

6.5: The Sample Variance

Descriptive Theory

Recall the basic model of statistics: we have a population of objects of interest, and we have various measurements (variables) that we make on these objects. We select objects from the population and record the variables for the objects in the sample; these become our data. Once again, our first discussion is from a descriptive point of view. That is, we do not assume that the data are generated by an underlying probability distribution. Remember however, that the data themselves form a probability distribution.

Variance and Standard Deviation

Suppose that $\mathbf{x} = (x_1, x_2, \dots, x_n)$ is a sample of size n from a real-valued variable x . Recall that the sample mean is

$$m = \frac{1}{n} \sum_{i=1}^n x_i \quad (6.5.1)$$

and is the most important measure of the center of the data set. The *sample variance* is defined to be

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - m)^2 \quad (6.5.2)$$

If we need to indicate the dependence on the data vector \mathbf{x} , we write $s^2(\mathbf{x})$. The difference $x_i - m$ is the *deviation* of x_i from the mean m of the data set. Thus, the variance is the *mean square deviation* and is a measure of the spread of the data set with respect to the mean. The reason for dividing by $n-1$ rather than n is best understood in terms of the inferential point of view that we discuss in the next section; this definition makes the sample variance an unbiased estimator of the distribution variance. However, the reason for the averaging can also be understood in terms of a related concept.

$$\sum_{i=1}^n (x_i - m) = 0.$$

Proof

$$\sum_{i=1}^n (x_i - m) = \sum_{i=1}^n x_i - \sum_{i=1}^n m = nm - nm = 0.$$

Thus, if we know $n-1$ of the deviations, we can compute the last one. This means that there are only $n-1$ freely varying deviations, that is to say, $n-1$ *degrees of freedom* in the set of deviations. In the definition of sample variance, we average the squared deviations, not by dividing by the number of terms, but rather by dividing by the number of degrees of freedom in those terms. However, this argument notwithstanding, it would be reasonable, *from a purely descriptive point of view*, to divide by n in the definition of the sample variance. Moreover, when n is sufficiently large, it hardly matters whether we divide by n or by $n-1$.

In any event, the square root s of the sample variance s^2 is the sample standard deviation. It is the *root mean square deviation* and is also a measure of the spread of the data with respect to the mean. Both measures of spread are important. Variance has nicer mathematical properties, but its physical unit is the square of the unit of x . For example, if the underlying variable x is the height of a person in inches, the variance is in square inches. On the other hand, the standard deviation has the same physical unit as the original variable, but its mathematical properties are not as nice.

Recall that the data set \mathbf{x} naturally gives rise to a probability distribution, namely the *empirical distribution* that places probability $\frac{1}{n}$ at x_i for each i . Thus, if the data are distinct, this is the uniform distribution on $\{x_1, x_2, \dots, x_n\}$. The sample mean m is simply the expected value of the empirical distribution. Similarly, if we were to divide by n rather than $n-1$, the sample variance would be the variance of the empirical distribution. Most of the properties and results this section follow from much more general properties and results for the variance of a probability distribution (although for the most part, we give independent proofs).

Measures of Center and Spread

Measures of center and measures of spread are best thought of together, in the context of an *error function*. The error function measures how well a single number a represents the entire data set \mathbf{x} . The values of a (if they exist) that minimize the error functions are our measures of center; the minimum value of the error function is the corresponding measure of spread. Of course, we hope for a *single* value of a that minimizes the error function, so that we have a unique measure of center.

Let's apply this procedure to the *mean square error function* defined by

$$\text{mse}(a) = \frac{1}{n-1} \sum_{i=1}^n (x_i - a)^2, \quad a \in \mathbb{R} \quad (6.5.3)$$

Minimizing mse is a standard problem in calculus.

The graph of mse is a parabola opening upward.

1. mse is minimized when $a = m$, the sample mean.
2. The minimum value of mse is s^2 , the sample variance.

Proof

We can tell from the form of mse that the graph is a parabola opening upward. Taking the derivative gives

$$\frac{d}{da} \text{mse}(a) = -\frac{2}{n-1} \sum_{i=1}^n (x_i - a) = -\frac{2}{n-1} (nm - na) \quad (6.5.4)$$

Hence $a = m$ is the unique value that minimizes mse. Of course, $\text{mse}(m) = s^2$.

Trivially, if we defined the mean square error function by dividing by n rather than $n-1$, then the minimum value would still occur at m , the sample mean, but the minimum value would be the alternate version of the sample variance in which we divide by n . On the other hand, if we were to use the *root mean square deviation* function $\text{rmse}(a) = \sqrt{\text{mse}(a)}$, then because the square root function is strictly increasing on $[0, \infty)$, the minimum value would again occur at m , the sample mean, but the minimum value would be s , the sample standard deviation. The important point is that with all of these error functions, the unique measure of center is the sample mean, and the corresponding measures of spread are the various ones that we are studying.

Next, let's apply our procedure to the *mean absolute error function* defined by

$$\text{mae}(a) = \frac{1}{n-1} \sum_{i=1}^n |x_i - a|, \quad a \in \mathbb{R} \quad (6.5.5)$$

The mean absolute error function satisfies the following properties:

1. mae is a continuous function.
2. The graph of mae consists of lines.
3. The slope of the line at a depends on where a is in the data set \mathbf{x} .

