Quantitative Economics: Regression

Regression Interpretation

Ian Crawford

Department of Economics

Nuffield College

Trinity Term, 2018

Weeks 3 & 4

Regression

1. Regression with the population
2. Regression with a sample
3. Regression Interpretation.

Interpreting Regressions

Interpreting linear regression coefficients

The effects of individual variables
Causal v Descriptive Interpretation
The Neyman/Rubin causal model
Regression and causal inference

Interpreting Regressions

$$\mathbb{E}\left[Y|X_1, X_2\right] = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

- We are interested in the interpretation of the least squares coefficients.
  - β_0 is the "intercept"
  - β_1 and β_2 are the "slopes" wrt to X_1 and X_2 .
- The slope coefficients have two possible interpretations
  - Literal/descriptive - this always valid.
  - Causal - this is typically controversial.

Interpreting Regressions

$$\mathbb{E}\left[Y|X_1, X_2\right] = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

- Their descriptive interpretation follows directly from the Frisch-Waugh-Lovel theorem: they are measures of the degree of linear association between Y and each individual explanatory variable, netting out the effect of the other variables.
- Their causal interpretation is that they measure the causal effect of *ceteris paribus* exogenous changes in the explanatory variable on the dependent variable Y.

Interpreting Regressions

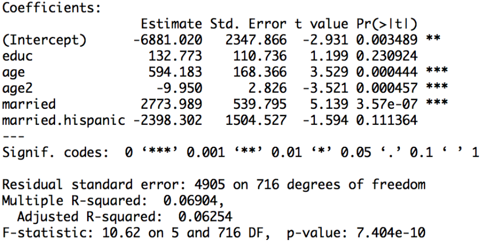
- A good place to start is to go back to the basic idea behind a regression - that it is a solution to a prediction problem.
- The justification for an interest in the CEF is that is it the best predictor, in the squared-loss sense, of Y given $X_1, X_2, ...$
- The justification for the LRM is that it is the best linear predictor, in the squared-loss sense, of Y given $X_1, X_2, ...$ amongst the class of linear predictors.

Interpreting Regressions

- Generating predicted values from regressions is mechanically very easy.
- If you want to know the predicted value of Y for $X_1 = x_1$, $X_2 = x_2$, etc just plug in the values of the X's at which you want to predict Y and do the arithmetic.

Interpreting Regressions

Examples



```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      -6881.020   2347.866  -2.931 0.003489 **
educ               132.773    110.736   1.199 0.230924
age                594.183    168.366   3.529 0.000444 ***
age2                -9.950      2.826  -3.521 0.000457 ***
married           2773.989    539.795   5.139 3.57e-07 ***
married.hispanic -2398.302   1504.527  -1.594 0.111364
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4905 on 716 degrees of freedom
Multiple R-squared:  0.06904,
  Adjusted R-squared:  0.06254
F-statistic: 10.62 on 5 and 716 DF,  p-value: 7.404e-10
```

Interpreting Regressions

Examples

Suppose we want to predict Earnings for the average values of the covariates. The averages values of the regressors are 10.27 for `educ`, 24.52 for `age`, 645.11 for `age2`, 0.16 for `married` and 0.02 for `married.hispanic`. The predicted value is obtained by multiplying each value of the regressor by its coefficient and summing:

$$-6.881.020 + 132.773 \times 10.27 + 594.183 \times 24.52$$
$$-9.950 \times 645.11 + 2773.989 \times 0.16 - 2398.302 \times 0.02$$

Which comes out at the overall mean for Earnings: $3,042.897. This is not a coincidence.

Interpreting Regressions

Examples

Suppose we want to predict Earnings for someone with 12 years of education (educ=12), who was 30 (age=30, age2=900), who was married (married =1) and not an Hispanic American (married.hispanic=0). The predicted value is obtained by multiplying each value of the regressor by its coefficient and summing:

$$-6.881.020 + 132.773 \times 12 + 594.183 \times 30$$
$$-9.950 \times 900 + 2773.989 \times 1 - 2398.302 \times 0$$

Which comes to \$6,356.812

Interpreting Regressions

Examples

Suppose we want to compare that person with someone who is unmarried. Originally we predicted

$$-6.881.020 + 132.773 \times 12 + 594.183 \times 30$$
$$-9.950 \times 900 + 2773.989 \times 1 - 2398.302 \times 0$$

Which came to $6,356.812. Now hold everything else constant and change the `married` value to 0

$$-6.881.020 + 132.773 \times 12 + 594.183 \times 30$$
$$-9.950 \times 900 + 2773.989 \times 0 - 2398.302 \times 0$$

Which comes out at a lot less: $3582.823.

Interpreting Regressions

- The LRM provides a prediction for various values of the X's.
- And as we change the values of the X's so the prediction changes according to the empirical associations in the data.
- The coefficients (the β's) give us information on that predictor:prediction relationship.

Interpretation of Regressions

Interpretation of Regressions

- Changes in covariates
  - Linear terms
  - Polynomials
  - Interactions
  - Logs

Interpreting Regressions

1. Treat the LRM as a simple linear-in-parameters equation.
2. Then think about partial differentials of this wrt to the X's as telling you the "effect" of each X on the predicted value of Y, holding other things fixed.
3. Adjust your interpretation accordingly for discrete regressors/log regressors etc.

Interpreting Regressions

- Consider our fairly elaborate, LRM:

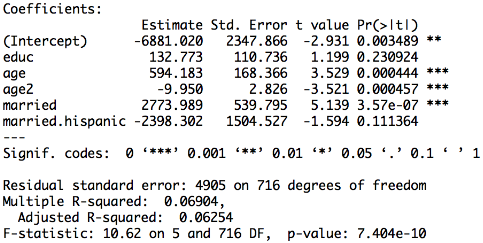
$$\mathbb{E}\left[Y|X_1,...,X_4\right]=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3\left(X_2\right)^2+\beta_4X_3+\beta_5\left(X_3X_4\right)$$

- The effects of each of the four variables on the conditional mean of Y are as follows

Variable	Effect
X_1	β_1
X_2	$\beta_2+2\beta_3X_2$
X_3	$\beta_4+\beta_5X_4$
X_4	β_5X_3

