Estimating a Population Mean

MATH 130, *Elements of Statistics I*

J. Robert Buchanan

Department of Mathematics

Fall 2018

Objectives

At the end of this lesson we will be able to:

- obtain a point estimate for the population mean,
- state the properties of Student's t-distribution,
- determine t-values,
- construct and interpret a confidence interval for a population mean, and
- find the sample size needed to estimate the population mean within a given margin of error.

Point Estimate of the Mean

- The sample mean $\overline{x}$ is the best point estimate of the population mean μ.
- Previously we constructed confidence interval estimates for the mean of a population assuming we knew the population standard deviation σ (which is unlikely in practice).

$$\text{Lower and upper estimates: } \overline{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Point Estimate of the Mean

- The sample mean $\overline{x}$ is the best point estimate of the population mean μ.
- Previously we constructed confidence interval estimates for the mean of a population assuming we knew the population standard deviation σ (which is unlikely in practice).

$$\text{Lower and upper estimates: } \overline{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

- What can we do when σ is unknown?

Point Estimate of the Mean

- The sample mean $\overline{x}$ is the best point estimate of the population mean μ.
- Previously we constructed confidence interval estimates for the mean of a population assuming we knew the population standard deviation σ (which is unlikely in practice).

$$\text{Lower and upper estimates: } \overline{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

- What can we do when σ is unknown?

$$\text{Lower and upper estimates: } \overline{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

is not correct since s, like $\overline{x}$, is a random variable (will depend on the sample chosen) and is not constant.

Student's t-Distribution

Definition

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

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

follows **Student's t-Distribution** with $n - 1$ degrees of freedom. The sample mean is $\overline{x}$ and the sample standard deviation is s.

Student's t-Distribution

Definition

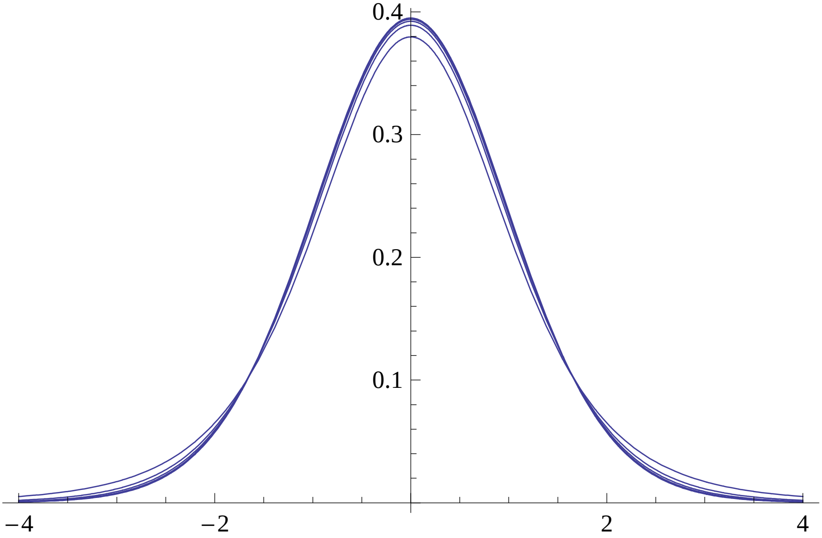
Suppose a simple random sample of size n is drawn from a population. If the population from which the sample is taken follows a normal distribution, the distribution of the random variable

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

follows **Student's t-Distribution** with $n - 1$ degrees of freedom. The sample mean is $\overline{x}$ and the sample standard deviation is s.

Remark: we will treat the random variable t similar to the Z-score.

Shape of Student's *t*-Distribution

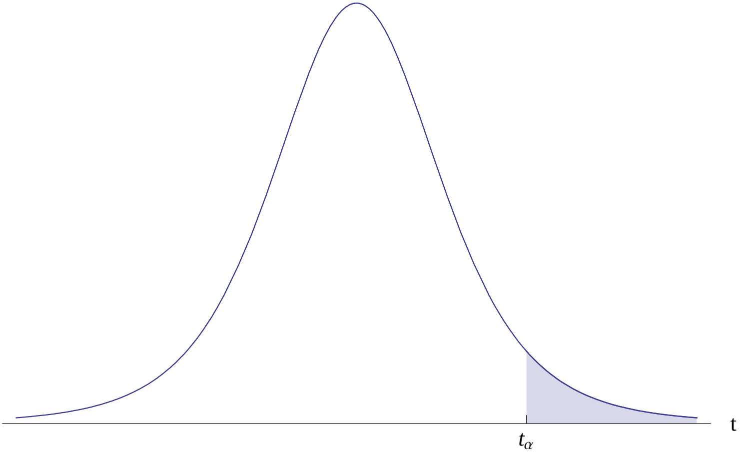


Properties of Student's t-Distribution

1. The t-distribution is different for different degrees of freedom.
2. The t-distribution is centered at 0 and symmetric about 0.
3. The total area under the curve is 1. The area to the left of 0 is $1/2$ and the area to the right of 0 is $1/2$.
4. As the magnitude of t increases the graph approaches but never equals 0.
5. The area in the tails of the t-distribution is larger than the area in the tails of the normal distribution.
6. As sample size n increases, the distribution becomes approximately normal.

t-Values

We will use the symbol t_α to represent the value of t for which the area to the right is α.



