2 Measures of Variability

Once we have located the central values of a set of data, it is important to measure the *variability*, whether the data values are tightly clustered or spread out. At the heart of Statistics lies variability: measuring it, reducing it, distinguishing random from real variability, identifying the various sources of real variability and making decisions in the presence of it. We need to know how "unstable" the data is and how much the values differ from its average or from other middle values. These numbers will have small values for closely grouped data (little variation) and larger values for more widely spread out data (large variation).

The measures of variation will also help us assess the reliability of our estimates and the accuracy of our forecasts.

2.1 Quantiles, percentiles and quartiles

Consider the primary data $X = \{x_1, \ldots, x_N\}$. The first two measures of variation give a very general idea of the spread in the data values.

Definition 2.1. *The **range** ($\boxed{\text{range}}$) of X is the difference*

$$x_{max} - x_{min}.$$

If the values of X are sorted in increasing order, then the range is $x_N - x_1$.

Definition 2.2. *The **mean absolute deviation** ($\boxed{\text{mad}}$) of X is the mean of the absolute value of the deviations from the mean, i.e. the value*

$$MAD_1 = \frac{1}{N}\sum_{i=1}^{N} |x_i - \overline{x}|.$$

*The **median absolute deviation** ($\boxed{\text{mad}}$) of X is the median of the absolute value of the deviations from the median, i.e. the value*

$$MAD_2 = \text{median}\{|x_i - \overline{M}|\}.$$

Like the median, the median absolute deviation is not influenced by extreme values, whereas the mean absolute deviation is.

Next, following the idea behind the definition of the median, we define values that divide the data into certain percentages. We simply replace 0.5 in its definition by some probability $0 < p < 1$.

Definition 2.3. *Let X be a set of data sorted increasingly, $p \in (0, 1)$ and $k = 1, 2, \ldots, 99$.*

(1) *A **sample p-quantile** ($\boxed{\text{quantile}}$) is any number that exceeds at most $100p\%$ of the sample and is exceeded by at most $100(1 - p)\%$ of the sample.*

(2) *A **k-percentile** ($\boxed{\text{prctile}}$) P_k is a $(k/100)$-quantile. So, P_k exceeds at most $k\%$ and is exceeded by at most $(100 - k)\%$ of the data*

(3) *The **quartiles** of X are the values*

$$Q_1 = P_{25}, \ \ Q_2 = P_{50} = \overline{M} \ \text{ and } \ Q_3 = P_{75}.$$

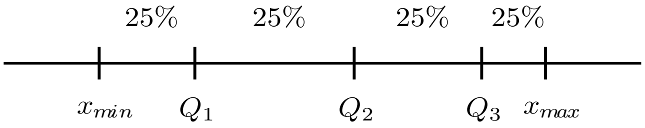


Fig. 1: Quartiles

Definition 2.4. *Let X be a set of sorted data with quartiles Q_1, Q_2 and Q_3.*

(1) *The **interquartile range** ($\boxed{\text{iqr}}$) is the difference between the third and the first quartile*

$$IQR = Q_3 - Q_1. \tag{2.1}$$

(2) *The **interquartile deviation** or the **semi interquartile range** is the value*

$$IQD = \frac{IQR}{2} = \frac{Q_3 - Q_1}{2}. \tag{2.2}$$

(3) *The **interquartile deviation coefficient** or the **relative interquartile deviation** is the value*

$$IQDC = \frac{IQD}{\overline{M}} = \frac{Q_3 - Q_1}{2Q_2}. \tag{2.3}$$

Remark 2.5.

1. The interquartile deviation is an absolute measure of variation and it has an important property: the range $\overline{M} \pm IQD$ contains approximately 50% of the data.
2. The interquartile deviation coefficient $IQDC$ varies between -1 and 1, taking values close to 0 for symmetrical distributions, with little variation and values close to ± 1 for skewed data with large variation.

Example 2.6. Let us use again the data from Example 1.6 from last time, about the CPU times (in seconds) for $N = 30$ randomly chosen jobs (sorted ascendingly):

$$\begin{array}{rrrrrrrrrr}
9 & 15 & 19 & 22 & 24 & 25 & 30 & \textbf{34} & 35 & 35 \\
36 & 36 & 37 & 38 & 42 & 43 & 46 & 48 & 54 & 55 \\
56 & 56 & \textbf{59} & 62 & 69 & 70 & 82 & 82 & 89 & 139
\end{array}$$

and compute various measures of variation.

