Medidas de Dispersión

Revisamos la tarea de la clase pasada
Distribución de Frecuencias de las distancias alcanzadas por las pelotas de golf nuevas:

Dato	Frecuencia	Dato	Frecuencia	Dato	Frecuencia	Dato	Frecuencia
223.7	1	239.9	1	256.3	1	269.6	1
224.4	1	243.6	1	256.5	1	271.4	1
226.9	1	247.2	1	258.8	1	278.7	1
232.3	1	248.3	1	260.4	1	294.1	1
232.7	1	249.2	1	264.3	1		
233.5	1	252.8	1	265.1	1	Total	25
237.4	1	253.6	1	267.5	1		

Moda: no hay moda

Mediana: $$n = \frac{N+1}{2} = \frac{25+1}{2} = 13 \qquad Mediana = 252.8$$

La media es: $$\bar{x} = \frac{\sum x_i}{N} = \frac{5831}{25} = 251.68$$

Ahora vamos a agrupar estos datos para ver si nos proporcionan más información:

1° Rango $R = 294.1 - 223.7 = 70.4$

2° Número de intervalos de clase: $n_i = \sqrt{25} = 5$ (como es impar así la dejamos)

3° Ancho de los intervalos de clase: $i = \frac{70.4}{5} = 14.08 \approx 14.1$ (como el orden de los datos es hasta el primer decimal, así lo dejamos).

4° Con esto vamos a determinar los intervalos de clase, la primera columna de nuestra tabla de distribución de frecuencias con datos agrupados, vamos a poner de una vez, las columnas que sabemos necesitamos para construir el diagrama de pastel, el histograma polígono de frecuencias y la ojiva de frecuencias relativas acumuladas:

Tabla de Distribución de Frecuencias de las distancias alcanzadas por las pelotas de golf nuevas

Intervalos de clase (en cm.)	Marca de clase x	Frecuencia Alumnos F	Frecuencias Acumuladas Fa	$Frel$	$Frel$ $acum$
223.65 – 237.75	**230.7**	**7**	**7**	**0.28**	**0.28**
237.75 – 251.85	**244.8**	**5**	**12**	**0.20**	**0.48**
251.85 – 265.95	**258.9**	**8**	**20**	**0.32**	**0.80**
265.95 – 280.05	**273.0**	**4**	**24**	**0.16**	**0.96**
280.05 – 294.15	**287.1**	**1**	**25**	**0.04**	**1**
Total		***N = 25***			

