Chapter 9: estimating the value of a parameter using confidence intervals

By: Wandi Ding

9.1 WHEN THE POPULATION STANDARD DEVIATION KNOWN

1. Point estimate: is the value of a statistic that estimates the value of a parameter. For example, the sample mean $\bar{x}$ is a point estimate of the population mean, μ .

Suppose: we want to estimate the average weight for all students in MTSU for this semester, we could take a random sample of 100 students and find the average weight of these students, say, 130 pounds, this kind of estimate is called a point estimate. Often, there is another question to be asked, it is, how good is a point estimate? There is no way of knowing how close a particular point estimate is to the population mean if the population is large. For this reason, statisticians prefer another type of estimate, called interval estimate.

2. Interval estimate: an interval for unknown parameter is an interval or a range of values used to estimate the parameter with the specific confidence level of estimate. It is also called confidence interval.

9.1 WHEN THE POPULATION STANDARD DEVIATION KNOWN

To make our interval estimate more reasonable and confident, we usually use a degree of confidence to describe interval estimate, like 95%, 98%, 99%...

3. The level of confidence: represents the expected proportion of intervals that will contain the parameter if a large number of different samples is obtained. The level of confidence is denoted $(1-\alpha)*100\%$. For example, a 95% level of confidence ($\alpha=0.05$) implies that if 100 different confidence intervals are constructed, each based on a different sample from the same population, then we will expect 95 of the intervals to include the parameter and 5 to not include the parameter.

4. Formula for the confidence interval of the population mean for a specific α, where σ is known and population is normally distribution or sample size $n \geq 30$

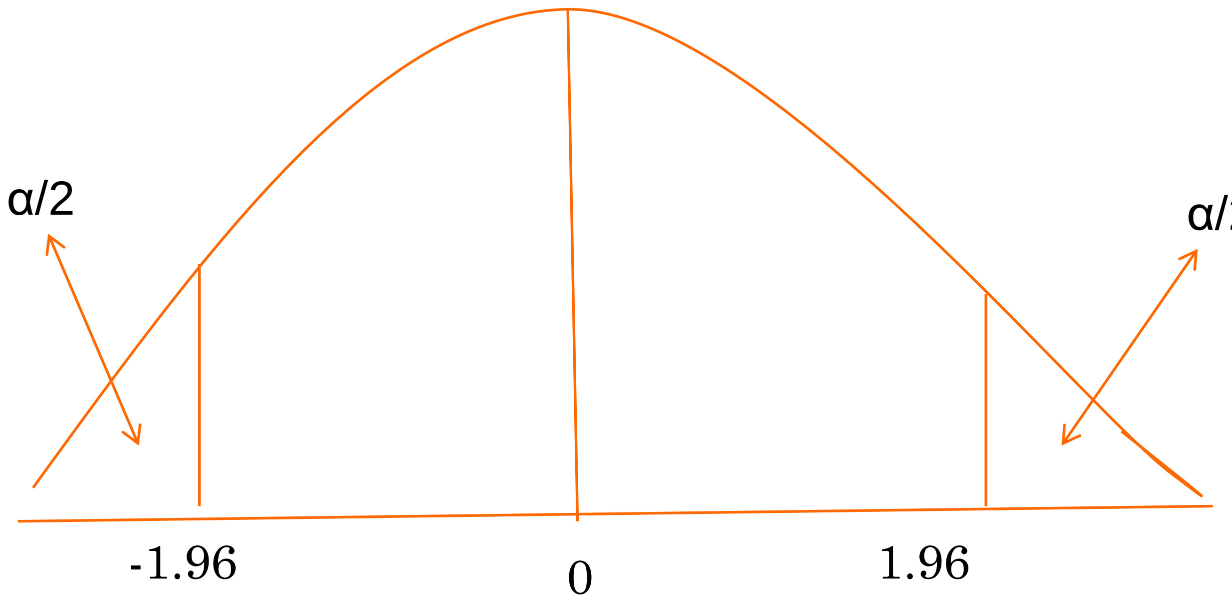
$$\bar{x} - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \bar{x} + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

9.1 When the Population Standard Deviation Known

Ex: for 95% confidence interval for a population mean:

α=1-95%=5%; $\alpha/2$=0.05/2=0.025,

$z_{\alpha/2} = z_{0.025} = 1.96$ (Based on the standard normal table)



9.1 WHEN THE POPULATION STANDARD DEVIATION KNOWN

Confidence interval estimates for the population mean can be written in this form, too.

point estimate $\pm$ margin of error

$z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$ =E is margin of error, also called maximum error of estimate. There are three factors which affect the margin of error.