Solution. For this example, the range is

$$139 - 9 = 130 \text{ seconds}$$

and the mean and median absolute deviations are

$$\begin{aligned}
MAD_1 &= 19.6133, \\
MAD_2 &= 13.5.
\end{aligned}$$

To determine the quartiles, notice that 25% of the sample equals $30/4 = 7.5$ and 75% of the sample is $90/4 = 22.5$ observations. From the ordered sample, we see that the 8th element, 34, has 7 observations to its left and 22 to its right, so it has *no more* than 7.5 observations to the left and *no more* than 22.5 observations to the right of it. Hence, $Q_1 = 34$.

Similarly, the third quartile is the 23rd smallest element, $Q_3 = 59$. Recall from last time that the second quartile (the median) si $Q_2 = \overline{M} = 42.5$. Then

$$\begin{aligned}
IQR &= 59 - 34 \; = \; 25, \\
IQD &= IQR/2 \; = \; 12.5, \\
IQDC &= IQD/Q_2 \; = \; 0.2941.
\end{aligned}$$

The interval

$$\overline{M} \pm IQD \;=\; [30, 55]$$

contains 14 observations.

The value of the $IQDC$ is close neither to 0, nor to the values ± 1. So the data doesn't show strong symmetry or strong asymmetry. This may be due to the extreme values 9 and/or 139. ■

Example 2.7. A computer maker sells extended warranty on the produced computers. It agrees to issue a warranty for x years if it knows that only 10% of computers will fail before the warranty expires. It is known from past experience that lifetimes of these computers have a Gamma distribution with parameters $\alpha = 60$ and $\lambda = 1/5$ years. Compute x and advise the company on the important decision under uncertainty about possible warranties.

Solution. We just need to find the tenth percentile of the specified Gamma distribution and let $x = P_{10}$. In Matlab, that would be computed (as the *inverse* of the cdf) by

$$x = gaminv(0.1, 60, 1/5) = 10.0624.$$

Thus, the company can issue a 10-year warranty rather safely. ■

Remark 2.8. For populations or very large data sets, calculating exact percentiles can be computationally very expensive since it requires sorting all the data values. Machine learning and statistical software use special algorithms (such as linear interpolation) to get an approximate percentile that can be calculated very quickly and is guaranteed to have a certain accuracy.

Outliers

The interquartile range is also involved in another important aspect of statistical analysis, namely the detection of outliers. An *outlier*, as the name suggests, is basically an atypical value, "far away" from the rest of the data, that does not seem to belong to the distribution of the rest of the values in the data set.

We have seen how the mean is very sensitive to outliers. Other statistical procedures can be gravely affected by the presence of outliers in the data. Thus, the problem of detecting and locating an outlier is an important part of any statistical data analysis process.

How to classify a value as being "extreme"? First, we could use a simple property, known as the "3σ rule". This is an application of Chebyshev's inequality

$$P(|X - E(X)| < \varepsilon) \geq 1 - \frac{V(X)}{\varepsilon^2},\ \forall \varepsilon > 0.$$

If we use the classical notations $E(X) = \mu, V(X) = \sigma^2,\ \text{Std}(X) = \sigma$ for the mean, variance and standard deviation of X and take $\varepsilon = 3\sigma$, we get

$$\begin{aligned} P(|X-\mu| < 3\sigma) &\geq 1 - \frac{\sigma^2}{9\sigma^2} \\ &= \frac{8}{9} \approx .89. \end{aligned}$$

This is saying that it is *very* probable (at least 0.89 probable) that $|X - \mu| < 3\sigma$, or, equivalently, that $\mu - 3\sigma < X < \mu + 3\sigma$. In words, the 3σ rule states that *most of the values that any random variable takes, at least* 89%*, lie within* 3 *standard deviations away from the mean*. This property is true in general, for any distribution, but especially for unimodal and symmetrical ones, where that percentage is even higher.

Based on that, one simple procedure would be to consider an outlier any value that is more than 2.5 standard deviations away from the mean, and an *extreme* outlier a value more than 3 standard deviations away from the mean.

