

Estadística 1

Andrea Gomez Vargas

Table of contents

1	Primera parte: Evaluación conceptual y teórica	2
1.1	Consigne las definiciones de Población, muestra y unidad de análisis	2
1.2	Consigne, defina y ejemplifique los niveles de medición de Stevens	2
1.3	Consigne las principales propiedades de la distribución normal	2
1.4	Concepto y elementos constitutivos de las pruebas de hipótesis	4
2	Segunda parte: Evaluación práctica	4
3	Tercera parte: Análisis estadístico descriptivo	4
3.1	Importar los datos a R.	4
3.2	Utilizando la función table() y sus funciones anexas, calcular	4
3.2.1	Las tablas de frecuencias absolutas para las variables ocio, genero, colegio, amor, religión	4
3.2.2	Las tablas de frecuencias relativas para las variables ocio, genero, colegio, amor, religión	6
3.3	Realizar gráficos de barras para cada una de las variables anteriores	7
3.3.1	Gráfico 1: frecuencia absoluta y porcentual por variable género	8
3.3.2	Gráfico 2: frecuencia absoluta y porcentual por variable ocio	9
3.3.3	Gráfico 3: frecuencia absoluta y porcentual por variable amor	10
3.3.4	Gráfico 4: frecuencia absoluta y porcentual por variable colegio	11
3.3.5	Gráfico 5: frecuencia absoluta y porcentual por variable religión	12
3.4	Calcular las tablas de contingencia para los siguientes cruces de variables	13
3.4.1	Ocio y genero	13
3.4.2	Ocio y colegio	13
3.4.3	Amor y religión	14
3.4.4	Colegio y religión	14
4	Cuarta parte: Contrastes de hipótesis	14

1 Primera parte: Evaluación conceptual y teórica

1.1 Consigne las definiciones de Población, muestra y unidad de análisis¹

- **Población:** es un conjunto de *todos* los elementos que estamos estudiando, acerca de los cuales intentamos sacar conclusiones. Debemos definir esa población de modo que quede claro cuándo cierto elemento pertenece o no a la población.
- **Muestra:** Colección de algunos elementos —no todos—, de la población bajo estudio, utilizada para describir poblaciones.
 - Muestra representativa: Muestra que contiene las características importantes de la población en las mismas proporciones en que están contenidas en la población.
- **Unidad de Análisis:** o la unidad de estudio u objeto/sujeto de estudio (los elementos que van a ser estudiados), mientras que la unidad de recolección es la fuente proveedora de la información requerida, son los elementos que funcionan como informantes

1.2 Consigne, defina y ejemplifique los niveles de medición de Stevens

Nivel	Definición
Nominal	Se caracteriza por datos que consisten exclusivamente en nombres, etiquetas o categorías. Los
Ordinal	Las categorías están ordenadas, pero no hay diferencias o estas carecen de significado.
De intervalo	Las diferencias tienen un significado, pero no hay punto de partida cero natural, y las razones
De razón	Hay un punto de partida cero natural y las razones tienen significado.

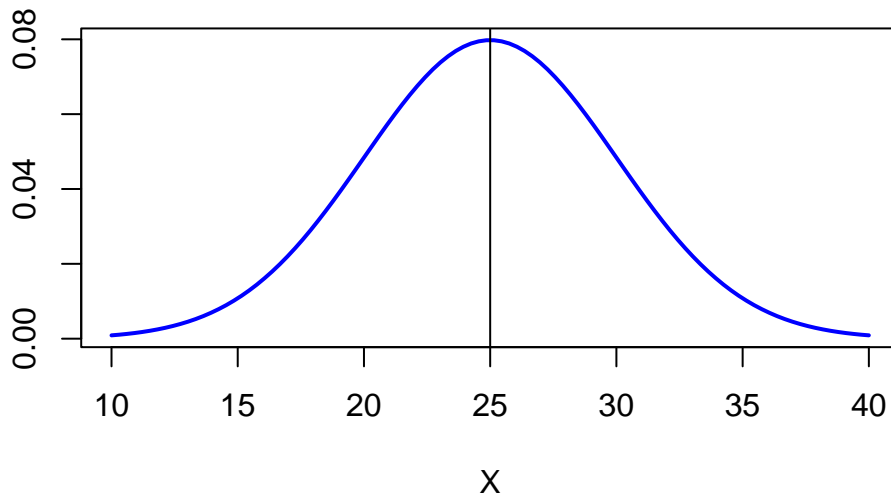
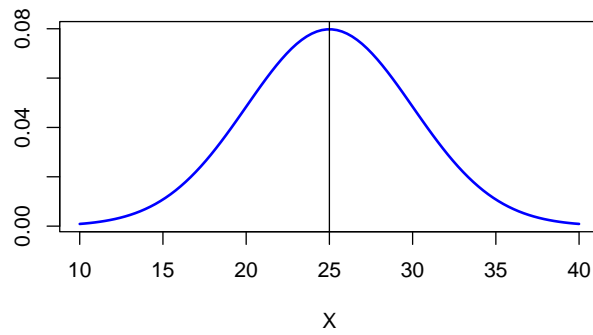
1.3 Consigne las principales propiedades de la distribución normal

