Lecture 11 - Hypothesis Tests for a Mean

Sta102 / BME102

March 3, 2016

Colin Rundel

Recap

Inference using CIs for sample means

When conditions for CLT are met and σ is unknown:

$$\frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t_{df=n-1}$$

Inference using CIs for sample means

When conditions for CLT are met and σ is unknown:

$$\frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t_{df=n-1}$$

Conditions (same as CLT):

- *Independent observations* - random sample, if sampling without replacement $n < 10\%$ of population
- *Sample size* - $> 20 - 30$ is usually reasonable, population not overly skewed or heavy/light tailed

Inference using CIs for sample means

When conditions for CLT are met and σ is unknown:

$$\frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t_{df=n-1}$$

Conditions (same as CLT):

- *Independent observations* - random sample, if sampling without replacement $n < 10\%$ of population
- *Sample size* - $> 20 - 30$ is usually reasonable, population not overly skewed or heavy/light tailed

Confidence interval:

$$\bar{X} \pm t^{\star}_{df=n-1} \frac{s}{\sqrt{n}}$$

Hypothesis Tests for one mean

Null Value Hypothesis Testing framework

- We start with a *null hypothesis* (H_0) that represents the status quo.

Null Value Hypothesis Testing framework

- We start with a *null hypothesis (H_0)* that represents the status quo.
- We develop an *alternative hypothesis (H_A)* that represents our research question (what we're testing for). It should be mutually exclusive to H_0.

Null Value Hypothesis Testing framework

- We start with a *null hypothesis* (H_0) that represents the status quo.
- We develop an *alternative hypothesis* (H_A) that represents our research question (what we're testing for). It should be mutually exclusive to H_0.
- We conduct the hypothesis test under the assumption that the null hypothesis is true, either via simulation or theoretical methods.

Null Value Hypothesis Testing framework

- We start with a *null hypothesis (H_0)* that represents the status quo.
- We develop an *alternative hypothesis (H_A)* that represents our research question (what we're testing for). It should be mutually exclusive to H_0.
- We conduct the hypothesis test under the assumption that the null hypothesis is true, either via simulation or theoretical methods.
- We examine how likely our data (or something more extreme) is under this assumption, and use that as evidence against the null hypothesis (and hence for the alternative).

Example - Grade inflation?

In 2001 the average GPA of students at Duke University was 3.37. Last semester Duke students in a Stats class were surveyed and ask for their current GPA. This survey had 63 respondents and yielded an average GPA of 3.56 with a standard deviation of 0.31.

Assuming that this sample is random and representative of all Duke students, do these data provide convincing evidence that the average GPA of Duke students has *changed* over the last decade?

Setting the hypotheses

- The population *parameter of interest* is the average GPA of current Duke students.

Setting the hypotheses

- The population *parameter of interest* is the average GPA of current Duke students.

- There may be two explanations why our sample mean is higher than the average GPA from 2001.
    - The true population mean has changed.
    - The true population mean remained at 3.37, the difference between the true population mean and the sample mean is simply due to natural sampling variability.

Setting the hypotheses

- The population *parameter of interest* is the average GPA of current Duke students.

- There may be two explanations why our sample mean is higher than the average GPA from 2001.
  - The true population mean has changed.
  - The true population mean remained at 3.37, the difference between the true population mean and the sample mean is simply due to natural sampling variability.

- We start with the assumption that nothing has changed.

$$H_0 : \mu = 3.37$$

Setting the hypotheses

- The population *parameter of interest* is the average GPA of current Duke students.

- There may be two explanations why our sample mean is higher than the average GPA from 2001.
    - The true population mean has changed.
    - The true population mean remained at 3.37, the difference between the true population mean and the sample mean is simply due to natural sampling variability.

- We start with the assumption that nothing has changed.

$$H_0 : \mu = 3.37$$

- We test the claim that average GPA has changed.

$$H_A : \mu \neq 3.37$$

Making a decision - p-values

We want to make a decision about whether we think H_0 or H_A is correct, to do this in a principled / quantitative way we calculate what is known as a *p-value*.

