INTRODUCTION TO CAUSAL INFERENCE

Michael Kühhirt

October 9, 2017

v.2017-11-22-9:17pm

Introduction and course organization

TODAY

1. What is causal inference?
2. Causal inference within empirical social research
3. A roadmap for causal inference
4. Course goals and learning outcomes
5. Requirements and grading
6. Course materials

WHAT IS CAUSAL INFERENCE?

THE SHORT ANSWER FOR NOW

It's the attempt to learn about causal relations between at least two variables from empirical data.

(see Heckman, 2000)

WHAT ARE CAUSAL RELATIONS?

Causal relations concern statements about cause and effect.

WHAT ARE CAUSAL RELATIONS?

Causal relations concern statements about cause and effect.

A causal statement claims that Y happened (or became more likely), because of X.

WHAT ARE CAUSAL RELATIONS?

Causal relations concern statements about cause and effect.

A causal statement claims that Y happened (or became more likely), because of X.

This implies a comparison of Y under alternative states of X.

WHAT ARE CAUSAL RELATIONS?

Causal relations concern statements about cause and effect.

A causal statement claims that Y happened (or became more likely), because of X.

This implies a comparison of Y under alternative states of X.

Namely, Y would be different, had X been x instead of x' (read: not x).

WHAT ARE CAUSAL RELATIONS?

Causal relations concern statements about cause and effect.

A causal statement claims that Y happened (or became more likely), because of X.

This implies a comparison of Y under alternative states of X.

Namely, Y would be different, had X been x instead of x' (read: not x).

In other words, this statement claims that changing X (and only X!) would result in a change of Y.

WHAT ARE CAUSAL RELATIONS?

Causal relations concern statements about cause and effect.

A causal statement claims that Y happened (or became more likely), because of X.

This implies a comparison of Y under alternative states of X.

Namely, Y would be different, had X been x instead of x' (read: not x).

In other words, this statement claims that changing X (and only X!) would result in a change of Y.

If this is the case, we say that X has a causal effect on Y.

WHAT ARE CAUSAL RELATIONS?

Causal relations concern statements about cause and effect.

A causal statement claims that Y happened (or became more likely), because of X.

This implies a comparison of Y under alternative states of X.

Namely, Y would be different, had X been x instead of x' (read: not x).

In other words, this statement claims that changing X (and only X!) would result in a change of Y.

If this is the case, we say that X has a causal effect on Y.

Put simply, learning about causal relations then means to answer the question whether X really has a causal effect on Y.

WHAT ARE CAUSAL RELATIONS?

Causal relations concern statements about cause and effect.

A causal statement claims that Y happened (or became more likely), because of X.

This implies a comparison of Y under alternative states of X.

Namely, Y would be different, had X been x instead of x' (read: not x).

In other words, this statement claims that changing X (and only X!) would result in a change of Y.

If this is the case, we say that X has a causal effect on Y.

Put simply, learning about causal relations then means to answer the question whether X really has a causal effect on Y.

Importantly, claiming that X has a causal effect on Y doesn't mean that there are no other causes of Y, only that X is one of its (potentially countless number of) causes.

EXAMPLES FOR CAUSAL STATEMENTS ON THE UNIT LEVEL

Causal statements can be formulated on the level of units of observations*, for example

* Units of observation in social research are mostly individuals, but can also be families, schools, regions, countries etc.

EXAMPLES FOR CAUSAL STATEMENTS ON THE UNIT LEVEL

Causal statements can be formulated on the level of units of observations*, for example

I received a good grade (Y), because I regularly attended class (X).

* Units of observation in social research are mostly individuals, but can also be families, schools, regions, countries etc.

EXAMPLES FOR CAUSAL STATEMENTS ON THE UNIT LEVEL

Causal statements can be formulated on the level of units of observations*, for example

I received a good grade (Y), because I regularly attended class (X).

Donald Trump won the election (Y), because he ran a negative campaign (X) against Hillary Clinton.

* Units of observation in social research are mostly individuals, but can also be families, schools, regions, countries etc.

EXAMPLES FOR CAUSAL STATEMENTS ON THE UNIT LEVEL

Causal statements can be formulated on the level of units of observations*, for example

I received a good grade (Y), because I regularly attended class (X).

Donald Trump won the election (Y), because he ran a negative campaign (X) against Hillary Clinton.

My neighborhood is safe (Y), because there are a lot of trees (X).

* Units of observation in social research are mostly individuals, but can also be families, schools, regions, countries etc.

EXAMPLES FOR CAUSAL STATEMENTS ON THE UNIT LEVEL

Causal statements can be formulated on the level of units of observations*, for example

I received a good grade (Y), because I regularly attended class (X).

Donald Trump won the election (Y), because he ran a negative campaign (X) against Hillary Clinton.

My neighborhood is safe (Y), because there are a lot of trees (X).

In Sweden, female employment rates (Y) are high, because of its high level of public childcare (X).

* Units of observation in social research are mostly individuals, but can also be families, schools, regions, countries etc.

EXAMPLES FOR CAUSAL STATEMENTS ON THE UNIT LEVEL

Causal statements can be formulated on the level of units of observations*, for example

I received a good grade (Y), because I regularly attended class (X).

Donald Trump won the election (Y), because he ran a negative campaign (X) against Hillary Clinton.

My neighborhood is safe (Y), because there are a lot of trees (X).

In Sweden, female employment rates (Y) are high, because of its high level of public childcare (X).

These statements or claims imply specific contrasts (in Y) between what happened or what is and what could have happened or could have been instead (with regard to X).

* Units of observation in social research are mostly individuals, but can also be families, schools, regions, countries etc.

EXAMPLES FOR CAUSAL STATEMENTS ON THE UNIT LEVEL

Causal statements can be formulated on the level of units of observations*, for example

I received a good grade (Y), because I regularly attended class (X).
Implies that I would have a worse grade, had I not attended class regularly.

Donald Trump won the election (Y), because he ran a negative campaign (X) against Hillary Clinton.

My neighborhood is safe (Y), because there are a lot of trees (X).

In Sweden, female employment rates (Y) are high, because of its high level of public childcare (X).

These statements or claims imply specific contrasts (in Y) between what happened or what is and what could have happened or could have been instead (with regard to X).

* Units of observation in social research are mostly individuals, but can also be families, schools, regions, countries etc.

EXAMPLES FOR CAUSAL STATEMENTS ON THE UNIT LEVEL

Causal statements can be formulated on the level of units of observations*, for example

I received a good grade (Y), because I regularly attended class (X).
Implies that I would have a worse grade, had I not attended class regularly.

Donald Trump won the election (Y), because he ran a negative campaign (X) against Hillary Clinton.
Implies that Donald Trump would not have won the election, had he not used a negative campaign.

My neighborhood is safe (Y), because there are a lot of trees (X).

In Sweden, female employment rates (Y) are high, because of its high level of public childcare (X).

These statements or claims imply specific contrasts (in Y) between what happened or what is and what could have happened or could have been instead (with regard to X).

* Units of observation in social research are mostly individuals, but can also be families, schools, regions, countries etc.

EXAMPLES FOR CAUSAL STATEMENTS ON THE UNIT LEVEL

Causal statements can be formulated on the level of units of observations*, for example

I received a good grade (Y), because I regularly attended class (X).
Implies that I would have a worse grade, had I not attended class regularly.

Donald Trump won the election (Y), because he ran a negative campaign (X) against Hillary Clinton.
Implies that Donald Trump would not have won the election, had he not used a negative campaign.

My neighborhood is safe (Y), because there are a lot of trees (X).
Implies that my neighborhood would be less safe, were there fewer trees.