A more general approach, that works well also for skewed data, is to consider an outlier any observation that is outside the range

$$\left[Q_1 - \frac{3}{2}IQR,\ Q_3 + \frac{3}{2}IQR\right] = \left[Q_1 - 3IQD,\ Q_3 + 3IQD\right].$$

Also, the coefficient $3/2$ can be replaced by some other number to decrease or enlarge the interval of "normal" values (or, equivalently, the domain that covers the outliers):

$$\left[Q_1 - w \cdot IQR,\ Q_3 + w \cdot IQR\right],\ w = 0.5, 1, 1.5.$$

For our example on CPU times of processors, we have

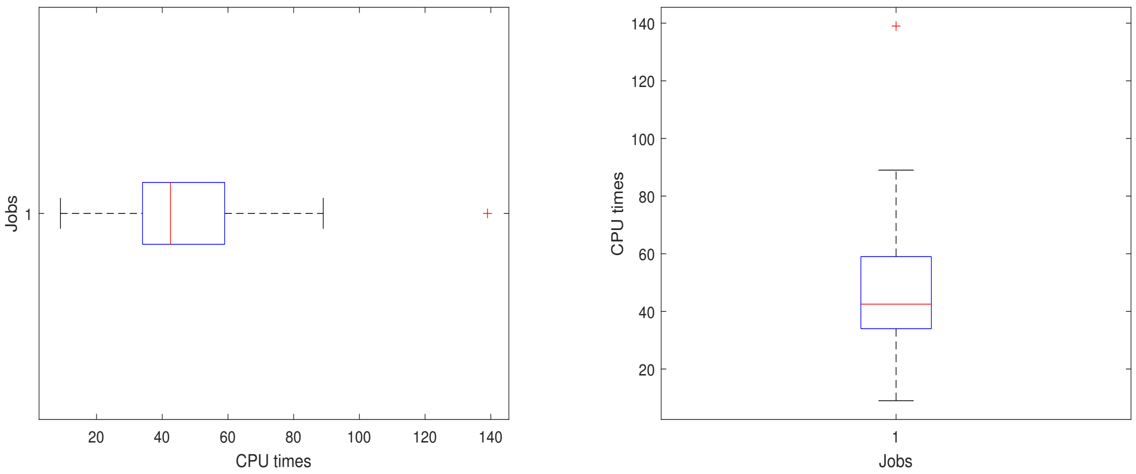
$$\begin{aligned} Q_1 - \frac{3}{2}IQR &= -3.5, \\ Q_3 + \frac{3}{2}IQR &= 96.5, \end{aligned}$$

so observations outside the interval $[-3.5, 96.5]$ are considered outliers. In this case, there is only one, the value 139.

Boxplots

All the information we discussed above is summarized in a graphical display, called a **boxplot** (boxplot), a plot in which a rectangle is drawn to represent the second and third quartiles (so the interquartile range), with a line inside for the median value and which indicates which values are considered extreme. The "whiskers" of the boxplot are the endpoints of the interval on which normal values lie (so everything outside the whiskers is considered an outlier).

For the data in Example 2.6, the boxplot is displayed in Figure 2.



