Unit 4: Inference for numerical data
3. Power

Sta 101 – Spring 2019

Duke University, Department of Statistical Science

Dr. Ellison

Slides posted at
https://www2.stat.duke.edu/courses/Spring19/sta101.001/

Outline

1. Housekeeping
2. Main ideas
The role of a statistician is not just in the analysis of data but also in planning and design of a study.
   1. Considerations when selecting **sample size:**
      1. Not every statistically significant result is practically significant
   2. Considerations when selecting **significance level:**
      1. Hypothesis tests have error rates associated with them
      2. Type 1 error rate = significance level
      3. Calculating the power is a two step process
      4. Power goes up with effect size and sample size, and is inversely proportional with significance level and standard error
      5. A priori power calculations determine desired sample size
3. Summary

Announcements

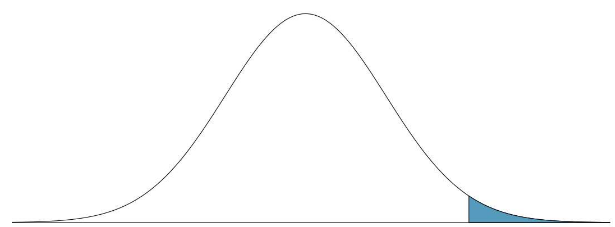
Coming up…

- Project Stage 1 is due **Thursday just before your lab section time.**
- Problem Set 4 is due **Friday 3/8 11:55pm**
- Performance Assessment 4 is due **Sunday 3/17 11:55pm** (opens Wednesday)
- Readiness Assessment 5 is **Monday 3/18**
- New TA for 10:05 and 3:05 sections… email me if you have questions/Thursday STINFs for now.

1

Warm Up

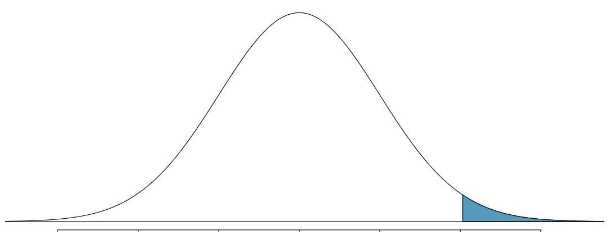
Below is the sampling distribution of sampling statistics for $\mu_1 - \mu_2$. What are the types of sample statistics plotted on the x-axis?



12

Warm Up

Below is the sampling distribution of sampling statistics for $\mu_1 - \mu_2$. What are the types of sample statistics plotted on the x-axis?

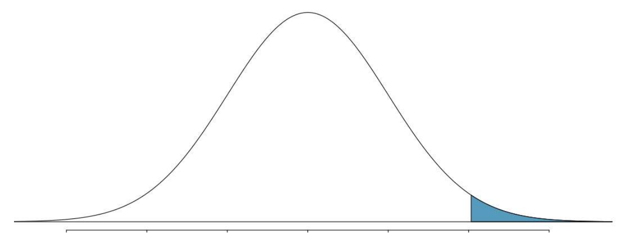


$Values\ of\ \bar{x}_1 - \bar{x}_2$(sample sizes are $n_1 = 10$ $n_2 = 20$)

12

Warm Up

Below is the sampling distribution of sampling statistics for population parameter $\mu_1 - \mu_2$. What is the mean of this distribution?

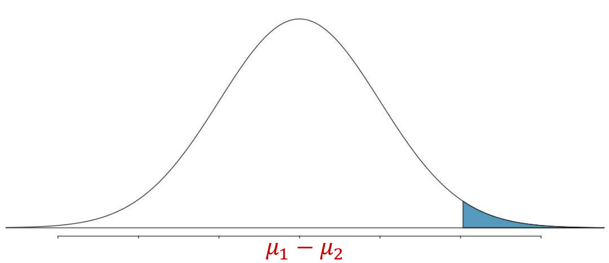


$Values\ of\ \bar{x}_1 - \bar{x}_2$(sample sizes are $n_1 = 10$ $n_2 = 20$)

12

Warm Up

Below is the sampling distribution of sampling statistics for population parameter $\mu_1 - \mu_2$. What is the mean of this distribution?



$Values\ of\ \bar{x}_1 - \bar{x}_2$(sample sizes are $n_1 = 10$ $n_2 = 20$)

12

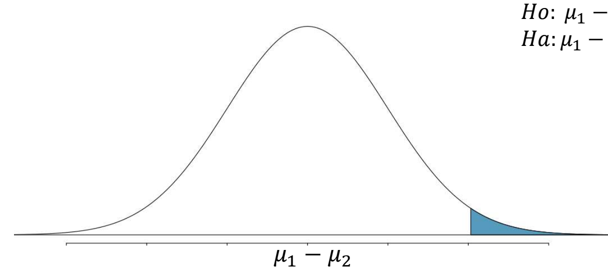
Warm Up

Below is the sampling distribution we would use to find the p-value for the following hypotheses. What is the center/mean?

Hypotheses

Ho: $\mu_1 - \mu_2 = 0$

Ha: $\mu_1 - \mu_2 > 0$



$Values\ of\ \bar{x}_1 - \bar{x}_2$(sample sizes are $n_1 = 10$ $n_2 = 20$)

12

Warm Up

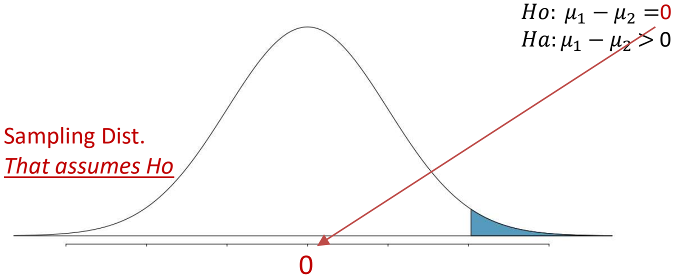
Below is the sampling distribution we would use to find the p-value for the following hypotheses. What is the center/mean?

Hypotheses

$Ho: \mu_1 - \mu_2 = 0$

$Ha: \mu_1 - \mu_2 > 0$

Sampling Dist.
That assumes Ho



$Values\ of\ \bar{x}_1 - \bar{x}_2$(sample sizes are $n_1 = 10\ n_2 = 20$)

12

Warm Up

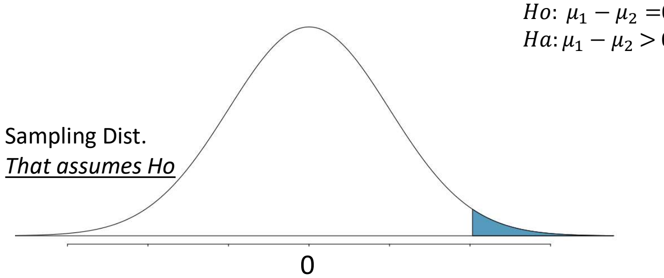
Below is the sampling distribution we would use to find the p-value for the following hypotheses. What is the standard deviation of this distribution?

Hypotheses

$Ho: \mu_1 - \mu_2 = 0$

$Ha: \mu_1 - \mu_2 > 0$

Sampling Dist.
That assumes Ho



$Values\ of\ \bar{x}_1 - \bar{x}_2$(sample sizes are $n_1 = 10\ n_2 = 20$)

12

Warm Up

Below is the sampling distribution we would use to find the p-value for the following hypotheses. What is the standard deviation of this distribution?

