

Ejercicios resueltos TP2 Parte 1.

Actividad 5: i) Realiza, para cada lote de datos, un gráfico que sea acorde a los datos recolectados.

i) Este ítem lo hacen ustedes. Piensen que tipo de variable es y que gráfico corresponde hacer con esa variable.

ii) Sin obtener ninguna medida, ordena a tu criterio las siguientes distribuciones de acuerdo al nivel de variabilidad que representan (la primera es la menos variable y la última es la más variable).

- a- 10 8 6 0 8 3 2 5 8 0
- b- 1 3 3 5 5 5 7 7 9
- c- 20 1 2 5 4 4 4 0
- d- 5 5 5 5 5 5 5 5 5

ii) 1-d, 2-b, 3-a y 4-c.

iii) Calcular las medidas de tendencia central y la desviación estándar de cada una y verificar si el orden que diste en el ítem anterior es correcto. Analizar a qué se deben las diferencias encontradas entre las medidas.

iii) En primer lugar, calculamos las medidas de tendencia central para las cuatro muestras:

Muestra a: Media: 5, Mediana: 5 y Moda: 8.

Muestra b: Media: 5, Mediana: 5 y Moda: 5.

Muestra c: Media: 5, Mediana: 4 y Moda: 4.

Muestra d: Media: 5, Mediana: 5 y Moda: 5.

Ahora, calculamos la desviación estándar de cada muestra:

Muestra a: Desviación estándar: 3,59.

Muestra b: Desviación estándar: 2,39.

Muestra c: Desviación estándar: 6,3.

Muestra d: Desviación estándar: 0.

Como se puede observar, con las medidas de tendencia central y la desviación estándar, que el orden que dimos en el ítem anterior es correcto. Ya que la muestra **d** tiene desviación estándar 0 nos dice que es el menos variable, es más, no hay variabilidad, luego le sigue la muestra **b**, después la muestra **a** y por último la muestra más variable es la **c**. **CUÁNTO MÁS ALTO LA DESVIACIÓN ESTÁNDAR MÁS VARIABLE ES LA MUESTRA.**

Las diferencias encontradas entre las medidas se deben a la **VARIABILIDAD** de cada muestra.

Actividad 6: Un profesor desea averiguar el nivel de conocimiento de literatura de sus dos grupos de alumnos A y B para planificar sus clases de acuerdo con dicho nivel. Para ello elige al azar a 15 alumnos de cada grupo y les aplica un test elaborado para tal fin, obteniendo los resultados siguientes:

Grupo A	4	3	7	5	6	4	5	4	5	6	7	7	3	4	5
Grupo B	8	9	1	2	8	8	4	3	2	2	10	7	8	2	1

- a) Analizando cada conjunto de datos, responde: ¿En cuál de los dos grupos le será más fácil la planificación de la asignatura? Justifica tu respuesta.

a) La planificación de la asignatura es más fácil en el grupo que resulte más homogéneo, es decir, habría que calcular los promedios de cada grupo y las desviaciones estándar y ver ahí un costo-beneficio porque en realidad le va a convenir trabajar en un grupo con rendimiento más homogéneo no con rendimiento más alto porque quiere decir que los alumnos son todos más o menos parejos. Por lo tanto, observando sólo las muestras, sin realizar cálculos, se puede decir que con el **grupo A** le será más fácil la planificación de la asignatura.

- b) ¿Qué medidas utilizarías para responder a la pregunta del ítem a)? Justifica tu elección y luego calcúlalas y verifica si la respuesta dada en a) coincide o no con lo obtenido numéricamente.

b) Las medidas que utilizaría para responder a la pregunta del ítem a) serían los promedios (o medias) y las desviaciones estándar de cada grupo ya que con sólo calcular los promedios no es suficiente, se necesita también analizar la variabilidad de las muestras.

Calcularemos los promedios y las desviaciones estándar de cada grupo:

GRUPO A: Promedio: 5, desviación estándar: 1,36.

GRUPO B: Promedio: 5, desviación estándar: 3,31.

Dado que ambos grupos tienen el mismo promedio, debemos ver su variabilidad y elegir el grupo que menos variabilidad tenga, es decir, el grupo con la desviación estándar más pequeña. Este grupo corresponde al **grupo A** que coincide con lo contestado en el ítem anterior.

