VII. Inference for a Single Population Mean

Example VII.A

A current area of research interest is the familial aggregation of cardiovascular risk factors in general and lipid levels in particular. To this end, a group of researchers measured the cholesterol levels of children with fathers who have died of heart disease. It is known that the average cholesterol level of children in the U.S. is 175 mg/dL.

What questions might the researchers want to answer?
To apply this familiar diagram to Example VII.A, what is the parameter associated with the population? What are the sample statistics of interest?
Statistical Inference

There are two aspects of inference:

- **Estimation**
- **Hypothesis Testing**

In terms of **Example VII.A**, estimation helps us to answer the question: “What is the actual mean cholesterol level of children whose fathers die of heart disease?” Hypothesis testing answers the question: “Is the mean cholesterol level of these children greater than the national average of 175 mg/dL?”
Revisiting the $t$ Distribution

To make inferences about an unknown mean $\mu$, where the underlying variance is unknown, we need to use the $t$ statistic

$$T = \frac{\sqrt{n}(\bar{X} - \mu)}{s},$$

which generally follows the $t$ distribution with $n - 1$ degrees of freedom.

Suppose that $n = 20$. What value of $c$ satisfies $P(|T| < c) = 0.99$? What value of $c$ satisfies $P(|T| < c) = 0.95$?

We refer to these values of $c$ respectively as $t_{0.005,19}$ and $t_{0.025,19}$ (the upper 0.005 and 0.025 critical values of the $t_{19}$ distribution).
Estimation: Computing Confidence Intervals

In general, based on a sample size $n$ and critical value $t_{\alpha/2,n-1}$, by definition

$$P\left(\left|\sqrt{n}(\bar{X} - \mu) / s\right| < t_{\alpha/2,n-1}\right) =$$

$$P\left(-t_{\alpha/2,n-1} < \sqrt{n}(\bar{X} - \mu) / s < t_{\alpha/2,n-1}\right) =$$

$$P\left(\frac{\bar{X} - t_{\alpha/2,n-1}s / \sqrt{n}}{\sqrt{n}} < \mu < \frac{\bar{X} + t_{\alpha/2,n-1}s / \sqrt{n}}{\sqrt{n}}\right) = 1 - \alpha.$$ 

Hence, the interval

$$\bar{X} \pm t_{\alpha/2,n-1} \frac{s}{\sqrt{n}}$$

covers the population mean $\mu$ with probability $1 - \alpha$. We refer to this interval as a $(1 - \alpha)100\%$ confidence interval.
Interpreting the Confidence Interval

• We refer to the value \((1 – \alpha)100\%\) as the **confidence level**. That is, we are \((1 – \alpha)100\%\) confident that the population mean \(\mu\) lies within the computed interval.

• As we generally want a high degree of confidence, \(\alpha\) is taken to be something relatively small, such as 0.05 or 0.01.

• The value \(\alpha\) represents the probability that the confidence interval does not cover \(\mu\). **We must accept some probability that we are wrong** (i.e., that the interval doesn’t cover \(\mu\)) in order to infer something meaningful about the population.

• Note that the interval represents a set of plausible values for \(\mu\) that are consistent with the observed data.

• The *interval* is random – \(\mu\) is fixed!!
Example VII.B

Suppose that in sampling from the population of children described in Example VII.A we observe the following cholesterol levels:

\[
\begin{array}{cccccc}
276 & 183 & 127 & 171 & 164 \\
254 & 268 & 180 & 236 & 283 \\
158 & 261 & 189 & 203 & 250 \\
136 & 223 & 160 & 173 & 221 \\
\end{array}
\]

Note that the sample mean is 205.8 mg/dL, and \( s^2 = 2341.747 \text{ (mg/dL)}^2 \).

Compute and interpret the 95% confidence interval.

Compute and interpret the 99% confidence interval.
Assumptions

Keep in mind that for our confidence interval to be valid (that is, in order to preserve our confidence level) we need to assume that:

• The sample is random.

