One and two sample estimation problems (Chap 9)

From Walpole book

9.2 Statistical Inference

In Chapter 1, we discussed the general philosophy of formal statistical inference. **Statistical inference** consists of those methods by which one makes inferences or generalizations about a population. The trend today is to distinguish between the **classical method** of estimating a population parameter, whereby inferences are based strictly on information obtained from a random sample selected from the population, and the **Bayesian method**, which utilizes prior subjective knowledge about the probability distribution of the unknown parameters in conjunction with the information provided by the sample data. Throughout most of this chapter, we shall use classical methods to estimate unknown population parameters such as the mean, the proportion, and the variance by computing statistics from random samples and applying the theory of sampling distributions, much of which was covered in Chapter 8. Bayesian estimation will be discussed in Chapter 18.

Statistical inference may be divided into two major areas: **estimation** and **tests of hypotheses**. We treat these two areas separately, dealing with theory and applications of estimation in this chapter and hypothesis testing in Chapter 10. To distinguish clearly between the two areas, consider the following examples. A candidate for public office may wish to estimate the true proportion of voters favoring him by obtaining opinions from a random sample of 100 eligible voters. The fraction of voters in the sample favoring the candidate could be used as an estimate of the true proportion in the population of voters. A knowledge of the sampling distribution of a proportion enables one to establish the degree of accuracy of such an estimate. This problem falls in the area of estimation.

Now consider the case in which one is interested in finding out whether brand A floor wax is more scuff-resistant than brand B floor wax. He or she might hypothesize that brand A is better than brand B and, after proper testing, accept or reject this hypothesis. In this example, we do not attempt to estimate a parameter, but instead we try to arrive at a correct decision about a prestated hypothesis. Once again we are dependent on sampling theory and the use of data to provide us with some measure of accuracy for our decision.

Interval Estimation

Even the most efficient unbiased estimator is unlikely to estimate the population parameter exactly. It is true that estimation accuracy increases with large samples, but there is still no reason we should expect a **point estimate** from a given sample to be exactly equal to the population parameter it is supposed to estimate. There are many situations in which it is preferable to determine an interval within which we would expect to find the value of the parameter. Such an interval is called an **interval estimate**.

An interval estimate of a population parameter θ is an interval of the form $\hat{\theta}_L < \theta < \hat{\theta}_U$, where $\hat{\theta}_L$ and $\hat{\theta}_U$ depend on the value of the statistic $\hat{\Theta}$ for a particular sample and also on the sampling distribution of $\hat{\Theta}$. For example, a random sample of SAT verbal scores for students in the entering freshman class might produce an interval from 530 to 550, within which we expect to find the true average of all SAT verbal scores for the freshman class. The values of the endpoints, 530 and 550, will depend on the computed sample mean $\bar{x}$ and the sampling distribution of $\bar{X}$. As the sample size increases, we know that $\sigma^2_{\bar{X}} = \sigma^2/n$ decreases, and consequently our estimate is likely to be closer to the parameter μ, resulting in a shorter interval. Thus, the interval estimate indicates, by its length, the accuracy of the point estimate. An engineer will gain some insight into the population proportion defective by taking a sample and computing the *sample proportion defective*. But an interval estimate might be more informative.

Interpretation of Interval Estimates

Since different samples will generally yield different values of $\hat{\Theta}$ and, therefore, different values for $\hat{\theta}_L$ and $\hat{\theta}_U$, these endpoints of the interval are values of corresponding random variables $\hat{\Theta}_L$ and $\hat{\Theta}_U$. From the sampling distribution of $\hat{\Theta}$ we shall be able to determine $\hat{\Theta}_L$ and $\hat{\Theta}_U$ such that $P(\hat{\Theta}_L < \theta < \hat{\Theta}_U)$ is equal to any

positive fractional value we care to specify. If, for instance, we find $\hat{\Theta}_L$ and $\hat{\Theta}_U$ such that