Making a decision - p-values

We want to make a decision about whether we think H_0 or H_A is correct, to do this in a principled / quantitative way we calculate what is known as a *p-value*.

- The *p-value* is the probability of observing data *at least* as favorable to the alternative hypothesis as our observed data, assuming the null hypothesis is true.

Making a decision - p-values

We want to make a decision about whether we think H_0 or H_A is correct, to do this in a principled / quantitative way we calculate what is known as a *p-value*.

- The *p-value* is the probability of observing data *at least* as favorable to the alternative hypothesis as our observed data, assuming the null hypothesis is true.
- *Small* p-value ($< \alpha$, usually 5%) we claim it is very unlikely to observe these data (or more extreme) if the null hypothesis were true, and therefore *reject H_0*.

Making a decision - p-values

We want to make a decision about whether we think H_0 or H_A is correct, to do this in a principled / quantitative way we calculate what is known as a *p-value*.

- The *p-value* is the probability of observing data *at least* as favorable to the alternative hypothesis as our observed data, assuming the null hypothesis is true.
- *Small* p-value ($< \alpha$, usually 5%) we claim it is very unlikely to observe these data (or more extreme) if the null hypothesis were true, and therefore *reject H_0*.
- *Large* p-value ($> \alpha$) we claim it is likely to observe these data if the null hypothesis were true, and therefore *do not reject H_0*.

Making a decision - p-values

We want to make a decision about whether we think H_0 or H_A is correct, to do this in a principled / quantitative way we calculate what is known as a *p-value*.

- The *p-value* is the probability of observing data *at least* as favorable to the alternative hypothesis as our observed data, assuming the null hypothesis is true.
- *Small* p-value ($< \alpha$, usually 5%) we claim it is very unlikely to observe these data (or more extreme) if the null hypothesis were true, and therefore *reject* H_0.
- *Large* p-value ($> \alpha$) we claim it is likely to observe these data if the null hypothesis were true, and therefore *do not reject* H_0.
- We never accept H_0 since we're not in the business of trying to prove it. We just want to know if the data provide *convincing* evidence against H_0.

What is a p-value

What is a p-value:

- The probability of the observed data (sample statistic) or something more extreme in favor of the null hypothesis given the null hypothesis is true.
- Indirect evidence against H_0.

What a p-value *isn't*:

- A p-value is not the probably H_0 is true
- A p-value is not the probably H_A is false

Conditions for inference - GPA

Back to the GPA example, in order to perform inference on these data we need to use the CLT, and therefore we need to check the conditions:

Conditions for inference - GPA

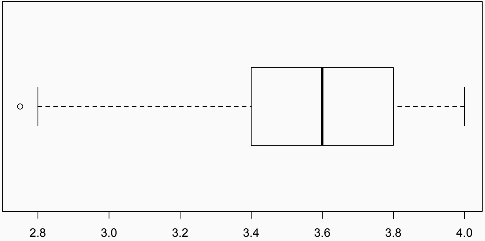
Back to the GPA example, in order to perform inference on these data we need to use the CLT, and therefore we need to check the conditions:

1. *Independence:*
   - We have already assumed this sample is random.
   - Assume sampling without replacement, but $63 < 10\%$ of all current Duke students.

   $\Rightarrow$ it appears reasonable to assume that GPA of one student in this sample is independent of another.

Conditions for inference - GPA

2. *Sample size / skew:* The distribution appears to be slightly left skewed (but not extremely) and $n = 63$ so we will assume that the sampling distribution of the sample means should be nearly normal by the CLT.



Calculating the p-value

p-value - probability of observing data at least as favorable to H_A as our current data set, if in fact H_0 is true (the true population mean $\mu = 3.37$).

In this case because we are not making any claims about GPAs going up or down, we need to consider GPA changes in both directions. E.g. a sample average GPA of 3.18 is just as much in favor of H_A as a sample average GPA of 3.56.

Calculating the p-value (cont.)

