Causal Inference

By: Miguel A. Hernán and James M. Robins

Part I: Causal inference without models

30th April, 2014

Outline

Chapter 1.1: Individual causal effects

"The purpose of this chapter is to introduce mathematical notation that formalizes the causal intuition that you already possess."

Some notation

- Dichotomous treatment variable: A (1: treated; 0: untreated)
- Dichotomous outcome variable: Y (1: death; 0: survival)
- $Y^{a=i}$: Outcome under treatment $a = i$, $i \in \{0, 1\}$.

Chapter 1.1: Individual causal effects

"The purpose of this chapter is to introduce mathematical notation that formalizes the causal intuition that you already possess."

Some notation

- Dichotomous treatment variable: A (1: treated; 0: untreated)
- Dichotomous outcome variable: Y (1: death; 0: survival)
- $Y^{a=i}$: Outcome under treatment $a = i$, $i \in \{0, 1\}$.

Definition

Causal effect for an individual: Treatment A has a causal effect if

$$Y^{a=1} \neq Y^{a=0}.$$

Chapter 1.1: Individual causal effects

"The purpose of this chapter is to introduce mathematical notation that formalizes the causal intuition that you already possess."

Some notation

- Dichotomous treatment variable: A (1: treated; 0: untreated)
- Dichotomous outcome variable: Y (1: death; 0: survival)
- $Y^{a=i}$: Outcome under treatment $a = i$, $i \in \{0, 1\}$.

Definition

Causal effect for an individual: Treatment A has a causal effect if

$$Y^{a=1} \neq Y^{a=0}.$$

However, in general, individual effects **cannot** be identified.

Chapter 1.2: Average causal effects

Definition

Average causal effect is present if

$$\Pr(Y^{a=1} = 1) \neq \Pr(Y^{a=0} = 1).$$

Chapter 1.2: Average causal effects

Definition

Average causal effect is present if

$$\Pr(Y^{a=1} = 1) \neq \Pr(Y^{a=0} = 1).$$

More generally (nondichotomous outcomes):

$$E(Y^{a=1}) \neq E(Y^{a=0}).$$

Chapter 1.2: Average causal effects

Definition

Average causal effect is present if

$$\Pr(Y^{a=1} = 1) \neq \Pr(Y^{a=0} = 1).$$

More generally (nondichotomous outcomes):

$$E(Y^{a=1}) \neq E(Y^{a=0}).$$

What we would like to observe:

$$\Pr(Y^{a=1} = 1) - \Pr(Y^{a=0} = 1) \qquad \text{(Causal risk difference)}$$

$$\frac{\Pr(Y^{a=1} = 1)}{\Pr(Y^{a=0} = 1)} \qquad \text{(Causal risk ratio)}$$

$$\frac{\Pr(Y^{a=1} = 1)/\Pr(Y^{a=1} = 0)}{\Pr(Y^{a=0} = 1)/\Pr(Y^{a=0} = 0)} \qquad \text{(Causal odds ratio)}$$

Chapter 1.2: Average causal effects

Definition

Average causal effect is present if

$$\Pr(Y^{a=1} = 1) \neq \Pr(Y^{a=0} = 1).$$

More generally (nondichotomous outcomes):

$$E(Y^{a=1}) \neq E(Y^{a=0}).$$

What we can estimate:

$$\Pr(Y = 1|A = 1) - \Pr(Y = 1|A = 0)$$ (Associational risk difference)

$$\frac{\Pr(Y = 1|A = 1)}{\Pr(Y = 1|A = 0)}$$ (Associational risk ratio)

$$\frac{\Pr(Y = 1|A = 1)/\Pr(Y = 0|A = 1)}{\Pr(Y = 1|A = 0)/\Pr(Y = 0|A = 0)}$$ (Associational odds ratio)

Chapter 1.5: Causation versus association

$\Pr(Y = 1|A = 1)$ is a conditional, $\Pr(Y^a = 1)$ an unconditional probability.

