*corinna.birner@stud-mail.uni-wuerzburg.de*
*max.mueller@stud-mail.uni-wuerzburg.de*

# Understanding Metrics based on Mastering Metrics

Corinna Birner & Max Müller

University of Würzburg

May 8, 2020

# Chapter 1: Randomized Trials

# Overview

Today we will start our journey on the path from cause to effect. Therefore, we will discuss randomized trials as one possible way of finding causal relations in data. Our topics will be the following:

Introduction

The RAND Experiment

The Oregon Trail

Appendix

# The path from cause to effect

This lecture will help you understand how to deal with econometrics. It is not always easy to get from data to a causal effect between things. We will go on the path from cause to effect, trying to reach a ceteris paribus condition.

ceteris paribus stands for a condition with causal interpretation!

On our journey, we will get to know the Furious Five of econometric research:

- random assignment
- regression
- instrumental variables
- regression discontinuity designs
- differences-in-differences

# The effects of health care

To start with, we are interested in the effects of health insurance. The main question we want to answer in our first lecture is whether health insurance leads to healthier people.

To find an answer to this question, we will look at two studies with random assignment: The RAND Health Insurance Experiment and the Oregon Trail
Both experiments took place in the USA, where people don't have a mandatory health insurance

# Health Care in the USA

- Medicare: health insurance for elderly Americans (over 65 years)
- Medicaid: health insurance for poor Americans
- emergency departments: have to accept everyone but can't provide long time care

To get to a causal effect, we would need to compare the health of a person without health insurance with the health of the same person had she chosen health insurance. As people can either have or have no health insurance (at least not at the same time), we can't use this comparison.
Comparing people who have chosen insurance to people who haven't is also not a good idea due to *selection bias*.

# The National Health Interview Survey

- in the survey participants ranked their health (poor, fair, good, very good or excellent)
- the rankings were coded from 1 to 5, using 1 for poor and 5 for excellent health
- this ranking shows the outcome we're interested in (how healthy is a person?)
- causal relation here is determined by having health insurance or not
- we divide the participants in treatment group (with insurance) and control group (without insurance)

Health and demographic characteristics of insured and uninsured couples in the NHIS

| | Husbands | | | Wives | | |
|---|---|---|---|---|---|---|
| | Some HI (1) | No HI (2) | Difference (3) | Some HI (4) | No HI (5) | Difference (6) |
| | | | A. Health | | | |
| Health index | 4.01 [.93] | 3.70 [1.01] | .31 (.03) | 4.02 [.92] | 3.62 [1.01] | .39 (.04) |
| | | | B. Characteristics | | | |
| Nonwhite | .16 | .17 | −.01 (.01) | .15 | .17 | −.02 (.01) |
| Age | 43.98 | 41.26 | 2.71 (.29) | 42.24 | 39.62 | 2.62 (.30) |
| Education | 14.31 | 11.56 | 2.74 (.10) | 14.44 | 11.80 | 2.64 (.11) |
| Family size | 3.50 | 3.98 | −.47 (.05) | 3.49 | 3.93 | −.43 (.05) |
| Employed | .92 | .85 | .07 (.01) | .77 | .56 | .21 (.02) |
| Family income | 106,467 | 45,656 | 60,810 (1,355) | 106,212 | 46,385 | 59,828 (1,406) |
| Sample size | 8,114 | 1,281 | | 8,264 | 1,131 | |

*Notes:* This table reports average characteristics for insured and uninsured married couples in the 2009 National Health Interview Survey (NHIS). Columns (1), (2), (4), and (5) show average characteristics of the group of individuals specified by the column heading. Columns (3) and (6) report the difference between the average characteristic for individuals with and without health insurance (HI). Standard deviations are in brackets; standard errors are reported in parentheses.

- Americans with insurance seem to be healthier
- But there are differences in people with and without coverage!
- those with insurance have higher education, higher income, ...

