Capítulo 3

Métodos numéricos para representar datos cuantitativos

3.1 Generalidades

Los métodos numéricos para representar datos cuantitativos son procedimientos aritméticos o algebraicos que se realizan a partir de datos que se representan a través de variables cuantitativas. A continuación, se definen términos que son fundamentales para la comprensión de la estadística descriptiva e inferencial.

- **Variable:** la necesidad de representar la característica específica de una población, la cual puede tomar diversos valores dependiendo de los individuos que la conforman.
- **Estadística:** es una medida calculada a partir de los datos de una muestra. Por ejemplo: se requiere conocer el promedio de edad y estatura de una muestra seleccionada de una población, en donde la representación de los datos se hace a través de las variables x e y respectivamente. Por consiguiente, el promedio de cada variable es $\bar{x}$ e $\bar{y}$.
- **Parámetro:** es una medida calculada a partir de los datos de una población. Por ejemplo: se requiere conocer el promedio y la varianza de la edad de una población; los datos son representados con la variable x. Luego el promedio poblacional y la varianza de la variable de interés serán μ y σ^2 respectivamente.

3.2 Clasificación

La clasificación de los métodos numéricos para representar los datos cuantitativos es (Mendenhall & Sincich, 2002):

Medidas de tendencia central

- Medidas de posición.
- Medidas de variación.

3.2.1 Medidas de tendencia central

Miden el comportamiento de los datos con respecto a un valor específico y permiten encontrar el centro de distribución de los datos. Entre ellas se tienen la moda, la mediana y la media aritmética o valor promedio.

3.2.1.1 La moda

Es una medida de tendencia central que identifica el dato que más se repite en un conjunto de datos.

Ejemplo 3.1. Se tiene la variable $x = \{1, 2, 3, 3, 7, 4, 8, 9, 3\}$, luego la moda de la variable x es $\{3\}$.

Ejemplo 3.2. Se tiene la variable $y = \{1, 2, 3, 3, 7, 4, 8, 9, 2\}$; se deduce que tiene dos modas la variable y, queson $\{2,\ 3\}$, por lo que se denomina bimodal.

Aquí se puede afirmar que una variable puede tener más de una moda y se nombra por el número de modas que posea, por ejemplo: 3 modas, 4 modas, etc.

3.2.1.2 La mediana

Es una medida de tendencia central que identifica el valor que genera el punto de equilibrio en un conjunto de datos en donde distribuye el 50 % de los datos a cada lado del valor generado. Existen dos posibilidades para su cálculo:

- Cuando **n es impar**

La base ordenada: $x = \{1, 2, 4, 7, 8, 10, 12\}$; n = 7

La mediana = $\frac{(n+1)}{2} = \frac{8}{2} = 4$; se selecciona el número de la posición. Es decir, 7.

Siendo n el tamaño de la variable en consideración.

- Cuando **n es par**

La base ordenada: $x = \{2, 4, 7, 8, 10, 12,\ 14, 20\}$; n = 8

La mediana = $\frac{x_{(n/2)} + x_{(n/2+1)}}{2}$. Las expresiones entre paréntesis representan la posi

ción del valor de la variable. Reemplazando se tiene la mediana = $\frac{8+10}{2} = \frac{18}{2} = 9$
Luego el punto de equilibrio se encuentra entre los valores 8 y 10.

Nota: los datos deben estar ordenados en forma ascendente o descendente.

3.2.1.3 La media aritmética

Es una medida de tendencia central que muestra el valor de tendencia de los datos; su ecuación se representa:

Para la muestra: $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ Para la población: $\mu = \frac{\sum_{i=1}^{N} x_i}{N}$.