Chapter 1.5: Causation versus association

$\Pr(Y = 1|A = 1)$ is a conditional, $\Pr(Y^a = 1)$ an unconditional probability.

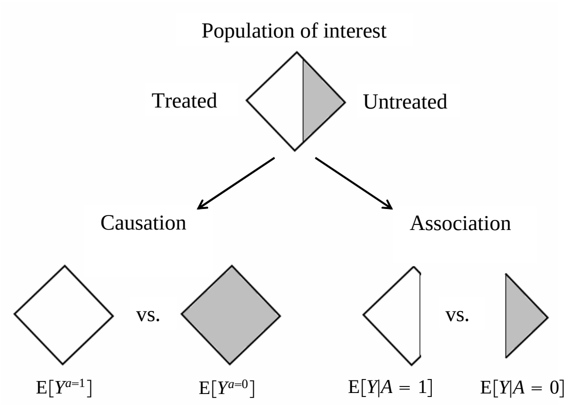

Figure : Association-causation difference (Figure 1.1 in the book)

Chapter 2: Randomized experiments

"This chapter describes why randomization results in convincing causal inferences."

Chapter 2: Randomized experiments

"This chapter describes why randomization results in convincing causal inferences."

Exchangeability

- Means that the outcome would be the same in both study groups if both received the treatment or if both did not receive it.

Chapter 2: Randomized experiments

"This chapter describes why randomization results in convincing causal inferences."

Exchangeability

- Means that the outcome would be the same in both study groups if both received the treatment or if both did not receive it.
- Formally: Exchangeability, $Y^a \coprod A$ for $a \in \{0, 1\}$, holds if

$$\Pr(Y^{a=0} = 1) = \underbrace{\Pr(Y^{a=0} = 1|A = 0)}_{\text{Observable}} = \underbrace{\Pr(Y^{a=0} = 1|A = 1)}_{\text{Counterfactual}},$$

$$\Pr(Y^{a=1} = 1) = \underbrace{\Pr(Y^{a=1} = 1|A = 0)}_{\text{Counterfactual}} = \underbrace{\Pr(Y^{a=1} = 1|A = 1)}_{\text{Observable}}.$$

Chapter 2: Randomized experiments

- Randomization is expected to produce exchangeability.

- Randomization is expected to produce exchangeability.
- Hence, in **ideal** randomized experiments, **association is causation**.

Chapter 2: Randomized experiments

- Randomization is expected to produce exchangeability.
- Hence, in **ideal** randomized experiments, **association is causation**.
- **Conditional exchangeability:** $Y^a \coprod A|L$.
  Present if exchangeability holds within the levels of variable L.

Chapter 2: Randomized experiments

- Randomization is expected to produce exchangeability.
- Hence, in **ideal** randomized experiments, **association is causation**.
- **Conditional exchangeability:** $Y^a \coprod A | L$.
  Present if exchangeability holds within the levels of variable L.
- How can the CRR be computed in a conditionally randomized experiment?
  $\rightsquigarrow$ **Standardization** or **inverse probability weighting**!

Chapter 2: Randomized experiments

- Randomization is expected to produce exchangeability.
- Hence, in **ideal** randomized experiments, **association is causation**.
- **Conditional exchangeability:** $Y^a \coprod A|L$.
  Present if exchangeability holds within the levels of variable L.
- How can the CRR be computed in a conditionally randomized experiment?
  $\rightsquigarrow$ **Standardization** or **inverse probability weighting**!
- Standardization

$$\text{CRR} = \frac{\Pr(Y^{a=1}=1)}{\Pr(Y^{a=0}=1)} = \frac{\sum_l \Pr(Y=1|L=l, A=1)\Pr(L=l)}{\sum_l \Pr(Y=1|L=l, A=0)\Pr(L=l)}$$

Hernán & Robins: Causal Inference.