$$P(\hat{\Theta}_L < \theta < \hat{\Theta}_U) = 1 - \alpha,$$

for $0 < \alpha < 1$, then we have a probability of $1-\alpha$ of selecting a random sample that will produce an interval containing θ. The interval $\hat{\theta}_L < \theta < \hat{\theta}_U$, computed from the selected sample, is called a $100(1-\alpha)\%$ **confidence interval**, the fraction $1-\alpha$ is called the **confidence coefficient** or the **degree of confidence**, and the endpoints, $\hat{\theta}_L$ and $\hat{\theta}_U$, are called the lower and upper **confidence limits**. Thus, when $\alpha = 0.05$, we have a 95% confidence interval, and when $\alpha = 0.01$, we obtain a wider 99% confidence interval. The wider the confidence interval is, the more confident we can be that the interval contains the unknown parameter. Of course, it is better to be 95% confident that the average life of a certain television transistor is between 6 and 7 years than to be 99% confident that it is between 3 and 10 years. Ideally, we prefer a short interval with a high degree of confidence. Sometimes, restrictions on the size of our sample prevent us from achieving short intervals without sacrificing some degree of confidence.

9.4 Single Sample: Estimating the Mean

Confidence Interval on μ, σ^2 Known

If $\bar{x}$ is the mean of a random sample of size n from a population with known variance σ^2, a $100(1-\alpha)\%$ confidence interval for μ is given by

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}},$$

where $z_{\alpha/2}$ is the z-value leaving an area of $\alpha/2$ to the right.

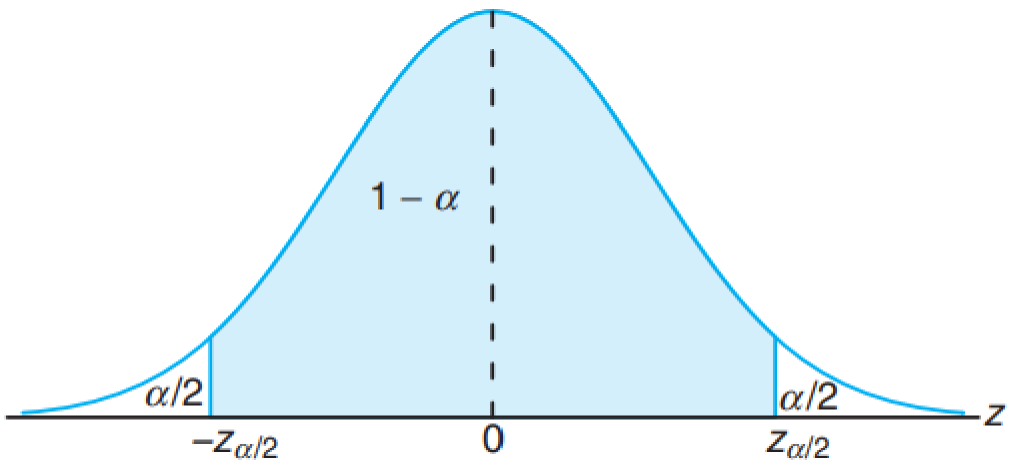

Figure 9.2: $P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$.

Example 9.2: The average zinc concentration recovered from a sample of measurements taken in 36 different locations in a river is found to be 2.6 grams per milliliter. Find the 95% and 99% confidence intervals for the mean zinc concentration in the river. Assume that the population standard deviation is 0.3 gram per milliliter.

***Solution*:** The point estimate of μ is $\bar{x} = 2.6$. The z-value leaving an area of 0.025 to the right, and therefore an area of 0.975 to the left, is $z_{0.025} = 1.96$ (Table A.3). Hence, the 95% confidence interval is

$$2.6 - (1.96)\left(\frac{0.3}{\sqrt{36}}\right) < \mu < 2.6 + (1.96)\left(\frac{0.3}{\sqrt{36}}\right),$$

which reduces to $2.50 < \mu < 2.70$. To find a 99% confidence interval, we find the z-value leaving an area of 0.005 to the right and 0.995 to the left. From Table A.3 again, $z_{0.005} = 2.575$, and the 99% confidence interval is

$$2.6 - (2.575)\left(\frac{0.3}{\sqrt{36}}\right) < \mu < 2.6 + (2.575)\left(\frac{0.3}{\sqrt{36}}\right),$$

or simply

$$2.47 < \mu < 2.73.$$

We now see that a longer interval is required to estimate μ with a higher degree of confidence.