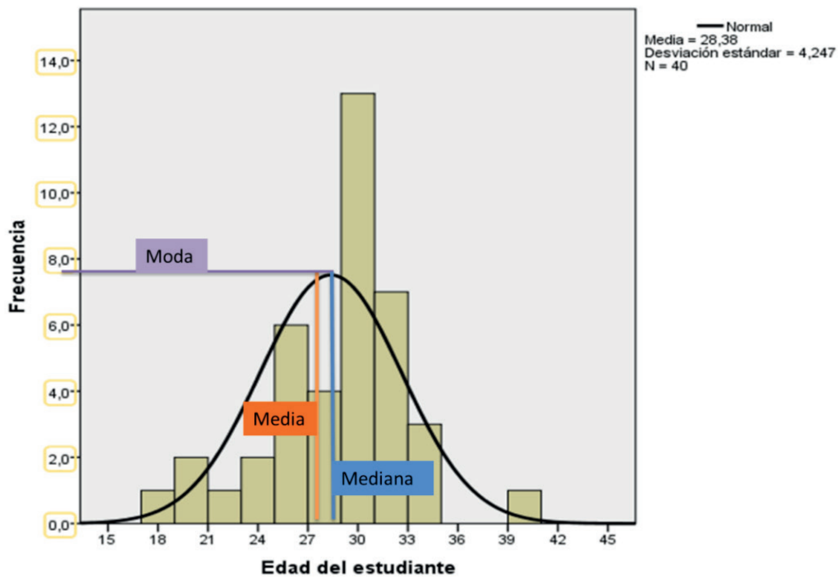
Siendo n y N los tamaños de la muestra y la población respectivamente.

Ejemplo 3.3. Se tiene la edad en años de un grupo de estudiantes de la secundaria: $x = \{15, 16, 16, 18, 20\}$

Su media aritmética es: $\bar{x} = \frac{15 + 16 + 16 + 18 + 20}{5}$

$\bar{x} = 17;$ es el promedio de edad del grupo, es decir, existe una proporción de la edad de los estudiantes que se encuentran alrededor de este valor.

Figura 3.1. Representación de las medidas de tendencia central

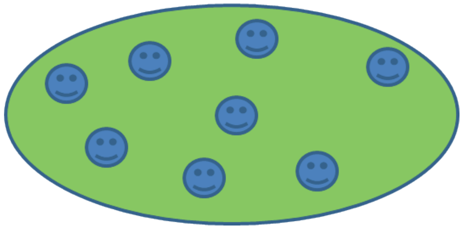


3.2.2 Medidas de variación

Miden la variabilidad de los datos, el cómo se encuentran dispersos los datos. Cuanto menos dispersos se encuentren los elementos del conjunto, habrá menos variabilidad (figura 3.2) y cuanto más dispersos se encuentren, habrá mayor variabilidad (figura 3.3).

Figura 3.2. Menor variabilidad de los datos

Figura 3.3. Mayor variabilidad de los datos



3.2.2.1 La varianza

Es una medida de variación que mide el grado de dispersión en que se encuentran los datos. Se define como el promedio de las desviaciones al cuadrado con respecto al promedio.

Para la muestra: $s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$ Para la población: $\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \bar{X})^2}{N-1}$

Se deben tener las siguientes consideraciones cuando se utiliza la varianza:

- No es una medida para comparar variables.
- Viene en unidades cuadradas.
- Su valor siempre es positivo.

3.2.2.2 La desviación estándar

Es una medida de variación que mide el grado o nivel de dispersión de los datos con respecto a la media.

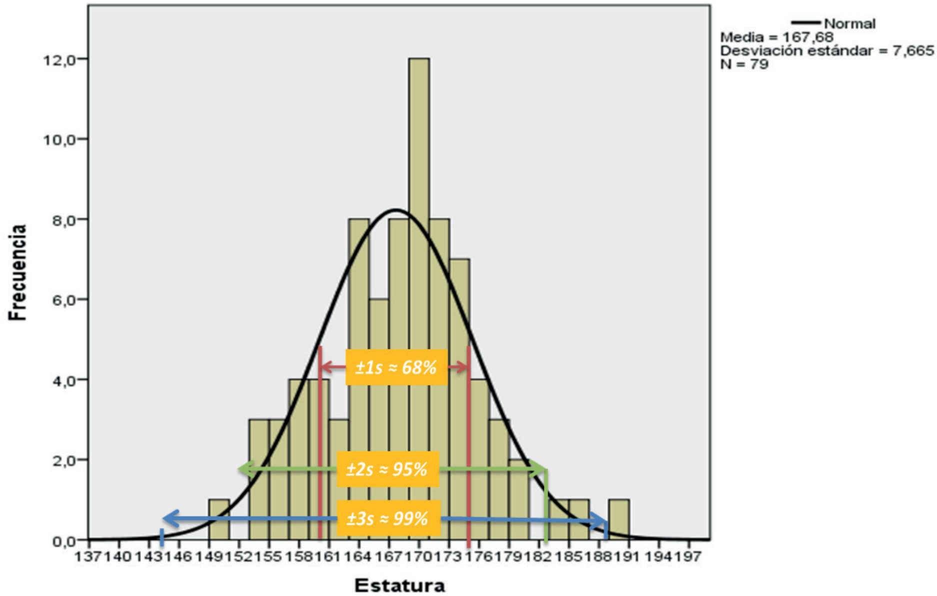
Para la muestra: $s = \sqrt[2]{s^2}$ Para la población: $\sigma = \sqrt[2]{\sigma^2}$