Interpreting Regressions

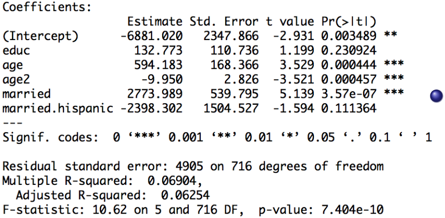
$$\mathbb{E}\left[Y|X_1, ..., X_4\right] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 \left(X_2\right)^2 + \beta_4 X_3 + \beta_5 \left(X_3 X_4\right)$$



```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)     -6881.020   2347.866  -2.931 0.003489 ** 
educ              132.773    110.736   1.199 0.230924    
age               594.183    168.366   3.529 0.000444 ***
age2               -9.950      2.826  -3.521 0.000457 ***
married          2773.989    539.795   5.139 3.57e-07 ***
married.hispanic -2398.302   1504.527  -1.594 0.111364    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4905 on 716 degrees of freedom
Multiple R-squared:  0.06904,
  Adjusted R-squared:  0.06254 
F-statistic: 10.62 on 5 and 716 DF,  p-value: 7.404e-10
```

Interpreting Regressions

$$\mathbb{E}\left[Y|X_1,...,X_4\right] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 \left(X_2\right)^2 + \beta_4 X_3 + \beta_5 \left(X_3 X_4\right)$$



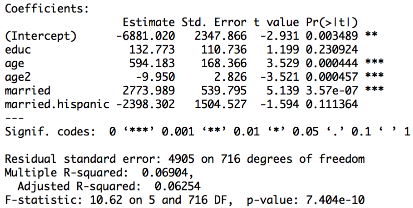
```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     -6881.020   2347.866  -2.931 0.003489 **
educ              132.773    110.736   1.199 0.230924
age               594.183    168.366   3.529 0.000444 ***
age2               -9.950      2.826  -3.521 0.000457 ***
married          2773.989    539.795   5.139 3.57e-07 ***
married.hispanic -2398.302   1504.527  -1.594 0.111364
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4905 on 716 degrees of freedom
Multiple R-squared:  0.06904,
  Adjusted R-squared:  0.06254
F-statistic: 10.62 on 5 and 716 DF,  p-value: 7.404e-10
```

- The effect of education on earnings is the coefficient 132.773.
- Now recall that in the NSW data education is measured in completed years.

- This means that a one year increase in years of education raises expected earnings by \$132.773.
- Since the LRM is linear with respect to education the effect of *n* additional years is $n \times$\$132.773.

Interpreting Regressions

$$\mathbb{E}\left[Y|X_1,\ldots,X_4\right]=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3\left(X_2\right)^2+\beta_4X_3+\beta_5\left(X_3X_4\right)$$



```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)      -6881.020   2347.866  -2.931 0.003489 ** 
educ               132.773    110.736   1.199 0.230924    
age                594.183    168.366   3.529 0.000444 ***
age2                -9.950      2.826  -3.521 0.000457 ***
married           2773.989    539.795   5.139 3.57e-07 ***
married.hispanic -2398.302   1504.527  -1.594 0.111364    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1 

Residual standard error: 4905 on 716 degrees of freedom
Multiple R-squared:  0.06904,
  Adjusted R-squared:  0.06254 
F-statistic: 10.62 on 5 and 716 DF,  p-value: 7.404e-10
```

- The LRM is quadratic with respect to age. So the effect of a change in age varies with age:

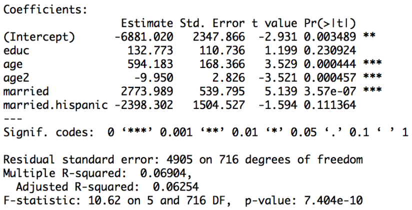
$$594.183-(2\times 9.950)\,age$$

- This means that the effect of age on earnings is non-linear; earnings increase with age but at a declining rate.
- The max is where the derivative is zero:

$$594.183-(2\times 9.950)\,age=0\Rightarrow age\approx 30$$

Interpreting Regressions

$$\mathbb{E}\left[Y|X_1, ..., X_4\right] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 \left(X_2\right)^2 + \beta_4 X_3 + \beta_5 \left(X_3 X_4\right)$$



```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)     -6881.020   2347.866  -2.931 0.003489 ** 
educ              132.773    110.736   1.199 0.230924    
age               594.183    168.366   3.529 0.000444 ***
age2               -9.950      2.826  -3.521 0.000457 ***
married          2773.989    539.795   5.139 3.57e-07 ***
married.hispanic -2398.302   1504.527  -1.594 0.111364    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4905 on 716 degrees of freedom
Multiple R-squared:  0.06904,
  Adjusted R-squared:  0.06254
F-statistic: 10.62 on 5 and 716 DF,  p-value: 7.404e-10
```

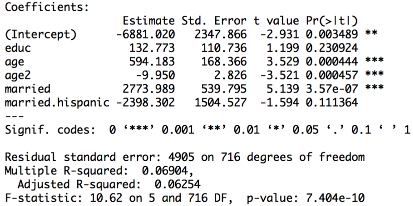
- The effect of being married interacts with whether or not you someone is Hispanic.
- The effect is

  $$2773.989 - 2398.302\ \textit{hispanic}$$

- So for non-Hispanics, $\textit{hispanic} = 0$, and the “marriage premium” is \$2773.989.
- But for Hispanics, $\textit{hispanic} = 1$, and the effect is 2773.989 - 2398.302 $=$ \$375.687.

Interpreting Regressions

$$\mathbb{E}\left[Y|X_1,...,X_4\right] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 \left(X_2\right)^2 + \beta_4 X_3 + \beta_5 \left(X_3 X_4\right)$$



```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     -6881.020   2347.866  -2.931 0.003489 **
educ              132.773    110.736   1.199 0.230924
age               594.183    168.366   3.529 0.000444 ***
age2               -9.950      2.826  -3.521 0.000457 ***
married          2773.989    539.795   5.139 3.57e-07 ***
married.hispanic -2398.302   1504.527  -1.594 0.111364
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4905 on 716 degrees of freedom
Multiple R-squared:  0.06904,
  Adjusted R-squared:  0.06254
F-statistic: 10.62 on 5 and 716 DF,  p-value: 7.404e-10
```

- The earnings differential for Hispanics, in this, specification of the model is mediated by marital status. The effect is

$$-2398.302 \; married$$

- So for married Hispanics, *married* $= 1$, and wage effect is -\$2398.302.
- But for Hispanics who are not married, *married* $= 0$, and the effect is 0.
- On this rather odd evidence it seems that the labour market discriminated against Hispanic Americans only if they were married. (But note the t-value.)

