PubHlth 540
Introductory Biostastics
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.

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

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 $\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 = \sum_{i=1}^{n} x_i^2 \; - \; n\,\overline{x}^2$:

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

$$= \sum_{i=1}^{n}\left(x_i^2\right) - 2\,\overline{x}\sum_{i=1}^{n}\left(x_i\right) + \left(\overline{x}^2\right)\sum_{i=1}^{n}(1)$$

$$= \sum_{i=1}^{n}\left(x_i^2\right) - 2\,\overline{x}\left(n\,\overline{x}\right) + \overline{x}^2(n)$$

$$= \sum_{i=1}^{n}\left(x_i^2\right) - 2n\,\overline{x}^2 + n\,\overline{x}^2$$

$$= \sum_{i=1}^{n}\left(x_i^2\right) - n\,\overline{x}^2$$ ☺