

## 6.8: Special Properties of Normal Samples

Random samples from normal distributions are the most important special cases of the topics in this chapter. As we will see, many of the results simplify significantly when the underlying sampling distribution is normal. In addition we will derive the distributions of a number of random variables constructed from normal samples that are of fundamental importance in inferential statistics.

### The One Sample Model

Suppose that  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  is a random sample from the normal distribution with mean  $\mu \in \mathbb{R}$  and standard deviation  $\sigma \in (0, \infty)$ . Recall that the term *random sample* means that  $\mathbf{X}$  is a sequence of independent, identically distributed random variables. Recall also that the normal distribution has probability density function

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \quad x \in \mathbb{R} \quad (6.8.1)$$

In the notation that we have used elsewhere in this chapter,  $\sigma_3 = \mathbb{E}[(X-\mu)^3] = 0$  (equivalently, the skewness of the normal distribution is 0) and  $\sigma_4 = \mathbb{E}[(X-\mu)^4] = 3\sigma^4$  (equivalently, the kurtosis of the normal distribution is 3). Since the sample (and in particular the sample size  $n$ ) is fixed in this subsection, it will be suppressed in the notation.

### The Sample Mean

First recall that the sample mean is

$$M = \frac{1}{n} \sum_{i=1}^n X_i \quad (6.8.2)$$

$M$  is normally distributed with mean and variance given by

1.  $\mathbb{E}(M) = \mu$
2.  $\text{var}(M) = \sigma^2/n$

**Proof**

This follows from basic properties of the normal distribution. Recall that the sum of independent normally distributed variables also has a normal distribution, and a linear transformation of a normally distributed variable is also normally distributed. The mean and variance of  $M$  hold in general, and were derived in the section on the Law of Large Numbers.

Of course, by the central limit theorem, the distribution of  $M$  is approximately normal, if  $n$  is large, even if the underlying sampling distribution is not normal. The *standard score* of  $M$  is given as follows:

$$Z = \frac{M - \mu}{\sigma/\sqrt{n}} \quad (6.8.3)$$

$Z$  has the standard normal distribution.

The standard score  $Z$  associated with the sample mean  $M$  plays a critical role in constructing interval estimates and hypothesis tests for the distribution mean  $\mu$  when the distribution standard deviation  $\sigma$  is known. The random variable  $Z$  will also appear in several derivations in this section.

### The Sample Variance

The main goal of this subsection is to show that certain multiples of the two versions of the sample variance that we have studied have *chi-square* distributions. Recall that the chi-square distribution with  $k \in \mathbb{N}_+$  degrees of freedom has probability density function

$$f(x) = \frac{1}{\Gamma(k/2)2^{k/2}} x^{k/2-1} e^{-x/2}, \quad 0 < x < \infty \quad (6.8.4)$$

and has mean  $k$  and variance  $2k$ . The moment generating function is

$$G(t) = \frac{1}{(1 - 2t)^{k/2}}, \quad -\infty < t < \frac{1}{2} \quad (6.8.5)$$

The most important result to remember is that the chi-square distribution with  $k$  degrees of freedom governs  $\sum_{i=1}^k Z_i^2$ , where  $(Z_1, Z_2, \dots, Z_k)$  is a sequence of independent, standard normal random variables.

Recall that if  $\mu$  is known, a natural estimator of the variance  $\sigma^2$  is the statistic

$$W^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \quad (6.8.6)$$

Although the assumption that  $\mu$  is known is almost always artificial,  $W^2$  is very easy to analyze and it will be used in some of the derivations below. Our first result is the distribution of a simple multiple of  $W^2$ . Let

$$U = \frac{n}{\sigma^2} W^2 \quad (6.8.7)$$

$U$  has the chi-square distribution with  $n$  degrees of freedom.

Proof

Note that

$$\frac{n}{\sigma^2} W^2 = \sum_{i=1}^n \left( \frac{X_i - \mu}{\sigma} \right)^2 \quad (6.8.8)$$

and the terms in the sum are independent standard normal variables.