In Sweden, female employment rates (Y) are high, because of its high level of public childcare (X).

These statements or claims imply specific contrasts (in Y) between what happened or what is and what could have happened or could have been instead (with regard to X).

* Units of observation in social research are mostly individuals, but can also be families, schools, regions, countries etc.

EXAMPLES FOR CAUSAL STATEMENTS ON THE UNIT LEVEL

Causal statements can be formulated on the level of units of observations*, for example

I received a good grade (Y), because I regularly attended class (X).
Implies that I would have a worse grade, had I not attended class regularly.

Donald Trump won the election (Y), because he ran a negative campaign (X) against Hillary Clinton.
Implies that Donald Trump would not have won the election, had he not used a negative campaign.

My neighborhood is safe (Y), because there are a lot of trees (X).
Implies that my neighborhood would be less safe, were there fewer trees.

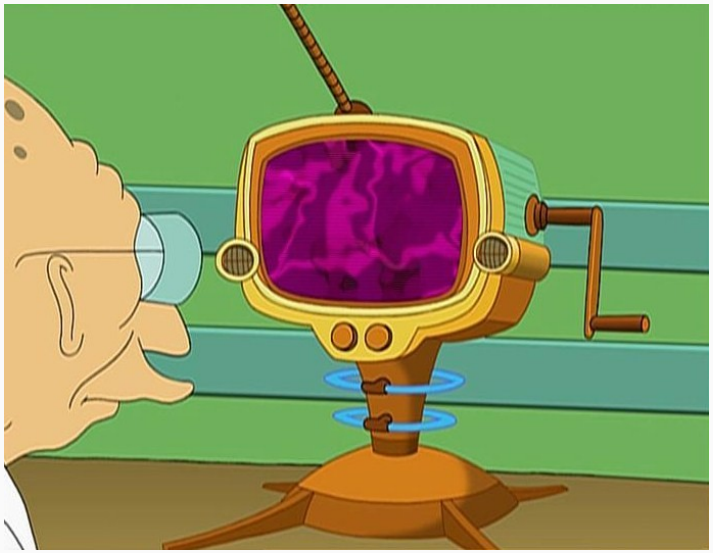
In Sweden, female employment rates (Y) are high, because of its high level of public childcare (X).
Implies that Sweden would have a lower female employment rate, if there was less public childcare.

These statements or claims imply specific contrasts (in Y) between what happened or what is and what could have happened or could have been instead (with regard to X).

* Units of observation in social research are mostly individuals, but can also be families, schools, regions, countries etc.

CAUSAL INFERENCE TRIES TO BUILD THE “WHAT-IF-MACHINE”

Source:
`http://futurama.wikia.com/wiki/What-If_Machine`



The What-If Machine was an invention created by Professor Hubert J. Farnsworth. The device resembled a television set, and could predict the outcome of any “What if?” phrased question, though it can only do this three times a year. The machine is made of gold and the method of activating is never the same.

“Alright, Professor! Lets do it. Make that machine show me what would happen if I was a little more impulsive. Just a little... Not too much.”

—Leela

“What if I never fell into that freezer-doodle and came to the future-jiggy?”

—Fry

“I wanna know what would happen if I were human. I mean, being a robot’s great, but we don’t have emotions and sometimes that makes me very sad.”

—Bender

CAUSAL STATEMENTS ON THE POPULATION-LEVEL

In (quantitative) empirical social research, we usually don't (or cannot) study causal relations on the unit-level.

CAUSAL STATEMENTS ON THE POPULATION-LEVEL

In (quantitative) empirical social research, we usually don't (or cannot) study causal relations on the unit-level.

Instead, we're interested in causal effects on the population-level, where population refers to the collection of units that is the target of investigation.

CAUSAL STATEMENTS ON THE POPULATION-LEVEL

In (quantitative) empirical social research, we usually don't (or cannot) study causal relations on the unit-level.

Instead, we're interested in causal effects on the population-level, where population refers to the collection of units that is the target of investigation.

So instead of studying...

CAUSAL STATEMENTS ON THE POPULATION-LEVEL

In (quantitative) empirical social research, we usually don't (or cannot) study causal relations on the unit-level.

Instead, we're interested in causal effects on the population-level, where population refers to the collection of units that is the target of investigation.

So instead of studying...

- ...ourselves, we're interested in all students at UoC (or in Germany or in...)

CAUSAL STATEMENTS ON THE POPULATION-LEVEL

In (quantitative) empirical social research, we usually don't (or cannot) study causal relations on the unit-level.

Instead, we're interested in causal effects on the population-level, where population refers to the collection of units that is the target of investigation.

So instead of studying...

- ...ourselves, we're interested in all students at UoC (or in Germany or in...)
- ...Donald Trump, we're interested in all US presidential candidates (or possibly all political candidates on earth...)

CAUSAL STATEMENTS ON THE POPULATION-LEVEL

In (quantitative) empirical social research, we usually don't (or cannot) study causal relations on the unit-level.

Instead, we're interested in causal effects on the population-level, where population refers to the collection of units that is the target of investigation.

So instead of studying...

- ...ourselves, we're interested in all students at UoC (or in Germany or in...)
- ...Donald Trump, we're interested in all US presidential candidates (or possibly all political candidates on earth...)
- ...our neighborhood, we're interested in all neighborhoods in NRW (or in Germany or in...)

CAUSAL STATEMENTS ON THE POPULATION-LEVEL

In (quantitative) empirical social research, we usually don't (or cannot) study causal relations on the unit-level.

Instead, we're interested in causal effects on the population-level, where population refers to the collection of units that is the target of investigation.

So instead of studying...

- ...ourselves, we're interested in all students at UoC (or in Germany or in...)
- ...Donald Trump, we're interested in all US presidential candidates (or possibly all political candidates on earth...)
- ...our neighborhood, we're interested in all neighborhoods in NRW (or in Germany or in...)
- ...Sweden, we're interested in all European countries (or all OECD countries or...)

CAUSAL STATEMENTS ON THE POPULATION-LEVEL

In (quantitative) empirical social research, we usually don't (or cannot) study causal relations on the unit-level.

Instead, we're interested in causal effects on the population-level, where population refers to the collection of units that is the target of investigation.

So instead of studying...

- ...ourselves, we're interested in all students at UoC (or in Germany or in...)
- ...Donald Trump, we're interested in all US presidential candidates (or possibly all political candidates on earth...)
- ...our neighborhood, we're interested in all neighborhoods in NRW (or in Germany or in...)
- ...Sweden, we're interested in all European countries (or all OECD countries or...)

Being precise about the population of interest is crucial in empirical social research, no less so in causal inference.

EXAMPLES FOR CAUSAL STATEMENTS ON THE POPULATION LEVEL

Causal statements on the population-level turn attention from specific units to a collection of units of observation, for example

EXAMPLES FOR CAUSAL STATEMENTS ON THE POPULATION LEVEL

Causal statements on the population-level turn attention from specific units to a collection of units of observation, for example

I received a good grade (Y), because I regularly attended class (X).

EXAMPLES FOR CAUSAL STATEMENTS ON THE POPULATION LEVEL

Causal statements on the population-level turn attention from specific units to a collection of units of observation, for example

The students received higher grades (Y), because they regularly attended class (X).

EXAMPLES FOR CAUSAL STATEMENTS ON THE POPULATION LEVEL

Causal statements on the population-level turn attention from specific units to a collection of units of observation, for example

The students received higher grades (Y), because they regularly attended class (X).
Implies that the students would have worse grades, had they not attended class regularly.

EXAMPLES FOR CAUSAL STATEMENTS ON THE POPULATION LEVEL

Causal statements on the population-level turn attention from specific units to a collection of units of observation, for example

The students received higher grades (Y), because they regularly attended class (X).
Implies that the students would have worse grades, had they not attended class regularly.

