

4. Estimation in the Two-Sample Normal Model

In this section, we will study estimation problems in the two-sample normal model and in the bivariate normal model. This section parallels the section on [Tests in the Two-Sample Normal Model](#) in the Chapter on [Hypothesis Testing](#).

The Two-Sample Normal Model

Suppose that $X = (X_1, X_2, \dots, X_m)$ is a random sample of size m from the [normal distribution](#) with [mean](#) μ and [standard deviation](#) σ . and that $Y = (Y_1, Y_2, \dots, Y_n)$ is a random sample of size n from the normal distribution with mean ν and standard deviation τ . Moreover, suppose that the samples X and Y are [independent](#). Usually, the parameters are unknown, so the parameter space for our vector of parameters (μ, ν, σ, τ) is $\mathbb{R}^2 \times (0, \infty)^2$

This type of situation arises frequently when the random variables represent a measurement of interest for the objects of the population, and the samples correspond to two different treatments. For example, we might be interested in the blood pressure of a certain population of patients. The X vector records the blood pressures of a control sample, while the Y vector records the blood pressures of the sample receiving a new drug. Similarly, we might be interested in the yield of an acre of corn. The X vector records the yields of a sample receiving one type of fertilizer, while the Y vector records the yields of a sample receiving a different type of fertilizer.

Usually our interest is in a comparison of the parameters (either the mean or standard deviation) for the two sampling distributions. In this section we will construct confidence sets for the parameter vector, which will in turn lead to confidence intervals for the ratio of the standard deviations and for the difference of the means. As with previous estimation problems, the construction depends on finding appropriate [pivot variables](#).

For a generic sample $U = (U_1, U_2, \dots, U_k)$ from a distribution with mean a . we will use our standard notation for the [sample mean](#) and for the [sample variances](#). Recall also the [special properties](#) of these statistics when the sampling distribution is normal.

$$M(U) = \frac{1}{k} \sum_{i=1}^k U_i, \quad W^2(U, a) = \frac{1}{k} \sum_{i=1}^k (U_i - a)^2, \quad S^2(U) = \frac{1}{k-1} \sum_{i=1}^k (U_i - M(U))^2$$

Confidence Sets Based on an F Pivot Variable

1. Show that the following random variable has the [F distribution](#) with m degrees of freedom in the numerator and n degrees of freedom in the denominator. Show that the variable is a pivot variable for (μ, ν, σ, τ) and for $(\mu, \nu, \frac{\tau}{\sigma})$.

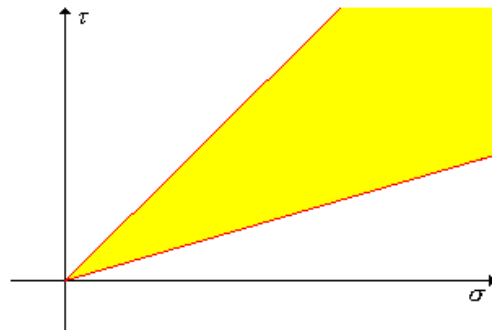
$$\frac{W^2(X, \mu) \tau^2}{W^2(Y, \nu) \sigma^2}$$

Now for $p \in (0, 1)$ and for $m > 0$ and $n > 0$, let $f_{m,n}(p)$ denote the [quantile](#) of order p for the F distribution with m degrees of freedom in the numerator and n degrees of freedom in the denominator. For selected values of m, n , and p , $f_{m,n}(p)$ can be computed using the [quantile applet](#) or from most statistical and mathematical software packages.

