Hypothesis Testing

Applied Regression in R

Aleš Vomáčka

09. 03. 2026

Faculty of Arts, Charles University

Logic of Hypothesis Testing

Logic of Hypothesis Testing

Usually, we are mostly interested in uncertainity of our estimates.

But sometimes, we want to check validity of a specific claim about the state of the world.

This **quantifiable claim about the state of the world** is called a (statistical) hypothesis.

Logic of Hypothesis Testing

Statistical hypothesis has to be a quantifiable claim about the state of the world.

Which one of these is a valid statistical hypothesis?

1. People with more casual attitudes towards sex tend to have more sex partners?
2. Correlation between sexual attitudes index and number of sexual partners is > 0.

Logic of Hypothesis Testing

Hypothesis testing can be done in both frequentist and bayesian way.

We will focus on **frequentist** approach (which uses **p values**).

Bayesian hypothesis testing is also possible (using **Bayes factor**), but not particularly popular.

Frequentist Hypothesis Testing

Frequentist Hypothesis Testing

Roughly speaking, all tests follow the same approach:

1. Formulate a statement about the world (hypothesis).
2. Create model of the world where the hypothesis is true.
3. Check how likely your actual data would be in this fictional world.

Frequentist Hypothesis Testing

Example: We think attitudes towards casual sex are unrelated to amount of sexual partners.

But we want to formally test this hypothesis.

Step 1: Formulate a hypothesis

We need to formulate a quantifiable claim about the state of the world. We'll call it a **null hypothesis**.

What could that be?

Step 1: Formulate a hypothesis

We need to formulate a quantifiable claim about the state of the world. We'll call it a **null hypothesis**.

What could that be?

The regression slope of model predicting number of sexual partners using sexual liberalism index is zero.

Step 2: Create a model

We need to create model representing a world where our null hypothesis is true. This requires making some **assumptions** (linearity, homoskedasticity, normality, etc.).

Step 2: Create a model

We need to create model representing a world where our null hypothesis is true. This requires making some **assumptions** (linearity, homoskedasticity, normality, etc.).

In our case, the model is

$$\text{sex partners} = \beta_0 + \beta_1 \cdot \text{sex index} + e$$

Step 2: Create a model

We need to create model representing a world where our null hypothesis is true. This requires making some **assumptions** (linearity, homoskedasticity, normality, etc.).

In our case, the model is

$$\text{sex partners} = \beta_0 + \beta_1 \cdot \text{sex index} + e$$

Where e is normally distributed with mean of 0 and variance estimated from data.

Step 3: Check model against data

Our model is

$$\text{sex partners} = \beta_0 + \beta_1 \cdot \text{sex index} + e$$

Our null hypothesis is that $\beta_1 = 0$

Step 3: Check model against data

Using our data, the actualy observed value is 3.99.

Now we can ask „*Assuming our null hypothesis and our assumptions are true, what is the probability we'd observe value of 3.99 or higher?*“

What is this probability called?

Parameter	Coefficient
(Intercept)	-2.87
sexlib_index	3.99

Step 3: Check model against data

P value*is the **probability** of seeing the observed test statistic or one more extreme, **assuming the null hypothesis** (and assumptions) are true.

In our case, the p value of $\beta_1 = 0$ is $5.111057e - 34$.

Step 3: Check model against data

P value is the **probability** of seeing the observered test statistic or one more extreme, **assuming the null hypothesis** (and assumptions) are true.

In our case, the p value of $\beta_1 = 0$ is $00000000000000000000000000000000005.111057\%$.

Step 3: Check model against data

P value of $00000000000000000000000000000000000005.111057\%$ is extremely low.

This could mean two things. Either:

1. We've just witnessed an extraordinary rare event.
2. Our null hypothesis is not true.

Most people pick option 2 if p value is < 0.05.

Questions?

Caveats (& Rants)

Common Misconceptions

P value is **not**:

1. Probability the null hypothesis is true (we need to assume it's true to compute the p value).
2. Probability that the result is (only) due to chance (assumptions can skew it).

Is It Actually Worth It?

