AP Statistics
Unit 06 – Day 03 Notes
Confidence Intervals: Means

Name Key
Period______

**Review:**

In January 2007 a Fox News poll of 900 ($n$) registered voters found that 82% ($\hat{p} = 0.82$) of the respondents believed that global warming exists.

a. Construct a 95% confidence interval to estimate the proportion of registered voters who believe that global warming exists.

State: we want to find the true proportion $p$ of all registered voters who believe that global warming exists with 95% confidence. $\hat{p} = 0.82$

Plan: Random: Fox News poll assumed random ✓
10% condition: $n = 900$ $9000 <$ all registered voters ✓
Large counts: $n\hat{p} = 900 \cdot 0.82 = 738 \geq 10$ ✓ $n\hat{q} = 900 \cdot 0.18 = 162 \geq 10$ ✓
Because our conditions are met, we will construct a 1-proportion z-interval to estimate $p$.

DO: $z^* = 1.96$ $0.82 \pm 1.96\sqrt{\frac{(0.82)(1-0.82)}{900}} = \boxed{(0.7949, 0.8451)}$

b. Interpret the results of the confidence interval in the context of the problem

Conclude: we are 95% confident that the interval from 0.7949 to 0.8451 captures the true proportion of all registered voters who believe that global warming exists.

c. Based on the confidence interval, Fox News should report, "Voter polls suggest that 82% of registered voters believe in global warming with a 0.025 margin of error."

$$ME = z^*\sqrt{\frac{\hat{p}\hat{q}}{n}} = 1.96\sqrt{\frac{(0.82)(1-0.82)}{900}} = 0.025$$

d. What is the meaning of "95% confidence" in the context of the problem? "confidence level"

If we take many samples of the same size from this population, about 95% of them will result in an interval that captures the true proportion of all registered voters who believe that global warming exists.

e. Note: Instead of using $\hat{p}$, pollsters often use a "worst case" assumption that p = 0.5 to calculate the margin of error. To understand this, re-calculate the ME for our survey results using p = 0.5. Why is this considered the "worst case" assumption?

$$ME = 1.96\sqrt{\frac{(0.5)(0.5)}{900}} = 0.0327$$

This is the largest our margin of error could possibly be with this sample size and this confidence level.

## Confidence Intervals for Means

Constructing a confidence interval for a mean is very similar to constructing one for a proportion. We always have to use the four step process: **State, Plan, Do,** and **Conclude.**

Means can be tricky because there are two ways to construct confidence intervals for them:

**If we somehow know $\sigma$:**

Use the z critical value and the standard Normal distribution to help calculate confidence intervals:

$$\boxed{\text{CI: } \bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}}$$

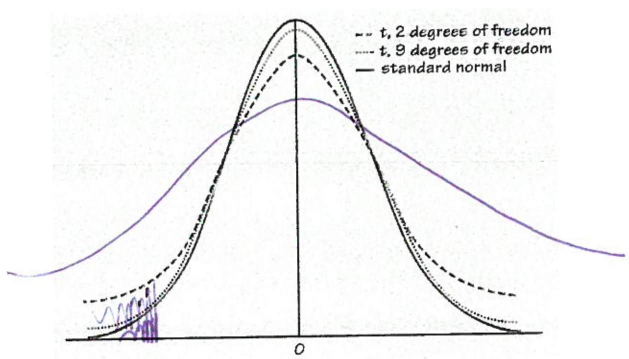
**In practice, we usually don't know $\sigma$:**

$t^*$

Replace the standard deviation with the standard error: $\boxed{SE_{\bar{x}} = \frac{s_{\bar{x}}}{\sqrt{n}}}$

Note that the standard error of the mean, SE, is same as the standard deviation of the sampling distribution for the sample mean, but with the sample standard deviation, $s_{\bar{x}}$, instead of $\sigma$.

When we use standard error, we use something called a t distribution instead of a Normal distribution. This is a special family of probability distribution distributions known as **Gosset's *t* distribution** (also known as Student's *t* distribution).




The *t* distributions are unimodal and symmetric, but change shape with different sample sizes.

**Degrees of freedom (df):** Each *t* distribution is defined by the number of degrees of freedom. We calculate degrees of freedom using the sample size:

$$df = n - 1$$

Hence, we often write a particular distribution as $t_{n-1}$ or $t_{df}$.

$t^*$

**Margin of Error (ME) for a sample mean:** When we calculate the margin of error, we use a t* critical value instead of a z* critical value. This value can be found on *Table B t distribution critical values*. It can also be found on the calculator:

1. 2nd distr
2. Option 4: invT(area, df)

Once you have calculated your t*, you can plug it into the margin of error equation:

$$\boxed{ME = t * \frac{s_{\bar{x}}}{\sqrt{n}}}$$

**Confidence Interval for a mean:** To calculate the confidence interval for a mean, we use our point estimate and add/subtract the margin of error. It is written the same way as a confidence interval for proportions. The critical value t* is chosen so that the t curve with n – 1 degrees of freedom has C% of the area between –t* and t*.

$$CI = \bar{x} \pm t * \frac{s_{\bar{x}}}{\sqrt{n}}$$

**On a calculator:**

1. STAT → TESTS
2. Choose Option #8: TInterval
3. Enter Stats or use a List
4. Press ENTER

**STATE:**

State what you are trying to do or find IN CONTEXT. Figure out what information you have and what information you need.

**PLAN:**

Check conditions!