2. Use the pivot variable in [Exercise 1](#) to show that for any $\alpha \in (0, 1)$ and $p \in (0, 1)$, a $1 - \alpha$ confidence set for (μ, ν, σ, τ) is

$$\left\{ (\mu, \nu, \sigma, \tau) : f_{m,n}(\alpha - p\alpha) \frac{W^2(Y, \nu)}{W^2(X, \mu)} < \frac{\tau^2}{\sigma^2} < f_{m,n}(1 - p\alpha) \frac{W^2(Y, \nu)}{W^2(X, \mu)} \right\}$$

It's hard to get much insight about the structure of this confidence set as subset of our four-dimensional parameter space. The set is unbounded, of course, since we are estimating four parameters with a single pivot variable. If μ and ν are known (admittedly an artificial assumption), then the set becomes a $1 - \alpha$ confidence set for (σ, τ) . This set is also unbounded, and has the general shape given in the picture below:



Note, however, that with μ and ν known, our confidence set gives a bounded confidence interval for $\frac{\tau^2}{\sigma^2}$, and by taking square roots, a bounded confidence interval for $\frac{\tau}{\sigma}$. As always, the number p controls the proportion of α in the right tail of the distribution of our pivot variable (so that $1 - p$ is the proportion in the left tail). As usual, the most important special cases are

- $p = 0$. When μ and ν are known, the confidence set gives the $1 - \alpha$ lower bound for $\frac{\tau^2}{\sigma^2}$
- $p = 1$. When μ and ν are known, the confidence set gives the $1 - \alpha$ upper bound for $\frac{\tau^2}{\sigma^2}$
- $p = \frac{1}{2}$. When μ and ν are known, the confidence set gives the equal-tailed $1 - \alpha$ confidence interval for $\frac{\tau^2}{\sigma^2}$

Confidence Sets Based on Another F Pivot Variable

This subsection will parallel the previous one, except that we will use the usual version of the sample variances instead of the special version.

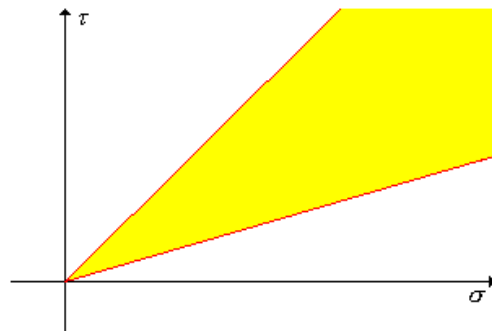
3. Show that the following random variable has the F distribution with $m - 1$ degrees of freedom in the numerator and $n - 1$ degrees of freedom in the denominator. Show that the variable is a pivot variable for (μ, ν, σ, τ) , for (σ, τ) , and for $\frac{\tau}{\sigma}$.

$$\frac{S^2(X) \tau^2}{S^2(Y) \sigma^2}$$

4. Use the pivot variable in Exercise 3 to show that for any $\alpha \in (0, 1)$ and $p \in (0, 1)$, a $1 - \alpha$ confidence set for (μ, ν, σ, τ) is

$$\left\{ (\mu, \nu, \sigma, \tau) : f_{m-1, n-1}(\alpha - p\alpha) \frac{S^2(Y)}{S^2(X)} < \frac{\tau^2}{\sigma^2} < f_{m-1, n-1}(1 - p\alpha) \frac{S^2(Y)}{S^2(X)} \right\}$$

By design, this confidence set gives no information at all about μ and ν . As a confidence set for (σ, τ) , the set has the same general shape as before:



Moreover, our confidence set gives a bounded confidence interval for $\frac{\tau^2}{\sigma^2}$, and by taking square roots, a bounded confidence interval for $\frac{\tau}{\sigma}$. As always, the number p controls the proportion of α in the right tail of the distribution of our pivot variable (so that $1 - p$ is the proportion of α in the left tail). As usual, the most important special cases are

- $p = 0$. The confidence set gives the $1 - \alpha$ lower bound for $\frac{\tau^2}{\sigma^2}$
- $p = 1$. The confidence set gives the $1 - \alpha$ upper bound for $\frac{\tau^2}{\sigma^2}$
- $p = \frac{1}{2}$. The confidence set gives the equal-tailed $1 - \alpha$ confidence interval for $\frac{\tau^2}{\sigma^2}$

Confidence Sets Based on a Normal Pivot Variable

Next we will study a pivot variable whose confidence sets are better in terms of the difference of the means.