The variable  $U$  associated with the statistic  $W^2$  plays a critical role in constructing interval estimates and hypothesis tests for the distribution standard deviation  $\sigma$  when the distribution mean  $\mu$  is known (although again, this assumption is usually not realistic).

The mean and variance of  $W^2$  are

1.  $\mathbb{E}(W^2) = \sigma^2$
2.  $\text{var}(W^2) = 2\sigma^4/n$

Proof

These results follow from the chi-square distribution of  $U$  and standard properties of expected value and variance.

As an estimator of  $\sigma^2$ , part (a) means that  $W^2$  is *unbiased* and part (b) means that  $W^2$  is *consistent*. Of course, these moment results are special cases of the general results obtained in the section on Sample Variance. In that section, we also showed that  $M$  and  $W^2$  are uncorrelated if the underlying sampling distribution has skewness 0 ( $\sigma_3 = 0$ ), as is the case here.

Recall now that the standard version of the sample variance is the statistic

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - M)^2 \quad (6.8.9)$$

The sample variance  $S^2$  is the usual estimator of  $\sigma^2$  when  $\mu$  is unknown (which is usually the case). We showed earlier that in general, the sample mean  $M$  and the sample variance  $S^2$  are uncorrelated if the underlying sampling distribution has skewness 0 ( $\sigma_3 = 0$ ). It turns out that if the sampling distribution is normal, these variables are in fact independent, a very important and useful property, and at first blush, a very surprising result since  $S^2$  appears to depend explicitly on  $M$ .

The sample mean  $M$  and the sample variance  $S^2$  are independent.

Proof

The proof is based on the vector of deviations from the sample mean. Let

$$\mathbf{D} = (X_1 - M, X_2 - M, \dots, X_{n-1} - M) \quad (6.8.10)$$

Note that  $S^2$  can be written as a function of  $\mathbf{D}$  since  $\sum_{i=1}^n (X_i - M) = 0$ . Next,  $M$  and the vector  $\mathbf{D}$  have a joint multivariate normal distribution. We showed earlier that  $M$  and  $X_i - M$  are uncorrelated for each  $i$ , and hence it follows that  $M$  and  $\mathbf{D}$  are independent. Finally, since  $S^2$  is a function of  $\mathbf{D}$ , it follows that  $M$  and  $S^2$  are independent.

We can now determine the distribution of a simple multiple of the sample variance  $S^2$ . Let

$$V = \frac{n-1}{\sigma^2} S^2 \quad (6.8.11)$$

$V$  has the chi-square distribution with  $n-1$  degrees of freedom.

Proof

We first show that  $U = V + Z^2$  where  $U$  is the chi-square variable associated with  $W^2$  and where  $Z$  is the standard score associated with  $M$ . To see this, note that

$$U = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - M + M - \mu)^2 \quad (6.8.12)$$

$$= \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - M)^2 + \frac{2}{\sigma^2} \sum_{i=1}^n (X_i - M)(M - \mu) + \frac{1}{\sigma^2} \sum_{i=1}^n (M - \mu)^2 \quad (6.8.13)$$

In the right side of the last equation, the first term is  $V$ . The second term is 0 because  $\sum_{i=1}^n (X_i - M) = 0$ . The last term is  $\frac{n}{\sigma^2} (M - \mu)^2 = Z^2$ . Now, from the result [above](#),  $U$  has the chi-square distribution with  $n$  degrees of freedom, and of course  $Z^2$  has the chi-square distribution with 1 degree of freedom. From the [previous result](#),  $V$  and  $Z^2$  are independent. Recall that the moment generating function of a sum of independent variables is the product of the MGFs. Thus, taking moment generating functions in the equation  $U = V + Z^2$  gives

$$\frac{1}{(1-2t)^{n/2}} = \mathbb{E}(e^{tV}) \frac{1}{(1-2t)^{1/2}}, \quad t < \frac{1}{2} \quad (6.8.14)$$

Solving we have  $\mathbb{E}(e^{tV}) = 1/(1-2t)^{(n-1)/2}$  for  $t < 1/2$  and therefore  $V$  has the chi-square distribution with  $n-1$  degrees of freedom.