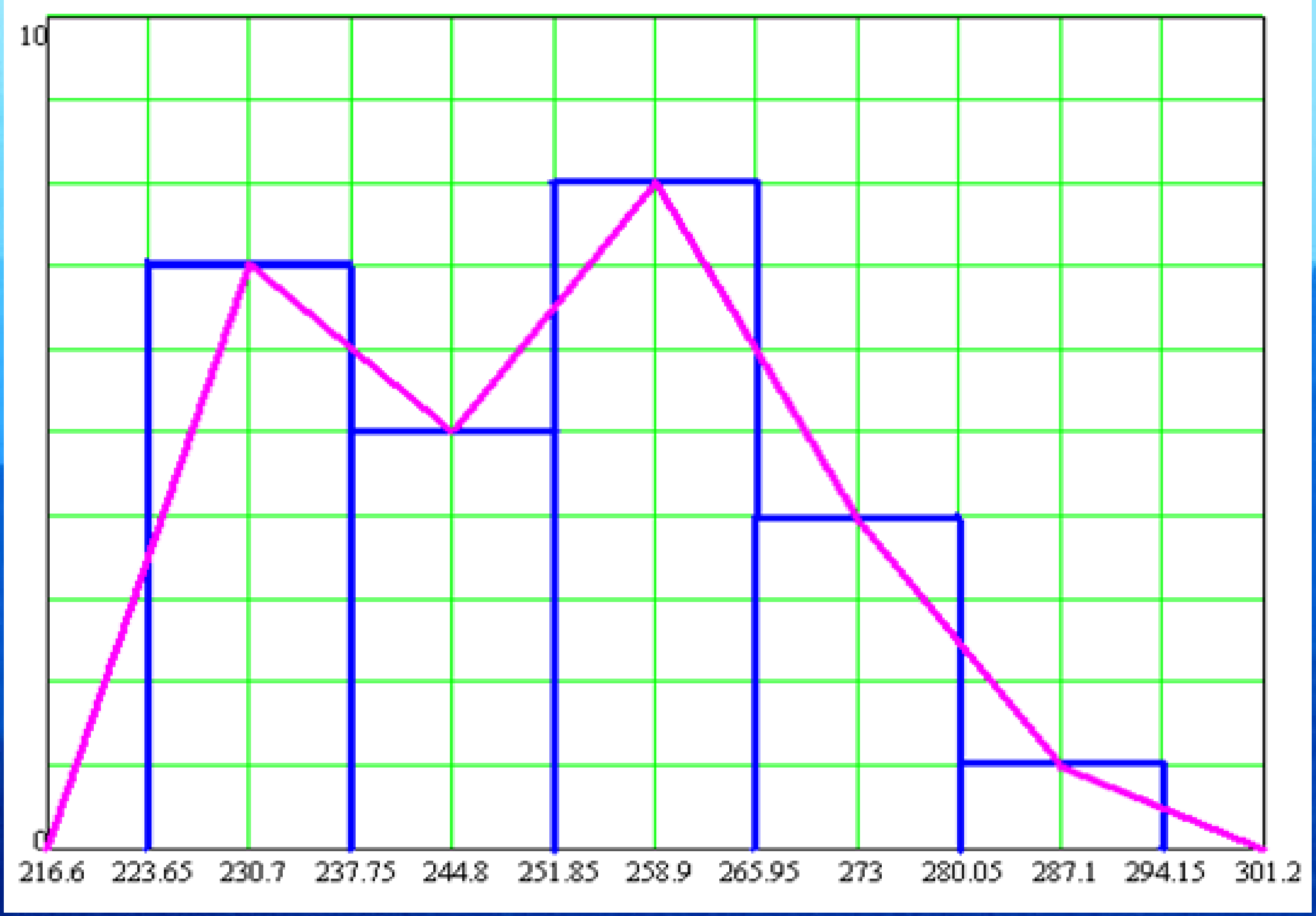
Medidas de Tendencia Central

Moda: ***258.9***

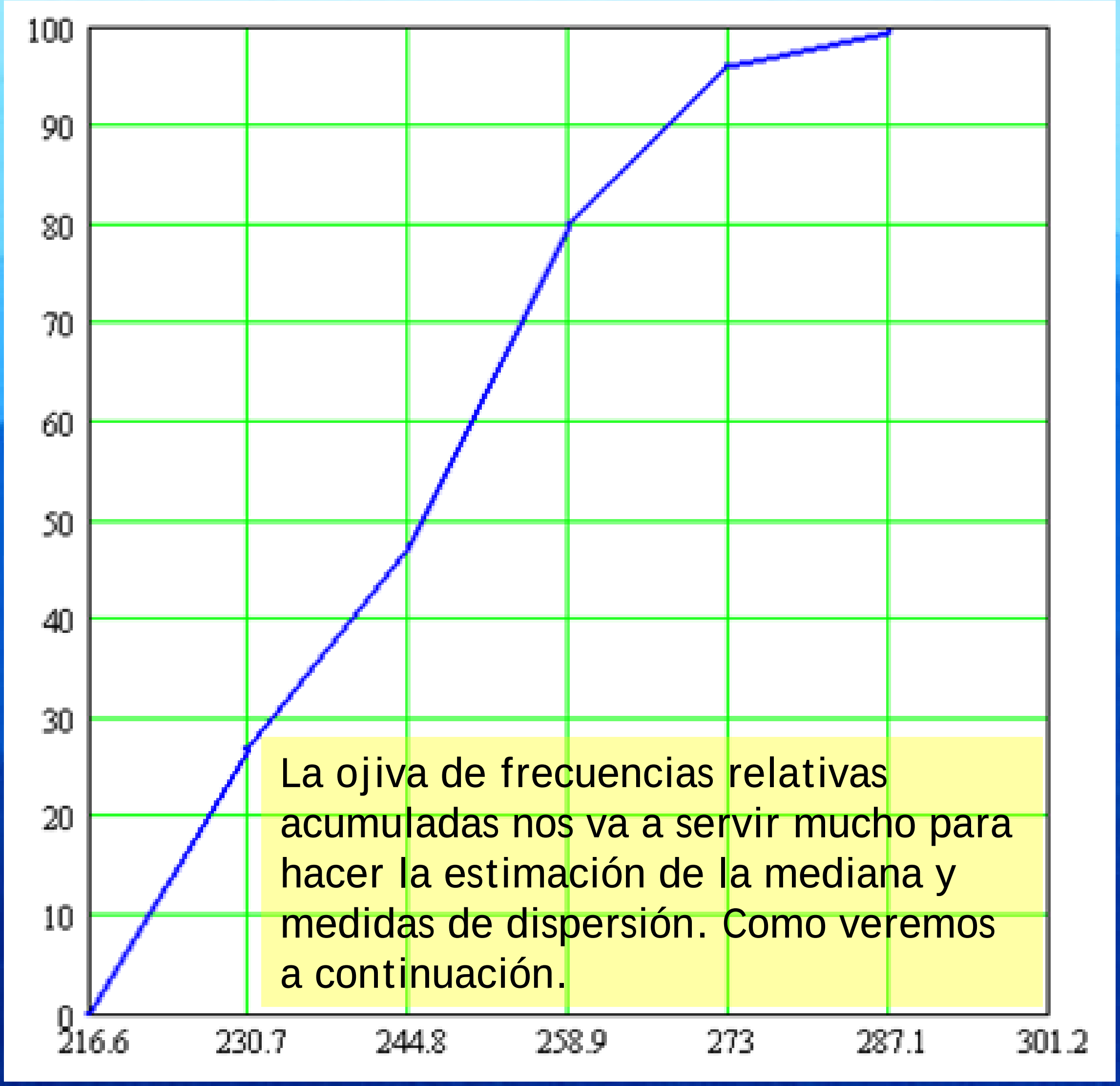
Media: $\bar{x} = 251.68$

Mediana: después aprenderemos a estimarla y a calcularla

Histograma y polígono de frecuencias de las distancias alcanzadas por las pelotas de golf nuevas



