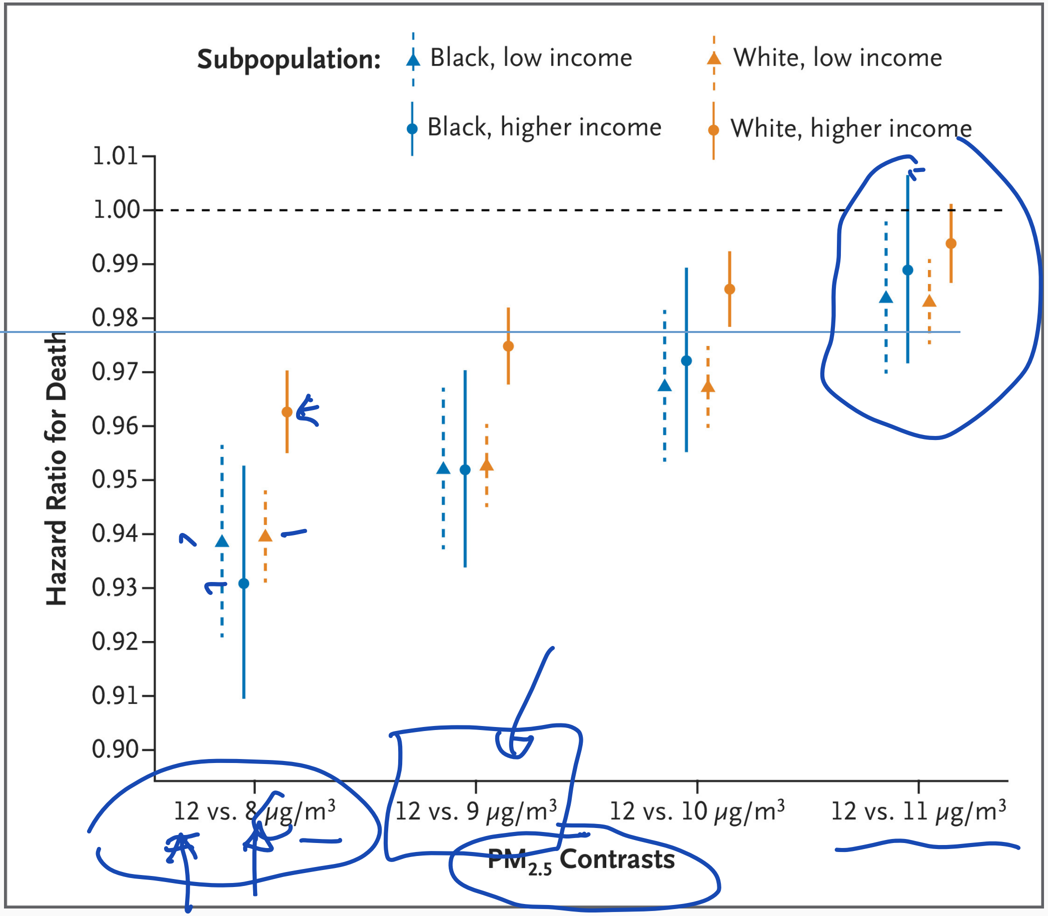
Mathematical Statistics II

STA2212H S LEC9101

Week 10

March 18 2025



SPECIAL ARTICLE

Air Pollution and Mortality at the Intersection of Race and Social Class

Kevin P. Josey, Ph.D., Scott W. Delaney, Sc.D., J.D., Xiao Wu, Ph.D., Rachel C. Nethery, Ph.D., Priyanka DeSouza, Ph.D., Danielle Braun, Ph.D., and Francesca Dominici, Ph.D.

ABSTRACT

BACKGROUND

Black Americans are exposed to higher annual levels of air pollution containing fine particulate matter (particles with an aerodynamic diameter of ≤2.5 μm [$PM_{2.5}$]) than White Americans and may be more susceptible to its health effects. Low-income Americans may also be more susceptible to $PM_{2.5}$ pollution than high-income Americans. Because information is lacking on exposure–response curves for $PM_{2.5}$ exposure and mortality among marginalized subpopulations categorized according to both race and socioeconomic position, the Environmental Protection Agency lacks important evidence to inform its regulatory rulemaking for $PM_{2.5}$ standards.

METHODS

We analyzed 623 million person-years of Medicare data from 73 million persons 65 years of age or older from 2000 through 2016 to estimate associations between annual $PM_{2.5}$ exposure and mortality in subpopulations defined simultaneously by racial identity (Black vs. White) and income level (Medicaid eligible vs. ineligible).

RESULTS

Lower $PM_{2.5}$ exposure was associated with lower mortality in the full population, but marginalized subpopulations appeared to benefit more as $PM_{2.5}$ levels decreased. For example, the hazard ratio associated with decreasing $PM_{2.5}$ from 12 μg per cubic meter to 8 μg per cubic meter for the White higher-income subpopulation was 0.963 (95% confidence interval [CI], 0.955 to 0.970), whereas equivalent hazard

From the Departments of Biostatistics (K.P.J., R.C.N., D.B., F.D.) and Environmental Health (S.W.D.), Harvard T.H. Chan School of Public Health, Boston; the Department of Biostatistics, Mailman School of Public Health, Columbia University, New York (X.W.); and the Department of Urban and Regional Planning, University of Colorado Denver, Denver (P.D.). Dr. Dominici can be reached at fdominic@hsph.harvard.edu or at the Department of Biostatistics, Harvard T.H. Chan School of Public Health, 655 Huntington Ave., Bldg. 2, 4th Flr., Boston, MA 02115.

Drs. Josey and Delaney and Drs. Braun and Dominici contributed equally to this article.

This article was published on March 24, 2023, at NEJM.org.