Actividad 7: En las siguientes situaciones ¿qué medida de tendencia central crees que es más adecuada para resumir los datos?

- a) La municipalidad de un pueblo está estudiando la posibilidad de aplicar un impuesto sobre la renta a sus habitantes. Para ello, quiere conocer el promedio de renta de estos ciudadanos y tomarlo como referencia para el cálculo de la base impositiva.

a) La medida de tendencia central que es más adecuada para resumir los datos es la **MEDIA** o el **PROMEDIO**.

- b) Un psicólogo desea estimar el nivel de estrés promedio de los obreros de la construcción sometidos a situaciones de riesgo. Con dicho objetivo, aplica un test a los obreros de una empresa constructora.

b) La medida de tendencia central que es más adecuada para resumir los datos es la **MEDIA** o el **PROMEDIO**.

Actividad 8: La edad a la que comenzaron a fumar los individuos de una muestra extraída de una población de fumadores se detalla a continuación en función del sexo:

Edad	Hombres	Mujeres
(10 – 15]	6	8
(15 – 18]	12	21
(18 – 20]	26	35
(20 – 25]	34	6
(25 – 30]	9	4
(30 – 40]	4	2
Total	91	76

- a) Representa las frecuencias relativas de cada sexo en un mismo gráfico que permita realizar comparaciones. Basándote en el gráfico expresa tus conclusiones en relación con las características principales de cada grupo.

a) Este ítem lo hacen ustedes. Piensen que tipo de variable es y que gráfico corresponde hacer con esa variable.

- b) Calcula el promedio y la desviación estándar en cada uno de los grupos y sacar conclusiones.

b) Para calcular el promedio y la desviación estándar en cada uno de los grupos, debemos, en primer lugar, calcular el punto medio de cada intervalo:

Edad	Hombres	Mujeres	P_M
(10 – 15]	6	8	12,5
(15 – 18]	12	21	16,5
(18 – 20]	26	35	19
(20 – 25]	34	6	22,5
(25 – 30]	9	4	27,5
(30 – 40]	4	2	35
Total	91	76	

Ahora, calculamos el promedio de edad de hombres y mujeres (Se puede realizar con calculadora también):

$$\bar{X}_{EH} = \sum_{i=1}^{n_H} \frac{P_{M_i} \cdot f_{aEH}}{n_H} = \sum_{i=1}^{91} \frac{P_{M_i} \cdot f_{aEH}}{91} = \frac{12,5 \cdot 6 + 16,5 \cdot 12 + 19 \cdot 26 + 22,5 \cdot 34 + 27,5 \cdot 9 + 35 \cdot 4}{91} = \frac{75 + 198 + 494 + 765 + 247,5 + 140}{91} = \frac{1919,5}{91} = 21,09 \Rightarrow \bar{X}_{EH} = 21,09$$

$$\bar{X}_{EM} = \sum_{i=1}^{n_M} \frac{P_{M_i} \cdot f_{aEM}}{n_M} = \sum_{i=1}^{76} \frac{P_{M_i} \cdot f_{aEM}}{76} = \frac{12,5 \cdot 8 + 16,5 \cdot 21 + 19 \cdot 35 + 22,5 \cdot 6 + 27,5 \cdot 4 + 35 \cdot 2}{76} = \frac{100 + 346,5 + 665 + 135 + 110 + 70}{76} = \frac{1426,5}{76} = 18,77 \Rightarrow \bar{X}_{EM} = 18,77$$

