

|                          |   |                    |                 |
|--------------------------|---|--------------------|-----------------|
| NOMBRE                   | MOISES MARIN  | MATRÍCULA          | 17980023        |
| MAESTRIA/DOCTORADO       | CIENCIA DE DATOS E INTELIGENCIA DE NEGOCIOS   | FECHA              | 24-FEB-2021     |
| PROFESOR                 | José Luis Avila Valdez  | GRUPO              | 2-22            |
|                          |   |                    |                 |
| ACTIVIDAD DE APRENDIZAJE |   | MODALIDAD          | TIEMPO ESTIMADO |
| Laboratorio 1            |   | Individual / Mixta | 180 minutos     |
|                          |   |                    |                 |
| TEMA                     | Medidas estadísticas descriptivas.  |                    |                 |
| PROPÓSITO                | Resume e interpreta un conjunto de datos, a través de medidas estadísticas, para la toma de decisiones. |                    |                 |
|                          |   |                    |                 |
| INDICACIONES             | Contesta cada uno de los ejercicios, para ello debes utilizar SPSS.                                     |                    |                 |

**Caso 1.** El correo electrónico no solicitado y el spam afectan la productividad de los empleados. Una empresa monitoreó a una muestra de dichos empleados para determinar el tiempo improductivo por día dedicado a correo electrónico no solicitado y spam. Los datos del archivo Spam.xlsx representan el tiempo en minutos dedicado a esta tarea por los empleados de dos departamentos de esta empresa.

1. ¿Quién es la población en estudio?

**todos los empleados de esta empresa**

2. ¿Quién es la unidad de análisis?

**cada uno de los emplea**

3. ¿De qué tamaño fue la muestra?

**403**

4. Identifica el tipo de variable y la escala de medición de cada una de ellas.

| Variable            | Tipo     | Escala  |
|---------------------|----------|---------|
| <b>Id</b>           | nominal  | nominal |
| <b>Departamento</b> | nominal  | nominal |
| <b>Sexo</b>         | nominal  | nominal |
| <b>Tiempo</b>       | numérico | escala  |

5. Calcula e interpreta las siguientes medidas estadísticas para el tiempo dedicado al spam para los empleados del departamento de calidad.

| Medida              | Valor  | Interpretación   |
|---------------------|--------|--|
| Media               | 21.70  | la muestra de empleados de la empresa dedican en promedio 21 minutos al día a correo no solicitado y spam                    |
| Mediana             | 16     | 16 minutos es el límite que divide el 50% del tiempo que la muestra de empleados dedica al día a correo no solicitado y spam |
| Varianza            | 374.26 | la dispersión de los datos tiene una varianza de 374   |
| Desviación estándar | 19.35  | la dispersión de los datos tiene una desviación estándar de 19 minutos   |
| Rango               | 118    | la diferencia entre el valor mínimo y máximo es de 118 minutos   |
| Q <sub>1</sub>      | 8.5    | el 25% de los empleados de la muestra dedican 8.5 minutos o menos al correo no solicitado y spam                             |
| Q <sub>2</sub>      | 16     | el 50% de los empleados de la muestra dedican 16 minutos o menos al correo no solicitado y spam.                             |