N Engl J Med 2023;388:1396-404.
DOI: 10.1056/NEJMsa2300523

Today

1. Recap Mar 11 goodness-of-fit tests
2. Introduction to causal inference
3. Project guidelines:

   if you are not sure how to fit your paper into the guidelines contact me

   office hour: Tuesday 3-4; Monday 7-8

   email: nancym.reid@utoronto.ca

Recap: multinomial goodness of fit statistics

$X_1, \ldots, X_n$ iid $F(x; \theta)$

- Pearson's χ^2 test

$(O - E)^2$

$$Q = \sum_{j=1}^{k} \frac{\{Y_j - np_j(\tilde{\theta})\}^2}{np_j(\tilde{\theta})} \xrightarrow{d} \chi^2_{k-1-p}$$

E

$Y_j = \#\{X_i \in A_j\}$

bins $A_1 \ldots A_k$

$\tilde{\theta}$ MLE in multinomial

$n \to \infty$

k fixed

- Likelihood ratio (deviance) test

$P_i(\mu, \sigma^2) = \int_{A_j} \varphi\left(\frac{x-\mu}{\sigma}\right)\frac{1}{\sigma}dx$

$$W = 2\sum_{j=1}^{k} Y_j \log\left(\frac{Y_j}{np_j(\tilde{\theta})}\right) \xrightarrow{d} \chi^2_{k-1-p}$$

e.g. $\Phi(a_j^{\mu,\sigma}) - \Phi(a_{j-1})$

$W = Q + o_p(1)$

$\boxed{\log(1+\varepsilon) \sim \varepsilon - \frac{1}{2}\varepsilon^2 + O(\varepsilon^3)}$

Recap: Smooth goodness-of-fit statistics

$H_0: F = F_0$. $\quad X_1, \dots, X_n$ iid F

$$K_n = \sup_t |\widehat{F}_n(t) - F_0(t)| \xrightarrow{d} K, \qquad \boxed{\mathrm{pr}(K > x) = 2\sum_{j=1}^{\infty}(-1)^{j+1}\exp(-2j^2x^2)}$$

$$W_n^2 = \int \{\widehat{F}_n(t) - F_0(t)\}^2 dF_0(t) \xrightarrow{d} \sum_{j=1}^{\infty} \frac{Z_j^2}{j^2\pi^2}$$

if $H_0: F = F_0(\,;\theta)$

? ?

$$A_n^2 = \int \frac{\{\widehat{F}_n(t) - F_0(t)\}^2}{F_0(t)\{1 - F_0(t)\}} dF_0(t) \xrightarrow{d} \sum_{j=1}^{\infty} \frac{Z_j^2}{j(j+1)}$$

SM : $F_0\left(\frac{x-\mu}{\sigma}\right)$ uses $\widehat{F}_n$ for $(x_i - \hat{\mu})/\hat{\sigma}$

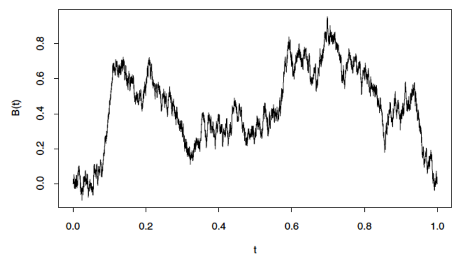


Figure 9.1 *A simulated realization of a Brownian bridge process.*

This just in

Phystat

Topic Introduction

Goodness-of-Fit tests via Machine Learning

GOF tests:

- Assess how well a model H0 describes the data
- No alternative model H1 specified, if it was, likelihood ratio L(data|H1)/L(data|H0) would provide optimal test (Neyman-Pearson)
- Standard GOF tests in HEP: χ^2 (most frequent), Kolgomorov- Smirnov (seldomly), others..
  - Difficulties arise for multi-dimensional distributions

→ Machine Learning offers various possibilities 😊

Todays topic!

link

This just in

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Statistical Methodology
Series B

J. R. Statist. Soc. B (2020)
82, *Part* 3, *pp.* 773–795

Goodness-of-fit testing in high dimensional generalized linear models

Jana Janková and Rajen D. Shah,
University of Cambridge, UK

Peter Bühlmann
Eidgenössische Technische Hochschule Zürich, Switzerland

and Richard J. Samworth
University of Cambridge, UK

[Received August 2019. Revised March 2020]

Summary. We propose a family of tests to assess the goodness of fit of a high dimensional

This just in

2. Methodology: generalized residual prediction tests

As mentioned in Section 1.1, our generalized residual prediction (GRP) testing methodology relies on an initial fit of the form

$$\hat{\beta} := \arg\min_{\beta \in \mathbb{R}^p} \left\{ \frac{1}{n} \sum_{i=1}^{n} \rho(Y_i, x_i^{\mathrm{T}} \beta) + \lambda \|\beta\|_1 \right\}.$$

In what follows, we refer to $\hat{\beta}$ as the GLM lasso, though it is not essential that the loss function $\rho : \mathcal{Y} \times \mathbb{R} \to \mathbb{R}$ is the negative log-likelihood that is obtained from a GLM, and indeed this definition incorporates penalized quasi-likelihood estimators, among others. Our general framework for goodness-of-fit testing will also assume that we have available an auxiliary data set $(X_A, Y_A) \in \mathbb{R}^{n_A \times p} \times \mathcal{Y}^{n_A}$ independent of (X, Y). In the rest of the paper, we take $n_A = n$ for simplicity, although this is not needed for our procedures. Consider the Pearson-type residuals

$$R_i = \frac{Y_i - \mu(x_i^{\mathrm{T}} \hat{\beta})}{\sqrt{V\{\mu(x_i^{\mathrm{T}} \tilde{\beta})\}}}, \qquad i = 1, \ldots, n.$$

Here $\tilde{\beta} \in \mathbb{R}^p$ is an additional estimate of β_0 that may be computed by using the auxiliary data set, or in certain circumstances may be taken as $\hat{\beta}$ itself: we discuss these two cases in the following sections. Given the vector R of residuals, the basic form of our test statistic is $w^{\mathrm{T}} R$; here $w \in \mathbb{R}^n$ is a direction that is typically derived by using the auxiliary data set. We describe in detail the construction of such a w in Section 2.1, where the goal is general goodness-of-fit testing.