1) level of confidence . Note: α=1-level of confidence.

2) sample size, n.

3) standard deviation of the population.

Note: The value of $z_{\alpha/2}$ is called critical value of the distribution and the next slide shows common critical values used lot in confidence intervals.

9.1 WHEN THE POPULATION STANDARD DEVIATION KNOWN

The common critical value for 90%, 95%, 99% confidence level:

Level of confidence (1-α)*100%	**Area in each tail, α/2**	**Critical vale** $z_{\alpha/2}$
90%	0.05	1.645
95%	0.025	1.96
99%	0.005	2.575

Interpretation of a confidence interval: A (1-α)*100% confidence interval indicates that (1-α)*100% of all simple random samples of size n from the population whose parameter is unknown will contain the parameter.

9.1 WHEN THE POPULATION STANDARD DEVIATION KNOWN

Comments for confidence interval:

$$\bar{x} - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \bar{x} + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

Lower bound:

$$\bar{x} - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

Upper bound:

$$\bar{x} + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

9.1 WHEN THE POPULATION STANDARD DEVIATION KNOWN

Ex: a survey of 30 adults found that the mean age of a person's primary vehicle is 5.6 years. Assuming the standard deviation of the population is 0.8 year, find the 99% confidence interval of the population mean and interpret it.

Ans:

$$\bar{x} - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \bar{x} + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

$\bar{x} = 5.6$ based on the 30 adults sample; n=30;

$\sigma = 0.8$ is the standard deviation of population.

confidence level is 99% and α=1-99%=1%,

$$z_{\alpha/2} = z_{0.01/2} = z_{0.005} = 2.575$$

9.1 WHEN THE POPULATION STANDARD DEVIATION KNOWN

and the final answer is:

$$\bar{x} - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \bar{x} + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$
$$5.6 - 2.575\left(\frac{0.8}{\sqrt{30}}\right) < \mu < 5.6 + 2.575\left(\frac{0.8}{\sqrt{30}}\right)$$
$$5.6 - 0.376 < \mu < 5.6 + 0.376$$
$$5.224 < \mu < 5.976 \quad \text{or } 5.22 < \mu < 5.98$$

Interpretation: one can be 99% confident that the mean age of all primary vehicles is between 5.2 years and 6.0 years, based on 30 vehicles.

Comments: the margin of error in this example is 0.376, if we increase sample size, and the margin of error will decreasing. If possible, collecting more data to reduce the margin of error.

9.1 WHEN THE POPULATION STANDARD DEVIATION KNOWN

5. Calculating the necessary sample size:

$$E = z_{a/2} \frac{\sigma}{\sqrt{n}}$$

$$E\sqrt{n} = z_{a/2} * \sigma$$

$$\sqrt{n} = \frac{z_{a/2} * \sigma}{E}$$

$$n = (\frac{z_{a/2} * \sigma}{E})^2$$

9.1 WHEN THE POPULATION STANDARD DEVIATION KNOWN

Ex: the college president asks the statistics teacher to estimate the average age of the students at their college. How large a sample is necessary? The statistics teacher would like to be 99% confident that the estimate should be accurate within 1 year. (the standard deviation of the age is known to be 3 years.)

Ans:

α=1-99%=1%=0.01, α/2=0.01/2=0.005, $Z_{a/2}$ =2.575, E=1

$$n = \left(\frac{z_{a/2} * \sigma}{E}\right)^2$$

$$= \left(\frac{2.575 * 3}{1}\right)^2$$

$$\approx 59.67 \text{ so, round up to } 60$$

The sample size at least is 60 students.