Donald Trump won the election (Y), because he ran a negative campaign (X) against Hillary Clinton.

EXAMPLES FOR CAUSAL STATEMENTS ON THE POPULATION LEVEL

Causal statements on the population-level turn attention from specific units to a collection of units of observation, for example

The students received higher grades (Y), because they regularly attended class (X).
Implies that the students would have worse grades, had they not attended class regularly.

The political candidates won their elections (Y), because they ran negative campaigns (X).

EXAMPLES FOR CAUSAL STATEMENTS ON THE POPULATION LEVEL

Causal statements on the population-level turn attention from specific units to a collection of units of observation, for example

The students received higher grades (Y), because they regularly attended class (X).
Implies that the students would have worse grades, had they not attended class regularly.

The political candidates won their elections (Y), because they ran negative campaigns (X).
Implies that the candidates would not have won their elections, had they not used negative campaigns.

EXAMPLES FOR CAUSAL STATEMENTS ON THE POPULATION LEVEL

Causal statements on the population-level turn attention from specific units to a collection of units of observation, for example

The students received higher grades (Y), because they regularly attended class (X).
Implies that the students would have worse grades, had they not attended class regularly.

The political candidates won their elections (Y), because they ran negative campaigns (X).
Implies that the candidates would not have won their elections, had they not used negative campaigns.

My neighborhood is safe (Y), because there are a lot of trees (X).

EXAMPLES FOR CAUSAL STATEMENTS ON THE POPULATION LEVEL

Causal statements on the population-level turn attention from specific units to a collection of units of observation, for example

The students received higher grades (Y), because they regularly attended class (X).
Implies that the students would have worse grades, had they not attended class regularly.

The political candidates won their elections (Y), because they ran negative campaigns (X).
Implies that the candidates would not have won their elections, had they not used negative campaigns.

The neighborhoods are safe (Y), because there are a lot of trees (X).

EXAMPLES FOR CAUSAL STATEMENTS ON THE POPULATION LEVEL

Causal statements on the population-level turn attention from specific units to a collection of units of observation, for example

The students received higher grades (Y), because they regularly attended class (X).
Implies that the students would have worse grades, had they not attended class regularly.

The political candidates won their elections (Y), because they ran negative campaigns (X).
Implies that the candidates would not have won their elections, had they not used negative campaigns.

The neighborhoods are safe (Y), because there are a lot of trees (X).
Implies that the neighborhoods would be less safe, were there fewer trees.

EXAMPLES FOR CAUSAL STATEMENTS ON THE POPULATION LEVEL

Causal statements on the population-level turn attention from specific units to a collection of units of observation, for example

The students received higher grades (Y), because they regularly attended class (X).
Implies that the students would have worse grades, had they not attended class regularly.

The political candidates won their elections (Y), because they ran negative campaigns (X).
Implies that the candidates would not have won their elections, had they not used negative campaigns.

The neighborhoods are safe (Y), because there are a lot of trees (X).
Implies that the neighborhoods would be less safe, were there fewer trees.

In Sweden, female employment rates (Y) are high, because of its high public childcare level (X).

EXAMPLES FOR CAUSAL STATEMENTS ON THE POPULATION LEVEL

Causal statements on the population-level turn attention from specific units to a collection of units of observation, for example

The students received higher grades (Y), because they regularly attended class (X).
Implies that the students would have worse grades, had they not attended class regularly.

The political candidates won their elections (Y), because they ran negative campaigns (X).
Implies that the candidates would not have won their elections, had they not used negative campaigns.

The neighborhoods are safe (Y), because there are a lot of trees (X).
Implies that the neighborhoods would be less safe, were there fewer trees.

The countries have high female employment rates (Y), because of high public childcare levels (X).

EXAMPLES FOR CAUSAL STATEMENTS ON THE POPULATION LEVEL

Causal statements on the population-level turn attention from specific units to a collection of units of observation, for example

The students received higher grades (Y), because they regularly attended class (X).
Implies that the students would have worse grades, had they not attended class regularly.

The political candidates won their elections (Y), because they ran negative campaigns (X).
Implies that the candidates would not have won their elections, had they not used negative campaigns.

The neighborhoods are safe (Y), because there are a lot of trees (X).
Implies that the neighborhoods would be less safe, were there fewer trees.

The countries have high female employment rates (Y), because of high public childcare levels (X).
Implies that the countries would have lower female employment rates, if there was less public childcare.

FROM CAUSAL STATEMENTS TO CAUSAL QUESTIONS

If we're not sure causal statements are true, we need to ask causal questions (which we then try to answer using causal inference), for example

FROM CAUSAL STATEMENTS TO CAUSAL QUESTIONS

If we're not sure causal statements are true, we need to ask causal questions (which we then try to answer using causal inference), for example

The students received higher grades (Y), because they regularly attended class (X).

FROM CAUSAL STATEMENTS TO CAUSAL QUESTIONS

If we're not sure causal statements are true, we need to ask causal questions (which we then try to answer using causal inference), for example

Are grades (Y) affected by class attendance (X)?

FROM CAUSAL STATEMENTS TO CAUSAL QUESTIONS

If we're not sure causal statements are true, we need to ask causal questions (which we then try to answer using causal inference), for example

Are grades (Y) affected by class attendance (X)?

The political candidates won their elections (Y), because they ran negative campaigns (X).

FROM CAUSAL STATEMENTS TO CAUSAL QUESTIONS

If we're not sure causal statements are true, we need to ask causal questions (which we then try to answer using causal inference), for example

Are grades (Y) affected by class attendance (X)?

Is the likelihood of winning an election (Y) affected by running negative campaigns (X)?

FROM CAUSAL STATEMENTS TO CAUSAL QUESTIONS

If we're not sure causal statements are true, we need to ask causal questions (which we then try to answer using causal inference), for example

Are grades (Y) affected by class attendance (X)?

Is the likelihood of winning an election (Y) affected by running negative campaigns (X)?

The neighborhoods are safe (Y), because there are a lot of trees (X).

FROM CAUSAL STATEMENTS TO CAUSAL QUESTIONS

If we're not sure causal statements are true, we need to ask causal questions (which we then try to answer using causal inference), for example

Are grades (Y) affected by class attendance (X)?

Is the likelihood of winning an election (Y) affected by running negative campaigns (X)?

Is neighborhood safety (Y) affected by the number of trees (X)?

FROM CAUSAL STATEMENTS TO CAUSAL QUESTIONS

If we're not sure causal statements are true, we need to ask causal questions (which we then try to answer using causal inference), for example

Are grades (Y) affected by class attendance (X)?

Is the likelihood of winning an election (Y) affected by running negative campaigns (X)?

Is neighborhood safety (Y) affected by the number of trees (X)?

The countries have high female employment rates (Y), because of high public childcare levels (X).

FROM CAUSAL STATEMENTS TO CAUSAL QUESTIONS

If we're not sure causal statements are true, we need to ask causal questions (which we then try to answer using causal inference), for example

Are grades (Y) affected by class attendance (X)?

Is the likelihood of winning an election (Y) affected by running negative campaigns (X)?

Is neighborhood safety (Y) affected by the number of trees (X)?

Are the countries' female employment rates (Y) affected by the level of public childcare (X)?

FROM CAUSAL STATEMENTS TO CAUSAL QUESTIONS

If we're not sure causal statements are true, we need to ask causal questions (which we then try to answer using causal inference), for example

Are grades (Y) affected by class attendance (X)?

Is the likelihood of winning an election (Y) affected by running negative campaigns (X)?