A further modification of the method can enable us to use multiple directions w to test simultaneously for different departures from the null or to aggregate over different directions derived by using flexible regression methods with different tuning parameters. Given a set $W \subseteq \mathbb{R}^n$ of direction vectors w, our proposed test statistic then takes the form

$$\sup_{w \in W} w^{\mathrm{T}} R.$$

This just in

The Annals of Statistics
2022, Vol. 50, No. 5, 2514–2544
https://doi.org/10.1214/22-AOS2187

TESTING GOODNESS-OF-FIT AND CONDITIONAL INDEPENDENCE WITH APPROXIMATE CO-SUFFICIENT SAMPLING

BY RINA FOYGEL BARBER[1,a] AND LUCAS JANSON[2,b]

[1]*Department of Statistics, University of Chicago,* [a]*rina@uchicago.edu*
[2]*Department of Statistics, Harvard University,* [b]*ljanson@fas.harvard.edu*

Goodness-of-fit (GoF) testing is ubiquitous in statistics, with direct ties to model selection, confidence interval construction, conditional independence testing, and multiple testing, just to name a few applications. While testing the GoF of a simple (point) null hypothesis provides an analyst great flexibility in the choice of test statistic while still ensuring validity, most GoF tests for composite null hypotheses are far more constrained, as the test statistic must have a tractable distribution over the entire null model space. A notable exception is *co-sufficient sampling* (CSS): resampling the data conditional on a sufficient statistic for the null model guarantees valid GoF testing using any test statistic the analyst chooses. But CSS testing requires the null model to

Causality

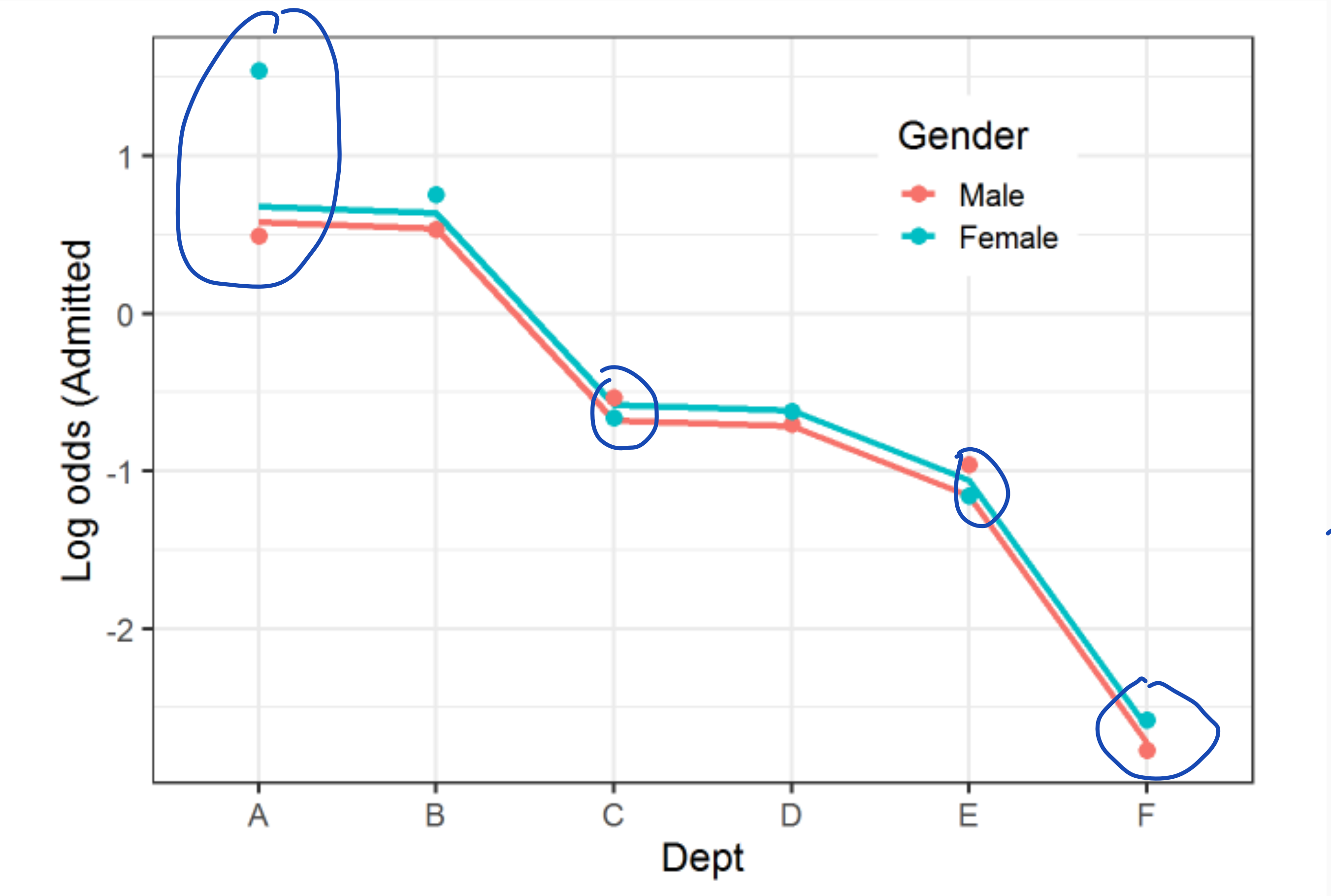
- randomization; confounding; observational studies; experiments; “correlation is not causation”, Simpson’s ‘paradox’
- counterfactuals; average treatment effect; conditional average treatment effect; ...
- graphical models; directed acyclic graphs; causal graphs; Markov assumptions...
- The Book