5. Show that $M(Y) - M(X)$ has the normal distribution with mean $\nu - \mu$ and variance $\frac{\sigma^2}{m} + \frac{\tau^2}{n}$.

6. Show that the following random variable has the standard normal distribution. Show that the variable is a pivotal variable for (μ, ν, σ, τ) and for $(\nu - \mu, \sigma, \tau)$

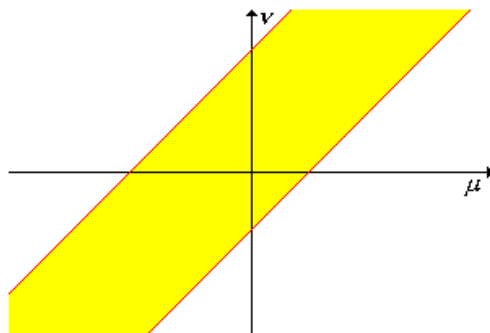
$$\frac{(M(Y) - M(X)) - (\nu - \mu)}{\sqrt{\frac{\sigma^2}{m} + \frac{\tau^2}{n}}}$$

As usual, for $p \in (0, 1)$, we will let $z(p)$ denote the quantile of order p for the standard normal distribution. Recall also that $z(p) = -z(1 - p)$

7. Use the pivot variable in [Exercise 6](#) to show that for any $\alpha \in (0, 1)$ and $p \in (0, 1)$, a $1 - \alpha$ confidence set for (μ, ν, σ, τ) is

$$\left\{ (\mu, \nu, \sigma, \tau) : (M(Y) - M(X)) - z(1 - \alpha p) \sqrt{\frac{\sigma^2}{m} + \frac{\tau^2}{n}} < \nu - \mu < (M(Y) - M(X)) - z(\alpha - \alpha p) \sqrt{\frac{\sigma^2}{m} + \frac{\tau^2}{n}} \right\}$$

Once again, it is hard to get much insight about this confidence set, as a subset of the four-dimensional parameter space. It is unbounded, of course, since we are estimating four parameters using a single pivot variable. If σ and τ are known, then the set is a $1 - \alpha$ confidence set for (μ, ν) . This set is also unbounded, but is easy to understand; the boundary curves are lines with slope 1, so the set has the general shape given in the picture below. Note that the width of the confidence strip is deterministic.



Moreover, if σ and τ are known, the confidence set gives bounded confidence *intervals* for the difference of the means $\nu - \mu$, and often this difference is our parameter of interest. The assumption that σ and τ are known is usually artificial, but less so than the assumption in our first subsection that μ and ν are known.

As always, the number p controls the proportion of α in the right tail of the distribution of our pivot variable (so that $1 - p$ is the proportion of α in the left tail). As usual, the most important special cases are

- $p = 0$. When σ and τ are known, the confidence set gives the $1 - \alpha$ upper bound for $\nu - \mu$
- $p = 1$. When σ and τ are known, the confidence set gives the $1 - \alpha$ lower bound for $\nu - \mu$
- $p = \frac{1}{2}$. When σ and τ are known, the confidence set gives the equal-tailed $1 - \alpha$ confidence interval for $\nu - \mu$

Confidence Sets Based on a Student t Pivot Variable

Our last construction will produce a pivot variable that is useful for estimating the difference of the means $\nu - \mu$ without needing to know the standard deviations σ and τ . However, there is a cost; we will assume that the standard deviations are the same, $\sigma = \tau$, but the common value is unknown. 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.