Is neighborhood safety (Y) affected by the number of trees (X)?

Are the countries' female employment rates (Y) affected by the level of public childcare (X)?

To answer these questions with certainty we would need to compare Y in the same population under different states of X (and only X).

FROM CAUSAL STATEMENTS TO CAUSAL QUESTIONS

If we're not sure causal statements are true, we need to ask causal questions (which we then try to answer using causal inference), for example

Are grades (Y) affected by class attendance (X)?
Implies comparison of average grade points in same students under different attendance rates.

Is the likelihood of winning an election (Y) affected by running negative campaigns (X)?

Is neighborhood safety (Y) affected by the number of trees (X)?

Are the countries' female employment rates (Y) affected by the level of public childcare (X)?

To answer these questions with certainty we would need to compare Y in the same population under different states of X (and only X).

FROM CAUSAL STATEMENTS TO CAUSAL QUESTIONS

If we're not sure causal statements are true, we need to ask causal questions (which we then try to answer using causal inference), for example

Are grades (Y) affected by class attendance (X)?
Implies comparison of average grade points in same students under different attendance rates.

Is the likelihood of winning an election (Y) affected by running negative campaigns (X)?
Implies comparison of likelihood of winning elections after same candidates have used vs. not used negative campaigns.

Is neighborhood safety (Y) affected by the number of trees (X)?

Are the countries' female employment rates (Y) affected by the level of public childcare (X)?

To answer these questions with certainty we would need to compare Y in the same population under different states of X (and only X).

FROM CAUSAL STATEMENTS TO CAUSAL QUESTIONS

If we're not sure causal statements are true, we need to ask causal questions (which we then try to answer using causal inference), for example

Are grades (Y) affected by class attendance (X)?
Implies comparison of average grade points in same students under different attendance rates.

Is the likelihood of winning an election (Y) affected by running negative campaigns (X)?
Implies comparison of likelihood of winning elections after same candidates have used vs. not used negative campaigns.

Is neighborhood safety (Y) affected by the number of trees (X)?
Implies comparison of safety in same neighborhoods under different number of trees.

Are the countries' female employment rates (Y) affected by the level of public childcare (X)?

To answer these questions with certainty we would need to compare Y in the same population under different states of X (and only X).

FROM CAUSAL STATEMENTS TO CAUSAL QUESTIONS

If we're not sure causal statements are true, we need to ask causal questions (which we then try to answer using causal inference), for example

Are grades (Y) affected by class attendance (X)?
Implies comparison of average grade points in same students under different attendance rates.

Is the likelihood of winning an election (Y) affected by running negative campaigns (X)?
Implies comparison of likelihood of winning elections after same candidates have used vs. not used negative campaigns.

Is neighborhood safety (Y) affected by the number of trees (X)?
Implies comparison of safety in same neighborhoods under different number of trees.

Are the countries' female employment rates (Y) affected by the level of public childcare (X)?
Implies comparison of female employment rates in same countries under different levels of public childcare.

To answer these questions with certainty we would need to compare Y in the same population under different states of X (and only X).

CAUSAL INFERENCE WITHIN EMPIRICAL SOCIAL RESEARCH

OTHER AREAS OF EMPIRICAL SOCIAL RESEARCH

Of course, causal inference is not the only area or goal of empirical social research.

OTHER AREAS OF EMPIRICAL SOCIAL RESEARCH

Of course, causal inference is not the only area or goal of empirical social research.

There are at least two other important goals for which researchers study empirical data.

OTHER AREAS OF EMPIRICAL SOCIAL RESEARCH

Of course, causal inference is not the only area or goal of empirical social research.

There are at least two other important goals for which researchers study empirical data.

1. Description: involves using data to answer questions about the occurrence or prevalence of phenomena and statistical relations between phenomena

OTHER AREAS OF EMPIRICAL SOCIAL RESEARCH

Of course, causal inference is not the only area or goal of empirical social research.

There are at least two other important goals for which researchers study empirical data.

1. Description: involves using data to answer questions about the occurrence or prevalence of phenomena and statistical relations between phenomena
2. Statistical prediction: involves using statistical relations found in empirical data to predict occurence or prevalence of phenomena elsewhere or at a different time

EXAMPLES FOR DESCRIPTIVE RESEARCH QUESTIONS

About the prevalence of phenomena, e.g.,

EXAMPLES FOR DESCRIPTIVE RESEARCH QUESTIONS

About the prevalence of phenomena, e.g.,

- What is the percentage of children living in poverty in Germany?

EXAMPLES FOR DESCRIPTIVE RESEARCH QUESTIONS

About the prevalence of phenomena, e.g.,

- What is the percentage of children living in poverty in Germany?
- What is the unemployment rate in NRW?

EXAMPLES FOR DESCRIPTIVE RESEARCH QUESTIONS

About the prevalence of phenomena, e.g.,

- What is the percentage of children living in poverty in Germany?
- What is the unemployment rate in NRW?
- What is the approval rate of Donald Trump among US voters?

EXAMPLES FOR DESCRIPTIVE RESEARCH QUESTIONS

About the prevalence of phenomena, e.g.,

- What is the percentage of children living in poverty in Germany?
- What is the unemployment rate in NRW?
- What is the approval rate of Donald Trump among US voters?

About statistical relations, e.g.,

EXAMPLES FOR DESCRIPTIVE RESEARCH QUESTIONS

About the prevalence of phenomena, e.g.,

- What is the percentage of children living in poverty in Germany?
- What is the unemployment rate in NRW?
- What is the approval rate of Donald Trump among US voters?

About statistical relations, e.g.,

- How did the birth rate in Europe develop over time?

EXAMPLES FOR DESCRIPTIVE RESEARCH QUESTIONS

About the prevalence of phenomena, e.g.,

- What is the percentage of children living in poverty in Germany?
- What is the unemployment rate in NRW?
- What is the approval rate of Donald Trump among US voters?

About statistical relations, e.g.,

- How did the birth rate in Europe develop over time?
- Do university graduates have a higher life satisfaction than nongraduates?

EXAMPLES FOR DESCRIPTIVE RESEARCH QUESTIONS

About the prevalence of phenomena, e.g.,

- What is the percentage of children living in poverty in Germany?
- What is the unemployment rate in NRW?
- What is the approval rate of Donald Trump among US voters?

About statistical relations, e.g.,

- How did the birth rate in Europe develop over time?
- Do university graduates have a higher life satisfaction than nongraduates?
- Is the rate of mental illness in a country higher, the higher the extent of income inequality?

RESEARCH QUESTIONS INVOLVING STATISTICAL PREDICTION

- If people bought a specific product, which other product are they likely to buy?

RESEARCH QUESTIONS INVOLVING STATISTICAL PREDICTION

- If people bought a specific product, which other product are they likely to buy?
- What will the birth rate in Germany be in 2050?

RESEARCH QUESTIONS INVOLVING STATISTICAL PREDICTION

- If people bought a specific product, which other product are they likely to buy?
- What will the birth rate in Germany be in 2050?
- Which people are likely to engage in criminal behavior?

RESEARCH QUESTIONS INVOLVING STATISTICAL PREDICTION

- If people bought a specific product, which other product are they likely to buy?
- What will the birth rate in Germany be in 2050?
- Which people are likely to engage in criminal behavior?
- How many people are likely to die of the flue in the coming winter?

SOME QUESTIONS FOR YOU

1. What's your name?
2. Where did you attain your BA?
3. What was the topic of your BA thesis?
4. Would you say the main question of your thesis was a causal question, a descriptive question, a question involving prediction, or something else?