Drawing a Conclusion / Inference

$$\text{p-value} = 4.2 \times 10^{-6}$$

$$\text{p-value} = 4.2 \times 10^{-6}$$

If the true average GPA of Duke students is 3.37, there is approximately a 4.2×10^{-6} chance of observing a random sample of 63 Duke students with an average GPA of 3.56 and above or 3.18 and below.

- This is a very small probability, it seems very unlikely that a 3.56 sample average GPA could have happened by chance.

Drawing a Conclusion / Inference

$$\text{p-value} = 4.2 \times 10^{-6}$$

If the true average GPA of Duke students is 3.37, there is approximately a 4.2×10^{-6} chance of observing a random sample of 63 Duke students with an average GPA of 3.56 and above or 3.18 and below.

- This is a very small probability, it seems very unlikely that a 3.56 sample average GPA could have happened by chance.
- Since the p-value is *small* (lower than 5%) we *reject* H_0.

Drawing a Conclusion / Inference

$$\text{p-value} = 4.2 \times 10^{-6}$$

If the true average GPA of Duke students is 3.37, there is approximately a 4.2×10^{-6} chance of observing a random sample of 63 Duke students with an average GPA of 3.56 and above or 3.18 and below.

- This is a very small probability, it seems very unlikely that a 3.56 sample average GPA could have happened by chance.
- Since the p-value is *small* (lower than 5%) we *reject H_0*.
- Claim - the data provide convincing evidence that Duke students' average GPA has changed since 2001. E.g. the difference between the null value of a 3.37 GPA and observed sample mean of 3.56 GPA is *not due to chance /*

Example - College applications

A similar survey asked how many colleges each student had applied to. 206 students responded to this question and the sample yielded an average of 9.7 college applications with a standard deviation of 7. The College Board website states that counselors recommend students apply to 8 colleges. What would be the correct set of hypotheses be to test if these data provide convincing evidence that the average number of colleges Duke students apply to is *greater* than the number recommended by the College Board.

Example - College applications

A similar survey asked how many colleges each student had applied to. 206 students responded to this question and the sample yielded an average of 9.7 college applications with a standard deviation of 7. The College Board website states that counselors recommend students apply to 8 colleges. What would be the correct set of hypotheses be to test if these data provide convincing evidence that the average number of colleges Duke students apply to is *greater* than the number recommended by the College Board.

$$H_0 : \mu = 8$$
$$H_A : \mu > 8$$

Example - College applications

A similar survey asked how many colleges each student had applied to. 206 students responded to this question and the sample yielded an average of 9.7 college applications with a standard deviation of 7. The College Board website states that counselors recommend students apply to 8 colleges. What would be the correct set of hypotheses be to test if these data provide convincing evidence that the average number of colleges Duke students apply to is *greater* than the number recommended by the College Board.

$$H_0 : \mu = 8$$
$$H_A : \mu > 8$$

Are the conditions for inference met?

Example - College applications

A similar survey asked how many colleges each student had applied to. 206 students responded to this question and the sample yielded an average of 9.7 college applications with a standard deviation of 7. The College Board website states that counselors recommend students apply to 8 colleges. What would be the correct set of hypotheses be to test if these data provide convincing evidence that the average number of colleges Duke students apply to is *greater* than the number recommended by the College Board.

$$H_0 : \mu = 8$$
$$H_A : \mu > 8$$

Are the conditions for inference met?

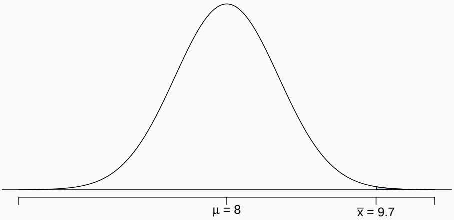
Yes - Independence ✓, Nearly Normal ✓

College Applications - p-value

p-value - probability of observing data at least as favorable to H_A as our current data set (a sample mean greater than 9.7), if in fact H_0 was true (the true population mean is 8).

