Lesson 10:
Chapter 6 Sections 3-4

Caleb Moxley

BSC Mathematics

7 October 15

§6.3 Use and Abuse of Tests

Tests of significance have limitations, and it's important to understand these limits. In §6.3, we'll discuss

§6.3 Use and Abuse of Tests

Tests of significance have limitations, and it's important to understand these limits. In §6.3, we'll discuss

1. what levels of significance mean,

§6.3 Use and Abuse of Tests

Tests of significance have limitations, and it's important to understand these limits. In §6.3, we'll discuss

1. what levels of significance mean,
2. the difference between statistical and practical significance,

§6.3 Use and Abuse of Tests

Tests of significance have limitations, and it's important to understand these limits. In §6.3, we'll discuss

1. what levels of significance mean,
2. the difference between statistical and practical significance,
3. what statistical significance does not mean,

§6.3 Use and Abuse of Tests

Tests of significance have limitations, and it's important to understand these limits. In §6.3, we'll discuss

1. what levels of significance mean,
2. the difference between statistical and practical significance,
3. what statistical significance does not mean,
4. what lack of significance means,

§6.3 Use and Abuse of Tests

Tests of significance have limitations, and it's important to understand these limits. In §6.3, we'll discuss

1. what levels of significance mean,
2. the difference between statistical and practical significance,
3. what statistical significance does not mean,
4. what lack of significance means,
5. when statistical inference is inappropriate to use, and

§6.3 Use and Abuse of Tests

Tests of significance have limitations, and it's important to understand these limits. In §6.3, we'll discuss

1. what levels of significance mean,
2. the difference between statistical and practical significance,
3. what statistical significance does not mean,
4. what lack of significance means,
5. when statistical inference is inappropriate to use, and
6. searching for significance.

What Levels of Significance Mean:

There's no cutoff for what is and is not significant.

What Levels of Significance Mean:

There's no cutoff for what is and is not significant. Our choice of α really is arbitrary. Obviously a P-value of 0.051 suggests that the data might be statistically significant, even though the level of significance might have been set at $\alpha = 0.05$.

§6.3 Use and Abuse of Tests

What Levels of Significance Mean:

There's no cutoff for what is and is not significant. Our choice of α really is arbitrary. Obviously a P-value of 0.051 suggests that the data might be statistically significant, even though the level of significance might have been set at $\alpha = 0.05$. But some data is obviously statistically significant.

What Levels of Significance Mean:

There's no cutoff for what is and is not significant. Our choice of α really is arbitrary. Obviously a P-value of 0.051 suggests that the data might be statistically significant, even though the level of significance might have been set at $\alpha = 0.05$. But some data is obviously statistically significant. If the P-value for a test were 0.000001, then this result would arise from the underlying assumptions in the test only one in a million times! So the assumptions are almost certainly wrong. Sometimes a very small P-value is denoted by $P < 0.001$.

§6.3 Use and Abuse of Tests

Statistical vs. Practical Significance:

Say a weight loss plan has a 99% chance of resulting in weight loss based on a study.

§6.3 Use and Abuse of Tests

Statistical vs. Practical Significance:

Say a weight loss plan has a 99% chance of resulting in weight loss based on a study. Would this be statistically significant?

§6.3 Use and Abuse of Tests

Statistical vs. Practical Significance:

Say a weight loss plan has a 99% chance of resulting in weight loss based on a study. Would this be statistically significant? What if you additionally knew that the amount of weight lost would be 0.25lbs over the course of 2 years?

§6.3 Use and Abuse of Tests

Statistical vs. Practical Significance:

Say a weight loss plan has a 99% chance of resulting in weight loss based on a study. Would this be statistically significant? What if you additionally knew that the amount of weight lost would be 0.25lbs over the course of 2 years? Would this still be a worthwhile weight loss program?