CAUSAL QUESTIONS VS. STATISTICAL QUESTIONS

In order to draw valid conclusions from empirical social research we must distinguish sharply between causal questions and statistical questions, for example

CAUSAL QUESTIONS VS. STATISTICAL QUESTIONS

In order to draw valid conclusions from empirical social research we must distinguish sharply between causal questions and statistical questions, for example

Are grades (Y) increased by class attendance (X)?

CAUSAL QUESTIONS VS. STATISTICAL QUESTIONS

In order to draw valid conclusions from empirical social research we must distinguish sharply between causal questions and statistical questions, for example

Are grades (Y) higher for students who attended class more often (X)?

CAUSAL QUESTIONS VS. STATISTICAL QUESTIONS

In order to draw valid conclusions from empirical social research we must distinguish sharply between causal questions and statistical questions, for example

Are grades (Y) higher for students who attended class more often (X)?

Is the likelihood of winning an election (Y) increased by running negative campaigns (X)?

CAUSAL QUESTIONS VS. STATISTICAL QUESTIONS

In order to draw valid conclusions from empirical social research we must distinguish sharply between causal questions and statistical questions, for example

Are grades (Y) higher for students who attended class more often (X)?

Is the likelihood of winning an election (Y) higher for candidates who ran a negative campaign (X)?

CAUSAL QUESTIONS VS. STATISTICAL QUESTIONS

In order to draw valid conclusions from empirical social research we must distinguish sharply between causal questions and statistical questions, for example

Are grades (Y) higher for students who attended class more often (X)?

Is the likelihood of winning an election (Y) higher for candidates who ran a negative campaign (X)?

Is neighborhood safety (Y) increased by the number of trees (X)?

CAUSAL QUESTIONS VS. STATISTICAL QUESTIONS

In order to draw valid conclusions from empirical social research we must distinguish sharply between causal questions and statistical questions, for example

Are grades (Y) higher for students who attended class more often (X)?

Is the likelihood of winning an election (Y) higher for candidates who ran a negative campaign (X)?

Is safety (Y) higher in neighborhoods with more trees (X)?

CAUSAL QUESTIONS VS. STATISTICAL QUESTIONS

In order to draw valid conclusions from empirical social research we must distinguish sharply between causal questions and statistical questions, for example

Are grades (Y) higher for students who attended class more often (X)?

Is the likelihood of winning an election (Y) higher for candidates who ran a negative campaign (X)?

Is safety (Y) higher in neighborhoods with more trees (X)?

Is a country's female employment rate (Y) increased by the level of public childcare (X)?

CAUSAL QUESTIONS VS. STATISTICAL QUESTIONS

In order to draw valid conclusions from empirical social research we must distinguish sharply between causal questions and statistical questions, for example

Are grades (Y) higher for students who attended class more often (X)?

Is the likelihood of winning an election (Y) higher for candidates who ran a negative campaign (X)?

Is safety (Y) higher in neighborhoods with more trees (X)?

Are female employment rates (Y) higher in countries with more public childcare (X)?

CAUSAL QUESTIONS VS. STATISTICAL QUESTIONS

In order to draw valid conclusions from empirical social research we must distinguish sharply between causal questions and statistical questions, for example

Are grades (Y) higher for students who attended class more often (X)?

Is the likelihood of winning an election (Y) higher for candidates who ran a negative campaign (X)?

Is safety (Y) higher in neighborhoods with more trees (X)?

Are female employment rates (Y) higher in countries with more public childcare (X)?

To answer these questions we compare Y in different subpopulations with different values of X.

CAUSAL QUESTIONS VS. STATISTICAL QUESTIONS

In order to draw valid conclusions from empirical social research we must distinguish sharply between causal questions and statistical questions, for example

Are grades (Y) higher for students who attended class more often (X)?
Implies comparison of average grades between students with different attendance rates.

Is the likelihood of winning an election (Y) higher for candidates who ran a negative campaign (X)?

Is safety (Y) higher in neighborhoods with more trees (X)?

Are female employment rates (Y) higher in countries with more public childcare (X)?

To answer these questions we compare Y in different subpopulations with different values of X.

CAUSAL QUESTIONS VS. STATISTICAL QUESTIONS

In order to draw valid conclusions from empirical social research we must distinguish sharply between causal questions and statistical questions, for example

Are grades (Y) higher for students who attended class more often (X)?
Implies comparison of average grades between students with different attendance rates.

Is the likelihood of winning an election (Y) higher for candidates who ran a negative campaign (X)?
Implies comparison of likelihood of winning elections between candidates who ran negative campaigns and those who did not.

Is safety (Y) higher in neighborhoods with more trees (X)?

Are female employment rates (Y) higher in countries with more public childcare (X)?

To answer these questions we compare Y in different subpopulations with different values of X.

CAUSAL QUESTIONS VS. STATISTICAL QUESTIONS

In order to draw valid conclusions from empirical social research we must distinguish sharply between causal questions and statistical questions, for example

Are grades (Y) higher for students who attended class more often (X)?
Implies comparison of average grades between students with different attendance rates.

Is the likelihood of winning an election (Y) higher for candidates who ran a negative campaign (X)?
Implies comparison of likelihood of winning elections between candidates who ran negative campaigns and those who did not.

Is safety (Y) higher in neighborhoods with more trees (X)?
Implies comparison of safety between neighborhoods with different number of trees.

Are female employment rates (Y) higher in countries with more public childcare (X)?

To answer these questions we compare Y in different subpopulations with different values of X.

CAUSAL QUESTIONS VS. STATISTICAL QUESTIONS

In order to draw valid conclusions from empirical social research we must distinguish sharply between causal questions and statistical questions, for example

Are grades (Y) higher for students who attended class more often (X)?
Implies comparison of average grades between students with different attendance rates.

Is the likelihood of winning an election (Y) higher for candidates who ran a negative campaign (X)?
Implies comparison of likelihood of winning elections between candidates who ran negative campaigns and those who did not.

Is safety (Y) higher in neighborhoods with more trees (X)?
Implies comparison of safety between neighborhoods with different number of trees.

Are female employment rates (Y) higher in countries with more public childcare (X)?
Implies comparison of female employment rates between countries with different rates of public childcare.

To answer these questions we compare Y in different subpopulations with different values of X.

WHY IS THIS DIFFERENCE IMPORTANT?

By answering statistical questions we gain knowledge about statistical relations, i.e., associations/correlations.

(see Pearl, 2009; Shalizi, 2016, p. 506)

WHY IS THIS DIFFERENCE IMPORTANT?

By answering statistical questions we gain knowledge about statistical relations, i.e., associations/correlations.

With statistical relations we study and exploit factual distributions (What is?).

(see Pearl, 2009; Shalizi, 2016, p. 506)

WHY IS THIS DIFFERENCE IMPORTANT?

By answering statistical questions we gain knowledge about statistical relations, i.e., associations/correlations.

With statistical relations we study and exploit factual distributions (What is?).

This involves passive observation (*seeing*).

(see Pearl, 2009; Shalizi, 2016, p. 506)

WHY IS THIS DIFFERENCE IMPORTANT?

By answering statistical questions we gain knowledge about statistical relations, i.e., associations/correlations.

With statistical relations we study and exploit factual distributions (What is?).

This involves passive observation (*seeing*).

To investigate statistical relations we only need data on the phenomena of interest.

(see Pearl, 2009; Shalizi, 2016, p. 506)

WHY IS THIS DIFFERENCE IMPORTANT?

By answering statistical questions we gain knowledge about statistical relations, i.e., associations/correlations.

With statistical relations we study and exploit factual distributions (What is?).

This involves passive observation (*seeing*).

To investigate statistical relations we only need data on the phenomena of interest.

By answering causal questions we gain knowledge about causal relations, i.e., (causal) effects.