La distribución normal tiene las siguientes características clave:

- Simetría: Es simétrica alrededor de la media (μ), lo que significa que las colas izquierda y derecha de la distribución son idénticas.
- Unimodal: Tiene un solo pico en la media (μ).
- Forma de Campana: La función de densidad de probabilidad forma una curva en forma de campana.
- Regla Empírica: Aproximadamente el 68% de los datos caen dentro de una desviación estándar de la media, el 95% dentro de dos desviaciones estándar y el 99.7% dentro de tres desviaciones estándar.

¹https://drive.google.com/file/d/1WLEdjR0WAI9UFv178iBQv4nhLVa8V_DX/view

$$X \sim N(\mu, \sigma)$$



Cuando se grafica, una distribución normal tiene la forma de “campana”. Las características de la curva en forma de campana son:

1. el aumento de las frecuencias, las cuales alcanzan un punto máximo y luego disminuyen;
y
2. la simetría, donde la mitad izquierda de la gráfica es casi una imagen especular de la mitad derecha.

1.4 Concepto y elementos constitutivos de las pruebas de hipótesis

2 Segunda parte: Evaluación práctica

Se ha realizado una encuesta a jóvenes españoles en el año 2000. Se muestran a continuación las 16 preguntas que contenía el cuestionario.

1. Para cada una de las preguntas definir:
 - a. Las variables que contienen información a analizar estadísticamente.
 - b. Identificar, para cada una de ellas el nivel de medición y las modalidades/categorías que la integran.

3 Tercera parte: Análisis estadístico descriptivo

En esta actividad se trabajará con el archivo de datos “cuestionario.xlsx”.

3.1 Importar los datos a R.

```
cuestionario <- openxlsx::read.xlsx("EVALUACION/cuestionario.xlsx")
```

3.2 Utilizando la función table() y sus funciones anexas, calcular

3.2.1 Las tablas de frecuencias absolutas para las variables ocio, genero, colegio, amor, religión

3.2.1.1 Ocio

```
addmargins(table(cuestionario$ocio, useNA = "always")) %>% kbl()
```

Var1	Freq
	28
1	195
2	166
3	109
4	169
5	32
6	18
NA	0
Sum	717

3.2.1.2 Género

```
addmargins(table(cuestionario$genero,useNA = "always")) %>% kbl()
```

Var1	Freq
1	346
2	371
NA	0
Sum	717

3.2.1.3 Colegio

```
addmargins(table(cuestionario$colegio,useNA = "always")) %>% kbl()
```

Var1	Freq
	3
1	475
2	59
3	180
NA	0
Sum	717

3.2.1.4 Amor

```
addmargins(table(cuestionario$amor,useNA = "always")) %>% kbl()
```

Var1	Freq
	68
1	294
2	138
3	149
4	68
NA	0
Sum	717

3.2.1.5 Religión

```
addmargins(table(cuestionario$religion,useNA = "always")) %>% kbl()
```

Var1	Freq
1	177
2	317
3	12
4	108
5	99
99999	4
NA	0
Sum	717

3.2.2 Las tablas de frecuencias relativas para las variables ocio, genero, colegio, amor, religión

3.2.2.1 Ocio

```
kbl(prop.table(table(cuestionario$ocio))*100)
```

Var1	Freq
	3.90516
1	27.19665
2	23.15202
3	15.20223
4	23.57043
5	4.46304
6	2.51046

