frecuencia	Frecuencia relativa (%)	Frec. acum.
[4.45, 5.95)	2	10	2
[5.95, 7.45)	8	40	10
[7.45, 8.95)	4	20	14
[8.95, 10.45)	3	15	17
[10.45, 11.95)	2	10	19
[11.95, 13.45]	1	5	20

d) Determina e interpreta las frecuencias acumulativas relativas.

	Frecuencia	Frecuencia relativa (%)	Frec. acum.	Frec. acum. relativa (%)
[4.45, 5.95)	2	10	2	10
[5.95, 7.45)	8	40	10	50
[7.45, 8.95)	4	20	14	70
[8.95, 10.45)	3	15	17	85
[10.45, 11.95)	2	10	19	95
[11.95, 13.45]	1	5	20	100

Analizando la columna de las frecuencias acumuladas relativas llegamos a la conclusión de que el 50% de los analistas establece un pronóstico inferior a 7.45\$ por acción.

Pregunta 3

¿Cómo se suele enterar la gente sobre un producto nuevo? A una muestra aleatoria de 200 compradores en una tienda en particular se les preguntó su edad y si se enteraron del producto por medio de un amigo o por medio de un anuncio en un periódico local. Los resultados indicaron que 50 encuestados eran menores de 21 años, 90 personas estaban en el grupo de edad entre 21 y 35 y 60 encuestados tenían más de 35 años. De los menores de 21 años, 30 se enteraron del producto a través de un amigo y el resto se enteró del producto a través de un anuncio en el periódico local. Un tercio de las personas en la categoría de edad de 21 a 35 se enteró por primera vez sobre el producto a través del anuncio del periódico local; el resto de este grupo de edad se enteró del producto por medio de un amigo. Un amigo informó al 30% de las personas en la categoría de mayores de 35 años sobre el producto; el resto se enteró por el anuncio del periódico local.

a) Describe los datos con una tabla cruzada.

	Amigo	Periódico local
<21	30	20
[21,35)	60	30
>35	18	42

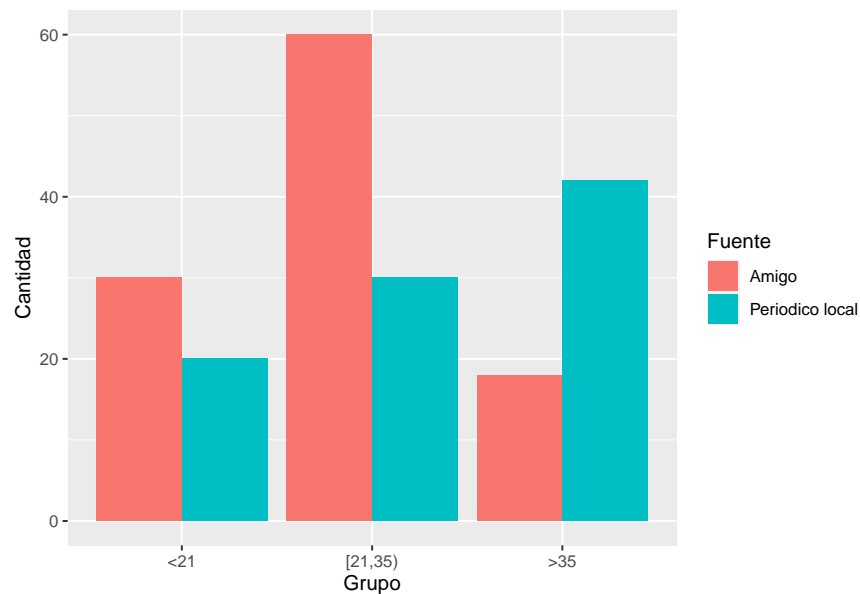
b) Describe los datos gráficamente.

Una forma muy útil de visualizar datos de una tabla cruzada es mediante el uso de gráfico de barras agrupadas:

```
library(tidyverse)

df <- data.frame(
  Grupo = c(rep("<21",2), rep("[21,35]",2), rep(">35",2)),
  Fuente = c(rep(c("Amigo", "Periodico local"),3)),
  Cantidad = c(30, 20, 60, 30, 18, 42)
)

df %>%
  ggplot(aes(x=Grupo, y=Cantidad, fill=Fuente)) +
  geom_bar(position="dodge", stat="identity") +
  scale_x_discrete(limits = c("<21", "[21,35]", ">35"))
```