(see Pearl, 2009; Shalizi, 2016, p. 506)

WHY IS THIS DIFFERENCE IMPORTANT?

By answering statistical questions we gain knowledge about statistical relations, i.e., associations/correlations.

With statistical relations we study and exploit factual distributions (What is?).

This involves passive observation (*seeing*).

To investigate statistical relations we only need data on the phenomena of interest.

By answering causal questions we gain knowledge about causal relations, i.e., (causal) effects.

With causal relations we study counterfactual distributions (What if?).

(see Pearl, 2009; Shalizi, 2016, p. 506)

WHY IS THIS DIFFERENCE IMPORTANT?

By answering statistical questions we gain knowledge about statistical relations, i.e., associations/correlations.

With statistical relations we study and exploit factual distributions (What is?).

This involves passive observation (*seeing*).

To investigate statistical relations we only need data on the phenomena of interest.

By answering causal questions we gain knowledge about causal relations, i.e., (causal) effects.

With causal relations we study counterfactual distributions (What if?).

This involves (at least hypothetical) manipulation/change of the phenomena under study (*doing*).

(see Pearl, 2009; Shalizi, 2016, p. 506)

WHY IS THIS DIFFERENCE IMPORTANT?

By answering statistical questions we gain knowledge about statistical relations, i.e., associations/correlations.

With statistical relations we study and exploit factual distributions (What is?).

This involves passive observation (*seeing*).

To investigate statistical relations we only need data on the phenomena of interest.

By answering causal questions we gain knowledge about causal relations, i.e., (causal) effects.

With causal relations we study counterfactual distributions (What if?).

This involves (at least hypothetical) manipulation/change of the phenomena under study (*doing*).

To investigate causal relations we need data and also knowledge/theory/assumptions about the processes that generated the data.

(see Pearl, 2009; Shalizi, 2016, p. 506)

WHY IS THIS DIFFERENCE IMPORTANT?

By answering statistical questions we gain knowledge about statistical relations, i.e., associations/correlations.

With statistical relations we study and exploit factual distributions (What is?).

This involves passive observation (*seeing*).

To investigate statistical relations we only need data on the phenomena of interest.

By answering causal questions we gain knowledge about causal relations, i.e., (causal) effects.

With causal relations we study counterfactual distributions (What if?).

This involves (at least hypothetical) manipulation/change of the phenomena under study (*doing*).

To investigate causal relations we need data and also knowledge/theory/assumptions about the processes that generated the data.

Therefore, causal inference is to be sharply distinguished from other (important!) tasks of empirical social research like description, exploration, and statistical/prediction forecasting.

(see Pearl, 2009; Shalizi, 2016, p. 506)

AN OBVIOUS EXAMPLE

When we ask whether the number of trees affects neighborhood safety, we want to know whether changing the number of trees would result in a change in neighborhood safety.

AN OBVIOUS EXAMPLE

When we ask whether the number of trees affects neighborhood safety, we want to know whether changing the number of trees would result in a change in neighborhood safety.

Suppose that data show that neighborhoods with more trees are safer?

AN OBVIOUS EXAMPLE

When we ask whether the number of trees affects neighborhood safety, we want to know whether changing the number of trees would result in a change in neighborhood safety.

Suppose that data show that neighborhoods with more trees are safer?

Does that mean there is a causal effect?

AN OBVIOUS EXAMPLE

When we ask whether the number of trees affects neighborhood safety, we want to know whether changing the number of trees would result in a change in neighborhood safety.

Suppose that data show that neighborhoods with more trees are safer?

Does that mean there is a causal effect?

No! It may be that richer neighborhoods can afford more green spaces and also are safer.

AN OBVIOUS EXAMPLE

When we ask whether the number of trees affects neighborhood safety, we want to know whether changing the number of trees would result in a change in neighborhood safety.

Suppose that data show that neighborhoods with more trees are safer?

Does that mean there is a causal effect?

No! It may be that richer neighborhoods can afford more green spaces and also are safer.

Changing the number of trees would then not affect safety.

AN OBVIOUS EXAMPLE

When we ask whether the number of trees affects neighborhood safety, we want to know whether changing the number of trees would result in a change in neighborhood safety.

Suppose that data show that neighborhoods with more trees are safer?

Does that mean there is a causal effect?

No! It may be that richer neighborhoods can afford more green spaces and also are safer.

Changing the number of trees would then not affect safety.

But we would still observe an association between number of trees and safety in neighborhoods.

A LESS OBVIOUS EXAMPLE

When we ask whether class attendance increases students' grades, we want to know whether changing class attendance would result in a change in grades.

A LESS OBVIOUS EXAMPLE

When we ask whether class attendance increases students' grades, we want to know whether changing class attendance would result in a change in grades.

Suppose that data show that students who attended class more regularly have better grades.

A LESS OBVIOUS EXAMPLE

When we ask whether class attendance increases students' grades, we want to know whether changing class attendance would result in a change in grades.

Suppose that data show that students who attended class more regularly have better grades.

Does that mean there is a causal effect?

A LESS OBVIOUS EXAMPLE

When we ask whether class attendance increases students' grades, we want to know whether changing class attendance would result in a change in grades.

Suppose that data show that students who attended class more regularly have better grades.

Does that mean there is a causal effect?

No! It may be that more motivated students attend more often and also have better grades.

A LESS OBVIOUS EXAMPLE

When we ask whether class attendance increases students' grades, we want to know whether changing class attendance would result in a change in grades.

Suppose that data show that students who attended class more regularly have better grades.

Does that mean there is a causal effect?

No! It may be that more motivated students attend more often and also have better grades.

Changing class attendance would then not affect grades.

A LESS OBVIOUS EXAMPLE

When we ask whether class attendance increases students' grades, we want to know whether changing class attendance would result in a change in grades.

Suppose that data show that students who attended class more regularly have better grades.

Does that mean there is a causal effect?

No! It may be that more motivated students attend more often and also have better grades.

Changing class attendance would then not affect grades.

But we would still observe an association between class attendance and grades.

WHY IS CAUSAL INFERENCE USEFUL?

Knowledge about statistical relations is not sufficient whenever we're interested in the consequences of phenomena and actions and, by extension, in changing specific aspects of the world (hopefully for the better).

WHY IS CAUSAL INFERENCE USEFUL?

Knowledge about statistical relations is not sufficient whenever we're interested in the consequences of phenomena and actions and, by extension, in changing specific aspects of the world (hopefully for the better).

Therefore, causal inference is of interest

WHY IS CAUSAL INFERENCE USEFUL?

Knowledge about statistical relations is not sufficient whenever we're interested in the consequences of phenomena and actions and, by extension, in changing specific aspects of the world (hopefully for the better).

Therefore, causal inference is of interest

- in many (but by far not all) questions in (social) science,

WHY IS CAUSAL INFERENCE USEFUL?

Knowledge about statistical relations is not sufficient whenever we're interested in the consequences of phenomena and actions and, by extension, in changing specific aspects of the world (hopefully for the better).

Therefore, causal inference is of interest

- in many (but by far not all) questions in (social) science,
- for evaluating the effectiveness of past actions,

WHY IS CAUSAL INFERENCE USEFUL?

Knowledge about statistical relations is not sufficient whenever we're interested in the consequences of phenomena and actions and, by extension, in changing specific aspects of the world (hopefully for the better).

Therefore, causal inference is of interest

- in many (but by far not all) questions in (social) science,
- for evaluating the effectiveness of past actions,
- in order to devise effective new actions.

WHY IS CAUSAL INFERENCE USEFUL?

Knowledge about statistical relations is not sufficient whenever we're interested in the consequences of phenomena and actions and, by extension, in changing specific aspects of the world (hopefully for the better).