- Su cálculo se basa en la obtención de la raíz cuadrada de la varianza.
- Existe una regla empírica para hallar la dispersión de los datos si se distribuyen en forma de joroba (figura 3.4):
  - ✓ $\bar{x} \pm 1s$: existe el 68 % de la concentración de los datos.
  - ✓ $\bar{x} \pm 2s$: existe el 95 % de la concentración de los datos.
  - ✓ $\bar{x} \pm 3s$: existe el 99 % de la concentración de los datos.

Ejemplo 3.4. Se desea conocer la dispersión de la estatura de un grupo de estudiantes de un colegio determinado; para tal efecto, se selecciona una muestra de 79 estudiantes y se les toma las medidas respectivas.

Los cálculos se han realizado utilizando el programa de R y los resultados se muestran en la figura 3.4.

Figura 3.4. Dispersión de los datos de acuerdo a la desviación estándar de la variable estatura



Interpretación:

- $167.68 \pm 1 * (7.665)$: existe el 68 % de la concentración de los datos.
- $167.68 \pm 2 * (7.665)$: existe el 95 % de la concentración de los datos.
- $167.68 \pm 3 * (7.665)$: existe el 99 % de la concentración de los datos.

3.2.2.3 Intervalo o rango

Se define como la diferencia entre el dato mayor y el dato menor.

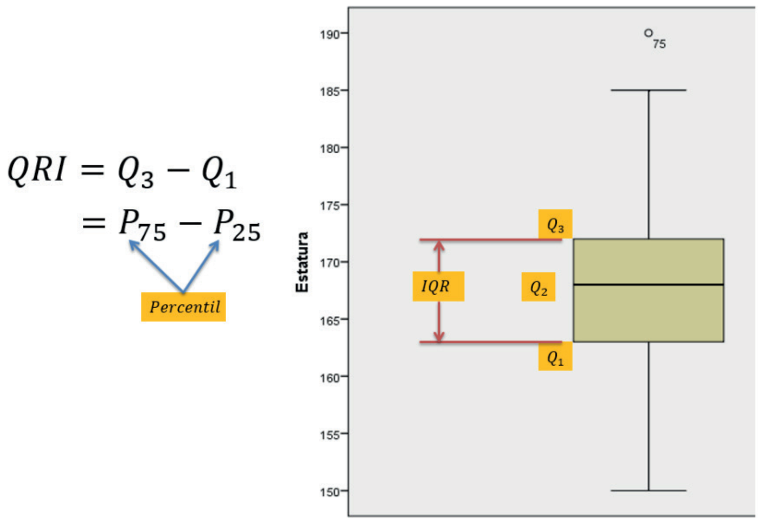
Ejemplo 3.5. Se tienen los siguientes datos ordenados de forma ascendente: $x = \{2, 3, 6, 7, 10, 15\}$. De donde: $Intervalo = 15 - 2 = 13$

3.2.2.4 Rango intercuartil

Se define como la distancia entre el primer y tercer cuartil.
Entre sus características se encuentran:

- Similar al rango, difiere en que no incluye datos extremos.
- No es sensible a valores extremos.
- El primer cuartil se encuentra en el percentil 25.
- El tercer cuartil se encuentra en el percentil 75.

Figura 3.5. Cálculo del rango intercuartil



3.2.2.5 Coeficiente de variación

Se define como el cociente entre la desviación estándar y la media de los datos de la variable cuantitativa (DANE, 2005).