3.2.2.2 género

```
kbl(prop.table(table(cuestionario$genero))*100)
```

Var1	Freq
1	48.25662
2	51.74338

3.2.2.3 colegio

```
kbl(prop.table(table(cuestionario$colegio))*100)
```

Var1	Freq
	0.418410
1	66.248257
2	8.228731
3	25.104602

3.2.2.4 amor

```
kbl(prop.table(table(cuestionario$amor))*100)
```

Var1	Freq
	9.483961
1	41.004184
2	19.246862
3	20.781032
4	9.483961

3.2.2.5 religión

```
kbl(prop.table(table(cuestionario$religion))*100)
```

Var1	Freq
1	24.6861925
2	44.2119944
3	1.6736402
4	15.0627615
5	13.8075314
99999	0.5578801

3.3 Realizar gráficos de barras para cada una de las variables anteriores

- Con la frecuencia relativa
- Con la frecuencia porcentual

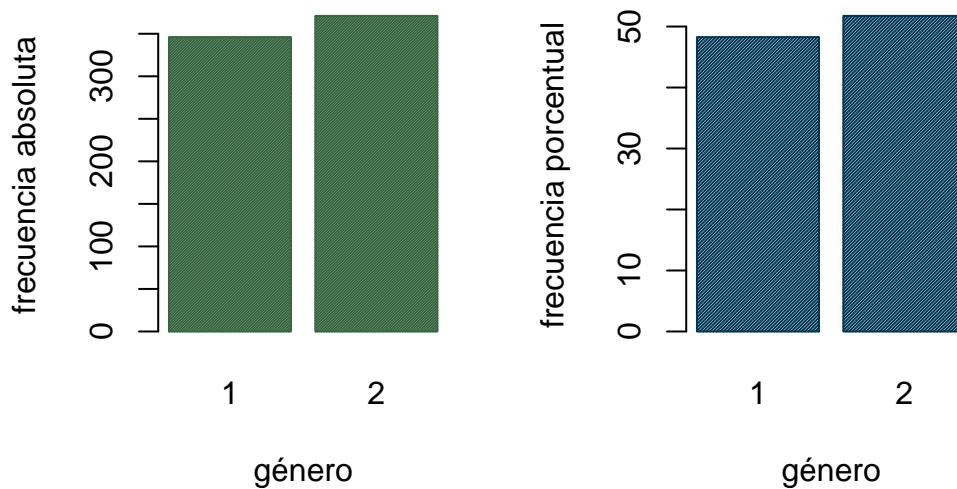
3.3.1 Gráfico 1: frecuencia absoluta y porcentual por variable género

```
par(mfrow = c(1, 2))

barplot(table(cuestionario$genero),
        main="frecuencia absoluta por género",
        xlab="género",
        ylab="frecuencia absoluta",
        border="#386641",
        col="#386641",
        density=100)

barplot(prop.table(table(cuestionario$genero))*100,
        main="frecuencia porcentual por género",
        xlab="género",
        ylab="frecuencia porcentual",
        border="#003554",
        col="#003554",
        density=80)
```

