

Gráficos

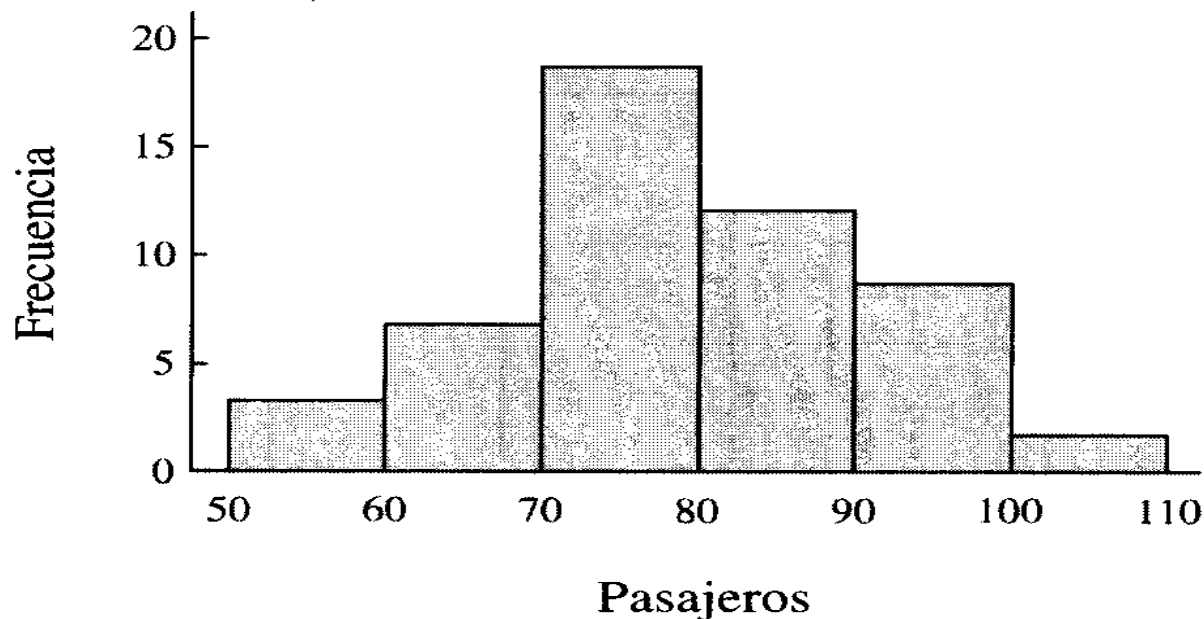
20

- Son muy útiles para describir los datos y “entenderlos” de manera rápida.
- Existen gráficos de distribución para variables categóricas como el *gráfico de barras*, y para proporciones relativas (*diagrama circular*).
- El *histograma* es la alternativa al gráfico de barras en variables continuas. Otros gráficos más avanzados (como el diagrama de caja o *boxplot*) se estudiarán más avanzado el ramo.

Gráficos(2)

21

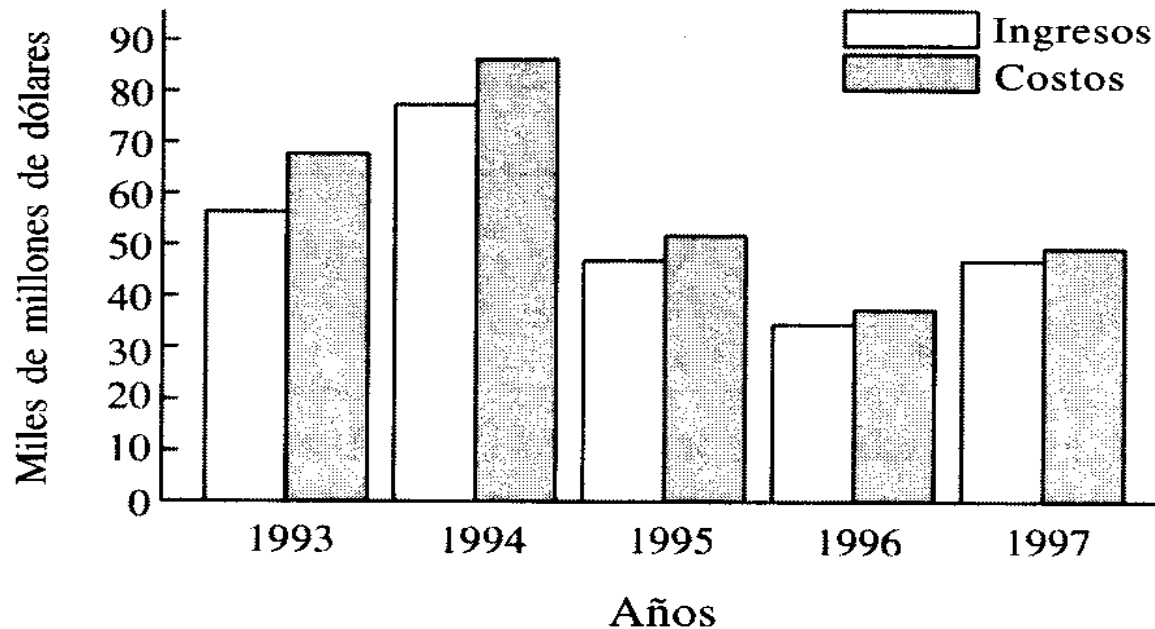
- El *histograma* ubica las clases de una distribución de frecuencia en el eje horizontal y las frecuencias en el eje vertical. Las frecuencias relativas se ilustran claramente:



Gráficos(3)

22

- El *gráfico de barras* en su versión más simple muestra categorías o valores numéricos (sin agrupar en clases) y cantidades de otra variable. Como ejemplo el desempeño de una empresa:

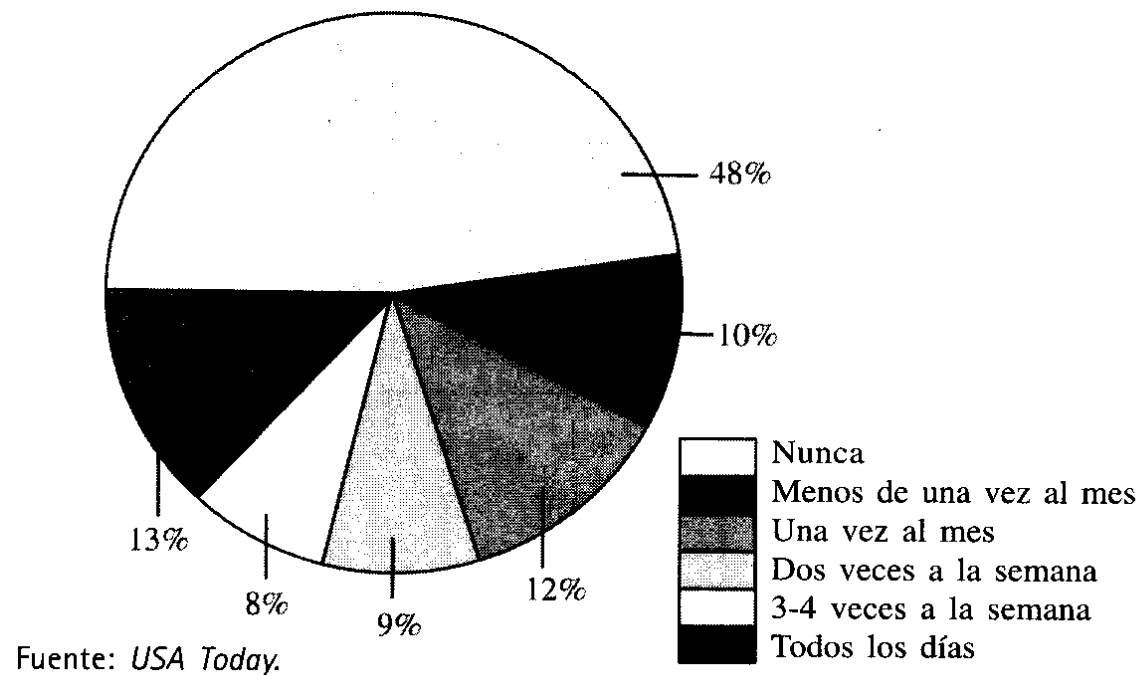


Gráficos(4)

23

- El *diagrama circular* es de particular utilidad para mostrar porcentajes de una variable, donde cada categoría se representa como una porción del círculo.

Con qué frecuencia los trabajadores llevan trabajo para hacer en la casa



Medidas de Tendencia Central y Dispersión

24

- Junto con tablas y gráficos, la estadística descriptiva incluye medidas de tendencia central y dispersión.
- Dentro de las medidas de tendencia central están la *media* o promedio, la *mediana*, la *moda*, el *promedio ponderado* y la *media geométrica*.
- Dentro de las medidas de dispersión se consideran el rango, la *varianza* y *desviación estándar* y los *percentiles*.

Medidas de Tendencia Central - Media

25

- Si se cuentan con n observaciones (muestra) de una variable X , la *media* (aritmética) de los valores observados es:

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

- El parámetro que define la media poblacional (promedio real de las N observaciones de una población) es:

$$\mu = \frac{\sum_{i=1}^N X_i}{N}$$

Medidas de Tendencia Central – Media(2)

26

Propiedades

- La suma de los cuadrados de las desviaciones de los ejemplos c/r a la media es menor que c/r a cualquier otro valor:

$$\sum_{i=1}^N (X_i - \bar{X})^2 < \sum_{i=1}^N (X_i - c)^2, c \neq \bar{X}$$

- Si se multiplican los ejemplos por una constante, la media quedará multiplicada por ésta:

$$Y_i = kX_i \rightarrow \bar{Y} = k\bar{X}$$

Medidas de Tendencia Central - Mediana

27

- La *mediana* es la **observación** de la mitad después de que se han colocado todos los elementos de manera ordenada.
- Si n es impar entonces la mediana es $\frac{n + 1}{2}$
- Si n es par hay dos enfoques: se consideran 2 medianas o éstas se promedian en una (por lo general se hace esto último).