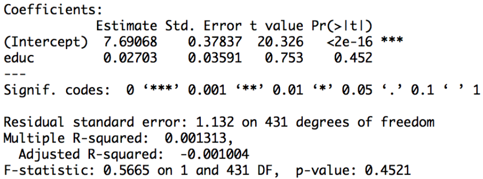
Interpreting Regressions - logs

- Sometimes we may wish to work with variables expressed in logs.
- logs are used to capture proportional relationships between variables.

Model	Interpretation of β_1
$\mathbb{E}\left[\log Y \mid X\right] = \beta_0 + \beta_1 X$	A change in X by 1 unit has a $100\beta_1$% effect on Y
$\mathbb{E}\left[Y \mid \log X\right] = \beta_0 + \beta_1 \log X$	A 1% change in X has an effect on Y of $0.01\beta_1$.
$\mathbb{E}\left[\log Y \mid \log X\right] = \beta_0 + \beta_1 \log X$	β_1 measures the elasticity of Y with respect to X

Interpreting Regressions - logs

$$\mathbb{E}\left[\log Y | X\right] = \beta_0 + \beta_1 X$$

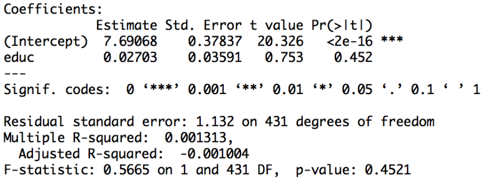


```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.69068    0.37837  20.326   <2e-16 ***
educ         0.02703    0.03591   0.753    0.452
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.132 on 431 degrees of freedom
Multiple R-squared:  0.001313,
  Adjusted R-squared:  -0.001004
F-statistic: 0.5665 on 1 and 431 DF,  p-value: 0.4521
```

- This just uses the sub-sample with positive earnings.
- Since there are no zeros mean earnings in these data are \$5073.84 which is higher than the full sample.

Interpreting Regressions - logs

$$\mathbb{E}\left[\log Y | X\right] = \beta_0 + \beta_1 X$$

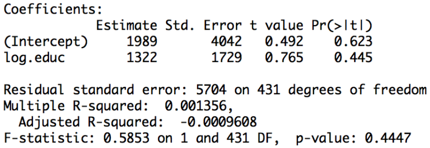


```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  7.69068    0.37837  20.326   <2e-16 ***
educ         0.02703    0.03591   0.753    0.452    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.132 on 431 degrees of freedom
Multiple R-squared:  0.001313,
  Adjusted R-squared:  -0.001004
F-statistic: 0.5665 on 1 and 431 DF,  p-value: 0.4521
```

- A one additional year in education is associated with a 2.703% increase in earnings.
- Since mean earnings in these data are $5073.84 that amounts to about $137.

Interpreting Regressions - logs

$$\mathbb{E}\left[Y \middle| \log X\right] = \beta_0 + \beta_1 \log X$$

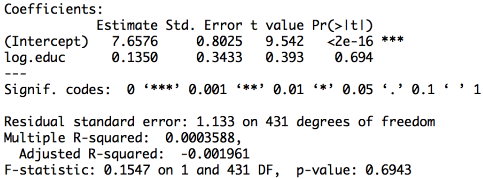


```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)     1989       4042   0.492    0.623
log.educ        1322       1729   0.765    0.445

Residual standard error: 5704 on 431 degrees of freedom
Multiple R-squared:  0.001356,
  Adjusted R-squared:  -0.0009608
F-statistic: 0.5853 on 1 and 431 DF,  p-value: 0.4447
```

- A one percent increase in education is associated with a \$13.22 increase in earnings.
- Since mean education in these data is 10.427 years, a one-percent increase is about 6 weeks.
- That amounts to an annual effect of about \$132.

Interpreting Regressions - logs

$$\mathbb{E}\left[\log Y \middle| \log X\right] = \beta_0 + \beta_1 \log X$$



```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   7.6576     0.8025   9.542   <2e-16 ***
log.educ      0.1350     0.3433   0.393    0.694
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.133 on 431 degrees of freedom
Multiple R-squared:  0.0003588,
  Adjusted R-squared:  -0.001961
F-statistic: 0.1547 on 1 and 431 DF,  p-value: 0.6943
```

- The elasticity of earnings with respect to education is 0.135.
- A 1% increase in education rates earnings by 0.135%

Interpreting Regressions

Summary

Interpreting regressions

Treat as a linear (in parameters) equation.

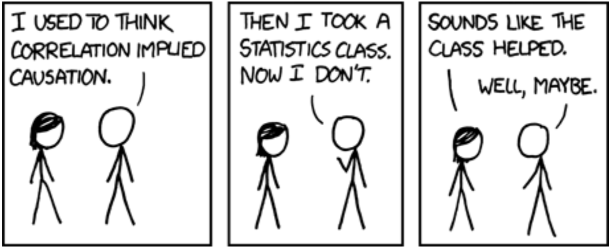
Partially differentiate wrt the variable of interest.

Adjust your interpretation when dealing with binary/discrete-ordered/logged regressors etc.

Remember that for non-linearly transformed variables the marginal effects depend on the point at which you evaluate them.

Bear in mind the distributions of the data themselves (the mean, range etc) when interpreting coefficients.

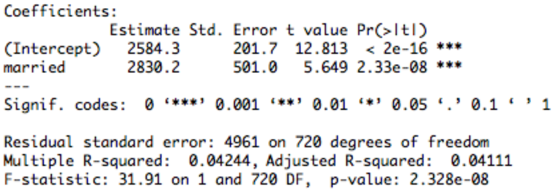
Interpreting Regressions - Causal Inference


I USED TO THINK CORRELATION IMPLIED CAUSATION.
THEN I TOOK A STATISTICS CLASS. NOW I DON'T.
SOUNDS LIKE THE CLASS HELPED.
WELL, MAYBE.

Source: xkcd

Interpreting Regressions

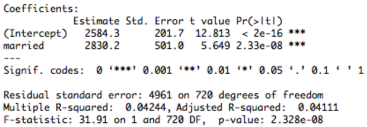
- Consider the following regression of Earnings on marriage from the NSW data



```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2584.3      201.7  12.813  < 2e-16 ***
married       2830.2      501.0   5.649 2.33e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4961 on 720 degrees of freedom
Multiple R-squared:  0.04244, Adjusted R-squared:  0.04111
F-statistic: 31.91 on 1 and 720 DF,  p-value: 2.328e-08
```

- This predicts that married people get paid more than unmarried people by $2830.20.
- It is statistically significant.
- Is this causal? Or merely an association? If someone exogenously got married (?!?!?) would their Earnings rise?