# How to measure causal effect

First, we have to establish how to write down our variables:

- Outcome: $Y$ (in our example: health)
- $Y_i$: health of individual $i$
- every person $i$ has two potential outcomes: $Y_{0i}$ (without insurance) and $Y_{1i}$ (with insurance)
- the causal effect of insurance: $Y_{1i}$ - $Y_{0i}$

# Example

Imagine two people: Max and Moritz. Max is very healthy and fit, Moritz has always had a fragile health. They can both decide whether to acquire health insurance or not. Let's have a look at their health index (ranked from poor (1) to excellent (5):

- $Y_{0,\,\mathrm{Max}} = 5$ and $Y_{1,\,\mathrm{Max}} = 5$
- $Y_{0,\,\mathrm{Moritz}} = 3$ and $Y_{1,\,\mathrm{Moritz}} = 4$
- the effect of getting insurance for Max $= 0$ and for Moritz $= 1$
- problem: *selection bias* between those two
- Moritz has worse health even if both of them decide not to get insurance

# Average Causal Effects

The problem of the selection bias is consistent also in groups. Therefore we look at the **average causal effects**. In a group of $n$ people, the average causal effect is:

$$Avg_n[Y_{1,i} - Y_{0,i}] = \frac{1}{n}\sum_{i=1}^{n}[Y_{1,i} - Y_{0,i}]$$

$$= \frac{1}{n}\sum_{i=1}^{n}[Y_{1,i}] - \frac{1}{n}\sum_{i=1}^{n}[Y_{0,i}]$$

- We compare average health in a scenario where everybody in the group does and does not have health insurance
- to make this comparison easier, we will use a *dummy variable*, $D_i$
- $D_i = 1$ means insured, $D_i = 0$ means uninsured

# Difference in group means

We can now use Dummies for the average in the two groups ($Avg_n[Y_i|D_i = 1]$ for the insured and $Avg_n[Y_i|D_i = 0]$ for the uninsured).

Therefore the difference in group means would be:

$$Avg_n[Y_i|D_i = 1] - Avg_n[Y_i|D_i = 0]$$

$$= Avg_n[Y_{1i}|D_i = 1] - Avg_n[Y_{0i}|D_i = 0]$$

# constant-effects assumption

We are interested in $Avg_n[Y_{1i} - Y_{0i}]$, therefore we can imagine that people with insurance are healthier than people without an insurance.
Let's say that everyone is healthier through the insurance by the amount $\kappa$.
We can use the *constant-effects assumption*

$$Y_{1i} = Y_{0i} + \kappa$$

$\kappa$ stands for the individual and average causal effect of insurance on health.

# Oranges & Apples

Using the constant-effects models we can substitute $Avg_n$ and get:

$$Avg_n[Y_{1i}|D_i = 1] - Avg_n[Y_{0i}|D_i = 0]$$

$$= \kappa + Avg_n[Y_{0i}|D_i = 1] - Avg_n[Y_{0i}|D_i = 0]$$

- But: difference in group means = average causal effect + selection bias!
- $Y_{0i}$ is any characterictic related to health that's not insurance and therefore comparing outcomes for two groups with different $Y_{0i}$ is comparing oranges with apples.
- Even if the insurance effect $\kappa$ was zero, our groups would differ in their health.

# Randomized Experiments

- Experimental random assignment eliminates the selection bias.
- Idea: we start with a sample of people that is currently uninsured
- we would give health insurance to a randomly chosen subset of this sample
- due to random assignment we can compare these groups ceteris paribus (they differ only in their health insurance status)
- Watch out: two make two randomly chosen groups comparable, they have to be large enough (Law of Large Numbers LLN)!

# Law of Large Numbers

- The LLN shows the behavior of sample averages in relation to sample size
- For example, imagine to play dice: everytime you throw the dice, you write down your result and average these results
- You have six possible outcomes (numbers 1 to 6) but after enough throws you would expect to get 3,5 on average!
- this average value represents the mathematical expectation

# mathematical and conditional expectations

## mathematical expectation

The mathematical expectation of a variable $E[Y_i]$ is the average obtained if everyone in the survey population from which the sample os drawn was to be enumerated.

## conditional expectation

The conditional expectation of a variable $Y_i$, given a dummy variable $D_i = 1$, is written $E[Y_i|D_i = 1]$. This expression stands for the average of $Y_i$ in the population that has $D_i = 1$.