Null hypothesis testing *very* common in sociology.

I don't think it's particularly useful. Two reasons:

- Hypthesis testing is often not enough.
- Formulating a good hypothesis is hard.

Hypthesis testing is often not enough.

Hypothesis testing often doesn’t provide enough information for responsible decision making.

Imagine you are testing the effect of subsidizing school lunches on math performance at elementary school.

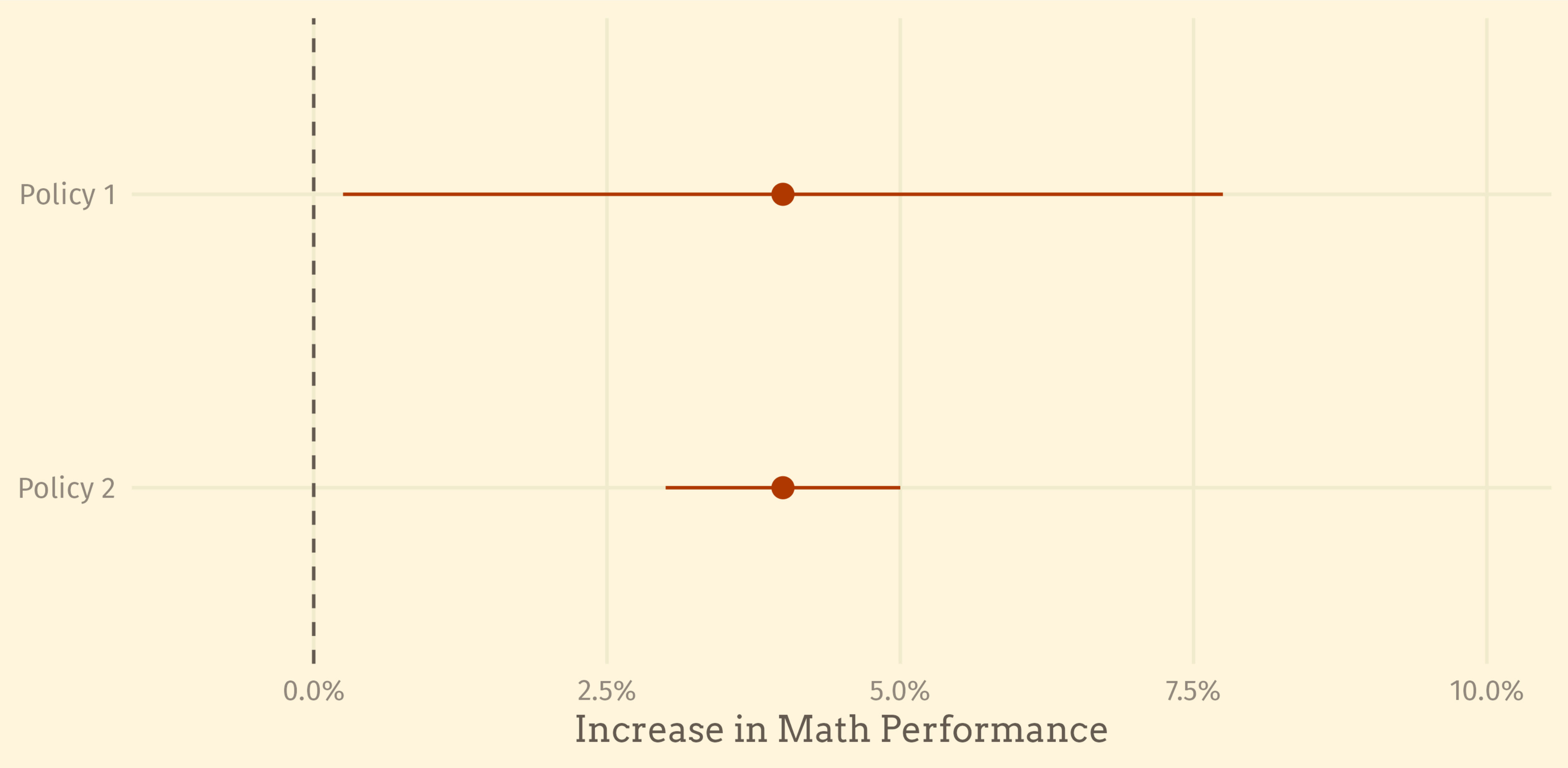
You run a test and report „*There is statistically significant, positive effect of subsidizing school lunches on math performance*“

Hypthesis testing is often not enough.