frecuencia absoluta por género | frecuencia porcentual por género

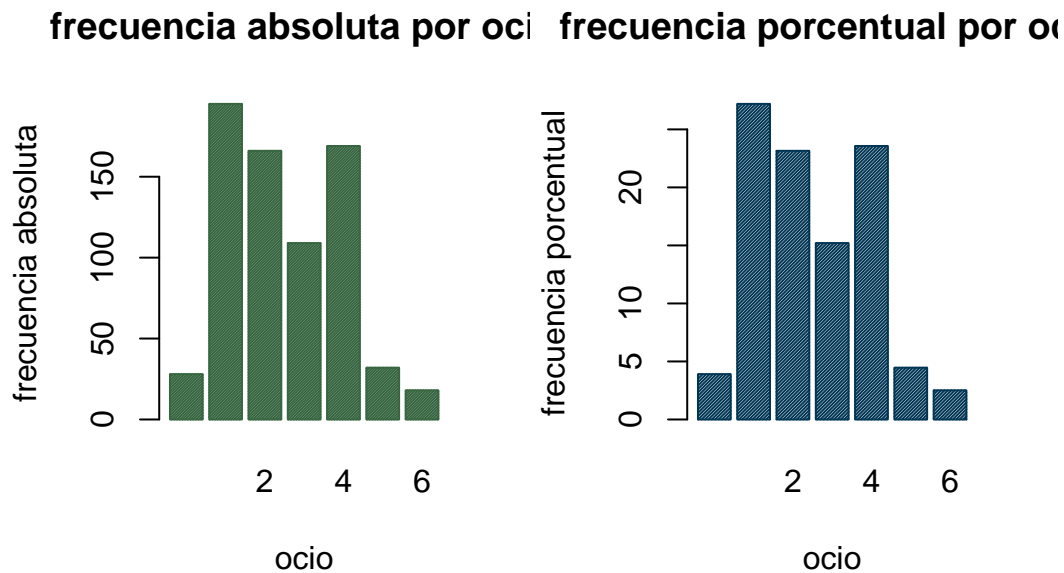


3.3.2 Gráfico 2: frecuencia absoluta y porcentual por variable ocio

```
par(mfrow = c(1, 2))

barplot(table(cuestionario$ocio),
        main="frecuencia absoluta por ocio",
        xlab="ocio",
        ylab="frecuencia absoluta",
        border="#386641",
        col="#386641",
        density=100)

barplot(prop.table(table(cuestionario$ocio))*100,
        main="frecuencia porcentual por ocio",
        xlab="ocio",
        ylab="frecuencia porcentual",
        border="#003554",
        col="#003554",
        density=80)
```

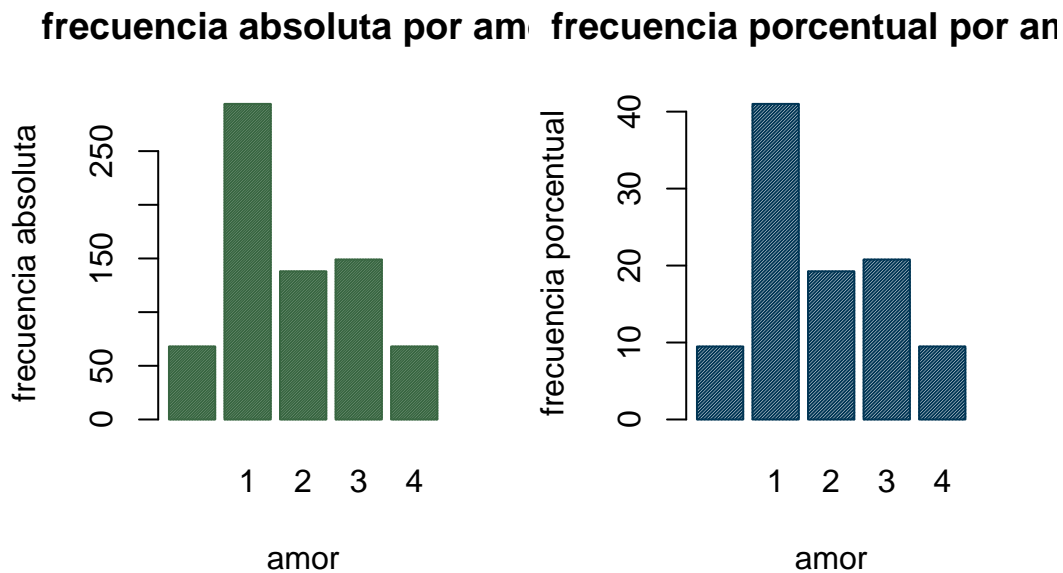


3.3.3 Gráfico 3: frecuencia absoluta y porcentual por variable amor

```
par(mfrow = c(1, 2))

barplot(table(cuestionario$amor),
        main="frecuencia absoluta por amor",
        xlab="amor",
        ylab="frecuencia absoluta",
        border="#386641",
        col="#386641",
        density=100)

barplot(prop.table(table(cuestionario$amor))*100,
        main="frecuencia porcentual por amor",
        xlab="amor",
        ylab="frecuencia porcentual",
        border="#003554",
        col="#003554",
        density=80)
```