## random assignment

As randomly assigned treatment and control group come from the same population, they are the same in every way, also in their expected $Y_{0i}$. Therefore, their conditional expectations $E[Y_{0i}|D_i = 1]$ and $E[Y_{0i}|D_i = 0]$ are the same.

## random assignment eliminates selection bias

When $D_i$ is randomly assigned, $E[Y_{0i}|D_i = 1] = E[Y_{0i}|D_i = 0]$ and therefore the difference in expectations by the treatment shows the causal effect of the treatment.

$$
\begin{aligned}
&E[Y_i|D_i = 1] - E[Y_i|D_i = 0] \\
&= E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0] \\
&= E[Y_{0i} + \kappa|D_i = 1] - E[Y_{0i}|D_i = 0] \\
&= \kappa
\end{aligned}
$$

- Conclusion: when the sample is large enough, due to the LLN, the selection bias disappears in a randomized experiment! Random assignment ensures that the mix of the individuals in the different groups is the same.
- when analyzing data from a randomized trial, you should always be checking for balance to make sure your treatment and control groups look similar.

# The RAND Health Insurance Experiment

- We will now talk about the RAND Health Insurance Experiment in which 3958 participants were randomly assigned to health different health insurance plans.
- The experiment was carried out from 1974 to 1982 in the USA.
- Participants were randomly assigned into 14 insuranceplans.
- These can be roughly categorized into 4 types:
  - Catastrophic: families pay 95 % of their healthcare costs themselves (capped at $1000)
  - uninsured: families pay 95 % of their healthcare costs themselves (capped at $450)
  - Coinsurance: families pay 25-50 % if their healthcare costs themselves (capped at $1000)
  - Free: would be similar to the German health insurance

# The RAND experiment: first glance

Demographic characteristics and baseline health in the RAND HIE

| | Means | Differences between plan groups | | | |
|---|---|---|---|---|---|
| | Catastrophic plan (1) | Deductible – catastrophic (2) | Coinsurance – catastrophic (3) | Free – catastrophic (4) | Any insurance – catastrophic (5) |
| **A. Demographic characteristics** | | | | | |
| Female | .560 | −.023 | −.025 | −.038 | −.030 |
| | | (.016) | (.015) | (.015) | (.013) |
| Nonwhite | .172 | −.019 | −.027 | −.028 | −.025 |
| | | (.027) | (.025) | (.025) | (.022) |
| Age | 32.4 | .56 | .97 | .43 | .64 |
| | [12.9] | (.68) | (.65) | (.61) | (.54) |
| Education | 12.1 | −.16 | −.06 | −.26 | −.17 |
| | [2.9] | (.19) | (.19) | (.18) | (.16) |
| Family income | 31,603 | −2,104 | 970 | −976 | −654 |
| | [18,148] | (1,384) | (1,389) | (1,345) | (1,181) |
| Hospitalized last year | .115 | .004 | −.002 | .001 | .001 |
| | | (.016) | (.015) | (.015) | (.013) |
| **B. Baseline health variables** | | | | | |
| General health index | 70.9 | −1.44 | .21 | −1.31 | −.93 |
| | [14.9] | (.95) | (.92) | (.87) | (.77) |
| Cholesterol (mg/dl) | 207 | −1.42 | −1.93 | −5.25 | −3.19 |
| | [40] | (2.99) | (2.76) | (2.70) | (2.29) |
| Systolic blood pressure (mm Hg) | 122 | 2.32 | .91 | 1.12 | 1.39 |
| | [17] | (1.15) | (1.08) | (1.01) | (.90) |
| Mental health index | 73.8 | −.12 | 1.19 | .89 | .71 |
| | [14.3] | (.82) | (.81) | (.77) | (.68) |
| Number enrolled | 759 | 881 | 1,022 | 1,295 | 3,198 |

*Notes:* This table describes the demographic characteristics and baseline health of subjects in the RAND Health Insurance Experiment (HIE). Column (1) shows the average for the group assigned catastrophic coverage. Columns (2)–(5) compare averages in the deductible, cost-sharing, free care, and any insurance groups with the average in column (1). Standard errors are reported in parentheses in columns (2)–(5); standard deviations are reported in brackets in column (1).

