Chapter 7

Inference for numerical data[1]

Department of Mathematics & Statistics
North Carolina A&T State University

[1] These notes use content from OpenIntro Statistics Slides by Mine Cetinkaya-Rundel.

Computing the power for a 2-sample test

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true		
	H_A true		

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true		Type 1 Error, α
	H_A true		

- Type 1 error is rejecting H_0 when you shouldn't have, and the probability of doing so is α (significance level)

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true		Type 1 Error, α
	H_A true	Type 2 Error, β	

- Type 1 error is rejecting H_0 when you shouldn't have, and the probability of doing so is α (significance level)
- Type 2 error is failing to reject H_0 when you should have, and the probability of doing so is β (a little more complicated to calculate)

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	$1-\alpha$	Type 1 Error, α
	H_A true	Type 2 Error, β	

- Type 1 error is rejecting H_0 when you shouldn't have, and the probability of doing so is α (significance level)
- Type 2 error is failing to reject H_0 when you should have, and the probability of doing so is β (a little more complicated to calculate)
- **Power** of a test is the probability of correctly rejecting H_0, and the probability of doing so is $1-\beta$

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

- Type 1 error is rejecting H_0 when you shouldn't have, and the probability of doing so is α (significance level)
- Type 2 error is failing to reject H_0 when you should have, and the probability of doing so is β (a little more complicated to calculate)
- **Power** of a test is the probability of correctly rejecting H_0, and the probability of doing so is $1 - \beta$
- In hypothesis testing, we want to keep α and β low, but there are inherent trade-offs.

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.
- 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.
- Clearly, β depends on the **effect size** (δ).

Example - Blood Pressure (BP), hypotheses

Suppose a pharmaceutical company has developed a new drug for lowering blood pressure, and they are preparing a clinical trial to test the drug's effectiveness. They recruit people who are taking a particular standard blood pressure medication, and half of the subjects are given the new drug (treatment) and the other half continue to take their current medication through generic-looking pills to ensure blinding (control). What are the hypotheses for a two-sided hypothesis test in this context?

Example - Blood Pressure (BP), hypotheses

Suppose a pharmaceutical company has developed a new drug for lowering blood pressure, and they are preparing a clinical trial to test the drug's effectiveness. They recruit people who are taking a particular standard blood pressure medication, and half of the subjects are given the new drug (treatment) and the other half continue to take their current medication through generic-looking pills to ensure blinding (control). What are the hypotheses for a two-sided hypothesis test in this context?

$$H_0 : \mu_{treatment} - \mu_{control} = 0$$
$$H_A : \mu_{treatment} - \mu_{control} \neq 0$$

Example - BP, standard error

Suppose researchers would like to run the clinical trial on patients with systolic blood pressures between 140 and 180 mmHg. Suppose previously published studies suggest that the standard deviation of the patients' blood pressures will be about 12 mmHg and the distribution of patient blood pressures will be approximately symmetric. If we had 100 patients per group, what would be the approximate standard error for difference in sample means of the treatment and control groups?

Example - BP, standard error

Suppose researchers would like to run the clinical trial on patients with systolic blood pressures between 140 and 180 mmHg. Suppose previously published studies suggest that the standard deviation of the patients' blood pressures will be about 12 mmHg and the distribution of patient blood pressures will be approximately symmetric. If we had 100 patients per group, what would be the approximate standard error for difference in sample means of the treatment and control groups?

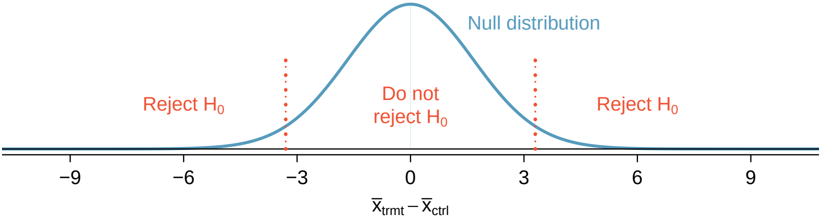
$$SE = \sqrt{\frac{12^2}{100} + \frac{12^2}{100}} = 1.70$$

Example - BP, minimum effect size required to reject H_0

For what values of the difference between the observed averages of blood pressure in treatment and control groups (effect size) would we reject the null hypothesis at the 5% significance level?

