Lecture 15: Type I, II errors, power of a hypothesis test

Announcements:

- Midterm exam 2 will be a week from today: Tuesday, 7 December, 14:00 - 16:00.

Lecture 15: Type I, II errors, power of a hypothesis test

Announcements:

- Midterm exam 2 will be a week from today: Tuesday, 7 December, 14:00 - 16:00.
- Exam will cover problem sets 5, 6, 7, and 8, and selected problems from Problem Set 4.

Lecture 15: Type I, II errors, power of a hypothesis test

Announcements:

- Midterm exam 2 will be a week from today: Tuesday, 7 December, 14:00 - 16:00.
- Exam will cover problem sets 5, 6, 7, and 8, and selected problems from Problem Set 4.
- Answer keys for Problem Sets 5, 6, and 7 are posted on the web. You'll write down the answers to Problem Set 8 yourselves this week. Tomislav will help you with this on Thursday.

Lecture 15: Type I, II errors, power of a hypothesis test

Announcements:

- Midterm exam 2 will be a week from today: Tuesday, 7 December, 14:00 - 16:00.
- Exam will cover problem sets 5, 6, 7, and 8, and selected problems from Problem Set 4.
- Answer keys for Problem Sets 5, 6, and 7 are posted on the web. You'll write down the answers to Problem Set 8 yourselves this week. Tomislav will help you with this on Thursday.
- Today: statistical power, and errors of a hypothesis test.

Lecture 15: Type I, II errors, power of a hypothesis test

Announcements:

- Midterm exam 2 will be a week from today: Tuesday, 7 December, 14:00 - 16:00.
- Exam will cover problem sets 5, 6, 7, and 8, and selected problems from Problem Set 4.
- Answer keys for Problem Sets 5, 6, and 7 are posted on the web. You'll write down the answers to Problem Set 8 yourselves this week. Tomislav will help you with this on Thursday.
- Today: statistical power, and errors of a hypothesis test.
- Thursday: review session on all material for the test, especially this week's problem set.

Problem

Problem

Statistical significance is not the same as biological "significance" (or meaning).

Problem

Statistical significance is not the same as biological "significance" (or meaning).

If you do not reject the null hypothesis, that could be because:

Problem

Statistical significance is not the same as biological "significance" (or meaning).

If you do not reject the null hypothesis, that could be because:
1) the null hypothesis is true, or

Problem

Statistical significance is not the same as biological "significance" (or meaning).

If you do not reject the null hypothesis, that could be because:
1) the null hypothesis is true, or
2) your test was not powerful enough: your sample size was too low.

Problem

Statistical significance is not the same as biological "significance" (or meaning).

If you do not reject the null hypothesis, that could be because:
1) the null hypothesis is true, or
2) your test was not powerful enough: your sample size was too low.

What is the power of a test?

Problem

As you recall, a screening test has four combinations of test outcome/condition.

Example

As you recall, a screening test has four combinations of test outcome/condition.

Errors

Situation	Decision	
	Accept H_0	Reject H_0
H_0 true	$1 - \alpha$	α
H_0 false	β	$1 - \beta$

Terminology

- **Type I error** is α.

Terminology

- **Type I error** is α.
- **Type II error** is β.

Terminology

- **Type I error** is α.
- **Type II error** is β.
- **Sensitivity** of the test is the power, which is $1 - \beta$.

Terminology

- **Type I error** is α.
- **Type II error** is β.
- **Sensitivity** of the test is the power, which is $1 - \beta$.
- **Specificity** of the test is $1 - \alpha$.