Confounding variables

Major	Men: Number of applicants	Men: Number admitted	Men: Percent admitted	Women: Number of applicants	Women: Number admitted	Women: Percent admitted
A	825	512	62	108	89	82
B	560	353	63	25	17	68
C	325	120	37	593	202	34
D	417	138	33	375	131	35
E	191	53	28	393	94	24
F	373	22	6	341	24	7
Total	2691	1198	44	1835	557	30

```
data(UCBAdmissions)
```

... Confounding variables



logistic regression

y = adm. / not

x gender + dept

… Confounding variables

race of defendant	death penalty imposed	death penalty not imposed	percentage
white	19	141	11.88%
black	17	149	10.24%

... Confounding variables

Radelet 1981

race of defendant	death penalty imposed	death penalty not imposed	percentage
white	19	141	11.88%
black	17	149	10.24%

white victim	race of defendant	death penalty imposed	death penalty not imposed	percentage
	white	19	132	12.58%
	black	11	52	17.46%

black victim	race of defendant	death penalty imposed	death penalty not imposed	percentage
	white	0	9	0%
	black	6	97	5.83%

Simpson's paradox

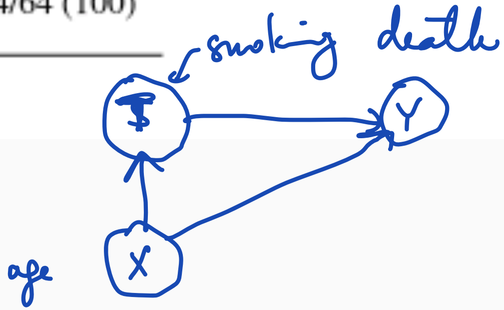
race of victim
confounding
factor

… Confounding variables

Age (years)	Smokers	Non-smokers
Overall	139/582 (24)	230/732 (31)
18–24	2/55 (4)	1/62 (2)
25–34	3/124 (2)	5/157 (3)
35–44	14/109 (13)	7/121 (6)
45–54	27/130 (21)	12/78 (15)
55–64	51/115 (44)	40/121 (33)
65–74	29/36 (81)	101/129 (78)
75+	13/13 (100)	64/64 (100)

Table 6.8 Twenty-year survival and smoking status for 1314 women (Appleton *et al.*, 1996). The smoker and non-smoker columns contain number dead/total (% dead).

age is a confounder



- A – binary treatment indicator   or T
- Y – binary outcome
- "A causes Y" to be distinguished from "A is associated with Y"

AoS uses X for tmt

could be continuous

- A – binary treatment indicator — AoS uses X for tmt
- Y – binary outcome — could be continuous
- "A causes Y" to be distinguished from "A is associated with Y"

- introduce potential outcomes $Y(0), Y(1)$ — AoS C_0, C_1; HR Y^a

Y^0 Y^1

- causal treatment effect $\theta = E(Y(1)) - E(Y(0))$ — want to estimate this

C_1

- association $\alpha = E(Y \mid A = 1) - E(Y \mid A = 0)$ — have data to estimate α

- Consistency assumption: $Y = Y(a)$ — we can learn about potential outcomes from observed values

if $A = 1$ $Y = Y(1)$ pot. outc.
$\quad\;\; 0 \quad\;\; Y(0)$ " "

$Y = C_Y$

Counterfactual: Examples

$Y = AY^1 + (1-A)Y^0$
$= AY(1) + (1-A)Y(0)$

Potential outcomes C_0, C_1

X (A)	Y (Y)	C_0 (Y(0))	C_1 (Y(1))
0	4	4	*
0	7	7	*
0	2	2	*
0	8	8	*
1	3	*	3
1	5	*	5
1	8	*	8
1	9	*	9

treatment X, response Y

Potential outcomes Y^0, Y^1

Table 2.1

obs'd

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

cons.

Causal Effect and Association

Potential outcomes

Table 1.1

	$Y^{a=0}$	$Y^{a=1}$
Rheia	0	1
Kronos	1	0
Demeter	0	0
Hades	0	0
Hestia	0	0
Poseidon	1	0
Hera	0	0
Zeus	0	1
Artemis	1	1
Apollo	1	0
Leto	0	1
Ares	1	1
Athena	1	1
Hephaestus	0	1
Aphrodite	0	1
Cyclope	0	1
Persephone	1	1
Hermes	1	0
Hebe	1	0
Dionysus	1	0

θ = causal effect

$E(Y^1 - Y^0)$

$\frac{10}{20} - \frac{10}{20}$

$= 0$

Observed outcomes

data

Table 1.2

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

$\frac{1+6}{13} = \frac{7}{13} =$

$E(Y \mid A=1)$

$\frac{1+2}{7} = \frac{3}{7} = E(Y \mid A=0)$

$\alpha = \frac{7}{13} - \frac{3}{7}$

Causal treatment effect

AoS Eq. (16.2)

