Supplement to Section 2.3. Centered Vectors and Variances and Covariances of Vectors: Formulas

Note. This supplement contains a list of formulas from Section 2.3 for quick reference.

Note. First, recall some background formulas: If $x_1, x_2, \ldots, x_n$ are data points from a population of size n, then

- the mean of the population is $\mu = \dfrac{\sum_{i=1}^n x_i}{n}$,
- the variance of the population is $\sigma^2 = \dfrac{\sum_{i=1}^n (x_i - \mu)^2}{n}$, and
- the standard deviation of the population is $\sigma = \sqrt{\dfrac{\sum_{i=1}^n (x_i - \mu)^2}{n}}$.

If $x_1, x_2, \ldots, x_n$ are data points *sampled* from a population, then

- the sample mean is $\overline{x} = \dfrac{\sum_{i=1}^n x_i}{n}$,
- the sample variance is $s^2 = \dfrac{\sum_{i=1}^n (x_i - \overline{x})^2}{(n-1)}$, and
- the sample standard deviation is $s = \sqrt{\sum_{i=1}^n (x_i - \overline{x})^2/(n-1)}$.

For sample (x_i, y_i) for $i = 1, 2, \ldots, n$ of discrete random variable pair (X, Y):

- the sample covariance is $\mathrm{Cov}(X, Y) = \dfrac{1}{n-1} \sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})$.
- the sample correlation is $\mathrm{Corr}(X, Y) = \dfrac{\mathrm{Cov}(X, Y)}{s_x s_y} = \dfrac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}\sqrt{\sum_{i=1}^n (y_i - \overline{y})^2}}$.

Note. Now for a list of new formulas introduced in this section. Here, x is an n-vector of data.

1. For a given n-vector x, its *centered counterpart*, denoted x_c, is $x_c = x - \overline{x}$ where $\overline{x}$.
2. $\overline{x} = \mathrm{proj}_{1_n}(x) = \dfrac{\langle 1_n, x \rangle 1_n}{\|1_n\|^2} = \left(\dfrac{1_n^T x}{n}\right) 1_n$ (recall that $\|1_n\|^2 = (\sqrt{n})^2 = n$).
3. $\|x_c\|^2 = \|x\|^2 - \|\overline{x}\|^2$.

4. The *scaled vector*, denoted x_s, is $x_s = \sqrt{n-1}x/\|x_c\|$.

5. The *centered and scaled vector* is $x_{cs} = \sqrt{n-1}x_c/\|x_c\|$.

6. The *standard deviation* of x is the scalar quantity $s_x = \|x_c\|/\sqrt{n-1}$ and

7. The *variance* of x is $V(x) = s_x^2 = \|x_c\|^2/(n-1)$.

8. For sample (x_i, y_i) for $i = 1, 2, \ldots, n$ of discrete random variable pair (X, Y), the *sample covariance* is $\text{Cov}(X,Y) = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$. The *sample correlation* is $\text{Corr}(X,Y) = \dfrac{\text{Cov}(X,Y)}{s_x s_y} = \dfrac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}}$.

9. The (sample) *covariance* between x and y is $\text{Cov}(x,y) = \dfrac{\langle x - \overline{x}, y - \overline{y}\rangle}{n-1} = \dfrac{\langle x_c, y_c\rangle}{n-1}$.

10. The *correlation* between x and y is $\text{Corr}(x,y) = \text{Cov}(x_{cs}, y_{cs}) = \left\langle \dfrac{x_c}{\|x_c\|}, \dfrac{y_c}{\|y_c\|} \right\rangle = \dfrac{\langle x_c, y_c\rangle}{\|x_c\|\|y_c\|}$.

    We can also express $\text{Corr}(x,y)$ as $\text{Corr}(x,y) = \dfrac{\text{Cov}(x,y)}{\sqrt{V(x)V(y)}}$.

11. $\text{Corr}(ax, y) = \text{sign}(a)\text{Corr}(x, y)$.

Revised: 5/27/2020