Estimating a Population Mean from a Large Sample

Our text introduces us to an experiment conducted by Gray and Sauerbeck and reported in *The Sport Journal's* Winter 2005 issue. (See p. 255) A sample of 38 athletes participated in the study to evaluate the effectiveness of a speed training program. The decrease in sprint times (in seconds) after completing the program is recorded in the table below. A negative increase indicates the athlete's time after completing the program was higher than his time prior to entering the program.

Decrease in 40-Yard Sprint Times for 38 Athletes									
0.32	0.10	0.10	0.24	0.25	0.05	0.28	0.25	0.20	0.21
-0.01	0.34	0.30	0.09	0.05	0.00	0.04	0.17	0.00	0.10
0.15	0.30	0.02	0.12	0.14	0.10	0.08	0.50	0.36	
0.01	0.90	0.34	0.38	0.44	0.08	0.00	0.00	0.14	

To address the effectiveness of the training program, on the basis of our sample, we would like to estimate mean improvement in sprint times for all similar athletes who would participate in the speed training program. We will use the sample mean $\bar{x}$ of the data in the table above to estimate μ, the mean of all similar athletes' decrease in sprint times. The sample mean $\bar{x}$ represents a ***point estimator*** of the population mean μ.

According to the Central Limit Theorem the sampling distribution for the sample mean is approximately normal for large samples as shown below.

We determine an interval four standard deviations wide—from two standard deviations below the sample mean to two standard deviations above the mean.

Before we draw our sample of 38 athletes, what are the chances that the interval $\bar{x} \pm 2\sigma_{\bar{x}} \approx \bar{x} \pm \frac{2\sigma}{\sqrt{n}}$ will contain μ, the population mean?
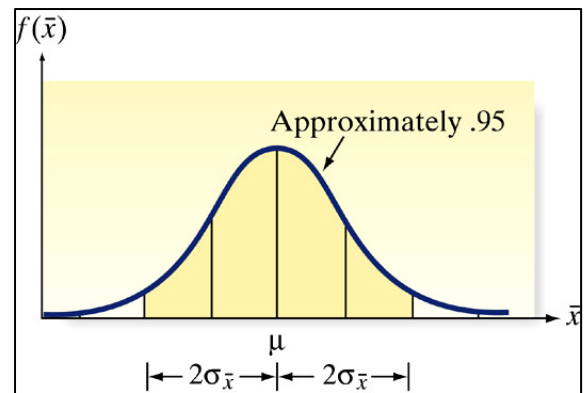
If the 38 measurements produce a value for $\bar{x}$ that lies between $\mu - 2\sigma_{\bar{x}}$ and $\mu + 2\sigma_{\bar{x}}$ then the interval $\bar{x} \pm 2\sigma_{\bar{x}}$ will contain μ; if $\bar{x}$ falls outside those boundaries, the interval $\mu \pm 2\sigma_{\bar{x}}$ will not contain μ. The area under the sampling distribution of $\bar{x}$ between the boundaries is approximately 0.95, the interval $\bar{x} \pm 2\sigma_{\bar{x}}$ will contain $\bar{\mu}$ with a probability of about 0.95.

MINITAB statistics for the data in the table above are shown below.

Descriptive Statistics: Decreased Time

```
Variable   N   N*    Mean   SE Mean    StDev
DecrTime   38   0   0.1879   0.0294   0.1814
```

We construct the interval $\bar{x} \pm 2\sigma_{\bar{x}}$ using the sample standard deviation s to approximate σ. So,

$$\bar{x} \pm 2\sigma_{\bar{x}} \approx 0.1879 \pm \frac{2\sigma}{\sqrt{38}} \approx 0.1879 \pm 2\left(\frac{0.1814}{6.1644}\right) \approx 0.1879 \pm 0.058854.$$

We can now estimate that the mean decrease in sprint time (sec) for athletes similar to those in the sample will fall in the interval from 0.1290 to 0.2468.

How sure are we that the true mean μ, is actually in the interval from 0.1290 to 0.2468? Of course we cannot be certain. Our confidence follows from the knowledge that if we draw repeated random samples of size 38 from a population of similar athletes and consider the interval $\bar{x} \pm 2\sigma_{\bar{x}}$ for each sample, about 95% of the intervals will contain μ. We have no way of knowing whether our sample interval is among the 95% that contain μ or is among the 5% that do not contain μ. The odds favor our sample interval being among the 95% that include μ. So, the interval from 0.1290 to 0.2468 gives us a reliable estimate of the mean expected improvement in sprint times for those who complete the training program.

The formula that tells us how to calculate an interval estimate on the basis of sample data is called an *interval estimator* or *confidence interval*. The probability 0.95 that is a measure of the confidence we place in our interval is called the *confidence coefficient*. The percent, 95%, is called the *confidence level* for the interval estimate.

Below is MINTAB output showing the 95% confidence interval for the speed training study. The figures differ slightly from our work above due to the number of significant digits carried.