(a) horizontally

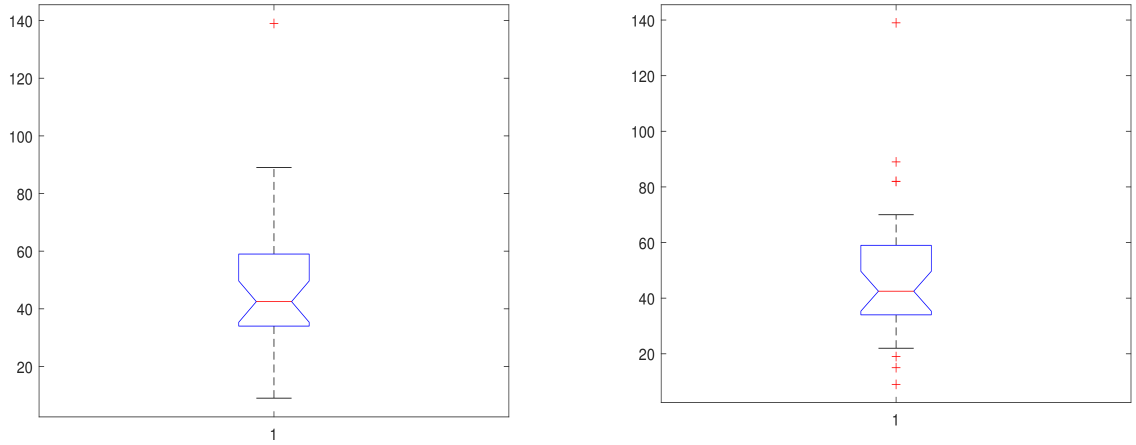
(b) vertically

Fig. 2: Quartiles, Interquartile Range, Outliers

A boxplot can be displayed vertically (default) or horizontally, as in Figure 2. The box can have a "notch" (indentation) at the value of the median, as in Figure 3(a). The width of the interval of the whiskers can be changed. The interval that determines the outliers (i.e., outside of which values are considered too extreme, outliers) is

$$[Q_1 - w \cdot IQR, Q_3 + w \cdot IQR].$$

The default value is $w = 1.5$. With the smaller whiskers, boxplot displays more data points as outliers. In Figure 3(b), the whisker size is set to $w = 0.5$. Then, outliers are all the values outside the interval $[Q_1 - 0.5 \cdot IQR, Q_3 + 0.5 \cdot IQR] = [21.5, 71.5]$. These would be $9, 15, 19$ (too small) and $82, 89, 139$ (too large).



(a) boxplot with a notch

(b) whisker $w = 0.5$

Fig. 3: Boxplots

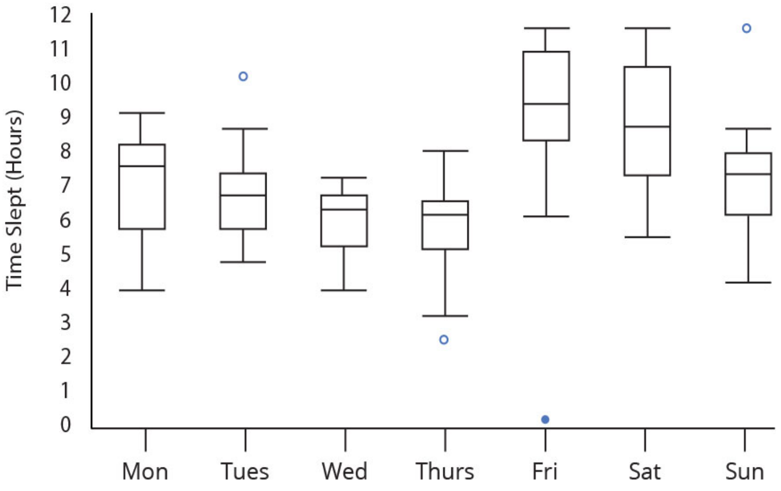


Fig. 4: Multiple boxplots

Boxplots are also very useful when we want to compare data from different samples (see Figure 4). We can compare the interquartile ranges, to examine how the data is dispersed between each sample. The longer the box, the more dispersed the data.

2.2 Moments, variance, standard deviation and coefficient of variation

The idea of the mean can be generalized, by taking various powers of the values in the data.

Definition 2.9.

(1) *The* ***moment of order k*** *is the value*

$$\overline{\nu}_k = \frac{1}{N}\sum_{i=1}^{N} x_i^k,\ \ \overline{\nu}_k = \frac{1}{N}\sum_{i=1}^{n} f_i x_i^k, \tag{2.4}$$

for primary and for grouped data, respectively.

(2) *The* ***central moment of order k*** ($\boxed{\text{moment}}$) *is the value*

$$\overline{\mu}_k = \frac{1}{N}\sum_{i=1}^{N} (x_i - \overline{x})^k,\ \ \overline{\mu}_k = \frac{1}{N}\sum_{i=1}^{n} f_i (x_i - \overline{x})^k \tag{2.5}$$

for primary and for grouped data, respectively.

(3) *The* ***variance*** ($\boxed{\text{var}}$) *is the value*

$$\overline{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N} (x_i - \overline{x})^2,\ \ \overline{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{n} f_i (x_i - \overline{x})^2 \tag{2.6}$$

for primary and for grouped data, respectively. The quantity $\overline{\sigma} = \sqrt{\overline{\sigma}^2}$ *is the* ***standard deviation*** ($\boxed{\text{std}}$).

Remark 2.10.