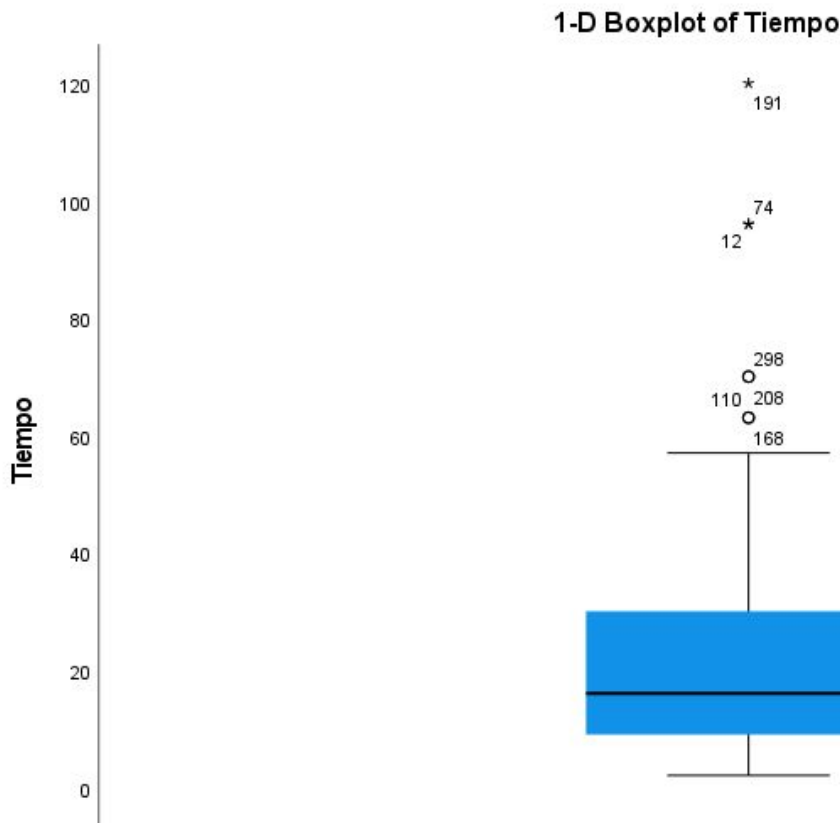
|                      |       |  |
|----------------------|-------|--|
| Q <sub>3</sub>       | 30.50 | el 75% de los empleados de la muestra dedican media hora o menos a correo no solicitado y spam     |
| P <sub>7</sub>       | 3.46  | el 7% de los empleados de la muestra dedican 3.46 minutos o menos a correo no solicitado y spam    |
| P <sub>56</sub>      | 17.68 | el 56% de los empleados de la muestra dedican 17.686 minutos o menos a correo no solicitado y spam |
| P <sub>75</sub>      | 30.50 | el 75% de los empleados de la muestra dedican 30.56 minutos o menos a correo no solicitado y spam  |
| D <sub>5</sub>       | 16    | el 50% de los empleados de la muestra dedican 166 minutos o menos a correo no solicitado y spam    |
| D <sub>9</sub>       | 49.20 | el 90% de los empleados de la muestra dedican 49.20 minutos o menos a correo no solicitado y spam  |
| Coeficiente de sesgo | 1.879 | la muestra presenta alargamiento de la distribución hacia la derecha, es decir, sesgo positivo     |

### Statistics

Tiempo

|                        |         |         |
|------------------------|---------|---------|
| N                      | Valid   | 177     |
|                        | Missing | 0       |
| Mean                   |         | 21.70   |
| Median                 |         | 16.00   |
| Std. Deviation         |         | 19.346  |
| Variance               |         | 374.256 |
| Skewness               |         | 1.879   |
| Std. Error of Skewness |         | .183    |
| Range                  |         | 118     |
| Percentiles            | 7       | 3.46    |
|                        | 25      | 8.50    |
|                        | 50      | 16.00   |
|                        | 56      | 17.68   |
|                        | 75      | 30.50   |
|                        | 90      | 49.20   |