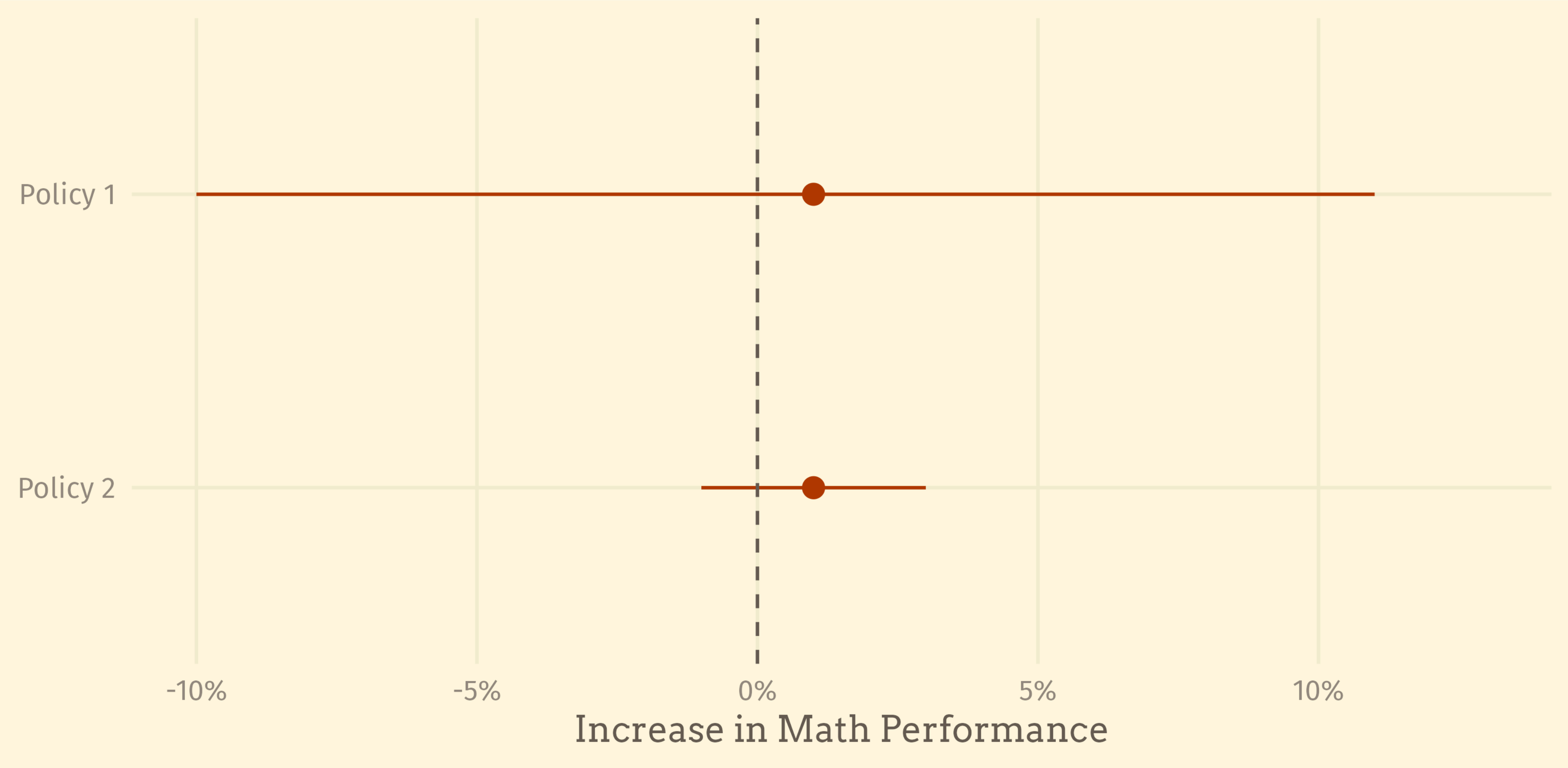
You run a test and report „*There is statistically significant, positive effect of subsidizing school lunches on math performance*“

Will this information be enough? Almost certainly not.