Proof

For parts (a) and (b), note that for each i , $|x_i - a|$ is a continuous function of a with the graph consisting of two lines (of slopes ± 1) meeting at x_i .

Mathematically, mae has some problems as an error function. First, the function will not be smooth (differentiable) at points where two lines of different slopes meet. More importantly, the values that minimize mae may occupy an entire interval, thus leaving us without a unique measure of center. The [error function exercises](#) below will show you that these pathologies can really happen. It turns out that mae is minimized at any point in the median interval of the data set \mathbf{x} . The proof of this result follows from a much more general result for probability distributions. Thus, the medians are the natural measures of center associated with mae as a measure of error, in the same way that the sample mean is the measure of center associated with the mse as a measure of error.

Properties

In this section, we establish some essential properties of the sample variance and standard deviation. First, the following alternate formula for the sample variance is better for computational purposes, and for certain theoretical purposes as well.

The sample variance can be computed as

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n x_i^2 - \frac{n}{n-1} m^2 \quad (6.5.6)$$

Proof

Note that

$$\sum_{i=1}^n (x_i - m)^2 = \sum_{i=1}^n (x_i^2 - 2mx_i + m^2) = \sum_{i=1}^n x_i^2 - 2m \sum_{i=1}^n x_i + \sum_{i=1}^n m^2 \quad (6.5.7)$$

$$= \sum_{i=1}^n x_i^2 - 2nm^2 + nm^2 = \sum_{i=1}^n x_i^2 - nm^2 \quad (6.5.8)$$

Dividing by $n - 1$ gives the result.

If we let $\mathbf{x}^2 = (x_1^2, x_2^2, \dots, x_n^2)$ denote the sample from the variable x^2 , then the computational formula in the last exercise can be written succinctly as

$$s^2(\mathbf{x}) = \frac{n}{n-1} [m(\mathbf{x}^2) - m^2(\mathbf{x})] \quad (6.5.9)$$

The following theorem gives another computational formula for the sample variance, directly in terms of the variables and thus without the computation of an intermediate statistic.

The sample variance can be computed as

$$s^2 = \frac{1}{2n(n-1)} \sum_{i=1}^n \sum_{j=1}^n (x_i - x_j)^2 \quad (6.5.10)$$

Proof

Note that

$$\frac{1}{2n} \sum_{i=1}^n \sum_{j=1}^n (x_i - x_j)^2 = \frac{1}{2n} \sum_{i=1}^n \sum_{j=1}^n (x_i - m + m - x_j)^2 \quad (6.5.11)$$

$$= \frac{1}{2n} \sum_{i=1}^n \sum_{j=1}^n [(x_i - m)^2 + 2(x_i - m)(m - x_j) + (m - x_j)^2] \quad (6.5.12)$$

$$= \frac{1}{2n} \sum_{i=1}^n \sum_{j=1}^n (x_i - m)^2 + \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n (x_i - m)(m - x_j) + \frac{1}{2n} \sum_{i=1}^n \sum_{j=1}^n (m - x_j)^2 \quad (6.5.13)$$

$$= \frac{1}{2} \sum_{i=1}^n (x_i - m)^2 + 0 + \frac{1}{2} \sum_{j=1}^n (m - x_j)^2 \quad (6.5.14)$$

$$= \sum_{i=1}^n (x_i - m)^2 \quad (6.5.15)$$

Dividing by $n - 1$ gives the result.

The sample variance is nonnegative:

1. $s^2 \geq 0$
2. $s^2 = 0$ if and only if $x_i = x_j$ for each $i, j \in \{1, 2, \dots, n\}$.

Proof

Part (a) is obvious. For part (b) note that if $s^2 = 0$ then $x_i = m$ for each i . Conversely, if \mathbf{x} is a constant vector, then m is that same constant.

Thus, $s^2 = 0$ if and only if the data set is constant (and then, of course, the mean is the common value).

If c is a constant then

1. $s^2(c\mathbf{x}) = c^2 s^2(\mathbf{x})$
2. $s(c\mathbf{x}) = |c| s(\mathbf{x})$

Proof

For part (a), recall that $m(c\mathbf{x}) = cm(\mathbf{x})$. Hence

$$s^2(c\mathbf{x}) = \frac{1}{n-1} \sum_{i=1}^n [cx_i - cm(\mathbf{x})]^2 = \frac{1}{n-1} \sum_{i=1}^n c^2 [x_i - m(\mathbf{x})]^2 = c^2 s^2(\mathbf{x}) \quad (6.5.16)$$

If \mathbf{c} is a sample of size n from a constant c then

1. $s^2(\mathbf{x} + \mathbf{c}) = s^2(\mathbf{x})$.
2. $s(\mathbf{x} + \mathbf{c}) = s(\mathbf{x})$

Proof

Recall that $m(\mathbf{x} + \mathbf{c}) = m(\mathbf{x}) + c$. Hence

$$s^2(\mathbf{x} + \mathbf{c}) = \frac{1}{n-1} \sum_{i=1}^n \{(x_i + c) - [m(\mathbf{x}) + c]\}^2 = \frac{1}{n-1} \sum_{i=1}^n [x_i - m(\mathbf{x})]^2 = s^2(\mathbf{x}) \quad (6.5.17)$$