Para la muestra: $cv = \frac{s}{\bar{x}}$: Para la población: $CV = \frac{\sigma}{\mu}$

Entre sus características se encuentran:

- Es una medida que identifica la homogeneidad de los datos.
- Generalmente, su cálculo se expresa en forma de porcentaje.
- Sirve para comparar las dispersiones entre un conjunto de datos de diferentes variables. Es decir, permite comparar el comportamiento entre los datos de dos variables. La escala de valores se categoriza de la siguiente manera:
  - ✓ Se considera homogéneo si $cv < 15\ \%$.
  - ✓ Valores entre: $15 \leq cv \leq 30$: se consideran como homogeneidad moderada.
  - ✓ Valores $cv > 30\ \%$: se consideran los valores con una significativa heterogeneidad en el grupo de datos.

Cuándo no funciona:

- Cuando la media es cero o cercana a cero.
- Cuando toma valores negativos (medidas de temperaturas negativas y su media es negativa).

Ejemplo 3.6. Se tiene la siguiente información de las características físicas de un grupo de estudiantes sobre su edad y el peso. ¿Cuál es la uniformidad de sus datos?

No.	Género	Edad	Peso
1	H	25	75
2	M	22	60
3	M	21	54
4	M	20	63
5	H	60	78
6	M	28	60
7	H	24	67
8	H	20	60
9	M	25	59
10	H	70	75

Nota: H: hombre, M: mujer, Edad: en años, Peso: en kilogramos.

Primero se calculan las medidas de tendencia central y de variación:

Estadísticas	Edad	Peso
Promedio aritmético	31,5	65,1
Desviación estándar	17,99	8,23
Coeficiente de variación	57,11	12,63

Luego se calculan los coeficientes de variación para cada variable:

$$cv_{edad} = \frac{17{,}99}{31{,}5} * 100 \qquad cv_{Peso} = \frac{8{,}22}{65{,}1} * 100$$

$$cv_{edad} = 57{,}11\ \% \qquad cv_{Peso} = 12{,}6\ \%$$

Se realiza el análisis por género.

Caso 1: hombres

No.	Género	Edad	Peso
1	H	25	75
5	H	60	78
7	H	24	67
8	H	20	60
10	H	70	75

Sus cálculos respectivos son:

Estadísticas	Edad	Peso
Media	33,4	78,6
Desviación estándar	14,4	21,8
Coeficiente de variación	43,11	27,73

Caso 2: mujeres

No.	Género	Edad	Peso
2	M	22	60
3	M	21	54
4	M	20	63
6	M	28	60
9	M	25	59

Sus cálculos respectivos son:

Estadísticas	Edad	Peso
Promedio	23,2	59,2
Desviación estándar	3,27	3,27
Coeficiente de variación	14,10	5,53

Análisis:

- En los hombres, existe mayor heterogeneidad en la variable edad que en el peso.
- En las mujeres, las dos variables son homogéneas. Sin embargo, es mayor en la variable peso.
- En el análisis de las variables por separado, se observa un comportamiento diferente al general; esto se debe a las variaciones en las características de las variables.

3.2.3 Medidas de posición relativa o localización

Miden la posición relativa de los datos con respecto a un valor o patrón determinado. Entre ellas se encuentran:

- Percentil.
- Cuartil.
- Cuantil.
- Decil.
- Valor Z.

3.2.3.1 Percentil

Es una medida de localización que ubica el valor P_i en un rango por encima y por debajo del 100 % de los datos. Cuándo se utiliza:

- Por ejemplo, en los resultados de las pruebas de Estado.
- Cuando se desea conocer posiciones en medidas determinadas.

Ejemplo 3.7. La calificación de un estudiante en una materia determinada estuvo en el 92° percentil.

Esto significa que el 92 % de las notas de los estudiantes fueron inferiores a las de él y el restante 8 % obtuvo notas superiores.