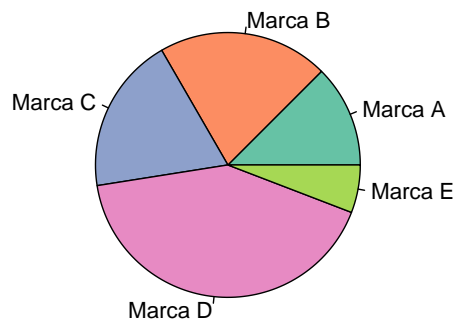
Pregunta 4

Se encuesta a una muestra aleatoria de clientes para que seleccionara su refresco favorito de una lista de cinco marcas. Los resultados mostraron que 30 preferían la Marca A, 50 preferían Marca B, 46 marca preferida C, 100 marca preferida D, y 14 marcas preferidas E.

```
refresco.df <- data.frame("Clientes" = c(30,50,46,100,14),
  "Marca" = c("Marca A","Marca B","Marca C","Marca D","Marca E"))
```

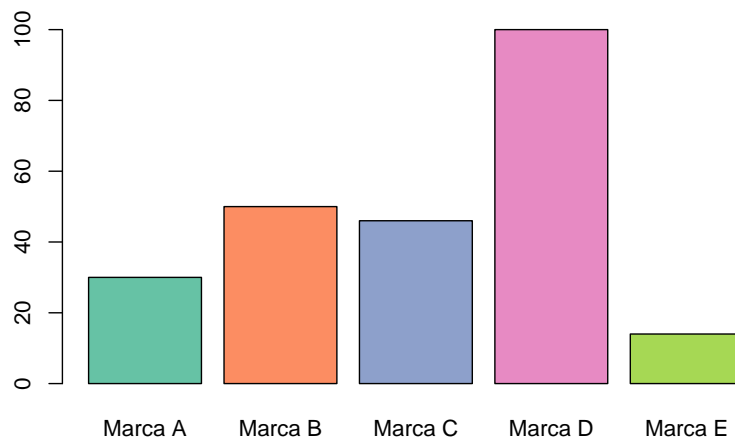
a) Construye un diagrama circular.

```
library(RColorBrewer)
coul <- brewer.pal(5, "Set2")
pie(refresco.df$Clientes, labels = refresco.df$Marca, col = coul)
```



b) Construye un diagrama de barras.

```
barplot(height = refresco.df$Clientes, names = refresco.df$Marca, col=coul)
```



Pregunta 5

Accede a la web del Censo de Estados Unidos (<https://www.census.gov>) y busca un conjunto de datos que consideres interesantes. Describe qué variables tiene y de qué tipo se trata cada una de ellas, así como qué tipo de gráfico utilizarías para realizar un resumen estadístico de la distribución de los mismos.

Dataset de impuestos en Estados Unidos obtenido de: <https://www.census.gov/data/datasets/2021/econ/stc/2021-annual.html>

Variables:

- T01: Property Taxes (numérica)
- T09: General Sales and Gross Receipts Taxes (numérica)
- T10: Alcoholic Beverages Sales Tax (numérica)
- T11: Amusements Sales Tax (numérica)
- T12: Insurance Premiums Sales Tax (numérica)
- T13: Motor Fuels Sales Tax (numérica)
- T14: Pari-mutuels Sales Tax (numérica)
- T15: Public Utilities Sales Tax (numérica)
- T16: Tobacco Products Sales Tax (numérica)
- T19: Other Selective Sales and Gross Receipts Taxes (numérica)
- T20: Alcoholic Beverages License (numérica)
- T21: Amusements License (numérica)
- T22: Corporations in General License (numérica)
- T23: Hunting and Fishing License (numérica)
- T24: Motor Vehicles License (numérica)
- T25: Motor Vehicle Operators License (numérica)
- T27: Public Utilities License (numérica)
- T28: Occupation and Businesses License, NEC (numérica)
- T29: Other License Taxes (numérica)
- T40: Individual Income Taxes (numérica)
- T41: Corporation Net Income Taxes (numérica)
- T50: Death and Gift Taxes (numérica)
- T51: Documentary and Stock Transfer Taxes (numérica)
- T53: Severance Taxes (numérica)
- T99: Taxes, NEC (numérica)

Sería interesante analizar mediante un diagrama de barras el valor de un impuesto para cada uno de los estados 52 de US.

