Decision errors

Prof. Dr. Jan Kirenz

Four possible scenarios in a hypothesis test

	Test conclusion	
Truth	**Reject null hypothesis**	**Fail to reject null hypothesis**
Null hypothesis is true	Type 1 Error	Good decision
Alternative hypothesis is true	Good decision	Type 2 Error

Significance levels should reflect consequences of errors.

- The **significance level** provides the **cutoff** for the p-value which will lead to a decision of “reject the null hypothesis.”
- Typical significance value is **0.05**

- If making a **Type 1 Error** is dangerous or especially costly, we should choose a _____ significance level
- If a **Type 2 Error** is relatively more dangerous or much more costly than a Type 1 Error, then we should choose a _____ significance level

One-sided vs two-sided hypotheses

- In earlier case studies, we've actually ignored some possibilities
- Problem: "**Confirmation bias**"

- What if men are actually discriminated against?
- What if the money trick actually makes students spend more?

Two sided hypotheses

Blood thinner experiment

- Here we consider an experiment with **patients** who underwent an operation for a heart attack and were subsequently admitted to a hospital.

- Each patient was randomly assigned to either receive a
  - **blood thinner** (**treatment group**) or
  - **not** receive a **blood thinner** (**control group**).
- The outcome variable of interest was whether the patient **survived** for at least 24 hours.

Böttiger, Bernd W, Christoph Bode, Sabine Kern, André Gries, René Gust, Rolf Glätzer, Harald Bauer, Johann Motsch, and Eike Martin. 2001. "Efficacy and Safety of Thrombolytic Therapy After Initially Unsuccessful Cardiopulmonary Resuscitation: A Prospective Clinical Trial." The Lancet 357 (9268): 1583–85.

Two sided hypothesis

- p_C = true survival rate of people who do not receive a blood thinner (**control group**)
- p_T = true survival rate of people receiving a blood thinner (**treatment group**)

H_0: Blood thinners do not have an overall survival effect,

i.e., the survival proportions are **the same** in each group. $p_T - p_C = 0$

H_1: Blood thinners have an impact on survival,

either positive or negative, but not zero. $p_T - p_C \neq 0$

One sided hypothesis

- p_C = true survival rate of people who do not receive a blood thinner (**control group**)
- p_T = true survival rate of people receiving a blood thinner (**treatment group**)

H_0: Blood thinners do not have an overall survival effect,

i.e., the survival proportions for the blood thinner group is **the same or lower** than the control group. $p_T - p_C <= 0$

H_1: Blood thinners have a **positive** impact on survival. $p_T - p_C > 0$

Group	Died	Survived	Total
Control	39	11	50
Treatment	26	14	40
Total	65	25	90

$p'_C = 11/50 = 0.22$

$p'_T = 14/40 = 0.35$

$p'_T - p'_C = 0.35 - 0.22 = 0.13$

There were **50 patients** in the experiment who did not receive a blood thinner and **40 patients** who did.

What is the observed survival rate in the control group? And in the treatment group?

Also, provide a point estimate

($p'_C - p'_T$) for the true difference in population survival proportions across the two groups: $p_C - p_T$

However, we wonder if this difference could be easily explainable by **chance**, if the treatment has no effect on survival…

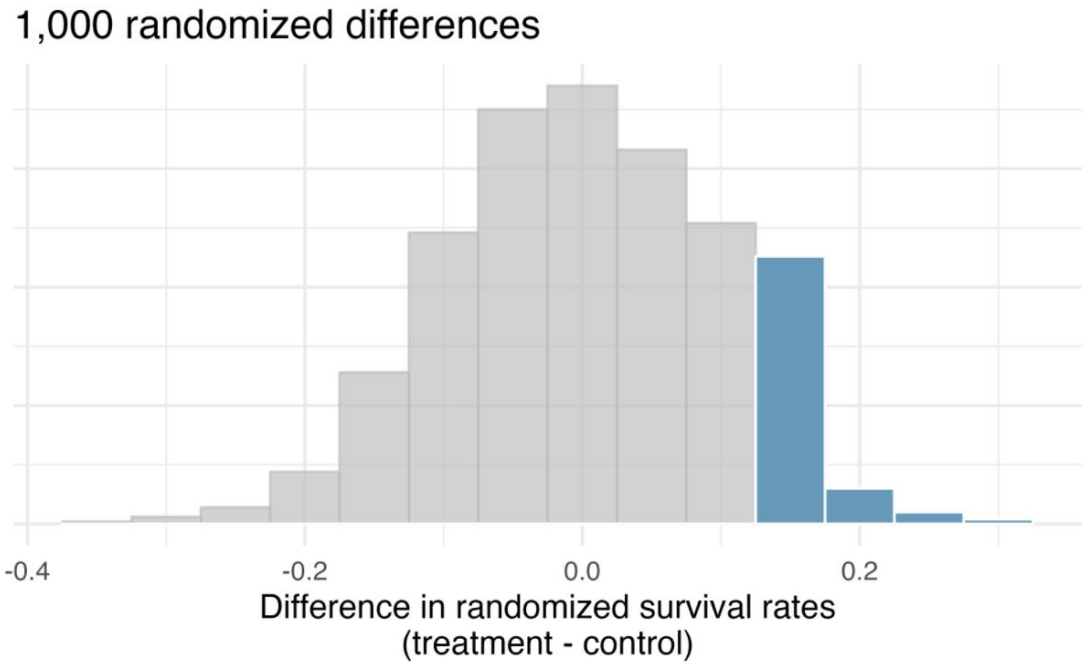


Figure 14.1: Null distribution of the point estimate for the difference in proportions, $\hat{p}_T - \hat{p}_C$. The shaded right tail shows observations that are at least as large as the observed difference, 0.13.