- first step: checking for balance
- the participants in the different plans seem to be similar (Panel A)
- values in the brackets are *standard errors* so that you can check whether a difference is large enough to be *statistically significant*
- differences larger than two standard errors are *statistically significant*

# The RAND experiment: first glance

Demographic characteristics and baseline health in the RAND HIE

| | Means | Differences between plan groups | | | |
|---|---|---|---|---|---|
| | Catastrophic plan (1) | Deductible – catastrophic (2) | Coinsurance – catastrophic (3) | Free – catastrophic (4) | Any insurance – catastrophic (5) |
| **A. Demographic characteristics** | | | | | |
| Female | .560 | −.023 | −.025 | −.038 | −.030 |
| | | (.016) | (.015) | (.015) | (.013) |
| Nonwhite | .172 | −.019 | −.027 | −.028 | −.025 |
| | | (.027) | (.025) | (.025) | (.022) |
| Age | 32.4 | .56 | .97 | .43 | .64 |
| | [12.9] | (.68) | (.65) | (.61) | (.54) |
| Education | 12.1 | −.16 | −.06 | −.26 | −.17 |
| | [2.9] | (.19) | (.19) | (.18) | (.16) |
| Family income | 31,603 | −2,104 | 970 | −976 | −654 |
| | [18,148] | (1,384) | (1,389) | (1,345) | (1,181) |
| Hospitalized last year | .115 | .004 | −.002 | .001 | .001 |
| | | (.016) | (.015) | (.015) | (.013) |
| **B. Baseline health variables** | | | | | |
| General health index | 70.9 | −1.44 | .21 | −1.31 | −.93 |
| | [14.9] | (.95) | (.92) | (.87) | (.77) |
| Cholesterol (mg/dl) | 207 | −1.42 | −1.93 | −5.25 | −3.19 |
| | [40] | (2.99) | (2.76) | (2.70) | (2.29) |
| Systolic blood pressure (mm Hg) | 122 | 2.32 | .91 | 1.12 | 1.39 |
| | [17] | (1.15) | (1.08) | (1.01) | (.90) |
| Mental health index | 73.8 | −.12 | 1.19 | .89 | .71 |
| | [14.3] | (.82) | (.81) | (.77) | (.68) |
| Number enrolled | 759 | 881 | 1,022 | 1,295 | 3,198 |

*Notes:* This table describes the demographic characteristics and baseline health of subjects in the RAND Health Insurance Experiment (HIE). Column (1) shows the average for the group assigned catastrophic coverage. Columns (2)–(5) compare averages in the deductible, cost-sharing, free care, and any insurance groups with the average in column (1). Standard errors are reported in parentheses in columns (2)–(5); standard deviations are reported in brackets in column (1).

- first step: checking for balance
- good balance in pre-treatment index of general health (Panel B)
- no statistically significant differences in pre-treatment cholesterol, etc.

# The RAND experiment

| | Means | Differences between plan groups | | | |
|---|---|---|---|---|---|
| | Catastrophic plan (1) | Deductible – catastrophic (2) | Coinsurance – catastrophic (3) | Free – catastrophic (4) | Any insurance – catastrophic (5) |
| **A. Health-care use** | | | | | |
| Face-to-face visits | 2.78 [5.50] | .19 (.25) | .48 (.24) | 1.66 (.25) | .90 (.20) |
| Outpatient expenses | 248 [488] | 42 (21) | 60 (21) | 169 (20) | 101 (17) |
| Hospital admissions | .099 [.379] | .016 (.011) | .002 (.011) | .029 (.010) | .017 (.009) |
| Inpatient expenses | 388 [2,308] | 72 (69) | 93 (73) | 116 (60) | 97 (53) |
| Total expenses | 636 [2,535] | 114 (79) | 152 (85) | 285 (72) | 198 (63) |
| **B. Health outcomes** | | | | | |
| General health index | 68.5 [15.9] | −.87 (.96) | .61 (.90) | −.78 (.87) | −.36 (.77) |
| Cholesterol (mg/dl) | 203 [42] | .69 (2.57) | −2.31 (2.47) | −1.83 (2.39) | −1.32 (2.08) |
| Systolic blood pressure (mm Hg) | 122 [19] | 1.17 (1.06) | −1.39 (.99) | −.52 (.93) | −.36 (.85) |
| Mental health index | 75.5 [14.8] | .45 (.91) | 1.07 (.87) | .43 (.83) | .64 (.75) |
| Number enrolled | 759 | 881 | 1,022 | 1,295 | 3,198 |