§6.3 Use and Abuse of Tests

Statistical vs. Practical Significance:

Say a weight loss plan has a 99% chance of resulting in weight loss based on a study. Would this be statistically significant? What if you additionally knew that the amount of weight lost would be 0.25lbs over the course of 2 years? Would this still be a worthwhile weight loss program?

Often in very large samples, even small deviations from the null hypothesis will be significant. But this statistical significance doesn't mean that the significance is practical!

Statistical vs. Practical Significance:

Say a weight loss plan has a 99% chance of resulting in weight loss based on a study. Would this be statistically significant? What if you additionally knew that the amount of weight lost would be 0.25lbs over the course of 2 years? Would this still be a worthwhile weight loss program?

Often in very large samples, even small deviations from the null hypothesis will be significant. But this statistical significance doesn't mean that the significance is practical! In particular, strong significance doesn't imply dramatic difference from the null hypothesis. It just means that the results imply that our assumption was almost certainly wrong — but perhaps only a little wrong.

What Lack of Significance Means:

A study of a vaccine's effectiveness determined that 99% CI for the proportion change in number of infections was $(-0.25, 0.30)$. A newspaper reported that the vaccine didn't cause any change in the infection rate — because the null hypothesis in this case would have been that there is no effect from the vaccine and that case (where the proportion change is 0) is included in the 99% CI. What's wrong with this?

§6.3 Use and Abuse of Tests

What Lack of Significance Means:

A study of a vaccine's effectiveness determined that 99% CI for the proportion change in number of infections was $(-0.25, 0.30)$. A newspaper reported that the vaccine didn't cause any change in the infection rate — because the null hypothesis in this case would have been that there is no effect from the vaccine and that case (where the proportion change is 0) is included in the 99% CI. What's wrong with this?

Absence of evidence is not evidence of absence! In particular, there could be at most a 25% decrease or a 30% increase due to the vaccine (with 99% confidence). What really should be said is that the effectiveness of the vaccine has not been proven. Further study is necessary.

Statistical Inference Isn't Always Appropriate:

A study is interested in determining the veracity of the claim that 65% of voters in a statewide election are republicans, so they conduct an exit poll at a randomly selected polling station. They survey the 2581 voters who voted at the polling station and created a 99.99% CI for the proportion of registered republicans: (0.7500,0.7501). Does this support the claim?

Statistical Inference Isn't Always Appropriate:

A study is interested in determining the veracity of the claim that 65% of voters in a statewide election are republicans, so they conduct an exit poll at a randomly selected polling station. They survey the 2581 voters who voted at the polling station and created a 99.99% CI for the proportion of registered republicans: (0.7500,0.7501). Does this support the claim? **Should this test of significance be conducted, though?**

Statistical Inference Isn't Always Appropriate:

A study is interested in determining the veracity of the claim that 65% of voters in a statewide election are republicans, so they conduct an exit poll at a randomly selected polling station. They survey the 2581 voters who voted at the polling station and created a 99.99% CI for the proportion of registered republicans: (0.7500,0.7501). Does this support the claim? **Should this test of significance be conducted, though?** No! The sample was bad! Tests of significance cannot correct for bad sampling!

§6.3 Use and Abuse of Tests

Searching for Significance:

Amazon mined its data and discovered that its customers who use Paypal to make purchases seemed to be more likely to make multiple purchases in the same month. Amazon sought verification of this hypothesis from a third party. Why shouldn't the third party use Amazon's data to test this hypothesis?

Searching for Significance:

Amazon mined its data and discovered that its customers who use Paypal to make purchases seemed to be more likely to make multiple purchases in the same month. Amazon sought verification of this hypothesis from a third party. Why shouldn't the third party use Amazon's data to test this hypothesis? **You should never use data which suggests a hypothesis to verify the same hypothesis!** Always conduct a new experiment/collect new data.

§6.3 Power and Inference as a Decision

What does α really mean?

