Alberto Andrés Valdés González.
Degree: Mathematical Engineer.
Work position: Data Scientist.
Mail: anvaldes@uc.cl/alberto.valdes.gonzalez.96@gmail.com
Location: Santiago, Chile.

Degrees of freedom

Consider the observations $X_1, ..., X_n \overset{\text{iid}}{\sim} D(\mu, \sigma)$.

When we assume a distribution for the observations, we are assumming values of the parameters how in this case is μ and σ.

For every parameter we estimate **we lost one degree of freedom.**

Estimator 1: $(n-1)$ degrees of freedom.

$$\hat{S}_1 = \frac{1}{(n-1)} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$$

We are going to demostrate that estimator is unbiased.

$$\mathbb{E}\left[\hat{S}_1\right] = \mathbb{E}\left[\frac{1}{(n-1)} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2\right] = \frac{1}{(n-1)} \cdot \sum_{i=1}^{n} \mathbb{E}[(x_i - \bar{x})^2]$$

$\Rightarrow$

$$\boxed{\mathbb{E}\left[\hat{S}_1\right] = \frac{1}{(n-1)} \cdot \sum_{i=1}^{n} \mathbb{E}[(x_i - \bar{x})^2]} \quad (1)$$

Note:

$$\mathbb{E}\left[(x_i - \bar{x})^2\right] = \mathbb{E}\left[(x_i)^2 - 2 \cdot x_i \cdot \bar{x} + (\bar{x})^2\right]$$

$$= \mathbb{E}\left[(x_i)^2 - 2 \cdot x_i \cdot \frac{1}{n} \cdot \sum_{j=1}^{n} x_j + \frac{1}{n} \cdot \sum_{j=1}^{n} x_j \cdot \frac{1}{n} \cdot \sum_{k=1}^{n} x_k\right]$$

$$= \mathbb{E}\left[(x_i)^2 - \frac{2}{n} \cdot \sum_{j=1}^{n} x_i \cdot x_j + \frac{1}{n^2} \cdot \sum_{k=1}^{n} \sum_{j=1}^{n} x_j \cdot x_k\right]$$

$$= \mathbb{E}\left[(x_i)^2\right] - \frac{2}{n} \cdot \sum_{j=1}^{n} \mathbb{E}\left[x_i \cdot x_j\right] + \frac{1}{n^2} \cdot \sum_{k=1}^{n} \sum_{j=1}^{n} \mathbb{E}\left[x_j \cdot x_k\right]$$

$$= \mathbb{E}\left[(x_i)^2\right] - \frac{2}{n} \cdot \left(\sum_{j=1:j\neq i}^{n} \mathbb{E}\left[x_i \cdot x_j\right] + \mathbb{E}\left[x_i \cdot x_i\right] \right) + \frac{1}{n^2} \cdot \left(\sum_{k=1:k\neq j}^{n} \sum_{j=1}^{n} \mathbb{E}\left[x_j \cdot x_k\right] + \sum_{j=1}^{n} \mathbb{E}\left[x_j \cdot x_j\right] \right)$$

Now we can see that:

$$\mathbb{E}\left[(x_i)^2\right] = \mathbb{V}\left[x_i\right] + \mathbb{E}\left[x_i\right]^2 = \sigma^2 + \mu^2$$

On the other hand:

$$\sum_{j=1:j\neq i}^{n} \mathbb{E}\left[x_i \cdot x_j\right] + \mathbb{E}\left[x_i \cdot x_i\right] = \sum_{j=1:j\neq i}^{n} \left(Cov(x_i, x_j) + \mathbb{E}[x_i] \cdot \mathbb{E}[x_j]\right) + \mathbb{V}\left[x_i\right] + (\mathbb{E}[X_i])^2$$

$$\sum_{j=1:j\neq i}^{n} (\mu \cdot \mu) + \sigma^2 + \mu^2 = (n-1) \cdot \mu^2 + \sigma^2 + \mu^2 = n \cdot \mu^2 + \sigma^2$$