Therefore, causal inference is of interest

- in many (but by far not all) questions in (social) science,
- for evaluating the effectiveness of past actions,
- in order to devise effective new actions.

This is also relevant in jobs outside academia or government.

WHY IS CAUSAL INFERENCE USEFUL?

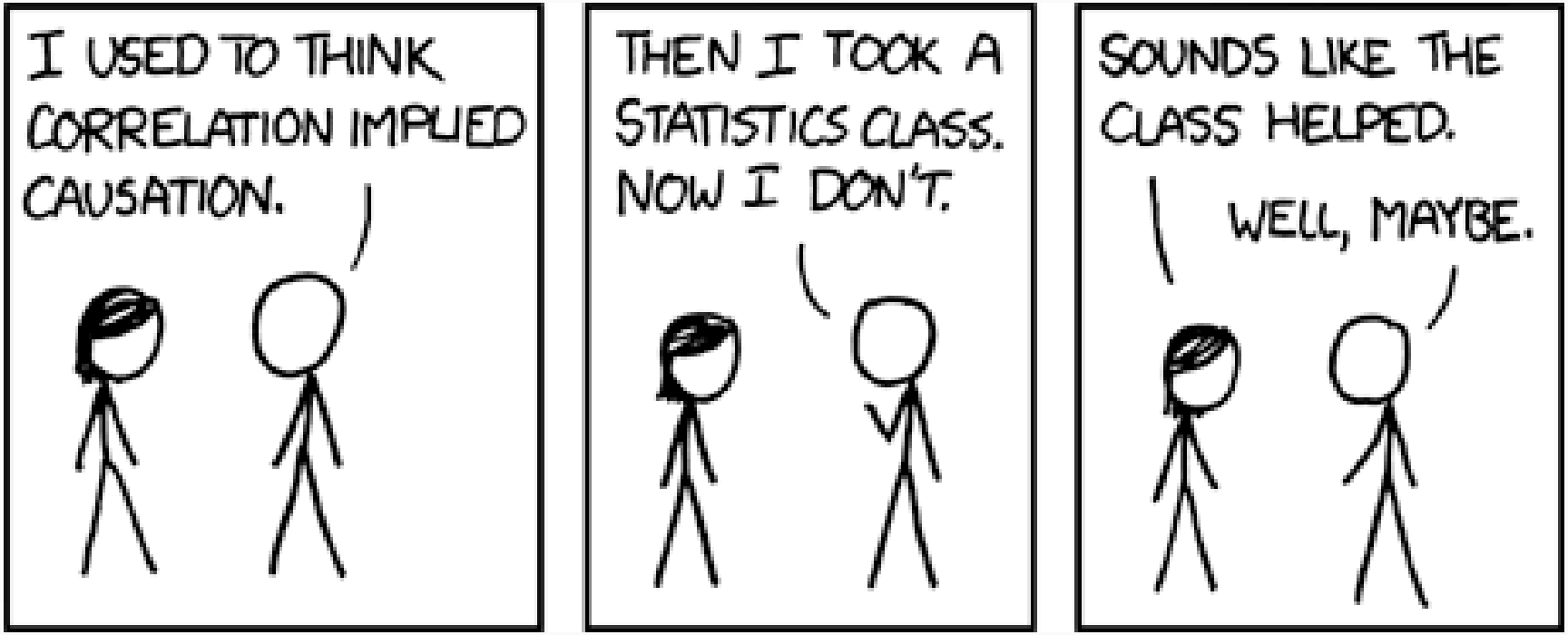
Knowledge about statistical relations is not sufficient whenever we're interested in the consequences of phenomena and actions and, by extension, in changing specific aspects of the world (hopefully for the better).

Therefore, causal inference is of interest

- in many (but by far not all) questions in (social) science,
- for evaluating the effectiveness of past actions,
- in order to devise effective new actions.

This is also relevant in jobs outside academia or government.

Can you think of some questions or problems for which we need knowledge about causal relations?



Source: `http://imgs.xkcd.com/comics/correlation.png`

Saying that correlation doesn't imply causation is true, but what now?

The course goes beyond this mantra and aims to equip you with tools to decide when causal claims in the literature, the media, or in everyday life are believable and when they aren't.

A ROADMAP FOR CAUSAL INFERENCE

A ROADMAP FOR CAUSAL INFERENCE?

To rigorously study causal relations it is helpful to follow specific steps.

Petersen and van der Laan (2014) proposed what they call a roadmap for causal inference.

The whole course is organized along this roadmap.

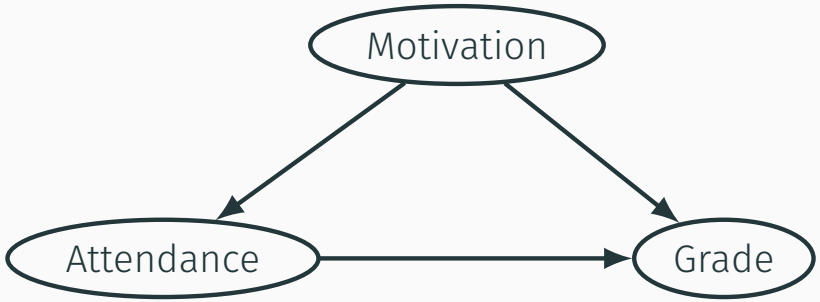
FIRST STEP

Specify a causal model (or possibly multiple alternative models)

- Causal models express knowledge, theory, and assumptions about the processes that generated the data that we can observe.
- They are models of how the world works (i.e., how we think it works).
- Causal models are not based on statistical relations in empirical data.
- But they imply statistical relations that we should be able to observe in empirical data (if the causal model is true).

(Petersen and van der Laan, 2014)

A SIMPLE EXAMPLE



This causal model expresses my belief that

- motivation may cause both class attendance and grades,
- class attendance may cause grades,
- there are no other common causes of motivation, class attendance, and grades.

SECOND STEP

Define the causal question of interest

- Which causal relation(s) in the model is(are) of interest?
- What is the specific change whose consequences are to be studied?
- What is the target population?

(Petersen and van der Laan, 2014)

THIRD STEP

Link the causal model to the available empirical data

- Which variables of the causal model are measured in the data and which aren't?
- What happened during data collection?

(Petersen and van der Laan, 2014)

FOURTH STEP

Assess whether the causal relation of interest can be identified with the available data

- Does the data include the variables necessary to learn about the causal relation of interest?
- If so, which statistical relation can be used to (approximately) estimate the causal relation?
- If not, which additional information (or assumptions) are necessary to identify the causal relation of interest?

(Petersen and van der Laan, 2014)

FURTHER STEPS

Only now, statistical methods (e.g., regression) come in!

5. Specify the statistical model used to estimate the statistical parameter
6. Estimate the statistical parameter (for example, using *Stata*)
7. Interpret the results
   Important: for causal interpretation of results (from any statistical method) causal model must be (approximately) true!

(Petersen and van der Laan, 2014)

A ROADMAP FOR CAUSAL INFERENCE!

1. Specify the causal model
2. Define the causal parameter of interest (along with the target population)
3. Link the causal model to the available empirical data
4. Assess whether the causal parameter of interest can be identified with the available data and define the respective statistical parameter
5. Specify the statistical model used to estimate the statistical parameter
6. Estimate the statistical parameter
7. Interpret the results and discuss assumptions

(Petersen and van der Laan, 2014)

COURSE GOALS AND LEARNING OUTCOMES

LEARNING GOALS AND OUTCOMES

1. Formalize a research question about causal relations using causal graphs and counterfactuals. This includes
   - ⋆ translating theoretical arguments about causal relations into a corresponding graphical model,
   - ⋆ specifying the causal relation of interest and defining this relation in terms of counterfactual contrasts,
   - ⋆ being able to conceptually distinguish this causal relation from statistical association.