La fórmula para localizar la posición del valor buscado se define como:

$$\frac{k * (n+1)}{p}$$

k: *es el percentil buscado* n: *número de elementos de la muestra*

p: *es el número de particiones* = 100

Ejemplo 3.8. Calcular el 10° percentil del siguiente conjunto de datos:
$x = \{14, 16, 17, 18, 20, 21, 22, 22, 24, 25, 27, 30\}$

Los datos deben encontrarse ordenados de manera ascendente o descendente.

Luego: 10° percentil = $\frac{10(12+1)}{100} = \frac{130}{100} = 1{,}3$ ***posición***

Por ende: el valor debe estar entre la 1ª y 2ª posición

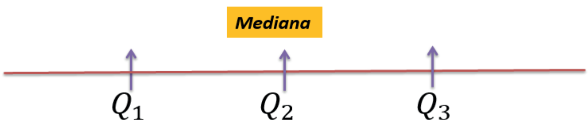
- El valor de la 1ª posición es 14 y la 2ª es 16.
- Se toma la diferencia: $16 - 14 = 2$.
- Este valor se multiplica por el excedente:
- $2 * 0{,}3 = 0{,}6$
- Este valor se le suma a la 1ª ***posición*** = 14,6.
- Luego el 10° percentil es: 14,6

{14, ,16,17,18,20,21,22,22,24,25,27,30}

3.2.3.2 Cuartil

Es una medida de localización que divide el grupo en consideración en cuatro partes y lo caracteriza con tres valores de la variable.

Figura 3.6 El cuartil se caracteriza por dividir el grupo de datos en cuatro partes



- Q_1: supera el 25 % de las observaciones y a su vez es superado por un 75 %.
- Q_2: corresponde al 50 % de las observaciones. Es equivalente a la mediana.
- Q_3: supera al 75 % de las observaciones y a su vez es superado por un 25 %.

Ejemplo 3.9. Calcular el cuartil 2 (Q_2) de la siguiente muestra:
$x = \{14, 16, 17, 18, 20, 21, 22, 22, 24, 25, 27, 30\}$
La posición es: $\frac{2(12+1)}{4} = 6{,}5$

6° posición = 21 7° posición = 22

Luego: $Q_2 = \frac{21+22}{2} = 21{,}5$ es el cuartil 2

3.2.3.3 Valor Z

Es una medida de posición estadística; también se conoce como puntuación Z o Z. Indica cuántas desviaciones estándar está un valor por encima o por debajo de la media de una distribución de datos. Tiene la propiedad de que permite comparaciones entre ellos. En estadística es muy útil cuando se desea comparar datos procedentes de variables que están en unidades de medida diferente; luego se procede a estandarizarlas, es decir, convertirlas a unidades de desviaciones estándar. La fórmula para estandarizar el dato o calcular el valor Z es:

$$Z = \frac{X - \mu}{\sigma}$$

En donde:
X: es el valor del dato μ : la media poblacional σ: la desviación estándar

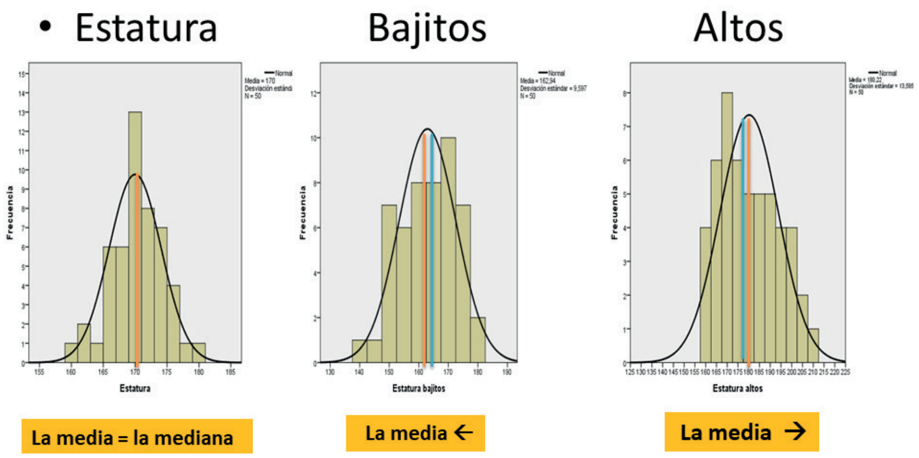
3.2.4 Medidas robustas

Se definen como aquellas que no son influenciadas por los datos extremos o atípicos en un conjunto de datos. Las estadísticas robustas guardan el comportamiento de los métodos comunes y son de utilidad cuando se detectan en la muestra datos que se encuentran fuera del rango o atípicos.