Pregunta 6

Se ha entrevistado a una muestra aleatoria de 50 empleados que trabajaban para un gran empresa internacional para establecer si existía un vínculo entre los trabajos de los empleados en casa durante el teletrabajo originado por la pandemia del COVID19 y el lugar donde normalmente comían a la hora del almuerzo, ya sea en el comedor o en sus escritorios de trabajo habitual. Los resultados de esta investigación mostraron que el 75% del personal que trabaja desde casa normalmente comía en el comedor de casa y el 54% de los empleados que seguían yendo a la oficina normalmente almorzaba en sus escritorios donde trabajaban. Crea una tabla cruzada y un gráfico de barras para mostrar esta información de forma clara.

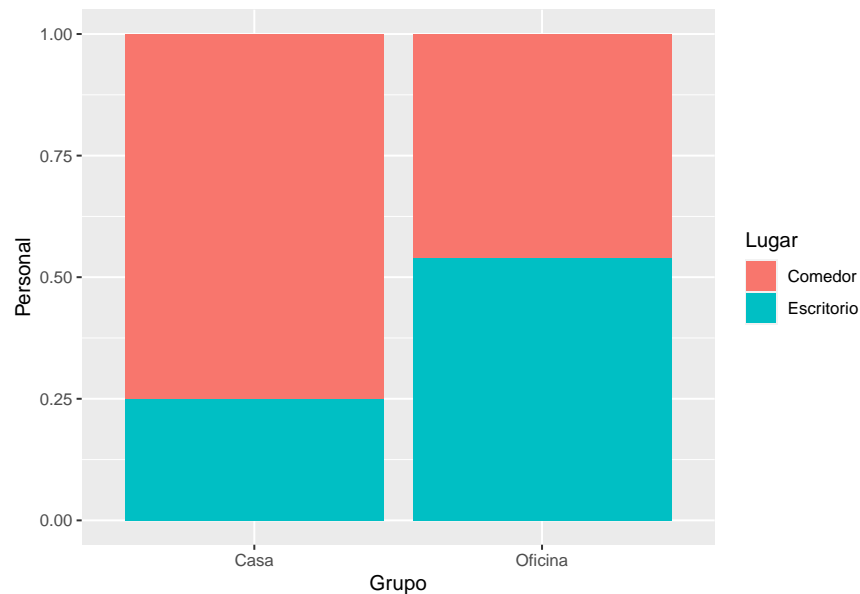
Tabla cruzada:

Lugar de trabajo	Comedor	Escritorio
Casa	0.75	0.25
Oficina	0.46	0.54

Una forma muy útil de visualizar datos de una tabla cruzada es mediante el uso de gráfico de barras apiladas:

```
df <- data.frame(
  Grupo = c(rep("Casa",2), rep("Oficina",2)),
  Lugar = c(rep(c("Comedor", "Escritorio"),2)),
  Personal = c(75, 25, 46, 54)
)
```

```
df %>%
  ggplot(aes(x=Grupo, y=Personal, fill=Lugar)) +
  geom_bar(position="fill", stat="identity") +
  scale_x_discrete(limits = c("Casa", "Oficina"))
```



Pregunta 7

Un banco ofrece cuatro tipos de cuentas corrientes. Recientemente hemos encuestado a una muestra aleatoria de 300 clientes bancarios y se le han realizado varias preguntas. Se encontró que el 60% de los encuestados prefería la cuenta fácil; el 12%, la cuenta inteligente; el 18%, la super cuenta corriente; y el resto, la cuenta business. De los que seleccionaron la cuenta fácil, 100 eran mujeres; un tercio de los encuestados que seleccionaron la cuenta inteligente eran hombres; la mitad de los encuestados que seleccionaron la super cuenta corriente eran hombres; y el 80% de los encuestados que seleccionaron la cuenta business eran hombres.

a) Describe los datos con una tabla cruzada.

	Hombre	Mujer
Cuenta fácil	80	100
Cuenta inteligente	12	24
Super cuenta corriente	27	27
Cuenta business	24	6

b) Describe los datos gráficamente.

Una forma muy útil de visualizar datos de una tabla cruzada es mediante el uso de gráfico de barras agrupadas:

```
df <- data.frame(
  Cuenta = c(rep("Cuenta fácil",2), rep("Cuenta inteligente",2),
             rep("Super cuenta corriente",2), rep("Cuenta business",2)),
  Sexo = c(rep(c("Hombre", "Mujer"),4)),
  Personas = c(80, 100, 12, 24, 27, 27, 24, 6)
)

df %>%
  ggplot(aes(x=Cuenta, y=Personas, fill=Sexo)) +
  geom_bar(position="dodge", stat="identity") +
  scale_x_discrete(limits = c("Cuenta fácil", "Cuenta inteligente",
                              "Super cuenta corriente", "Cuenta business"))
```