Luego, calculamos la desviación estándar de edad de hombres y mujeres (lo realizamos con la calculadora):

$$S_{EH} = 4,74 \text{ y } S_{EM} = 4,23$$

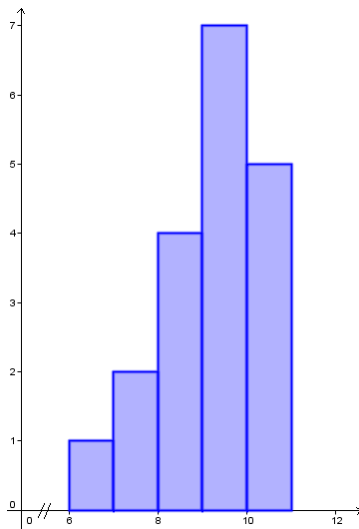
Al hacer las medias y desviaciones estándar de la edad de hombres y mujeres, podemos concluir que, en promedio, la edad a la que comenzaron a fumar las mujeres es de 18,77 mucho más bajo que la edad promedio a la que empezaron a fumar los hombres que es de 21,9 pero los datos varían un poco menos en la muestra de edades de mujeres (desvío: 4,23) que en la muestra de edades de hombres (desvío: 4,74).

Actividad 9: Si la edad media de las mujeres que han contraído matrimonio en un determinado juzgado en los últimos 20 días es 26 con una desviación típica de 0, ¿cuál es la edad de la mujer que se ha casado en último lugar? ¿Cuál es la edad de la más joven?

Tenemos una desviación típica de 0, es decir, que no hay variabilidad lo que significa que los valores de la muestra son todos iguales y puesto que la edad media de las mujeres es 26, podemos decir que cada valor de la muestra es **26**.

Por tanto, la edad de la mujer que se ha casado en último lugar es **26** y la edad de la más joven es **26**.

Actividad 15: i) El tratamiento de los niños con desórdenes de la conducta puede ser complejo. El mismo se puede proveer en una variedad de escenarios dependiendo de la severidad de los comportamientos. Además del reto que ofrece el tratamiento, se encuentran la falta de cooperación del niño/niña y el miedo y la falta de confianza de los adultos. Para poder diseñar un plan integral de tratamiento, los psicólogos de niños y adolescentes puede utilizar la información del niño, la familia, los profesores y de otros especialistas médicos para entender las causas del desorden. Para ello, un psicólogo local ha considerado una muestra aleatoria de 19 niños, anotando el tiempo, medido en horas, necesario que requiere en cada niño para lograr un plan integral del tratamiento, obteniéndose lo siguiente:



Decidir si son verdaderas o falsas las afirmaciones. Justificar en todos los casos.

a) Se puede afirmar que, en promedio, el tiempo requerido, en cada niño, para lograr un plan integral del tratamiento es de 7 horas.

a) **FALSO.** Como en la gráfica se observa que los datos están agrupados, debemos calcular el promedio en datos agrupados, por tanto, el promedio es $\bar{x} = \frac{6,5 \cdot 1 + 7,5 \cdot 2 + 8,5 \cdot 4 + 9,5 \cdot 7 + 10,5 \cdot 5}{19} = 9,18$ horas. Donde 6,5, 7,5, 8,5, 9,5 y 10,5 representan los puntos medios de cada intervalo.

b) El 50% inferior de los tiempos observados es de 9 horas.

b) **FALSO.** El 50% inferior de los tiempos observados se refiere a que debemos calcular la mediana. Por tanto, debemos calcular la mediana cuando los datos están agrupados:

En primer lugar, se divide $n/2$ para obtener la posición de la mediana. De esta manera identificamos el intervalo donde se encuentra la mediana. Se determina el límite inferior del intervalo, la frecuencia acumulada del intervalo anterior y la frecuencia absoluta del intervalo actual. Todos estos datos los aplicaremos en la siguiente fórmula:

$$Me = Li + \left(\frac{\frac{n}{2} - Fa(ant)}{fa} \right) \cdot A$$

donde Li = límite inferior del intervalo que contiene a la mediana, $Fa(ant)$ es la frecuencia acumulada del intervalo anterior, fa es la frecuencia absoluta del intervalo actual y A es el ancho del intervalo.

Entonces,

$$Pos(Me) = \frac{n}{2} = \frac{19}{2} = 9,5$$

El intervalo que contiene a la mediana es [9;10).