Robustez: su aplicación obedece a la aparición de datos extremos (atípicos) o que, al momento de su registro, fueron mal digitados en una muestra y para disminuir el error en los cálculos se utilizan modelos estadísticos robustos que no incluyen estos datos atípicos, también llamados outlayers.

Una medida es robusta cuando es resistente a los cambios abruptos, atípicos, *outlayers* o mal digitados que existen en la variable de análisis. Entre estas medidas se encuentra la mediana, mientras que la media no lo es. La siguiente gráfica muestra claramente la sensibilidad de la mediana y la media cuando se tiene datos atípicos:

Figura 3.7. Influencia de los datos atípicos cuando se aplican las estadísticas de la media y la mediana a un conjunto de datos



Nota: el primer gráfico muestra la distribución de los datos; la media y la mediana tienen el mismo valor. En el segundo gráfico, se observa una tendencia de las estaturas hacia la izquierda; la media tiende hacia la izquierda. Y en el último, se observa la media con una desviación hacia la derecha.

3.2.4.1 *Trimean*

Se define como la media recortada; calcula una especie de media con unos percentiles que no se afectan por los valores extremos (*outlayer*).

$$Trimean = \frac{P_{25} + 2P_{50} + P_{75}}{4}$$

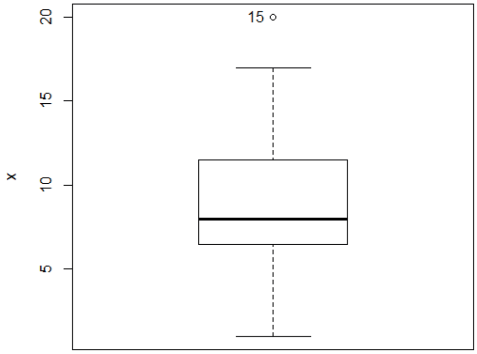
Se ordenan los datos de menor a mayor: $x_1, x_2, \dots, x_{n-1}, x_n$, se quitan los extremos (colas) y se les calcula la media a los internos. Se recomienda máximo el 15 % a lado y lado.

Ejemplo 3.10. Se tiene el siguiente conjunto de datos:
$y = \{1, 1, 5, 6, 7, 7, 8, 8, 9, 10, 11, 12, 12, 17, 20\}$.

$n = 15 \qquad P_{25} = 6{,}5 \quad P_{50} = 8 \qquad P_{75} = 11{,}5$

$$Trimean = \frac{6{,}5 + 2(8) + 11{,}5}{4} = 8{,}5$$

Figura 3.8. En el diagrama de caja y bigotes se evidencia la no inclusión de los datos atípicos en el cálculo del *trimean*



3.2.4.2 *Trimmed mean*

$$\boldsymbol{Trimmed\,mean} = \tilde{x}_\alpha = \frac{1}{n-2k}\sum_{i=k+1}^{n-k} x_i$$

- k es el entero más cercano de $\alpha * n$ y equivale al número de datos que se eliminan de cada cola.
- α es el porcentaje de datos que se desea desechar.

Ejemplo 3.11. Se tiene el siguiente conjunto de datos:
$y = \{1, 1, 5, 6, 7, 7, 8, 8, 9, 10, 11, 12, 12, 17, 20\}$

$$n = 15 \qquad \alpha = 10 \qquad k = n * \alpha = 1{,}5 \approx 2{,}0$$

$$\tilde{x}_\alpha = \frac{1}{15-4} * (5 + 6 + 7 + \cdots + 12 + 12)$$

$$\boldsymbol{Trimmed\,mean} = \tilde{x}_\alpha = 8{,}63$$

Ejercicios

3.1. De acuerdo con la información recolectada respecto a las condiciones climatológicas por una central meteorológica ubicada en una isla de la región del Pacífico colombiano, se pretende realizar una caracterización del comportamiento del clima en el periodo de tiempo en que fueron tomados los datos, como se evidencia en la siguiente tabla:

Temperatura (ºC)	Velocidad del viento (km/h)	% Humedad	Precipitación (mm/m^2)
8	300	70	520
10	160	92	580
11	130	78	540
13	125	70	510
13	120	84	555
14	115	93	600
15	110	85	560
17	105	30	200
21	100	60	480
22	100	40	400
25	90	40	410
26	85	60	485
37	85	55	460
26	80	50	440
27	70	65	500
29	70	60	480
27	70	45	420
35	70	45	425
30	65	35	350
41	45	40	390

a. Realizar la estadística descriptiva para las cuatro variables con su análisis.
b. Realizar el histograma correspondiente a cada variable con el análisis.
c. Llevar a cabo un análisis comparativo de las variables y explicar si existe alguna relación o, por el contrario, carece de relación alguna.
d. Realizar los diagramas de cajas y bigotes y explicar la existencia de datos *outlayers* de las variables.
e. En las variables en donde se encuentren datos *outlayers*, realizar los cálculos de las medidas robustas (media recortada) y compararlas con el valor promedio.

3.2. Se desea realizar un estudio sobre los resultados de las pruebas de Estado Saber Pro respecto a las competencias genéricas que obtuvieron los estudiantes de una universidad durante un semestre determinado. Para tal efecto, se seleccionó una muestra de 150 registros, recopilando la información como se muestra en la tabla. Para efectos de la comparación, se tuvo en cuenta el promedio nacional en la competencia genérica de análisis, que fue de 165 puntos con una desviación estándar de 10 puntos.

198	183	170	166	160	155	153	145	173	158	185	192	178
195	184	170	166	160	155	153	145	172	158	181	172	181
180	184	170	166	160	155	153	145	173	158	168	168	168
178	184	170	166	160	155	153	145	170	158	164	164	164
178	184	170	166	160	155	153	145	170	155	158	158	158
175	184	170	166	160	155	150	145	168	155	155	155	155
175	184	170	166	160	155	150	130	166	155	150	150	
170	181	168	165	158	155	150	130	166	153	183	150	
177	181	168	165	158	155	150	123	166	153	160	183	
184	181	168	165	158	155	150	115	166	153	145	162	
191	181	168	165	158	155	150	183	163	150	183	145	
198	181	168	164	158	155	150	183	162	145	160	145	

a. Realizar la estadística descriptiva con su análisis.
b. ¿Cuál es el coeficiente de asimetría de Pearson de los resultados? Definir la tendencia de los datos.
c. Calcular el 25 y 75 percentil.
d. ¿Qué tanto se desvió la mediana de los universitarios respecto a la media nacional?
e. Existen algunos resultados atípicos. Determinarlos y explicar el efecto que causa dentro de la muestra seleccionada.
f. ¿Los estudiantes obtuvieron un puntaje superior a la media nacional?
g. ¿En qué rango de puntaje se encuentra el 68 % de los resultados con respecto a la media de los estudiantes de la universidad?

3.3. Se desea conocer la caracterización socioeconómica en que se encuentra una población de profesionales recién egresados de una institución de educación superior. El estudio busca identificar las condiciones en que se hallan los recién egresados a la hora de encontrarse como profesionales. El propósito es seguirle la ruta en su desarrollo social y económico para identificar si ha existido algún cambio gracias al hecho de haber adquirido un título profesional. Para tal efecto, se ha seleccionado una muestra aleatoria de un grupo de 50 egresados, al cual se les pidió que diligenciaran una encuesta que solicitaba información básica como su género, edad, peso, estatura, cuál es su estado civil en este momento, cuántos hijos tiene y cuál era el ingreso promedio que tiene al momento de obtener su título profesional. La primera parte del estudio solicita llevar a cabo la estadística descriptiva que caracterice a la población en estudio.