Interpreting Regressions



```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2584.3      201.7  12.813  < 2e-16 ***
married       2830.2      501.0   5.649 2.33e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4961 on 720 degrees of freedom
Multiple R-squared:  0.04244, Adjusted R-squared:  0.04111
F-statistic: 31.91 on 1 and 720 DF,  p-value: 2.328e-08
```

- In this model we have a single binary regressor which we will denote by $D \in \{0, 1\}$

$$\mathbb{E}\left[Y|D\right] = \beta_0 + \beta_1 D$$

- Our estimated model is

$$\widehat{\text{Earnings}} = 2584.3 + 2830.2\text{married}$$

Interpreting Regressions

- Find the conditional expectations for $D = 0$ and $D = 1$ in the normal way

$$\begin{aligned} \mathbb{E}\left[Y|D\right] &= \beta_0 + \beta_1 D \\ \mathbb{E}\left[Y|D=0\right] &= \beta_0 \\ \mathbb{E}\left[Y|D=1\right] &= \beta_0 + \beta_1 \end{aligned}$$

- The difference between them is

$$\mathbb{E}\left[Y|D=1\right] - \mathbb{E}\left[Y|D=0\right] = \beta_1$$

- This is the difference between the mean outcome for the $D = 1$ group and the $D = 0$ group. In this case

$$\mathbb{E}\left[\text{Earnings}|\text{married}=1\right] - \mathbb{E}\left[\text{Earnings}|\text{married}=0\right] = \$2830.20$$

Causal Inference

> *"Causality is not mystical or metaphysical. It can be understood in terms of simple processes, and it can be expressed in a friendly mathematical language."*

Judea Pearl, *Causality*.

Causal Inference

- Causal Inference is not the same as Statistical Inference.
- Inferring cause and effect remains a challenge even when you have data on an entire population.
- To emphasise this we are going to take the "population first" approach again.

Causal Inference

Causal Inference

- Potential outcomes
- Treatment effects
- Selection bias
- Randomisation
- The CIA

Causal Inference

The Problem

- What is the effect of some "treatment" on some outcome of interest compared to what would have happened had the treatment not taken place?

- The fundamental problem is one of missing data: we do not observe the counterfactual outcome.

What is the effect of marriage on earnings?

- Why not simply compare the earnings across marital status? i.e. use the two groups as counterfactuals for each other?

Effect = earnings for married - earnings for unmarried

Observables

We have (population) data on individuals which record.
The *outcome* variable $Y \in \mathbb{R}$
The *treatment* variable $D \in \{0, 1\}$

- To keep things as simple as possible we will stick to the idea that the treatment is binary:
  - either you get it $(D = 1)$
  - or you do not $(D = 0)$.
- However *everything* in these lectures extends in the natural/obvious way to encompass a continuous treatment (a dosage) with $D \in \mathbb{R}$.

Potential Outcomes

For each individual we postulate the existence of two *potential outcomes:*

$$Y(0) \text{ and } Y(1).$$

$Y(0)$: the outcome under no treatment.

$Y(1)$: the outcome under treatment.

The difference

$$Y(1) - Y(0)$$

is *defined* to be the **causal effect** of the treatment.

Causal Inference

Potential Outcomes

Person	D	Potential Outcomes	
1	1	$Y_1(0)$	$Y_1(1)$
2	0	$Y_2(0)$	$Y_2(1)$
3	1	$Y_3(0)$	$Y_3(1)$
4	1	$Y_4(0)$	$Y_4(1)$
5	0	$Y_5(0)$	$Y_5(1)$
6	0	$Y_6(0)$	$Y_6(1)$
7	0	$Y_7(0)$	$Y_7(1)$

Causal Inference

Potential Outcomes

Potential Outcomes

The treatment determines which of the two potential outcomes actually happens and is therefore observed/observable.

$$\begin{tikzcd}
 & D = 0 \longrightarrow \text{we observe } Y(0) \\
\text{Treatment} \arrow[ur] \arrow[dr] & \\
 & D = 1 \longrightarrow \text{we observe } Y(1)
\end{tikzcd}$$

Potential Outcomes

The observed outcome is linked to potential outcomes as follows:

$$Y = Y(1)\,D + Y(0)\,(1 - D)$$

Objects of interest

We are generally interested in the population values of summaries of the distribution of causal effects:

1. The average treatment effect (ATE).

$$ATE = \mathbb{E}\left[Y\left(1\right)\right] - \mathbb{E}\left[Y\left(0\right)\right]$$

2. The average effect treatment on the treated (ATT).

$$ATT = \mathbb{E}\left[Y\left(1\right)|D = 1\right] - \mathbb{E}\left[Y\left(0\right)|D = 1\right]$$

Causal Inference
Potential Outcomes

What do we get if we just compare averages across groups?

Person	D	Observed Outcomes Y	
1	1		$Y_1(1)$
2	0	$Y_2(0)$	
3	1		$Y_3(1)$
4	1		$Y_4(1)$
5	0	$Y_5(0)$	
6	0	$Y_6(0)$	
7	0	$Y_7(0)$	

$$\underbrace{\mathbb{E}\left[Y(1)\,|D=1\right]\text{-}\mathbb{E}\left[Y(0)\,|D=0\right]}_{\text{Obs difference in group averages}}$$

Causal Inference

Potential Outcomes

$\mathbb{E}[Y(1)|D=1] - \mathbb{E}[Y(0)|D=0]$ is not necessarily the same as

The ATE: $\mathbb{E}[Y(1)] - \mathbb{E}[Y(0)]$

The ATT: $\mathbb{E}[Y(1)|D=1] - \mathbb{E}[Y(0)|D=1]$

Causal Inference

Selection bias

Observables and counterfactuals

The following identity is crucial:

$$\underbrace{\mathbb{E}\left[Y\left(1\right)|D=1\right]\text{-}\mathbb{E}\left[Y\left(0\right)|D=0\right]}_{\text{Obs difference in group averages}}=$$

$$\underbrace{\mathbb{E}\left[Y\left(1\right)|D=1\right]\text{-}\mathbb{E}\left[Y\left(0\right)|D=1\right]}_{\text{Ave. effect of treatment on the treated}}+\underbrace{\mathbb{E}\left[Y\left(0\right)|D=1\right]\text{-}\mathbb{E}\left[Y\left(0\right)|D=0\right]}_{\text{Selection effect/bias}}$$