1. A more efficient computational formula for the variance is

$$\overline{\sigma}^2 = \frac{1}{N}\left(\sum_{i=1}^{N} x_i^2 - \frac{1}{N}\Big(\sum_{i=1}^{N} x_i\Big)^2\right) = \frac{1}{N}\left(\sum_{i=1}^{N} x_i^2 - N\overline{x}^2\right), \tag{2.7}$$

which follows straight from the definition.

2. We will see later that when the data represents a sample (not the entire population), a better

formula is

$$\begin{aligned} s^2 &= \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\overline{x})^2 &&= \frac{1}{N-1}\Big(\sum_{i=1}^{N} x_i^2 - N\overline{x}^2\Big), \\ s^2 &= \frac{1}{N-1}\sum_{i=1}^{n} f_i(x_i-\overline{x})^2 &&= \frac{1}{N-1}\Big(\sum_{i=1}^{N} f_i x_i^2 - N\overline{x}^2\Big), \end{aligned} \tag{2.8}$$

for the *sample* variance for primary or grouped data. The reason the sum is divided by $N-1$ instead of N will have to do with the "bias" of an estimator and will be explained later on in the next chapter. To fully explain why using N leads to a biased estimate involves the notion of *degrees of freedom*, which takes into account the number of constraints in computing an estimate. The sample observations $x_1, \ldots, x_N$ are independent (by the definition of a random sample), but when computing the variance, we use the variables $x_1 - \overline{x}, \ldots, x_N - \overline{x}$. Notice that by subtracting the sample mean $\overline{x}$ from each observation, there exists a linear relation among the elements, namely

$$\sum_{k=1}^{N}(x_k - \overline{x}) = 0$$

and, thus, we lose 1 degree of freedom due to this constraint. Hence, there are only $N-1$ degrees of freedom. So, we will use (2.7) to compute the variance of a set of data that represents a population and (2.8) for the variance of a sample.

Example 2.11. Consider again our previous example on CPU times (in seconds) for $N = 30$ randomly chosen jobs:

70	36	43	69	82	48	34	62	35	15
59	139	46	37	42	30	55	56	36	82
38	89	54	25	35	24	22	9	56	19

Recall that for this data the sample mean was $\overline{x} = 48.2333$ seconds. The sample variance is

$$s^2 = \frac{(70-48.2333)^2 + \ldots + (19-48.2333)^2}{30-1} = \frac{20391}{29} \approx 703.1506 \text{ sec}^2.$$

Alternatively, using (2.7),

$$s^2 = \frac{70^2 + \ldots + 19^2 - 30\cdot 48.2333^2}{30-1} = \frac{90185 - 69794}{29} \approx 703.1506 \text{ sec}^2.$$

The sample standard deviation is

$$s = \sqrt{703.1506} \approx 26.1506 \text{ sec}.$$

By the 3σ rule, using $\overline{x}$ and s as estimates for the population mean μ and population standard deviation σ, we may infer that at least 89% of the tasks performed by this processor require between $\overline{x} - 3s = -30.2185$ and $\overline{x} + 3s = 126.6851$ (so less than 126.6851) seconds of CPU time.

Definition 2.12. *The* ***coefficient of variation*** *is the value*

$$CV = \frac{s}{\overline{x}}.$$

Remark 2.13.
1. The coefficient of variation is also known as the **relative standard deviation (RSD)**.
2. It can be expressed as a ratio or as a percentage. It is useful in comparing the degrees of variation of two sets of data, even when their means are different.
2. The coefficient of variation is used in fields such as Analytical Chemistry, Engineering or Physics when doing quality assurance studies. It is also widely used in Business Statistics. For example, in the investing world, the coefficient of variation helps brokers determine how much volatility (risk) they are assuming in comparison to the amount of return they can expect from a certain investment. The lower the value of the CV, the better the risk-return trade off.

3 Sample Theory

In inferential Statistics, we will have the following situation: we are interested in studying a characteristic (a random variable) X, relative to a population P of (known or unknown) size N. The difficulty or even the impossibility of studying the entire population, as well as the merits of choosing and studying a random sample from which to make inferences about the population of interest, have already been discussed in the previous chapter. Now, we want to give a more rigorous and precise definition of a random sample, in the framework of random variables, one that can then employ probability theory techniques for making inferences.