*Notes:* This table reports means and treatment effects for health expenditure and health outcomes in the RAND Health Insurance Experiment (HIE). Column (1) shows the average for the group assigned catastrophic coverage. Columns (2)–(5) compare averages in the deductible, cost-sharing, free care, and any insurance groups with the average in column (1). Standard errors are reported in parentheses in columns (2)–(5); standard deviations are reported in brackets in column (1).

- Now we observe the health expenditure & outcomes
- with more generous insurance plan, more health care use (Panel A)
- the health care use was different among insurance groups
- but we can't find a great difference between groups in the health outcomes

# Does health insurance make people healthier?

- after looking at the results of the RAND Health Insurance Experiment, generous health care might have some unintended and undesirable consequences.
- due to increased health care use the costs would go up without improving health care outcomes.
- before trying to give a final answer to whether health insurance makes people healthier we will have a look at another experiment.

# The Oregon Trail

- RAND might have missed the mark
- Today´s uninsured Americans differ considerably from the RAND population: most of the uninsured are younger, less educated, poorer, and less likely to be working.
- The value of extra health care in such a group might be very different than for the middle class families that participated in the RAND Study
- What´s the effect of insuring the currently uninsured?
- Solution: Oregon Health Plan (OHP) lottery

# Oregon Health Plan

- State of Oregon offered Medicaid of thousands of randomly chosen people in a lottery
- Only Oregon residents 19-64 years old could apply
- Winners could apply for Medicaid $\rightarrow$ coverage was not automatic, even for lottery winners
- Out of 75000 Applicants 30000 were randomly selected and invited to apply. The other 45000 constitute the control group.
- Are OHP lottery winners more likely to end up insured as a result of winning?

OHP effects on insurance coverage and health-care use

| Outcome | Oregon | | Portland area | |
|---|---|---|---|---|
| | Control mean (1) | Treatment effect (2) | Control mean (3) | Treatment effect (4) |
| A. Administrative data | | | | |
| Ever on Medicaid | .141 | .256 (.004) | .151 | .247 (.006) |
| Any hospital admissions | .067 | .005 (.002) | | |
| Any emergency department visit | | | .345 | .017 (.006) |
| Number of emergency department visits | | | 1.02 | .101 (.029) |
| Sample size | 74,922 | | 24,646 | |
| B. Survey data | | | | |
| Outpatient visits (in the past 6 months) | 1.91 | .314 (.054) | | |
| Any prescriptions? | .637 | .025 (.008) | | |
| Sample size | 23,741 | | | |

*Notes:* This table reports estimates of the effect of winning the Oregon Health Plan (OHP) lottery on insurance coverage and use of health care. Odd-numbered columns show control group averages. Even-numbered columns report the regression coefficient on a dummy for lottery winners. Standard errors are reported in parentheses.



- OHP winners rather insured than lottery losers
- Treatment group used more healthcare services
- More emergency dep. visits

# But did that make them healthier?

OHP effects on health indicators and financial health

| Outcome | Oregon: Control mean (1) | Oregon: Treatment effect (2) | Portland area: Control mean (3) | Portland area: Treatment effect (4) |
|---|---|---|---|---|
| A. Health indicators | | | | |
| Health is good | .548 | .039 (.008) | | |
| Physical health index | | | 45.5 | .29 (.21) |
| Mental health index | | | 44.4 | .47 (.24) |
| Cholesterol | | | 204 | .53 (.69) |
| Systolic blood pressure (mm Hg) | | | 119 | −.13 (.30) |
| B. Financial health | | | | |
| Medical expenditures >30% of income | | | .055 | −.011 (.005) |
| Any medical debt? | | | .568 | −.032 (.010) |
| Sample size | 23,741 | | 12,229 | |

*Notes:* This table reports estimates of the effect of winning the Oregon Health Plan (OHP) lottery on health indicators and financial health. Odd-numbered columns show control group averages. Even-numbered columns report the regression coefficient on a dummy for lottery winners. Standard errors are reported in parentheses.