Determining t-Values

We can look up t_α in Table VII.

To use this table we must know:

- the degrees of freedom $df = n - 1$,
- the area in the right tail of the distribution.

Confidence Intervals

Suppose a simple random sample of size n is taken from a population with unknown mean μ and an unknown standard deviation σ. A $(1-\alpha)\cdot 100\%$ confidence interval for μ is given by

$$\text{Lower and upper estimates: } \overline{x} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

where $t_{\alpha/2}$ is computed with $n-1$ degrees of freedom.

Remark: this type of confidence interval is sometimes called a ***t*-interval**.

Example

The EPA is suing the city of Rochester for noncompliance with carbon monoxide standards. The standard level is set at 4.9 parts per million (PPM). A random sample of 22 carbon monoxide levels yields $\overline{x} = 5.1$ PPM with $s = 1.2$ PPM. Construct the 95% confidence interval for the mean carbon monoxide level in Rochester.

Confidence Interval

Make note of the quantities mentioned in the example:

$$\begin{aligned} \overline{x} &= 5.1 \\ s &= 1.2 \\ n &= 22 \end{aligned}$$

Confidence Interval

Make note of the quantities mentioned in the example:

$$\begin{aligned} \overline{x} &= 5.1 \\ s &= 1.2 \\ n &= 22 \end{aligned}$$

The 95% CI
implies

$$\begin{aligned} \alpha &= 0.05 \\ \alpha/2 &= 0.025 \\ \text{df} &= 21 \end{aligned}$$

thus $t_{\alpha/2} = 2.080$.

Confidence Interval

Make note of the quantities mentioned in the example:

$$\begin{aligned} \overline{x} &= 5.1 \\ s &= 1.2 \\ n &= 22 \end{aligned}$$

The 95% CI implies

$$\begin{aligned} \alpha &= 0.05 \\ \alpha/2 &= 0.025 \\ \text{df} &= 21 \end{aligned}$$

thus $t_{\alpha/2} = 2.080$.

The margin of error is

$$E = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} = 2.080 \cdot \frac{1.2}{\sqrt{22}} = 0.532.$$

The 95% confidence interval estimate of μ is

$$\begin{aligned} (\overline{x} - E, \overline{x} + E) &= (5.1 - 0.5, 5.1 + 0.5) \\ &= (4.6, 5.6) \text{ PPM.} \end{aligned}$$

Example

A student organization maintains that the average student must travel for at least 25 minutes to reach college each day. The college admissions office obtained a random sample of 31 one-way travel times from students. The sample had a mean of 19.4 minutes and a standard deviation of 9.6 minutes. Construct the 90% confidence interval for the average time students spend traveling one-way to college.

Confidence Interval

Make note of the quantities mentioned in the example:

$$\begin{aligned} \overline{x} &= 19.4 \\ s &= 9.6 \\ n &= 31 \end{aligned}$$

Confidence Interval

Make note of the quantities mentioned in the example:

$$\begin{aligned} \overline{x} &= 19.4 \\ s &= 9.6 \\ n &= 31 \end{aligned}$$

The 90% CI
implies

$$\begin{aligned} \alpha &= 0.10 \\ \alpha/2 &= 0.05 \\ \text{df} &= 30 \end{aligned}$$

thus $t_{\alpha/2} = 1.697$.

Confidence Interval

Make note of the quantities mentioned in the example:

$$\begin{aligned} \bar{x} &= 19.4 \\ s &= 9.6 \\ n &= 31 \end{aligned}$$

The 90% CI implies

$$\begin{aligned} \alpha &= 0.10 \\ \alpha/2 &= 0.05 \\ \text{df} &= 30 \end{aligned}$$

thus $t_{\alpha/2} = 1.697$.