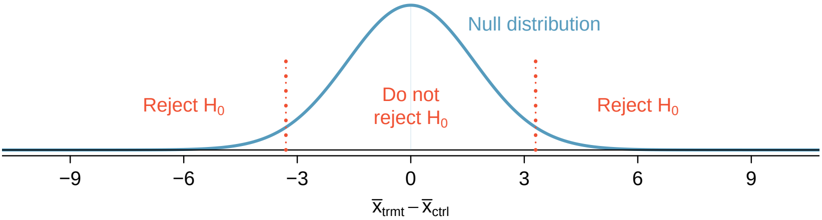
Example - BP, minimum effect size required to reject H_0

For what values of the difference between the observed averages of blood pressure in treatment and control groups (effect size) would we reject the null hypothesis at the 5% significance level?



Example - BP, minimum effect size required to reject H_0

For what values of the difference between the observed averages of blood pressure in treatment and control groups (effect size) would we reject the null hypothesis at the 5% significance level?



The difference should be at least

$$1.96 \times 1.70 = 3.332.$$

or at most

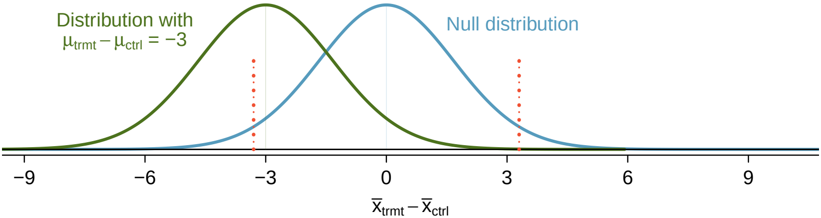
$$-1.96 \times 1.70 = -3.332.$$

Example - BP, power

Suppose that the company researchers care about finding any effect on blood pressure that is 3 mmHg or larger vs the standard medication. What is the power of the test that can detect this effect?

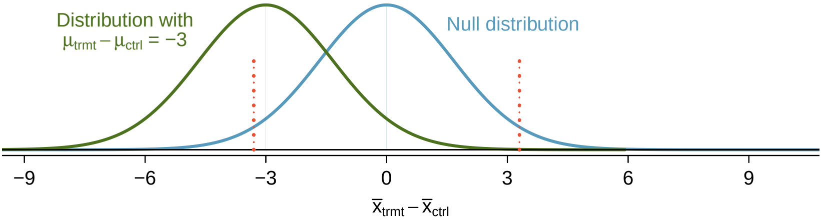
Example - BP, power

Suppose that the company researchers care about finding any effect on blood pressure that is 3 mmHg or larger vs the standard medication. What is the power of the test that can detect this effect?



Example - BP, power

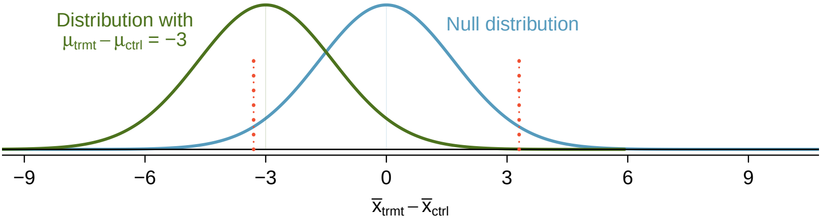
Suppose that the company researchers care about finding any effect on blood pressure that is 3 mmHg or larger vs the standard medication. What is the power of the test that can detect this effect?



$$Z = \frac{-3.332 - (-3)}{1.70} = 0.20$$

Example - BP, power

Suppose that the company researchers care about finding any effect on blood pressure that is 3 mmHg or larger vs the standard medication. What is the power of the test that can detect this effect?



$$Z = \frac{-3.332 - (-3)}{1.70} = 0.20$$

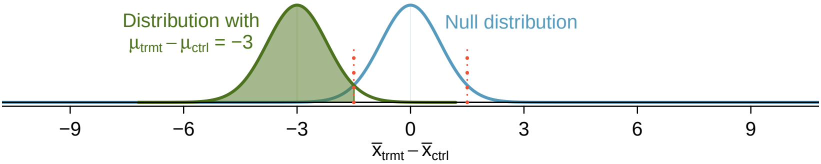
$$P(Z < -0.20) = 0.4207$$

Example - BP, required sample size for 80% power

What sample size will lead to a power of 80% for this test?