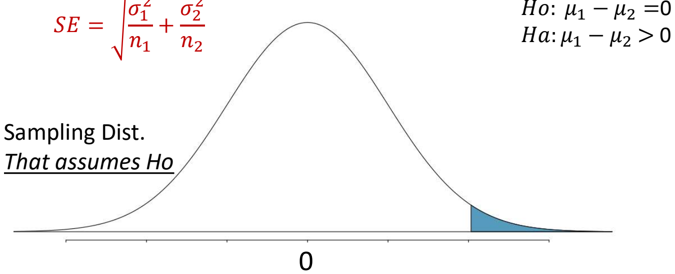
$$SE = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Hypotheses

$Ho: \mu_1 - \mu_2 = 0$

$Ha: \mu_1 - \mu_2 > 0$

Sampling Dist.
That assumes Ho



$Values\ of\ \bar{x}_1 - \bar{x}_2$(sample sizes are $n_1 = 10\ n_2 = 20$)

12

Warm Up

Below is the sampling distribution we would use to find the p-value for the following hypotheses. How many values of $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ do we need to be away from $\mu_1 - \mu_2$ =0 to create the right tail of area = 0.025 below?

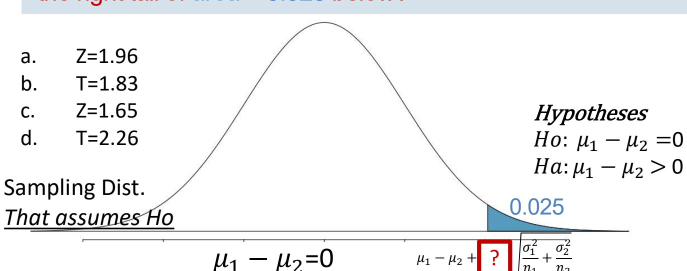
a. Z=1.96
b. T=1.83
c. Z=1.65
d. T=2.26

Hypotheses

$Ho: \mu_1 - \mu_2 = 0$

$Ha: \mu_1 - \mu_2 > 0$

Sampling Dist.
That assumes Ho



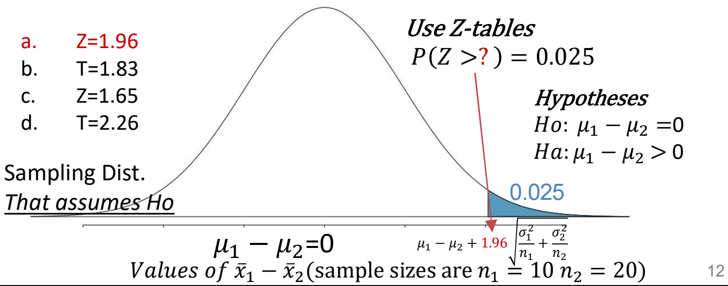
$Values\ of\ \bar{x}_1 - \bar{x}_2$(sample sizes are $n_1 = 10\ n_2 = 20$)

12

Warm Up

Below is the sampling distribution we would use to find the p-value for the following hypotheses. How many values of $\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$ do we need to be away from $\mu_1 - \mu_2$ =0 to create the right tail of area = 0.025 below?

a. Z=1.96
b. T=1.83
c. Z=1.65
d. T=2.26

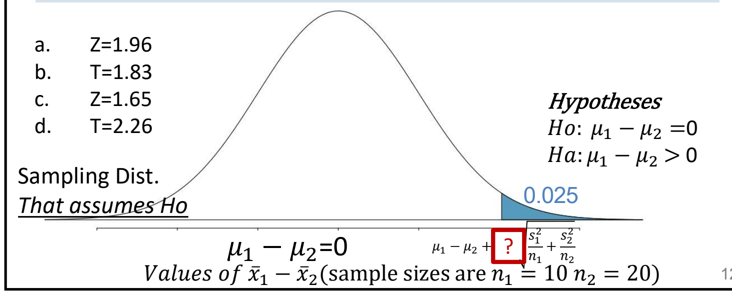


12

Warm Up

Below is the sampling distribution we would use to find the p-value for the following hypotheses. How many values of $\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$ do we need to be away from $\mu_1 - \mu_2$ =0 to create the right tail of area = 0.025 below?

a. Z=1.96
b. T=1.83
c. Z=1.65
d. T=2.26

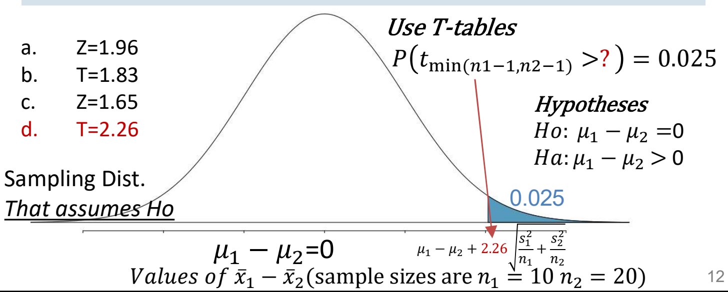


12

Warm Up

Below is the sampling distribution we would use to find the p-value for the following hypotheses. How many values of $\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$ do we need to be away from $\mu_1 - \mu_2$ =0 to create the right tail of area = 0.025 below?

a. Z=1.96
b. T=1.83
c. Z=1.65
d. T=2.26



12

Outline

1. Housekeeping
2. Main ideas

The role of a statistician is not just in the analysis of data but also in planning and design of a study.

1. <u>Considerations when selecting **sample size:**</u>
    1. Not every statistically significant result is practically significant
2. <u>Considerations when selecting **significance level:**</u>
    1. Hypothesis tests have error rates associated with them
    2. Type 1 error rate = significance level
    3. Calculating the power is a two step process
    4. Power goes up with effect size and sample size, and is inversely proportional with significance level and standard error
    5. A priori power calculations determine desired sample size

3. Summary

Outline

Would we expect the average height of all NC men to be *exactly* the same as the average height of all US men?

Would we expect the average height of all NC men to be *practically* the same as the height of all US men?

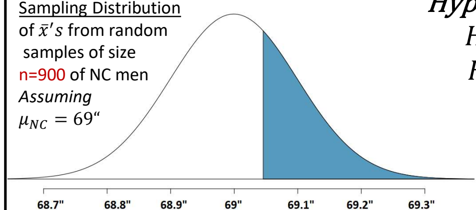
Outline

Would we expect the average height of all NC men to be *exactly* the same as the average height of all US men (assume 69")?

Would we expect the average height of all NC men to be *practically* the same as the height of all US men (assume 69") ?

Outline

Sampling Distribution
of $\bar{x}'s$ from random samples of size n=900 of NC men
Assuming $\mu_{NC} = 69$"



Hypotheses

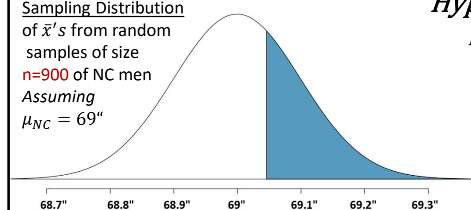
$Ho\colon \mu_{NC} = 69$"

$Ha\colon \mu_{NC} > 69$"

$\bar{x} = 69.05$", would not give us enough evidence to suggest $\mu_{NC} > 69$" if $\bar{x}$ came from a sample of size n=900.

Outline

Sampling Distribution
of $\bar{x}'s$ from random samples of size n=900 of NC men
Assuming $\mu_{NC} = 69$"



Hypotheses

$Ho\colon \mu_{NC} = 69$"

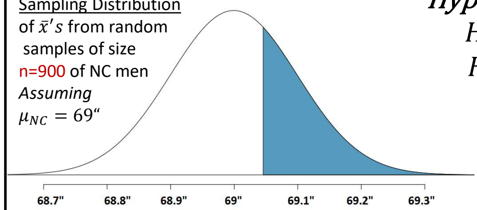
$Ha\colon \mu_{NC} > 69$"

$\bar{x} = 69.05$", would not give us enough evidence to suggest $\mu_{NC} > 69$" if $\bar{x}$ came from a sample of size n=900.