The right tail area is **0.135.**

(Note: it is only a coincidence that we also have $p'_T - p'_C = 0.13$)

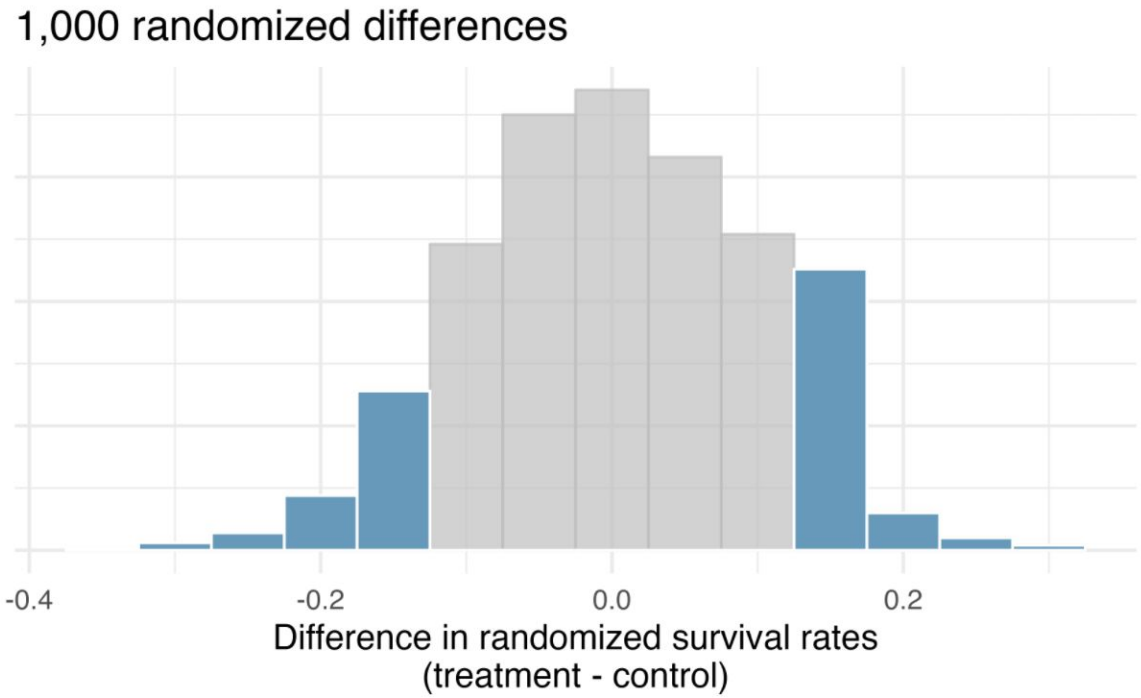


Figure 14.2: Null distribution of the point estimate for the difference in proportions, $\hat{p}_T - \hat{p}_C$. All values that are at least as extreme as +0.13 but in either direction away from 0 are shaded.

For a **two-sided test**, take the single tail (in this case, 0.131) and double it to get the **p-value: 0.262.**

Since this p-value is larger than 0.05, we do not reject the null hypothesis. That is, we do not find convincing evidence that the blood thinner has any influence on survival of patients who undergo CPR prior to arriving at the hospital.

Default to a two-sided test.

We want to be rigorous and keep an open mind when we analyze data and evidence. Use a one-sided hypothesis test only if you truly have interest in only one direction.

Computing a p-value for a two-sided test.

First compute the p-value for one tail of the distribution, then double that value to get the two-sided p-value. That's it!

Power

Often times in experiment planning, there are two competing considerations:

- We want to collect **enough data** that we can detect important effects.
- Collecting data can be **expensive**, and, in experiments involving people, there may be some risk to patients.

When planning a study, we want to know how likely we are to **detect** an **effect** we care about.

Power.

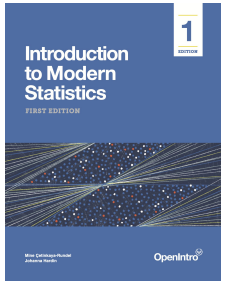
The power of the test is the probability of rejecting the null claim when the alternative claim is true.

How easy it is to detect the effect depends on both how big the effect is (e.g., how good the medical treatment is) as well as the sample size.

Terms you should know

confirmation bias	power	type 1 error
null distribution	significance level	type 2 error
one-sided hypothesis test	two-sided hypothesis test	

Resources



The content of this lecture is mainly based on the excellent book (can be accessed for free)

- “Introduction to Modern Statistics” by Mine Çetinkaya-Rundel and Johanna Hardin (2021)

  https://openintro-ims.netlify.app/index.html