Práctica 7.3

Estimación de la varianza

Aprenderemos a estimar la varianza mediante un ejemplo:

A un profesor de Matemáticas de 2º de Bachillerato le ha comentado un colega que él obtiene en sus exámenes, desde hace años, una media 4,6 y una desviación típica de 2,1 con bastante regularidad. Este dato le hace caer en la cuenta de que llevaba tiempo preocupado porque veía mucha dispersión en sus calificaciones. Para comprobar este dato, elige al azar uno de los exámenes del curso, y obtiene este resultado:

0	0	1	1	1	2	2	3	3	3
4	4	4	4	4	5	5	5	5	5
6	6	6	7	7	7	7	8	9	9

A partir de esta muestra, que no es totalmente aleatoria, ¿qué dispersión puede suponer que ha tenido en los últimos años?

Ante todo hay que advertir que las muestras obtenidas en la enseñanza no son aleatorias puras. Lo que se infiere de ellas no es tan válido como lo obtenido mediante diseño de experimentos. Usamos estos ejemplos por su cercanía, renunciando al rigor propio de otro tipo de curso.

Supuestos de la estimación:

- Estimador: La cuasivarianza (como estimador de la varianza de la población)

- Población: Podemos suponerla infinita, porque el profesor piensa en varios años. Tampoco se comete error suponiéndola normal.
- Muestra: La supondremos aleatoria, aunque no totalmente, ya que se ha elegido un examen al azar.

Abre de nuevo **estima.ods** y en su zona de datos copia las calificaciones del examen. Es posible que tengas que pegar como HTM y después trasponer. Puede quedar así:

Entrada de datos		
0	4	6
0	4	6
1	4	6
1	4	7
1	4	7
2	5	7
2	5	7
3	5	8
3	5	9
3	5	9

Aquí merece la pena detenerse en las cuatro medidas que contiene, además de la media de 4,43, más baja que la de su compañero:

Media	4,43	Desviación típica	2,47
Varianza	6,11	Cuasivarianza	6,32
Desviación estándar o cuasidesviación típica			2,51
Número de datos	30		

Desviación típica: Es la desviación típica usual

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2 \cdot n_i}{N}} = \sqrt{\frac{x_i{}^2 \cdot n_i}{N} - \bar{x}^2}$$

y su valor en este caso es de 2,47, pero hemos explicado en la teoría, que no es un buen estimador de la desviación típica de la población, porque está sesgada.

Varianza: Es la varianza usual, el cuadrado de la anterior: 6,11 = 2,47^2. No nos vale, por la misma razón, su sesgo.

Desviación estándar o insesgada: Es la desviación típica en la que se divide entre n-1 en la fórmula, en lugar de entre n. Se puede calcular multiplicando la desviación típica usual por la raíz de n/(n-1). Este sí es un buen estimador de la desviación típica de los exámenes de ese profesor: 2,51

Cuasivarianza: Es el cuadrado de la desviación estándar, en este caso 6,32 = 2,51^2. Es el estimador insesgado de la varianza de la población.

Según estos resultados, su dispersión en las notas, tal como él sospechaba, puede ser mayor que la de su compañero, ya que ha obtenido una estimación de 2,51 y su compañero suele obtener 2,1.

Intervalo de confianza para la varianza

Para construir la horquilla o intervalo de confianza se suele usar mejor la cuasivarianza, pero existe una fórmula alternativa para el uso de la varianza. En el modelo **estima.ods** se usa la siguiente:

Intervalo para la varianza

$$\left(\frac{nS_n^2}{\chi^2_{1-\alpha/2}}, \frac{nS_n^2}{\chi^2_{\alpha/2}}\right)$$

donde la chi-cuadrado se toma con n-1 grados de libertad

Abre el modelo estima.ods y abre la Hoja *Varianza*. Rellena la celda de tamaño de muestra con el número 30, número de exámenes que ha estudiado, y en la celda de varianza el valor 6,11. Puedes dejar como nivel de confianza el 95%.

Deberás obtener una horquilla de **(2,00 , 3,38)** para la desviación típica de la población de alumnos que examina. Como el límite inferior es 2,00 y su compañero obtiene un 2,1, **no tenemos absoluta confianza** en que sus variabilidades sean distintas.

Intervalo de confianza	**(4,01 , 11,42) para la varianza**
Intervalo de confianza	**(2,00 , 3,38) para la desviación típica**