Introductory Statistics Lectures

Estimating a population mean

Confidence intervals for means

Anthony Tanbakuchi
Department of Mathematics
Pima Community College

(Compile date: Tue May 19 14:50:24 2009)

Contents

1 Estimating a population mean

1.1 Introduction

Example 1. We would like to estimate the mean height of US adults using our class data (assuming it is a representative random sample). Moreover, we wish to determine the margin of error for our estimate to have a measure of its precision.

```
R: summary(height)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   62.0    65.0    68.0    67.6    69.8    77.0
```

Question 1. What do we need to know to determine our margin of error?

Recall from last lecture

- Confidence interval tells us the margin of error in estimating a population parameter with a statistic. The margin of error depends on (1) the sample size, (2) the confidence level, (3) sampling distribution of the sample statistic.
- confidence level $= 1 - \alpha$

- $z_{\alpha/2}$ is z-score with $\alpha/2$ area to the **right**.
- CLT: $\bar{x}$ is normally distributed with $\sigma_{\bar{x}}$ if either (1) x is normal or (2) $n > 30$.
- Poor sampling leads to useless and potentially misleading results!

1.2 Confidence intervals for $\bar{x}$

USE

Often used to answer:
1. What is a reasonable estimate for the population mean?
2. How much variability is there in the estimate for the population mean?
3. Does a given target value fall within the confidence interval?

COMPUTATION

DEFINITION 1.1 CONFIDENCE INTERVAL FOR μ WHEN σ IS KNOWN.
Requirements: (1) Simple random samples, (2) CLT applies (x normal or $n > 30$).

$$\boxed{\bar{x} \pm E} \tag{1}$$

where

$$\boxed{E = z_{\alpha/2} \cdot \sigma_{\bar{x}}} \tag{2}$$

and

$$\boxed{\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}} \tag{3}$$

DEFINITION 1.2 CONFIDENCE INTERVAL FOR μ WHEN σ IS UNKNOWN.
Requirements: (1) Simple random samples, (2) CLT applies (x normal or $n > 30$).
Just like before except:

$$\boxed{E = \boldsymbol{t}_{\alpha/2} \cdot \hat{\sigma}_{\bar{x}}} \tag{4}$$

and

$$\boxed{\hat{\sigma}_{\bar{x}} = \frac{\boldsymbol{s}}{\sqrt{n}}} \tag{5}$$

Since we also have to estimate $\sigma_{\bar{x}}$ (hence the hat), a margin of error is associated with $\hat{\sigma}_{\bar{x}}$ so the distribution of $\bar{x}$ should be broader than when σ is known.

DEFINITION 1.3 STUDENT t DISTRIBUTION.

A bell shape symmetrical distribution that describes

$$t = \frac{\bar{x} - \mu}{\hat{\sigma}_{\bar{x}}} \tag{6}$$

with increasing dispersion (variation) as the sample size decreases in terms of the **degrees of freedom**:

$$\boxed{df = n - 1} \tag{7}$$

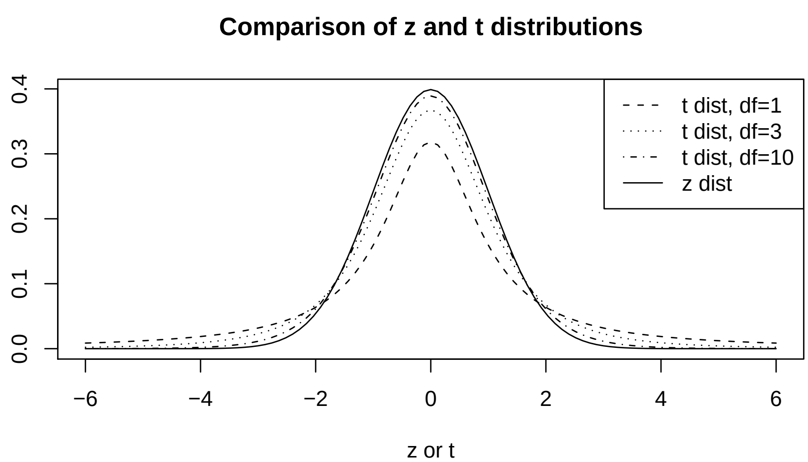
Think of the t distribution as the z distribution but with an adjusted standard deviation that increases for smaller sample sizes to account for a larger margin of error.

STUDENT t CDF:
`p=pt(t, df)`
Where p is the area to the left and df is the degrees of freedom.

R COMMAND

STUDENT t INVERSE CDF:
`t=qt(p, df)`
Where p is the area to the left and df is the degrees of freedom.

R COMMAND



Comparison of critical values for z and t distribution
Critical values $t_{\alpha/2}$ when $\alpha = 0.05$ for the

```
R: qt(1 - 0.05/2, df = 9)
[1] 2.2622
R: qt(1 - 0.05/2, df = 29)
[1] 2.0452
R: qt(1 - 0.05/2, df = 99)
[1] 1.9842
```

As compared to the $z_{\alpha/2}$

```
R: qnorm(1 - 0.05/2)
[1] 1.9600
```

Determining required sample size given desired E

Solving equation 2 for n:

$$\boxed{n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2} \tag{8}$$

You should always determine the required n **before** conducting a study! If σ is unknown do a pilot study to estimate it or find applicable prior data.