As a special case of these results, suppose that $\mathbf{x} = (x_1, x_2, \dots, x_n)$ is a sample of size n corresponding to a real variable x , and that a and b are constants. The sample corresponding to the variable $y = a + bx$, in our vector notation, is $\mathbf{a} + b\mathbf{x}$. Then $m(\mathbf{a} + b\mathbf{x}) = a + bm(\mathbf{x})$ and $s(\mathbf{a} + b\mathbf{x}) = |b|s(\mathbf{x})$. Linear transformations of this type, when $b > 0$, arise frequently when physical units are changed. In this case, the transformation is often called a *location-scale* transformation; a is the location parameter and b is the scale parameter. For example, if x is the length of an object in inches, then $y = 2.54x$ is the length of the object in centimeters. If x is the temperature of an object in degrees Fahrenheit, then $y = \frac{5}{9}(x - 32)$ is the temperature of the object in degree Celsius.

Now, for $i \in \{1, 2, \dots, n\}$, let $z_i = (x_i - m)/s$. The number z_i is the *standard score* associated with x_i . Note that since x_i , m , and s have the same physical units, the standard score z_i is *dimensionless* (that is, has no physical units); it measures the directed distance from the mean m to the data value x_i in standard deviations.

The sample of standard scores $\mathbf{z} = (z_1, z_2, \dots, z_n)$ has mean 0 and variance 1. That is,

1. $m(\mathbf{z}) = 0$
2. $s^2(\mathbf{z}) = 1$

Proof

These results follow from Theorems 7 and 8. In vector notation, note that $\mathbf{z} = (\mathbf{x} - \mathbf{m})/s$. Hence $m(\mathbf{z}) = (m - m)/s = 0$ and $s(\mathbf{z}) = s/s = 1$.

Approximating the Variance

Suppose that instead of the actual data \mathbf{x} , we have a frequency distribution corresponding to a partition with classes (intervals) (A_1, A_2, \dots, A_k) , class marks (midpoints of the intervals) (t_1, t_2, \dots, t_k) , and frequencies (n_1, n_2, \dots, n_k) . Recall that the relative frequency of class A_j is $p_j = n_j/n$. In this case, approximate values of the sample mean and variance are, respectively,

$$m = \frac{1}{n} \sum_{j=1}^k n_j t_j = \sum_{j=1}^k p_j t_j \quad (6.5.18)$$

$$s^2 = \frac{1}{n-1} \sum_{j=1}^k n_j (t_j - m)^2 = \frac{n}{n-1} \sum_{j=1}^k p_j (t_j - m)^2 \quad (6.5.19)$$

These approximations are based on the hope that the data values in each class are well represented by the class mark. In fact, these are the standard definitions of sample mean and variance for the data set in which t_j occurs n_j times for each j .

Inferential Statistics

We continue our discussion of the sample variance, but now we assume that the variables are random. Thus, suppose that we have a basic random experiment, and that X is a real-valued random variable for the experiment with mean μ and standard deviation σ . We will need some higher order moments as well. Let $\sigma_3 = \mathbb{E}[(X - \mu)^3]$ and $\sigma_4 = \mathbb{E}[(X - \mu)^4]$ denote the 3rd and 4th moments about the mean. Recall that $\sigma_3/\sigma^3 = \text{skew}(X)$, the skewness of X , and $\sigma_4/\sigma^4 = \text{kurt}(X)$, the kurtosis of X . We assume that $\sigma_4 < \infty$.

We repeat the basic experiment n times to form a new, compound experiment, with a sequence of independent random variables $\mathbf{X} = (X_1, X_2, \dots, X_n)$, each with the same distribution as X . In statistical terms, \mathbf{X} is a random sample of size n from the distribution of X . All of the statistics above make sense for \mathbf{X} , of course, but now these statistics are random variables. We will use the same notation, except for the usual convention of denoting random variables by capital letters. Finally, note that the deterministic properties and relations established above still hold.

In addition to being a measure of the center of the data \mathbf{X} , the sample mean

$$M = \frac{1}{n} \sum_{i=1}^n X_i \quad (6.5.20)$$

is a natural estimator of the distribution mean μ . In this section, we will derive statistics that are natural estimators of the distribution variance σ^2 . The statistics that we will derive are different, depending on whether μ is known or unknown; for this reason, μ is referred to as a *nuisance parameter* for the problem of estimating σ^2 .

A Special Sample Variance

First we will assume that μ is known. Although this is almost always an artificial assumption, it is a nice place to start because the analysis is relatively easy and will give us insight for the standard case. A natural estimator of σ^2 is the following statistic, which we will refer to as the *special sample variance*.

$$W^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \quad (6.5.21)$$

W^2 is the sample mean for a random sample of size n from the distribution of $(X - \mu)^2$, and satisfies the following properties:

1. $\mathbb{E}(W^2) = \sigma^2$
2. $\text{var}(W^2) = \frac{1}{n}(\sigma_4 - \sigma^4)$
3. $W^2 \rightarrow \sigma^2$ as $n \rightarrow \infty$ with probability 1
4. The distribution of $\sqrt{n}(W^2 - \sigma^2) / \sqrt{\sigma_4 - \sigma^4}$ converges to the standard normal distribution as $n \rightarrow \infty$.