§6.3 Power and Inference as a Decision

What does α really mean? Well, α describes the "fringe" values of the sampling distribution — it's related to the values that would be unusual under the assumptions of the null hypothesis.

§6.3 Power and Inference as a Decision

What does α really mean? Well, α describes the "fringe" values of the sampling distribution — it's related to the values that would be unusual under the assumptions of the null hypothesis. If the null hypothesis were indeed true, then, with probability α, one would reject the null hypothesis.

§6.3 Power and Inference as a Decision

What does α really mean? Well, α describes the "fringe" values of the sampling distribution — it's related to the values that would be unusual under the assumptions of the null hypothesis. If the null hypothesis were indeed true, then, with probability α, one would reject the null hypothesis.

Definition (type one error)

A **Type I Error** is the error of rejecting the null hypothesis when it is actually true — the probability of making a Type I Error is exactly α.

§6.3 Power and Inference as a Decision

You might also be interested in knowing how sensitive a test is, i.e. how often it will reject H_0 when an alternative to H_0 is actually true.

§6.3 Power and Inference as a Decision

You might also be interested in knowing how sensitive a test is, i.e. how often it will reject H_0 when an alternative to H_0 is actually true.

Definition (power)

The **power** of an α-level test is the probability that the test will reject H_0 when H_a is true.

§6.3 Power and Inference as a Decision

You might also be interested in knowing how sensitive a test is, i.e. how often it will reject H_0 when an alternative to H_0 is actually true.

Definition (power)

The **power** of an α-level test is the probability that the test will reject H_0 when H_a is true.

You can improve power by increasing sample size, but it's not an easy task to compute the power of an α-level test.

§6.3 Power and Inference as a Decision

If power measures how often you'll correctly reject H_0 when H_a is true, then obviously $1 - \text{(power)}$ measures how often you'll support H_0 when H_a is true! We call this a Type II Error.

§6.3 Power and Inference as a Decision

If power measures how often you'll correctly reject H_0 when H_a is true, then obviously $1 - \text{(power)}$ measures how often you'll support H_0 when H_a is true! We call this a Type II Error.

Definition (type two error)

A **Type II Error** is the error of failing to reject the null hypothesis when the alternative hypothesis is actually true. If β is the power of a test, then the probability of a Type II Error is $1 - \beta$.

§6.3 Power and Inference as a Decision

These ideas are particularly useful when we use statistical inference to make a decision.

§6.3 Power and Inference as a Decision

These ideas are particularly useful when we use statistical inference to make a decision.

Example (bad apples)

Remember our example of bad apples? An orchard shipped 2000 apples to an applesauce company, and the applesauce company would only accept the apples if the proportion of bad apples was under some value.

§6.3 Power and Inference as a Decision

These ideas are particularly useful when we use statistical inference to make a decision.

Example (bad apples)

Remember our example of bad apples? An orchard shipped 2000 apples to an applesauce company, and the applesauce company would only accept the apples if the proportion of bad apples was under some value. Instead of testing all 2000 apples, the applesauce company could have simply performed an α- and β-level test to test the claim that the proportion was no more than 4.1%.

§6.3 Power and Inference as a Decision

These ideas are particularly useful when we use statistical inference to make a decision.

Example (bad apples)

Remember our example of bad apples? An orchard shipped 2000 apples to an applesauce company, and the applesauce company would only accept the apples if the proportion of bad apples was under some value. Instead of testing all 2000 apples, the applesauce company could have simply performed an α- and β-level test to test the claim that the proportion was no more than 4.1%. What is the probability that the shipment would be accepted even though the proportion was over 4.1%?

§6.3 Power and Inference as a Decision

These ideas are particularly useful when we use statistical inference to make a decision.

Example (bad apples)

Remember our example of bad apples? An orchard shipped 2000 apples to an applesauce company, and the applesauce company would only accept the apples if the proportion of bad apples was under some value. Instead of testing all 2000 apples, the applesauce company could have simply performed an α- and β-level test to test the claim that the proportion was no more than 4.1%. What is the probability that the shipment would be accepted even though the proportion was over 4.1%? $1 - \beta$.