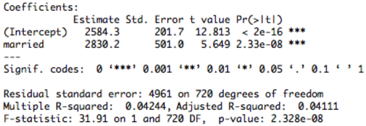
$$\text{Obs. difference in averages} = ATT + \text{ Selection effect/bias}$$

Selection bias

$$\text{Selection effect/bias} = \mathbb{E}\left[Y\left(0\right)\left|D=1\right.\right] - \mathbb{E}\left[Y\left(0\right)\left|D=0\right.\right]$$

Regression and Causal Inference

Selection bias



```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   2584.3      201.7  12.813  < 2e-16 ***
married       2830.2      501.0   5.649 2.33e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4961 on 720 degrees of freedom
Multiple R-squared:  0.04244, Adjusted R-squared:  0.04111 
F-statistic: 31.91 on 1 and 720 DF,  p-value: 2.328e-08
```

- In this model the coefficient on `married` is

$$\mathbb{E}\left[Y|D=1\right]-\mathbb{E}\left[Y|D=0\right]=2830.2$$

- To evaluate whether this is causal you need to think in terms of potential outcomes

$$\beta_1=\mathbb{E}\left[Y\left(1\right)|D=1\right]-\mathbb{E}\left[Y\left(0\right)|D=0\right]$$

and the standard decomposition:

$$\beta_1=ATT+\underbrace{\mathbb{E}\left[Y\left(0\right)|D=1\right]\text{-}\mathbb{E}\left[Y\left(0\right)|D=0\right]}_{\text{Selection effect/bias}}$$

Causal Inference

Selection bias

- On average, men who were married earned \$2830.20 more than mean who were unmarried.
- Is this causal? Does something about the organisation of married partnerships somehow allow men to earn more?

$$\text{Treatment} = \begin{cases} D = 1 \text{ (married)} \\ D = 0 \ \text{ (unmarried)} \end{cases}$$

$$\text{Potential Outcomes} = \begin{cases} Y(0) \\ Y(1) \end{cases}$$

Causal Inference

Selection bias

$$\$2830.20 = \text{Observed difference in average earnings}$$

$$\$2830.20 = ATT + \underbrace{\mathbb{E}\left[Y(0)\,|D=1\right]\text{-}\mathbb{E}\left[Y(0)\,|D=0\right]}_{\text{Selection effect/bias}}$$

$\mathbb{E}\left[Y(0)\,|D=1\right]$: how much would married people have earned, on average, if they hadn't got married?
$\mathbb{E}\left[Y(0)\,|D=0\right]$: how much, on average, do unmarried people earn?

The data do not reveal the counterfactual: how much a married man would earn had they been unmarried.

High potential male earnings may be something which matters in the decision to get married. Men with high earning potential are perhaps more likely to get married

$$\mathbb{E}\left[Y(0)\,|D=1\right]-\mathbb{E}\left[Y(0)\,|D=0\right]>0$$
$$\$2830.20 = ATT + \text{a positive term}$$
$$\text{Obs. difference in averages} > ATT$$

The observed marriage premium could be due to selection in the "marriage market".

Causal Inference
Selection bias

- You observe that, on average, people who study at the University of Oxford are paid £4,000 more five years after graduation than those who attended the University of X.
- Is this causal? Can Oxford use this to argue for charging higher tuition fees than the University of X?

$$\text{Treatment} = \begin{cases} D = 1 \text{ (U of O)} \\ D = 0 \ \text{ (U of X)} \end{cases}$$

$$\text{Potential Outcomes} = \begin{cases} Y(0) \\ Y(1) \end{cases}$$

Causal Inference

Selection bias

$$\text{£}4,000 = \text{Observed difference in average earnings}$$

$$\text{£}4,000 = ATT + \underbrace{\mathbb{E}\left[Y(0)\,|D=1\right] \text{-} \mathbb{E}\left[Y(0)\,|D=0\right]}_{\text{Selection effect/bias}}$$

$\mathbb{E}\left[Y(0)\,|D=1\right]$: how well would a U of O student have done had they gone to U of X?
$\mathbb{E}\left[Y(0)\,|D=0\right]$: how well does someone from U of X do?

The data do not reveal the counterfactual: how much an Oxford grad would have earned had they not been admitted and gone elsewhere.

Oxford selects high ability students with good potential outcomes. U of O students would do better than average at U of X because they are, on average, smarter than students of U of X

$$\mathbb{E}\left[Y\left(0\right)|D=1\right]-\mathbb{E}\left[Y\left(0\right)|D=0\right]>0$$
$$£4,000 = ATT + \text{a positive term}$$
$$\text{Obs. difference in averages} > ATT$$

The observed earning premium could be due to selection of those admitted.

Causal Inference

Selection bias

In the mid 1980's it was observed that young people who had been on a youth training schemes had lower subsequent earnings (by about £8.80pw) than those who hadn't.
Is this causal? Did the scheme harm labour market outcomes? (Some argued it did and that it should be shut down)

$$\begin{aligned}\text{Treatment} &= \begin{cases} D = 1 \text{ (YTS)} \\ D = 0 \quad \text{(Not YTS)} \end{cases} \\ \text{Potential Outcomes} &= \begin{cases} Y(0) \\ Y(1) \end{cases}\end{aligned}$$

Causal Inference

Selection bias

$$-\pounds 8.80 = \text{Observed difference in average earnings}$$

$$-\pounds 8.80 = ATT + \underbrace{\mathbb{E}\left[Y\left(0\right)|D=1\right]\text{-}\mathbb{E}\left[Y\left(0\right)|D=0\right]}_{\text{Selection effect/bias}}$$

$\mathbb{E}\left[Y\left(0\right)|D=1\right]$: how much would a YTS participant have earned had they not gone on the scheme?
$\mathbb{E}\left[Y\left(0\right)|D=0\right]$: how much does someone not on the YTS scheme earn?

The data do not reveal the counterfactual: how much a YTS trainee would have earned had they not been on the scheme.

Causal Inference

Selection bias

YTS participants had extremely poor potential labour market outcomes. They were selected onto the scheme because their potential outcomes were so poor.

$$\mathbb{E}\left[Y\left(0\right)|D=1\right]-\mathbb{E}\left[Y\left(0\right)|D=0\right]<0$$

$$\text{-£8.80}=ATT+\text{a negative term}$$

$$\text{Obs. difference in averages}<ATT$$

This negative earning effect was taken as evidence against the scheme - but it could all be selection and the true ATT could be positive.