Ojiva de frecuencias relativas acumuladas de las distancias alcanzadas por las pelotas de golf nuevas



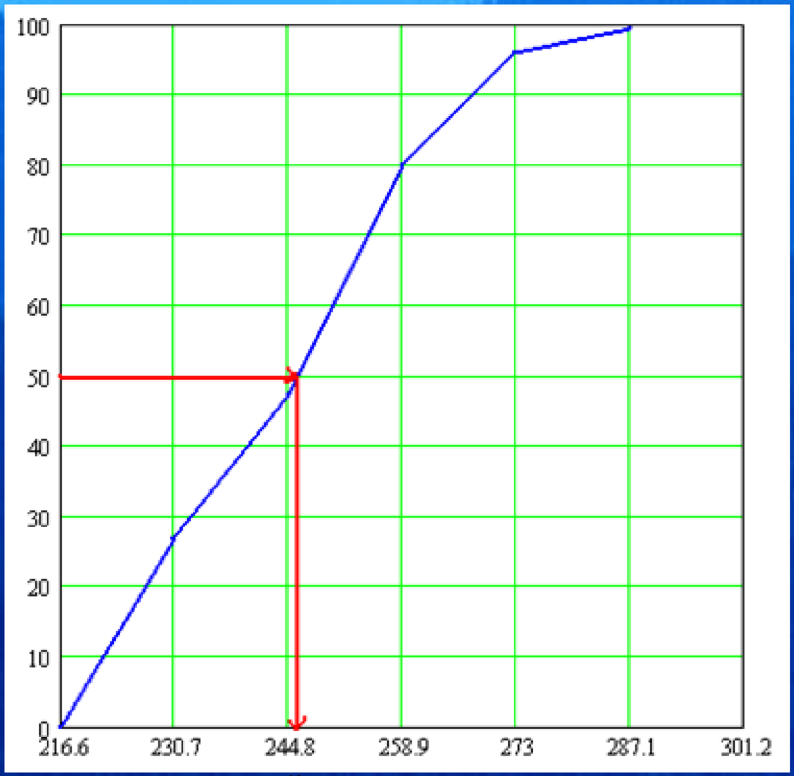
La ojiva de frecuencias relativas acumuladas nos va a servir mucho para hacer la estimación de la mediana y medidas de dispersión. Como veremos a continuación.

Mediana

Para estimar la mediana en la ojiva de frecuencias relativas acumuladas, si tenemos en el eje *Y* los porcentajes, localizamos el 0.5; si tenemos en el eje Y datos en porciento, localizamos el 50%. De ahí vamos hacia la curva horizontalmente y donde cruzamos la curva nos bajamos hacia el eje X y anotamos el valor correspondiente, ese es el valor estimado de la Mediana

Ojiva de frecuencias relativas acumuladas de las distancias alcanzadas por las pelotas de golf nuevas

Mediana 245



Mediana

Medidas de dispersión

Ahora vamos a ver las medidas de dispersión las cuales nos dicen de cómo están diseminados nuestros datos en la distribución. Para esto estudiaremos los conceptos que a continuación vamos a definir.

Rango

Ya lo habíamos visto, es la diferencia entre el valor máximo y el mínimo en nuestros datos. Esta medida de dispersión aunque es la más fácil de obtener, es muy poco usada ya que no nos dice poco acerca de la mayoría de los datos.