College Applications - p-value

p-value - probability of observing data at least as favorable to H_A as our current data set (a sample mean greater than 9.7), if in fact H_0 was true (the true population mean is 8).



$$p - value < 0.005$$

College Applications - Making a decision

$$p - value < 0.005$$

If the true average of the number of colleges Duke students applied to is 8, there is less than a 0.005 chance of observing a random sample of 206 Duke students who on average apply to 9.7 *or more* schools.

$$p - value < 0.005$$

If the true average of the number of colleges Duke students applied to is 8, there is less than a 0.005 chance of observing a random sample of 206 Duke students who on average apply to 9.7 *or more* schools.

- This is a very small probability, it seems very unlikely that a sample mean of 9.7 or more schools is likely to happen simply by chance.

College Applications - Making a decision

$$p - value < 0.005$$

If the true average of the number of colleges Duke students applied to is 8, there is less than a 0.005 chance of observing a random sample of 206 Duke students who on average apply to 9.7 *or more* schools.

- This is a very small probability, it seems very unlikely that a sample mean of 9.7 or more schools is likely to happen simply by chance.
- Since p-value is *low* (lower than 5%) we *reject* H_0.

College Applications - Making a decision

$$p - value < 0.005$$

If the true average of the number of colleges Duke students applied to is 8, there is less than a 0.005 chance of observing a random sample of 206 Duke students who on average apply to 9.7 *or more* schools.

- This is a very small probability, it seems very unlikely that a sample mean of 9.7 or more schools is likely to happen simply by chance.
- Since p-value is *low* (lower than 5%) we *reject* H_0.
- The data provide convincing evidence that Duke students apply on average to more than 8 schools.

What about a confidence interval?

We can also assess this claim using a confidence interval.

What about a confidence interval?

We can also assess this claim using a confidence interval.

$$\bar{X} = 9.7 \qquad s^2 = 7^2 \qquad n = 206$$

What about a confidence interval?

We can also assess this claim using a confidence interval.

$$\bar{X} = 9.7 \qquad s^2 = 7^2 \qquad n = 206$$

We construct a 95% confidence interval using $t^*_{df=205} \approx t^*_{df=200} = 1.97$,

Example - Sleep

A poll by the National Sleep Foundation found that college students average about 7 hours of sleep per night. A sample of 169 Duke students (you!) yielded an average of 6.88 hours, with a standard deviation of 0.94 hours. Assuming that this is a random sample representative of all Duke students, a hypothesis test was conducted to evaluate if Duke students on average sleep *less than* 7 hours per night. The p-value for this hypothesis test is 0.0485.

Example - Sleep

A poll by the National Sleep Foundation found that college students average about 7 hours of sleep per night. A sample of 169 Duke students (you!) yielded an average of 6.88 hours, with a standard deviation of 0.94 hours. Assuming that this is a random sample representative of all Duke students, a hypothesis test was conducted to evaluate if Duke students on average sleep *less than* 7 hours per night. The p-value for this hypothesis test is 0.0485.

What are the hypotheses being tested?

Example - Sleep

A poll by the National Sleep Foundation found that college students average about 7 hours of sleep per night. A sample of 169 Duke students (you!) yielded an average of 6.88 hours, with a standard deviation of 0.94 hours. Assuming that this is a random sample representative of all Duke students, a hypothesis test was conducted to evaluate if Duke students on average sleep *less than* 7 hours per night. The p-value for this hypothesis test is 0.0485.

What are the hypotheses being tested?

$$H_0 : \mu = 7$$
$$H_A : \mu < 7$$

Example - Sleep

A poll by the National Sleep Foundation found that college students average about 7 hours of sleep per night. A sample of 169 Duke students (you!) yielded an average of 6.88 hours, with a standard deviation of 0.94 hours. Assuming that this is a random sample representative of all Duke students, a hypothesis test was conducted to evaluate if Duke students on average sleep *less than* 7 hours per night. The p-value for this hypothesis test is 0.0485.

What are the hypotheses being tested?

$$H_0 : \mu = 7$$
$$H_A : \mu < 7$$

What is the correct inference for this situation?

Two-sided hypothesis test

If the research question had been "Do the data provide convincing evidence that the average amount of sleep Duke students get per night is *different* than the national average?", how would the null and alternative hypotheses change?