No.	Sexo	Edad	Estatura (cm)	Peso (kg)	Nivel máximo de formación	Estado civil	Número de hijos	Ingresos SMLV	Estrato socioeconómico
1	H	20	180	75	Universitario	Soltero	1	3	2
2	H	18	176	68	Tecnólogo	Soltero	0	1	2
3	H	27	169	67	Universitario	Unión libre	2	3	3
4	H	25	173	76	Universitario	Unión libre	1	3	2
5	H	28	178	79	Tecnólogo	Unión libre	1	2	2
6	H	32	165	85	Universitario	Soltero	0	3	3
7	H	35	172	75	Universitario	Soltero	0	5	4
8	H	19	166	65	Tecnólogo	Soltero	0	1	2
9	H	22	169	72	Tecnólogo	Soltero	1	2	2
10	H	24	178	75	Universitario	Soltero	0	2	2
11	H	25	176	70	Universitario	Soltero	1	3	2
12	H	21	160	62	Tecnólogo	Soltero	0	1	1
13	H	24	165	74	Tecnólogo	Soltero	0	3	2
14	H	29	174	77	Universitario	Unión libre	1	4	3
15	H	20	169	70	Tecnólogo	Soltero	0	1	2
16	H	22	165	72	Tecnólogo	Soltero	0	0	1
17	H	25	178	77	Universitario	Soltero	0	3	2
18	H	28	166	73	Tecnólogo	Casado	1	3	2
19	H	25	172	75	Universitario	Soltero	1	4	3
20	H	24	168	71	Universitario	Soltero	0	0	2
21	H	23	166	69	Universitario	Casado	1	3	2
22	H	22	160	66	Tecnólogo	Soltero	1	2	2
23	H	21	155	61	Tecnólogo	Unión libre	1	1	2
24	H	21	166	73	Tecnólogo	Soltero	0	0	3
25	H	21	169	80	Tecnólogo	Soltero	1	2	2
26	M	18	160	61	Tecnólogo	Soltero	0	0	2
27	M	19	163	63	Tecnólogo	Soltero	0	0	1
28	M	20	158	56	Tecnólogo	Soltero	0	0	2
29	M	21	166	65	Tecnólogo	Separado	1	1	2
30	M	24	167	66	Universitario	Casado	1	3	2
31	M	27	169	68	Universitario	Unión libre	1	3	2
32	M	24	162	63	Tecnólogo	Soltero	0	2	2
33	M	23	169	64	Tecnólogo	Soltero	0	1	2
34	M	21	155	57	Tecnólogo	Soltero	0	0	2
35	M	22	162	60	Tecnólogo	Soltero	0	0	2
36	M	28	168	72	Universitario	Separado	1	4	2
37	M	25	163	68	Universitario	Soltero	0	3	2
38	M	23	178	73	Tecnólogo	Separado	1	2	2
39	M	22	167	68	Universitario	Soltero	0	3	3
40	M	24	161	66	Universitario	Soltero	0	0	2

41	M	29	172	74	Tecnólogo	Soltero	2	3	2
42	M	24	159	63	Universitario	Soltero	1	3	2
43	M	20	160	62	Tecnólogo	Soltero	0	0	2
44	M	--	155	56	Tecnólogo	Soltero	0	0	2
45	M	21	161	60	Tecnólogo	Soltero	0	2	2
46	M	20	159	58	Tecnólogo	Soltero	1	2	2
47	M	23	175	70	Tecnólogo	Soltero	0	0	2
48	M	27	169	65	Universitario	Soltero	1	4	3
49	M	25	166	69	Universitario	Soltero	2	4	3
50	M	28	172	--	Universitario	Separado	2	3	2

a. Realizar un análisis estadístico descriptivo e identificar las características principales del grupo de recién egresados.
b. ¿Existe alguna relación entre la edad, el peso y la estatura de los profesionales?, ¿cómo es el comportamiento por género?
c. ¿Existe algún tipo de relación entre el número de hijos y el ingreso promedio de los profesionales?
d. Se puede afirmar que el nivel de formación se encuentra directamente relacionado con el ingreso promedio de los profesionales. Es decir, cuanto más alto sea el nivel de formación, ¿mayores serán los ingresos promedio percibidos por los recién egresados?
e. Existen datos atípicos que afectan el estudio. Utilizar las medidas robustas para realizar el análisis y comparar los resultados con los obtenidos en el inciso a). Realizar las conclusiones a que den lugar.
f. ¿Son los egresados con nivel de formación universitaria los que tienen mejor estabilidad económica y su vínculo marital es más estable que los de formación tecnológica?