Chapter 3: Observational Studies

“Looking up at the sky”: Version II

“Does one’s looking up at the sky make other pedestrians look up too?”

"Looking up at the sky": Version II

"Does one's looking up at the sky make other pedestrians look up too?"

An **observational study**:

- Find pedestrian that is not looking up.

“Looking up at the sky”: Version II

“Does one’s looking up at the sky make other pedestrians look up too?”

An **observational study**:

- Find pedestrian that is not looking up.
- Identify 2nd pedestrian walking towards him/her.

"Looking up at the sky": Version II

"Does one's looking up at the sky make other pedestrians look up too?"

An **observational study**:

- Find pedestrian that is not looking up.
- Identify 2nd pedestrian walking towards him/her.
- Record behavior in next 10 seconds.

“Looking up at the sky”: Version II

“Does one’s looking up at the sky make other pedestrians look up too?”

An **observational study**:

- Find pedestrian that is not looking up.
- Identify 2nd pedestrian walking towards him/her.
- Record behavior in next 10 seconds.
- Compare $\widehat{\Pr}(2^{nd}\text{ l. up}|1^{st}\text{ l. up})$ with $\widehat{\Pr}(2^{nd}\text{ l. up}|1^{st}\text{ l. not up})$.

“Looking up at the sky”: Version II

“Does one’s looking up at the sky make other pedestrians look up too?”

An **observational study**:

- Find pedestrian that is not looking up.
- Identify 2nd pedestrian walking towards him/her.
- Record behavior in next 10 seconds.
- Compare $\widehat{\Pr}(2^{nd}\text{ l. up}|1^{st}\text{ l. up})$ with $\widehat{\Pr}(2^{nd}\text{ l. up}|1^{st}\text{ l. not up})$.

“This chapter reviews some conditions under which observational studies lead to valid causal inferences.”

3.1 The randomized experiment paradigm

Identifiability conditions for causal inference

Three conditions must hold so that an observational study can be conceptualized as a conditionally randomized experiment:

1. The values of treatment under comparison correspond to well-defined interventions.

3.1 The randomized experiment paradigm

Identifiability conditions for causal inference

Three conditions must hold so that an observational study can be conceptualized as a conditionally randomized experiment:

1. The values of treatment under comparison correspond to well-defined interventions.
2. The conditional probability of receiving every value of treatment, though not decided by the investigators, depends only on the measured covariates.

3.1 The randomized experiment paradigm

Identifiability conditions for causal inference

Three conditions must hold so that an observational study can be conceptualized as a conditionally randomized experiment:

1. The values of treatment under comparison correspond to well-defined interventions.
2. The conditional probability of receiving every value of treatment, though not decided by the investigators, depends only on the measured covariates.
3. The conditional probability of receiving every value of treatment is greater than zero, i.e., positive.

3.1 The randomized experiment paradigm

Identifiability conditions for causal inference

Three conditions must hold so that an observational study can be conceptualized as a conditionally randomized experiment:

1. The values of treatment under comparison correspond to well-defined interventions.
2. The conditional probability of receiving every value of treatment, though not decided by the investigators, depends only on the measured covariates.
3. The conditional probability of receiving every value of treatment is greater than zero, i.e., positive.

If these three (identifiability) conditions hold,

> *"...causal effects can be identified from observational studies by using IP weighting or standardization."*

3.1 The randomized experiment paradigm

Comments:

3.1 The randomized experiment paradigm

Comments:

- Other possible approach to causal inference: "…hoping that a predictor of treatment, referred to as an **instrumental variable**, was randomly assigned conditional on the measured covariates."

