Introduction to Biostatistics

Topic 1 – Summarizing Data

FAQ – Computational Formula for Sample Variance Calculation

Question – Please show how the two equivalent formulae for the sample variance $S^2$ are actually the same thing.

$$S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{(n-1)} \quad \text{and} \quad S^2 = \frac{\sum_{i=1}^{n} x_i^2 - (n)(\bar{x})^2}{(n-1)}$$

Proof

Since both formulae have $(n-1)$ in the denominator, a proof is obtained by verifying that the two numerators are the same.

Proof that $\sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n \bar{x}^2$:

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} (2\bar{x}x_i) + \sum_{i=1}^{n} (\bar{x}^2)$$

$$= \sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i + (\bar{x}^2)\sum_{i=1}^{n} (1)$$

$$= \sum_{i=1}^{n} x_i^2 - 2\bar{x}(n\bar{x}) + \bar{x}^2(n)$$

$$= \sum_{i=1}^{n} x_i^2 - 2n\bar{x}^2 + n\bar{x}^2$$

$$= \sum_{i=1}^{n} x_i^2 - n\bar{x}^2 \smile$$