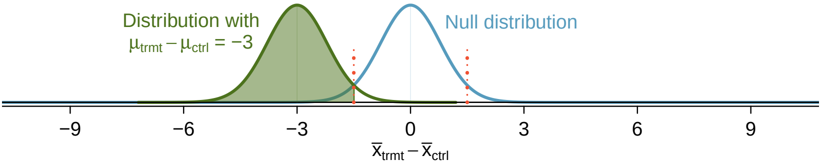
Example - BP, required sample size for 80% power

What sample size will lead to a power of 80% for this test?



Example - BP, required sample size for 80% power

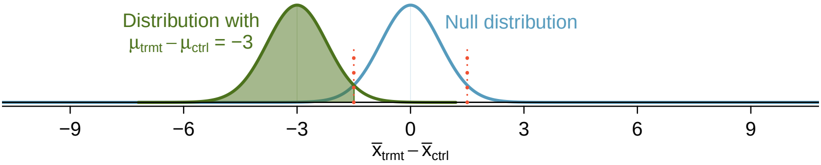
What sample size will lead to a power of 80% for this test?



$$SE = \frac{3}{2.8} = 1.07142$$

Example - BP, required sample size for 80% power

What sample size will lead to a power of 80% for this test?

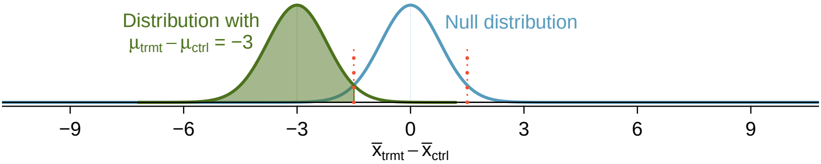


$$SE = \frac{3}{2.8} = 1.07142$$

$$1.07142 = \sqrt{\frac{12^2}{n} + \frac{12^2}{n}}$$

Example - BP, required sample size for 80% power

What sample size will lead to a power of 80% for this test?



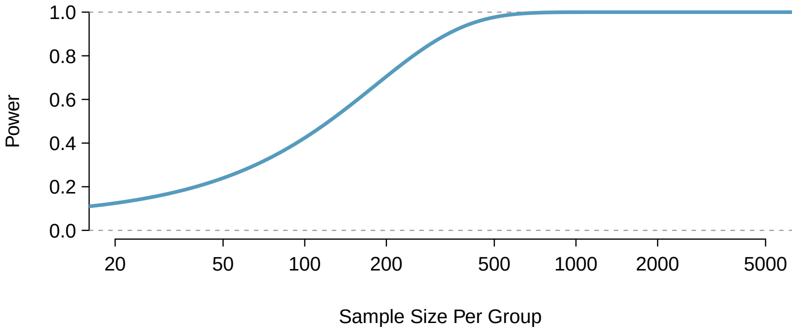
$$SE = \frac{3}{2.8} = 1.07142$$

$$1.07142 = \sqrt{\frac{12^2}{n} + \frac{12^2}{n}}$$

$$n = 250.88 \rightarrow n \geq 251$$

Recap

- Calculate required sample size for a desired level of power.
- Calculate power for a range of sample sizes, then choose the sample size that yields the target power (usually 80% or 90%).



Achieving desired power

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

Achieving desired power

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

1. Increase the sample size.

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 essentially has the same effect as 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 homogeneous may help.

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 essentially has the same effect as 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 homogeneous 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).

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 essentially has the same effect as 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 homogeneous 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.