9.2 WHEN THE POPULATION STANDARD DEVIATION UNKNOWN

As we know, if the standard deviation of population is known and

1)the sample is drawn from a normal distribution

or

2) sample size $n \geq 30$ when parent population is unknown.

We will use z-value based on standard normal distribution to construct confidence interval.

If the population standard deviation is unknown and sample size $n<30$, then we will use t-value based on student's t distribution NOT use z-value based on standard normal distribution.

9.2 WHEN THE POPULATION STANDARD DEVIATION UNKNOWN

Student's t-distribution:

Suppose that a simple random sample of size n is taken from a population. If the population from which the sample is drawn follows a normal distribution. The distribution of

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

follows a student's t-distribution with n-1

Degree of freedom, where $\bar{x}$ is the sample mean and s is the sample standard deviation.

Note: the t-statistic represents the number of sample standard errors $\bar{x}$ is from the population mean, μ.

9.2 WHEN THE POPULATION STANDARD DEVIATION UNKNOWN

Properties of the t-distribution:

1) The t-distribution is different for different degree of freedom.
2) The t-distribution is centered at 0 and is symmetric about 0.
3) The area under the curve is 1. The area under the curve to the right of 0 equals to the area under the curve to the left of 0, which equals ½=0.5.
4) As t increases without bound, the graph approaches, but never equals, zero. As t decreases without bound, the graph approaches, but never equals, zero.
5) The area is the tails of the t-distribution is a little greater than the area in the tails of the standard normal distribution, because we are using s as an estimate of σ , thereby introducing further variability into the t-statistic.

9.2 WHEN THE POPULATION STANDARD DEVIATION UNKNOWN

6) As the sample size n increases, the density curve of t gets closer to the standard normal density curve. This result occurs because, as the sample size increases, the values of s get closer to the value of σ , by the law of large numbers.

Page 426, figure 8 shows these properties.

A. How to find t-values:

Let's go over the example 2 on page 426. After this example, you should know how to find the t-values.

9.2 WHEN THE POPULATION STANDARD DEVIATION UNKNOWN

Constructing a (1-α)*100% confidence interval for μ, σ **unknown:**

$$\bar{x} - t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right) < \mu < \bar{x} + t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

$$\text{lower bound}: \ \bar{x} - t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

$$\text{upper bound}: \bar{x} + t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

9.2 WHEN THE POPULATION STANDARD DEVIATION UNKNOWN

Ex: the below data represents a sample of the number of home fires started by candles for the past several years, find the 99% confidence interval for the mean number of home fires started by candles each year.

5460, 5900, 6090, 6310, 7160, 8440, 9930

Answer: step 1: find the sample mean and standard deviation based on the sample

$$\bar{x} = 7041.4 \quad s = 1610.3$$

Step 2: find the $t_{\alpha/2}$

sample size n=7, so degree of freedom is n-1=7-1=6

$t_{\alpha/2} = 3.707$

9.2 WHEN THE POPULATION STANDARD DEVIATION UNKNOWN

Step 3: formula used:

$$\bar{x} - t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right) < \mu < \bar{x} + t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

$$7041.4 - 3.707\left(\frac{1610.3}{\sqrt{7}}\right) < \mu < 7041.4 + 3.707\left(\frac{1610.3}{\sqrt{7}}\right)$$

$$4785.2 < \mu < 9297.6$$

How to interpret this result?

9.2 WHEN THE POPULATION STANDARD DEVIATION UNKNOWN

Summary:

1) If population standard deviation σ is known, and

   a) sample size n ≥30, using z-interval.

   b) sample size n <30, but the sample is drawn from a normal population, using z-interval. If the sample not from normal population, using non-parametric method.

2) If population deviation σ unknown, and

   a) sample size n≥30, using t-interval.

   b) sample size n<30, if sample is from normal population, then using t-interval, if not from normal population, using non-parametric method.