Proof

These result follow immediately from standard results in the section on the Law of Large Numbers and the section on the Central Limit Theorem. For part (b), note that

$$\text{var}[(X - \mu)^2] = \mathbb{E}[(X - \mu)^4] - (\mathbb{E}[(X - \mu)^2])^2 = \sigma_4 - \sigma^4 \quad (6.5.22)$$

In particular part (a) means that W^2 is an *unbiased* estimator of σ^2 . From part (b), note that $\text{var}(W^2) \rightarrow 0$ as $n \rightarrow \infty$; this means that W^2 is a *consistent* estimator of σ^2 . The square root of the special sample variance is a special version of the *sample standard deviation*, denoted W .

$\mathbb{E}(W) \leq \sigma$. Thus, W is a *negatively biased* estimator that tends to underestimate σ .

Proof

This follows from the unbiased property and Jensen's inequality. Since $w \mapsto \sqrt{w}$ is concave downward on $[0, \infty)$, we have $\mathbb{E}(W) = \mathbb{E}(\sqrt{W^2}) \leq \sqrt{\mathbb{E}(W^2)} = \sqrt{\sigma^2} = \sigma$.

Next we compute the covariance and correlation between the sample mean and the special sample variance.

The covariance and correlation of M and W^2 are

1. $\text{cov}(M, W^2) = \sigma_3/n$.
2. $\text{cor}(M, W^2) = \sigma^3 / \sqrt{\sigma^2(\sigma_4 - \sigma^4)}$

Proof

1. From the bilinearity of the covariance operator and by independence,

$$\text{cov}(M, W^2) = \text{cov} \left[\frac{1}{n} \sum_{i=1}^n X_i, \frac{1}{n} \sum_{j=1}^n (X_j - \mu)^2 \right] = \frac{1}{n^2} \sum_{i=1}^n \text{cov} [X_i, (X_i - \mu)^2] \quad (6.5.23)$$

But $\text{cov} [X_i, (X_i - \mu)^2] = \text{cov} [X_i - \mu, (X_i - \mu)^2] = \mathbb{E} [(X_i - \mu)^3] - \mathbb{E}(X_i - \mu) \mathbb{E} [(X_i - \mu)^2] = \sigma_3$. Substituting gives the result.

2. This follows from part (a), the unbiased property, and our previous result that $\text{var}(M) = \sigma^2/n$.

Note that the correlation does not depend on the sample size, and that the sample mean and the special sample variance are uncorrelated if $\sigma_3 = 0$ (equivalently $\text{skew}(X) = 0$).

The Standard Sample Variance

Consider now the more realistic case in which μ is unknown. In this case, a natural approach is to average, in some sense, the squared deviations $(X_i - M)^2$ over $i \in \{1, 2, \dots, n\}$. It might seem that we should average by dividing by n . However, another approach is to divide by whatever constant would give us an unbiased estimator of σ^2 . This constant turns out to be $n - 1$, leading to the *standard sample variance*:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - M)^2 \quad (6.5.24)$$

$\mathbb{E}(S^2) = \sigma^2$.

Proof

By expanding (as was shown in the last section),

$$\sum_{i=1}^n (X_i - M)^2 = \sum_{i=1}^n X_i^2 - nM^2 \quad (6.5.25)$$

Recall that $\mathbb{E}(M) = \mu$ and $\text{var}(M) = \sigma^2/n$. Taking expected values in the displayed equation gives

$$\mathbb{E} \left(\sum_{i=1}^n (X_i - M)^2 \right) = \sum_{i=1}^n (\sigma^2 + \mu^2) - n \left(\frac{\sigma^2}{n} + \mu^2 \right) = n(\sigma^2 + \mu^2) - n \left(\frac{\sigma^2}{n} + \mu^2 \right) = (n-1)\sigma^2 \quad (6.5.26)$$

Of course, the square root of the sample variance is the *sample standard deviation*, denoted S .

$\mathbb{E}(S) \leq \sigma$. Thus, S is a negatively biased estimator than tends to underestimate σ .

Proof

The proof is exactly the same as for the [special standard variance](#).

$S^2 \rightarrow \sigma^2$ as $n \rightarrow \infty$ with probability 1.

Proof

This follows from the strong law of large numbers. Recall again that

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n X_i^2 - \frac{n}{n-1} M^2 = \frac{n}{n-1} [M(\mathbf{X}^2) - M^2(\mathbf{X})] \quad (6.5.27)$$

But with probability 1, $M(\mathbf{X}^2) \rightarrow \sigma^2 + \mu^2$ as $n \rightarrow \infty$ and $M^2(\mathbf{X}) \rightarrow \mu^2$ as $n \rightarrow \infty$.