Causal Inference

Selection bias

- Selection effects are endemic in observational microeconomic data.
- Selection bias is often related to rational *choice*; if treatment is chosen on the basis of potential outcomes (e.g. to maximise the outcome) then there is *potential* for selection bias.
- Economists tend to see rational, self-interested choice everywhere so are very concerned about selection bias.

Causal Inference

Randomisation

If treatment is *independent* of potential outcomes.

$$D \perp\!\!\!\perp \{Y(0), Y(1)\}$$

it means that

1. the probability of assignment to treatment does not vary with the potential outcomes.
2. the distribution of potential outcomes does not vary with treatment status.

Independence

Two random variables $\{X, Y\}$ are independent

$$X \perp\!\!\!\perp Y$$

if the realisation of one does not affect the probability distribution of the other.

Independence and correlatedness

$$\begin{aligned}
X \text{ and } Y \text{ are independent} &\Rightarrow \mathbb{E}[XY] = \mathbb{E}[Y]E[X] \\
X \text{ and } Y \text{ are independent} &\Rightarrow cov[X, Y] = 0 \\
X \text{ and } Y \text{ are independent} &\Rightarrow corr[X, Y] = 0
\end{aligned}$$

The converse, that if two random variables have a covariance of 0 or are uncorrelated they must be independent, is not true.

Mean-independence

Mean-independence is a notion of independence which lies "between" independence and uncorrelatedness.

$$\mathbb{E}[Y|X] = \mathbb{E}[Y] \Leftrightarrow Y \text{ is mean-independent of } X$$

$$\mathbb{E}[X|Y] = \mathbb{E}[X] \Leftrightarrow X \text{ is mean-independent of } Y$$

Causal Inference

Randomisation

Mean Independence

The following is important to remember

$$\text{Independence} \Rightarrow \text{Mean-independence}$$

$$\text{Independence} \nLeftarrow \text{Mean-independence}$$

Randomisation

Random assignment of treatment ⇒ $D \perp\!\!\!\perp \{Y(0), Y(1)\}$

⇒ mean-independence of the potential outcomes wrt treatment

$$\mathbb{E}\left[Y(0)\,|D\right] = \mathbb{E}\left[Y(0)\right]$$

⇒

$$\mathbb{E}\left[Y(0)\,|D=1\right] = \mathbb{E}\left[Y(0)\,|D=0\right]$$

⇒

Diff in group means $= ATT + \underbrace{\mathbb{E}\left[Y(0)\,|D=1\right]\text{-}\mathbb{E}\left[Y(0)\,|D=0\right]}_{\text{Selection effect/bias}} \leftarrow$ Zero

Randomisation

Randomisation ensures

$$\underbrace{\mathbb{E}\left[Y\left(1\right)|D=1\right]\text{-}\mathbb{E}\left[Y\left(0\right)|D=0\right]}_{\text{Obs diff. between groups}}=\underbrace{\mathbb{E}\left[Y\left(1\right)|D=1\right]\text{-}\mathbb{E}\left[Y\left(0\right)|D=1\right]}_{ATT}$$

Randomisation

Randomisation ensures

$$\underbrace{\mathbb{E}\left[Y\left(1\right)|D=1\right]\text{-}\mathbb{E}\left[Y\left(0\right)|D=0\right]}_{\text{Obs diff. between groups}}=\underbrace{\mathbb{E}\left[Y\left(1\right)\right]\text{-}\mathbb{E}\left[Y\left(0\right)\right]}_{ATE}$$

Causal Inference

Randomised Controlled Trials

- Random assignment of treatment is a mechanism which ensures independence and thus eliminates selection bias from a comparison of observed means across treatment groups.
- Randomised Controlled Trials are therefore usually considered the best possible approach to the study of causal effects.

Causal Inference
Randomised Controlled Trials

- Influential work by Campbell and Stanley (1963), *Experimental and Quasi-Experimental Designs for Research*, distinguished between the internal and external validity of a study of treatment response.
- A study is said to have *internal validity* if its findings for the sample are credible.
- A study is said to have *external validity* if its findings can be credibly extrapolated to the population of interest.

- **Individualistic treatment response/stable unit treatment value assumption (SUTVA)** The experimental ideal only works perfectly if there are no interaction effects between subjects - each person's outcome depends only on his own treatment, not on those received by other members of the study population.

- **Contamination:** People in the control group access the treatment anyway. In other words, some of the untreated turn out to be treated too.
- **Non-compliance:** Individuals who are offered a treatment refuse to take it. In other words, some of the treated turn out to be untreated.

Causal Inference

Threats to Internal Validity

- **Hawthorne Effect**: This refers to a phenomenon in which participants alter their behaviour as a result of being part of an experiment or study.
- **Placebo Effect**: The placebo effect impacts final outcomes because of perceived changes (different from the Hawthorne effect where the outcome changes due to imperceptible changes).

- Properly executed RCT's have high internal validity, but you need external validity too.
- The internal validity of RCT's does not imply external validity.
- Credible policy evaluation requires **both**.

Causal Inference

Threats to External Validity

- If an RCT has external validity it means that we can expect the distribution of outcomes that would occur in the population under the policy of interest would be the same as the distribution of outcomes realised by a specific experimental treatment group.
- But there are often reasons why this is not a credible assumption.

Causal Inference

Threats to External Validity

- The sample used in an RCT may differ from the population of policy interest due to the small scale/local nature of the RCT, (e.g. the trail is carried out in a particular geographic area, institutional environment or demographic group)
- Then establishing external validity requires more work. Essentially enumerative and eliminative induction: doing lots of RCTs and varying the circumstances under which the RCT takes place and seeing which aspects of the environment matter/don't matter to the outcome.

Causal Inference

Threats to External Validity

- The treatments assigned in an RCT may differ from those that would be assigned in actual policies.
- The assumption of individualistic treatment response may be valid within the design of the RCT but may not hold in the population where there is the potential for spillover effects.
- E.g. a vaccine is internally effective if it generates an immune response that prevents a vaccinated person from becoming ill or infectious. It is externally effective to the extent that it prevents transmission of disease to members of the population who are unvaccinated or unsuccessfully vaccinated.

- A serious measurement problem often occurs when studies have short durations.
- We often want to learn long-term outcomes of treatments, but short studies reveal only immediate surrogate outcomes.
- Even well-conducted RCT's may only reveal the distribution of surrogate outcomes in the study population, and this may not be the population of policy interest

Working with observational data

Observational data

- RCTs are not always available to us. This can be because they not feasible or (very often) unethical.
- In these circumstances we need to try to disentangle causal inferences from observational data.
- Observational data are data which are not generated by a randomised controlled trial.