One-Sided Confidence Bounds

One-Sided Confidence Bounds on μ, σ^2 Known

If $\bar{X}$ is the mean of a random sample of size n from a population with variance σ^2, the one-sided $100(1-\alpha)\%$ confidence bounds for μ are given by

$$\text{upper one-sided bound:} \quad \bar{x} + z_\alpha \sigma/\sqrt{n};$$
$$\text{lower one-sided bound:} \quad \bar{x} - z_\alpha \sigma/\sqrt{n}.$$

Example 9.4: In a psychological testing experiment, 25 subjects are selected randomly and their reaction time, in seconds, to a particular stimulus is measured. Past experience suggests that the variance in reaction times to these types of stimuli is 4 sec^2 and that the distribution of reaction times is approximately normal. The average time for the subjects is 6.2 seconds. Give an upper 95% bound for the mean reaction time.

***Solution*:** The upper 95% bound is given by

$$\begin{aligned}\bar{x} + z_\alpha \sigma/\sqrt{n} &= 6.2 + (1.645)\sqrt{4/25} = 6.2 + 0.658 \\ &= 6.858 \text{ seconds.}\end{aligned}$$

Hence, we are 95% confident that the mean reaction time is less than 6.858 seconds.

The Case of σ Unknown

Confidence Interval on μ, σ^2 Unknown

If $\bar{x}$ and s are the mean and standard deviation of a random sample from a normal population with unknown variance σ^2, a $100(1-\alpha)\%$ confidence interval for μ is

$$\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}},$$

where $t_{\alpha/2}$ is the t-value with $v = n-1$ degrees of freedom, leaving an area of $\alpha/2$ to the right.

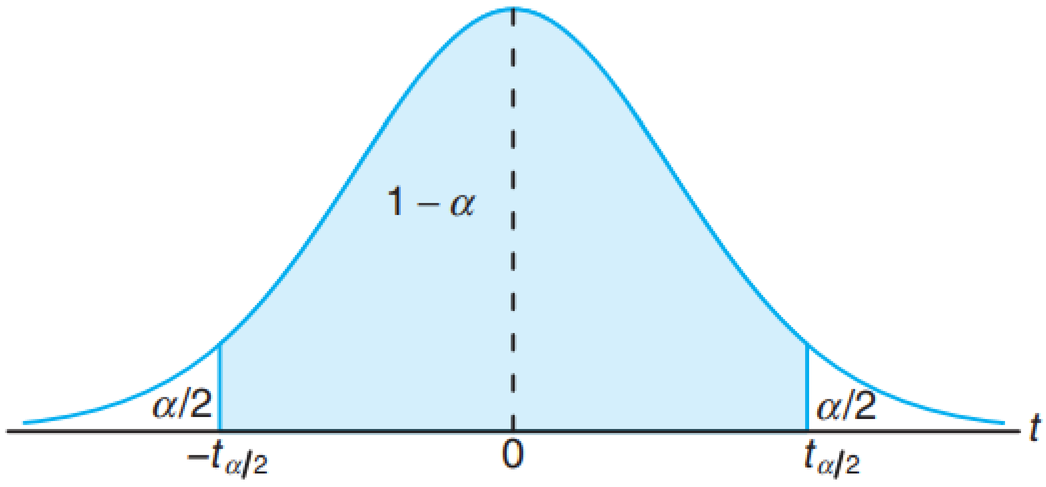

Figure 9.5: $P(-t_{\alpha/2} < T < t_{\alpha/2}) = 1 - \alpha$.

Computed one-sided confidence bounds for μ with σ unknown are as the reader would expect, namely

$$\bar{x} + t_\alpha \frac{s}{\sqrt{n}} \qquad \text{and} \qquad \bar{x} - t_\alpha \frac{s}{\sqrt{n}}.$$

They are the upper and lower $100(1-\alpha)\%$ bounds, respectively. Here t_α is the t-value having an area of α to the right.

Example 9.5: The contents of seven similar containers of sulfuric acid are 9.8, 10.2, 10.4, 9.8, 10.0, 10.2, and 9.6 liters. Find a 95% confidence interval for the mean contents of all such containers, assuming an approximately normal distribution.