3.1 Random Samples and Sample Functions

We choose n objects from the population and actually study $X_i,\ i = \overline{1,n}$, the characteristic of interest *for the i^{th} object selected.* Since the n objects were randomly selected, it makes sense that for $i = \overline{1,n}$, X_i is a random variable, one that has *the same* distribution (pdf) as X, the characteristic relative to the entire population. Furthermore, these random variables are independent, since the value assumed by one of them has no effect on the values assumed by the others. Once the n objects have been selected, we will have n numerical values available, $x_1, \dots, x_n$, the observed values of the sample variables $X_1, \dots, X_n$.

Definition 3.1. *A* ***random sample of size n*** *from the distribution of X, a characteristic relative to a population P, is a collection of n independent random variables $X_1, \dots, X_n$, having the same distribution as X. The variables $X_1, \dots, X_n$, are called* ***sample variables*** *and their observed values $x_1, \dots, x_n$, are called* ***sample data****.*

Remark 3.2. The term *random sample* may refer to the objects selected, to the sample variables, or to the sample data. It is usually clear from the context which meaning is intended. In general, we use capital letters to denote sample variables and corresponding lowercase letters for their observed values, the sample data.

We are able now to define sample functions, or statistics, in the more precise context of random variables.

Definition 3.3. *A* ***sample function*** *or* ***statistic*** *is a random variable*

$$Y_n = h_n(X_1, \dots, X_n),$$

where $h_n : \mathbb{R}^n \to \mathbb{R}$ is a measurable function. The value of the sample function Y_n is $y_n = h_n(x_1, \dots, x_n)$.

We will revisit now some sample numerical characteristics discussed in the previous sections and define them as sample functions. That means they will have a pdf, a cdf, a mean value, variance, standard deviation, etc. A sample function will, in general, be an approximation for the corresponding population characteristic. In that context, the standard deviation of the sample function is usually referred to as the **standard error**.

In what follows, $\{X_1, \dots, X_n\}$ denotes a sample of size n drawn from the distribution of some population characteristic X.

3.2 Sample Mean

Definition 3.4. *The* ***sample mean*** *is the sample function defined by*

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{3.1}$$

and its value is $\overline{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i$.

Now that the sample mean is defined as a random variable, we can discuss its numerical characteristics.

Proposition 3.5. *Let X be a population characteristic with mean $E(X) = \mu$ and variance $V(X) = \sigma^2$. Then*

$$E\left(\overline{X}\right) = \mu \text{ and } V\left(\overline{X}\right) = \frac{\sigma^2}{n}. \tag{3.2}$$

Proof. Since $X_1, \dots, X_n$ are identically distributed, with the same distribution as X, $E(X_i) = E(X) = \mu$ and $V(X_i) = V(X) = \sigma^2$, $\forall i = \overline{1, n}$. Then, by the usual properties of expectation, we have

$$E\left(\overline{X}\right) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\, n\mu = \mu.$$

Further, since $X_1, \dots, X_n$ are also independent, by the properties of variance, it follows that

$$V\left(\overline{X}\right) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{1}{n^2}\, n\sigma^2 = \frac{\sigma^2}{n}.$$

□

Remark 3.6. As a consequence, the standard deviation of $\overline{X}$ is

$$\mathrm{Std}(\overline{X}) \;=\; \sqrt{V(\overline{X})} \;=\; \frac{\sigma}{\sqrt{n}}.$$

So, when estimating the population mean μ from a sample of size n by the sample mean $\overline{X}$, the *standard error* of the estimate is $\sigma/\sqrt{n}$, which oftentimes is estimated by $s/\sqrt{n}$. Either way, notice that as n increases and tends to ∞, the standard error decreases and approaches 0. That means that the larger the sample on which we base our estimate, the more accurate the approximation.

3.3 Sample Moments and Sample Variance

Definition 3.7. *The statistic*

$$\overline{\nu}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \tag{3.3}$$

is called the ***sample moment of order k*** *and its value is* $\frac{1}{n}\sum_{i=1}^{n} x_i^k$.

The statistic

$$\overline{\mu}_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k \tag{3.4}$$

is called the ***sample central moment of order k*** *and its value is* $\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^k$.

Remark 3.8. Just like for theoretical (population) moments, we have

$$\begin{aligned} \overline{\nu}_1 &= \overline{X}, \\ \overline{\mu}_1 &= 0, \\ \overline{\mu}_2 &= \overline{\nu}_2 - \overline{\nu}_1^2. \end{aligned}$$

Next we discuss the characteristics of these new sample functions.