Since S^2 is an unbiased estimator of σ^2 , the variance of S^2 is the *mean square error*, a measure of the quality of the estimator.

$$\text{var}(S^2) = \frac{1}{n} \left(\sigma_4 - \frac{n-3}{n-1} \sigma^4 \right).$$

Proof

Recall from the [result above](#) that

$$S^2 = \frac{1}{2n(n-1)} \sum_{i=1}^n \sum_{j=1}^n (X_i - X_j)^2 \quad (6.5.28)$$

Hence, using the bilinear property of covariance we have

$$\text{var}(S^2) = \text{cov}(S^2, S^2) = \frac{1}{4n^2(n-1)^2} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n \text{cov}[(X_i - X_j)^2, (X_k - X_l)^2] \quad (6.5.29)$$

We compute the covariances in this sum by considering disjoint cases:

- $\text{cov}[(X_i - X_j)^2, (X_k - X_l)^2] = 0$ if $i = j$ or $k = l$, and there are $2n^3 - n^2$ such terms.
- $\text{cov}[(X_i - X_j)^2, (X_k - X_l)^2] = 0$ if i, j, k, l are distinct, and there are $n(n-1)(n-2)(n-3)$ such terms.
- $\text{cov}[(X_i - X_j)^2, (X_k - X_l)^2] = 2\sigma_4 + 2\sigma^4$ if $i \neq j$ and $\{k, l\} = \{i, j\}$, and there are $2n(n-1)$ such terms.
- $\text{cov}[(X_i - X_j)^2, (X_k - X_l)^2] = \sigma_4 - \sigma^4$ if $i \neq j, k \neq l$ and $\#\{i, j\} \cap \{k, l\} = 1$, and there are $4n(n-1)(n-2)$ such terms.

Substituting gives the result.

Note that $\text{var}(S^2) \rightarrow 0$ as $n \rightarrow \infty$, and hence S^2 is a consistent estimator of σ^2 . On the other hand, it's not surprising that the variance of the standard sample variance (where we assume that μ is unknown) is greater than the variance of the special standard variance (in which we assume μ is known).

$$\text{var}(S^2) > \text{var}(W^2).$$

Proof

From the [formula above](#) for the variance of W^2 , the previous result for the variance of S^2 , and simple algebra,

$$\text{var}(S^2) - \text{var}(W^2) = \frac{2}{n(n-1)} \sigma^4 \quad (6.5.30)$$

Note however that the difference goes to 0 as $n \rightarrow \infty$.

Next we compute the covariance between the sample mean and the sample variance.

The covariance and correlation between the sample mean and sample variance are

1. $\text{cov}(M, S^2) = \sigma_3/n$
2. $\text{cor}(M, S^2) = \frac{\sigma_3}{\sigma \sqrt{\sigma_4 - \sigma^4(n-3)/(n-1)}}$

Proof

1. Recall again that

$$M = \frac{1}{n} \sum_{i=1}^n X_i, \quad S^2 = \frac{1}{2n(n-1)} \sum_{j=1}^n \sum_{k=1}^n (X_j - X_k)^2 \quad (6.5.31)$$

Hence, using the bilinear property of covariance we have

$$\text{cov}(M, S^2) = \frac{1}{2n^2(n-1)} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \text{cov}[X_i, (X_j - X_k)^2] \quad (6.5.32)$$

We compute the covariances in this sum by considering disjoint cases:

- $\text{cov}[X_i, (X_j - X_k)^2] = 0$ if $j = k$, and there are n^2 such terms.
- $\text{cov}[X_i, (X_j - X_k)^2] = 0$ if i, j, k are distinct, and there are $n(n-1)(n-2)$ such terms.
- $\text{cov}[X_i, (X_j - X_k)^2] = \sigma_3$ if $j \neq k$ and $i \in \{j, k\}$, and there are $2n(n-1)$ such terms.

Substituting gives the result.

2. This follows from part(a), the result above on the variance of S^2 , and $\text{var}(M) = \sigma^2/n$.

In particular, note that $\text{cov}(M, S^2) = \text{cov}(M, W^2)$. Again, the sample mean and variance are uncorrelated if $\sigma_3 = 0$ so that $\text{skew}(X) = 0$. Our last result gives the covariance and correlation between the special sample variance and the standard one. Curiously, the covariance the same as the variance of the special sample variance.

The covariance and correlation between W^2 and S^2 are

$$\begin{aligned} 1. \text{cov}(W^2, S^2) &= (\sigma_4 - \sigma^4)/n \\ 2. \text{cor}(W^2, S^2) &= \sqrt{\frac{\sigma_4 - \sigma^4}{\sigma_4 - \sigma^4(n-3)/(n-1)}} \end{aligned}$$

Proof

1. Recall again that

$$W^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2, \quad S^2 = \frac{1}{2n(n-1)} \sum_{j=1}^n \sum_{k=1}^n (X_j - X_k)^2 \quad (6.5.33)$$

so by the bilinear property of covariance we have

$$\text{cov}(W^2, S^2) = \frac{1}{2n^2(n-1)} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \text{cov}[(X_i - \mu)^2, (X_j - X_k)^2] \quad (6.5.34)$$

Once again, we compute the covariances in this sum by considering disjoint cases:

- $\text{cov}[(X_i - \mu)^2, (X_j - X_k)^2] = 0$ if $j = k$, and there are n^2 such terms.
- $\text{cov}[(X_i - \mu)^2, (X_j - X_k)^2] = 0$ if i, j, k are distinct, and there are $n(n-1)(n-2)$ such terms.
- $\text{cov}[(X_i - \mu)^2, (X_j - X_k)^2] = \sigma_4 - \sigma^4$ if $j \neq k$ and $i \in \{j, k\}$, and there are $2n(n-1)$ such terms.