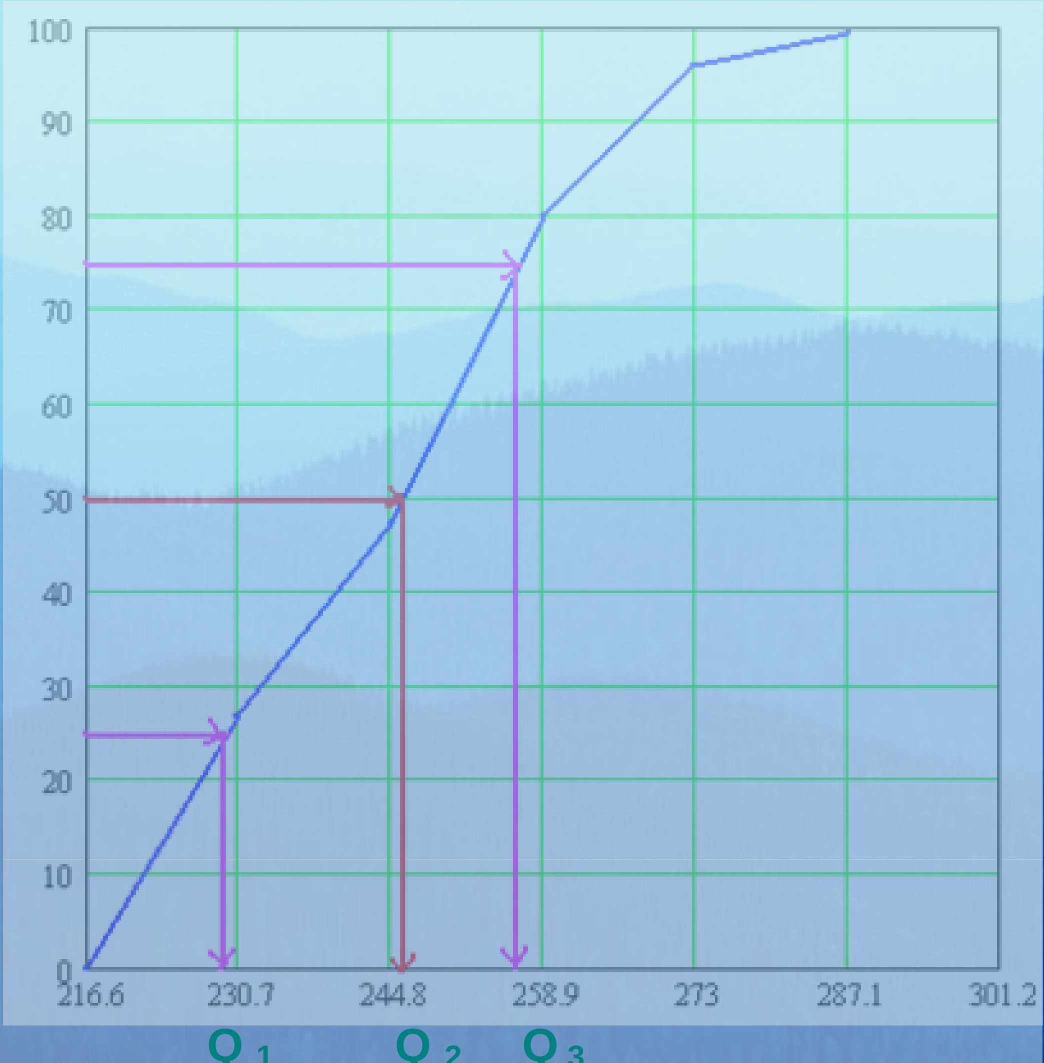
Cuartiles y Deciles

Estas medidas de dispersión se parecen mucho a la mediana en cuanto a que dividen a la distribución en partes iguales y se encuentra el valor que corresponde, los cuarteles la dividen en cuatro y los deciles en diez.

Cuartiles

Al dividir a la distribución en cuatro partes iguales, los cuartiles contendrán entre uno y otro al 25% del total de datos. Al primer cuartil se le denota Q_1 y separa al primer 25% del total de datos; el segundo cuartil, Q_2 , separa al primer 50% de los datos, (por lo que coincide con la mediana; el tercer cuartil , Q_3 , separa al 75% de los datos.

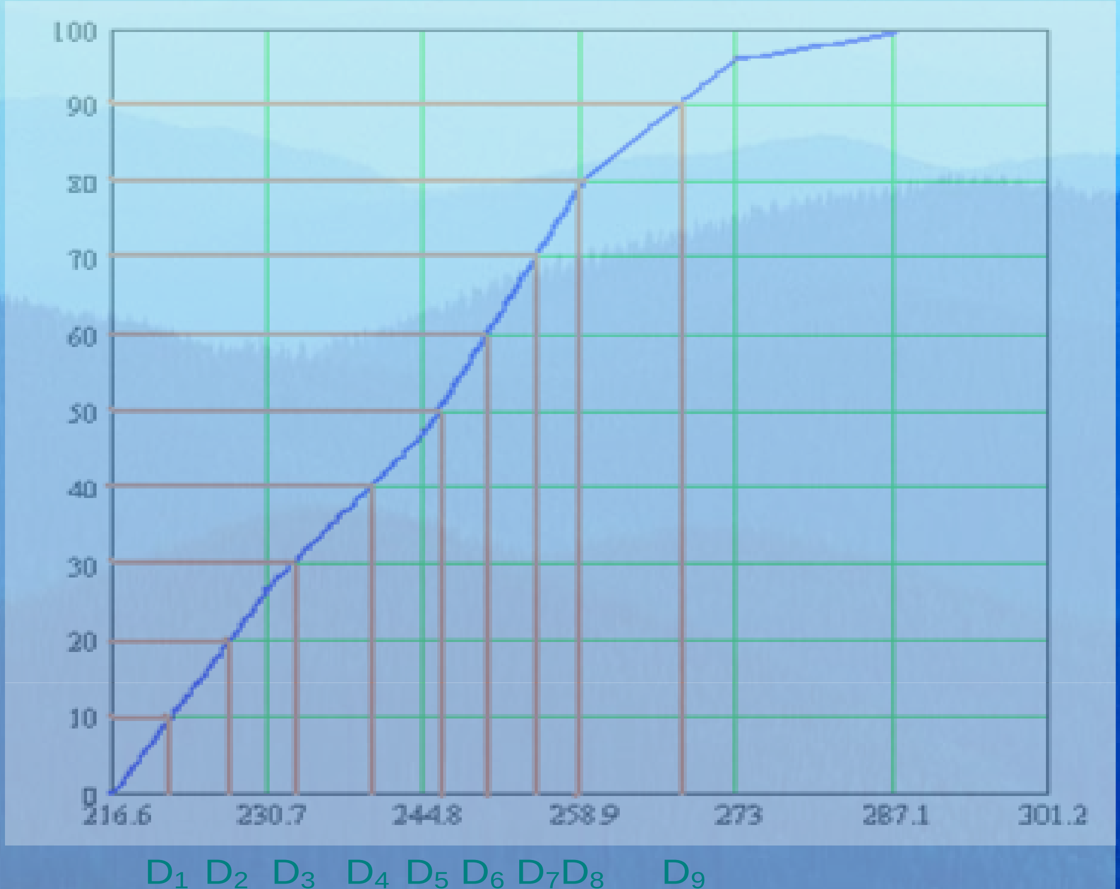
- el cuartil 1 (Q_1) divide a la población en 25% - 75%.
- el cuartil 2 (Q_2) divide a la población en 50% - 50%. Por lo que es igual a la Mediana
- el cuartil 3 (Q_3) divide a la población en 75% - 25%.