And finally:

$$\sum_{k=1:k\neq j}^{n} \sum_{j=1}^{n} \mathbb{E}\left[x_j \cdot x_k\right] + \sum_{j=1}^{n} \mathbb{E}\left[x_j \cdot x_j\right]$$

$$= \sum_{k=1:k\neq j}^{n} \sum_{j=1}^{n} (Cov(x_j, x_k) + \mathbb{E}[x_j] \cdot \mathbb{E}[x_k]) + \sum_{j=1}^{n} (\mathbb{V}\left[x_j\right] + \mathbb{E}\left[x_j\right]^2)$$

$$= \sum_{k=1:k\neq j}^{n} \sum_{j=1}^{n} \mu^2 + \sum_{j=1}^{n} (\sigma^2 + \mu^2) = (n^2 - n) \cdot \mu^2 + n \cdot \sigma^2 + n \cdot \mu^2 = n \cdot [n \cdot \mu^2 + \sigma^2]$$

Thus:

$$\mathbb{E}\left[(x_i - \bar{x})^2\right] = \mathbb{E}\left[(x_i)^2\right] - \frac{2}{n} \cdot \left(\sum_{j=1:j\neq i}^{n} \mathbb{E}\left[x_i \cdot x_j\right] + \mathbb{E}\left[x_i \cdot x_i\right] \right) + \frac{1}{n^2} \cdot \left(\sum_{k=1:k\neq j}^{n} \sum_{j=1}^{n} \mathbb{E}\left[x_j \cdot x_k\right] + \sum_{j=1}^{n} \mathbb{E}\left[x_j \cdot x_j\right] \right)$$

$$= (\sigma^2 + \mu^2) - \frac{2}{n} \cdot (n \cdot \mu^2 + \sigma^2) + \frac{1}{n^2} \cdot n \cdot [n \cdot \mu^2 + \sigma^2]$$

$$= \sigma^2 + \mu^2 - 2 \cdot \mu^2 - \frac{2}{n} \cdot \sigma^2 + \mu^2 + \frac{\sigma^2}{n} = \sigma^2 - \frac{\sigma^2}{n} = \frac{(n-1)}{n} \cdot \sigma^2$$

$\Rightarrow$

$$\mathbb{E}\left[(x_i - \bar{x})^2\right] = \frac{(n-1)}{n} \cdot \sigma^2$$

Substituting this on (1) we have that:

$$\mathbb{E}\left[\hat{S}_1\right] = \frac{1}{(n-1)} \cdot \sum_{i=1}^{n} \mathbb{E}[(x_i - \bar{x})^2] = \frac{1}{(n-1)} \cdot \sum_{i=1}^{n} \frac{(n-1)}{n} \cdot \sigma^2$$
$$= \frac{1}{(n-1)} \cdot \frac{(n-1)}{n} \cdot n \cdot \sigma^2 = \sigma^2$$

$\Rightarrow$

$$\boxed{\mathbb{E}\left[\hat{S}_1\right] = \sigma^2}$$

Estimator 2: n degrees of freedom.

$$\hat{S}_2 = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \mu)^2$$

We are going to demostrate that estimator is unbiased.

$$\mathbb{E}\left[\hat{S}_2\right] = \mathbb{E}\left[\frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \mu)^2\right] = \frac{1}{n} \cdot \sum_{i=1}^{n} \mathbb{E}\left[(x_i - \mu)^2\right] = \frac{1}{n} \cdot \sum_{i=1}^{n} \mathbb{E}\left[(x_i)^2 - 2 \cdot \mu \cdot x_i + \mu^2\right]$$
$$= \frac{1}{n} \cdot \left(\sum_{i=1}^{n} (\mathbb{E}\left[(x_i)^2\right] - 2 \cdot \mu \cdot \mathbb{E}\left[x_i\right] + \mu^2)\right) = \frac{1}{n} \cdot \left(\sum_{i=1}^{n} (\mathbb{V}\left[x_i\right] + \mathbb{E}[x_i]^2 - 2 \cdot \mu \cdot \mu + \mu^2)\right)$$
$$= \frac{1}{n} \cdot \left(\sum_{i=1}^{n} (\sigma^2 + \mu^2 - 2 \cdot \mu^2 + \mu^2)\right) = \frac{1}{n} \cdot \left(\sum_{i=1}^{n} \sigma^2\right) = \frac{1}{n} \cdot n \cdot \sigma^2 = \sigma^2$$

$\Rightarrow$

$$\boxed{\mathbb{E}\left[\hat{S}_2\right] = \sigma^2}$$