***Solution*:** The sample mean and standard deviation for the given data are

$$\bar{x} = 10.0 \qquad \text{and} \qquad s = 0.283.$$

Using Table A.4, we find $t_{0.025} = 2.447$ for $v = 6$ degrees of freedom. Hence, the

95% confidence interval for μ is

$$10.0 - (2.447)\left(\frac{0.283}{\sqrt{7}}\right) < \mu < 10.0 + (2.447)\left(\frac{0.283}{\sqrt{7}}\right),$$

which reduces to $9.74 < \mu < 10.26$.

Table A.4 Critical Values of the t-Distribution

	α						
v	**0.40**	**0.30**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**
1	0.325	0.727	1.376	1.963	3.078	6.314	12.706
2	0.289	0.617	1.061	1.386	1.886	2.920	4.303
3	0.277	0.584	0.978	1.250	1.638	2.353	3.182
4	0.271	0.569	0.941	1.190	1.533	2.132	2.776
5	0.267	0.559	0.920	1.156	1.476	2.015	2.571
6	0.265	0.553	0.906	1.134	1.440	1.943	2.447
7	0.263	0.549	0.896	1.119	1.415	1.895	2.365
8	0.262	0.546	0.889	1.108	1.397	1.860	2.306
9	0.261	0.543	0.883	1.100	1.383	1.833	2.262
10	0.260	0.542	0.879	1.093	1.372	1.812	2.228
11	0.260	0.540	0.876	1.088	1.363	1.796	2.201
12	0.259	0.539	0.873	1.083	1.356	1.782	2.179
13	0.259	0.538	0.870	1.079	1.350	1.771	2.160
14	0.258	0.537	0.868	1.076	1.345	1.761	2.145
15	0.258	0.536	0.866	1.074	1.341	1.753	2.131
16	0.258	0.535	0.865	1.071	1.337	1.746	2.120
17	0.257	0.534	0.863	1.069	1.333	1.740	2.110
18	0.257	0.534	0.862	1.067	1.330	1.734	2.101
19	0.257	0.533	0.861	1.066	1.328	1.729	2.093
20	0.257	0.533	0.860	1.064	1.325	1.725	2.086
21	0.257	0.532	0.859	1.063	1.323	1.721	2.080
22	0.256	0.532	0.858	1.061	1.321	1.717	2.074
23	0.256	0.532	0.858	1.060	1.319	1.714	2.069
24	0.256	0.531	0.857	1.059	1.318	1.711	2.064
25	0.256	0.531	0.856	1.058	1.316	1.708	2.060
26	0.256	0.531	0.856	1.058	1.315	1.706	2.056
27	0.256	0.531	0.855	1.057	1.314	1.703	2.052
28	0.256	0.530	0.855	1.056	1.313	1.701	2.048
29	0.256	0.530	0.854	1.055	1.311	1.699	2.045
30	0.256	0.530	0.854	1.055	1.310	1.697	2.042
40	0.255	0.529	0.851	1.050	1.303	1.684	2.021
60	0.254	0.527	0.848	1.045	1.296	1.671	2.000
120	0.254	0.526	0.845	1.041	1.289	1.658	1.980
∞	0.253	0.524	0.842	1.036	1.282	1.645	1.960

Concept of a Large-Sample Confidence Interval

Often statisticians recommend that even when normality cannot be assumed, σ is unknown, and $n \geq 30$, s can replace σ and the confidence interval

$$\bar{x} \pm z_{\alpha/2}\frac{s}{\sqrt{n}}$$

Example 9.6: Scholastic Aptitude Test (SAT) mathematics scores of a random sample of 500 high school seniors in the state of Texas are collected, and the sample mean and standard deviation are found to be 501 and 112, respectively. Find a 99% confidence interval on the mean SAT mathematics score for seniors in the state of Texas.

***Solution*:** Since the sample size is large, it is reasonable to use the normal approximation. Using Table A.3, we find $z_{0.005} = 2.575$. Hence, a 99% confidence interval for μ is

$$501 \pm (2.575)\left(\frac{112}{\sqrt{500}}\right) = 501 \pm 12.9,$$

which yields $488.1 < \mu < 513.9$.