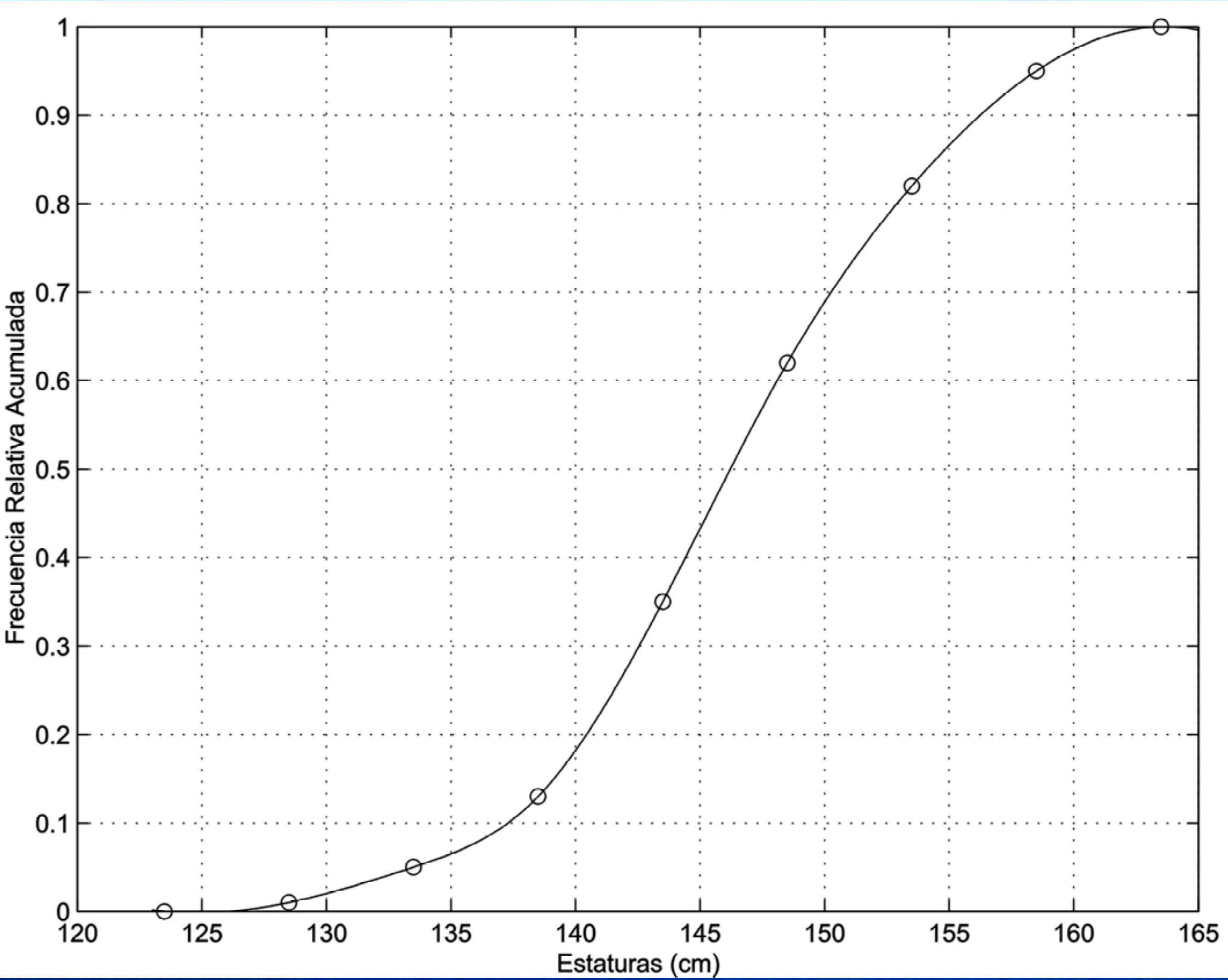
Los deciles

Son muy similares a los cuartiles pero dividen a la distribución en diez partes iguales:

- el decil 1 (D_1)
- el decil 2 (D_2)
- el decil 3 (D_3)
- el decil 4 (D_4)
- el decil 5 (D_5)
- el decil 6 (D_6)
- el decil 7 (D_7)
- el decil 8 (D_8)
- el decil 9 (D_9)



Actividad 1 Marcar los cuartiles y los deciles en la ojiva de frecuencias relativas de las estaturas



Contestar lo siguiente:

a)¿Cuál es la mediana de esta distribución de frecuencias?

b)¿Cuál es la estatura del primer cuartil?

c)¿Cuál es la estatura del tercer cuartil?

d)¿Cuál es el porcentaje de las alumnas que miden entre 1.50 y 1.60 metros?

e)¿Cuál es la estatura máxima?

f)¿Qué porcentaje mide menos de 1.40 metros?

g)¿Qué porcentaje mide más de 1.55 metros?

Continuamos con las medidas de dispersión.
Ahora vamos a ver una de las más usadas (tal vez por ser de las más útiles).

Desviación media absoluta

En datos no agrupados, se llama desviación media al promedio del valor absoluto de las diferencias entre cada dato y la media, o sea el promedio de las desviaciones de la media en valor absoluto.

$$DM = \frac{\sum_{i=1}^{N} |x_i - \mu|}{N}$$

donde:

DM = desviación media
N = número total de datos
x_i = dato i
μ = media de la distribución de la población

Actividad 2 Encontrar la desviación media de la población siguiente:

10, 12, 2, 9, 15, 6, 7, 8, 12, 9

$$\mu = \frac{10+12+2+9+15+6+7+8+12+9}{10} = 9$$

$$DM = \frac{\sum_{i=1}^{10} |10-9| + |12-9| + |2-9| + |9-9| + |15-9| + |6-9| + |7-9| + |8-9| + |12-9| + |9-9|}{10}$$

$$DM = 2.6$$

Varianza

En datos no agrupados, la varianza es la sumatoria del cuadrado de las desviaciones respecto a la media entre el número total de datos

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$$

¿Porqué al cuadrado ?

σ^2 = varianza de la población

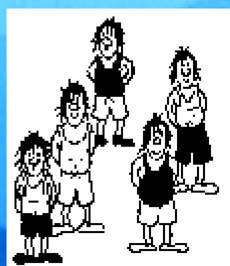
x_i = dato i

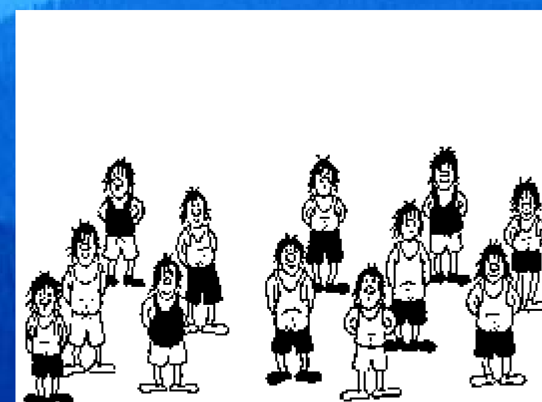
N = número total de datos de la población

μ = media de la población

Nota: Se puede calcular la varianza de la población o de una muestra, a ésta última se denomina generalmente con la letra S.

Cabe mencionar que con lo que hemos estado trabajando es con la idea de que a pesar de que puede existir una población de un tamaño específico (generalmente grande), lo que tenemos a la mano es **una parte de dicha población**, o sea, **una muestra**.





Desviación Estándar

Se llama desviación típica o estándar a la raíz de la varianza, tanto para datos de la población o para una muestra:

Esta medida de dispersión es de las más usadas porque permite comparar mejor la dispersión de los datos en poblaciones diferentes.