→ Roadmap 1 & 2

LEARNING GOALS AND OUTCOMES

2. Use graphical models to devise strategies for identifying the causal relation of interest. For this, students
   - ⋆ demonstrate that they are capable to derive empirical implications from a graphical causal model,
   - ⋆ understand the theoretical assumptions necessary to test these implications,
   - ⋆ critically evaluate whether these assumptions hold in applied social research.

→ Roadmap 3 & 4

LEARNING GOALS AND OUTCOMES

3. Estimate the causal relation of interest and, if feasible, test underlying assumptions. To do so, students
   - ⋆ adapt existing *Stata* code for their purposes,
   - ⋆ correctly interpret the resulting estimates,
   - ⋆ understand and perform tests of the validity of the analyses.

→ Roadmap 5–7

COURSE STRUCTURE AND CONTENT

After a refresher on statistical relations (next week) and an introduction to graphical causal models (in two weeks) we will follow the roadmap for causal inference to study

1. total effects (Oct 30–Nov 27)
2. direct and indirect effects (Dec 4–Dec 18)
3. cumulative effects (Jan 8–Jan 22).

REQUIREMENTS AND GRADING

FINAL GRADE

1. 60-min exam on January 29, 2018 (60 points), based on required readings and class contents (not the lecture slides alone!)
2. max. of 30 bonus points by completing three labs (10 points each), in which you implement the roadmap for causal inference to study (1) total effects (2) direct and indirect effects (3) cumulative effects using empirical data

!DON'T FORGET TO REGISTER FOR THE EXAM ON TIME!

GRADING

Points from labs are added to exam points until you reach max. 60 points

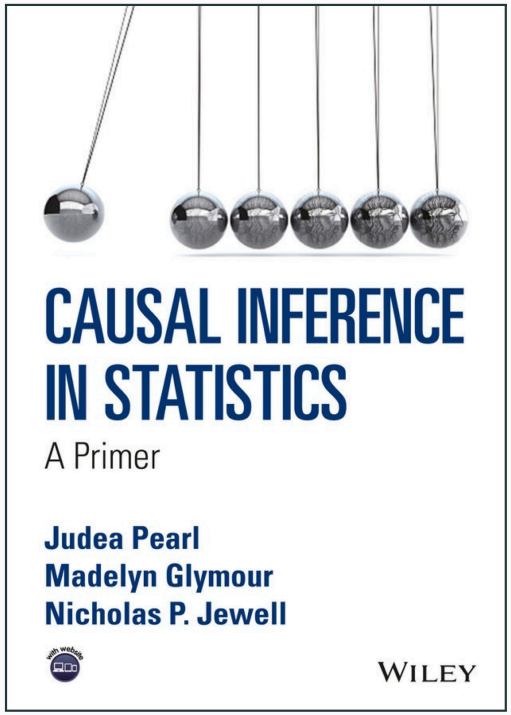
Grades are awared as follows:

1.0: $60 \geq \text{points} \leq 58$
1.3: $58 > \text{points} \leq 55$
1.7: $55 > \text{points} \leq 51$
2.0: $51 > \text{points} \leq 48$
2.3: $48 > \text{points} \leq 45$
2.7: $45 > \text{points} \leq 42$
3.0: $42 > \text{points} \leq 39$
3.3: $39 > \text{points} \leq 36$
3.7: $36 > \text{points} \leq 33$
4.0: $33 > \text{points} \leq 30$
n.p.: $30 > \text{points} \leq 0$

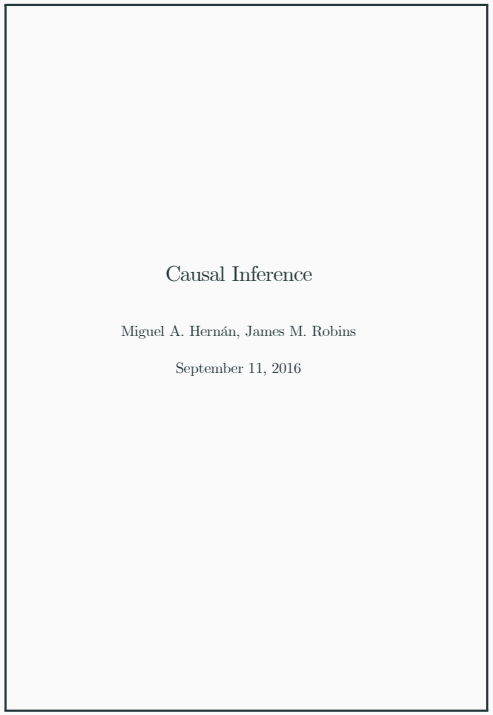
COURSE MATERIALS

TEXTBOOKS AND REQUIRED READINGS

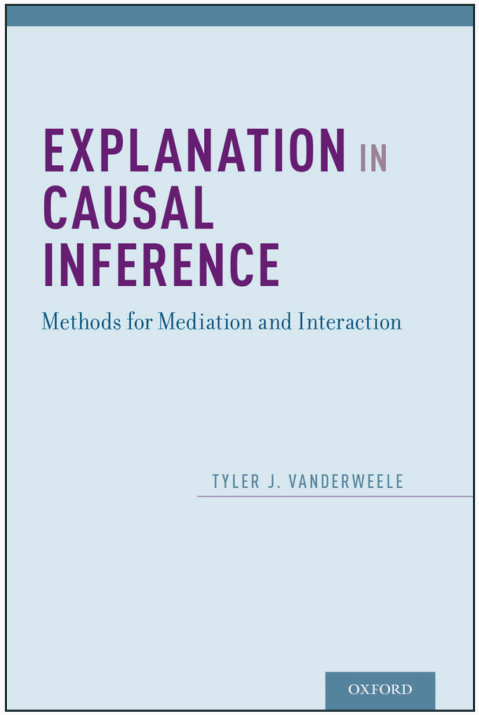
PGJ



HR



VDW



Please use PGJ with errata!

Other required readings (and some further readings) on *Ilias*

ILIAS

- course materials (readings, slides, data etc.)
- submission of completed labs
- short quizzes
- forum for questions and discussion
  Please use this forum for all questions regarding course organization and content that are possibly of general interest to course members and use email only for personal matters such as office hour appointments, illness, and questions regarding grading.

NEXT WEEK: STATISTICAL RELATIONS AND THEIR INTERPRETATION

1. Variables and distributions
2. Conditioning and conditional distributions
3. Statistical dependence and independence
4. Statistical models
5. Simpson's Paradox and the data generating process

THANK YOU FOR YOUR ATTENTION!

REFERENCES

Heckman, J. J. (2000). Causal parameters and policy analysis in economics: A twentieth century retrospective. *Quarterly Journal of Economics*, 115, 45–97.

Pearl, J. (2009). Causal inference in statistics: An overview. *Statistics Surveys*, 3, 96–146.

Petersen, M. L. and van der Laan, M. J. (2014). Causal models and learning from data: Integrating causal modeling and statistical estimation. *Epidemiology*, 25(3), 418–426.

Shalizi, C. R. (2016). *Advanced Data Analysis from an Elementary Point of View*. Cambridge University Press, New York. URL `http://www.stat.cmu.edu/~cshalizi/ADAfaEPoV/ADAfaEPoV.pdf`.

Get these slides from

`https://osf.io/jgrsy/`

The METROPOLIS theme was created by Matthias Vogelgesang and is available from

`https://github.com/matze/mtheme`