Proposition 3.9. *Let X be a characteristic with the property that for $k \in \mathbb{N}$, the theoretical moment $\nu_{2k} = \nu_{2k}(X) = E\left(X^{2k}\right)$ exists. Then*

$$E\left(\overline{\nu}_k\right) = \nu_k \text{ and } V\left(\overline{\nu}_k\right) = \frac{1}{n}\left(\nu_{2k} - \nu_k^2\right). \tag{3.5}$$

Proof. First off, the condition that ν_{2k} exists for X ensures the fact that all theoretical moments of X of order up to k also exist. The rest follows as before. We have

$$E\left(\overline{\nu}_k\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i^k) = \frac{1}{n}\sum_{i=1}^{n} E(X^k) = \frac{1}{n}\, n\nu_k = \nu_k$$

and

$$\begin{aligned} V\left(\overline{\nu}_k\right) &= \frac{1}{n^2}\sum_{i=1}^{n} V(X_i^k) &&= \frac{1}{n^2}\sum_{i=1}^{n} V(X^k) \\ &= \frac{1}{n^2}\, n\left(\nu_{2k} - \nu_k^2\right) &&= \frac{1}{n}\left(\nu_{2k} - \nu_k^2\right). \end{aligned}$$

□

Proposition 3.10. *Let X be a characteristic with variance $V(X) = \mu_2 = \sigma^2$ and for which the theoretical moment $\nu_4 = E\left(X^4\right)$ exists. Then*

$$\begin{aligned} E\left(\overline{\mu}_2\right) &= \frac{n-1}{n}\,\sigma^2, \\ V\left(\overline{\mu}_2\right) &= \frac{n-1}{n^3}\Big[(n-1)\mu_4 - (n-3)\sigma^4\Big]. \end{aligned} \tag{3.6}$$

Proof. We only prove the first assertion, as it is the most important and oftenly used property of $\overline{\mu}_2$. Using Proposition 3.9, Remark 3.8, the properties of expectation and the fact that $X_1, \dots, X_n$ are independent and identically distributed, we have

$$\begin{aligned} E\left(\overline{\mu}_2\right) &= E\left(\overline{\nu}_2\right) - E\left(\overline{\nu}_1^2\right) = \nu_2 - E\left(\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2\right) \\ &= \nu_2 - \frac{1}{n^2}E\Big(\sum_{i=1}^{n} X_i^2 + 2\sum_{i<j} X_i X_j\Big) \\ &= \nu_2 - \frac{1}{n^2}\left[\sum_{i=1}^{n} E\left(X_i^2\right) + 2\sum_{i<j} E(X_i)E(X_j)\right] \\ &= \nu_2 - \frac{1}{n^2}\left[n\nu_2 + 2\,\frac{n(n-1)}{2}\,\nu_1^2\right] = \nu_2 - \frac{1}{n}\nu_2 - \frac{n-1}{n}\nu_1^2 \\ &= \frac{n-1}{n}\left(\nu_2 - \nu_1^2\right) = \frac{n-1}{n}\,\sigma^2. \end{aligned}$$

□

Remark 3.11. Notice that the sample central moment of order 2 is the first statistic whose expected value *is not* the corresponding population function, in this case the theoretical variance. This is the motivation for the next definition.

Definition 3.12. *The statistic*

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \tag{3.7}$$

is called the ***sample variance*** *and its value is* $\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$.

The statistic $s = \sqrt{s^2}$ is called the ***sample standard deviation****.*

Remark 3.13. Notice that the sample central moment of order 2 is no longer equal to the sample variance, as we are used. In fact, we have

$$s^2 = \frac{n}{n-1}\,\overline{\mu}_2.$$

Then, by Proposition 3.10, we have for the sample variance

$$\begin{aligned} E\left(s^2\right) &= \mu_2 \;=\; \sigma^2, \\ V\left(s^2\right) &= \frac{1}{n(n-1)}\Big[(n-1)\mu_4 - (n-3)\sigma^4\Big] \end{aligned} \tag{3.8}$$

and, again, the estimation of σ^2 by s^2 (or of σ by s) has a standard error that decreases as the sample size increases:

$$\mathrm{Std}(s^2) \;=\; \sqrt{\frac{1}{n(n-1)}\Big((n-1)\mu_4 - (n-3)\sigma^4\Big)} \;\longrightarrow\; 0, \;\text{ as }\; n \to \infty.$$