9.8 Two Samples: Estimating the Difference between Two Means

Confidence Interval for $\mu_1 - \mu_2$, σ_1^2 and σ_2^2 Known

If $\bar{x}_1$ and $\bar{x}_2$ are means of independent random samples of sizes n_1 and n_2 from populations with known variances σ_1^2 and σ_2^2, respectively, a $100(l - \alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is given by

$$(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}},$$

where $z_{\alpha/2}$ is the z-value leaving an area of $\alpha/2$ to the right.

Example 9.10: A study was conducted in which two types of engines, A and B, were compared. Gas mileage, in miles per gallon, was measured. Fifty experiments were conducted using engine type A and 75 experiments were done with engine type B. The gasoline used and other conditions were held constant. The average gas mileage was 36 miles per gallon for engine A and 42 miles per gallon for engine B. Find a 96% confidence interval on $\mu_B - \mu_A$, where μ_A and μ_B are population mean gas mileages for engines A and B, respectively. Assume that the population standard deviations are 6 and 8 for engines A and B, respectively.

***Solution*:** The point estimate of $\mu_B - \mu_A$ is $\bar{x}_B - \bar{x}_A = 42 - 36 = 6$. Using $\alpha = 0.04$, we find $z_{0.02} = 2.05$ from Table A.3. Hence, with substitution in the formula above, the 96% confidence interval is

$$6 - 2.05\sqrt{\frac{64}{75} + \frac{36}{50}} < \mu_B - \mu_A < 6 + 2.05\sqrt{\frac{64}{75} + \frac{36}{50}},$$

or simply $3.43 < \mu_B - \mu_A < 8.57$.

Pooled Estimate of Variance	$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}.$$

Confidence Interval for $\mu_1 - \mu_2$, $\sigma_1^2 = \sigma_2^2$ but Both Unknown	If $\bar{x}_1$ and $\bar{x}_2$ are the means of independent random samples of sizes n_1 and n_2, respectively, from approximately normal populations with unknown but equal variances, a $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is given by $$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}},$$ where s_p is the pooled estimate of the population standard deviation and $t_{\alpha/2}$ is the t-value with $v = n_1 + n_2 - 2$ degrees of freedom, leaving an area of $\alpha/2$ to the right.

Confidence Interval for $\mu_1 - \mu_2$, $\sigma_1^2 \neq \sigma_2^2$ and Both Unknown

If $\bar{x}_1$ and s_1^2 and $\bar{x}_2$ and s_2^2 are the means and variances of independent random samples of sizes n_1 and n_2, respectively, from approximately normal populations with unknown and unequal variances, an approximate $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is given by

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}},$$

where $t_{\alpha/2}$ is the t-value with

$$v = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{[(s_1^2/n_1)^2/(n_1-1)] + [(s_2^2/n_2)^2/(n_2-1)]}$$

degrees of freedom, leaving an area of $\alpha/2$ to the right.

Example 9.11: The article "Macroinvertebrate Community Structure as an Indicator of Acid Mine Pollution," published in the *Journal of Environmental Pollution*, reports on an investigation undertaken in Cane Creek, Alabama, to determine the relationship between selected physiochemical parameters and different measures of macroinvertebrate community structure. One facet of the investigation was an evaluation of the effectiveness of a numerical species diversity index to indicate aquatic degradation due to acid mine drainage. Conceptually, a high index of macroinvertebrate species diversity should indicate an unstressed aquatic system, while a low diversity index should indicate a stressed aquatic system.

Two independent sampling stations were chosen for this study, one located downstream from the acid mine discharge point and the other located upstream. For 12 monthly samples collected at the downstream station, the species diversity index had a mean value $\bar{x}_1 = 3.11$ and a standard deviation $s_1 = 0.771$, while 10 monthly samples collected at the upstream station had a mean index value $\bar{x}_2 = 2.04$ and a standard deviation $s_2 = 0.448$. Find a 90% confidence interval for the difference between the population means for the two locations, assuming that the populations are approximately normally distributed with equal variances.