§6.3 Power and Inference as a Decision

These ideas are particularly useful when we use statistical inference to make a decision.

Example (bad apples)

Remember our example of bad apples? An orchard shipped 2000 apples to an applesauce company, and the applesauce company would only accept the apples if the proportion of bad apples was under some value. Instead of testing all 2000 apples, the applesauce company could have simply performed an α- and β-level test to test the claim that the proportion was no more than 4.1%. What is the probability that the shipment would be accepted even though the proportion was over 4.1%? $1-\beta$. What is the probability that it would rejected even though it was no more than 4.1%?

§6.3 Power and Inference as a Decision

These ideas are particularly useful when we use statistical inference to make a decision.

Example (bad apples)

Remember our example of bad apples? An orchard shipped 2000 apples to an applesauce company, and the applesauce company would only accept the apples if the proportion of bad apples was under some value. Instead of testing all 2000 apples, the applesauce company could have simply performed an α- and β-level test to test the claim that the proportion was no more than 4.1%. What is the probability that the shipment would be accepted even though the proportion was over 4.1%? $1-\beta$. What is the probability that it would rejected even though it was no more than 4.1%? α.

§6.3 Power and Inference as a Decision

These ideas are particularly useful when we use statistical inference to make a decision.

Example (bad apples)

Remember our example of bad apples? An orchard shipped 2000 apples to an applesauce company, and the applesauce company would only accept the apples if the proportion of bad apples was under some value. Instead of testing all 2000 apples, the applesauce company could have simply performed an α- and β-level test to test the claim that the proportion was no more than 4.1%. What is the probability that the shipment would be accepted even though the proportion was over 4.1%? $1 - \beta$. What is the probability that it would rejected even though it was no more than 4.1%? α. What's the probability that it would be correctly rejected?

§6.3 Power and Inference as a Decision

These ideas are particularly useful when we use statistical inference to make a decision.

Example (bad apples)

Remember our example of bad apples? An orchard shipped 2000 apples to an applesauce company, and the applesauce company would only accept the apples if the proportion of bad apples was under some value. Instead of testing all 2000 apples, the applesauce company could have simply performed an α- and β-level test to test the claim that the proportion was no more than 4.1%. What is the probability that the shipment would be accepted even though the proportion was over 4.1%? $1 - \beta$. What is the probability that it would rejected even though it was no more than 4.1%? α. What's the probability that it would be correctly rejected? β.

§6.3 Power and Inference as a Decision

These ideas are particularly useful when we use statistical inference to make a decision.

Example (bad apples)

Remember our example of bad apples? An orchard shipped 2000 apples to an applesauce company, and the applesauce company would only accept the apples if the proportion of bad apples was under some value. Instead of testing all 2000 apples, the applesauce company could have simply performed an α- and β-level test to test the claim that the proportion was no more than 4.1%. What is the probability that the shipment would be accepted even though the proportion was over 4.1%? $1 - \beta$. What is the probability that it would rejected even though it was no more than 4.1%? α. What's the probability that it would be correctly rejected? β. Correctly accepted?

These ideas are particularly useful when we use statistical inference to make a decision.

Example (bad apples)

Remember our example of bad apples? An orchard shipped 2000 apples to an applesauce company, and the applesauce company would only accept the apples if the proportion of bad apples was under some value. Instead of testing all 2000 apples, the applesauce company could have simply performed an α- and β-level test to test the claim that the proportion was no more than 4.1%. What is the probability that the shipment would be accepted even though the proportion was over 4.1%? $1 - \beta$. What is the probability that it would rejected even though it was no more than 4.1%? α. What's the probability that it would be correctly rejected? β. Correctly accepted? $1 - \alpha$.