Hypthesis testing is often not enough.



Hypthesis testing is often not enough.



Questions?

Formulating a good hypothesis is hard.

Vast majority of sociology studies use hypothesis testing to check whether two phenomena are completly unrelated.

But in a system as complicated as human society, how likely it is for two things to have absolutely no connection.

The logic of a usual sociology paper

1. Define research question („*Are single mothers at higher risk of poverty than single fathers?*“)

The logic of a usual sociology paper

1. Define research question („*Are single mothers at higher risk of poverty than single fathers?*“)
2. Present a literature review showing there almost certainly is a relationship („*40 years of research show women are systematically at a disadvantage during parenthood*“)

The logic of a usual sociology paper

1. Define research question („*Are single mothers at higher risk of poverty than single fathers?*“)
2. Present a literature review showing there almost certainly is a relationship („*40 years of research show women are systematically at a disadvantage during parenthood*“)
3. Test a null hypothesis that parent gender and risk of poverty are complety unrelated.

The logic of a usual sociology paper

1. Define research question („*Are single mothers at higher risk of poverty than single fathers?*“)
2. Present a literature review showing there almost certainly is a relationship („*40 years of research show women are systematically at a disadvantage during parenthood*“)
3. Test a null hypothesis that parent gender and risk of poverty are completly unrelated.
4. See low p value, conclude the hypothesis no one believed in the first place is most likely false...

Formulating a good hypothesis is hard.

Very often, we don't have good, nontrivial hypothesis to test. We are just interested in how strong a relationship is. That's ok.

To be clear, not a problem of p values etc. It is defintely possible to do hypothesis testing in a useful way.

Haven't see it in sociology yet...

Questions?

Bonus: Tests as Models

Bonus: Tests as Models

Classical statistics classes focus on rote memorization of selected set of simple tests.

This takes a lot of memory space and makes it hard to learn advanced topics.

Fortunately, most popular tests are actualy linear models!

Bonus: Tests as Models

(Equal variance) t test is just linear regression with binary predictor.

This

```
t.test(sex_partners ~ gender, var.equal = TRUE)
```

is equal to:

```
lm(sex_partners ~ gender)
```

Bonus: Tests as Models

Unequal variance t test:

```r
t.test(sex_partners ~ gender, var.equal = FALSE)
```

is (mostly) equal to:

```r
lm(sex_partners ~ gender) %>% parameters(vcov = "HC3")
```

Slighly differences because there are different ways to account for unequal variance. Welch's t test isn't necessarily the best.

Bonus: Tests as Models

Mann-Whitney/Wilcoxon Sum Rank Test:

```
wilcox.test(sex_partners ~ gender)
```

Is (mostly) equal:

```
lm(rank(sex_partners) ~ gender)
```

Bonus: Tests as Models

ANOVA:

```
aov(sex_partners ~ edu)
```

is just test of regression residuals:

```
lm(sex_partners ~ edu) %>% anova()
```

Bonus: Tests as Models

Even things that don't look like linear models are often linear models.

Including Chi Squared test for contingency tables

gender	Elementary	High School	University
Men	99	154	118
Women	168	98	40

Bonus: Tests as Models

Chi squared test:

```
table(sex$gender, sex$education) |> chisq.test()
```

is equivalent to this:

```
sex |>
    count(gender, education) |>
    glm(n ~ gender * education, family = poisson(), data = _) |>
    anova(test = "Rao")
```

Bonus: Tests as Models

Why does it matter?

1. You don't need to waste mental capacity on remembering a ton of different names (t test, Mann-Whitney,...)
2. I t makes learning new things much easier. Majority of applied statistics boils down „*Linear model, but...*"

Questions?