Two-sided hypothesis test

If the research question had been "Do the data provide convincing evidence that the average amount of sleep Duke students get per night is *different* than the national average?", how would the null and alternative hypotheses change?

$$H_0 : \mu = 7$$

$$H_A : \mu \neq 7$$

Two-sided hypothesis test

If the research question had been “Do the data provide convincing evidence that the average amount of sleep Duke students get per night is *different* than the national average?”, how would the null and alternative hypotheses change?

$$H_0 : \mu = 7$$
$$H_A : \mu \neq 7$$

How would the p-value change?

Two-sided hypothesis test

If the research question had been "Do the data provide convincing evidence that the average amount of sleep Duke students get per night is *different* than the national average?", how would the null and alternative hypotheses change?

$$H_0 : \mu = 7$$

$$H_A : \mu \neq 7$$

How would the p-value change?

What about a confidence interval?

Once again, we can also assess this claim using a confidence interval.

What about a confidence interval?

Once again, we can also assess this claim using a confidence interval.

$$\bar{X} = 6.88 \qquad s^2 = 0.94^2 \qquad n = 169$$

What about a confidence interval?

Once again, we can also assess this claim using a confidence interval.

$$\bar{X} = 6.88 \qquad s^2 = 0.94^2 \qquad n = 169$$

We construct a 95% confidence interval using $t^*_{df=168} \approx t^*_{df=150} = 1.98$,

Recap: Null Value Hypothesis Testing

Regardless of the sample statistic of interest, all null value hypothesis testing takes exactly the same form.

1. Set the hypotheses
2. Check assumptions and conditions
3. Calculate a *test statistic* and a p-value (draw a picture!)
4. Make a decision, and interpret it in context of the research question

Recap: Null Value Hypothesis Testing - Sample Means

1. Set the hypotheses
2. Check assumptions and conditions
3. Calculate a *test statistic* and a p-value (draw a picture!)
4. Make a decision, and interpret it in context of the research question

Recap: Null Value Hypothesis Testing - Sample Means

1. Set the hypotheses
   - $H_0 : \mu = \textit{null value}$
   - $H_A : \mu <$ or $>$ or $\neq$ *null value*
2. Check assumptions and conditions
3. Calculate a *test statistic* and a p-value (draw a picture!)
4. Make a decision, and interpret it in context of the research question

Recap: Null Value Hypothesis Testing - Sample Means

1. Set the hypotheses
   - $H_0 : \mu = \textit{null value}$
   - $H_A : \mu <$ or $>$ or $\neq$ *null value*
2. Check assumptions and conditions
   - Independence: random sample/assignment, 10% condition when sampling without replacement
   - Normality/Sample size: nearly normal population or n large enough, w/ no extreme skew or tail weirdness
3. Calculate a *test statistic* and a p-value (draw a picture!)
4. Make a decision, and interpret it in context of the research question

Recap: Null Value Hypothesis Testing - Sample Means

1. Set the hypotheses
   - $H_0 : \mu = \textit{null value}$
   - $H_A : \mu < \text{or} > \text{or} \neq \textit{null value}$
2. Check assumptions and conditions
   - Independence: random sample/assignment, 10% condition when sampling without replacement
   - Normality/Sample size: nearly normal population or n large enough, w/ no extreme skew or tail weirdness
3. Calculate a *test statistic* and a p-value (draw a picture!)

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \qquad T = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

4. Make a decision, and interpret it in context of the research question

Recap: Null Value Hypothesis Testing - Sample Means

1. Set the hypotheses
    - $H_0 : \mu = null\ value$
    - $H_A : \mu <$ or $>$ or $\neq$ $null\ value$
2. Check assumptions and conditions
    - Independence: random sample/assignment, 10% condition when sampling without replacement
    - Normality/Sample size: nearly normal population or n large enough, w/ no extreme skew or tail weirdness
3. Calculate a *test statistic* and a p-value (draw a picture!)

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \qquad T = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

4. Make a decision, and interpret it in context of the research question
    - If p-value $< \alpha$, reject H_0
    - If p-value $> \alpha$, do not reject H_0