6. Elabora un diagrama de caja-brazos para el tiempo dedicado al spam para los empleados del departamento de calidad.

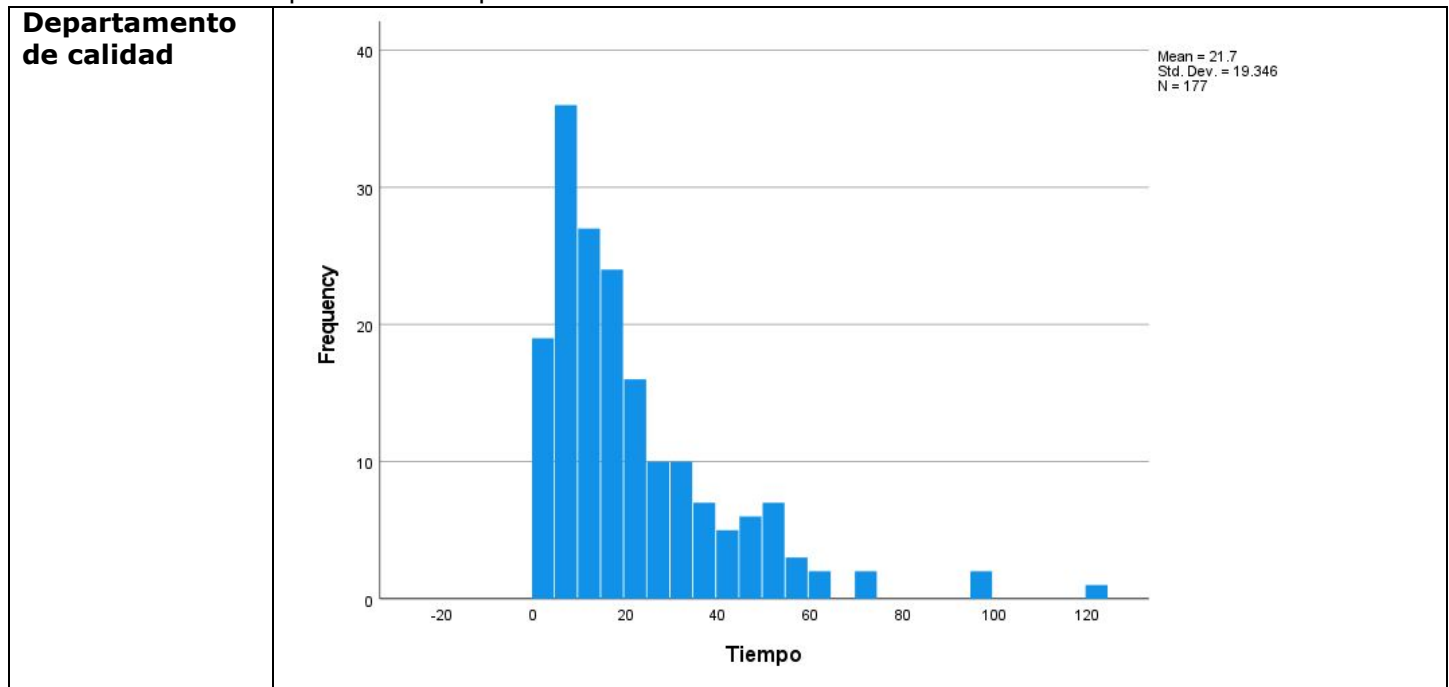


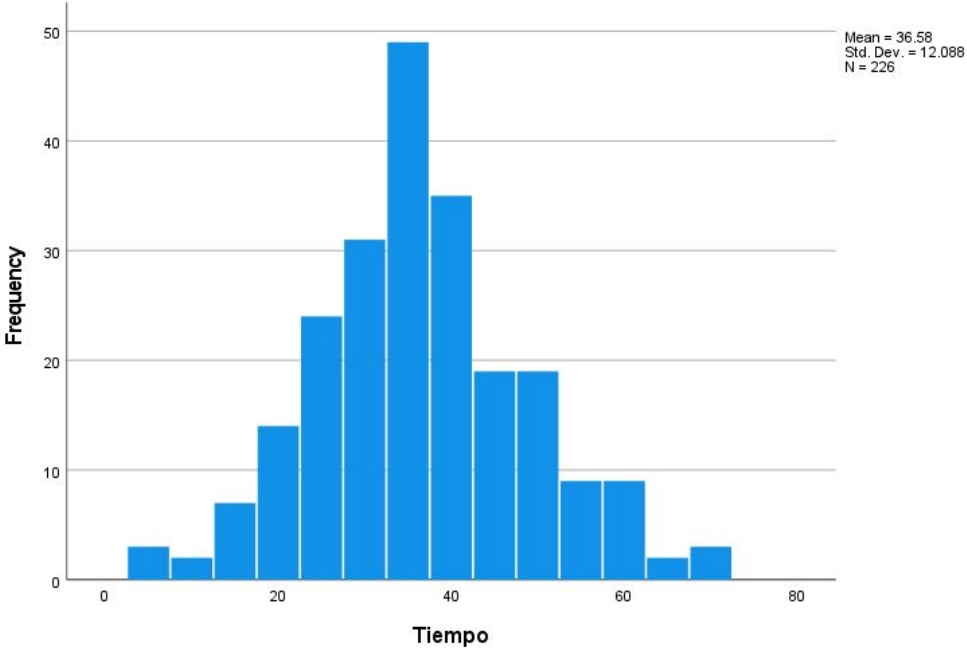
7. Calcula e interpreta las siguientes medidas estadísticas para el tiempo dedicado al spam para los empleados de cada departamento y realiza una interpretación comparativa.