The margin of error is

$$E = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} = 1.697 \cdot \frac{9.6}{\sqrt{31}} = 2.926.$$

The 90% confidence interval estimate of μ is

$$\begin{aligned} (\bar{x} - E, \bar{x} + E) &= (19.4 - 2.9, 19.4 + 2.9) \\ &= (16.5, 22.3) \text{ minutes.} \end{aligned}$$

Example

The National Adoption Information Clearinghouse tracks and posts information about child adoptions in the United States. Twenty states were randomly sampled and the percentage change in the number of adoptions per year from 2003 to 2007 was recorded:

$$\begin{array}{rrrrrrrrrr} 6 & 8 & -17 & 18 & 8 & 11 & 2 & 13 & 14 & 22 \\ -5 & -11 & 0 & -20 & -23 & 12 & -1 & 32 & 5 & 5 \end{array}$$

Find the 99% confidence interval for the average percent change in the number of adoptions per year from 2003 to 2007 for all states.

Example

The National Adoption Information Clearinghouse tracks and posts information about child adoptions in the United States. Twenty states were randomly sampled and the percentage change in the number of adoptions per year from 2003 to 2007 was recorded:

6	8	−17	18	8	11	2	13	14	22
−5	−11	0	−20	−23	12	−1	32	5	5

Find the 99% confidence interval for the average percent change in the number of adoptions per year from 2003 to 2007 for all states.

Hint: $\bar{x} = 4.0$ and $s = 14.0$.

Confidence Interval

Make note of the quantities mentioned in the example:

$$\begin{aligned} \overline{x} &= 4.0 \\ s &= 14.0 \\ n &= 20 \end{aligned}$$

Confidence Interval

Make note of the quantities mentioned in the example:

$$\begin{aligned} \bar{x} &= 4.0 \\ s &= 14.0 \\ n &= 20 \end{aligned}$$

The 99% CI
implies

$$\begin{aligned} \alpha &= 0.01 \\ \alpha/2 &= 0.005 \\ \text{df} &= 19 \end{aligned}$$

thus $t_{\alpha/2} = 2.861$.

Confidence Interval

Make note of the quantities mentioned in the example:

$$\begin{aligned}\overline{x} &= 4.0\\ s &= 14.0\\ n &= 20\end{aligned}$$

The 99% CI implies

$$\begin{aligned}\alpha &= 0.01\\ \alpha/2 &= 0.005\\ \text{df} &= 19\end{aligned}$$

thus $t_{\alpha/2} = 2.861$.

The margin of error is

$$E = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} = 2.861 \cdot \frac{14.0}{\sqrt{20}} = 8.956.$$

The 99% confidence interval estimate of μ is

$$\begin{aligned}(\overline{x} - E, \overline{x} + E) &= (4.0 - 9.0, 4.0 + 9.0)\\ &= (-5.0, 13.0) \text{ percent.}\end{aligned}$$

Determining Sample Size (1 of 2)

Recall: the margin of error in constructing a confidence interval for the population mean is

$$E = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}.$$

If we want to solve this for n in order to determine the sample size to collect we get

$$n = \left(\frac{t_{\alpha/2} \cdot s}{E} \right)^2$$

but there is a problem with this approach because in order to look up $t_{\alpha/2}$ we must know the number of degrees of freedom and therefore the sample size.

Determining Sample Size (2 of 2)

Instead we will use the standard normal distribution.

Determining the Sample Size n

The sample size required to estimate the population mean, μ, with a level of confidence $(1-\alpha)\cdot 100\%$ within a specified margin of error, E, is given by

$$n = \left(\frac{z_{\alpha/2}\cdot s}{E}\right)^2$$

where n is rounded up to the nearest whole number.

Example

Suppose a previous study has shown the sample standard deviation of a particular model of automobile to be 3.12 mpg.

How large a sample is necessary to estimate the mean miles per gallon to within 0.25 mpg with 95% confidence?

Example

Suppose a previous study has shown the sample standard deviation of a particular model of automobile to be 3.12 mpg.

How large a sample is necessary to estimate the mean miles per gallon to within 0.25 mpg with 95% confidence?

Reading carefully above we have $s = 3.12$, $E = 0.25$, and $\alpha = 0.05$ which implies $z_{\alpha/2} = 1.96$.

$$n = \left(\frac{z_{\alpha/2} \cdot s}{E} \right)^2 =$$

Example

Suppose a previous study has shown the sample standard deviation of a particular model of automobile to be 3.12 mpg.

How large a sample is necessary to estimate the mean miles per gallon to within 0.25 mpg with 95% confidence?

Reading carefully above we have $s = 3.12$, $E = 0.25$, and $\alpha = 0.05$ which implies $z_{\alpha/2} = 1.96$.

$$n = \left(\frac{z_{\alpha/2} \cdot s}{E} \right)^2 = \left(\frac{1.96 \cdot 3.12}{0.25} \right)^2 = 598.33 \approx 599$$