MATH 3339 - 03 15951
Statistics for the Sciences
Chapter 7: Statistical Inferences

Wendy Wang
wwang60@central.uh.edu

Lecture 28 - 3339
Outline

1. Review
2. T-distribution
3. Confidence Interval for Population Mean
4. Examples of Confidence Intervals
Popper Set Up

- Fill in all of the proper bubbles.
- Make sure your ID number is correct.
- Make sure the filled in circles are very dark.
- This is popper number 24.
Recap: The confidence interval for population mean

The $1 - \alpha$ confidence interval for $\mu$, given that we know the population standard deviation is:

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

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

$$M.E = 2 \cdot \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}}$$

Width of confidence interval: $2 \cdot M.E$
Recap: Critical Value when $\sigma$ is known

- We are assuming we know the population standard deviation.

- For means if the population standard deviation is known, the critical value is $z^*$. where the area under the Normal curve is between $-z_{\alpha/2}$ and $+z_{\alpha/2}$ is the confidence level $C = 1 - \alpha$.

- This can be found using the z-table or `qnorm((1 + C)/2)` in R.

- The following table is the common confidence levels with their z-score

<table>
<thead>
<tr>
<th>C</th>
<th>80%</th>
<th>90%</th>
<th>95%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$z_{\alpha/2}$</td>
<td>$z_{0.10} = 1.28$</td>
<td>$z_{0.05} = 1.645$</td>
<td>$z_{0.025} = 1.96$</td>
<td>$z_{0.005} = 2.576$</td>
</tr>
</tbody>
</table>
Introducing T-distribution
A coffee machine dispenses coffee into paper cups. Here are the amounts measured in a random sample of 20 cups:

9.9, 9.7, 10.0, 10.1, 9.9, 9.6, 9.8, 9.8, 10.0, 9.5, 9.7, 10.1, 9.9, 9.6, 10.2, 9.8, 10.0, 9.9, 9.5, 9.9

Determine a 90% confidence interval for the mean amount of coffee dispensed from this machine.

\[ \bar{x} \pm z_{\alpha/2} \left( \frac{S}{\sqrt{n}} \right) \]

\[ \bar{x} = \frac{9.9 + 9.7 + \cdots + 9.9}{20} = 9.845 \]

\[ n = 20 \quad \sigma \text{ is not provided!} \]
Standard Error when $\sigma$ is unknown

- When the standard deviation of a statistic is estimated from the data, the result is called the **standard error** of the statistic.

- The standard error of the sample mean is

$$SE\bar{x} = \frac{s}{\sqrt{n}}$$

where $s$ is the computed **sample** standard deviation from the data.

- From our example: $SE\bar{x} = \frac{0.1986}{\sqrt{20}} = 0.0444$. 
The T-distribution

- The problem is that the sample standard deviation $s$ varies from sample to sample.
- William Gosset, (a quality control engineer for the Guinness Brewery) discovered this problem and figured out a new distribution that changes the critical value based on the sample size.
- This new distribution is called *Students T* distribution, because Guinness would not allow Gosset to publish his findings since he was their employee.
- The shape of this distribution changes with different sample sizes. So it depends on a parameter called the degrees of freedom ($df$).
- The degrees of freedom for the T-distribution of the sample mean is the sample size minus one: $(n - 1)$. Because we are using the sample standard deviation $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}$.
T distribution

- Used for the inference of the population mean. When population standard deviation $\sigma$ is unknown.
- The distribution of the population is basically **bell-shape**.
- Formula for $t$:
  \[ t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t(df=n-1) \]
- Use t-table, or `qt(probability,df)` in R.
- Degrees of freedom: $df = n - 1$. 
Normal Distribution vs T distribution

The red graph is the Normal density curve and the blue graph is the T density curve with a degrees of freedom of 4.
Using T-table


- The top margin is the area in the right tail.

- The left margin is the degrees of freedom $n - 1$.

- The values inside the table are the $t$ values.
Critical value when $\sigma$ unknown

- When $\sigma$ is **unknown** we use $t$-distribution.

- With degrees of freedom, $df = n - 1$.

- The critical value is $t_{\alpha/2}$, where the area between $-t_{\alpha/2}$ and $+t_{\alpha/2}$ under the T-curve is the confidence level $C = 1 - \alpha$.

- $t_{\alpha/2}$ is found in T-table using the row according to the degrees of freedom and the column according to the confidence level at the bottom of the table.

- In R use `qt((1 + C)/2, df)`.
Mean Amount of Coffee Dispensed

A coffee machine dispenses coffee into paper cups. Here are the amounts measured in a random sample of 20 cups.

