

Universidad Católica del Uruguay
Facultad de Ciencias Humanas

Autotexto

Serie Estadística

Medidas de dispersión



Universidad
Católica

DAMASO A. LARRAÑAGA • URUGUAY

Laboratorio Metodológico

Versión original:
Claudia Rafaniello

Autores revisión 2008:
A. de León – M. Dodel – C. Rafaniello

Módulo de Práctica de Análisis

Serie Estadística

Tema: Análisis de datos, medidas de dispersión.

Descripción

En esta guía explicaremos los pasos que deben seguirse tanto en forma manual como en SPSS para calcular e interpretar los valores de las medidas de dispersión. Es en ese sentido que trabajaremos con 2 medidas de dispersión: el rango y el desvío estándar.

Marco Conceptual

Las medidas de dispersión, o de amplitud, para un número discreto de valores numéricos o medidas toma muchas formas en el análisis de datos. La medida de dispersión más simple es llamada *rango*, la cual se define como la medida mayor menos la medida menor.

Por ejemplo: para la distribución 5,8,8,14,15 y 10 el rango para los 6 valores es: 15 menos 5 es decir 10.

Mientras que el rango es afectado sólo por los dos valores extremos en el set de medidas, las otras medidas de dispersión revelan que se encuentran afectadas por todas las medidas o valores. La más común de estas otras medidas es llamada desviación estándar o desvío estándar. El valor del desvío estándar para un set de n medidas puede ser calculado por 1) primero calculando la media aritmética, 2) encontrando la diferencia entre la media y cada medida, 3) elevando al cuadrado cada una de esas diferencias, 4) sumando los cuadrados, 5) dividiendo la suma de los cuadrados entre n , y finalmente 6) calculando la raíz cuadrada del resultado de 5).

Atienda la siguiente demostración para el ejemplo dado:

x	$x-10$	$(x-10)^2$
5	-5	25
8	-2	4
8	-2	4
10	0	0
14	4	16
15	5	25
		74

$$\text{Desviación estándar: } \sqrt{\frac{74}{6}} \approx 3,5$$

El desvío estándar puede ser interpretado como la distancia promedio desde la media aritmética y las n medidas. El desvío estándar no puede ser negativo y

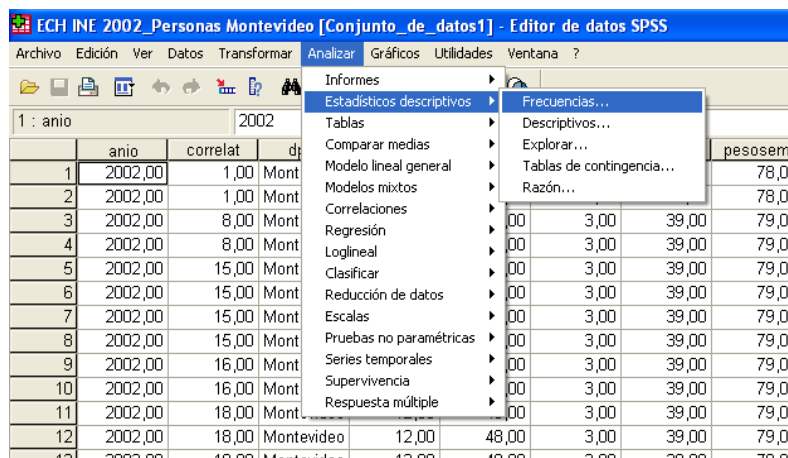
cuando dos set de medidas son comparadas, el que tiene la dispersión más grande tiene la desviación estándar más grande.

Procedimiento

El primer paso es abrir la base de datos de la Encuesta Continua de Hogares de Personas de Montevideo. Al igual que para el caso de las medidas de tendencia central, las medidas de dispersión pueden encontrarse siguiendo el siguiente procedimiento:

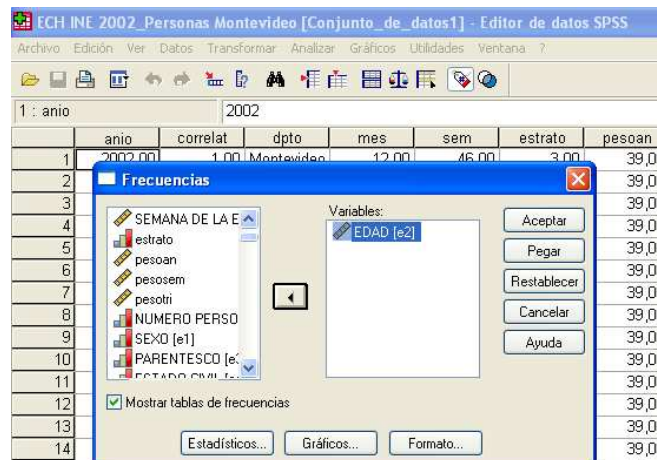
- **Pasos / Ventanas**

Se selecciona la distribución de frecuencia, la cual se ubica, como muestra la imagen, en *Análisis, Estadísticos descriptivos* y luego *frecuencia*.