8. Show that the normal pivot variable in [Exercise 6](#) becomes

$$\frac{(M(Y) - M(X)) - (\nu - \mu)}{\sigma \sqrt{\frac{1}{m} + \frac{1}{n}}}$$

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

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

9. Show that the following variable has the [chi-square distribution](#) with $m + n - 2$ degrees of freedom. *Hint:* Express the variable as the sum of independent chi-square variables.

$$\frac{(m - 1)S^2(X) + (n - 1)S^2(Y)}{\sigma^2}$$

10. Show that $M(Y) - M(X)$ and $S^2(X, Y)$ are independent. *Hint:* Show or recall independence for each of the following pairs of variables:

- a. $(M(X), S(X))$ and $(M(Y), S(Y))$
- b. $M(X)$ and $S(X)$
- c. $M(Y)$ and $S(Y)$

11. Show that the following random variable has the [t distribution](#) with $m + n - 2$ degrees of freedom. Show that the variable is a pivot variable for (μ, ν, σ, τ) , for (μ, ν) , and for $\nu - \mu$.

$$\frac{(M(Y) - M(X)) - (\nu - \mu)}{S(X, Y) \sqrt{\frac{1}{m} + \frac{1}{n}}}$$

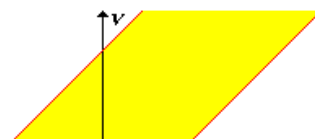
Hint: Show that the random variable can be written as $\frac{Z}{\sqrt{V/(m+n-2)}}$ where Z is the random variable in [Exercise 8](#) and V is the random variable in [Exercise 9](#). Moreover, Z and V are independent by [Exercise 10](#).

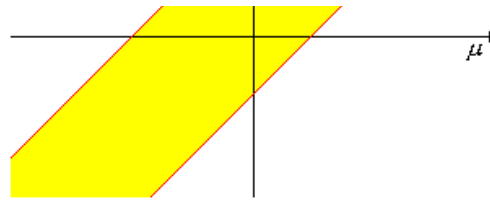
For $k > 0$ and $p \in (0, 1)$, let $t_k(p)$ denote the [quantile](#) of order p for the t distribution with k degrees of freedom. For selected values of k and p , values of $t_k(p)$ are given in the [table of the Student t distribution](#), from the [quantile applet](#), or from most statistical software packages. Recall also that, by symmetry, $t_k(p) = -t_k(1 - p)$

12. Use the pivot variable in [Exercise 11](#) to show that for any $\alpha \in (0, 1)$ and $p \in (0, 1)$, a $1 - \alpha$ confidence set for (μ, ν, σ, τ) is

$$\left\{ (\mu, \nu, \sigma, \tau) : (M(Y) - M(X)) - t_{m+n-2}(1 - \alpha p) S(X, Y) \sqrt{\frac{1}{m} + \frac{1}{n}} < \nu - \mu < (M(Y) - M(X)) - t_{m+n-2}(\alpha - \alpha p) S(X, Y) \sqrt{\frac{1}{m} + \frac{1}{n}} \right\}$$

By design, the confidence set gives no information about σ and τ . As a $1 - \alpha$ confidence set for (μ, ν) , the set has the same general shape as in the last subsection, except that the width of the strip is random.





Finally, of course, the confidence set gives bounded confidence *intervals* for the difference of the means $\nu - \mu$, and again, this difference is often our parameter of interest.

As always, the number p controls the proportion of α in the right tail of the distribution of our pivot variable (so that $1 - p$ is the proportion of α in the left tail). The most important special cases are

- $p = 0$. The confidence set gives the $1 - \alpha$ upper bound for $\nu - \mu$
- $p = 1$. The confidence set gives the $1 - \alpha$ lower bound for $\nu - \mu$
- $p = \frac{1}{2}$. The confidence set gives the equal-tailed $1 - \alpha$ confidence interval for $\nu - \mu$

Estimation in the Bivariate Normal Model

In this subsection, we consider a model that is superficially similar to the two-sample normal model, but is actually much simpler. Suppose 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](#) of (X, Y) with $\mathbb{E}(X) = \mu$, $\mathbb{E}(Y) = \nu$, $\text{var}(X) = \sigma^2$, $\text{var}(Y) = \tau^2$, and $\text{cov}(X, Y) = \delta$.

Thus, instead of a *pair of samples*, we have a *sample of pairs*. This type of model frequently arises in **before and after experiments**, in which a measurement of interest is recorded for a sample of n objects from the population, both before and after a treatment. For example, we could record the blood pressure of a sample of n patients, before and after the administration of a certain drug. As with the two-sample normal model, the interest is usually in estimating the difference of the means.