| Medida              | Valor   |        | Interpretación comparativa  |
|---------------------|---------|--------|---|
|                     | Calidad | Ventas |   |
| Media               | 21.70   | 36.58  | la muestra de empleados del departamento de ventas dedica en promedio 15 minutos más al día a correo no deseado y spam  |
| Mediana             | 16      | 36     | la muestra de empleados del departamento de ventas tiene 36 minutos como límite que divide el 50% de empleados, es decir, el 50% de los empleados del departamento de ventas utiliza hasta 20 minutos mas en correo no deseado y spam que los empleados del departamento de calidad de la muestra |
| Desviación estándar | 19.35   | 12.10  | el departamento de calidad presenta una mayor dispersión en los minutos que los empleados dedican a correo no deseado y spam  |

|                      |       |      |   |
|----------------------|-------|------|---|
| Rango                | 118   | 66   | el departamento de calidad tiene una mayor diferencia entre el valor máximo y mínimo de minutos que empleados de la muestra dedican a correo no deseado y spam, en otras palabras, empleados del departamento de ventas de la muestra están más parejos en la cantidad de minutos que dedican al día a correo no deseado y spam |
| Coeficiente de sesgo | 1.879 | .182 | ambas distribuciones tienen sesgo positivo, siendo el sesgo del departamento de calidad mayor, la distribución del departamento de ventas está más cerca a una distribución simétrica   |

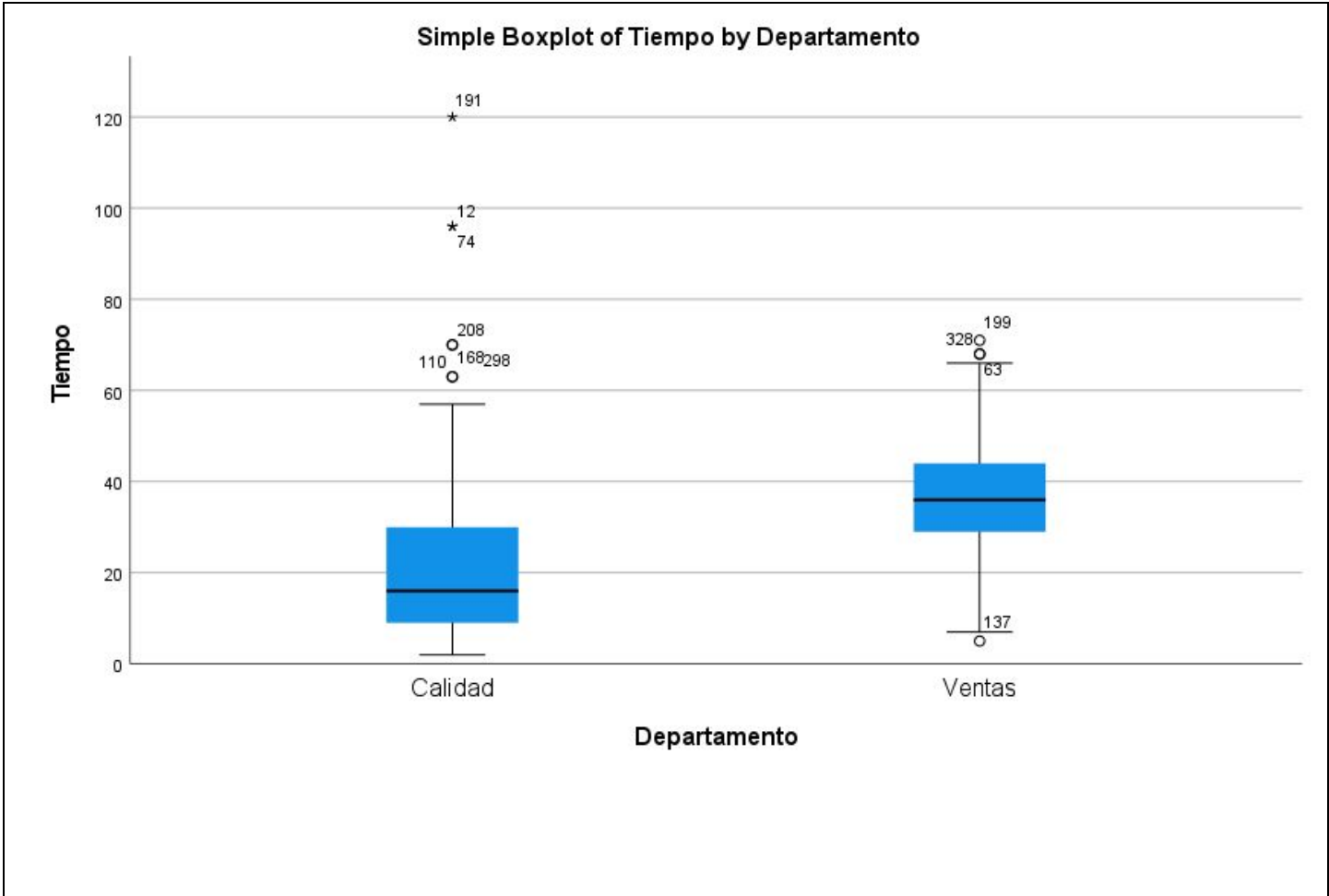
8. Elabora un histograma para el tiempo dedicado al spam para los empleados de cada departamento y realiza una interpretación comparativa.



