

In Chapter 16-17, we learned how to do confidence intervals and test hypothesis for proportions. In this chapter we will do the same for means.

18.1 The Central Limit Theorem (Again)

The Central Limit Theorem (Chap. 15) gave us the sampling distribution for means as Normal with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$.

- How do you find the population standard deviation, σ ?
- We will use the sample standard deviation, s , which is the estimate for σ to get the Standard Error of $SE(\bar{x}) = \frac{s}{\sqrt{n}}$.

18.2 Gosset's t

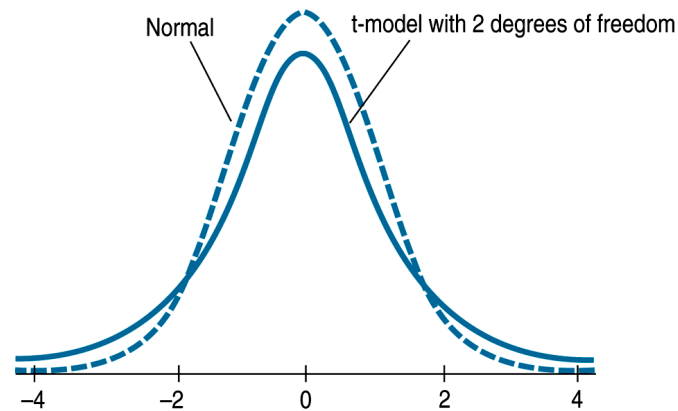
What sort of distribution can be used? The distribution is no longer Normal since the Standard Error introduces extra variation from s .

- William S. Gosset found the sampling model while working at Guinness Brewery in Dublin, Ireland.
- The model Gosset found is commonly referred to as the **Student's t distribution**.
- This model is actually a family of related distributions that depend on a parameter known as **degrees of freedom** ($df = n - 1$)
- A sampling distribution model for means, when the conditions are met, the standardized sample mean

$$t = \frac{\bar{x} - \mu}{SE(\bar{x})} \quad \text{where} \quad SE(\bar{x}) = \frac{s}{\sqrt{n}}$$

follows a Student's t distribution.

- Correcting for the extra variation causes this model to give us a larger margin of error which will make our intervals wider and our P-values will be larger than from the Normal Model.
- Student's t-models are unimodal, symmetric, and bell shaped like the Normal, but t-models with only a few degrees of freedom have fatter tails than the Normal.



- As the degrees of freedom increase, the t-models look more like the Normal. The t-model with infinite (∞) degrees of freedom is the Normal.

Assumptions and Conditions

- **Plausible Independence:** As before this is hard to check but you should at least think if independence is reasonable.
- **Randomization Condition:** The data comes from a random sample or randomized experiment. This helps with independence.
- **10% Condition:** When sample is drawn without replacement, the sample should be no more than 10% of the population.
- **Nearly Normal Condition:** The data come from a distribution that is unimodal and symmetric.
 - ◆ The smaller the sample size ($n < 15$ or so), the more closely the data should follow a Normal model.
 - ◆ For moderate sample sizes (in between 15 and 40 or so), the t works well as long as the data are unimodal and symmetric.
 - ◆ For larger sample sizes, the t models are safe to use even if the data are skewed.

One Sample t-Interval for the Mean: $\bar{x} \pm t_{n-1}^* \frac{s}{\sqrt{n}}$
 Estimate $\underbrace{\quad}_{\text{Margin of Error}}$

Example (p. 499, #24)

Parking Hoping to lure more shoppers downtown, a city builds a new public parking garage in the central business district. The city plans to pay for the structure through parking fees. During a two-month period (44 weekdays), daily fees collected averaged \$126, with a standard deviation of \$15.

$$n = 44 \quad \bar{x} = \$126 \quad s = \$15$$

a) What assumptions must you make in order to use these statistics for inference?

Not necessarily random, Days should be independent.