3.1 The randomized experiment paradigm

Comments:

- Other possible approach to causal inference: "…hoping that a predictor of treatment, referred to as an **instrumental variable**, was randomly assigned conditional on the measured covariates."
- What we should do: Describe carefully

3.1 The randomized experiment paradigm

Comments:

- Other possible approach to causal inference: "...hoping that a predictor of treatment, referred to as an **instrumental variable**, was randomly assigned conditional on the measured covariates."
- What we should do: Describe carefully
  (i) "the randomized experiment that we would like to, but cannot, conduct."

3.1 The randomized experiment paradigm

Comments:

- Other possible approach to causal inference: "...hoping that a predictor of treatment, referred to as an **instrumental variable**, was randomly assigned conditional on the measured covariates."
- What we should do: Describe carefully
  (i) "the randomized experiment that we would like to, but cannot, conduct."
  (ii) "how the observational study emulates that randomized experiment."

3.1 The randomized experiment paradigm

Comments:

- Other possible approach to causal inference: "...hoping that a predictor of treatment, referred to as an **instrumental variable**, was randomly assigned conditional on the measured covariates."
- What we should do: Describe carefully
  - (I) "the randomized experiment that we would like to, but cannot, conduct."
  - (II) "how the observational study emulates that randomized experiment."
- "In ideal randomized experiments, the data contain sufficient information to identify causal effects. In contrast, (...), the information contained in observational data is insufficient to identify causal effects."

3.1 The randomized experiment paradigm

Comments:

- Other possible approach to causal inference: "…hoping that a predictor of treatment, referred to as an **instrumental variable**, was randomly assigned conditional on the measured covariates."
- What we should do: Describe carefully
  - (i) "the randomized experiment that we would like to, but cannot, conduct."
  - (ii) "how the observational study emulates that randomized experiment."
- "In ideal randomized experiments, the data contain sufficient information to identify causal effects. In contrast, (…), the information contained in observational data is insufficient to identify causal effects."
- Two sources of information are required: data and identifiability assumptions.

3.2 Exchangeability

The "real world" example with a 3rd variable

	L	A	Y		L	A	Y
Rheia	0	0	0	Leto	1	0	0
Kronos	0	0	1	Ares	1	1	1
Demeter	0	0	0	Athena	1	1	1
Hades	0	0	0	Hephaestus	1	1	1
Hestia	0	1	0	Aphrodite	1	1	1
Poseidon	0	1	0	Cyclope	1	1	1
Hera	0	1	0	Persephone	1	1	1
Zeus	0	1	1	Hermes	1	1	0
Artemis	1	0	1	Hebe	1	1	0
Apollo	1	0	1	Dionysus	1	1	0

L is supposed to be a prognosis factor (1, critical situation; 0, otherwise).

3.2 Exchangeability

- If L is the only outcome predictor with unequal distribution in $A = 0$ and $A = 1$, then $Y^a \coprod A | L$ holds.
  $\rightsquigarrow$ Standardization or inverse probability weighting to estimate CRR.

3.2 Exchangeability

- If L is the only outcome predictor with unequal distribution in $A = 0$ and $A = 1$, then $Y^a \coprod A | L$ holds.
  $\rightsquigarrow$ Standardization or inverse probability weighting to estimate CRR.
- **But**: In observational studies, the value of A likely depends on some outcome predictors $(L_1, L_2, \ldots)$.

3.2 Exchangeability

- If L is the only outcome predictor with unequal distribution in $A = 0$ and $A = 1$, then $Y^a \coprod A | L$ holds.
  $\rightsquigarrow$ Standardization or inverse probability weighting to estimate CRR.
- **But**: In observational studies, the value of A likely depends on some outcome predictors $(L_1, L_2, \dots)$.
- **Crucial question**: Are all L_i with unequal distribution among treatment groups observed?

3.2 Exchangeability

- If L is the only outcome predictor with unequal distribution in $A = 0$ and $A = 1$, then $Y^a \coprod A|L$ holds.
  $\rightsquigarrow$ Standardization or inverse probability weighting to estimate CRR.
- **But**: In observational studies, the value of A likely depends on some outcome predictors $(L_1, L_2, \dots)$.
- **Crucial question**: Are all L_i with unequal distribution among treatment groups observed?
- We cannot know the answer to the previous question. There is **no** guarantee that $Y^a \coprod A|L$ holds.