|                                   |   |
|-----------------------------------|---|
| <b>Departamento de ventas</b>     |  <p>Mean = 36.58<br/> Std. Dev. = 12.088<br/> N = 226</p>         |
| <b>Interpretación comparativa</b> | el departamento de ventas tiene una distribución cercana a la simétrica, el departamento de calidad tiene un alargamiento a la derecha muy marcado. |

9. Dibuja en una misma figura los dos diagramas de caja-brazos de la variable tiempo dedicado al spam, uno para cada departamento y realiza tres interpretaciones comparativas sobresalientes de la variable en estudio.

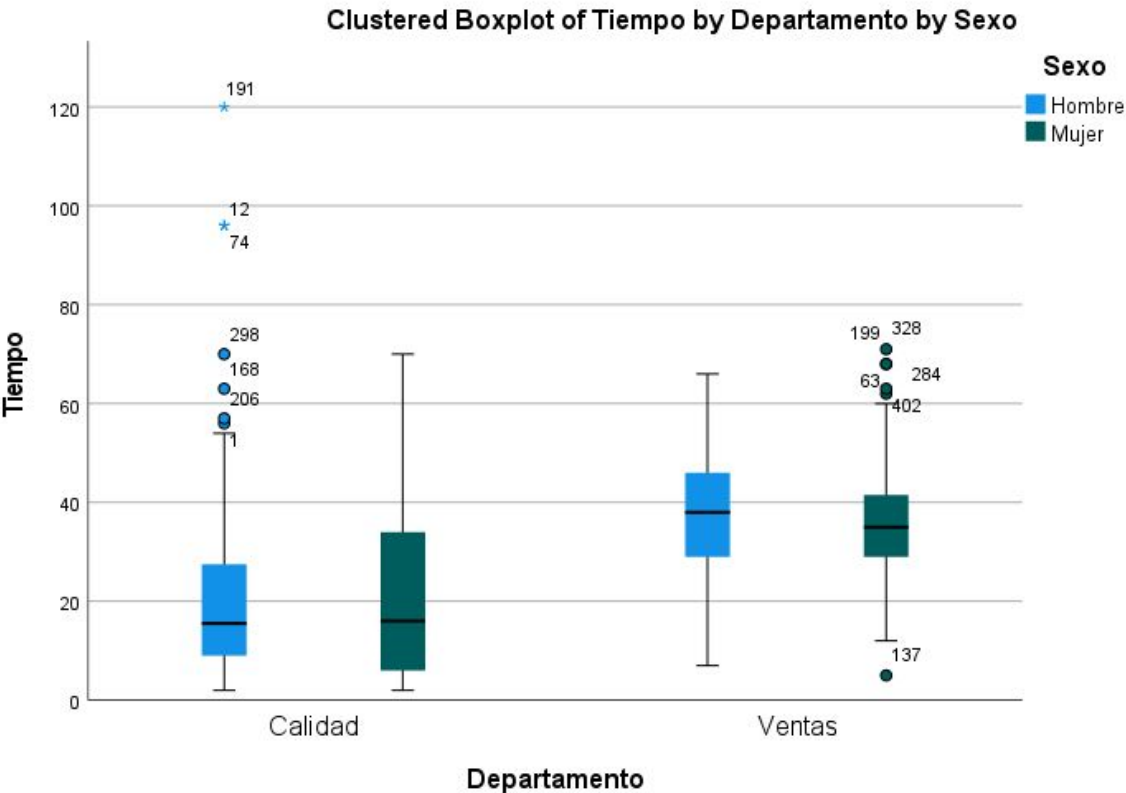
Diagrama



|                  |   |
|------------------|---|
| Interpretación 1 | el departamento de calidad tiene dos valores extremos mientras que el departament de ventas no tiene ninguno  |
| Interpretación 2 | el departamento de ventas tiene un valor atípico inferior mientras que el departament de calidad no tiene ninguno                                     |
| Interpretación 3 | 50% de los empleados del departamento de ventas dedican más minutos a correo n deseado y spam que el 75% de los empleados del departamento de calidad |

10. Dibuja en una misma figura cuatro diagramas de caja-brazos de la variable tiempo dedicado al spam, uno para hombres y otro para mujeres dentro de cada departamento. Redacta un párrafo de tal forma que resuma los hallazgos encontrados.

Diagrama



Interpretación

las mujeres del departamento de calidad tienen mayor diferencia entre ellas en el número de minutos dedicados a correo no deseado y spam que hombres en calidad que hombres y mujeres en el departamento de ventas.

**Caso 2.** Consumer Reports proporciona calificaciones de satisfacción del clientes en general para los servicios de telefonía celular AT&T, Sprint, T-Movil y Verizon en zonas metropolitanas importantes de todo Estados Unidos. La calificación de cada servicio refleja la satisfacción del clientes considerando una variedad de factores como el costo, los problemas de conectividad, las llamadas suspendidas, la interferencia estática y el soporte técnico. Se utilizó una escala de satisfacción de 0 a 100, en la cual 0 indica una insatisfacción total y 100 una satisfacción total. La calificación para los cuatro servicios de telefonía celular en 20 zonas metropolitanas se muestran en el archivo CellService.xlsx.