The variable  $V$  associated with the statistic  $S^2$  plays a critical role in constructing interval estimates and hypothesis tests for the distribution standard deviation  $\sigma$  when the distribution mean  $\mu$  is unknown (almost always the case).

The mean and variance of  $S^2$  are

1.  $\mathbb{E}(S^2) = \sigma^2$
2.  $\text{var}(S^2) = 2\sigma^4/(n-1)$

Proof

These results follow from the chi-square distribution of  $V$  and standard properties of expected value and variance.

As before, these moment results are special cases of the general results obtained in the section on Sample Variance. Again, as an estimator of  $\sigma^2$ , part (a) means that  $S^2$  is unbiased, and part (b) means that  $S^2$  is consistent. Note also that  $\text{var}(S^2)$  is larger than  $\text{var}(W^2)$  (not surprising), by a factor of  $\frac{n}{n-1}$ .

In the special distribution simulator, select the chi-square distribution. Vary the degree of freedom parameter and note the shape and location of the probability density function and the mean, standard deviation bar. For selected values of the parameter, run the experiment 1000 times and compare the empirical density function and moments to the true probability density function and moments.

The covariance and correlation between the special sample variance and the standard sample variance are

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

Proof

These results follow from general results obtained in the section on sample variance and the fact that  $\sigma_4 = 3\sigma^4$ .

Note that the correlation does not depend on the parameters  $\mu$  and  $\sigma$ , and converges to 1 as  $n \rightarrow \infty$ ,

### The $T$ Variable

Recall that the Student  $t$  distribution with  $k \in \mathbb{N}_+$  degrees of freedom has probability density function

$$f(t) = C_k \left(1 + \frac{t^2}{k}\right)^{-(k+1)/2}, \quad t \in \mathbb{R} \quad (6.8.15)$$

where  $C_k$  is the appropriate normalizing constant. The distribution has mean 0 if  $k > 1$  and variance  $k/(k-2)$  if  $k > 2$ . In this subsection, the main point to remember is that the  $t$  distribution with  $k$  degrees of freedom is the distribution of

$$\frac{Z}{\sqrt{V/k}} \quad (6.8.16)$$

where  $Z$  has the standard normal distribution;  $V$  has the chi-square distribution with  $k$  degrees of freedom; and  $Z$  and  $V$  are independent. Our goal is to derive the distribution of

$$T = \frac{M - \mu}{S/\sqrt{n}} \quad (6.8.17)$$

Note that  $T$  is similar to the standard score  $Z$  associated with  $M$ , but with the sample standard deviation  $S$  replacing the distribution standard deviation  $\sigma$ . The variable  $T$  plays a critical role in constructing interval estimates and hypothesis tests for the distribution mean  $\mu$  when the distribution standard deviation  $\sigma$  is unknown.

As usual, let  $Z$  denote the standard score associated with the sample mean  $M$ , and let  $V$  denote the chi-square variable associated with the sample variance  $S^2$ . Then

$$T = \frac{Z}{\sqrt{V/(n-1)}} \quad (6.8.18)$$

and hence  $T$  has the student  $t$  distribution with  $n-1$  degrees of freedom.

#### Proof

In the definition of  $T$ , divide the numerator and denominator by  $\sigma/\sqrt{n}$ . The numerator is then  $(M - \mu)/(\sigma/\sqrt{n}) = Z$  and the denominator is  $S/\sigma = \sqrt{V/(n-1)}$ . Since  $Z$  and  $V$  are independent,  $Z$  has the standard normal distribution, and  $V$  has the chi-square distribution with  $n-1$  degrees of freedom, it follows that  $T$  has the student  $t$  distribution with  $n-1$  degrees of freedom.

In the special distribution simulator, select the  $t$  distribution. Vary the degree of freedom parameter and note the shape and location of the probability density function and the mean  $\pm$  standard deviation bar. For selected values of the parameters, run the experiment 1000 times and compare the empirical density function and moments to the distribution density function and moments.