T Confidence Intervals

Variable	N	Mean	StDev	SE Mean	95.0 % CI
DecrTime	38	0.1879	0.1814	0.0294	(0.1283, 0.2475)

Here is a histogram produced by MINITAB that shows the confidence interval.
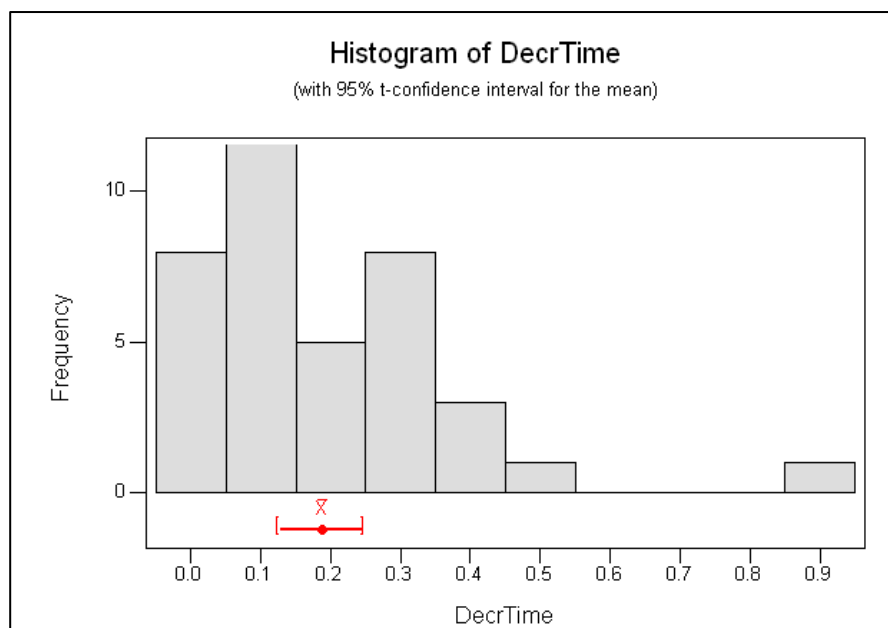
Suppose we want to use a confidence coefficient other than 0.95. We recall that 0.95 is 0.05 less than the area under the sampling distribution. That 0.05 is equally distributed between the two tails of the distribution with 0.025 in each tail. Suppose we want to construct a confidence interval with a confidence coefficient of 0.90. In this case we would place an area of 0.10/2 = 0.05 in each tail. We denote the z value such that an area of 0.10/2 = 0.05 lies to its right by $z_{0.10/2}$. The confidence interval with confidence coefficient 0.90 is $\bar{x} \pm z_{0.10/2}\, \sigma_{\bar{x}}$. In this example, our 90% confidence interval is

approximately $0.1879 \pm 1.65\left(\dfrac{0.1814}{\sqrt{38}}\right) \approx 0.1879 \pm 0.0486$. S our 90% confidence interval is about

(0.139, 0.237); MINITAB produces a 90% confidence interval of (0.1382, 0.2375).

Instructions for using MINITAB to construct a confidence interval are given on page 297 in our text. Instructions for using a TI graphing calculator to construct a confidence interval are given on page 262 in our text.

Summary

The ***confidence coefficient*** is equal to 1 - α, and is split between the two tails of the distribution.
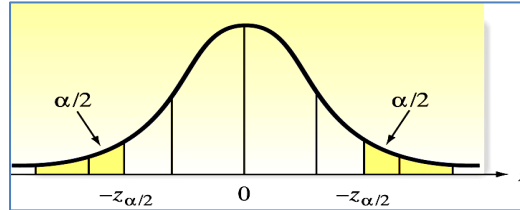
The large sample ***confidence interval*** is expressed generally as

$$\bar{x} \pm z_{\alpha/2}\sigma_{\bar{x}} = \bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

For samples of size > 30, the confidence interval is expressed as

$$\bar{x} \pm z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$ where s is the sample standard deviation, n is the number of measurements in the sample, and $z_{\alpha/2}$ is the z value with an area of $\alpha/2$ to its right.

The sample used is required to be random.

The table below was copied from our text and provides the most commonly used confidence coefficients with the corresponding values of σ and.

TABLE 7.2   Commonly Used Values of $z_{\alpha/2}$

Confidence Level 100(1 − α)	α	$\alpha/2$	$z_{\alpha/2}$
90%	.10	.05	1.645
95%	.05	.025	1.96
99%	.01	.005	2.575

Example. A random sample of 100 observations from a normally distributed population possesses a mean equal to 83.2 and a standard deviation of 6.4.

a.   Find a 95% confidence interval for μ.

n = _____; $\bar{x}$ = _____; s = _____; confidence coefficient = _____; α = _____;

$\dfrac{\alpha}{2}$ = _____; $z_{\alpha/2}$ = _____; $z_{\alpha/2}\left(\dfrac{s}{\sqrt{n}}\right)$ = _____.

So, the 95% confidence interval is _____.

b.   What do we mean when we say a confidence level is 95%?

Applet Exercise 5.1