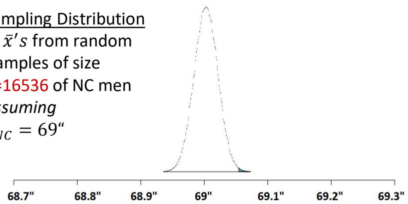
Sampling Distribution
of $\bar{x}'s$ from random samples of size n=16536 of NC men
Assuming $\mu_{NC} = 69$"

Outline

Sampling Distribution
of $\bar{x}'s$ from random
samples of size
n=900 of NC men
Assuming
$\mu_{NC} = 69$"



Sampling Distribution
of $\bar{x}'s$ from random
samples of size
n=16536 of NC men
Assuming
$\mu_{NC} = 69$"



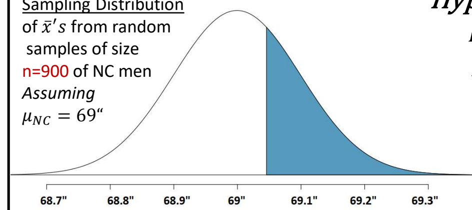
Hypotheses

$Ho: \mu_{NC} = 69$"
$Ha: \mu_{NC} > 69$"

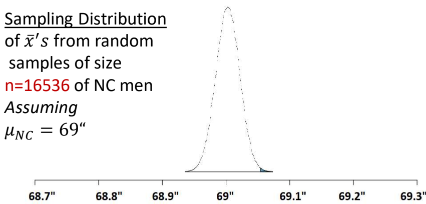
$\bar{x} = 69.05$", would not give us enough evidence to suggest $\mu_{NC} > 69$" if $\bar{x}$ came from a sample of size n=900.

Outline

Sampling Distribution
of $\bar{x}'s$ from random
samples of size
n=900 of NC men
Assuming
$\mu_{NC} = 69$"



Sampling Distribution
of $\bar{x}'s$ from random
samples of size
n=16536 of NC men
Assuming
$\mu_{NC} = 69$"



Hypotheses

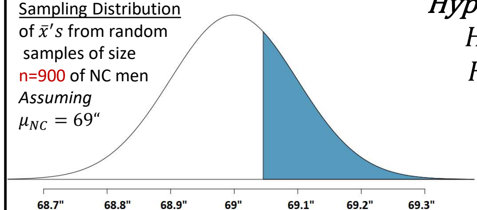
$Ho: \mu_{NC} = 69$"
$Ha: \mu_{NC} > 69$"

$\bar{x} = 69.05$", would not give us enough evidence to suggest $\mu_{NC} > 69$" if $\bar{x}$ came from a sample of size n=900.

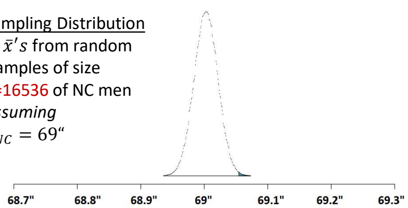
$\bar{x} = 69.05$", would give us enough evidence to suggest $\mu_{NC} > 69$" if $\bar{x}$ came from a sample of size n=16536.

Outline

Sampling Distribution
of $\bar{x}'s$ from random
samples of size
n=900 of NC men
Assuming
$\mu_{NC} = 69$"



Sampling Distribution
of $\bar{x}'s$ from random
samples of size
n=16536 of NC men
Assuming
$\mu_{NC} = 69$"



Hypotheses

$Ho: \mu_{NC} = 69$"
$Ha: \mu_{NC} > 69$"

$\bar{x} = 69.05$", would not give us enough evidence to suggest $\mu_{NC} > 69$" if $\bar{x}$ came from a sample of size n=900.

$\bar{x} = 69.05$", would give us enough evidence to suggest $\mu_{NC} > 69$" if $\bar{x}$ came from a sample of size n=16536.

BUT, is it practical to say men in NC are taller on average?

Reminder: Not every statistically significant result is practically significant

- Real differences between the point estimate and null value are easier to detect with larger samples
- However, very large samples will result in statistical significance even for tiny differences between the sample mean and the null value (***effect size***), even when the difference is not practically significant
- This is especially important to research: if we conduct a study, we want to focus on finding meaningful results (we want observed differences to be real but also large enough to matter).
- The role of a statistician is not just in the analysis of data but also in planning and design of a study.

"To call in the statistician after the experiment is done may be no more than asking him to perform a post-mortem examination: he may be able to say what the experiment died of." – R.A. Fisher

2

Reminder: Not every statistically significant result is practically significant

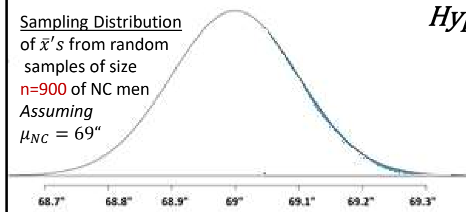
What is a **meaningful result**?

- Subject matter experts will usually give an effect size δ that they find meaningful.
- **effect size δ** = |actual pop. parameter value – null value|

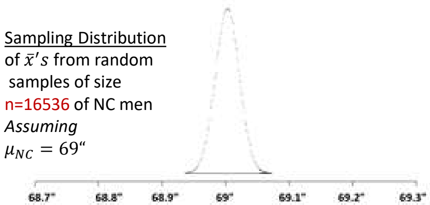
2

Outline

Sampling Distribution of $\bar{x}'s$ from random samples of size n=900 of NC men *Assuming* $\mu_{NC} = 69$"



Sampling Distribution of $\bar{x}'s$ from random samples of size n=16536 of NC men *Assuming* $\mu_{NC} = 69$"



Hypotheses

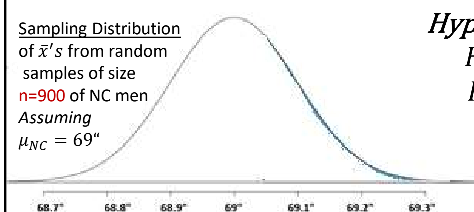
Ho: $\mu_{NC} = 69$"

Ha: $\mu_{NC} > 69$"

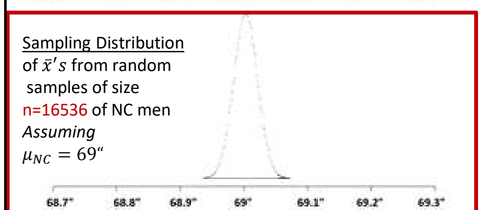
Which sample size would be better for detecting an effect size of δ = 0.2"?

Outline

Sampling Distribution of $\bar{x}'s$ from random samples of size n=900 of NC men *Assuming* $\mu_{NC} = 69$"



Sampling Distribution of $\bar{x}'s$ from random samples of size n=16536 of NC men *Assuming* $\mu_{NC} = 69$"



Hypotheses

Ho: $\mu_{NC} = 69$"

Ha: $\mu_{NC} > 69$"

Which sample size would be better for detecting an effect size of δ = 0.2"?

Outline

For a smaller effect size, a hypothesis test's ability to detect actual differences between the **null value** and **(null value +/- effect size)** *decreases*.

We can increase the sample size to try to detect this small difference.

Outline

Outline

Recap of:
- Type 1 Error
- Type 2 Error
- Power

Reminder: Hypothesis tests have error rates associated with them

There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make a decision about which might be true, but our choice might be incorrect.

		Decision: fail to reject H_0	Decision: reject H_0
Truth	H_0 true	✓	*Type 1 Error*
	H_A true	*Type 2 Error*	✓

- A *Type 1 Error* is rejecting the null hypothesis when H_0 is true.
- A *Type 2 Error* is failing to reject the null hypothesis when H_A is true.
- We (almost) never know if H_0 or H_A is true, but we need to consider all possibilities.