What if CLT does not apply?

We have only discussed methods for estimating μ when the Central Limit Theorem applies (x is normally distributed or $n > 30$). When looking at the original data to asses normality, it should be somewhat symmetric and have only one mode with no outliers. If the population severely deviates from a normal, sample sizes may need to be more than 50 to 100.

If the CLT does not apply you cannot use these methods. You need to use a (1) nonparametric method or (2) bootstrap method which makes no assumption about the population's distribution.

A COMPLETE EXAMPLE

Example 2. We would like to estimate the mean height of US adults using our class data (assuming it is a representative random sample). Moreover, we wish to determine the margin of error for our estimate to have a measure of its precision.

1. What is known:

   Point estimate for mean height (in inches):

```
R: x.bar = mean(height)
R: x.bar
[1] 67.611
```

   Sample standard deviation of heights:

```
R: s = sd(height)
R: s
[1] 3.8370
```

   Sample size:

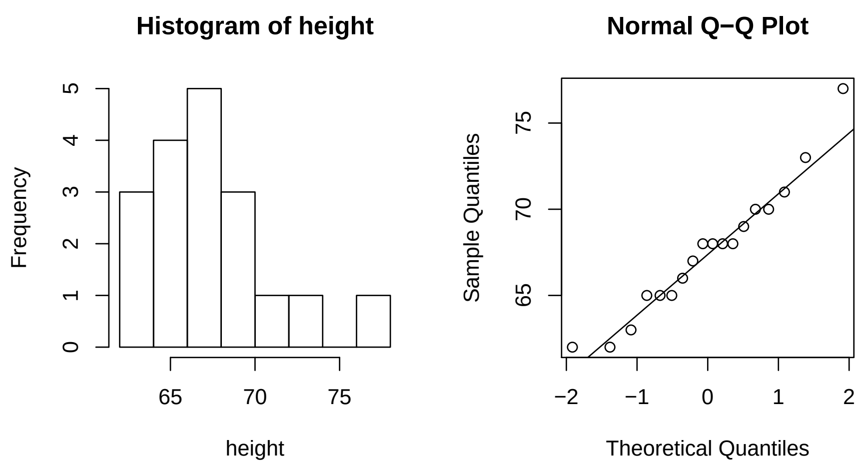
```
R: n = length(height)
R: n
[1] 18
```

   Since confidence level is unspecified, assume 95%:

```
R: alpha = 0.05
```

2. To construct a CI, determine if CLT applies:

```
R: par(mfrow = c(1, 2))
R: hist(height)
R: qqnorm(height)
R: qqline(height)
```



Question 2. Is the CLT satisfied?

To continue, assume population is normally distributed.

3. Determine which sampling distribution to use: since σ is unknown, use t distribution.

4. Find the margin of error and construct CI: $E = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$

```
R: t.critical = qt(1 − alpha/2, df = n − 1)
R: t.critical
[1] 2.1098
R: E = t.critical * s/sqrt(n)
R: E
[1] 1.9081
```

Thus our 95% confidence interval estimate for the mean height of US adults is (in inches): 67.6 ± 1.91 or $(65.7, 69.5)$. (The National Health Survey estimates the mean height as 66.3 inches.)

1.3 Summary

Confidence intervals for $\bar{x}$

If CLT applies: (if it does not apply you cannot use these methods)

- Sample size: (requires some estimate of σ)

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

- Confidence interval

$$\text{CI: } \bar{x} \pm E$$

1. If σ known:

$$E = z_{\alpha/2} \cdot \overbrace{\frac{\sigma}{\sqrt{n}}}^{\sigma_{\bar{x}}}$$

   $z_{\alpha/2}$= `qnorm(1-alpha/2)`

2. If σ unknown use s:

$$E = t_{\alpha/2} \cdot \overbrace{\frac{s}{\sqrt{n}}}^{\hat{\sigma}_{\bar{x}}}, \qquad df = n - 1$$

   $t_{\alpha/2}$= `qt(1-alpha/2, df=n-1)`

The method presented here for choosing the z or t distribution slightly differs from the book. This method is arguably simpler and more accurate.

1.4 Additional Examples

Question 3. Nelson Media Research wants to estimate the mean amount of time (in minutes) that full-time college students spend watching television each weekday. Find the sample size necessary to estimate that mean with a 15-minute margin of error. Assume that a 98% confidence level is desired. Also assume prior data indicates that the population is normally distributed with a standard deviation is 112.2 minutes.

Question 4. Find the 95% confidence interval for the mean pulse rate of adult males using the book data set `Mhealth` .

The world's smallest mammal is the bumblebee bat, also known as the Kitti's hog-nosed bat. Such bats are roughly the size of a large bumblebee. Listed below are weights (in grams) from a sample of these bats.

`1.7, 1.6, 1.5, 2.0, 2.3, 1.6, 1.6, 1.8, 1.5, 1.7, 2.2, 1.4, 1.6, 1.6, 1.6`

Question 5. Are the requirements met? How can we check?

Question 6. Construct a 90% confidence interval estimate of their mean weight (assuming that their weights are normally distributed).