## The Two Sample Model

Suppose that  $\mathbf{X} = (X_1, X_2, \dots, X_m)$  is a random sample of size  $m$  from the normal distribution with mean  $\mu \in \mathbb{R}$  and standard deviation  $\sigma \in (0, \infty)$ , and that  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)$  is a random sample of size  $n$  from the normal distribution with mean  $\nu \in \mathbb{R}$  and standard deviation  $\tau \in (0, \infty)$ . Finally, suppose that  $\mathbf{X}$  and  $\mathbf{Y}$  are independent. Of course, all of the results above in the [one sample model](#) apply to  $\mathbf{X}$  and  $\mathbf{Y}$  separately, but now we are interested in statistics that are helpful in inferential procedures that compare the two normal distributions. We will use the basic notation established above, but we will indicate the dependence on the sample.

The two-sample (or more generally the multi-sample model) occurs naturally when a basic variable in the statistical experiment is *filtered* according to one or more other variable (often nominal variables). For example, in the cicada data, the weights of the male cicadas and the weights of the female cicadas may fit observations from the two-sample normal model. The basic variable *weight* is

filtered by the variable *gender*. If weight is filtered by gender and species, we might have observations from the 6-sample normal model.

### The Difference in the Sample Means

We know from our work [above](#) that  $M(\mathbf{X})$  and  $M(\mathbf{Y})$  have normal distributions. Moreover, these sample means are independent because the underlying samples  $\mathbf{X}$  and  $\mathbf{Y}$  are independent. Hence, it follows from a basic property of the normal distribution that any linear combination of  $M(\mathbf{X})$  and  $M(\mathbf{Y})$  will be normally distributed as well. For inferential procedures that compare the distribution means  $\mu$  and  $\nu$ , the linear combination that is most important is the difference.

$M(\mathbf{X}) - M(\mathbf{Y})$  has a normal distribution with mean and variance given by

1.  $\mathbb{E}[(M(\mathbf{X}) - M(\mathbf{Y}))] = \mu - \nu$
2.  $\text{var}[(M(\mathbf{X}) - M(\mathbf{Y}))] = \sigma^2/m + \tau^2/n$

Hence the standard score

$$Z = \frac{[(M(\mathbf{X}) - M(\mathbf{Y}))] - (\mu - \nu)}{\sqrt{\sigma^2/m + \tau^2/n}} \quad (6.8.19)$$

has the standard normal distribution. This standard score plays a fundamental role in constructing interval estimates and hypothesis test for the difference  $\mu - \nu$  when the distribution standard deviations  $\sigma$  and  $\tau$  are known.

### Ratios of Sample Variances

Next we will show that the ratios of certain multiples of the sample variances (both versions) of  $\mathbf{X}$  and  $\mathbf{Y}$  have *F distributions*. Recall that the *F* distribution with  $j \in \mathbb{N}_+$  degrees of freedom in the numerator and  $k \in \mathbb{N}_+$  degrees of freedom in the denominator is the distribution of

$$\frac{U/j}{V/k} \quad (6.8.20)$$

where  $U$  has the chi-square distribution with  $j$  degrees of freedom;  $V$  has the chi-square distribution with  $k$  degrees of freedom; and  $U$  and  $V$  are independent. The *F* distribution is named in honor of Ronald Fisher and has probability density function

$$f(x) = C_{j,k} \frac{x^{(j-2)/2}}{[1 + (j/k)x]^{(j+k)/2}}, \quad 0 < x < \infty \quad (6.8.21)$$

where  $C_{j,k}$  is the appropriate normalizing constant. The mean is  $\frac{k}{k-2}$  if  $k > 2$ , and the variance is  $2\left(\frac{k}{k-2}\right)^2 \frac{j+k-2}{j(k-4)}$  if  $k > 4$ .

The random variable given below has the *F* distribution with  $m$  degrees of freedom in the numerator and  $n$  degrees of freedom in the denominator:

$$\frac{W^2(\mathbf{X})/\sigma^2}{W^2(\mathbf{Y})/\tau^2} \quad (6.8.22)$$

Proof

Using the notation in the subsection on the special sample variances, note that  $W^2(\mathbf{X})/\sigma^2 = U(\mathbf{X})/m$  and  $W^2(\mathbf{Y})/\tau^2 = U(\mathbf{Y})/n$ . The result then follows immediately since  $U(\mathbf{X})$  and  $U(\mathbf{Y})$  are independent chi-square variables with  $m$  and  $n$  degrees of freedom, respectively.