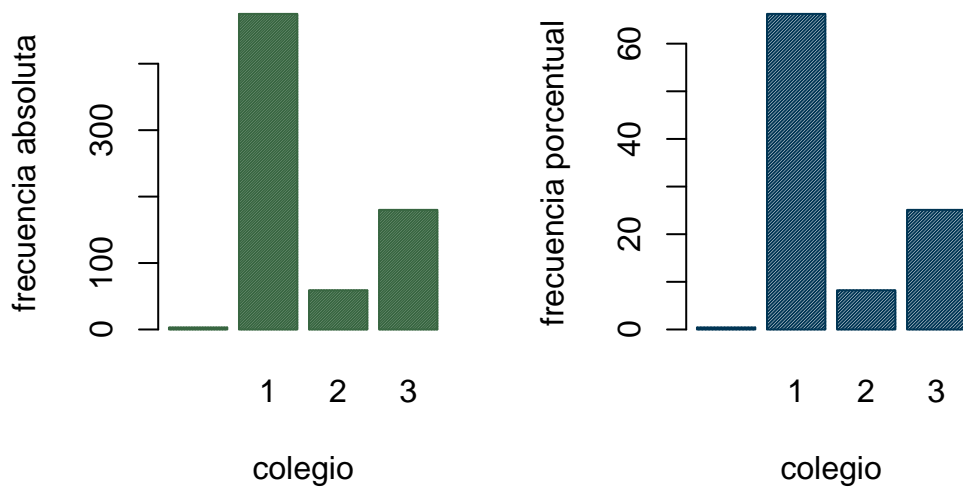
3.3.4 Gráfico 4: frecuencia absoluta y porcentual por variable colegio

```
par(mfrow = c(1, 2))

barplot(table(cuestionario$colegio),
        main="frecuencia absoluta por colegio",
        xlab="colegio",
        ylab="frecuencia absoluta",
        border="#386641",
        col="#386641",
        density=100)

barplot(prop.table(table(cuestionario$colegio))*100,
        main="frecuencia porcentual por colegio",
        xlab="colegio",
        ylab="frecuencia porcentual",
        border="#003554",
        col="#003554",
        density=80)
```

frecuencia absoluta por cole frecuencia porcentual por cole



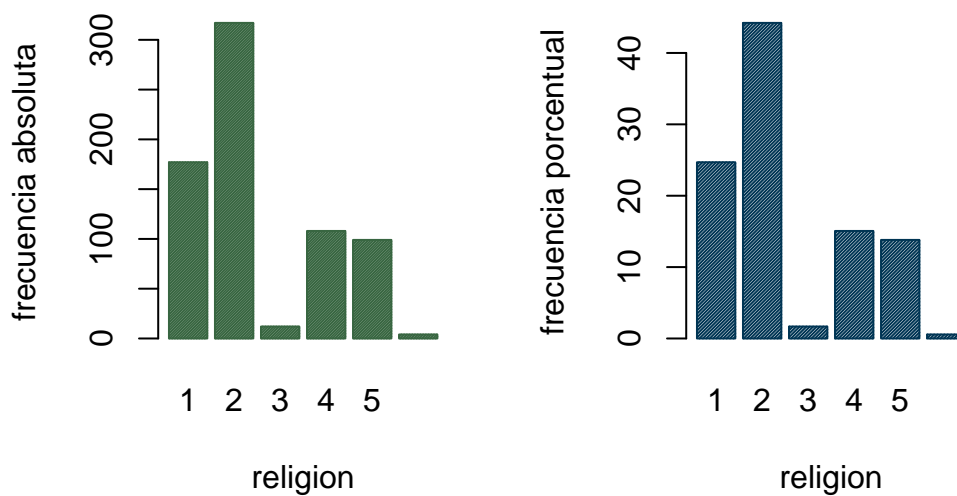
3.3.5 Gráfico 5: frecuencia absoluta y porcentual por variable religión

```
par(mfrow = c(1, 2))

barplot(table(cuestionario$religion),
        main="frecuencia absoluta por religion",
        xlab="religion",
        ylab="frecuencia absoluta",
        border="#386641",
        col="#386641",
        density=100)

barplot(prop.table(table(cuestionario$religion))*100,
        main="frecuencia porcentual por religion",
        xlab="religion",
        ylab="frecuencia porcentual",
        border="#003554",
        col="#003554",
        density=80)
```

frecuencia absoluta por religion frecuencia porcentual por religion



3.4 Calcular las tablas de contingencia para los siguientes cruces de variables

3.4.1 Ocio y genero

```
# absoluta
addmargins(table(cuestionario$ocio, cuestionario$genero,useNA = "always")) %>% kbl()
# relativa
prop.table(table(cuestionario$ocio, cuestionario$genero))*100
```