Aplicamos la fórmula y nos queda:

$$Me = Li + \left(\frac{\frac{n}{2} - Fa(ant)}{fa} \right) \cdot A = 9 + \left(\frac{\frac{19}{2} - 7}{4} \right) \cdot 1 = 9 + \left(\frac{9,5 - 7}{4} \right) = 9 + \left(\frac{2,5}{4} \right) = 9 + 0,625 = 9,625 \text{ horas.}$$

c) Podríamos decir que la distribución de los tiempos necesarios que requiere en cada niño para lograr un plan integral del tratamiento es aproximadamente simétrica.

c) **FALSO.** Observando el gráfico, podemos decir que es asimétrica a la izquierda.

ii) En el colegio X de Buenos Aires, se está realizando un experimento para comprobar si realmente la enseñanza de las matemáticas por computadora supera la enseñanza tradicional, en cuanto a resultados inmediatos de los niños. Dicho experimento se realizó con los estudiantes de cuarto grado de primaria y ocurrió así: se tomó un grupo control formado por 17 de esos niños, elegidos de su curso aleatoriamente, y se les enseñó un cierto tema a la manera tradicional. También se tomó un grupo experimental formado por otros 17 niños, elegidos del mismo curso en forma aleatoria y se les enseñó el mismo tema con asistencia de la computadora. Cuando terminaron de ver ese tema, se les hizo a los 34 alumnos el mismo examen. Las calificaciones obtenidas por los niños se presentan a continuación:

Grupo control 36 35 36 41 41 36 32 29 28 40 33 40 40 38 40 30 41

Grupo Experimental 40 38 38 44 44 40 36 39 38 40 38 44 44 38 40 38 40

a. ¿Cuál es la variable en estudio?

a. La variable en estudio es: **CALIFICACIONES OBTENIDAS POR LOS NIÑOS.**

b. ¿Cuál de las medidas de tendencia central podría describir el comportamiento de la variable en cada uno de los dos grupos? ¿Sería representativa? Justifique.

c. La medida de tendencia central que podría describir el comportamiento de la variable en cada uno de los dos grupos sería la **MEDIA**. Sería representativa pues **NO** hay valores alejados.

d. ¿Cuál de las dos muestras es más dispersa?

d. Para responder a esta pregunta, debemos ver el **COEFICIENTE DE VARIACIÓN** de cada grupo:

$$CV_{GC} = \frac{S_{GC}}{\bar{X}_{GC}} \cdot 100\% = \frac{4,44}{36,29} \cdot 100\% = 0,1223 \cdot 100\% = 12,23\%$$

$$CV_{GE} = \frac{S_{GE}}{\bar{X}_{GE}} \cdot 100\% = \frac{2,56}{39,94} \cdot 100\% = 0,0640 \cdot 100\% = 6,40\%$$

Como en la muestra de Grupo Control, el coeficiente de variación es más grande que el del grupo experimental significa que la muestra del Grupo Control es más dispersa.

d. ¿Cree usted que las calificaciones obtenidas en los dos grupos corroboran la hipótesis de que la enseñanza por computadora supera en resultados inmediatos a la enseñanza tradicional? Explique su respuesta.

Si, corroboran la hipótesis de que la enseñanza por computadora supera en resultados inmediatos a la enseñanza tradicional ya que como hemos observado, el promedio de calificaciones del Grupo Experimental es más alto que la media de calificaciones del Grupo Control, además ayuda el hecho de que el coeficiente de variación del Grupo Experimental sea mucho más bajo (la mitad aproximadamente) que el Grupo Control, es decir, los datos de la muestra del Grupo Experimental están menos dispersos que del grupo Control y por tanto están más concentrados los datos alrededor de la media 39,94.

Actividad 1: Se han medido los niveles de ozono alrededor de Madrid. Las concentraciones de magnitud alrededor de 220 unidades por billón pueden ocasionar quemaduras en los ojos y son peligrosas para la vida animal. Se han obtenido los siguientes datos:

160 176 160 180 167 164 165 163 162 168 173 179 170 196 185 163 162 163 172 162
167 161 169 178 161

- a) Diseñar un gráfico de tallos y hojas para estos datos, especificando la unidad correspondiente.

a) A continuación, diseñamos un gráfico de tallos y hojas:

16		0	0	1	1	2	2	2	3	3	3	4	5	7	7	8	9
17		0	2	3	6	8	9										
18		0	5														

Unidad: Billones

- b) Realizar una tabla de distribución de frecuencias y redactar un breve informe a partir de ella y utilizando también el diagrama realizado en el ítem anterior.