3.2 Exchangeability

- If L is the only outcome predictor with unequal distribution in $A = 0$ and $A = 1$, then $Y^a \coprod A|L$ holds.
  $\rightsquigarrow$ Standardization or inverse probability weighting to estimate CRR.
- **But**: In observational studies, the value of A likely depends on some outcome predictors $(L_1, L_2, \dots)$.
- **Crucial question**: Are all L_i with unequal distribution among treatment groups observed?
- We cannot know the answer to the previous question. There is **no** guarantee that $Y^a \coprod A|L$ holds.
- "Thus when we analyze an observational study under the assumption of conditional exchangeability, **we must hope** that the assumption is at least approximately true."

Fine Point 3.1: Attributable fraction

Measure that compares observed risk with counterfactual risk (under either $a = 0$ or $a = 1$):

$$\frac{\Pr(Y = 1) - \Pr(Y^a = 1)}{\Pr(Y = 1)}.$$

Fine Point 3.1: Attributable fraction

Measure that compares observed risk with counterfactual risk (under either $a = 0$ or $a = 1$):

$$\frac{\Pr(Y = 1) - \Pr(Y^a = 1)}{\Pr(Y = 1)}.$$

Example:

- 10 individuals receive ambrosia ($A = 1$), 10 receive nectar ($A = 0$).

Fine Point 3.1: Attributable fraction

Measure that compares observed risk with counterfactual risk (under either $a = 0$ or $a = 1$):

$$\frac{\Pr(Y = 1) - \Pr(Y^a = 1)}{\Pr(Y = 1)}.$$

Example:

- 10 individuals receive ambrosia ($A = 1$), 10 receive nectar ($A = 0$).
- Next day: $\Pr(\text{sick}|A = 1) = 7/10$, $\Pr(\text{sick}|A = 0) = 1/10$.

Fine Point 3.1: Attributable fraction

Measure that compares observed risk with counterfactual risk (under either $a = 0$ or $a = 1$):

$$\frac{\Pr(Y = 1) - \Pr(Y^a = 1)}{\Pr(Y = 1)}.$$

Example:

- 10 individuals receive ambrosia ($A = 1$), 10 receive nectar ($A = 0$).
- Next day: $\Pr(\text{sick}|A = 1) = 7/10$, $\Pr(\text{sick}|A = 0) = 1/10$.
- Assuming exchangeability:

$$\text{CRR} = 0.7/0.1 = 7, \quad \text{CRD} = 0.7 - 0.1 = 0.6.$$

Fine Point 3.1: Attributable fraction

Measure that compares observed risk with counterfactual risk (under either $a = 0$ or $a = 1$):

$$\frac{\Pr(Y = 1) - \Pr(Y^a = 1)}{\Pr(Y = 1)}.$$

Example:

- 10 individuals receive ambrosia ($A = 1$), 10 receive nectar ($A = 0$).
- Next day: $\Pr(\text{sick}|A = 1) = 7/10$, $\Pr(\text{sick}|A = 0) = 1/10$.
- Assuming exchangeability:

$$\text{CRR} = 0.7/0.1 = 7, \quad \text{CRD} = 0.7 - 0.1 = 0.6.$$

- What fraction of cases is attributable to $A = 1$?

$$\frac{\Pr(Y = 1) - \Pr(Y^{a=0} = 1)}{\Pr(Y = 1)} = (0.4 - 0.1)/0.4 = 0.75.$$

3.3 Positivity

- **Positivity condition**: In an experiment, the CRR can only be estimated if some subjects are assigned to each treatment level.