Medidas de Tendencia Central – Mediana(2)

28

- Ej: se tienen los ingresos por ventas mensuales en miles de dólares para 5 meses: 56, 67, 52, 45, 67

- Media: $\bar{X} = \frac{56 + 67 + 52 + 45 + 67}{5} = 57.4$

- Mediana: primero se ordenan los valores : 45, 52, 56, 67, 67. La posición del valor de la mediana se vuelve:

$$\text{Posición de la mediana} = \frac{5 + 1}{2} = 3$$

- La mediana vale entonces 56

Medidas de Tendencia Central – Mediana(3)

29

- Ej caso par: si los ingresos para el 6to mes son 35, la serie ordenada se vuelve 35, 45, 52, 56, 67, 67.
- La nueva posición de la mediana es $(6+1)/2$
- Los dos valores de las posiciones 3ra y 4ta se promedian para producir una mediana de $52+56=54$.
- Esto significa que la mitad de los meses las ventas estuvieron por debajo de US\$54.000 y en la mitad de los meses los ingresos excedieron esta suma.

Medidas de Tendencia Central - Moda

30

- La *moda* es la observación que más que ocurre con mayor frecuencia.
- La moda puede ser única, pueden haber más de una moda o puede no calcularse cuando todos los valores tienen la misma frecuencia (*distribución amodal*)
- Cuando se dispone de una distribución de frecuencias, se toma como moda el punto medio del intervalo de mayor frecuencia.

Medidas de Tendencia Central – Media P.

31

- En ciertos casos se desea darle un mayor peso a algunas observaciones, por ejemplo, una prueba con mayor ponderación. La fórmula de la *media ponderada* es la siguiente:

$$\bar{X}_w = \frac{\sum_{i=1}^N w_i X_i}{\sum_{i=1}^N w_i}$$

donde w es el peso o ponderación asignada a cada observación

Medidas de Tendencia Central – Media G.

32

- La *media geométrica* proporciona una medida precisa de un cambio porcentual promedio de una serie de números:

$$MG = \sqrt[n]{X_1 X_2 \cdots X_n}$$

- Es de utilidad en series económicas y ventas. Ej: tasa de crecimiento promedio en los ingresos (para tomar decisiones como campañas publicitarias)

Medidas de Tendencia Central – Media G.(2)

33

Año	Ingreso	Porcentaje del año anterior
1992	US\$ 50,000	-- --
1993	55,000	55/50 = 1.10
1994	66,000	66/55 = 1.20
1995	60,000	60/66 = 0.91
1996	78,000	78/60 = 1.30

$$MG = \sqrt[4]{(1.10)(1.2)(0.91)(1.3)} = 1.1179$$

Medidas de Dispersión

34

- Miden que tanto se dispersan o desvían los datos en torno a la media.
- El *rango* es la medida de dispersión más simple (y menos útil). El rango es simplemente la diferencia entre la observación más alta y la más baja.
- La desventaja es que sólo considera dos de los (posiblemente) cientos de observaciones, ignorando el resto de los datos.

Medidas de Dispersión - Varianza

35

- La varianza es el promedio de las desviaciones de las observaciones con respecto a su media al cuadrado.

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

- La desviación estándar es la raíz de la varianza. Es una medida muy útil de dispersión ya que tiene las mismas unidades que la variable estudiada.

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

Medidas de Dispersión – Varianza(2)

36

- La *varianza muestral* sigue la misma lógica:

$$s^2 = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{n-1} \quad s = \sqrt{s^2}$$

- Llama la atención que se divida por $n-1$, lo que se debe a que este estadístico tiene $n-1$ *grados de libertad*.
- Los grados de libertad equivalen al número de observaciones menos el número de *restricciones* impuesta en tales observaciones.

Medidas de Dispersión – Varianza(3)

37

- Una restricción es cualquier valor que deba calcularse de dichas observaciones. En este caso la restricción es el cálculo de la media muestral.
- Ej: Se tienen $n=4$ obs. que producen una media de 10, la media de 10 sirve como restricción y hay $n-1=3$ g.l.
- Se pueden escoger 3 obs. Cualquiera, por ej. 8, 9 y 11.
- Después ya no hay libertad para escoger la última obs., que debe ser 12 si se quiere tener un promedio de 10.