- not really
- statistically insignificant changes for physical and mental health
- changes in financial health could account for improved mental health

# Results

- Both experiments produces similar results
- Use of healthcare services increases sharply in response to insurance coverage
- But neither experiment reveals much of an insurance effect on physical health.
- These studies suggest that subsidized public health insurance should not be expected to yield a dramatic health dividend.

# In a nutshell

- **Causal inference** compares potential outcomes, descriptions of the world when alternative roads are taken.
- We cannot compare those who took one road with those who took another
- These comparisons are often contaminated by **selection bias**, that is, differences between treated and control subjects that exist even in the absence of a treatment effect.
- **Random assignment** to treatment and control conditions eliminates selection bias.
- But we should still **check for balances**.

# Mastering Interference

- How can we decide whether statistical results constitute strong evidence or are a lucky draw, unlikely to be replicated in repeated samples?
- How much sampling variance should we expect?
- Goal: The quantification of the uncertainty associated with a particular sample average and groups of averages and the differences among them.
- For example the RAND: Instead of studying the many millions of families, just about 2000 Families were selected at random and then randomly allocated to one of 14 plans or treatment groups.
- Two sorts of randomness: Random sampling and random assignment.

# Random Sampling

- Population mean of a variable is called mathematical expectation
- The Expectation of a variable $Y_i$ we write $E[Y_i]$.
- Unbiasedness of the Sample Mean $E[\bar{Y}] = E[Y_i]$

# Random Sampling

- Population mean of a variable is called mathematical expectation
- The Expectation of a variable $Y_i$ we write $E[Y_i]$.
  - $E[Y_i] \triangleright WholePopulation \triangleright Parameter$, *fixed feature*
  - $E[\bar{Y}] \triangleright Sample\ n \triangleright Sample\ Statistic$
- The sample average, $E[\bar{Y}]$, is a good estimator of $E[Y_i]$ (in statistics, an estimator is any function of sample data used to estimate parameters).

# Random Sampling

- LLN tells us that in large samples, the sample average is likely to be very close to the corresponding population mean.
- A related property is that the expectation of $E[\bar{Y}]$ is also $E[Y_i]$
- if we were to draw infinitely many random samples, the average of the resulting $E[\bar{Y}]$ across draws would be the underlying population mean.
- When a sample statistic has expectation equal to the corresponding population parameter, its said to be an unbiased estimator of that parameter.
- Unbiasedness of the Sample Mean $E[\bar{Y}] = E[Y_i]$

# Random Sampling

- Unbiasedness tells us that these deviations are not systematically up or down; rather, in repeated samples they average out to zero
- This is different from the LLN (Law of large Numbers)
- LLN says that the sample mean gets closer and closer to the population mean as the sample size grows.

# Measuring Variability

- To measure variability we need to look at average squared deviations from the mean, in which positive and negative gaps get equal weight.
- The resulting summary of variability is called variance.
- Sampling variance of sample average depends on the variance of the underlying observations $\sigma_y^2$ and the sample size $n$.

The sample variance of $Y_i$ in a sample size of n is defined as:

$$V(Y_i)^2 = \frac{1}{n}\sum_{i=1}^{n}(\mathrm{Y}_i - \bar{Y})^2$$

Because the expectation of the sample mean $E(\bar{Y})$ is $E[Y_i]$ the population variance of the sample mean can be written as:

$$V(\bar{Y}) = E[(\bar{Y} - E[\bar{Y}])^2] = E[(\bar{Y} - E[Y_i])^2]$$

- $V(Y_i) or(\sigma_y^2)$ denotes the variance of the underlying data, while $V(\bar{Y})$ is written for the variance of the sample mean