Basic power equations, one-sample Z-test

$$Z_{pow} = |Z_{crit}| - \frac{\delta}{\frac{\sigma}{\sqrt{n}}} = |Z_{crit}| - \frac{\delta\sqrt{n}}{\sigma}$$

$$n = \sigma^2 \frac{(Z_{pow} - Z_{crit})^2}{\delta^2}$$

$$\delta = \sigma \frac{(Z_{crit} - Z_{pow})}{\sqrt{n}}$$

Calculating power using R

Calculating power using R

$$\text{Power} = 1 - \text{pnorm}(Z_{pow})$$

Calculating power using R

$$\text{Power} = 1 - \text{pnorm}(Z_{pow})$$

$$Z_{pow} = \text{qnorm}(1 - \text{power})$$

Calculating power using R

$$\text{Power} = 1 - \text{pnorm}(Z_{pow})$$

$$Z_{pow} = \text{qnorm}(1 - \text{power})$$

$$Z_{crit} = \pm\text{qnorm}(1 - \alpha/T),$$

Calculating power using R

$$\text{Power} = 1 - \text{pnorm}(Z_{pow})$$

$$Z_{pow} = \text{qnorm}(1 - \text{power})$$

$$Z_{crit} = \pm\text{qnorm}(1 - \alpha/T),$$

T is the number of tails in the test. Note that the Z_{crit} is negative for a left-tailed test. + for right - tailed

Example

A buyer wants to buy a large batch of apples from a grower. But he does not want to buy if the true mean weight is less than 150 g. He can't weigh all the apples, but he can weigh a sample of n, and calculate the sample mean. Let's say the true standard deviation is 30g. If he weighs 20 apples and does a Z test he may or may not reject the null hypothesis. If the **true** mean weight of apples is 150 g, what is the probability that he rejects the null hypothesis, and rejects the apples? Assume he is doing a one-tailed Z-test, with a significance level of 0.05 ($\alpha = 0.05$). He rejects the apples only if he rejects the null hypothesis.

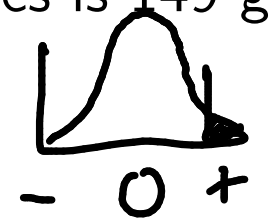
Example

$\sigma = 30 \quad n = 20$

$\delta = 1g \quad \alpha = 0.05$

$z_{crit} = 1.64$

What if the true mean weight of the grower's apples is 149 g? Will he reject the null hypothesis? 145 g? 140 g?



$$z_{pow} = z_{crit} - \frac{\delta\sqrt{n}}{\sigma}$$

$$= 1.64 - \frac{1\sqrt{20}}{30} = 1.49$$

power = 0.068

Example

$$LSD = z_{crit} \frac{\sigma}{\sqrt{n}}$$

What is the least significant difference (LSD) that the buyer must see before he rejects the grower's apples?

$$LSD = \underline{1.64} \cdot \frac{30}{\sqrt{20}} = 11.0$$

Decision: reject apples if $\bar{x} < 139$
if $n = 20$, $z_{crit} = 1.64$, $\sigma = 30$

Example

What is the probability that he will see the LSD if the true mean weight of the grower's apples is 149 g? 145 g? 140 g?

Example

$z_{crit} = 1.64$ $n = 20$ $\sigma = 30$ power $= 0.8$

$$\delta = \sigma \frac{(z_{crit} - z_{pow})}{\sqrt{n}}$$

z_{pow} = qnorm(1 − pow) = −0.84

Statistical power

If his sample size has to be 20, then what true difference (δ) will he be able to detect with his test 80% of the time? What if the sample size is 30? 50? 100? 500?

$$\delta = 30 \frac{(1.64 - (-0.84))}{\sqrt{20}} = 16.6$$

Example

$\sigma = 30$ $z_{pow} = -0.84$ $z_{crit} = 1.64$

$$n = \frac{\sigma^2 (z_{pow} - z_{crit})^2}{\delta^2}$$

$\delta = 2$

If he wants to detect, at a power of 80%, a true difference of only 2 g, what sample size will he need? 1 g? 0.5 g?

Example

If he wants to detect a true difference of 10 g, at a power of 80%, what should his sample size be? At a power of 90%? 95%? 99%?

Example

If he wants to detect a true difference of 15 g, at a power of 80%, with a sample size of 20. What effect α should he use?