- **Random:** The data comes form a well-designed random sample or randomized experiment.
- **10% Condition:** When sampling without replacement, check that the population is at least 10 times as large as the sample.
- **Normal/Large Sample:** The population distribution is Normal or the sample size is large (n≥30). When the sample size is small (n<30), examine a graph of the sample data for any possible departures from Normality in the population. You should be safe using a t distribution as long as there is no strong skewness and no outliers are present. rough sketch

**DO:** df = t* = #s plugged ( , )

List all of your values, list all of your equations, and plug it all in to get your answer!

**CONCLUDE:** sentence frame

Put your answer back in context. This usually means using your two sentences. It could also include answering additional questions.

**Example 1:**

n

In 1960, census results indicated that the age at which Americans first married had a mean of 23.3 years. We want to find out if the mean has increased during the past 53 years. We select a random sample of 40 American adults who have just gotten married and record their age. The mean age is 24.2 years with a standard deviation of 5.3 years.

$\bar{x}$ $s_x$

1. Construct a 95% confidence interval to estimate the mean age of first marriage for all Americans. Assume all conditions are met.

   a. State what you want to find:

   State: We want to find the true mean age ($\mu$) of American adults at first marriage with 95% confidence.

   $\bar{x} = 24.2$ years

   b. Check your conditions and state what process you will use (interval, test, etc.):

   Plan: Random: random sample of 40 American adults ✓
   10% condition: $n = 40$ $400 <$ all recently married American adults ✓
   Normal/Large: $n = 40 \geq 30$ ✓
   Because our conditions are met, we will construct a 1-sample t-interval to estimate $\mu$.

   c. Degrees of freedom (df):

   DO: $df = n - 1 = 40 - 1 = 39$

   d. Critical value, t* at 95% confidence level:

   $t^*_{39} = 2.023$

   e. Confidence Interval:

   $$24.2 \pm 2.023 \frac{5.3}{\sqrt{40}} = \boxed{(22.505, 25.895)}$$

   f. Interpret the confidence interval in the context of the problem:

   Conclude: We are 95% confident that the interval from 22.505 years to 25.895 years captures the true mean age at first marriage for all adult Americans.

2. What does "95% confidence" mean in context? "confidence level"
   If we take many samples of the same size from this population, about 95% of them will result in an interval that captures the true mean age at first marriage of all American adults.

3. Based on the confidence interval, do you think the average age of first marriage has increased over the last 53 years? Nope. 23.3 years is in our interval of plausible values, so there is not convincing evidence that the true mean age at first marriage for all American adults has increased.

**Example 2:**

Environmentalists, government officials, and vehicle manufacturers are all interested in studying the auto exhaust emissions produced by motor vehicles. The major pollutants in auto exhaust from gasoline engines are hydrocarbons, carbon monoxide, and nitrogen oxides (NOX). Researchers collected data on the NOX levels (in grams/mile) for a random sample of 40 light-duty engines of the same type. The mean NOX reading was 1.2675 and the standard deviation was 0.3332.

a. Construct and interpret a 95% confidence interval for the mean amount of NOX emitted by light-duty engines of this type.

State: we want to find the true mean amount of NOX emitted by light-duty engines of this type with 95% confidence. $\bar{x} = 1.2675$ grams/mi

Plan: Random: random sample of 40 light-duty engines ✓

10% condition: $n = 40$ $400 <$ all light-duty engines of this type ✓

Normal/Large: $n = 40 \geq 30$ ✓

Because our conditions are met, we will construct a 1-sample t-interval to estimate $\mu$.

DO: $df = 39$

$t^*_{39} = 2.023$

$1.2675 \pm 2.023 \cdot \frac{0.3332}{\sqrt{40}} = \boxed{(1.1609, 1.3741)}$

Conclude: we are 95% confident that the interval from 1.1609 g/mi to 1.3741 g/mi captures the true mean amount of NOX emitted by light-duty engines of this type.

b. The environmental Protection Agency (EPA) sets a limit of 1.0 gram/mile for average NOX emissions. Are you convinced that this type of engine violates the EPA limit? Use your interval from (a) to support your answer.

Yes, our entire interval of plausible values is above 1.0 g/mi so there is convincing evidence that this type of engine violates the EPA limit for NOX emissions.

**Calculating sample size:**

The sample size needed to obtain a confidence interval with approximate margin of error ME for a population mean involves solving

$$\boxed{z^* \frac{\sigma}{\sqrt{n}} \leq ME}$$

for n, where the standard deviation $\sigma$ is a reasonable value from a previous or pilot study, and z* is the critical value for the level of confidence we want.

**Example 3:**

C-level

ME

Researchers would like to estimate the mean cholesterol level $\mu$ of a particular variety of monkey that is often used in laboratory experiments. They would like their estimate to be within 1 milligram per deciliter (mg/dl) of the true value of $\mu$ at a 95% confidence level. A previous study involving this variety of monkey suggests that the standard deviation of cholesterol level is about 5 mg/dl.

$\sigma$

Obtaining monkeys for research is time-consuming, expensive, and controversial. What is the minimum number of monkeys the researchers will need to get a satisfactory estimate?

$$\sqrt{n} \cdot 1 \geq 1.96 \cdot \frac{5}{\cancel{\sqrt{n}}} \ \cancel{\sqrt{n}}$$

$$\sqrt{n}^2 \geq (1.96 \cdot 5)^2$$

$$n \geq (1.96 \cdot 5)^2$$

$$n \geq 96.04$$

$$\boxed{n \geq 97 \text{ monkeys}}$$