3.3 Positivity

- **Positivity condition**: In an experiment, the CRR can only be estimated if some subjects are assigned to each treatment level.
- Positivity holds if

$$\Pr(A = a | L = l) > 0, \forall l \text{ with } \Pr(L = l) \neq 0.$$

3.3 Positivity

- **Positivity condition**: In an experiment, the CRR can only be estimated if some subjects are assigned to each treatment level.
- Positivity holds if

$$\Pr(A = a | L = l) > 0, \forall l \text{ with } \Pr(L = l) \neq 0.$$

- If exchangeability is achieved conditional on some variables, then positivity must only hold for these.

3.3 Positivity

- **Positivity condition**: In an experiment, the CRR can only be estimated if some subjects are assigned to each treatment level.
- Positivity holds if

$$\Pr(A = a | L = l) > 0, \forall l \text{ with } \Pr(L = l) \neq 0.$$

- If exchangeability is achieved conditional on some variables, then positivity must only hold for these.
- In observational studies, neither positivity nor exchangeability are guaranteed.

3.3 Positivity

- **Positivity condition**: In an experiment, the CRR can only be estimated if some subjects are assigned to each treatment level.
- Positivity holds if

$$\Pr(A = a|L = l) > 0, \forall l \text{ with } \Pr(L = l) \neq 0.$$

- If exchangeability is achieved conditional on some variables, then positivity must only hold for these.
- In observational studies, neither positivity nor exchangeability are guaranteed.
- Standardization and IP weighted risk are only meaningful if positivity holds.

3.3 Positivity

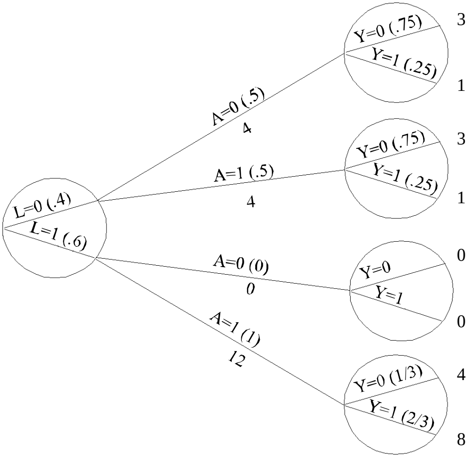

FIGURE : Hernán & Robins: Figure 3.1

3.4 Well-defined interventions

- If multiple versions of a treatment are present, the interventions are not well defined.

3.4 Well-defined interventions

- If multiple versions of a treatment are present, the interventions are not well defined.
- "...treatment-variation irrelevance may be reasonable in ideal randomized studies ..." but in observational studies, the investigators may not have control over the versions of the treatment.

3.4 Well-defined interventions

- If multiple versions of a treatment are present, the interventions are not well defined.
- "...treatment-variation irrelevance may be reasonable in ideal randomized studies ..." but in observational studies, the investigators may not have control over the versions of the treatment.
- "Because treatment-variation irrelevance cannot be taken for granted in observational studies, the interpretation of the causal effect is not always straightforward. At the very least, investigators need to characterize the versions of treatment that operate in the population."

3.4 Well-defined interventions

- If multiple versions of a treatment are present, the interventions are not well defined.
- "...treatment-variation irrelevance may be reasonable in ideal randomized studies ..." but in observational studies, the investigators may not have control over the versions of the treatment.
- "Because treatment-variation irrelevance cannot be taken for granted in observational studies, the interpretation of the causal effect is not always straightforward. At the very least, investigators need to characterize the versions of treatment that operate in the population."
- **Section 3.5:** Well-defined interventions are a pre-requisite for causal inference.

3.4 Well-defined interventions

- If multiple versions of a treatment are present, the interventions are not well defined.
- "…treatment-variation irrelevance may be reasonable in ideal randomized studies …" but in observational studies, the investigators may not have control over the versions of the treatment.
- "Because treatment-variation irrelevance cannot be taken for granted in observational studies, the interpretation of the causal effect is not always straightforward. At the very least, investigators need to characterize the versions of treatment that operate in the population."
- **Section 3.5:** Well-defined interventions are a pre-requisite for causal inference.
- "The problems generated by unspecified interventions cannot be dealt with by applying sophisticated statistical methods."