Working with observational data

Quasi- or Natural experiments

- Sometimes institutions or nature provide investigators with variation in treatments which is independent with respect to potential outcomes.
- Examples include
    - the Vietnam War draft
    - localised smoking bans
    - natural disasters
    - school admissions lotteries
- Of course with quasi/natural experiments nature decides what you can study.

Working with observational data

Adding co-variates

- So far the only data we have observed has been outcomes and treatment.
- Suppose that we can now observe a list of other things about the individual (“covariates” or “conditioning variables”).
- As with the NSW data these could be made up of personal characteristics, information on educational attainment, past work history, etc.
- Formally X denotes a list of variables but you’ll be fine thinking of it as just one.
- Now the data are

$$\{Y, D, X\}$$

Working with observational data

Conditional Independence

- The original independence requirement was that assignment to treatment was independent of potential outcomes

  $$D \perp\!\!\!\perp \{Y(0), Y(1)\}$$

  Randomisation delivered this.

Working with observational data

Conditional Independence

- The conditional independence assumption (CIA) says assignment to treatment is independent of potential outcomes *conditional on covariates.*

$$D \perp\!\!\!\perp \{Y(0), Y(1)\} \,|X$$

- In other words potential outcomes are independent of treatment *holding other factors about the individual fixed*.

Working with observational data
Conditional Independence

- Recall that independence ⇒ mean independence

$$D \perp\!\!\!\perp \{Y(0), Y(1)\} \Rightarrow \mathbb{E}\left[Y\left(0\right)|D\right] = \mathbb{E}\left[Y\left(0\right)\right]$$

and so

$$\mathbb{E}\left[Y\left(0\right)|D=1\right] = \mathbb{E}\left[Y\left(0\right)|D=0\right]$$

Working with observational data

Conditional Independence

- Equally, conditional independence ⇒ conditional mean independence

$$D \perp\!\!\!\perp \{Y(0), Y(1)\} \left| X \Rightarrow \mathbb{E}\left[Y(0) \left| D, X\right.\right] = \mathbb{E}\left[Y(0) \left| X\right.\right]\right.$$

and so

$$\mathbb{E}\left[Y(0) \left| D = 1, X = x\right.\right] = \mathbb{E}\left[Y(0) \left| D = 0, X = x\right.\right]$$

Working with observational data

Conditional Independence

- The conditional-on-covariates version of the identity linking observed outcomes to the ATT plus selection bias is

$$\underbrace{\mathbb{E}\left[Y(1)\,|D=1, X=x\right]-\mathbb{E}\left[Y(0)\,|D=0, X=x\right]}_{\text{Obs. difference in averages at } X=x}=$$

$$\underbrace{\mathbb{E}\left[Y(1)\,|D=1, X=x\right]-\mathbb{E}\left[Y(0)\,|D=1, X=x\right]}_{\text{Ave. effect of treatment on the treated at } X=x}$$

$$+\underbrace{\mathbb{E}\left[Y(0)\,|D=1, X=x\right]-\mathbb{E}\left[Y(0)\,|D=0, X=x\right]}_{\text{Conditional selection bias for } X=x}$$

Working with observational data

Conditional Independence

- Using *exactly the same logic as before*, what conditional independence buys us is

$$\mathbb{E}\left[Y\left(0\right)|D=1,X=x\right]=\mathbb{E}\left[Y\left(0\right)|D=0,X=x\right]$$

- This makes the the conditional selection bias term disappear and

$$\left(\text{Obs. difference in averages}|X=x\right)=ATT\left(x\right)=ATE\left(x\right)$$

Working with observational data

Conditional Independence and Empirical Practice

- You might consider dividing the data into cells and looking within each cell to get each $ATE(\mathbf{x})$ and then averaging those over the different cells. But many of those cells would be empty (something called the Curse of Dimensionality)
- Instead, in practice, we "run regressions" in which we include the causal variable of interest plus a set of co-variates.
- These are designed to, or assumed to, "control for" all of the non-random variation in assignment to treatment such that what variation is then left over (a la Frisch-Waugh-Lovell) is plausibly independent of potential outcomes.
- If we can do this credibly, then assignment to treatment is conditionally independent of potential outcomes and we can then measure causal effects.

Regression and Causal Inference

Univariate regression

- The coefficient β_1 in the linear regression model

$$\mathbb{E}\left[Y|D\right] = \beta_0 + \beta_1 D$$

1. always supports a descriptive interpretation as the difference between average observed outcomes between the $D = 1$ group and the $D = 0$ group.
2. only supports the interpretation of the *causal effect* of D on Y if D is *independent* of potential outcomes $\left\{Y\left(0\right), Y\left(1\right)\right\}$ so that selection bias is absent.

- But what if D is not plausibly independent of potential outcomes?
- The CIA may be able to help.
- Suppose that we now have another regressor X and we believe that

$$D \perp\!\!\!\perp \{Y(0), Y(1)\} \, | X$$

- Then we can add X to the LRM:

$$\mathbb{E}\left[Y | D, X\right] = \beta_0 + \beta_1 D + \beta_2 X$$

Regression and Causal Inference

Multivariate regression and the CIA

- Remember the Frisch-Waugh-Lovell theorem:

1. we can get β_1 from $\mathbb{E}\left[Y|\tilde{D}\right] = \beta_0 + \beta_1 \tilde{D}$ where $\tilde{D}$ is the variation in D which is not (linearly) related to variation in the X variable:

$$D = \alpha_0 + \alpha_1 X + \tilde{D}$$

2. we can get β_2 from $\mathbb{E}\left[Y|\tilde{X}\right] = \beta_0 + \beta_2 \tilde{X}$ where $\tilde{X}$ is the variation in X which is not (linearly) related to variation in the D variable:

$$X = \gamma_0 + \gamma_1 D + \tilde{X}$$

- If D is conditionally independent of potential outcomes

  $$D \perp\!\!\!\perp \{Y(0), Y(1)\} \,|X$$

  then $\tilde{D}$ (the variation in D which is not related to variation in the X variable) is independent of potential outcomes

  $$\tilde{D} \perp\!\!\!\perp \{Y(0), Y(1)\}$$

- As a result, due to Frisch-Waugh-Lovell, the multivariate LRM coefficient on D will have a causal interpretation.