Substituting gives the results.

2. This follows from part (a) and the formulas above for the variance of W^2 and the variance of V^2

Note that $\text{cor}(W^2, S^2) \rightarrow 1$ as $n \rightarrow \infty$, not surprising since with probability 1, $S^2 \rightarrow \sigma^2$ and $W^2 \rightarrow \sigma^2$ as $n \rightarrow \infty$.

A particularly important special case occurs when the sampling distribution is normal. This case is explored in the section on Special Properties of Normal Samples.

Exercises

Basic Properties

Suppose that x is the temperature (in degrees Fahrenheit) for a certain type of electronic component after 10 hours of operation. A sample of 30 components has mean 113° and standard deviation 18° .

1. Classify x by type and level of measurement.
2. Find the sample mean and standard deviation if the temperature is converted to degrees Celsius. The transformation is $y = \frac{5}{9}(x - 32)$.

Answer

1. continuous, interval
2. $m = 45^\circ$, $s = 10^\circ$

Suppose that x is the length (in inches) of a machined part in a manufacturing process. A sample of 50 parts has mean 10.0 and standard deviation 2.0.

1. Classify x by type and level of measurement.
2. Find the sample mean if length is measured in centimeters. The transformation is $y = 2.54x$.

Answer

1. continuous, ratio
2. $m = 25.4$, $s = 5.08$

Professor Moriarity has a class of 25 students in her section of Stat 101 at Enormous State University (ESU). The mean grade on the first midterm exam was 64 (out of a possible 100 points) and the standard deviation was 16. Professor Moriarity thinks the grades are a bit low and is considering various transformations for increasing the grades. In each case below give the mean and standard deviation of the transformed grades, or state that there is not enough information.

1. Add 10 points to each grade, so the transformation is $y = x + 10$.
2. Multiply each grade by 1.2, so the transformation is $z = 1.2x$
3. Use the transformation $w = 10\sqrt{x}$. Note that this is a non-linear transformation that curves the grades greatly at the low end and very little at the high end. For example, a grade of 100 is still 100, but a grade of 36 is transformed to 60.

One of the students did not study at all, and received a 10 on the midterm. Professor Moriarity considers this score to be an outlier.

4. Find the mean and standard deviation if this score is omitted.

Answer

1. $m = 74, s = 16$
2. $m = 76.8, s = 19.2$
3. Not enough information
4. $m = 66.25, s = 11.62$

Computational Exercises

All statistical software packages will compute means, variances and standard deviations, draw dotplots and histograms, and in general perform the numerical and graphical procedures discussed in this section. For real statistical experiments, particularly those with large data sets, the use of statistical software is essential. On the other hand, there is some value in performing the computations by hand, with small, artificial data sets, in order to master the concepts and definitions. In this subsection, do the computations and draw the graphs with minimal technological aids.

Suppose that x is the number of math courses completed by an ESU student. A sample of 10 ESU students gives the data $\mathbf{x} = (3, 1, 2, 0, 2, 4, 3, 2, 1, 2)$

1. Classify x by type and level of measurement.
2. Sketch the dotplot.
3. Construct a table with rows corresponding to cases and columns corresponding to i , x_i , $x_i - m$, and $(x_i - m)^2$. Add rows at the bottom in the i column for totals and means.

Answer

1. discrete, ratio

3. i	x_i	$x_i - m$	$(x_i - m)^2$
1	3	1	1
2	1	-1	1
3	2	0	0
4	0	-2	4
5	2	0	0
6	4	2	4
7	3	1	1
8	2	0	0
9	1	-1	1
Total	20	0	14
Mean	2	0	14/9

i	x_i	$x_i - m$	$(x_i - m)^2$
10	2	0	0
Total	20	0	14
Mean	2	0	14/9

Suppose that a sample of size 12 from a discrete variable x has empirical density function given by $f(-2) = 1/12$, $f(-1) = 1/4$, $f(0) = 1/3$, $f(1) = 1/6$, $f(2) = 1/6$.

1. Sketch the graph of f .
2. Compute the sample mean and variance.
3. Give the sample values, ordered from smallest to largest.

Answer

2. $m = 1/12$, $s^2 = 203/121$
3. $(-2, -1, -1, -1, 0, 0, 0, 0, 1, 1, 2, 2)$

The following table gives a frequency distribution for the commuting distance to the math/stat building (in miles) for a sample of ESU students.

Class	Freq	Rel Freq	Density	Cum Freq	Cum Rel Freq	Midpoint
(0, 2]	6					
(2, 6]	16					
(6, 10]	18					
(10, 20]	10					
Total						

1. Complete the table
2. Sketch the density histogram
3. Sketch the cumulative relative frequency ogive.
4. Compute an approximation to the mean and standard deviation.

Answer

1.	Class	Freq	Rel Freq	Density	Cum Freq	Cum Rel Freq	Midpoint
	(0, 2]	6	0.12	0.06	6	0.12	1
	(2, 6]	16	0.32	0.08	22	0.44	4
	(6, 10]	18	0.36	0.09	40	0.80	8
	(10, 20]	10	0.20	0.02	50	1	15
	Total	50	1				

4. $m = 7.28$, $s = 4.549$

Error Function Exercises

In the error function app, select root mean square error. As you add points, note the shape of the graph of the error function, the value that minimizes the function, and the minimum value of the function.