The random variable given below has the *F* distribution with  $m - 1$  degrees of freedom in the numerator and  $n - 1$  degrees of freedom in the denominator:

$$\frac{S^2(\mathbf{X})/\sigma^2}{S^2(\mathbf{Y})/\tau^2} \quad (6.8.23)$$

Proof

Using the notation in the subsection on the standard sample variances, note that  $S^2(\mathbf{X})/\sigma^2 = V(\mathbf{X})/(m-1)$  and  $S^2(\mathbf{Y})/\tau^2 = V(\mathbf{Y})/(n-1)$ . The result then follows immediately since  $V(\mathbf{X})$  and  $V(\mathbf{Y})$  are independent chi-square variables with  $m-1$  and  $n-1$  degrees of freedom, respectively.

These variables are useful for constructing interval estimates and hypothesis tests of the ratio of the standard deviations  $\sigma/\tau$ . The choice of the  $F$  variable depends on whether the means  $\mu$  and  $\nu$  are known or unknown. Usually, of course, the means are unknown and so the statistic in [above](#) is used.

In the special distribution simulator, select the  $F$  distribution. Vary the degrees of freedom parameters and note the shape and location of the probability density function and the mean  $\pm$  standard deviation bar. For selected values of the parameters, run the experiment 1000 times and compare the empirical density function and moments to the true distribution density function and moments.

### The $T$ Variable

Our final construction in the two sample normal model will result in a variable that has the student  $t$  distribution. This variable plays a fundamental role in constructing interval estimates and hypothesis test for the difference  $\mu - \nu$  when the distribution standard deviations  $\sigma$  and  $\tau$  are unknown. The construction requires the additional assumption that the distribution standard deviations are the same:  $\sigma = \tau$ . This assumption is reasonable if there is an inherent variability in the measurement variables that does not change even when different treatments are applied to the objects in the population.

Note first that the standard score associated with the difference in the sample means becomes

$$Z = \frac{[M(\mathbf{Y}) - M(\mathbf{X})] - (\nu - \mu)}{\sigma \sqrt{1/m + 1/n}} \quad (6.8.24)$$

To construct our desired variable, we first need an estimate of  $\sigma^2$ . A natural approach is to consider a weighted average of the sample variances  $S^2(\mathbf{X})$  and  $S^2(\mathbf{Y})$ , with the degrees of freedom as the weight factors (this is called the *pooled estimate* of  $\sigma^2$ ). Thus, let

$$S^2(\mathbf{X}, \mathbf{Y}) = \frac{(m-1)S^2(\mathbf{X}) + (n-1)S^2(\mathbf{Y})}{m+n-2} \quad (6.8.25)$$

The random variable  $V$  given below has the chi-square distribution with  $m+n-2$  degrees of freedom:

$$V = \frac{(m-1)S^2(\mathbf{X}) + (n-1)S^2(\mathbf{Y})}{\sigma^2} \quad (6.8.26)$$

Proof

The variable can be expressed as the sum of independent chi-square variables.

The variables  $M(\mathbf{Y}) - M(\mathbf{X})$  and  $S^2(\mathbf{X}, \mathbf{Y})$  are independent.

Proof

The following pairs of variables are independent:  $(M(\mathbf{X}), S^2(\mathbf{X}))$  and  $(M(\mathbf{Y}), S^2(\mathbf{Y}))$ ;  $M(\mathbf{X})$  and  $S^2(\mathbf{X})$ ;  $M(\mathbf{Y})$  and  $S^2(\mathbf{Y})$

The random variable  $T$  given below has the student  $t$  distribution with  $m+n-2$  degrees of freedom.

$$T = \frac{[M(\mathbf{Y}) - M(\mathbf{X})] - (\nu - \mu)}{S(\mathbf{X}, \mathbf{Y}) \sqrt{1/m + 1/n}} \quad (6.8.27)$$

Proof

The random variable can be written as  $Z/\sqrt{V/(m+n-2)}$  where  $Z$  is the standard normal variable given above and  $V$  is the chi-square variable given above. Moreover,  $Z$  and  $V$  are independent by the previous result.