Regression and Causal Inference

Multivariate regression and the CIA

$$\mathbb{E}\left[Y|D,X\right] = \beta_0 + \beta_1 D + \beta_2 X$$

- If we fix the value of X at x then

$$\mathbb{E}\left[Y|D=0,X=x\right] = \beta_0 + \beta_2 x$$

$$\mathbb{E}\left[Y|D=1,X=x\right] = \beta_0 + \beta_1 + \beta_2 x$$

- And β_1 is the difference between the $D=1$ and $D=0$ given $X=x$

$$\beta_1 = \mathbb{E}\left[Y|D=1,X=x\right] - \mathbb{E}\left[Y|D=0,X=x\right]$$

Regression and Causal Inference

Multivariate regression and the CIA

- In terms of potential outcomes

$$\beta_1 = \mathbb{E}\left[Y(1)|D=1, X=x\right] - \mathbb{E}\left[Y(0)|D=0, X=x\right]$$

- If the CIA holds then

$$\beta_1 = \mathbb{E}\left[Y(1)|D=1, X=x\right] - \mathbb{E}\left[Y(0)|D=1, X=x\right]$$

- So the coefficient β_1 measures the ATT given $X = x$.

Regression and Causal Inference

Multivariate regression and the CIA

- Furthermore, thanks to the fact that this regression model is purely additive (e.g no interactions) the causal interpretation of β_1 conditional on $X = x$, also holds for any value of X.
- Therefore not only is β_1 equal to the ATT conditional on $X = x$, but it is also the ATT for any value of X and so β_1 captures the unconditional ATT and the ATE.
- We will return to this when we look at heterogeneous treatment effects.

Regression and Causal Inference

Multivariate regression and the CIA

- In observational data the CIA is often used to try to identify a causal effect.
- The trick is essentially to load up with other regressors to "control for" selection on these observables which is non-random and assume/hope that the residual variability in your treatment variable of interest is thus cleansed of selection bias.
- This is not as clean as a well designed and well executed RCT and it does rest on an untestable assumption.

Correlation versus causation: when you're right you're right

Sir Richard Doll vs Ronald Fisher

- Sir Richard Doll (1912-2005) was appointed Regius Professor of Medicine in Oxford in 1969. Oxford's Cancer Epidemiology Unit and Department of Public Health are located in the Richard Doll Building.
- He was credited with being the first to prove that smoking caused lung cancer and increased the risk of heart disease.

Correlation versus causation: when you’re right you’re right

Sir Richard Doll vs Ronald Fisher

- At the end of the Second World War, Britain had the highest incidence of lung cancer in the world – and no one knew why.
- The figures were staggering. Between 1922 and 1947, the number of deaths attributed to lung cancer increased 15-fold across England and Wales.
- Similar trends were documented around the world.

Correlation versus causation: when you're right you're right

Sir Richard Doll vs Ronald Fisher

- The question of whether smoking (or anything else) causes cancer was not one you could or should answer with a randomised control trial.
- So Doll and his coauthor A. Bradford Hill were forced to use observational data and a method which today we would describe as being based on the conditional independence assumption: they controlled for every factor they thought could possibly be relevant and once those were accounted for there was still a significant difference between smokers and non-smokers.

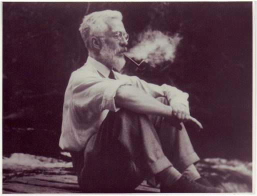
Correlation versus causation: when you're right you're right

Sir Richard Doll vs Ronald Fisher

- Doll and Hill undertook a study of lung cancer patients in 20 London hospitals in 1950.
- Doll himself originally thought the increase in incidence of cancers might be due to occupational factors, or to the tarring of roads, as there were known carcinogens in tar.
- But they rapidly discovering that, after controlling for other influences, tobacco smoking was the only factor these patients had in common.
- They then attributed a causal connection.

Correlation versus causation: when you're right you're right

Sir Richard Doll vs Ronald Fisher



- Ronald Fisher (1890-1962) is sometimes called the father of modern statistics.
- He spent much of his career devising ways statistically to evaluate causal claims and did a great deal to extend the use of RCT's particularly in agriculture.

Correlation versus causation: when you're right you're right

Sir Richard Doll vs Ronald Fisher

- According to Fisher, Doll's paper in the *BMJ* was statistically illiterate fearmongering: surely the danger posed to the smoking masses was "not the mild and soothing weed," he wrote, "but the organised creation of states of frantic alarm."
- Fisher did not dispute that smoking and lung cancer were correlated. But Hill and Doll and the entire British medical establishment had committed "an error. . . of an old kind, in arguing from correlation to causation," he wrote in a letter to *Nature*.

Correlation versus causation: when you're right you're right

Sir Richard Doll vs Ronald Fisher

- Fisher argued that the correlation was also consistent with the opposite conclusion that lung cancer caused smoking.
- What if the development of acute lung cancer was preceded by an undiagnosed "chronic inflammation," he wrote.
- And what if this inflammation led to a mild discomfort, but no conscious pain?
- If that were the case, wrote Fisher, then one would expect those suffering from pre-diagnosed lung cancer to turn to cigarettes for relief.

Correlation versus causation: when you're right you're right

Sir Richard Doll vs Ronald Fisher

- He also offer another explanation: both cancer and smoking could be caused by a third factor. Genetics struck him as a possibility.
- Fisher gathered data on identical twins in Germany and showed that twin siblings were more likely to mimic one another's smoking habits than randomly matched individuals.
- Perhaps, Fisher speculated, certain people were genetically predisposed to crave cigarettes. Was there a similar familial pattern for lung cancer? Did these two predispositions come from the same hereditary trait?

Causal vs Casual inference

Daily Mail

DIANA: WHO THE FUCK STILL CARES?

IMMIGRATION SPIRALS AS WHITEHALL PARTIES

SEE PAGES 12-13

THE MAIL GET UPSET ON YOUR BEHALF

Health minister comes clean and tells Mail readers

CANCER: READING THIS PAPER GIVES YOU DISEASE

Skinny Kiera finally eats

- Causal claims are always contestable.
- Fisher is often painted as the bad guy in this story who came to a sticky end.
- He was wrong about the substantive question, but Fisher never objected to the *possibility* that smoking caused cancer, only the *certainty* with which public health advocates asserted this conclusion.

Summary

- When presented with some regression output you need to be able to carry out two forms of triage:

1. Statistical inference regarding the robustness of the results to sampling variation.
2. Causal inference regarding the robustness of any causal claims either made or implied by the model.