	1	2	NA	Sum
	13	15	0	28
1	84	111	0	195
2	130	36	0	166
3	38	71	0	109
4	52	117	0	169
5	23	9	0	32
6	6	12	0	18
NA	0	0	0	0
Sum	346	371	0	717

	1	2
	1.8131102	2.0920502
1	11.7154812	15.4811715
2	18.1311018	5.0209205
3	5.2998605	9.9023710
4	7.2524407	16.3179916
5	3.2078103	1.2552301
6	0.8368201	1.6736402

3.4.2 Ocio y colegio

```
# absoluta
addmargins(table(cuestionario$ocio, cuestionario$colegio,useNA = "always")) %>% kbl()
# relativa
kbl(prop.table(table(cuestionario$ocio, cuestionario$colegio))*100)
```

		1	2	3	NA	Sum
	0	18	2	8	0	28
1	1	135	11	48	0	195
2	1	112	19	34	0	166
3	0	70	12	27	0	109
4	1	106	13	49	0	169
5	0	23	1	8	0	32
6	0	11	1	6	0	18
NA	0	0	0	0	0	0
Sum	3	475	59	180	0	717

		1	2	3
	0.00000	2.510460	0.27894	1.1157601
1	0.13947	18.828452	1.53417	6.6945607
2	0.13947	15.620642	2.64993	4.7419805
3	0.00000	9.762901	1.67364	3.7656904
4	0.13947	14.783822	1.81311	6.8340307
5	0.00000	3.207810	0.13947	1.1157601
6	0.00000	1.534170	0.13947	0.8368201

3.4.3 Amor y religión

```
# absoluta
addmargins(table(cuestionario$amor, cuestionario$religion, useNA = "always")) %>% kbl()
# relativa
kbl(prop.table(table(cuestionario$amor, cuestionario$religion))*100)
```

	1	2	3	4	5	99999	NA	Sum						
	14	17	3	16	16	2	0	68	1	2	3	4	5	999
1	70	149	2	40	32	1	0	1294	2580	2.370990	0.41841	2.231520	2.23152	0.278
2	32	64	3	17	21	1	0	913	2901	20.781032	0.27894	5.578801	4.46304	0.139
3	40	63	3	22	21	0	0	414	3040	8.926081	0.41841	2.370990	2.92887	0.139
4	21	24	1	13	9	0	0	55	8801	8.786611	0.41841	3.068340	2.92887	0.000
NA	0	0	0	0	0	0	0	29	28870	3.347280	0.13947	1.813110	1.25523	0.000
Sum	177	317	12	108	99	4	0	717						

3.4.4 Colegio y religión

```
# absoluta
addmargins(table(cuestionario$colegio, cuestionario$religion, useNA = "always")) %>% kbl()
# relativa
kbl(prop.table(table(cuestionario$colegio, cuestionario$religion))*100)
```

	1	2	3	4	5	99999	NA	Sum						
	1	2	0	0	0	0	0	3	1	2	3	4	5	
1	114	225	7	64	62	3	0	475	39470	0.27894	0.0000000	0.000000	0.000000	0.
2	13	15	3	18	10	0	0	155	99582	31.38075	0.9762901	8.926081	8.647141	0.
3	49	75	2	26	27	1	0	180	13110	2.09205	0.4184100	2.510460	1.394700	0.
NA	0	0	0	0	0	0	0	6	834031	10.46025	0.2789400	3.626220	3.765690	0.
Sum	177	317	12	108	99	4	0	717						

4 Cuarta parte: Contrastes de hipótesis

Utilizando los datos del archivo cuestionario y técnicas paramétricas y no paramétricas,

- Determinar con un nivel de significación del 5% si el promedio de horas libres a la semana para ocio o diversión en los hombres es mayor a la de las mujeres. Concluir en términos del problema

- b) Determinar si hay diferencias en la cantidad promedio de libros leídos si tenemos en cuenta la asistencia de los jóvenes a colegio estatal o privado. Utilizar un nivel de significación del 5%. Concluir en términos del problema
- c) Determinar en cada caso, cuál es el test más adecuado. Justificar.