Solution: Let μ_1 and μ_2 represent the population means, respectively, for the species diversity indices at the downstream and upstream stations. We wish to find a 90% confidence interval for $\mu_1 - \mu_2$. Our point estimate of $\mu_1 - \mu_2$ is

$$\bar{x}_1 - \bar{x}_2 = 3.11 - 2.04 = 1.07.$$

The pooled estimate, s_p^2, of the common variance, σ^2, is

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{(11)(0.771^2) + (9)(0.448^2)}{12+10-2} = 0.417.$$

Taking the square root, we obtain $s_p = 0.646$. Using $\alpha = 0.1$, we find in Table A.4 that $t_{0.05} = 1.725$ for $v = n_1 + n_2 - 2 = 20$ degrees of freedom. Therefore, the 90% confidence interval for $\mu_1 - \mu_2$ is

$$1.07 - (1.725)(0.646)\sqrt{\frac{1}{12} + \frac{1}{10}} < \mu_1 - \mu_2 < 1.07 + (1.725)(0.646)\sqrt{\frac{1}{12} + \frac{1}{10}},$$

which simplifies to $0.593 < \mu_1 - \mu_2 < 1.547$.

Example 9.12: A study was conducted by the Department of Zoology at the Virginia Tech to estimate the difference in the amounts of the chemical orthophosphorus measured at two different stations on the James River. Orthophosphorus was measured in milligrams per liter. Fifteen samples were collected from station 1, and 12 samples were obtained from station 2. The 15 samples from station 1 had an average orthophosphorus content of 3.84 milligrams per liter and a standard deviation of 3.07 milligrams per liter, while the 12 samples from station 2 had an average content of 1.49 milligrams per liter and a standard deviation of 0.80 milligram per liter. Find a 95% confidence interval for the difference in the true average orthophosphorus contents at these two stations, assuming that the observations came from normal populations with different variances.

***Solution*:** For station 1, we have $\bar{x}_1 = 3.84$, $s_1 = 3.07$, and $n_1 = 15$. For station 2, $\bar{x}_2 = 1.49$, $s_2 = 0.80$, and $n_2 = 12$. We wish to find a 95% confidence interval for $\mu_1 - \mu_2$.

Since the population variances are assumed to be unequal, we can only find an approximate 95% confidence interval based on the t-distribution with v degrees of freedom, where

$$v = \frac{(3.07^2/15 + 0.80^2/12)^2}{[(3.07^2/15)^2/14] + [(0.80^2/12)^2/11]} = 16.3 \approx 16.$$

Our point estimate of $\mu_1 - \mu_2$ is

$$\bar{x}_1 - \bar{x}_2 = 3.84 - 1.49 = 2.35.$$

Using $\alpha = 0.05$, we find in Table A.4 that $t_{0.025} = 2.120$ for $v = 16$ degrees of freedom. Therefore, the 95% confidence interval for $\mu_1 - \mu_2$ is

$$2.35 - 2.120\sqrt{\frac{3.07^2}{15} + \frac{0.80^2}{12}} < \mu_1 - \mu_2 < 2.35 + 2.120\sqrt{\frac{3.07^2}{15} + \frac{0.80^2}{12}},$$

which simplifies to $0.60 < \mu_1 - \mu_2 < 4.10$. Hence, we are 95% confident that the interval from 0.60 to 4.10 milligrams per liter contains the difference of the true average orthophosphorus contents for these two locations.

Confidence Interval for $\mu_D = \mu_1 - \mu_2$ for Paired Observations

If $\bar{d}$ and s_d are the mean and standard deviation, respectively, of the normally distributed differences of n random pairs of measurements, a $100(1-\alpha)\%$ confidence interval for $\mu_D = \mu_1 - \mu_2$ is

$$\bar{d} - t_{\alpha/2}\frac{s_d}{\sqrt{n}} < \mu_D < \bar{d} + t_{\alpha/2}\frac{s_d}{\sqrt{n}},$$

where $t_{\alpha/2}$ is the t-value with $v = n-1$ degrees of freedom, leaving an area of $\alpha/2$ to the right.