10% cond: $n = 44$ days is less than 10% of all days

Nearly Normal Condition: $n > 40$ so data can come from any distribution.

b) Write a 90% confidence interval for the mean daily income this parking garage will generate.

$$C = .9 \text{ \& } df = 44 - 1 = 43 \Rightarrow t^* = 1.684 \text{ use } df = 40 \text{ from Table}$$

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}} = 126 \pm 1.684 \frac{15}{\sqrt{44}} = \$122.19 \text{ to } \$129.81$$

c) Interpret this confidence interval in context.

Actual Daily Average

We are 90% confident that population mean (μ) is between \$122.19 to \$129.81.

d) Explain what "90% confidence" means in this context.

90% of similarly constructed intervals will contain μ .

e) The consultant who advised the city on this project predicted that parking revenues would average \$130 per day. Based on your confidence interval, do you think the consultant was correct? Why?

Consultant was not correct since entire interval is below \$130.

18.3 Interpreting Confidence Intervals

See p. 485. Confidence Interval is for a mean and not an individual value.

18.4 A Hypothesis Test for the Mean

One Sample t-Test for the Mean:

We test the hypothesis $H_0: \mu = \mu_0$ using the test statistic

$$t_{n-1} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

and the Student's t model with n-1 df.

Option 2 in Tests Menu from Stats Button.
T-Test

Example (p. 501, #38)

Saving gas Congress regulates corporate fuel economy and sets an annual gas mileage for cars. A company with a large fleet of cars hopes to meet the 2011 goal of 30.2 mpg or better for their fleet of cars. To see if the goal is being met, they check the gasoline usage for 50 company trips chosen at random, finding a mean of 32.12 mpg and a standard deviation of 4.83 mpg. Is this strong evidence that they have attained their fuel economy goal?

$n = 50$ $\bar{x} = 32.12 \text{ mpg}$ $S = 4.83 \text{ mpg}$
o.c.: Company is meeting fleet fuel economy goal. ($\mu > 30.2$)
a) $H_0: \mu = 30.2$ $\mu_0 = 30.2$
 $H_A: \mu > 30.2$ Right tailed test

b) trips are random and should be independent
 $n = 50$ should be less than 10% of all company's trips
 $n > 40$ Nearly Normal Condition says data can be from any distribution.

c) t distribution with $df = 50 - 1 = 49$ with $\mu_{\bar{x}} = \mu_0 = 30.2$
and Standard Error = $\frac{S}{\sqrt{n}} = \frac{4.83}{\sqrt{50}} = 0.683$

$$d) \text{ test Statistic: } t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{(32.12 - 30.2)}{(4.83/\sqrt{150})} = 2.81$$

check t.s and find p-value using option 2: T-test
in **TESTS** Menu of **STATS** button option.

$$P\text{-value} = .0035$$

f) p-value α
 $.0035 < .05$ Reject H_0 & Accept H_A

Data suggests company is meeting the national
fleet goal of 30.2 mpg.

18.5 Choosing the Sample Size

$$ME = t_{n-1}^* \frac{s}{\sqrt{n}} \Rightarrow n = \left(\frac{t_{n-1}^* s}{ME} \right)^2 \quad \text{have to use } n = \left(\frac{z^* s}{ME} \right)^2 \text{ first. Round up!}$$

Example (p. 497, #14)

An English professor is attempting to estimate the mean number of novels that the student body reads during their time in college. He is conducting an exit survey with seniors. He hopes to have a margin of error of 3 books with 95% confidence. From reading previous studies, he expects a large standard deviation and is going to assume it is 10. How many students should he survey?

Use Normal Distribution to get
n for degrees of freedom.

$$C = .95 \Rightarrow z^* = 1.96$$

$$s = 10 \quad ME = 3$$

$$n = \left(\frac{1.96 \times 10}{3} \right)^2$$

$$42.68$$

$$\text{use } n = 43$$

$$df = 43 - 1 = 42 \quad C = .95$$

$$\Rightarrow t^* = 2.021$$

$$n = \left(\frac{2.021 \times 10}{3} \right)^2$$

$$45.38$$

$$\text{use } n = 46$$