

EJERCICIOS PARA REALIZAR CON MINITAB 17.1.1

Trabajo Práctico Nro. 1: ESTADÍSTICA DESCRIPTIVA

Apellido y nombre:

Resuelva todas las consignas en el software correspondiente, copie y pegue los resultados obtenidos y extraiga las conclusiones que se pidan en cada caso. Todos los trabajos prácticos deberán ser presentados individualmente, **juntos e impresos en papel**, cuidando la prolijidad de los mismos.

1. Los números de tablets vendidas semanalmente por una casa de informática durante el período de 8 semanas fueron: 3, 0, 6, 4, 1, 5, 4, 1. Calcule la media, la varianza y el desvío estándar de esta muestra de tablets vendidas.

2. La cantidad de automóviles vendidos por dos agentes (por encima y por debajo de su compromiso adquirido por contrato) en seis períodos semanales se muestra a continuación:

vendedor 1:	-3	5	0	-2	3	2
vendedor 2:	-7	-5	4	0	-4	3

- a) Calcule la media y todas las medidas de dispersión que conozca para cada uno de los vendedores.
- b) ¿Qué vendedor muestra el desempeño más consistente? **¿Qué medida de las calculadas en (a) empleará para responder? ¿por qué?**

3. Los tamaños de dieciséis archivos con extensión .csv son:

47, 43, 42, 40, 38, 36, 33, 33, 33, 32, 32, 32, 27, 27, 26, 22 KB.

- a) Calcule la media, la mediana, el desvío estándar, el coeficiente de variación y el rango de estos datos.
- b) Vuelva a determinar las mismas medidas eliminando los 4 más grandes. **¿Qué medidas se modificaron y cuáles se mantuvieron invariantes?**

4. Una muestra de 5 latas de una marca de nueces mixtas contiene los siguientes porcentajes de nueces diferentes:

	Lata 1	Lata 2	Lata 3	Lata 4	Lata 5
Almendras	14,6	12,1	13,8	15,0	11,3
Nuez de Brasil	10,2	9,8	11,0	11,1	11,8
Nuez de la India	30,7	31,4	34,0	31,6	29,1
Maní	24,3	26,1	23,3	22,5	27,1
Avellanas	20,2	21,0	17,9	19,8	20,7

Calcule la media, la mediana, la variancia, el desvío estándar, el coeficiente de variación y el rango de cada uno de los cinco tipos de nuez. ¿Cuál conjunto de mediciones presenta mayor variabilidad?

5. Se ha realizado un estudio de la cantidad de autos que llegan a un puesto de control caminero durante intervalos elegidos al azar de 10 minutos:

2, 2, 3, 4, 0, 3, 5, 2, 1, 0, 3, 5, 4, 0, 1, 2, 3, 1, 1, 2,
1, 2, 4, 3, 0, 1, 3, 2, 4, 1, 2, 1, 0, 0, 1, 1, 2, 2, 3, 3

- a) Grafique la distribución de frecuencias absolutas.
- b) Determine la media, la mediana, la moda, la variancia y el desvío estándar. **Interprete la mediana.**

- c) Obtenga la distribución de frecuencias absolutas, relativas porcentuales, acumuladas y acumuladas relativas porcentuales.
Para contestar los siguientes ítems observe la tabla construida en el inciso c.
- d) ¿Qué porcentaje de lapsos de tiempo contienen al menos un automóvil?
- e) ¿Qué porcentaje de lapsos de tiempo contienen menos de cuatro automóviles?
- f) ¿Qué porcentaje de lapsos de tiempo contienen más de dos automóviles?

6. Driven Data (<https://www.drivendata.org/>) es una página web de competencias mundiales de análisis de datos, donde se dejan disponibles bases para el estudio estadístico de ellas con el fin de resolver problemas de predicción. Periódicamente, se publican competencias con un problema a resolver, ganando aquél científico que realice el mejor modelo estadístico de predicción.

Se cree que donar sangre es importante, y los buenos sistemas basados en datos para rastrear y predecir las donaciones y las necesidades de suministros pueden mejorar toda la cadena de suministro, asegurando que más pacientes reciban las transfusiones de sangre que necesitan.

La base de datos *The Blood Transfusion Service Center Dataset* (donaciones.xlsx) pertenece a la competencia *Warm Up: Predict Blood Donations* en Driven Data. La misma contiene columnas con: ID (número de paciente que dona sangre), *Months since Last Donation*, *Number of Donations*, *Volume Donated (cc)*, *Months since First Donation*, *Made donation in March 2007*, esta última vale 1 en caso afirmativo y 0 en otro caso.

Para más información acerca de la base de datos puede consultarse:

<https://www.drivendata.org/competitions/2/warm-up-predict-blood-donations/>.

Descargue la base de datos "donaciones.xlsx" en:

<https://www.dropbox.com/s/fm196ntw2z44yw1/donaciones.xlsx?dl=0>.

Cargue la base de datos **donaciones.xlsx** en el programa simplemente arrastrando el ícono del documento descargado a la hoja de trabajo de Minitab.

- a) ¿Cuántos donadores integran la base de datos?
- b) Clasifique las cinco variables de la base de datos (cuantitativa continua, cuantitativa discreta, cualitativa, ...).
- c) Realice los gráficos que considere adecuados (según el tipo de variable y la cantidad de datos) para las variables: *Meses desde la última donación*, *Número de donaciones*, *Total de volumen donado (cc)* y *Meses desde la última donación*. **Modifique los títulos y nombres de los ejes** para que el gráfico sea fácil de leer.
- d) ¿Alguna de las gráficas del inciso c es simétrica? En caso contrario, diga qué tipo de asimetría presentan.
- e) Presenta mediante un gráfico adecuado la variable *Donó en Marzo de 2007*. **Modifique los títulos y nombres de los ejes** para que el gráfico sea fácil de leer.
- f) Calcule los estadísticos descriptivos de la variable *Número de donaciones* para los grupos que donaron o no sangre en Marzo de 2007.
- g) Elabore un diagrama de caja con grupos para la variable *Número de donaciones* en función de la variable *Donó en Marzo de 2007*.