3

4. Hypothesis tests are prone to decision errors

		Decision: fail to reject H_0	Decision: reject H_0
Truth	H_0 true	✓	*Type 1 Error*, α
	H_A true	*Type 2 Error*, β	*Power*, $1-\beta$

- A *Type 1 Error* is rejecting the null hypothesis when H_0 is true.
  - P(Type 1 Error) = α = P(reject Ho|Ho is true)
- A *Type 2 Error* is failing to reject the null hypothesis when H_A is true.
  - P(Type 2 Error) = β = P(fail to reject Ho|Ha is true)
- *Power* is the probability of *correctly* rejecting H_0, and hence the complement of the probability of a Type 2 Error
  - Power = 1 – β = P(reject Ho|Ha is true)

Outline

1. Housekeeping
2. Main ideas
The role of a statistician is not just in the analysis of data but also in planning and design of a study.
   1. Considerations when selecting **sample size:**
      1. Not every statistically significant result is practically significant
   2. Considerations when selecting **significance level:**
      1. Hypothesis tests have error rates associated with them
      2. Type 1 error rate = significance level
      3. Calculating the power is a two step process
      4. Power goes up with effect size and sample size, and is inversely proportional with significance level and standard error
      5. A priori power calculations determine desired sample size

3. Summary

Outline

Why do we denote the significance level and the P(Type 1 Error) as α?

Outline

Why do we denote the significance level and the P(Type 1 Error) as α?

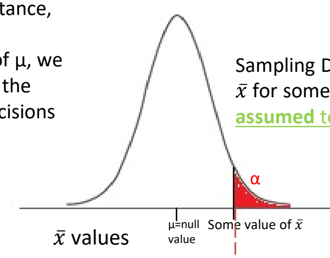
-They're the same!
P(Type 1 Error)=Significance Level = α

Reminder: Type 1 error rate = significance level

Example: For instance, for a right-tailed hypothesis test of μ, we use the graph to the right to make decisions about our null hypothesis.

Hypotheses
$Ho: \mu = null\ value$
$Ha: \mu > null\ value$

Sampling Distribution of $\bar{x}$ for some n where Ho is **assumed** to be true.



Decision: If our sample statistic $\bar{x}$ falls in this range, we __________.

Decision: If our sample statistic $\bar{x}$ falls in this range, we __________

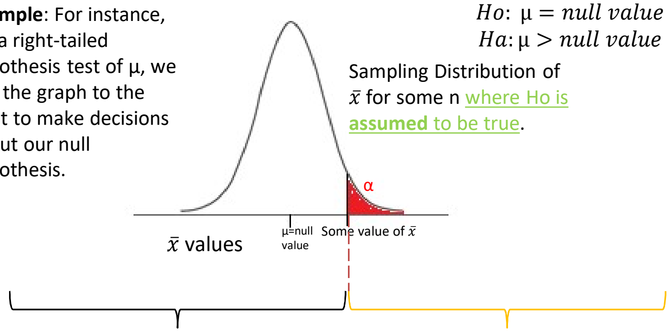
4

Reminder: Type 1 error rate = significance level

Hypotheses
Ho: $\mu = null\ value$
Ha: $\mu > null\ value$

Example: For instance, for a right-tailed hypothesis test of μ, we use the graph to the right to make decisions about our null hypothesis.

Sampling Distribution of $\bar{x}$ for some n where Ho is **assumed** to be true.



Decision: If our sample statistic $\bar{x}$ falls in this range, we fail to reject Ho.

Decision: If our sample statistic $\bar{x}$ falls in this range, we reject Ho.

4

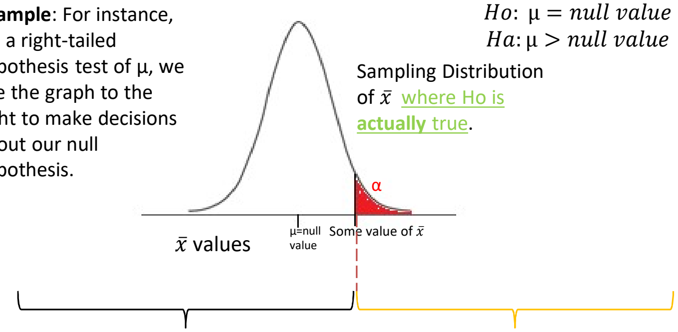
Reminder: Type 1 error rate = significance level

P(Type 1 error) = P(Reject H_0| H_0 is true) = α

Hypotheses
Ho: $\mu = null\ value$
Ha: $\mu > null\ value$

Example: For instance, for a right-tailed hypothesis test of μ, we use the graph to the right to make decisions about our null hypothesis.

Sampling Distribution of $\bar{x}$ where Ho is **actually** true.



Decision: If our sample statistic $\bar{x}$ falls in this range, we fail to reject Ho.

Decision: If our sample statistic $\bar{x}$ falls in this range, we reject Ho.

4

Reminder: Type 1 error rate = significance level

- As a general rule we reject H_0 when the p-value is less than 0.05, i.e. we use a *significance level* of 0.05, α = 0.05.

4

Reminder: Type 1 error rate = significance level

- As a general rule we reject H_0 when the p-value is less than 0.05, i.e. we use a *significance level* of 0.05, α = 0.05.
- This means that, for those cases where H_0 is actually true, we will incorrectly reject it at most 5% of the time.

4

Reminder: Type 1 error rate = significance level

- As a general rule we reject H_0 when the p-value is less than 0.05, i.e. we use a *significance level* of 0.05, $\alpha = 0.05$.
- This means that, for those cases where H_0 is actually true, we will incorrectly reject it at most 5% of the time.
- In other words, when using a 5% significance level there is about 5% chance of making a Type 1 error.

$$P(\text{Type 1 error}) = P(\text{Reject } H_0 | H_0 \text{ is true}) = \alpha$$

4

Reminder: Type 1 error rate = significance level

- As a general rule we reject H_0 when the p-value is less than 0.05, i.e. we use a *significance level* of 0.05, $\alpha = 0.05$.
- This means that, for those cases where H_0 is actually true, we will incorrectly reject it at most 5% of the time.
- In other words, when using a 5% significance level there is about 5% chance of making a Type 1 error.

$$P(\text{Type 1 error}) = P(\text{Reject } H_0 | H_0 \text{ is true}) = \alpha$$

- This is why we prefer small values of α – increasing α increases the Type 1 error rate.

4

Outline

Which of the following would have a lower Type 2 Error Rate? (Ie: which has a lower β?) **(Assume standard dev. is the same for both populations).**

Hypotheses
$Ho: \mu_{NC} = 69“$
$Ha: \mu_{NC} \neq 69“$

Hypotheses
$Ho: \mu_{NBA} = 69“$
$Ha: \mu_{NBA} \neq 69“$

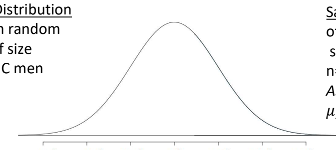
Outline

Which of the following would have a lower Type 2 Error Rate? (Ie: which has a lower β?) **(Assume standard dev. is the same for both populations).**

Hypotheses
$Ho: \mu_{NC} = 69“$
$Ha: \mu_{NC} \neq 69“$

Hypotheses
$Ho: \mu_{NBA} = 69“$
$Ha: \mu_{NBA} \neq 69“$

Sampling Distribution of $\bar{x}'s$ from random samples of size n=900 of NC men *Assuming* $\mu_{NC} = 69“$



Sampling Distribution of $\bar{x}'s$ from random samples of size n=900 of NBA men *Assuming* $\mu_{NBA} = 69“$

Type 2 error rate

If the alternative hypothesis is actually true, what is the chance that we make a Type 2 Error, i.e. we fail to reject the null hypothesis even when we should reject it?