b) A continuación, hacemos la tabla de distribución de frecuencias:

Niveles de ozono (Variable)	fa	fr	fp	Fa	Fr	Fp
[160;167,3)	14	0,56	56	14	0,56	56
[167,3;174,6)	5	0,2	20	19	0,76	76
[174,6;181,9)	4	0,16	16	23	0,92	92
[181,9;189,2)	1	0,04	4	24	0,96	96
[189,2;196,5)	1	0,04	4	25	1	100
TOTAL	25	1	100			

Cantidad de intervalos: $k=5$

Amplitud del intervalo: $A=7,2$

Amplitud conveniente: $A=7,3$

INFORME:

Se tomó una muestra de 25 medidas de los niveles de ozono alrededor de Madrid y se obtuvo lo siguiente:

- La mayoría de los niveles de ozono (más del 50%) alrededor de Madrid es de entre 160 y 167 unidades por billón.
- Hay muchos lugares con poco nivel de ozono y pocos lugares con muchos niveles de ozono.
- No hay valores atípicos o alejados.
- Hay una asimetría a la derecha.
- El diagrama de tallo y hojas es unimodal pues tiene un solo pico.
- El centro se localiza entre 160 y 169 unidades por billón.
- La mínima cantidad de ozono es 160 y la máxima 196 unidades por billón.

El diagrama de tallo y hojas muestra de forma más adecuada la información ya que en la tabla de frecuencias agrupadas, los datos originales se pierden en el proceso de agrupamiento y, además, hay que realizar un gráfico para analizar la información, en cambio, en el diagrama de tallo y hojas, ya se puede analizar el centro, la forma, los valores atípicos, etc.

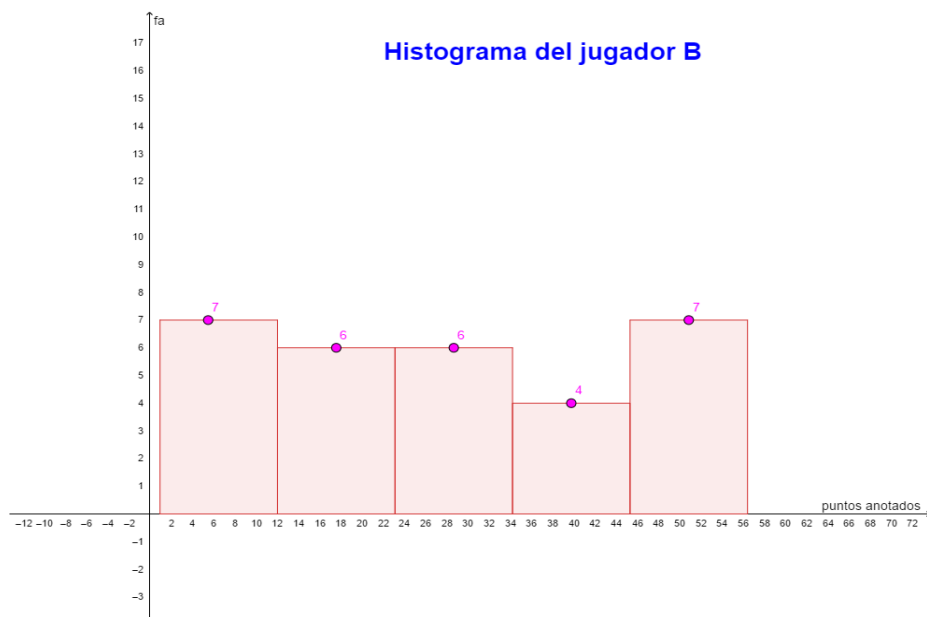
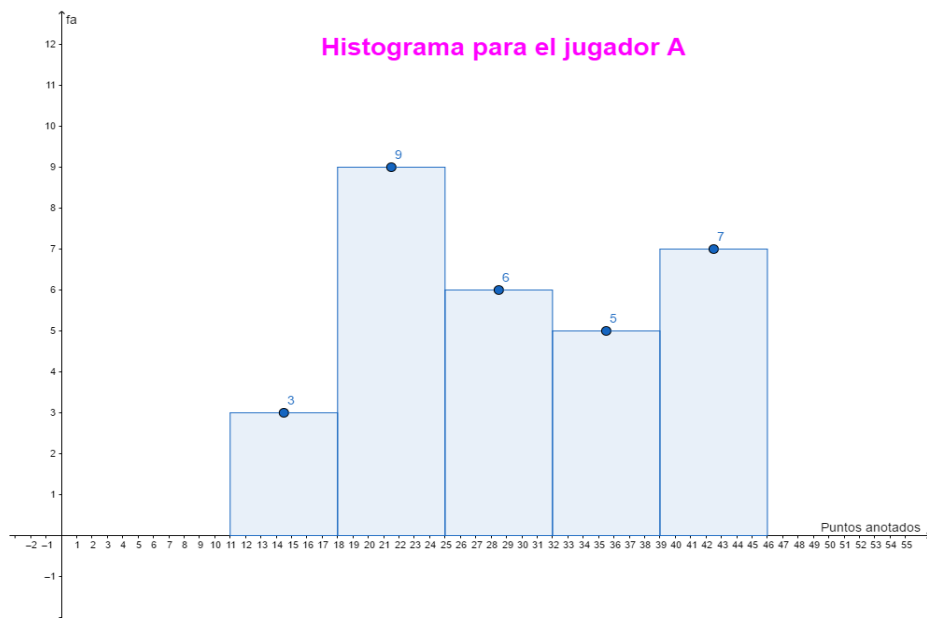
Actividad 2: Los datos que figuran a continuación corresponden a los puntos anotados por dos jugadores de baloncesto a lo largo de 30 partidos.