• The sample size $n$ is sufficiently large enough to ensure that the sample mean is approximately normally distributed.

• If $n$ is relatively small, then we should at least be assured that the underlying sample is approximately normally distributed.
Interval Width

Note that the width of a confidence interval is two times the margin of error. If $L$ represents the interval width, then

$$L = \frac{2t_{\alpha/2,n-1}s}{\sqrt{n}} = 2(\text{critical value})(\text{s.e.}(\bar{X})).$$

The width is therefore determined by what three values?

1.

2.

3.

What is the effect of each of these on width? Which of these factors can we control directly?
Example VII.C

All else being equal, approximately what sample size would we require in order to ensure that the 95% confidence interval computed in Example VII.B is no wider than 30 mg/dL?
Comments on One-Sided Intervals and $z$ Intervals

We will largely ignore the material in sections 8.1.5 and 8.1.6 in the text.

Section 8.1.5 discusses one-sided intervals, which have limited application in practice. These are used when only an upper or lower confidence bound is required, but not both. In Example VII.B, for instance, we may just want a lower bound in order to determine whether the children have a mean cholesterol level that is above the national average of 175 mg/dL.

Section 8.1.6 discusses $z$ intervals. The distinction here is that we assume we know the variance $\sigma^2$, so that $\sigma$ is used instead of $s$, and a $z$ critical value is used rather than a $t$ critical value. However, in practice we seldom know $\sigma^2$. 
Hypothesis Testing

Remember, the confidence interval yields a set of values for $\mu$ that are consistent with the data.

In practice, sometimes we are interested in whether $\mu$ is equal to some specific value, say $\mu_0$.

**Example VII.D**

In Example VII.A, the researchers are interested in whether the mean cholesterol level of children with fathers who died of heart disease is the same as the national average. In other words, if $\mu$ represents the true mean cholesterol level of these children, we want to know if $\mu = 175$ mg/dL.
Hypothesis Testing: Applying the Scientific Method

Hypothesis testing follows these steps:

• Construct competing hypotheses, which we generally refer to as the null hypothesis and the alternative hypothesis.
• Sample from the population of interest, observing the sample mean and its standard error.
• Assume for the sake of argument that the null hypothesis is true. Compute a test statistic, which determines how far away from the hypothesized mean our sample mean lies.
• Compute a probability – called a \( p \)-value – that represents the likelihood of having observed our data given that the null hypothesis is true. Small \( p \)-values indicate that our data are inconsistent with the null hypothesis, and that the evidence favors the alternative hypothesis.
Determining Hypotheses

The null hypothesis is denoted by $H_0$. It is generally a simple hypothesis – that is, it is the hypothesis that $\mu$ is equal to the specific value of interest, say $\mu_0$. We express this as

$$H_0: \mu = \mu_0.$$

The alternative hypothesis is denoted by $H_A$. It is the hypothesis that $H_0$ is not true, meaning literally: “Not $H_0$”. We express this as

$$H_A: \mu \neq \mu_0.$$

Our object is to determine which of these hypotheses the evidence supports.
Example VII.E

State the null and alternative hypotheses for Example VII.A. Explain what these hypotheses mean in words.
Deciding Between Competing Hypotheses

Once we have constructed our hypotheses and observed the data, we compute a test statistic and associated $p$-value. Note that we are weighing the evidence for or against $H_0$.

What are the possible outcomes of this exercise?
The Test Statistic

In light of our hypotheses, our next step is to measure how “extreme” our observed data is, assuming that the null hypothesis is true. In particular, the test statistic measures how many standard errors the sample mean is away from the expected mean under the null hypothesis. The further away from $\mu_0$ that $\bar{X}$ is observed to be, the more evidence this provides against the null.

