

Sesión 2: Medidas de Dispersión

Módulo 1: Estadística Descriptiva

Magdalena Cornejo

Medidas de Dispersión

Si nos limitáramos solamente a fijarnos en las medidas de tendencia central, no tendríamos una idea acabada de cómo se distribuyen los datos.

Medidas de Dispersión

Si nos limitáramos solamente a fijarnos en las medidas de tendencia central, no tendríamos una idea acabada de cómo se distribuyen los datos.

Imaginemos que queremos comparar la cantidad de llamadas recibidas en un call center por trimestre en 2011 y en 2012 y observamos lo siguiente:

Trimestre	2011	2012
Primer trimestre	1200	1070
Segundo trimestre	1500	2694
Tercer trimestre	1350	6
Cuarto trimestre	1220	1500
Media	1317.5	1317.5
Mediana	1285	1285

La media y la mediana son las mismas en 2011 y 2012,
¿se puede decir que los dos grupos son parecidos?

Medidas de Dispersión

Definición

Las **medidas de dispersión** son medidas estadísticas que muestran la variabilidad en la distribución de los datos.

Principales medidas:

- Varianza y desvío estándar
- Coeficiente de variación
- Rango y rango intercuartil

La varianza

- Dado que estamos interesados en analizar la dispersión de los valores, resulta natural fijarnos en sus diferencias respecto de la media.
- Es decir, cuánto se aleja cada observación respecto de la media.
- Como algunos de los valores son mayores que la media y otros menores, algunas de estas diferencias serán positivas y otras negativas, aunque estarán “equilibradas” (suman 0).
- Para analizar la dispersión de los datos no nos interesa el signo de las diferencias.
- Las trataremos de igual forma elevando las diferencias al cuadrado.
- El promedio de los cuadrados de las diferencias nos proporciona la medida de la **varianza**

Ejemplo

Retomando el ejemplo de los salarios de los analistas:

Analista	X	$X - \bar{X}$	$(X - \bar{X})^2$
1	35	$(35-33.7)=1.3$	1.69
2	31	$(31-33.7)=-2.7$	7.29
3	33	$(33-33.7)=-0.7$	0.49
4	36	$(36-33.7)=2.3$	5.29
5	34	$(34-33.7)=0.3$	0.09
6	34	$(34-33.7)=0.3$	0.09
7	35	$(35-33.7)=-0.7$	0.49
Suma	236	0	15.43

La **varianza** (S^2) resulta:

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1} = \frac{15.43}{6} = 2.6$$

Cuando $n < 30$, se divide por n en vez de $n - 1$.

El desvío estándar

- Si bien la varianza es una medida de dispersión, resulta difícil de interpretar. ¡Las unidades quedan elevadas al cuadrado!
- Una forma sencilla de volver a las unidades de medida originales es tomar la raíz cuadrada (positiva) de la varianza.
- Esto se conoce como el **desvío estándar** (S):

$$S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}} = \sqrt{\frac{15.43}{6}} = \sqrt{2.6} = 1.61$$

El coeficiente de variación

- Es una medida de dispersión que se utiliza fundamentalmente para comparar la variabilidad entre dos o más conjuntos de datos con distintas unidades de medida o distintas medias.
- El coeficiente de variación (CV) se define como:

$$CV = \frac{S}{\bar{X}}$$

- Se puede multiplicar por 100 para expresarlo en términos porcentuales.

Ejemplo

¿En qué caso hay más dispersión?

Semana	1	2	3	4	5	6	7	8	9	10	Media	Var.	Desvío
Temperatura (°C)	31	29	27	23	21	12	25	18	22	25	23.3	30.5	5.5
Cantidad de lluvia (mm)	12	11	6	6	6	14	6	3	3	2	6.9	16.8	4.1

El rango o recorrido

- Es quizás la medida de dispersión más simple y obvia.
- Es la diferencia entre el valor máximo y el mínimo de las observaciones.
- En el caso de los salarios de los analistas,

$$R = \text{max} - \text{min} = 36 - 31 = 5$$

- A pesar de su sencillez, el rango tiene el problema de que puede verse muy influido por la presencia de algún valor extremo.
- Además no tiene en cuenta la dispersión del resto de las observaciones que no están en ninguno de los dos extremos.

Cuartiles, Quintiles, Deciles o Percentiles

Ordenando los datos en forma ascendente (o descendente), se pueden calcular las siguientes medidas de posición:

- **Cuartiles:** son los 3 valores de la variable que dividen al conjunto de datos en 4 partes iguales (cada parte representa el 25%).
- **Quintiles:** son los 4 valores de la variable que dividen al conjunto de datos en 5 partes iguales (cada parte representa el 20%).
- **Deciles:** son los 9 valores de la variable que dividen al conjunto de datos en 10 partes iguales (cada parte representa el 10%).
- **Percentiles:** son los 99 valores de la variable que dividen al conjunto de datos en 100 partes iguales (cada parte representa el 1%).



Rango intercuartil

- Dado que el rango puede ser una medida insatisfactoria ya que compara el valor más alto con el más bajo,
- una medida alternativa que solventa (en parte) este problema es el **rango intercuartil**.
- El rango intercuartil descarta el 25% de las observaciones más altas y más bajas y halla el rango de las restantes.
- En el ejemplo de los salarios de los analistas:

31	33	33	34	34	35	36
Q1			Q2		Q3	

- La diferencia entre el tercer y primer cuartil nos da una medida de dispersión que llamamos **rango intercuartil**.

$$\text{Rango Intercuartil} = 35 - 33 = 2$$

Aplicación en Excel

- Bajar de **Yahoo Finance** (finance.yahoo.com) el precio de cierre ajustado de Intel e IBM en forma diaria entre el 4 de enero de 2016 y el 22 de julio de 2016.
- Compute el retorno diario de cada acción a través de la siguiente fórmula:

$$R_t = \frac{(P_t - P_{t-1})}{P_{t-1}} \times 100$$

- Utilice el complemento de Excel “Herramienta para análisis” y obtenga las principales medidas de estadística descriptiva de los retornos durante ese período.
- ¿Cuál de los dos retornos tiene una mayor dispersión?