$$\theta = \mathrm{E}(Y(1)) - \mathrm{E}(Y(0))$$

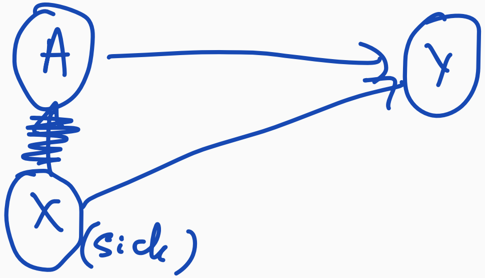
risk difference; ratio; odds

also called “ATE” and “ACE”: average treatment/causal effect

$\alpha \neq \theta$ in general

$$\alpha = \mathrm{E}(Y \mid A = 1) - \mathrm{E}(Y \mid A = 0)$$

this can be estimated from the data



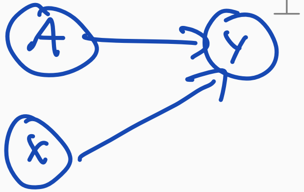
c_0 c_1

If A is is independent of $(Y(0), Y(1))$, then $\boxed{\theta = \alpha}$

trt ass't

If treatment is randomly assigned, then $A \perp (Y(0), Y(1))$

$\perp \equiv$independent



Example 16.2

Y^0 Y^1

X	Y	C_0	C_1
0	0	0	0*
0	0	0	0*
0	0	0	0*
0	0	0	0*
1	1	1*	1
1	1	1*	1
1	1	1*	1
1	1	1*	1

$\theta = 0$; $\alpha = 1$

(C_0, C_1) not independent of X

$Y(a)$ Y^a

$(C_0, C_1) \perp\!\!\!\perp (Y(0), Y(1))$

$\bar{Y}^0 = 0$

$\bar{Y}^1 = 1$

X	Y	C_0	C_1
0	0	0	0*
1	0	0	0*
1	0	0	0*
1	0	0	0*
1	1	1*	1
1	1	1*	1
1	1	1*	1
1	1	1*	1

$\theta = 0, \quad \alpha = 4/7 < 1$

thought experiment

actionable item deduced from $\alpha = 1$,

Potential outcomes

Table 1.1

	$Y^{a=0}$	$Y^{a=1}$
Rheia	0	1
Kronos	1	0
Demeter	0	0
Hades	0	0
Hestia	0	0
Poseidon	1	0
Hera	0	0
Zeus	0	1
Artemis	1	1
Apollo	1	0
Leto	0	1
Ares	1	1
Athena	1	1
Hephaestus	0	1
Aphrodite	0	1
Cyclope	0	1
Persephone	1	1
Hermes	1	0
Hebe	1	0
Dionysus	1	0

Observed outcomes

Table 1.2

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

Three types of causal statements

1. A well-understood evidence-based mechanism, or set of mechanisms, that links a cause to its effect

2. two phenomena are linked by a stable association, whose direction is established and which cannot be explained by mutual dependence on some other allowable variable

3. observed association may be linked to causal effect via counterfactuals if
   $(Y(0), Y(1)) \perp A$

not usually testable

- typically have additional explanatory variables (covariates) X

AoS uses Z; HR use L

- causal effect of treatment when $X = x$

$$\theta(x) = \mathrm{E}(Y(1) \mid X = x) - \mathrm{E}(Y(0) \mid X = x)$$

not estimable

- marginal causal effect

$$\theta = \mathrm{E}_X\{\mathrm{E}(Y(1) \mid X) - \mathrm{E}(Y(0) \mid X)\}$$

- association function

$$r(x) = \mathrm{E}(Y \mid A = 1, X = x) - \mathrm{E}(Y \mid A = 0, X = x)$$

$r(x, A=1) - r(x, A=0)$

- marginal association

$$\mathrm{E}_X\{r(X)\}$$

Example

Table 2.2

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

8

6

$E[Y|A=1,L=0] - E[Y|A=0,L=0]$

$\theta_{L=0} = \alpha_{L=0} = \frac{1}{4} - \frac{1}{4} = 0$

assoc =

$\theta_{L=1} = \alpha_{L=1} = \frac{6}{9} - \frac{2}{3} = 0$

$$\begin{tikzcd} & L \arrow[dl] \arrow[dr] & \\ A \arrow[rr] & & Y \end{tikzcd}$$

$A \perp Y \mid L$
random ass't
↑ claim

$L = 1$ critical condition

$L = 0$ stable condition

conditional randomization

If not $\theta \neq \alpha$ averaged over L (ntbc)

- in observational studies treatment is not randomly assigned $\implies \theta(x) \neq r(x)$
- No unmeasured confounding:

$$\{Y(a); a \in \mathcal{A}\} \perp A \mid X$$

can learn about $Y(a)$ even if $A \neq a$ by using observed Y for 'similar' people from $A = a$ group

- under the assumption of no unmeasured confounding, marginal causal effect

causal effect of trt

$E\{Y(1)\} - E\{Y(0)\}$

causal effect of trt among treated $(a=1)$

$$E(Y(a)) = \int E(Y \mid A = a, X = x) dF_X(x)$$

$Y = \beta_0 + \beta_1 A + \beta_2 X + \varepsilon$

can be estimated by the association function

marginal causal effect θ est'd by

$$\widehat{E}(Y(a)) = \frac{1}{n}\sum_{i=1}^{n} \hat{r}(a, X_i) = \hat{\beta}_0 + \hat{\beta}_1 a + \hat{\beta}_2 \bar{X}_n$$