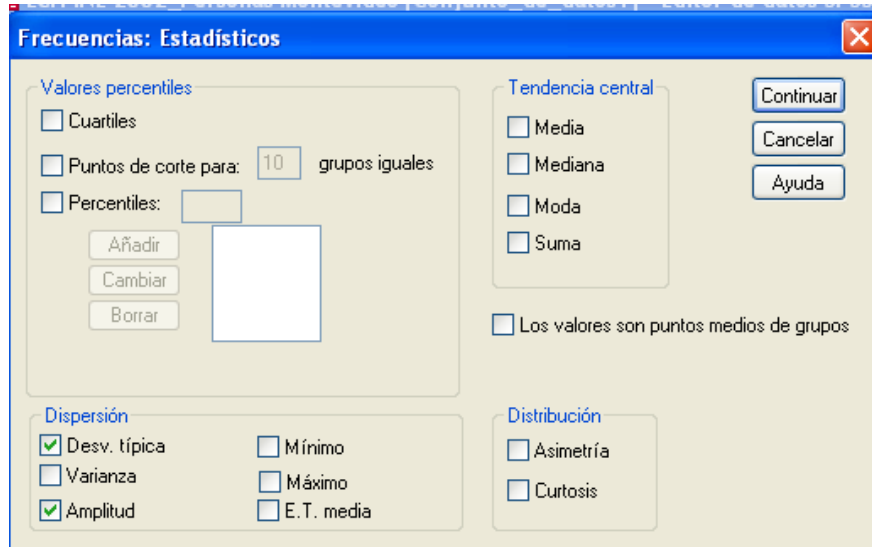


Un vez seleccionada la distribución de frecuencia, se despliega la siguiente ventana, en la que hay que seleccionar la/s variable/s a las que uno va a querer aplicarle las medidas de dispersión, para luego presionar el botón "Estadísticos".

Supongamos en este caso que trabajaremos con la Edad:



En ese momento de despliega la siguiente ventana en la que encontramos, sobre el lado izquierdo de la ventana las medidas de dispersión. Selecciono a través del puntero del mouse cuál de todas ellas voy a querer utilizar y le doy “Continuar” y luego “Aceptar” o “Pegar” según vaya a trabajar directamente con los cuadros o quiera ir guardando la Sintaxis de los procedimientos efectuados.



○ **Sintaxis**

A partir del ejemplo presentado hasta el momento se genera la siguiente sintaxis:

```
FRECUENCIAS
VARIABLES=e2
/STATISTICS=STDDEV RANGE
/ORDER= ANALYSIS .
```

Como se indica en el autotexto referido a medidas de tendencia central, para pedir estadísticos a partir de la frecuencia se los debe escribir luego de la línea: “/STATISTICS=”.
En este caso “STDDEV” refiere al desvío estandar y “RANGE” a rango.

○ **Resultados – Salidas de SPSS**

A partir de lo expuesto, se abre una nueva pantalla en mi SPSS donde aparecen 2 cuadros, el primero con los valores de las medidas de dispersión seleccionadas y el segundo con la distribución de frecuencia de la/s variable/s elegidas.

➔ Frecuencias

[Conjunto_de_datos1] H:\UCU\Bas

Estadísticos

EDAD

N	Válidos	6051
	Perdidos	0
Desv. típ.		23,55036
Rango		99,00

EDAD

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos ,00	62	1,0	1,0	1,0
1,00	70	1,2	1,2	2,2
2,00	70	1,2	1,2	3,3
3,00	64	1,1	1,1	4,4
4,00	83	1,4	1,4	5,8
5,00	77	1,3	1,3	7,0
6,00	81	1,3	1,3	8,4
7,00	73	1,2	1,2	9,6
8,00	85	1,4	1,4	11,0
9,00	71	1,2	1,2	12,2
10,00	93	1,5	1,5	13,7
11,00	84	1,4	1,4	15,1
12,00	75	1,2	1,2	16,3
13,00	87	1,4	1,4	17,8
14,00	86	1,4	1,4	19,2
15,00	74	1,2	1,2	20,4
16,00	73	1,2	1,2	21,6
17,00	89	1,5	1,5	23,1
18,00	93	1,5	1,5	24,6

Recomendaciones

¡Las medidas de dispersión pueden aplicarse **únicamente** a las variables intervalales!

Ejercicios

Ejercicios manuales

1. Las edades, en años, de los empleados de una pequeña compañía son 22, 33, 21, 28, 22, 31, 44, y 19.

- a) Si cada empleada fuera 10 años más grande, cuál sería el rango y el desvío estándar?

2. Encuentra el rango y desvío estándar para x , dadas las siguientes frecuencias

X	f
0	2
1	6
2	3
3	2
4	4

Ejercicio en SPSS

Abrir la base de Encuesta continuas de hogares, personas de Montevideo. Tírar la distribución de frecuencia de las variables **edad** e **ingreso**.

- Hallar el rango y el desvío estándar a través de la utilización del SPSS.
- ¿Cuál de ellas tiene el desvío estándar mayor?, ¿Por qué le parece que esto es así?
- Tome la variable ingreso y seleccione los casos¹ entre 14 y 65 años. Calcule la media, la mediana y el desvío estándar. ¿Qué hipótesis le sugiere el análisis de la distribución del ingreso obtenida?

Bibliografía de referencia

- Peña, D. Romo, J - Introducción a la Estadística para las Ciencias Sociales. Mc Graw Hill 1997.

- Mason y Lind - Estadística para administración y economía. Alfaomega 1998. México, D.F. 8ª edición.

¹ Ver autotexto de Seleccionar casos