## The Bivariate Sample Model

Suppose now that  $((X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n))$  is a random sample of size  $n$  from the bivariate normal distribution with means  $\mu \in \mathbb{R}$  and  $\nu \in \mathbb{R}$ , standard deviations  $\sigma \in (0, \infty)$  and  $\tau \in (0, \infty)$ , and correlation  $\rho \in [0, 1]$ . Of course,  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  is a random sample of size  $n$  from the normal distribution with mean  $\mu$  and standard deviation  $\sigma$ , and  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)$  is a random sample of size  $n$  from the normal distribution with mean  $\nu$  and standard deviation  $\tau$ , so the results above in the [one sample model](#) apply to  $\mathbf{X}$  and  $\mathbf{Y}$  individually. Thus our interest in this section is in the relation between various  $\mathbf{X}$  and  $\mathbf{Y}$  statistics and properties of sample covariance.

The bivariate (or more generally multivariate) model occurs naturally when considering two (or more) variables in the statistical experiment. For example, the heights of the fathers and the heights of the sons in Pearson's height data may well fit observations from the bivariate normal model.

In the notation that we have used previously, recall that  $\sigma^3 = \mathbb{E}[(X - \mu)^3] = 0$ ,  $\sigma_4 = \mathbb{E}[(X - \mu)^4] = 3\sigma^4$ ,  $\tau_3 = \mathbb{E}[(Y - \nu)^3] = 0$ ,  $\tau_4 = \mathbb{E}[(Y - \nu)^4] = 3\tau^4$ ,  $\delta = \text{cov}(X, Y) = \sigma\tau\rho$ , and  $\delta_2 = \mathbb{E}[(X - \mu)^2(Y - \nu)^2] = \sigma^2\tau^2(1 + 2\rho^2)$ .

The data vector  $((X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n))$  has a multivariate normal distribution.

1. The mean vector has a block form, with each block being  $(\mu, \nu)$ .
2. The variance-covariance matrix has a block-diagonal form, with each block being  $\begin{bmatrix} \sigma^2 & \sigma\tau\rho \\ \sigma\tau\rho & \tau^2 \end{bmatrix}$ .

Proof

This follows from standard results for the multivariate normal distribution. Of course the blocks in parts (a) and (b) are simply the mean and variance-covariance matrix of a single observation  $(X, Y)$ .

## Sample Means

$(M(\mathbf{X}), M(\mathbf{Y}))$  has a bivariate normal distribution. The covariance and correlation are

1.  $\text{cov}[M(\mathbf{X}), M(\mathbf{Y})] = \sigma\tau\rho/n$
2.  $\text{cor}[M(\mathbf{X}), M(\mathbf{Y})] = \rho$

Proof

The bivariate normal distribution follows from previous result since  $(M(\mathbf{X}), M(\mathbf{Y}))$  can be obtained from the data vector by a linear transformation. Parts (a) and (b) follow from our previous general results.

Of course, we know the individual means and variances of  $M(\mathbf{X})$  and  $M(\mathbf{Y})$  from the [one-sample model](#) above. Hence we know the complete distribution of  $(M(\mathbf{X}), M(\mathbf{Y}))$ .

## Sample Variances

The covariance and correlation between the special sample variances are

1.  $\text{cov}[W^2(\mathbf{X}), W^2(\mathbf{Y})] = 2\sigma^2\tau^2\rho^2/n$
2.  $\text{cor}[W^2(\mathbf{X}), W^2(\mathbf{Y})] = \rho^2$

Proof

These results follow from our previous general results and the special form of  $\delta_2$ ,  $\sigma_4$ , and  $\tau_4$ .

The covariance and correlation between the standard sample variances are

1.  $\text{cov}[S^2(\mathbf{X}), S^2(\mathbf{Y})] = 2\sigma^2\tau^2\rho^2/(n-1)$
2.  $\text{cor}[S^2(\mathbf{X}), S^2(\mathbf{Y})] = \rho^2$

Proof

These results follow from our previous general results and the special form of  $\delta$ ,  $\delta_2$ ,  $\sigma_4$ , and  $\tau_4$ .