$r(x,a) = \beta_0 + \beta_1 a + \beta_2 x$

causal reg function $\equiv$ adjusted treatment effect

Effect of confounding

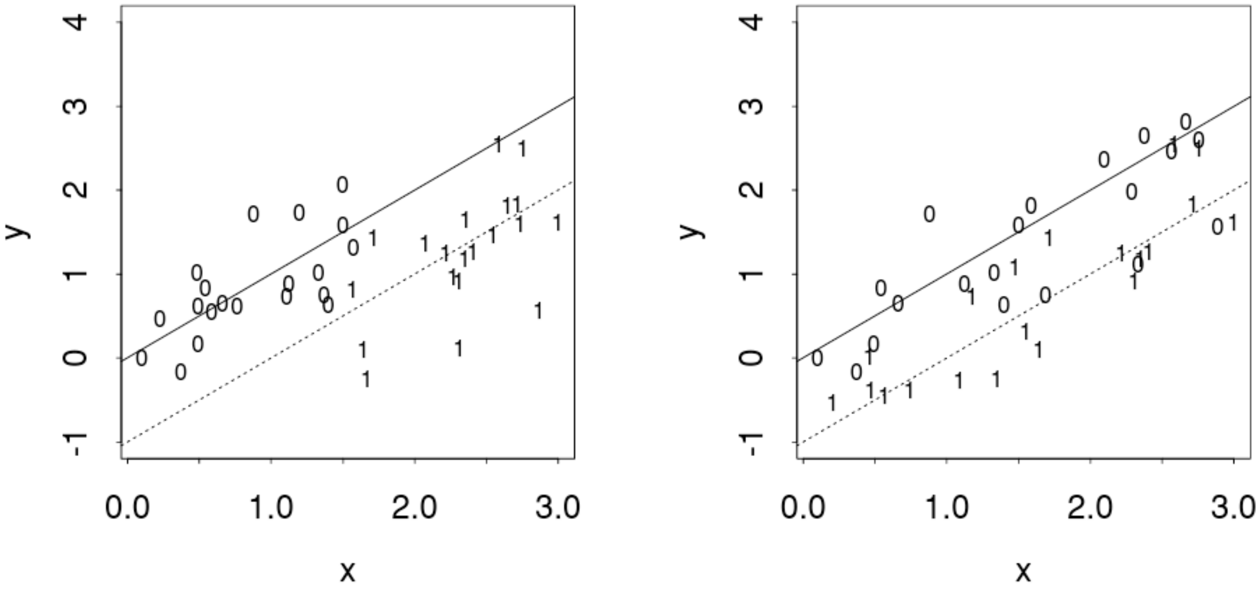


Figure 9.2 Simulated results from experiments to compare the effect of a treatment T on a response Y that varies with a covariate X. The lines show the mean response for $T = 0$ (solid) and $T = 1$ (dots). Left: the effect of T is confounded with dependence on X. Right: the experiment is balanced, with random allocation of T dependent on X.

true causal eff ≡ 1

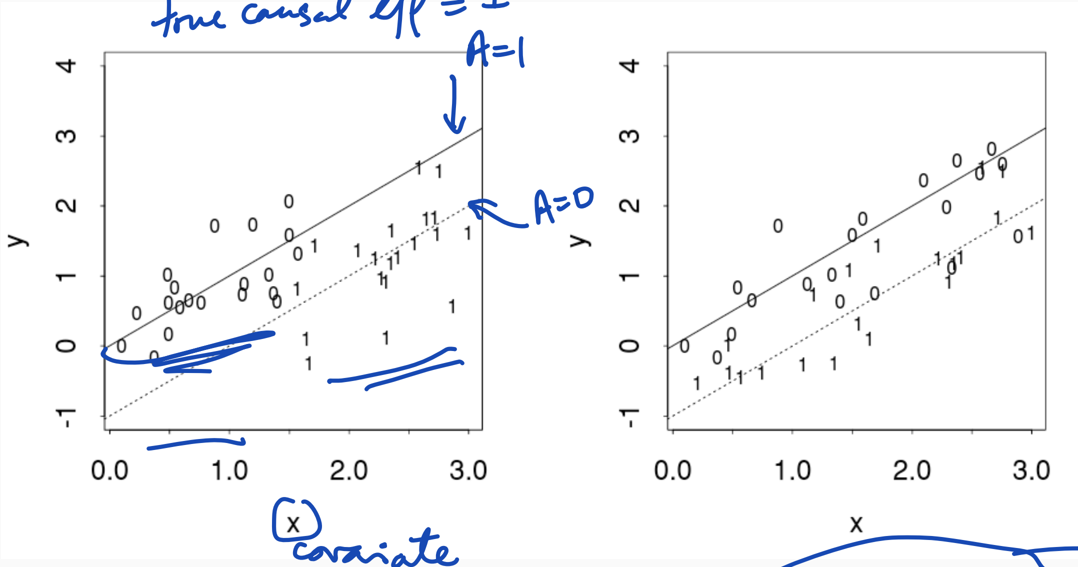


Figure 9.2 Simulated results from experiments to compare the effect of a treatment T on a response Y that varies with a covariate X. The lines show the mean response for $T = 0$ (solid) and $T = 1$ (dots). Left: the effect of T is confounded with dependence on X. Right: the experiment is balanced, with random allocation of T dependent on X.

Causal effect $\equiv 1$

Left: $\bar{y}_1 - \bar{y}_0 = 0.2 \pm 0.3$ Right: $\bar{y}_1 - \bar{y}_0 = -1.2 \pm 0.3$

$\beta_1 A$ $\hat{\sigma}^2 = \sum_{i=1} \hat{\varepsilon}_i^2 / (n-2)$

adjust for covariate: $y = \beta_0 + \beta_1 x + \delta t + \epsilon$

Left: $\hat{\delta} = -0.7 \pm 0.3$ Right: $\hat{\delta} = -1.25 \pm 0.16$

effic.

right randomized within pairs; matched on x

"Bradford-Hill guidelines" Evidence that an observed association is causal is strengthened if:

- the association is strong
- the association is found consistently — over a number of independent studies
- the association is specific to the outcome studied
- the observation of a potential cause occurs earlier in time than the outcome
- there is a dose-response relationship
- there is subject-matter theory that makes a causal effect plausible
- the association is based on a suitable natural experiment

see also AoS §16.3

Simpson's paradox revisited

AoS 16.4

highest prop$^{=}$ of
$Z=1$ (men)
have $X=1$

	$Y=1$	$Y=0$	$Y=1$	$Y=0$
$X=1$	.1500	.2250	.1000	.0250
$X=0$	.0375	.0875	.2625	.1125
	$Z=1$ (men)		$Z=0$ (women)	

low prop$^{=}$ of
$Z=0$ (women)
have $X=1$

The marginal distribution for (X, Y) is

	$Y=1$	$Y=0$	
$\cancel{X} A=1$	.25	.25	.50
$\cancel{X} A=0$	.30	.20	.50
	.55	.45	1

From these tables we find that,

$$\begin{aligned}
\mathbb{P}(Y=1|\cancel{X}A=1)-\mathbb{P}(Y=1|\cancel{X}A=0) &= -0.1\\
\mathbb{P}(Y=1|X=1,Z=1)-\mathbb{P}(Y=1|X=0,Z=1) &= 0.1\\
\mathbb{P}(Y=1|X=1,Z=0)-\mathbb{P}(Y=1|X=0,Z=0) &= 0.1.
\end{aligned}$$

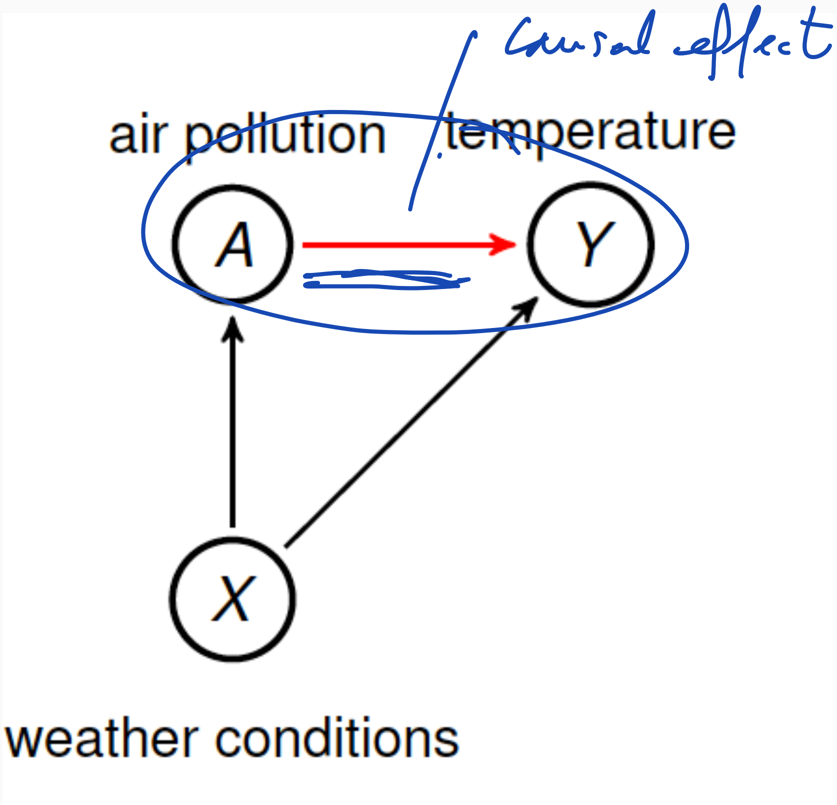
To summarize, we *seem* to have the following information:

Mathematical Statement	English Statement?

$P(Y=1 | A=1)$
$= P(Y=1, A=1)/P(A=1)$
$= .25/.50$
$= 1/2 = .5$

confusion of causal effect
with association

$P(Y=1 | A=0)$
$= .30/.5 = .6$



- assume no unmeasured confounding
- want to estimate

$\theta(x)$ $E(Y(1) \mid X) - E(Y(0) \mid X)$

causal regression function

- or possibly $E_X\{E(Y(1) \mid X) - E(Y(0) \mid X)\}$ θ

marginal effect of A

causal eff.

- regression model

$$E(Y \mid X, A) = \beta_0 + \beta_1 A + \beta_2 X$$

- or something more complex

$$E(Y \mid X, A) = f(X, A)$$

neural network ...

- estimand average causal effect or average treatment effect (ATE)

Version 1.

$$\theta = \underline{E\{Y(1)\} - E\{Y(0)\}}$$

averaged over Xs

estimand: something we estimate

- under the linear model $\underline{E(Y \mid X, A)} = \underline{\beta_0 + \beta_1 A + \beta_2 X}$, the ATE is β_1 if the linear model is correct

(interaction X, A messes this up)

- $$\widehat{\mathrm{E}}(Y(a)) = \underline{\frac{1}{n}\sum_{i=1}^{n} \widehat{\mathrm{E}}(Y \mid A = a, X_i)}$$

estimate θ

- recovers $\hat{\beta}_1$ in a linear model

Version 2

$$E_1(Y(1)) = E\left\{\frac{AY}{p_n(A=1 \mid X)}\right\}$$

X, Y, A

- treat $Y_i(1)$ as missing data, if $A_i = 0$ (and v.v.)
- write

$$\mathrm{E}(Y(a)) = \mathrm{E}\left\{\frac{1\{A=a\}Y}{\mathrm{pr}(A=a \mid x)}\right\}$$