Type 2 error rate

If the alternative hypothesis is actually true, what is the chance that we make a Type 2 Error, i.e. we fail to reject the null hypothesis even when we should reject it?

- The answer is not obvious, but
  - If the true population average is very close to the null hypothesis value, it will be difficult to detect a difference (and reject H_0).
  - If the true population average is very different from the null hypothesis value, it will be easier to detect a difference.

Type 2 error rate

If the alternative hypothesis is actually true, what is the chance that we make a Type 2 Error, i.e. we fail to reject the null hypothesis even when we should reject it?

- The answer is not obvious, but
  - If the true population average is very close to the null hypothesis value, it will be difficult to detect a difference (and reject H_0).
  - If the true population average is very different from the null hypothesis value, it will be easier to detect a difference.
- Therefore, β must depend on the *effect size* (δ) in some way

To ↑ power/↓ β you can:
- ↑ n
- ↑ δ
- ↑ α

Outline

1. Housekeeping
2. Main ideas

The role of a statistician is not just in the analysis of data but also in planning and design of a study.

1. Considerations when selecting **sample size:**
   1. Not every statistically significant result is practically significant
2. Considerations when selecting **significance level:**
   1. Hypothesis tests have error rates associated with them
   2. Type 1 error rate = significance level
   3. Calculating the power is a two step process
   4. Power goes up with effect size and sample size, and is inversely proportional with significance level and standard error
   5. A priori power calculations determine desired sample size

3. Summary

Let's calculate the **power** of a test for a given:

- Sample size
- Effect size.

A medical research group is recruiting people to complete short surveys about their medical history. **So far**, people complete an **average of 4 surveys**, with the **standard deviation of 2.2** surveys.

The research group wants to try a new interface that they think will encourage new enrollees to complete more surveys, where they will randomize a total of 300 enrollees to either get the new interface or the current interface (equally distributed between the two groups).

What is the **power** of the test that can detect an increase of 0.5 surveys per enrollee for the new interface compared to the old interface? Assume that the new interface does not affect the standard deviation of completed surveys, and $\alpha = 0.05$.

Hypothesis Tests we would use: $\alpha = 0.05$

$H_0: \mu_{new} - \mu_{current} = 0$

$H_A: \mu_{new} - \mu_{current} > 0$

Info:

$n_{new} = n_{current} = 150$

$s_{new} = s_{current} = 2.2$

What is the **power** of the test that can detect an increase of 0.5 surveys per enrollee for the new interface compared to the old interface?

Calculating power

What is the **power** of the test that can detect an increase of 0.5 surveys per enrollee for the new interface compared to the old interface?

Effect size = |0.5-0|

8

Calculating power

What is the **power** of the test that can detect an increase of 0.5 surveys per enrollee for the new interface compared to the old interface
(ie: What is the **power** that detects Ha: $\mu_{new} - \mu_{current} = 0 + 0.5$)?

8

Calculating power

What is the **power** of the test that can detect an increase of 0.5 surveys per enrollee for the new interface compared to the old interface
(ie: What is the **power** that detects Ha: $\mu_{new} - \mu_{current} = 0.5$)?

Let's break this down into two simpler problems:

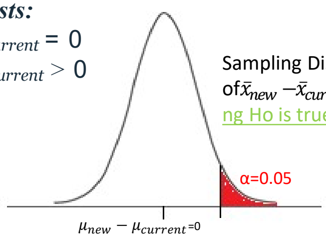
1. **Problem 1**: Which values of $\bar{x}_{new} - \bar{x}_{current}$ represent sufficient evidence to reject this H_0?

8

Reminder: Type 1 error rate = significance level

P(Type 1 error) = P(Reject H_0| H_0 is true) = α = 0.05

Hypothesis Tests:
$H_0: \mu_{new} - \mu_{current} = 0$
$H_A: \mu_{new} - \mu_{current} > 0$



Problem 1: Which values of ($\bar{x}_{new} - \bar{x}_{current}$) represent sufficient evidence to reject H_0?

4

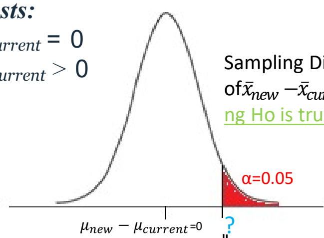
Reminder: Type 1 error rate = significance level

P(Type 1 error) = P(Reject H_0| H_0 is true) = α $= 0.05$

Hypothesis Tests:

$H_0: \mu_{new} - \mu_{current} = 0$

$H_A: \mu_{new} - \mu_{current} > 0$



Sampling Distribution of $\bar{x}_{new} - \bar{x}_{current}$ assuming Ho is true.

Decision: If our sample statistic $\bar{x}_{new} - \bar{x}_{current}$ falls in this range, we __________

Decision: If our sample statistic $\bar{x}_{new} - \bar{x}_{current}$ falls in this range, we ____________

Problem 1: Which values of $(\bar{x}_{new} - \bar{x}_{current})$ represent sufficient evidence to reject H_0?

4

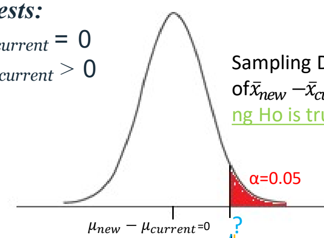
Reminder: Type 1 error rate = significance level

P(Type 1 error) = P(Reject H_0| H_0 is true) = α $= 0.05$

Hypothesis Tests:

$H_0: \mu_{new} - \mu_{current} = 0$

$H_A: \mu_{new} - \mu_{current} > 0$



Sampling Distribution of $\bar{x}_{new} - \bar{x}_{current}$ assuming Ho is true.

Decision: If our sample statistic $\bar{x}_{new} - \bar{x}_{current}$ falls in this range, we fail to reject Ho.

Decision: If our sample statistic $\bar{x}_{new} - \bar{x}_{current}$ falls in this range, we reject Ho.

Problem 1: Which values of $(\bar{x}_{new} - \bar{x}_{current})$ represent sufficient evidence to reject H_0?

Aka: What is this sample statistic "cutoff" value

4

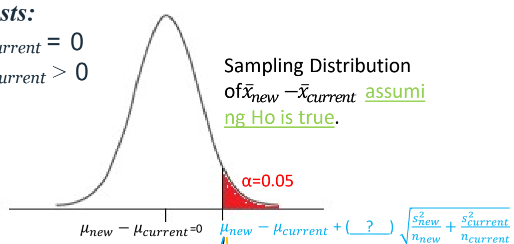
Reminder: Type 1 error rate = significance level

P(Type 1 error) = P(Reject H_0| H_0 is true) = α $= 0.05$

Hypothesis Tests:

$H_0: \mu_{new} - \mu_{current} = 0$

$H_A: \mu_{new} - \mu_{current} > 0$



Sampling Distribution of $\bar{x}_{new} - \bar{x}_{current}$ assuming Ho is true.

Decision: If our sample statistic $\bar{x}_{new} - \bar{x}_{current}$ falls in this range, we fail to reject Ho.

Decision: If our sample statistic $\bar{x}_{new} - \bar{x}_{current}$ falls in this range, we reject Ho.

Aka: How many values of $\sqrt{\frac{s^2_{new}}{n_{new}} + \frac{s^2_{current}}{n_{current}}}$ do we need to be away from $\mu_{new} - \mu_{current} = 0$ to create the right tail of area = 0.05 above?