In the error function app, select mean absolute error. As you add points, note the shape of the graph of the error function, the values that minimize the function, and the minimum value of the function.

Suppose that our data vector is $(2, 1, 5, 7)$. Explicitly give mae as a piecewise function and sketch its graph. Note that

1. All values of $a \in [2, 5]$ minimize mae.
2. mae is not differentiable at $a \in \{1, 2, 5, 7\}$.

Suppose that our data vector is $(3, 5, 1)$. Explicitly give mae as a piecewise function and sketch its graph. Note that

1. mae is minimized at $a = 3$.
2. mae is not differentiable at $a \in \{1, 3, 5\}$.

Simulation Exercises

Many of the apps in this project are simulations of experiments with a basic random variable of interest. When you run the simulation, you are performing independent replications of the experiment. In most cases, the app displays the standard deviation of the distribution, both numerically in a table and graphically as the radius of the blue, horizontal bar in the graph box. When you run the simulation, the sample standard deviation is also displayed numerically in the table and graphically as the radius of the red horizontal bar in the graph box.

In the binomial coin experiment, the random variable is the number of heads. For various values of the parameters n (the number of coins) and p (the probability of heads), run the simulation 1000 times and compare the sample standard deviation to the distribution standard deviation.

In the simulation of the matching experiment, the random variable is the number of matches. For selected values of n (the number of balls), run the simulation 1000 times and compare the sample standard deviation to the distribution standard deviation.

Run the simulation of the gamma experiment 1000 times for various values of the rate parameter r and the shape parameter k . Compare the sample standard deviation to the distribution standard deviation.

Probability Exercises

Suppose that X has probability density function $f(x) = 12x^2(1-x)$ for $0 \leq x \leq 1$. The distribution of X is a member of the beta family. Compute each of the following

1. $\mu = \mathbb{E}(X)$
2. $\sigma^2 = \text{var}(X)$
3. $d_3 = \mathbb{E}[(X - \mu)^3]$
4. $d_4 = \mathbb{E}[(X - \mu)^4]$

Answer

1. $3/5$
2. $1/25$
3. $-2/875$
4. $33/8750$

Suppose now that $(X_1, X_2, \dots, X_{10})$ is a random sample of size 10 from the beta distribution in the previous problem. Find each of the following:

1. $\mathbb{E}(M)$
2. $\text{var}(M)$
3. $\mathbb{E}(W^2)$
4. $\text{var}(W^2)$
5. $\mathbb{E}(S^2)$
6. $\text{var}(S^2)$

7. $\text{cov}(M, W^2)$
8. $\text{cov}(M, S^2)$
9. $\text{cov}(W^2, S^2)$

Answer

1. $3/5$
2. $1/250$
3. $1/25$
4. $19/87500$
5. $1/25$
6. $199/787500$
7. $-2/8750$
8. $-2/8750$
9. $19/87500$

Suppose that X has probability density function $f(x) = \lambda e^{-\lambda x}$ for $0 \leq x < \infty$, where $\lambda > 0$ is a parameter. Thus X has the exponential distribution with rate parameter λ . Compute each of the following

1. $\mu = \mathbb{E}(X)$
2. $\sigma^2 = \text{var}(X)$
3. $d_3 = \mathbb{E}[(X - \mu)^3]$
4. $d_4 = \mathbb{E}[(X - \mu)^4]$

Answer

1. $1/\lambda$
2. $1/\lambda^2$
3. $2/\lambda^3$
4. $9/\lambda^4$

Suppose now that (X_1, X_2, \dots, X_5) is a random sample of size 5 from the exponential distribution in the previous problem. Find each of the following:

1. $\mathbb{E}(M)$
2. $\text{var}(M)$
3. $\mathbb{E}(W^2)$
4. $\text{var}(W^2)$
5. $\mathbb{E}(S^2)$
6. $\text{var}(S^2)$
7. $\text{cov}(M, W^2)$
8. $\text{cov}(M, S^2)$
9. $\text{cov}(W^2, S^2)$

Answer

1. $1/\lambda$
2. $1/5\lambda^2$
3. $1/\lambda^2$
4. $8/5\lambda^4$
5. $1/\lambda^2$
6. $17/10\lambda^4$
7. $2/5\lambda^3$
8. $2/5\lambda^3$
9. $8/5\lambda^4$

Recall that for an *ace-six flat die*, faces 1 and 6 have probability $\frac{1}{4}$ each, while faces 2, 3, 4, and 5 have probability $\frac{1}{8}$ each. Let X denote the score when an ace-six flat die is thrown. Compute each of the following:

1. $\mu = \mathbb{E}(X)$
2. $\sigma^2 = \text{var}(X)$
3. $d_3 = \mathbb{E}[(X - \mu)^3]$
4. $d_4 = \mathbb{E}[(X - \mu)^4]$

Answer

1. $7/2$
2. $15/4$
3. 0
4. $333/16$

Suppose now that an ace-six flat die is tossed 8 times. Find each of the following:

1. $\mathbb{E}(M)$
2. $\text{var}(M)$
3. $\mathbb{E}(W^2)$
4. $\text{var}(W^2)$
5. $\mathbb{E}(S^2)$
6. $\text{var}(S^2)$
7. $\text{cov}(M, W^2)$
8. $\text{cov}(M, S^2)$
9. $\text{cov}(W^2, S^2)$

Answer

1. $7/2$
2. $15/32$
3. $15/4$
4. $27/32$
5. $15/4$
6. $207/512$
7. 0
8. 0
9. $27/32$

Data Analysis Exercises

Statistical software should be used for the problems in this subsection.

