

# HOJAS DE CÁLCULO (IX)

## TUTORIALES INFORMÁTICA DOCENCIA RAFALAFENA

Como ya os anunciamos, esta semana vamos a por la estadística un poco más *hardcore*, calculando desviaciones estándar y otras causas de cefalea del residente. Recordad que, si os perdéis, tenéis los tutoriales de investigación que preparó en su momento Elena Herrero para guiaros (y, cómo no, también los comentarios). Pero antes, como siempre, repasemos por donde íbamos tras el tutorial de la semana pasada y sus correspondientes deberes:

	A	B	C	D	E	F	G	H
1	SIP	Sexo	TAS	TAD	TAM	Grupos		
2	819883	M	119	81	98,87	Control	Media TAS:	125,333333333333
3	388755	M	107	62	77,80	Control	Media TAD:	74,4848484848485
4	838204	F	152	72	98,87	Intervención	Número de individuos:	33
5	356577	F	146	75	98,87	Intervención	Número de hombres:	19
6	287952	F	116	90	98,87	Intervención	Porcentaje de hombres:	57,58%
7	312188	F	112	76	88,00	Intervención	Número de mujeres:	14
8	272288	M	127	89	101,67	Control	Porcentaje de mujeres:	42,42%
9	224116	F	122	87	98,87	Intervención	TAS mínima:	156
10	514035	M	150	70	98,87	Intervención	TAS máxima:	100
11	474334	M	119	75	89,67	Control	TAD mínima:	90
12	494971	F	115	73	87,00	Control	TAD máxima:	55
13	446689	M	121	89	99,67	Control	Número HTAS:	8
14	837355	F	106	67	80,00	Control	Porcentaje HTAS:	24,24%
15	497926	M	119	62	81,00	Control	Número HTAD:	12
16	426277	F	100	89	92,67	Control	Porcentaje HTAD:	36,36%
17	321971	F	100	55	70,00	Intervención		
18	949341	F	138	64	88,67	Control	TENDENCIA CENTRAL TAS:	
19	348560	M	149	58	87,00	Control	Media aritmética:	125,333333333333
20	640875	M	103	60	74,33	Intervención	Mediana:	122
21	542324	F	130	65	86,67	Intervención	Moda:	119
22	379191	M	119	86	97,00	Control	TENDENCIA CENTRAL TAD:	
23	289568	M	101	59	73,00	Control	Media aritmética:	74,4848484848485
24	294874	M	121	79	93,00	Control	Mediana:	75
25	760163	F	126	86	109,67	Control	Moda:	89
26	161376	F	146	82	103,33	Intervención		
27	338796	M	108	84	92,00	Control		
28	470335	M	128	79	95,33	Intervención		
29	867644	M	128	56	88,00	Control		
30	593113	M	147	80	109,00	Intervención		
31	864337	F	135	80	105,00	Intervención		
32	129662	F	155	63	93,87	Control		
33	613512	F	151	78	101,00	Intervención		
34	610420	M	129	69	89,00	Intervención		

## MEDIDAS DE DISPERSIÓN

Las medidas de dispersión complementan la información sobre la distribución de la variable, indicando si los valores de la variable están muy dispersos o se concentran alrededor de la medida de centralización.

## RANGO

Cuando se quieren señalar valores extremos en una distribución de datos, se suele utilizar la amplitud como medida de dispersión. La amplitud o rango es la diferencia entre el valor mayor y el menor de la distribución. Antes de empezar a calcularlo, vamos a crear el gran título DISPERSIÓN TAS en G26, justo debajo de nuestros deberes de la semana pasada.

En este caso, no tenemos un operador concreto que calcule el rango, pero si recordáis si que teníamos los operadores MAX y MIN, que nos darán los valores extremos de la distribución. Si acabamos de decir que el rango es la diferencia entre el mayor y el menor valor, tan sólo necesitamos escribir:

$$=MAX(C2:C34)-MIN(C2:C34)$$

Para obtener el valor del rango.

## VARIANZA ( $S^2$ )

Es la media de los cuadrados de las diferencias entre cada valor de la variable y la media aritmética de la distribución. Teóricamente, podríamos hacer esto a mano; es decir, añadir una columna al lado de TAS en que nos calculara la diferencia de la siguiente manera (por ejemplo para el valor C2)

$$=((C2)-AVERAGE(C2:C34))^2$$

Después, simplemente haríamos la media de esa nueva columna y tendríamos el resultado. Pero esto es poco productivo, sobre todo si tenemos un operador que calcula la varianza. Así que introduciremos el operador VAR de la siguiente manera:

$$=VAR(C2:C34)$$

Y GDocs nos dará la varianza.

## DESVIACIÓN TÍPICA (S)

La desviación típica (S) es la raíz cuadrada de la varianza. Expresa la dispersión de la distribución y se expresa en las mismas unidades de medida de la variable. La desviación típica es la medida de dispersión más utilizada en estadística. Para ello, dado que en la fila superior tenemos la varianza, sería tan sencillo como decirle al programa:

$$=SQRT(H28)$$

Por si no lo habíamos comentado, SQRT es el operador que se usa en GDocs para hacer "square root" o, lo que es lo mismo, raíces cuadradas. Pero estaréis de acuerdo conmigo en que este método es poco elegante, porque siempre dependeremos de otras celdas que, si cambian o son borradas, nos trastocarán todas nuestras fórmulas. Mejor usamos el operador específico para la desviación típica, que es STDEV, de la siguiente forma:

$$=STDEV(C2:C34)$$

## CUASIDESVIACIÓN TÍPICA

En ocasiones, cuando queramos hacer inferencias sobre la población, necesitamos cambiar el denominador de la fórmula original de la desviación típica (n), por n-1, de la siguiente forma:

$$S_x = \sqrt{\frac{\sum_{j=1}^n (X_j - \text{Media}(X))^2}{n}} \xrightarrow{\text{se convierte en}} S_x = \sqrt{\frac{\sum_{j=1}^n (X_j - \text{Media}(X))^2}{n - 1}}$$

Para ello, necesitamos otro operador, que aplique esta sutil variación de la fórmula. En este caso, utilizamos STDEVP (recordad, la "P" extra es para poblacional):

$$=STDEVP(C2:C34)$$

## Y LA SEMANA QUE VIENE...

La semana que viene continuaremos con las medidas de dispersión, hablando de percentiles. Como deberes (seguro que ya os lo imaginábais), repetid lo hecho hoy con la TAD.

*Javier Sorribes Monfort*

*MIR-3 de MFyC, CS Rafalafena, Castellón.*