**UNIVERSIDAD TECNOLÓGICA DE PEREIRA - FACULTAD CIENCIAS DE LA SALUD**

**PROGRAMA CIENCIAS DEL DEPORTE Y LA RECREACIÓN - SEMINARIOS DE INVESTIGACIÓN**

|  |
| --- |
| **1. ANÁLISIS DESCRIPTIVO DE LOS DATOS** |

**1. VARIABLES**

Una variable es una característica cambiante y cuantificable del objeto de estudio.

**ESCALAS DE MEDICIÓN DE LAS VARIABLES**

Grado de precisión con que se cuantifican las variables. Son las siguientes:

* **Cualitativas**: cuando la variable es una cualidad (letras), puede ser:
	+ **Nominal**: sus valores son nombres.
	+ **Ordinal**: sus valores son nombres que además tienen un orden.
* **Numéricas**: cuando la variable es un número, puede ser:
	+ Según la secuencia de los datos: **discreta** (sólo números enteros) y **continua** (números enteros y decimales).
	+ Según el valor del punto cero: **intervalar** (el 0 es convencional por lo que acepta valores negativos, por ejemplo, 0 grados centígrados es el punto de congelación del agua, no la ausencia absoluta de calor) y de **razón** (el 0 es absoluto, por ejemplo, talla, peso, etc.).

**2. ANÁLISIS DESCRIPTIVO**

**2.1 VARIABLES CUALITATIVAS**

* **Moda**: valor que más se repite (y más de 1 vez).
* **Porcentajes**: cantidad proporcional a cien.

**2.2 VARIABLES NUMÉRICAS**

**MEDIDAS DE TENDENCIA CENTRAL**

Estimación del punto medio alrededor del cual se agrupan los datos. Son diferentes según la escala de medición así:

* **Media**: es el promedio aritmético de las observaciones. Es la que más se utiliza porque: (1) es la más **representativa,** dado que tiene en cuenta los valores todas las observaciones, y (2) en una muestra, es la más **confiable** en la estimación de la tendencia central de la población. No se usa: cuando existen datos imprecisos que no se pueden sumar (como “mayor a 10”) o datos inusualmente extremos (por ejemplo, un adulto mayor en una población juvenil) con lo cual la media no indica adecuadamente la tendencia central; en estos casos se utiliza la mediana. Convencionalmente, las medidas de una población (parámetros) se escriben con letras griegas, por ejemplo: µ (que se lee “mi” y es la letra “m” minúscula en griego) para la media poblacional; y las de una muestra (estadísticos) se escriben con letras romanas, por ejemplo: $\overbar{x}$ (que se lee “x barra”) para la media muestral. Las fórmulas son:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *De una muestra:* |  | *De una población:* |  | *O sea, la media (*$\overbar{x}$ *o µ) es igual a la sumatoria (∑) de todos los datos individuales (x) dividida entre el tamaño de la muestra: (n) o de la población (N).* |
|  | $\overbar{x}$ = | ∑x |  |  | *µ =* | ∑x |  |  |
|  | n |  | N |  |

* **Mediana**: es el punto medio de una distribución (colocación de los datos en orden de menor a mayor).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Ejemplos:** |  |  |  |
|  | *Los datos:* | *Se ordenan:* | *Se identifica el puesto: (n+1)/2* | *Y se identifica la mediana:* |
|  | *4, 3, 5, 4, 6 🡪*  | *3, 4, 4, 5, 6* | *(5+1)/2 = 3 puesto* | *El tercero es: 4.**La mediana es:* ***4****.* |
|  | *7, 5, 9, 10 🡪* | *5, 7, 9, 10* | *(4+1)/2 = 2½ puesto* | *Segundo: 7. Tercero: 9. Mitad entre segundo y tercero: (7+9)/2 = 8. La mediana es* ***8****.* |

**MEDIDAS DE DISPERSIÓN**

Estimación el grado de separación de los datos. Las principales son las siguientes:

* **Desviación estándar:** es la raíz cuadrada de la **varianza**, esto es, de la “media de las desviaciones elevadas al cuadrado”; se utiliza conjuntamente con la media. No obstante, para obtener la desviación estándar de una muestra se usa el número total de datos menos 1, porque así aumenta la desviación estándar muestral (a más pequeña la muestra, más se incrementa), con lo cual se asemeja más a la desviación estándar de la población, ya que las muestras pequeñas tienden a no incluir los datos extremos de la población por su baja frecuencia (a más pequeñas, más aún), con lo cual tienden a mostrar una dispersión menor que la del universo.En fórmula:

|  |  |  |
| --- | --- | --- |
| s = | $$\sqrt{\frac{\sum\_{}^{}(x-\overbar{x})^{2}}{n-1}}$$ | *O sea, la desviación estándar de una muestra (s) es igual a la raíz cuadrada de: la suma (∑) de las desviaciones (cada dato individual menos la media: x-*$\overbar{x}$*) elevadas al cuadrado, dividida entre el número total de datos (n), menos 1.* |
| σ = | $$\sqrt{\frac{\sum\_{}^{}(x-μ)^{2}}{N}}$$ | *O sea, la desviación estándar de una población (*σ, letra griega “sigma” equivalente a la S mayúscula*) es igual a la raíz cuadrada de: la suma (∑) de las desviaciones (cada dato individual menos la media: x-*$\overbar{x}$*) elevadas al cuadrado, dividida entre el número total de datos (N).* |

