A2. Statistics from a Geometric Viewpoint

Mean and Variance

Many of the concepts we will encounter can be unified in a very elegant geometric way, which yields additional insight and understanding. If you relate to visual ideas, then you might benefit from reading this. First, recall some basic facts from elementary vector analysis:

For any two column vectors \( \mathbf{v} = (v_1, v_2, \ldots, v_n)^T \) and \( \mathbf{w} = (w_1, w_2, \ldots, w_n)^T \) in \( \mathbb{R}^n \), the standard Euclidean dot product \( \mathbf{v} \cdot \mathbf{w} \) is defined as \( \mathbf{v}^T \mathbf{w} = \sum_{i=1}^{n} v_i w_i \), hence is a scalar. Technically, the dot product is a special case of a more general mathematical object known as an inner product, denoted by \( \langle \mathbf{v}, \mathbf{w} \rangle \), and these notations are often used interchangeably. The length, or norm, of a vector \( \mathbf{v} \) can therefore be characterized as \( \| \mathbf{v} \| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} = \sqrt{\sum_{i=1}^{n} v_i^2} \), and the included angle \( \theta \) between two vectors \( \mathbf{v} \) and \( \mathbf{w} \) can be calculated via the formula

\[
\cos \theta = \frac{\langle \mathbf{v}, \mathbf{w} \rangle}{\| \mathbf{v} \| \| \mathbf{w} \|}, \quad 0 \leq \theta \leq \pi.
\]

From this relation, it is easily seen that two vectors \( \mathbf{v} \) and \( \mathbf{w} \) are orthogonal (i.e., \( \theta = \pi/2 \)), written \( \mathbf{v} \perp \mathbf{w} \), if and only if their dot product is equal to zero, i.e., \( \langle \mathbf{v}, \mathbf{w} \rangle = 0 \).

Now suppose we have \( n \) random sample observations \( \{x_1, x_2, x_3, \ldots, x_n\} \), with mean \( \overline{x} \). As shown below, let \( \mathbf{x} \) be the vector consisting of these \( n \) data values, and \( \overline{x} \) be the vector composed solely of \( \overline{x} \). Note that \( \overline{x} \) is simply a scalar multiple of the vector \( \mathbf{1} = (1, 1, 1, \ldots, 1)^T \). Finally, let \( \mathbf{x} - \overline{x} \) be the vector difference; therefore its components are the individual deviations between the observations and the overall mean. (It’s useful to think of \( \overline{x} \) as a sample taken from an ideal population that responds exactly the same way to some treatment, hence there is no variation; \( \mathbf{x} \) is the sample of actual responses, and \( \mathbf{x} - \overline{x} \) measures the error between them.)
Recall that the sum of the individual deviations is equal to zero, i.e., \( \sum_{i=1}^{n} (x_i - \bar{x}) = 0 \), or in vector notation, the dot product \( \mathbf{1} \cdot (\mathbf{x} - \bar{x}) = 0 \). Therefore, \( \mathbf{1} \perp (\mathbf{x} - \bar{x}) \), and the three vectors above form a right triangle.

Let the scalars \( a, b, \) and \( c \) represent the lengths of the corresponding vectors, respectively. That is,

\[
a = \| \mathbf{x} - \bar{x} \| = \sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}, \quad b = \| \mathbf{x} \| = \sqrt{\sum_{i=1}^{n} x_i^2} = \sqrt{n \bar{x}^2}, \quad c = \| \mathbf{x} \| = \sqrt{\sum_{i=1}^{n} x_i^2}.
\]

Therefore, \( a^2, b^2, \) and \( c^2 \) are all “sums of squares,” denoted by

\[
\text{SS}_{\text{Error}} = a^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2, \quad \text{SS}_{\text{Treatment}} = b^2 = n \bar{x}^2, \quad \text{SS}_{\text{Total}} = c^2 = \sum_{i=1}^{n} x_i^2.
\]

via algebra, \( = \frac{1}{n} \left( \sum_{i=1}^{n} x_i \right)^2 \)

Now via the Pythagorean Theorem, we have \( c^2 = b^2 + a^2 \), referred to in this context as a “partitioning of sums of squares”:

\[
\text{SS}_{\text{Total}} = \text{SS}_{\text{Treatment}} + \text{SS}_{\text{Error}}.
\]

Note also that, by definition, the sample variance is

\[
s^2 = \frac{\text{SS}_{\text{Error}}}{n-1},
\]

and that combining both of these boxed equations yields the equivalent “alternate formula”:

\[
s^2 = \frac{1}{n-1} \left[ \text{SS}_{\text{Total}} - \text{SS}_{\text{Treatment}} \right],
\]

i.e.,

\[
s^2 = \frac{1}{n-1} \left[ \sum_{i=1}^{n} x_i^2 - \frac{1}{n} \left( \sum_{i=1}^{n} x_i \right)^2 \right]
\]

This formula, because it only requires one subtraction rather than \( n \), is computationally more stable than the original; however, it is less enlightening.

**Exercise:** Verify that \( \text{SS}_{\text{Total}} = \text{SS}_{\text{Treatment}} + \text{SS}_{\text{Error}} \) for the sample data values \( \{3, 8, 17, 20, 32\} \), and calculate \( s^2 \) both ways, showing equality. Be especially careful about roundoff error!