$$\sigma = \sqrt{\sigma^2} \qquad s = \sqrt{s^2}$$

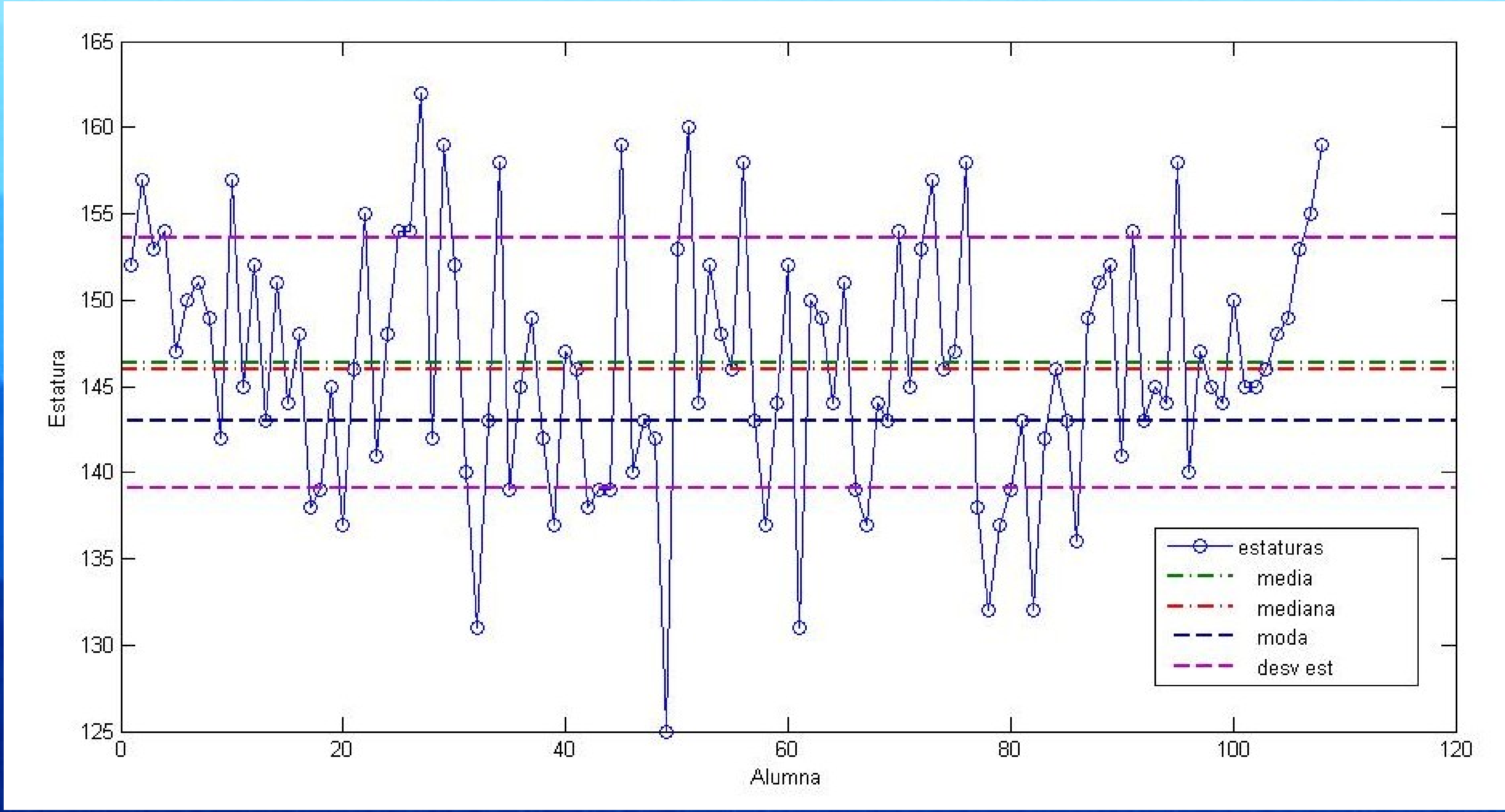
Actividad 3 Encontrar la varianza de la población del ejemplo anterior:

10, 12, 2, 9, 15, 6, 7, 8, 12, 9

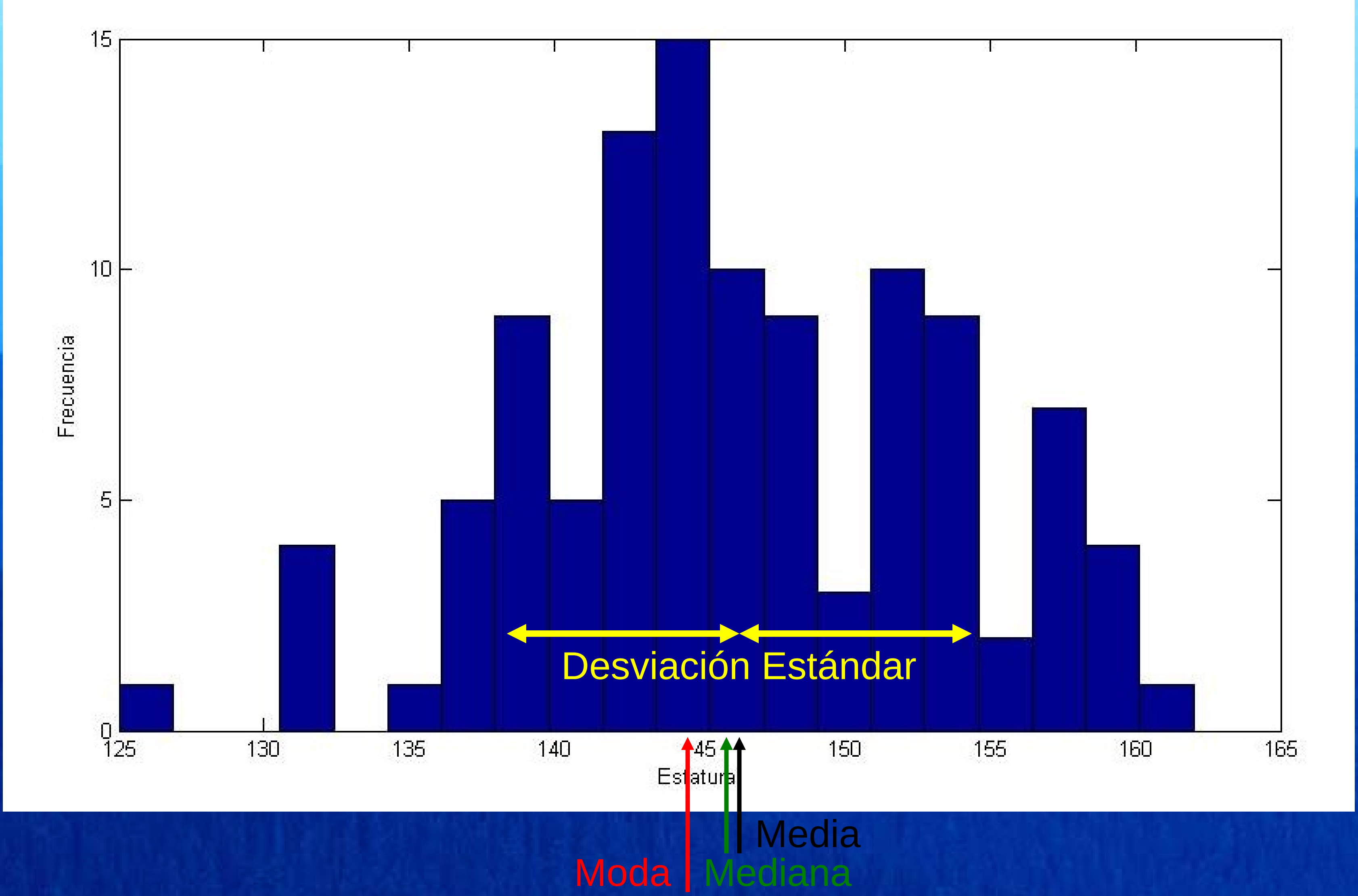
$$\sigma^2 = \frac{\sum_{i=1}^{10} (10-9)^2 + (12-9)^2 + (2-9)^2 + (9-9)^2 + (15-9)^2 + (6-9)^2 + (7-9)^2 + (8-9)^2 + (12-9)^2 + (9-9)^2}{10}$$

$$\sigma^2 = 11.8$$

Veamos qué significan todas estas medidas tanto de tendencia central como de dispersión usando los datos de las estaturas de alumnas de secundaria. En este caso sólo graficamos cada estatura una por una.



Si graficamos el histograma con 20 intervalos y ponemos las medidas de tendencia central y dispersión ¿Cómo nos queda?.



Varianza en datos agrupados

En datos agrupados de una muestra, la varianza es la sumatoria del cuadrado de las desviaciones respecto a la media de la muestra entre el número total de datos de la muestra menos 1.

$$s^2 = \frac{\sum_{i=1}^{t} f\left(mc_i - \bar{x}\right)^2}{n-1}$$

donde

s^2 = varianza de una muestra

t = número total de intervalos de clase

f = frecuencia de la marca de clase

mc_i = marca de clase del intervalo i

n = número total de datos de la muestra

$\bar{x}$ = media de la muestra

Varianza en datos agrupados de una población

En datos agrupados de una población, la varianza es la sumatoria del cuadrado de las desviaciones respecto a la media de la población entre el número total de datos de la población.

$$\sigma^2 = \frac{\sum_{i=1}^{t} f\left(mc_i - \mu\right)^2}{N}$$

donde

σ^2 = varianza de una población

t = número total de intervalos de clase

mc_i = dato i

N = número total de datos de la población

μ = media de la población

Actividad 4 Encontrar la varianza y la desviación estándar de los datos de la siguiente muestra:

Intervalo de clase	Marca de clase	Frecuencia años	$(mc_i - \overline{X})^2$	$f(mc_i - \overline{X})^2$
26.5 – 29.5	28	1	81	81
29.5 - 32.5	31	10	36	360
32.5 – 35.5	34	14	9	126
35.5 – 38.5	37	33	0	0
38.5 – 41.5	40	14	9	126
41.5 – 44.5	43	7	36	252
44.5 - 47.5	46	3	81	243
		N = 82		$\sum 1188$

$\overline{x} = 37$

Varianza: $s^2 = \dfrac{1188}{82-1} = 14.666$

Desviación estándar: $s = \sqrt{14.667} = 3.83$

Coeficiente de Variación

Se llama coeficiente de variación a la razón entre la desviación estándar y la media:

Para una población: $$CV = \frac{\sigma}{\mu} \times 100$$

Para una muestra: $$CV = \frac{s}{\bar{x}} \times 100$$

Actividad 5 Encontrar el coeficiente variación de la población anterior:

$$CV = \frac{3.83}{37} \times 100 \qquad CV = 10.35$$

Tarea 2. Calcular la Varianza, Desviación Estándar y Coeficiente de variación de los datos del ejercicio sobre las pelotas de golf, primero considerándolos como si fueran de la población completa y luego como si fueran de una muestra.