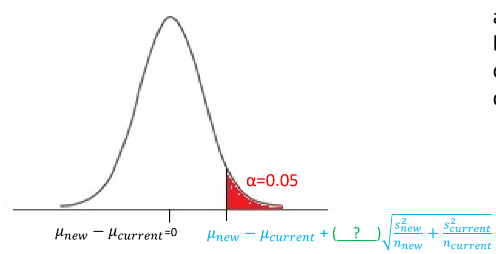
4

Reminder: Type 1 error rate = significance level

Clicker question

Aka: How many values of $\sqrt{\frac{s^2_{new}}{n_{new}} + \frac{s^2_{current}}{n_{current}}}$ do we need to be away from $\mu_{new} - \mu_{current} = 0$ to create the right tail of area = 0.05 below? (Remember $n_{new} = n_{current} = 150$).

a. Z=1.96
b. Z=1.65
c. T=1.66
d. T=1.98



4

Reminder: Type 1 error rate = significance level

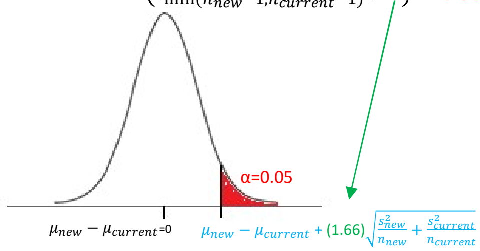
Clicker question

Aka: How many values of $\sqrt{\frac{s_{new}^2}{n_{new}} + \frac{s_{current}^2}{n_{current}}}$ do we need to be away from $\mu_{new} - \mu_{current} = 0$ to create the right tail of area = 0.05 below? (Remember $n_{new} = n_{current} = 150$).

Use T-tables

$P\left(t_{\min(n_{new}-1, n_{current}-1)} > ?\right) = 0.05$

a. Z=1.96
b. Z=1.65
c. T=1.66
d. T=1.98



4

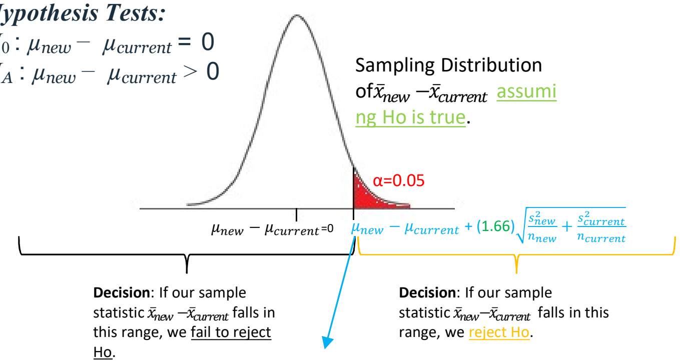
Reminder: Type 1 error rate = significance level

P(Type 1 error) = P(Reject H_0 | H_0 is true) = α = 0.05

Hypothesis Tests:

$H_0: \mu_{new} - \mu_{current} = 0$

$H_A: \mu_{new} - \mu_{current} > 0$



Problem 1: Which values of ($\bar{x}_{new} - \bar{x}_{current}$) represent sufficient evidence to reject H_0?

4

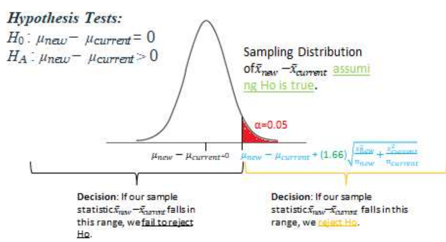
Problem 1 - cont.

Clicker question

Problem 1: Which values of ($\bar{x}_{new} - \bar{x}_{current}$) represent sufficient evidence to reject H_0?

$n_{new} = n_{current} = 150$, $\alpha = 0.05$, $s_{new} = 2.2 = s_{current} = 2.2$

(a) $\bar{x}_{new} - \bar{x}_{current} < -0.42$
(b) $\bar{x}_{new} - \bar{x}_{current} > -0.42$
(c) $\bar{x}_{new} - \bar{x}_{current} < 0.42$
(d) $\bar{x}_{new} - \bar{x}_{current} > 0.42$
(e) $\bar{x}_{new} - \bar{x}_{current} > 1.66$



11

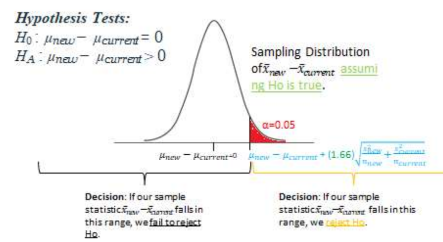
Problem 1 - cont.

Clicker question

Problem 1: Which values of ($\bar{x}_{new} - \bar{x}_{current}$) represent sufficient evidence to reject H_0?

$n_{new} = n_{current} = 150$, $\alpha = 0.05$, $s_{new} = 2.2 = s_{current} = 2.2$

(a) $\bar{x}_{new} - \bar{x}_{current} < -0.42$
(b) $\bar{x}_{new} - \bar{x}_{current} > -0.42$
(c) $\bar{x}_{new} - \bar{x}_{current} < 0.42$
(d) $\bar{x}_{new} - \bar{x}_{current} > 0.42$
(e) $\bar{x}_{new} - \bar{x}_{current} > 1.66$



$$\bar{x}_{new} - \bar{x}_{current} > \mu_{new} - \mu_{current} + 1.66\sqrt{\frac{s_{new}^2}{n_{new}} + \frac{s_{current}^2}{n_{current}}} = 0 + 1.66\sqrt{\frac{2.2^2}{150} + \frac{2.2^2}{150}} = 0.42$$

11

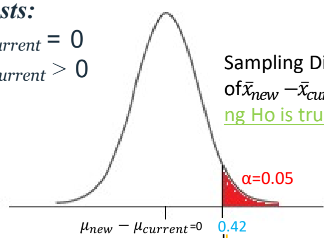
Reminder: Type 1 error rate = significance level

P(Type 1 error) = P(Reject H_0| H_0 is true) = α = 0.05

Hypothesis Tests:

$H_0: \mu_{new} - \mu_{current} = 0$

$H_A: \mu_{new} - \mu_{current} > 0$



Sampling Distribution of $\bar{x}_{new} - \bar{x}_{current}$ assuming Ho is true.

Decision: If our sample statistic $\bar{x}_{new} - \bar{x}_{current}$ falls in this range, we fail to reject Ho.

Decision: If our sample statistic $\bar{x}_{new} - \bar{x}_{current}$ falls in this range, we reject Ho.

4

Calculating power

What is the **power** of the test that can detect an increase of 0.5 surveys per enrollee for the new interface compared to the old interface
(ie: What is the **power** that detects $\mu_{new} - \mu_{current}$ = 0.5)?

Let's break this down intro two simpler problems:

Problem 1: Which values of $\bar{x}_{new} - \bar{x}_{current}$ represent sufficient evidence to reject this H_0?

Answer 1: $\bar{x}_{new} - \bar{x}_{current} > 0.42$

Problem 2: What is the **power** that detects $\mu_{new} - \mu_{current}$ = 0.5?

8

Reminder: Type 1 error rate = significance level

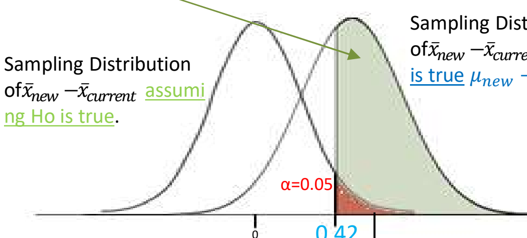
$Power = 1\text{-}\beta$ = P(Reject H_0| H_a is true)

4

Reminder: Type 1 error rate = significance level

$Power = 1\text{-}\beta$ = P(Reject H_0| H_a is true)

Sampling Distribution of $\bar{x}_{new} - \bar{x}_{current}$ assuming Ho is true.