3.6 Causation or prediction

- "Is everything lost when the observational data cannot be used to emulate an interesting randomized experiment? Not really. Observational data may still be quite useful by focusing on prediction."

3.6 Causation or prediction

- "Is everything lost when the observational data cannot be used to emulate an interesting randomized experiment? Not really. Observational data may still be quite useful by focusing on prediction."
- However, "when causal inference is the ultimate goal, prediction may be unsatisfying."

Fine Point 3.4

Identifiability of causal effects

Fine Point 3.4

Identifiability of causal effects

- "...an average causal effect is (non parametrically) identifiable when the distribution of the observed data is compatible with a **single value** of the effect measure."

Fine Point 3.4

Identifiability of causal effects

- "...an average causal effect is (non parametrically) identifiable when the distribution of the observed data is compatible with a **single value** of the effect measure."
- It is nonidentifiable, when the observed data are compatible with several values of the effect measure.

Fine Point 3.4

Identifiability of causal effects

- "...an average causal effect is (non parametrically) identifiable when the distribution of the observed data is compatible with a **single value** of the effect measure."
- It is nonidentifiable, when the observed data are compatible with several values of the effect measure.
- If the data from the table on slide 12 come from an observational study **without** the assumption of $Y^a \coprod A|L$, then they are consistent with a CRR

Fine Point 3.4

Identifiability of causal effects

- "...an average causal effect is (non parametrically) identifiable when the distribution of the observed data is compatible with a **single value** of the effect measure."
- It is nonidentifiable, when the observed data are compatible with several values of the effect measure.
- If the data from the table on slide 12 come from an observational study **without** the assumption of $Y^a \coprod A|L$, then they are consistent with a CRR
  - < 1 if risk factors other than L are more frequent among $A = 1$,

Fine Point 3.4

Identifiability of causal effects

- "...an average causal effect is (non parametrically) identifiable when the distribution of the observed data is compatible with a **single value** of the effect measure."
- It is nonidentifiable, when the observed data are compatible with several values of the effect measure.
- If the data from the table on slide 12 come from an observational study **without** the assumption of $Y^a \coprod A | L$, then they are consistent with a CRR
  - < 1 if risk factors other than L are more frequent among $A = 1$,
  - > 1 if risk factors other than L are more frequent among $A = 0$,

Fine Point 3.4

Identifiability of causal effects

- "...an average causal effect is (non parametrically) identifiable when the distribution of the observed data is compatible with a **single value** of the effect measure."
- It is nonidentifiable, when the observed data are compatible with several values of the effect measure.
- If the data from the table on slide 12 come from an observational study **without** the assumption of $Y^a \coprod A|L$, then they are consistent with a CRR
  - < 1 if risk factors other than L are more frequent among $A = 1$,
  - > 1 if risk factors other than L are more frequent among $A = 0$,
  - $= 1$ if risk factors other than L are equally distributed between $A = 0$ and $A = 1$.

Fine Point 3.4

Identifiability of causal effects

- "...an average causal effect is (non parametrically) identifiable when the distribution of the observed data is compatible with a **single value** of the effect measure."
- It is nonidentifiable, when the observed data are compatible with several values of the effect measure.
- If the data from the table on slide 12 come from an observational study **without** the assumption of $Y^a \coprod A|L$, then they are consistent with a CRR
  - < 1 if risk factors other than L are more frequent among $A = 1$,
  - > 1 if risk factors other than L are more frequent among $A = 0$,
  - $= 1$ if risk factors other than L are equally distributed between $A = 0$ and $A = 1$.

CONTINUARÁ...