Example 9.13: A study published in *Chemosphere* reported the levels of the dioxin TCDD of 20 Massachusetts Vietnam veterans who were possibly exposed to Agent Orange. The TCDD levels in plasma and in fat tissue are listed in Table 9.1.

Find a 95% confidence interval for $\mu_1 - \mu_2$, where μ_1 and μ_2 represent the true mean TCDD levels in plasma and in fat tissue, respectively. Assume the distribution of the differences to be approximately normal.

Veteran	TCDD Levels in Plasma	TCDD Levels in Fat Tissue	d_i
1	2.5	4.9	−2.4
2	3.1	5.9	−2.8
3	2.1	4.4	−2.3
4	3.5	6.9	−3.4
5	3.1	7.0	−3.9
6	1.8	4.2	−2.4
7	6.0	10.0	−4.0
8	3.0	5.5	−2.5
9	36.0	41.0	−5.0
10	4.7	4.4	0.3
11	6.9	7.0	−0.1
12	3.3	2.9	0.4
13	4.6	4.6	0.0
14	1.6	1.4	0.2
15	7.2	7.7	−0.5
16	1.8	1.1	0.7
17	20.0	11.0	9.0
18	2.0	2.5	−0.5
19	2.5	2.3	0.2
20	4.1	2.5	1.6

$$\bar{d} - t_{\alpha/2}\frac{s_d}{\sqrt{n}} < \mu_D < \bar{d} + t_{\alpha/2}\frac{s_d}{\sqrt{n}},$$

$$\bar{d} = -0.87. \qquad s_d = 2.9773$$

Using $\alpha = 0.05$, we find in Table A.4 that $t_{0.025} = 2.093$ for $v = n - 1 = 19$ degrees of freedom. Therefore, the 95% confidence interval is

$$-0.8700 - (2.093)\left(\frac{2.9773}{\sqrt{20}}\right) < \mu_D < -0.8700 + (2.093)\left(\frac{2.9773}{\sqrt{20}}\right),$$

$$-2.2634 < \mu_D < 0.5234,$$

Summary of all estimation formulae

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}},$$

When σ is known and n is large where for small n, population must be normal

$$\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}},$$

When σ is unknown, s is known and n is small (population must be normal) population

$v = n - 1$ degrees of freedom.

For large n, t is replaced by z

$$(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}},$$

When σ_1 and σ_2 are known and n_1 *and* n_2 are large where for small n_1 *and* n_2, populations must be normal

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2}s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2}s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}},$$

When σ_1 and σ_2 are unknown but $\sigma_1 = \sigma_2$ and n_1 *and* n_2 are small (populations must be normal) where for large n_1 *and* n_2, t is replaced by z

$v = n_1 + n_2 - 2$ degrees of freedom,

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}.$$

Summary of all estimation formulae

$$(\bar{x}_1-\bar{x}_2)-t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}<\mu_1-\mu_2<(\bar{x}_1-\bar{x}_2)+t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}},$$

When σ_1 and σ_2 are unknown but $\sigma_1 \neq \sigma_2$ and n_1 *and* n_2 are small (populations must be normal) where for large n_1 *and* n_2, t is replaced by z

$$v=\frac{(s_1^2/n_1+s_2^2/n_2)^2}{[(s_1^2/n_1)^2/(n_1-1)]+[(s_2^2/n_2)^2/(n_2-1)]}$$

$$\bar{d}-t_{\alpha/2}\frac{s_d}{\sqrt{n}}<\mu_D<\bar{d}+t_{\alpha/2}\frac{s_d}{\sqrt{n}},$$

If $\bar{d}$ and s_d are the mean and standard deviation, respectively, of the normally distributed differences of n random pairs of measurements, a $100(1-\alpha)\%$ confidence interval for $\mu_D=\mu_1-\mu_2$ is

$$\bar{d}-t_{\alpha/2}\frac{s_d}{\sqrt{n}}<\mu_D<\bar{d}+t_{\alpha/2}\frac{s_d}{\sqrt{n}},$$

where $t_{\alpha/2}$ is the t-value with $v=n-1$ degrees of freedom, leaving an area of $\alpha/2$ to the right.

Note: For only lower or upper intervals, $\frac{\alpha}{2}$ is replaced by α in all formulas