## Sample Covariance

If  $\mu$  and  $\nu$  are known (again usually an artificial assumption), a natural estimator of the distribution covariance  $\delta$  is the special version of the sample covariance

$$W(\mathbf{X}, \mathbf{Y}) = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)(Y_i - \nu) \quad (6.8.28)$$

The mean and variance of  $W(\mathbf{X}, \mathbf{Y})$  are

1.  $\mathbb{E}[W(\mathbf{X}, \mathbf{Y})] = \sigma\tau\rho$
2.  $\text{var}[W(\mathbf{X}, \mathbf{Y})] = \sigma^2\tau^2(1 + \rho^2)/n$

Proof

These results follow from our previous general results and the special form of  $\delta$  and  $\delta_2$ .

If  $\mu$  and  $\nu$  are unknown (again usually the case), then a natural estimator of the distribution covariance  $\delta$  is the standard sample covariance

$$S(\mathbf{X}, \mathbf{Y}) = \frac{1}{n-1} \sum_{i=1}^n [X_i - M(\mathbf{X})][Y_i - M(\mathbf{Y})] \quad (6.8.29)$$

The mean and variance of the sample variance are

1.  $\mathbb{E}[S(\mathbf{X}, \mathbf{Y})] = \sigma\tau\rho$
2.  $\text{var}[S(\mathbf{X}, \mathbf{Y})] = \sigma^2\tau^2(1 + \rho^2)/(n-1)$

Proof

These results follow from our previous general results and the special form of  $\delta$  and  $\delta_2$ .

## Computational Exercises

We use the basic notation established above for samples  $\mathbf{X}$  and  $\mathbf{Y}$ , and for the statistics  $M$ ,  $W^2$ ,  $S^2$ ,  $T$ , and so forth.

Suppose that the net weights (in grams) of 25 bags of M&Ms form a random sample  $\mathbf{X}$  from the normal distribution with mean 50 and standard deviation 4. Find each of the following:

1. The mean and standard deviation of  $M$ .
2. The mean and standard deviation of  $W^2$ .
3. The mean and standard deviation of  $S^2$ .
4. The mean and standard deviation of  $T$ .
5.  $\mathbb{P}(M > 49, S^2 < 20)$ .
6.  $\mathbb{P}(-1 < T < 1)$ .

Answer

1. 50, 4/5
2. 16,  $16\sqrt{2}/5$
3. 16,  $8/\sqrt{3}$
4. 0,  $2\sqrt{3/11}$
5. 0.7291
6. 0.6727

Suppose that the SAT math scores from 16 Alabama students form a random sample  $\mathbf{X}$  from the normal distribution with mean 550 and standard deviation 20, while the SAT math scores from 25 Georgia students form a random sample  $\mathbf{Y}$  from the normal distribution with mean 540 and standard deviation 15. The two samples are independent. Find each of the following:

1. The mean and standard deviation of  $M(\mathbf{X})$ .
2. The mean and standard deviation of  $M(\mathbf{Y})$ .
3. The mean and standard deviation of  $M(\mathbf{X}) - M(\mathbf{Y})$ .



4.  $\mathbb{P}[M(\mathbf{X}) > M(\mathbf{Y})]$ .
5. The mean and standard deviation of  $S^2(\mathbf{X})$ .
6. The mean and standard deviation of  $S^2(\mathbf{Y})$ .
7. The mean and standard deviation of  $S^2(\mathbf{X})/S^2(\mathbf{Y})$
8.  $\mathbb{P}[S(\mathbf{X}) > S(\mathbf{Y})]$ .

Answer

1. 550, 5
2. 540, 3
3. 10,  $\sqrt{34}$
4. 0.9568
5. 400,  $80\sqrt{10/3}$
6. 225,  $75\sqrt{3}/2$
7.  $64/33$ ,  $\frac{32}{165}\sqrt{74/3}$
8. 0.8750

This page titled [6.8: Special Properties of Normal Samples](#) 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.