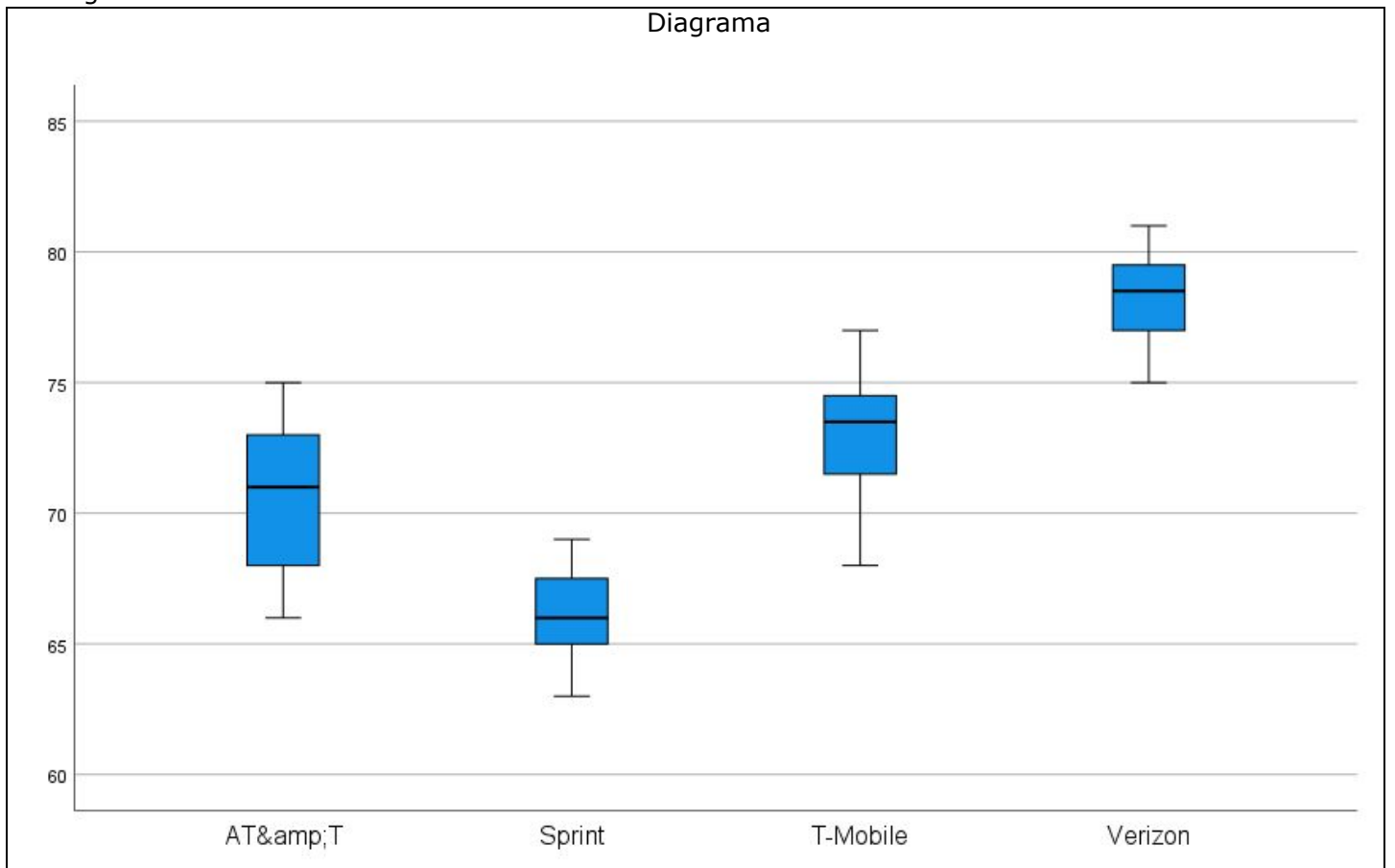
1. ¿Quién es la variable en estudio?

**la satisfacción de clientes**

2. Calcula las siguientes medidas de posición.

| Medida | Compañía |        |          |         |
|--------|----------|--------|----------|---------|
|        | AT&T     | Sprint | T-Mobile | Verizon |
| $Q_1$  | 68       | 65     | 71.25    | 77      |
| $Q_2$  | 71       | 66     | 73.50    | 78.50   |
| $Q_3$  | 73       | 67.75  | 74.75    | 79.75   |

3. Dibuje en una misma figura los cuatro diagramas de caja-brazos, uno para cada compañía y realiza cuatro interpretaciones comparativas sobresalientes de la satisfacción del cliente observadas en la figura realizada.





|                  |  |
|------------------|--|
|                  |  |
| Interpretación 1 | Verizon es la numero 1 en calificaciones de satisfacción         |
| Interpretación 2 | Sprint tiene el peor desempeño en calificaciones de satisfacción |
| Interpretación 3 | El 100% de calificaciones de Verizon son de 75 o superior.       |
| Interpretación 4 | 25% de calificaciones de Sprint estan abajo de 65                |

**Caso 3.** Heavenly Chocolates fabrica y vende productos de chocolate de calidad en su planta y tienda minorista ubicada en Nueva York. Hace dos años la empresa desarrolló un sitio web y comenzó a vender sus productos por Internet. Las ventas electrónicas han excedido de las expectativas de la empresa y la gerencia ahora está considerando estrategias para incrementar aún más. Para saber más sobre los clientes del sitio web, se seleccionó una muestra de transacciones de Heavenly Chocolate de las ventas del mes anterior, archivo Shoppers.xlsx. Los datos recabados fueron: día de la semana (Day) en que se realizó cada transacción, el tipo de navegador (Browser) usado por el cliente, el tiempo invertido en el sitio web (Time, en minutos), el número de páginas visitadas (Page Viewed) y la cantidad gastada (Amount Spent, en dólares) para cada uno de los clientes.