- model $\mathrm{pr}(A=a \mid X)$, e.g. by logistic regression
- doubly robust estimator

of $\mathrm{E}(Y(1))$

$$\hat{\mu}^{AIPW} = \underbrace{\frac{1}{n}\sum_{i=1}^{n}\frac{A_iY_i}{\widehat{\mathrm{pr}}(A=1 \mid X_i)}}_{\text{est. of } E\{Y(1)\} \text{ IPW}} + \left\{1 - \frac{A_i}{\widehat{\mathrm{pr}}(A=1 \mid X_i)}\right\}\underbrace{\widehat{\mathrm{E}}_2(Y(1))}_{\text{use linear model}}$$

propensity score

est. of $E\{Y(1)\}$

IPW

use linear model

Directed graphs and randomization

graphs can be useful for clarifying dependence relations among random variables

Fig 9.1 SM

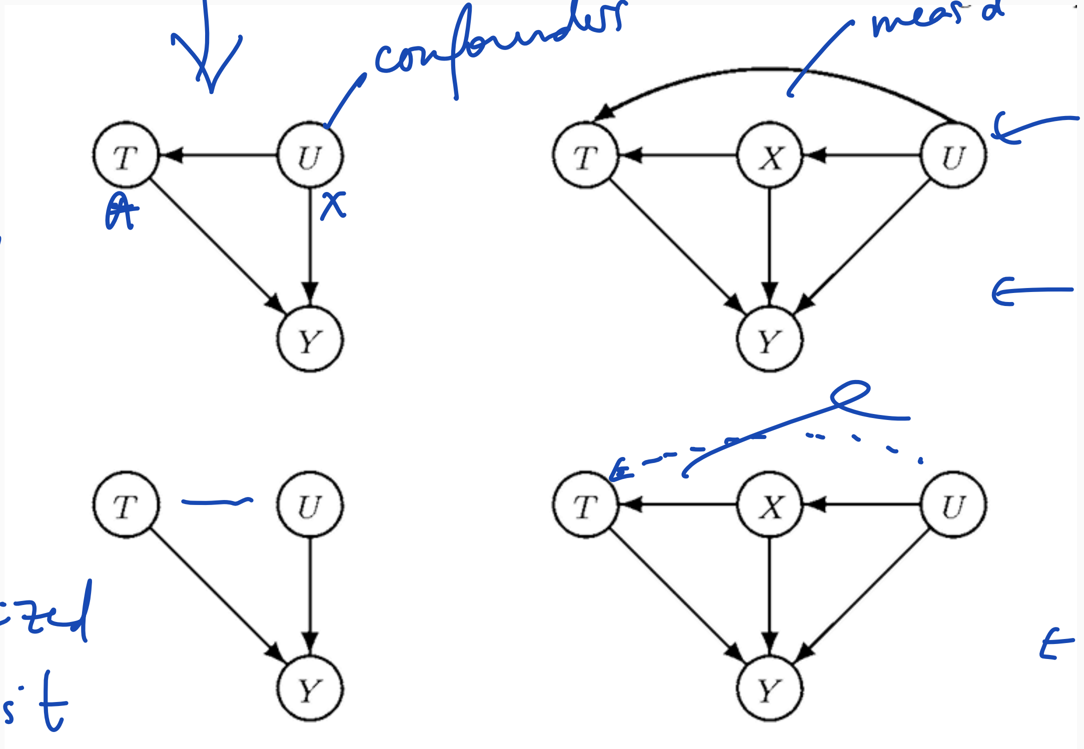
confounders

meas'd

unmeasured conf.

obs'l

obs'l



randomized trt ass't

randomized

DAGs and confounders

AoS 17.8

276 17. Directed Graphs and Conditional Independence

not measured (here)

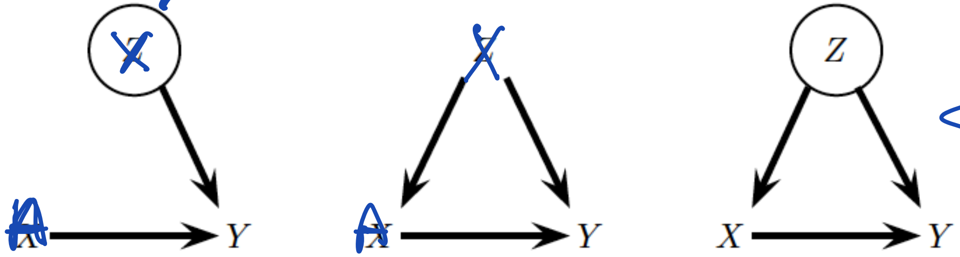


FIGURE 17.11. Randomized study; Observational study with measured confounders; Observational study with unmeasured confounders. The circled variables are unobserved.

randomized study

observational study $\mathrm{E}(Y \mid x) = \int \mathrm{E}(Y \mid X, Z = z) dF_Z(z)$

unobserved confounder: $\theta \neq \alpha$

no causal interpr.

The NEW ENGLAND JOURNAL of MEDICINE

SPECIAL ARTICLE

Air Pollution and Mortality at the Intersection of Race and Social Class

Kevin P. Josey, Ph.D., Scott W. Delaney, Sc.D., J.D., Xiao Wu, Ph.D., Rachel C. Nethery, Ph.D., Priyanka DeSouza, Ph.D., Danielle Braun, Ph.D., and Francesca Dominici, Ph.D.

ABSTRACT

BACKGROUND

Black Americans are exposed to higher annual levels of air pollution containing fine particulate matter (particles with an aerodynamic diameter of ≤2.5 μm [$PM_{2.5}$]) than White Americans and may be more susceptible to its health effects. Low-income Americans may also be more susceptible to $PM_{2.5}$ pollution than high-income Americans. Because information is lacking on exposure–response curves for $PM_{2.5}$ exposure and mortality among marginalized subpopulations categorized

From the Departments of Biostatistics (K.P.J., R.C.N., D.B., F.D.) and Environmental Health (S.W.D.), Harvard T.H. Chan School of Public Health, Boston; the Department of Biostatistics, Mailman School of Public Health, Columbia University, New York (X.W.); and the Department of Urban and Regional Planning,