The test statistic in the case of a single population mean is given by

$$T = \frac{\sqrt{n} (\bar{X} - \mu_0)}{s}.$$ 

Note that under $H_0$, $T \sim t_{n-1}$. 
Example VII.F

What is the value of the test statistic for the cholesterol data, given the hypotheses in Example VII.E? What is the approximate distribution of this test statistic?
The $p$-value

The $p$-value (for “probability value”) is a measure of the probability of our having observed data as “extreme” as ours, given that the null hypothesis is true. Against a two-sided alternative hypothesis such as $H_A : \mu \neq \mu_0$, a sample mean that is either larger or smaller than the hypothesized mean $\mu_0$ provides evidence against the null hypothesis. Therefore, the two-sided $p$-value is given by

$$p = 2P(T > |t|),$$

where $T \sim t_{n-1}$. 
Example VII.G

What is the $p$-value for the cholesterol data in testing the hypotheses in Example VII.E?

What does this $p$-value represent in words?

Do these data provide evidence against the null hypothesis?
Interpreting the $p$-value

- A small $p$-value indicates that our data are improbable *assuming that the null hypothesis is true*. Therefore, a small $p$-value is evidence *against* the null. If the $p$-value is sufficiently small, then we **reject the null hypothesis $H_0$ in favor of the alternative $H_A$**. This result is often referred to as **statistically significant**.

- A relatively large $p$-value indicates that our data are *consistent* with the null hypothesis. In this case, we **fail to reject $H_0$**. Such a result is often designated as **not statistically significant**.

- Review the material in Section 8.2.2 carefully.
The Use of Significance Levels

How small should a $p$-value be in order for us to reject the null hypothesis $H_0$?

In practice, many investigators use the threshold level of 0.05, although this is entirely subjective.

If we use a cut-off value against which to compare the $p$-value in order to determine statistical significance, such a cut-off is called a significance level. We denote the significance level by $\alpha$.

Always keep in mind, however, that statistical significance does not automatically imply practical significance!!
Example VII.H

Is the $p$-value computed in Example VII.G significant at the $\alpha = 0.10$ significance level?

Is it significant at the $\alpha = 0.05$ level?

Is it significant at the $\alpha = 0.01$ level?
Comments on One-Sided Tests and \( z \)-Tests

The text spends some time discussing so-called one-sided tests and \( z \)-tests.

One-sided tests are those with alternative hypotheses such as \( H_A: \mu < \mu_0 \), or \( H_A: \mu > \mu_0 \). However – like one-sided confidence intervals – such tests are not often used in practice, and so they’re not really worth mentioning within the scope of this class.

The \( z \)-test is the version of the test statistic with known population variance \( \sigma^2 \). As with the \( z \)-interval, the \( z \)-test is not worth describing at length, since generally we do not know \( \sigma^2 \). We simply rely on the estimated variance \( s^2 \).
Duality of Confidence Intervals and Tests

It’s no coincidence that we use the symbol $\alpha$ both in the context of $t$-intervals and $t$-tests. If we use a significance level $\alpha$ in testing a hypothesis $H_0: \mu = \mu_0$, then the resulting $t$-test is directly related to the $(1 - \alpha)100\%$ $t$-interval. Consider:

- What does $\alpha$ represent in computing a $(1 - \alpha)100\%$ confidence interval?

- What does $\alpha$ represent in the context of hypothesis testing?

It turns out that all of the values within the $(1 - \alpha)100\%$ $t$-interval represent possible values of $\mu_0$ for which we would fail to reject $H_0: \mu = \mu_0$ at the significance level $\alpha$. All of the values outside of the confidence interval are values for which we would reject at the $\alpha$-level (i.e., the $p$-value for each of these tests would be $< \alpha$).
Example VII.I

Based solely on the confidence interval computed in Example VII.B, would you reject the null hypothesis $H_0: \mu = 175$, versus the alternative $H_0: \mu = 175$, using a significance level of $\alpha = 0.05$?

Would you reject at a significance level of $\alpha = 0.10$?

Would you reject at a significance level of $\alpha = 0.01$?