Sampling Distribution of $\bar{x}_{new} - \bar{x}_{current}$ assuming Ha is true $\mu_{new} - \mu_{current}$=0.50

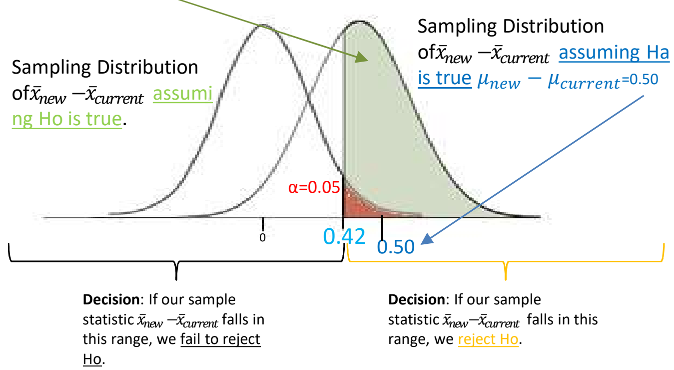
Decision: If our sample statistic $\bar{x}_{new} - \bar{x}_{current}$ falls in this range, we fail to reject Ho.

Decision: If our sample statistic $\bar{x}_{new} - \bar{x}_{current}$ falls in this range, we reject Ho.

4

Reminder: Type 1 error rate = significance level

$Power = 1\text{-}\beta = P(\text{ Reject } H_0 | \text{Ha: } \mu_{new} - \mu_{current} = 0.50)$



4

Problem 2

Clicker question

What is the **power** if $\mu_{new} - \mu_{current} = 0.5$?

$n_{new} = n_{current} = 150, \alpha = 0.05, s_{new} = 2.2 = s_{current} = 2.2$

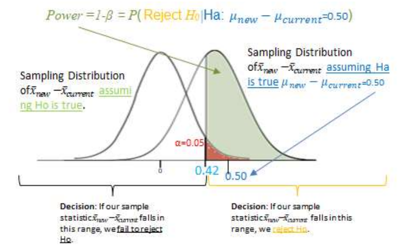
Hint: use pt( ) function in R

(a) 5%

(b) 38%

(c) 62%

(d) 80%



12

Problem 2

Clicker question

What is the **power** if $\mu_{new} - \mu_{current} = 0.5$?

$n_{new} = n_{current} = 150, \alpha = 0.05, s_{new} = 2.2 = s_{current} = 2.2$

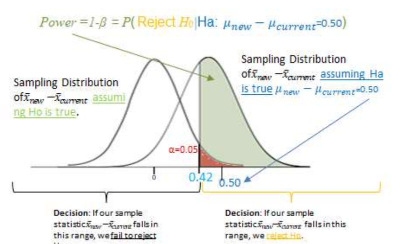
Hint: use pt( ) function in R

(a) 5%

(b) 38%

(c) 62%

(d) 80%

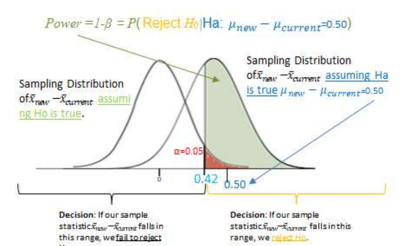


12

Reminder: Type 1 error rate = significance level

$Power = 1\text{-}\beta = P(\text{ Reject } H_0 | H_a \text{ is true})$

$= P(\bar{x}_{new} - \bar{x}_{current} > 0.42 | \mu_{new} - \mu_{current} = 0.50)$

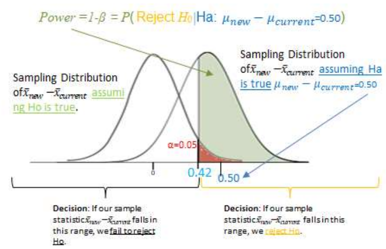


Reminder: Type 1 error rate = significance level

Power $=1\text{-}\beta = P(\text{ Reject } H_0 | \; H_a \text{ is true})$

$=P(\bar{x}_{new} - \bar{x}_{current} > 0.42 | \mu_{new} - \mu_{current} = 0.50)$

$$=P\left(t_{\min(n_1-1,n_2-1)} > \frac{0.42-(\mu_{new}-\mu_{current})}{\sqrt{\frac{s^2_{new}}{n_{new}}+\frac{s^2_{current}}{n_{current}}}} \middle| \mu_{new} - \mu_{current} = 0.50\right)$$



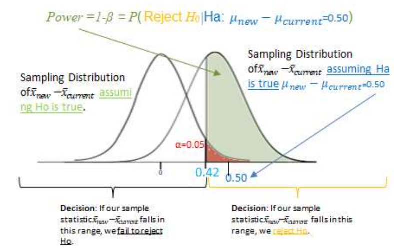
Reminder: Type 1 error rate = significance level

Power $=1\text{-}\beta = P(\text{ Reject } H_0 | \; H_a \text{ is true})$

$=P(\bar{x}_{new} - \bar{x}_{current} > 0.42 | \mu_{new} - \mu_{current} = 0.50)$

$$=P\left(t_{\min(n_1-1,n_2-1)} > \frac{0.42-(\mu_{new}-\mu_{current})}{\sqrt{\frac{s^2_{new}}{n_{new}}+\frac{s^2_{current}}{n_{current}}}} \middle| \mu_{new} - \mu_{current} = 0.50\right)$$

$$=P\left(t_{\min(n_1-1,n_2-1)} > \frac{042-0.50}{0.25}\right) = P\left(t_{\min(150-1,150-1)} > -0.315\right)$$



Reminder: Type 1 error rate = significance level

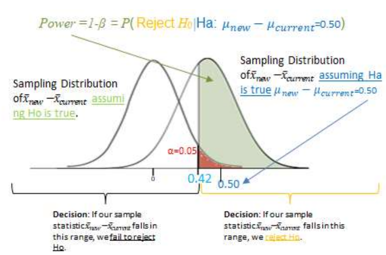
Power $=1\text{-}\beta = P(\text{ Reject } H_0 | \; H_a \text{ is true})$

$=P(\bar{x}_{new} - \bar{x}_{current} > 0.42 | \mu_{new} - \mu_{current} = 0.50)$

$$=P\left(t_{\min(n_1-1,n_2-1)} > \frac{0.42-(\mu_{new}-\mu_{current})}{\sqrt{\frac{s^2_{new}}{n_{new}}+\frac{s^2_{current}}{n_{current}}}} \middle| \mu_{new} - \mu_{current} = 0.50\right)$$

pt(-.315,df=149,lower.tail=FALSE)

$$=P\left(t_{\min(n_1-1,n_2-1)} > \frac{042-0.50}{0.25}\right) = P\left(t_{\min(150-1,150-1)} > -0.315\right) = 0.62$$



Calculating power

What is the **power** of the test that can detect an <u>increase of 0.5</u> surveys per enrollee for the new interface compared to the old interface

(ie: What is the **power** that detects $\mu_{new} - \mu_{current} = 0.5$)?

Let's break this down intro two simpler problems:

Problem 1: Which values of $\bar{x}_{new} - \bar{x}_{current}$ represent sufficient evidence to reject this H_0?

Answer 1: $\bar{x}_{new} - \bar{x}_{current} > 0.42$

Problem 2: What is the **power** that detects $\mu_{new} - \mu_{current} = 0.5$?

Answer 2: $power = 0.62$

8

Problem 2 - cont.

Clicker question

What is β, the Type 2 error rate?

(a) 5%
(b) 38%
(c) 62%
(d) 80%
(e) 95%

13

Problem 2 - cont.

Clicker question

What is β, the Type 2 error rate?

(a) 5%
(b) *38%*
(c) 62%
(d) 80%
(e) 95%

β=1-power=1-0.62

13

Outline

1. Housekeeping
2. Main ideas

The role of a statistician is not just in the analysis of data but also in planning and design of a study.