# Measuring Variability

- Simplified, the standard error of the sample mean can be written as:

$$SE(\bar{Y}) = \frac{\sigma_y}{\sqrt{n}}$$

- This summarizes the variability in an estimate due to random sampling
- Every estimate has an associated standard error
- But: Usually most quantities are unknown and must be estimated
- Therefore: We work with an estimated standard error:

$$\hat{SE}(\bar{Y}) = \frac{S(Y_i)}{\sqrt{n}}$$

# t-Statistic

- Now that we know how to measure variability, we just need to interpret it.
- Here: $E[Y_i] = \mu$
- This constitutes a working hypothesis and is a reference point called the null hypothesis
- Any t-Statistic for the sample mean under the working hypothesis is constructed as:

$$t(\mu) = \frac{\bar{Y} - \mu}{SE(\bar{Y})}$$

- The t-statistic is the ratio of the sample mean to its estimated standard error.
- With the t-statistic, we can interpret the results.

# Central Limit Theorem

- If $E[Y_i] = \mu$ and the sample size is large:
- The quantity $t(\mu)$ has a sampling distribution that is very close to a bell-shaped standard normal distribution.
- This is called the Central Limit Theorem (CLT)
- This allows us to make an empirically informed decision as to whether the available data support or cast doubt on the hypothesis that $E[Y_i]$ equals $\mu$.

# When is it significant?

- The standard normal distribution has a mean of 0 and standard deviation of 1.
- With any standard normal variable, values larger than $\pm 2$ are highly unlikely.
- Realizations larger than 2 in absolute value appear only about 5% of the time.
- Because the t-statistic is close to normally distributed, we similarly expect it to fall between about $\pm 2$ most of the time.
- Therefore, any t-statistic larger than about 2 (in absolute value) is too unlikely to be consistent with the null hypothesis used to construct it.
- When the null hypothesis is $= 0$ and the t-statistic exceeds 2 in absolute value, we say the sample mean is **significantly different from zero (statistically significant)**.

# Confidence intervalls

- The other side of this looks as follows
- Instead of checking whether the sample is consistent with a specific value of $\mu$, we can construct the set of all values of $\mu$ that are consistent with the data.
- That is called a confidence interval for $E[Y_i]$:

$$[\bar{Y} - 2 \times \hat{SE}(\bar{Y}), \bar{Y} + 2 \times \hat{SE}(\bar{Y})]$$

- When calculated in repeated samples, the interval should contain $E[Y_i]$ about 95% of the time.
- This interval is therefore said to be a 95% confidence interval for the population mean.

# Pairing Off

- Another option to measure if treatment changes outcomes:
- Hereby we use the treatment and the control group:
- $\bar{Y}^1$ is the treatment group mean, $\bar{Y}^0$ is the control group mean, with the sample size being $n = n^1 + n^0$
- The difference between $\bar{Y}^1$ and $\bar{Y}^0$ is an estimate of the causal effect of the treatment.
- If the treatment changes outcomes, there also must be different values for $\mu^1$ and $\mu^0$.

# Pairing Off

- Under the null hypothesis that $\mu^1$-$\mu^0$ is equal to the value $\mu$, the t-statistic for a difference in means is:

$$t(\mu) = \frac{\bar{Y}^1 - \bar{Y}^0 - \mu}{\hat{SE}(\bar{Y}^1 - \bar{Y}^0)}$$

- With the $\hat{SE}$ being the estimated standard error:

$$\hat{SE}(\bar{Y}^1 - \bar{Y}^0) = S(Y_i)\sqrt{\frac{1}{n_1} + \frac{1}{n_0}}$$

- $S(Y_i)$ in this case means the pooled sample standard deviation = both treatment and control group combined.
- We use this t-statistic to test working hypotheses about $\mu_1 - \mu_0$ and to construct confidence intervals for this difference.

# Pairing Off

- When the null hypothesis is one of equal means ($\mu = 0$), the statistic $t(\mu)$ equals the difference in sample means divided by the estimated standard error of this difference.
- When the t-statistic is large enough to reject a difference of zero, we say the estimated difference is statistically significant.
- Lack of statistical significance reflects lack of statistical precision (high sampling variance).

# Thank you very much for listening