1. ¿De qué tamaño fue la muestra?

**50**

2. Especifica todas las variables en estudio, el tipo de variable y la escala o nivel de medición de cada una de ellas. Además, para cada variable cualitativa especifica si corresponde a una variable discreta o continua.

| <b>Variable</b>                  | <b>Tipo</b>         | <b>Escala o nivel de medición</b> |
|----------------------------------|---------------------|-----------------------------------|
| dia de la semana                 | string - discreta   | nominal                           |
| tipo de navegador                | string - discreta   | nominal                           |
| tiempo invertido en el sitio web | numérico - continua | escala                            |
| numero de paginas visitadas      | numérico - discreta | escala                            |
| cantidad gastada                 | numérico - continua | escala                            |

3. Obtenga e interprete las siguientes medidas estadísticas para la cantidad gastada por cada cliente.

| Medida               | Valor  | Interpretación   |
|----------------------|--------|--|
| Media                | 68.13  | en promedio las transacciones de la muestra gastaron 68.13 dólares   |
| Mediana              | 62.15  | 62.15 dólares es el límite que divide el 50% de las transacciones    |
| Moda                 | 17.84  | la cantidad gastada más frecuente es 17.84 dólares                   |
| Desviación estándar  | 32.34  | la dispersión de los datos tiene una desviación estándar de 32.34    |
| Rango                | 140.67 | la diferencia entre el valor mínimo y máximo es de 140.67 dólares    |
| Q <sub>1</sub>       | 43.85  | el 25% de transacciones en la muestra gastaron 43.85 dólares o menos |
| Q <sub>2</sub>       | 62.15  | el 50% de transacciones en la muestra gastaron 62.15 dólares o menos |
| Q <sub>3</sub>       | 84.13  | el 75% de transacciones en la muestra gastaron 84.13 dólares o menos |
| P <sub>9</sub>       | 32.71  | el 9% de transacciones en la muestra gastaron 32.71 dólares o menos  |
| P <sub>34</sub>      | 52.79  | el 34% de transacciones en la muestra gastaron 52.79 dólares o menos |
| P <sub>75</sub>      | 84.13  | el 75% de transacciones en la muestra gastaron 84.13 dólares o menos |
| D <sub>6</sub>       | 68.34  | el 60% de transacciones en la muestra gastaron 68.34 dólares o menos |
| D <sub>8</sub>       | 94.24  | el 80% de transacciones en la muestra gastaron 94.24 dólares o menos |
| Coeficiente de sesgo | .337   | la muestra presenta distribución cercana a simétrica                 |

### Statistics

Amount Spent (\$)

|                        |         |                         |
|------------------------|---------|-------------------------|
| N                      | Valid   | 50                      |
|                        | Missing | 0                       |
| Mean                   |         | 68.12820000             |
| Median                 |         | 62.15000000             |
| Mode                   |         | 17.8400000 <sup>a</sup> |
| Std. Deviation         |         | 32.34375786             |
| Skewness               |         | 1.049                   |
| Std. Error of Skewness |         | .337                    |
| Range                  |         | 140.6700000             |
| Percentiles            | 9       | 32.71360000             |
|                        | 25      | 43.84750000             |
|                        | 34      | 52.79260000             |
|                        | 50      | 62.15000000             |
|                        | 60      | 68.33800000             |
|                        | 75      | 84.13250000             |
|                        | 80      | 94.24400000             |

a. Multiple modes exist. The smallest value is shown

4. Dibuje en la misma figura 7 gráficas de caja-brazos (uno para cada día) de la cantidad gastada por cada cliente y realice tres interpretaciones sobresalientes de lo que observa.