1. Considerations when selecting **sample size:**
   1. Not every statistically significant result is practically significant
2. Considerations when selecting **significance level:**
   1. Hypothesis tests have error rates associated with them
   2. Type 1 error rate = significance level
   3. Calculating the power is a two step process
   4. Power goes up with effect size and sample size, and is inversely proportional with significance level and standard error
   5. A priori power calculations determine desired sample size

3. Summary

Outline

What is the relationship between power and:

- **Sample size?**
- **Standard deviation?**
- **Significance Level?**
- **Effect Size?**

https://rpsychologist.com/d3/NHST/

Achieving desired power

NEW

There are several ways to increase power (and hence decrease Type 2 error rate):

Achieving desired power

NEW

There are several ways to increase power (and hence decrease Type 2 error rate):

1. Increase the sample size.

Achieving desired power

NEW

There are several ways to increase power (and hence decrease Type 2 error rate):

1. Increase the sample size.
2. Decrease the standard deviation of the sample, which is equivalent to increasing the sample size (it will decrease the standard error). With a smaller s we have a better chance of distinguishing the null value from the observed point estimate. This is difficult to ensure but cautious measurement process and limiting the population so that it is more homogenous may help.

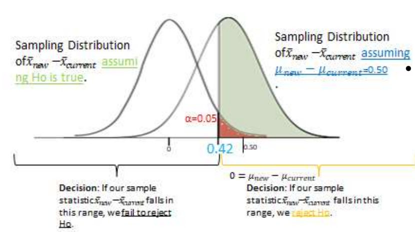
Achieving desired power

NEW

There are several ways to increase power (and hence decrease Type 2 error rate):

1. Increase the sample size(s).
2. Decrease the standard deviation(s) of the sample.

$$SE=\sqrt{\frac{s_{ne}^2}{n_{new}}+\frac{s_{current}^2}{n_{current}}}$$



Can we see why in this example?

- Both changes would decrease the spread (SE) of both of these sampling distributions.
- The "cut-off value" 0.42 would decrease to something lower, and a higher percentage of values in the distribution on the right would be higher than this new "cut-off value".

NEW

Achieving desired power

There are several ways to increase power (and hence decrease Type 2 error rate):

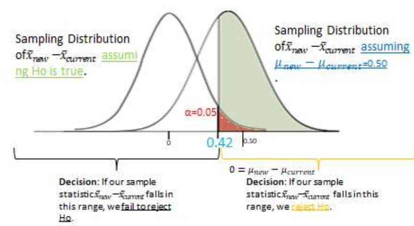
1. Increase the sample size.
2. Decrease the standard deviation of the sample, which is equivalent to increasing the sample size (it will decrease the standard error). With a smaller s we have a better chance of distinguishing the null value from the observed point estimate. This is difficult to ensure but cautious measurement process and limiting the population so that it is more homogenous may help.
3. Increase α, which will make it more likely to reject H_0 (but note that this has the side effect of increasing the Type 1 error rate).

NEW

Achieving desired power

There are several ways to increase power (and hence decrease Type 2 error rate):

3. Increase α, which will make it more likely to reject H_0 (but note that this has the side effect of increasing the Type 1 error rate).



Can we see why in this example?

- The "cut-off value" 0.42 would decrease to something lower, and thus a higher percentage of values in the distribution on the right would be higher than this new "cut-off value".

NEW

Achieving desired power

There are several ways to increase power (and hence decrease Type 2 error rate):

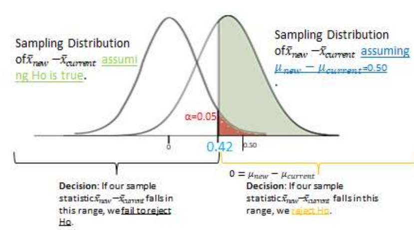
1. Increase the sample size.
2. Decrease the standard deviation of the sample, which is equivalent to increasing the sample size (it will decrease the standard error). With a smaller s we have a better chance of distinguishing the null value from the observed point estimate. This is difficult to ensure but cautious measurement process and limiting the population so that it is more homogenous may help.
3. Increase α, which will make it more likely to reject H_0 (but note that this has the side effect of increasing the Type 1 error rate).
4. Consider a larger effect size. If the true mean of the population is in the alternative hypothesis but close to the null value, it will be harder to detect a difference.

NEW

Achieving desired power

There are several ways to increase power (and hence decrease Type 2 error rate):

4. Consider a larger effect size. If the true mean of the population is in the alternative hypothesis but close to the null value, it will be harder to detect a difference.



Can we see why in this example?

- The distribution on the right would shift further to the right, and thus the percentage of values in this distribution that are higher than 0.42 would increase.

Recap - Calculating Power

- *Step 0:* Pick a meaningful effect size δ and a significance level α
- *Step 1:* Calculate the range of values for the point estimate beyond which you would reject H_0 at the chosen α level.
- *Step 2:* Calculate the probability of observing a value from preceding step if the sample was derived from a population where $\mu = \text{null value} + \delta$

15

Outline

1. Housekeeping
2. Main ideas

The role of a statistician is not just in the analysis of data but also in planning and design of a study.

1. Considerations when selecting **sample size:**
   1. Not every statistically significant result is practically significant
2. Considerations when selecting **significance level:**
   1. Hypothesis tests have error rates associated with them
   2. Type 1 error rate = significance level
   3. Calculating the power is a two step process
   4. Power goes up with effect size and sample size, and is inversely proportional with significance level and standard error
   5. A priori power calculations determine desired sample size

3. Summary

Back to medical surveys...

How large a sample size would you need if you wanted 90% power to detect a 0.5 increase in average number of surveys taken at the 5% significance level?

$H_0 : \mu_{new} - \mu_{current} = 0, H_A : \mu_{new} - \mu_{current} > 0$

$n_{new} = n_{current} = ?, s_{new} = 2.2 = s_{current} = 2.2$

$\delta = 0.5, \alpha = 0.05$, power = 90%, $\beta = 0.10$

16

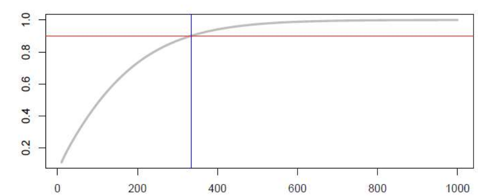
Back to medical surveys...

How large a sample size would you need if you wanted 90% power to detect a 0.5 increase in average number of surveys taken at the 5% significance level?

$H_0 : \mu_{new} - \mu_{current} = 0, H_A : \mu_{new} - \mu_{current} > 0$

$n_{new} = n_{current} = ?, s_{new} = 2.2 = s_{current} = 2.2$

$\delta = 0.5, \alpha = 0.05$, power = 90%, $\beta = 0.10$



See videos for discussion on how to calculate sample size n that results in a certain power level.

When $n > 334$, power is at least 90%.

16

If you're interested...

```
s = 2.2
mu = 0
delta = 0.5

ns = 10:1000
power = rep(NA, length(ns))

for(i in 10:1000){
  n = i
  t_star = qt(0.95, df = n-1)
  se = sqrt((s^2 / n) + (s^2 / n))
  cutoff = t_star * se
  t_cutoff = (cutoff - (mu+delta)) / se
  power[i-9] = pt(t_cutoff, df = n-1, lower.tail = FALSE)
}

which_n = which.min(abs(power - 0.9))
power[which_n]
power[which_n + 1]
ns[which_n + 1]
```

17

Application exercise: 4.3

See course website for details.

18

Summary of main ideas

1. Not every statistically significant result is practically significant
2. Hypothesis tests have error rates associated with them
3. Type 1 error rate = significance level
4. Calculating the power is a two step process
5. Power goes up with effect size and sample size, and is inversely proportional with significance level and standard error
6. A priori power calculations determine desired sample size

19