9.9, 9.7, 10.0, 10.1, 9.9, 9.6, 9.8, 9.8, 10.0, 9.5,
9.7, 10.1, 9.9, 9.6, 10.2, 9.8, 10.0, 9.9, 9.5, 9.9

Determine a 90% confidence interval for the mean amount of coffee dispensed from this machine.

\[ \bar{x} \pm t(\alpha/2, df) \cdot \frac{s}{\sqrt{n}} \]

\[ \bar{x} = 9.845 \]

\[ s = \text{sample st. dev.} \]

\[ n = 20 \]

\[ C = 90\% \]
Determine the 90% Confidence Interval
R code

```r
coffee=c(9.9, 9.7, 10, 10.1, 9.9, 9.6, 9.8, 9.8, 10, 9.5, 9.7, 10.1, 9.9, 9.6, 10.2, 9.8, 10, 9.9, 9.5, 9.9)
t.test(coffee, conf.level = 0.9)
```

One Sample t-test

data:  coffee

t = 221.68, df = 19, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
90 percent confidence interval:
9.768207 9.921793
sample estimates:
mean of x
9.845
Example 1

A soft-drink machine is regulated so that the amount of drink dispensed is approximately normally distributed with a standard deviation equal to 0.53 ounces. Find a 99% confidence interval for the mean of all drinks dispensed by this machine if a random sample of 36 drinks has an average content of 7.94 ounces.

\[
X \sim N \left( \mu, \sigma^2 \right)
\]

\[
\sigma = 0.53
\]

\[
C = 99\%, z = 0.99
\]

\[
n = 36, \quad \bar{x} = 7.94
\]

\[
\bar{x} \pm \frac{z\sigma}{\sqrt{n}}
\]

\[
7.94 \pm qnorm \left( \frac{1 + 0.99}{2} \right) \cdot \frac{0.53}{\sqrt{36}}
\]

\[
\left[ 7.712, \ 8.168 \right]
\]
Example 2

The heights of a random sample of 50 college students showed a mean of 174.5 centimeters and a standard deviation of 6.9 centimeters. Construct a 98% confidence interval for the mean height of all college students.

\[ n = 50 \quad \bar{x} = 174.5 \quad s = 6.9 \]

\[ C = 98\% = 0.98 \]

\( \sigma \) is not given \( \implies t\text{-method} \)

\[ \bar{x} \pm t(\alpha/2, df) \cdot \frac{s}{\sqrt{n}} \]

\[ > 174.5 + c(-1,1) \times qt(1.98/2,49) \times 6.9/\sqrt{50} \]

[1] 172.1533 176.8467
What will happen if changing Confidence Levels $C$

Suppose we have a 99% confidence interval for the population mean, $\mu$ of (2.272, 17.728). If we change the confidence level to 92% what is the confidence interval?

$$LL = 2.272$$
$$UL = 17.728$$

$$M.E = \frac{UL - LL}{2} = 7.728$$

$$M.E = \frac{Z_{\alpha/2} \cdot \text{St. error}}{2}$$

$$C = 99\% \Rightarrow Z_{\alpha/2} = \text{norm}(\frac{1.99}{2}) = 2.5758$$

$$\Rightarrow \text{St. error} = \frac{M.E}{Z_{\alpha/2}} = \frac{2.5758}{2}$$

If $C = 92\% \Rightarrow Z_{\alpha/2} = \text{norm}(\frac{1.92}{2}) = 1.75$$

$$M.E = Z_{\alpha/2} \cdot \text{St. error} = 1.75 \cdot (3) = 5.25$$
Changing Sample Size

The mean of a random sample of \( n \) measurements is equal to \( \bar{x} = 33.9 \). Assume \( \sigma = 3.3 \). Determine the margin of error for a 95% confidence interval for the population mean when the sample size is

1. \( n = 100 \)
   
   \[
   M.E = \frac{z_{\alpha/2} \sigma}{\sqrt{n}} = \frac{1.96 \cdot 3.3}{\sqrt{100}} = 0.6468
   \]

2. \( n = 400 \)

   \[
   M.E = \frac{z_{\alpha/2} \sigma}{\sqrt{n}} = \frac{1.96 \cdot 3.3}{\sqrt{400}} = 0.3234
   \]
Behavior of Confidence Intervals with different $n$

Notice that as the sample size increases the width of the interval decreases. Or the confidence interval becomes thinner.

- First: Mathematically, notice that we are dividing by a larger number, so that will decrease the quotient.

- Second: Intuitively, as the sample size increases the accuracy of the estimation becomes better, thus the point estimate is getting closer to the population mean and the interval does not need to be as wide.
Choosing Sample Size

You can have both a high confidence while at the same time a small margin of error by taking enough observations.

- Sample size for confidence intervals of means.
Starting Salary

- We want to estimate annual starting salaries for college graduates with degrees in business administration. To determine this we need a sample.

- Assume that a 95% confidence interval estimate of the population mean annual starting salary is desired.

- Assume the standard deviation is $\sigma = 7,500$.

- How large a sample should be taken if the desired margin of error is $m = 500$?