We will use our usual notation for the sample means and variances of $\mathbf{X} = (X_1, X_2, \dots, X_n)$ and $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)$. Recall also that the [sample covariance](#) of (\mathbf{X}, \mathbf{Y}) , is

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

13. Show that $\mathbf{Y} - \mathbf{X} = (Y_1 - X_1, Y_2 - X_2, \dots, Y_n - X_n)$ is a random sample of size n from the distribution of $Y - X$, which is normal with

- $\mathbb{E}(Y - X) = \nu - \mu$
- $\text{var}(Y - X) = \sigma^2 + \tau^2 - 2\delta$

14. Show that

- $M(\mathbf{Y} - \mathbf{X}) = M(\mathbf{Y}) - M(\mathbf{X})$
- $S^2(\mathbf{Y} - \mathbf{X}) = S^2(\mathbf{X}) + S^2(\mathbf{Y}) - 2S(\mathbf{X}, \mathbf{Y})$

The sample of differences $\mathbf{Y} - \mathbf{X}$ fits the normal model for a single variable. The section on [Estimation in the Normal Model](#) could be used to obtain confidence sets and intervals for the parameters $(\nu - \mu, \sigma^2 + \tau^2 - 2\delta)$.

15. Suppose that $\mathbf{X} = (X_1, X_2, \dots, X_n)$ and $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)$ are independent samples from normal distributions. This data fits both models--the two-sample normal model and the bivariate normal model. Which procedure would work better for estimating the

difference of means $\nu - \mu$?

- If the standard deviations σ and τ are known.
- If the standard deviations σ and τ are unknown.



Computational Exercises

16. A new drug is being developed to reduce a certain blood chemical. A sample of 36 patients are given a placebo while a sample of 49 patients are given the drug. Let X denote the measurement for a patient given the placebo and Y the measurement for a patient given the drug (in mg). The statistics are $m(\mathbf{X}) = 87$, $s(\mathbf{X}) = 4$, $m(\mathbf{Y}) = 63$, $s(\mathbf{Y}) = 6$.

- Compute the 90% confidence interval for $\frac{\tau}{\sigma}$.
- Assuming that $\sigma = \tau$, compute the 90% confidence interval for $\nu - \mu$.
- Based on (a), is the assumption that $\sigma = \tau$ reasonable?
- Based on (b), is the drug effective?



17. A company claims that an herbal supplement improves intelligence. A sample of 25 persons are given a standard IQ test before and after taking the supplement. Let X denote the IQ of a subject before taking the supplement and Y the IQ of the subject after the supplement. The before and after statistics are $m(\mathbf{X}) = 105$, $s(\mathbf{X}) = 13$, $m(\mathbf{Y}) = 110$, $s(\mathbf{Y}) = 17$, $s(\mathbf{X}, \mathbf{Y}) = 190$. Do you believe the company's claim?



18. In **Fisher's iris data**, let X denote consider the petal length of a Versicolor iris and Y the petal length of a Virginica iris.

- Compute the 90% confidence interval for $\frac{\tau}{\sigma}$.
- Assuming that $\sigma = \tau$, compute the 90% confidence interval for $\nu - \mu$.
- Based on (a), is the assumption that $\sigma = \tau$ reasonable?



19. A plant has two machines that produce a circular rod whose diameter (in cm) is critical. Let X denote the diameter of a rod from the first machine and Y the diameter of a rod from the second machine. A sample of 100 rods from the first machine as mean 10.3 and standard deviation 1.2. A sample of 100 rods from the second machine has mean 9.8 and standard deviation 1.6.

- Compute the 90% confidence interval for $\frac{\tau}{\sigma}$.
- Assuming that $\sigma = \tau$, compute the 90% confidence interval for $\nu - \mu$.
- Based on (a), is the assumption that $\sigma = \tau$ reasonable?