**Ejemplo**: los datos 1, 2, 3, 4, 5. La media es: 1 + 2 + 3 + 4 + 5 = 15; 15 ÷ 5 = 3. La desviación estándar se obtiene así:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *Se obtiene la desviación de cada dato (se le resta la media):* |  | *Se elevan las desviaciones al cuadrado:* |  | *Se obtiene la media de las desviaciones elevadas al cuadrado (varianza):* |  | *Se obtiene la raíz cuadrada de la varianza:* |
| *1 – 3 = -2* *2 – 3 = -1**3 – 3 = 0**4 – 3 = 1**5 – 3 = 2* |  | *-22 = 4**-12 = 1* *02 = 0* *12 = 1* *22 = 4* |  | *4 + 1 + 0 + 1 + 4* | *= 2.5* |  | *s2 = 2.5**s = √2.5 = 1.58* |
| *5-1* |

* **Coeficiente de variación:** porcentaje de dispersión en relación con la media. Se calcula dividiendo la desviación estándar entre la media y multiplicando por 100. Permite comparar unidades diferentes. En fórmulas:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *De una muestra:* |  | *De una población:* |  | ***Ejemplos:*** |
|  | V = | s | x 100% |  | *V =* | σ | *x 100%* |  | V = | *0.1* | *x 100% = 0.5%* | V = | *0.006* | *x 100% = 0.8%* |
|  | $$\overbar{x}$$ | *µ* | 20.0 | 0.787 |

**Ejemplo:** con dos adipómetros de diferente marca, uno en milímetros y otro en pulgadas, se midió repetidamente el pliegue tricipital de un voluntario con los siguientes resultados: para un adipómetro µ = 20.0 mm y σ = 0.1 mm; y para el otro µ = 0.787" y σ = 0.006"; los coeficientes fueron: 0.5% y 0.8% respectivamente, por lo que se puede concluir que el primero es más preciso.

* **Rango intercuartil:** es la distancia entre el percentil 25 (primer cuartil, Q1) y el 75 (tercer cuartil, Q3), o sea, los puntos que separa los 25 y 75 valores menores de los mayores, respectivamente, de una distribución. Se utiliza conjuntamente con la mediana (que es igual al percentil 50 o segundo cuartil, Q2). El cuartil inferior es la mediana de los valores a la izquierda de la mediana general y el cuartil superior, de los valores a la derecha. La fórmula del rango intercuartil es: Q3 – Q1. **Ejemplos**:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | *Posición:* | *1º* | *2º* | *3º* | *4º* | *5º* | *6º* | *7º* | *8º* | *9º* | *10º* | *11º* | *12º* |  | *Rango intercuartil:* |
|  | **Datos:** | **72** | **74** | **75** | **77** | **78** | **79** | **82** | **85** | **86** | **90** | **93** | **94** |  |  |
|  | *Cuartil:* |  |  | *Q1  =* ***76*** |  | *Q2 =* ***80.5*** |  | *Q3 =* ***88*** |  |  |  | *88 – 76 =* ***12*** |
|  |
|  | *Posición:* | *1º* | *2º* | *3º* | *4º* | *5º* | *6º* | *7º* | *8º* | *9º* | *10º* | *11º* |  |  | **Rango intercuartil:** |
|  | **Datos:** | **72** | **74** | **75** | **78** | **79** | **82** | **85** | **86** | **90** | **93** | **94** |  |  |  |
|  | *Cuartil:* |  | *Q1  =* ***75*** | *Q2 =* ***82*** | *Q3 =* ***90*** |  |  |  | *90 – 75 =* ***15*** |

* **Rango:** es la distancia entre el número mayor y el menor; se utiliza para destacar los valores extremos. En el ejemplo anterior es: 94 – 72 = 22.

**FIGURAS**

|  |  |
| --- | --- |
| **Diagrama de barras** | **Histograma** |
| Figura 1. Tipos de transporte. Población estudiada, 2007. | Figura 2. Edad. Población estudiada, 2007. |
| * Con datos cualitativos o numéricos discretos.
* La altura de cada barra es proporcional a la frecuencia de cada respuesta.
 | * Con datos numéricos continuos (agrupados en intervalos).
* El área de cada barra es proporcional a la frecuencia de cada grupo y la etiqueta inferior es el punto central del intervalo.
 |
| Ambas se utilizan para: comparar las opciones de respuesta entre sí. |

|  |  |
| --- | --- |
| **Gráfica circular** | **Polígono de frecuencias** |
| Figura 1. Tipos de transporte. Población estudiada, 2007. | Figura 2. Edad según género. Población estudiada, 2007. |
| * El área de cada porción es proporcional al porcentaje de cada opción de respuesta.
* Se utiliza para: comparar cada opción de respuesta con el total.
 | * Se forma uniendo los puntos centrales superiores de cada barra de un histograma.
* Se utiliza para: comparar el comportamiento de una variable numérica con respecto a otra (cualitativa o numérica), por ejemplo, edad vs. género.
 |