Jugador A: 34, 11, 19, 13, 20, 20, 33, 43, 22, 30, 25, 17, 28, 21, 22, 40, 40, 27, 41, 39, 39, 31, 18, 27, 35, 33, 35, 21, 19, 44.

Jugador B: 37, 1, 10, 2, 12, 13, 34, 54, 16, 31, 21, 6, 28, 15, 16, 48, 48, 25, 51, 47, 47, 32, 8, 25, 38, 35, 38, 13, 11, 56.

a) Realizar el gráfico que creas conveniente para comparar ambos grupos.

a) Como las muestras son grandes (tienen 30 valores cada uno) nos conviene agrupar los datos y hacer dos histogramas o polígonos de frecuencias. En este caso, hacemos dos histogramas:



b) Realizar un diagrama de tallo y hojas comparativo y destacar las características más sobresalientes de ambos grupos a partir de él.

b) A continuación, hacemos un diagrama de tallo y hojas comparativo:



Podemos observar que la mayoría de los puntos anotados por el jugador A está entre 11 y 28 mientras que la mayoría de los puntos anotados por el jugador B está entre 1 y 28.

El jugador A, el punto mínimo que obtuvo fue 11 mientras que el jugador B el mínimo es 1.

El punto máximo que anotó el jugador A fue 44 y el máximo de jugador B fue 56.

El diagrama de tallo y hojas para el jugador A es aproximadamente simétrica mientras que el diagrama de tallo y hojas para el jugador B es asimétrica a la derecha.

Ninguna de las dos muestras tiene valores atípicos o alejados.

Los dos diagramas de tallos son unimodales, es decir, tienen un solo pico.

El centro del diagrama para el jugador A se localiza entre 20 y 28 puntos mientras que el centro del diagrama para el jugador B se encuentra entre 10 y 16 puntos.

Se puede observar que la moda para la muestra del jugador A es 19, 20, 21, 22, 27, 33, 35, 39 y 40 mientras que la moda para la muestra del jugador B es 13, 16, 25, 38, 47 y 48.

Además, la mediana para la muestra del jugador A es 27,5 puntos, en cambio, la mediana para la muestra del jugador B es 26,5 puntos.

La media para la muestra del jugador A es 28,23 puntos, por otro lado, la media para la muestra del jugador B es 27,3 puntos.

Por último, se puede observar que en el diagrama de tallo y hojas para el jugador A, los datos están más concentrados en el centro mientras que en el diagrama para el jugador B los datos están más dispersos. Esto significa que la muestra del jugador B es más variable que la del jugador A.