Consider the petal length and species variables in Fisher's iris data.

1. Classify the variables by type and level of measurement.
2. Compute the sample mean and standard deviation, and plot a density histogram for petal length.
3. Compute the sample mean and standard deviation, and plot a density histogram for petal length by species.

Answers

1. petal length: continuous, ratio. species: discrete, nominal
2. $m = 37.8$, $s = 17.8$
3. $m(0) = 14.6$, $s(0) = 1.7$; $m(1) = 55.5$, $s(1) = 30.5$; $m(2) = 43.2$, $s(2) = 28.7$

Consider the erosion variable in the Challenger data set.

1. Classify the variable by type and level of measurement.
2. Compute the mean and standard deviation
3. Plot a density histogram with the classes $[0, 5)$, $[5, 40)$, $[40, 50)$, $[50, 60)$.

Answer

1. continuous, ratio

2. $m = 7.7$, $s = 17.2$

Consider Michelson's velocity of light data.

1. Classify the variable by type and level of measurement.
2. Plot a density histogram.
3. Compute the sample mean and standard deviation.
4. Find the sample mean and standard deviation if the variable is converted to km/hr. The transformation is $y = x + 299\,000$

Answer

1. continuous, interval
3. $m = 852.4$, $s = 79.0$
4. $m = 299\,852.4$, $s = 79.0$

Consider Short's parallax of the sun data.

1. Classify the variable by type and level of measurement.
2. Plot a density histogram.
3. Compute the sample mean and standard deviation.
4. Find the sample mean and standard deviation if the variable is converted to degrees. There are 3600 seconds in a degree.
5. Find the sample mean and standard deviation if the variable is converted to radians. There are $\pi/180$ radians in a degree.

Answer

1. continuous, ratio
3. $m = 8.616$, $s = 0.749$
4. $m = 0.00239$, $s = 0.000208$
5. $m = 0.0000418$, $s = 0.00000363$

Consider Cavendish's density of the earth data.

1. Classify the variable by type and level of measurement.
2. Compute the sample mean and standard deviation.
3. Plot a density histogram.

Answer

1. continuous, ratio
2. $m = 5.448$, $s = 0.221$

Consider the M&M data.

1. Classify the variables by type and level of measurement.
2. Compute the sample mean and standard deviation for each color count variable.
3. Compute the sample mean and standard deviation for the total number of candies.
4. Plot a relative frequency histogram for the total number of candies.
5. Compute the sample mean and standard deviation, and plot a density histogram for the net weight.

Answer

1. color counts: discrete ratio. net weight: continuous ratio.
2. $m(r) = 9.60$, $s(r) = 4.12$; $m(g) = 7.40$, $s(g) = 0.57$; $m(bl) = 7.23$, $s(bl) = 4.35$; $m(o) = 6.63$, $s(o) = 3.69$;
 $m(y) = 13.77$, $s(y) = 6.06$; $m(br) = 12.47$, $s(br) = 5.13$
3. $m(n) = 57.10$, $s(n) = 2.4$
5. $m(w) = 49.215$, $s(w) = 1.522$

Consider the body weight, species, and gender variables in the Cicada data.

1. Classify the variables by type and level of measurement.
2. Compute the relative frequency function for species and plot the graph.

3. Compute the relative frequency function for gender and plot the graph.
4. Compute the sample mean and standard deviation, and plot a density histogram for body weight.
5. Compute the sample mean and standard deviation, and plot a density histogram for body weight by species.
6. Compute the sample mean and standard deviation, and plot a density histogram for body weight by gender.

Answer

1. body weight: continuous, ratio. species: discrete, nominal. gender: discrete, nominal.
2. $f(0) = 0.423$, $f(1) = 0.519$, $f(2) = 0.058$
3. $f(0) = 0.567$, $f(1) = 0.433$
4. $m = 0.180$, $s = 0.059$
5. $m(0) = 0.168$, $s(0) = 0.054$; $m(1) = 0.185$, $s(1) = 0.185$; $m(2) = 0.225$, $s(2) = 0.107$
6. $m(0) = 0.206$, $s(0) = 0.052$; $m(1) = 0.145$, $s(1) = 0.051$

Consider Pearson's height data.

1. Classify the variables by type and level of measurement.
2. Compute the sample mean and standard deviation, and plot a density histogram for the height of the father.
3. Compute the sample mean and standard deviation, and plot a density histogram for the height of the son.

Answer

1. continuous ratio
2. $m(x) = 67.69$, $s(x) = 2.75$
3. $m(y) = 68.68$, $s(y) = 2.82$

This page titled [6.5: The Sample Variance](#) is shared under a [CC BY 2.0](#) license and was authored, remixed, and/or curated by [Kyle Siegrist \(Random Services\)](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.