Bush 631-603: Quantitative Methods

Lecture 8 (03.08.2022): Prediction vol. III

Rotem Dvir

The Bush school of Government and Public Policy

Texas A&M University

Spring 2022

What is today's plan?

- Task 1 lessons...
- Predictions: Linear model and causal inference.
- Binary predictors and randomized experiments.
- Multiple predictors, heterogeneous treatment effects.
- Writing: the 5 Ps.
- R work: lm(), levels(), coef().
- Final project prep.

Task 1

- Main issues:
  - Details, details...
  - Use your own words, no direct quoting.
- Internal and external validity:
  - Aspects to evaluate a research design, not results.
- Internal: how design helps answer research question?
  - Strong internal: experiments.
- External: can we generalize the results?
  - Strong: observational studies.

Least squared

THE LINEAR MODEL

$$Y = \alpha + \beta * X_i + \epsilon$$

Elements of model:

- *Intercept* (α): the average value of Y when X is zero.
- *Slope* (β): the average change in Y when X increases by 1 unit.
- *Error/disturbance term* (ϵ): the deviation of an observation from a perfect linear relationship.

Minimize the prediction error

Confused by data?

Regression to the mean - its everywhere

How sure are we?

- What does our model tell us?
- Do the results mean anything?
- **Causal inference**:
  - Predicting the counter-factual.
  - Assumptions $\rightarrow$ use regression models for prediction.

Causal inference

Randomized experiments: women politicians and policy outcomes

Causal inference

QSS example: West Bengal (1990's)

```
dim(women)
```

```
## [1] 322   6
```

```
head(women)
```

```
##   GP village reserved female irrigation water
## 1  1       2        1      1          0    10
## 2  1       1        1      1          5     0
## 3  2       2        1      1          2     2
## 4  2       1        1      1          4    31
## 5  3       2        0      0          0     0
## 6  3       1        0      0          0     0
```

Causal inference

Promoting women's issues

```
## drinking-water facilities
mean(women$water[women$reserved == 1]) -
    mean(women$water[women$reserved == 0])
```

```
## [1] 9.252423
```

```
## Irrigation facilities
mean(women$irrigation[women$reserved == 1]) -
    mean(women$irrigation[women$reserved == 0])
```

```
## [1] -0.3693319
```

Causal inference

Promoting women's issues: regression analysis

```r
# Drinking water model
lm(water ~ reserved, data = women)
```

```
##
## Call:
## lm(formula = water ~ reserved, data = women)
##
## Coefficients:
## (Intercept)     reserved
##      14.738        9.252
```

```r
# Irrigation facilities model
lm(irrigation ~ reserved, data = women)
```

```
##
## Call:
## lm(formula = irrigation ~ reserved, data = women)
##
## Coefficients:
## (Intercept)     reserved
##      3.3879      -0.3693
```

Causal inference

Binary dependent variable:

- slope coefficient (β) $=$ diff-in-means estimator
- $\hat{\beta}$: estimated average treatment effect
- Effect with/without women leaders.

- Why works?
  - Randomization $\rightarrow$ causal interpretation

Distributing foreign aid

US Foreign Aid: 2021

Total obligations: $38B
182 Countries

Main sectors:
Health: $15.75B
Humanitarian Assistance: $10.1B

Main agency:
USAID: 31.66
DoD: $1.79B

Why foreign aid?

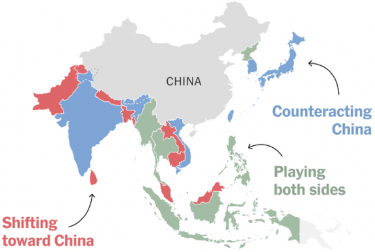
National interest vs. Moral motives

The New York Times

Trump Embraces Foreign Aid to Counter China's Global Influence

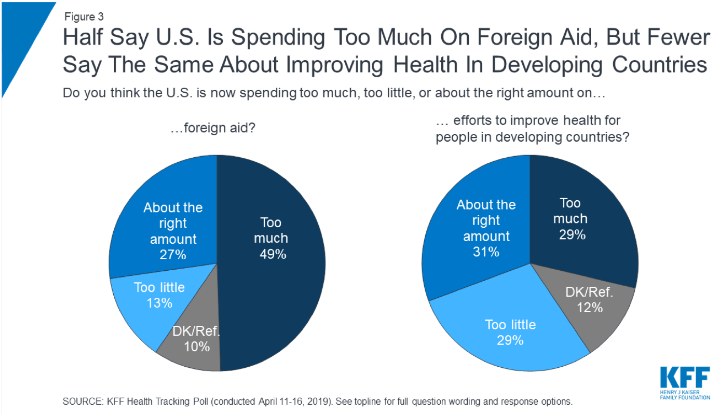
How China Is Challenging American Dominance in Asia

Every Asian country now trades more with China than with the United States, often by a factor of two to one. Here's how the outlines of the rivalry are defining the future of the continent.



Public views of aid

US public opinion on aid (2019)



Public views of aid

Wood, Hoy and Pryke (2020)

- Public attitudes towards foreign aid
- Context → Australia and the Pacific region
- More support for *national interest* objectives?
- Invoke strategic competition - China aid spike in Pacific

Public attitudes towards aid

- Design: Experiment.
- Sample: 2000 Australians (2019-2020).
- Treatments:
  1. Control - no info
  2. Measured - China increases aid to Pacific.
  3. Forceful - China's aid with focus on increased influence.
- Outcome measures:
  1. AUS gives too much.
  2. AUS more aid to Pacific.
  3. Aid focus on AUS or support poor countries.

Foreign aid data

```r
# Our Aussie data
dim(aus)
```

```
## [1] 2001   19
```

```r
# Experimental groups counts ~ equal size
table(aus$treatment_group)
```

```
##
##   1   2   3
## 673 660 668
```

```r
# Experimental groups proportions
prop.table(table(aus$treatment_group))
```

```
##
##         1         2         3
## 0.3363318 0.3298351 0.3338331
```

Foreign aid and public attitudes

General support for main measures

```
# Calculate means across all respondents (tidyverse)
gen.means <- aus %>%
  summarise(Too_much = mean(aus$too_much_aid, na.rm = T),
            Too_little = mean(aus$too_little_aid, na.rm = T),
            more_pac = mean(aus$more_to_pac, na.rm = T),
            Aussie_first = mean(aus$favour_aus, na.rm = T),
            Poor_first = mean(aus$favour_poor, na.rm = T)) %>%
  gather(Measure, mn_prop, Too_much:Poor_first) %>%
  mutate(mn_prop = mn_prop * 100) %>%
  arrange(-mn_prop)

gen.means
```

```
## # A tibble: 5 x 2
##   Measure      mn_prop
##   <chr>          <dbl>
## 1 Aussie_first    54.4
## 2 Too_much        46.0
## 3 Poor_first      45.6
## 4 more_pac        30.5
## 5 Too_little      17.3
```

Foreign aid and public attitudes

- Compare experimental groups: diff-in-means estimator

```
# Diff-in-means estimators: AUS provides too much foreign aid
mean(aus$too_much_aid[aus$treatment_group == 1], na.rm = T) -
  mean(aus$too_much_aid[aus$treatment_group == 2], na.rm = T)
```

```
## [1] 0.07894105
```

```
mean(aus$too_much_aid[aus$treatment_group == 1], na.rm = T) -
  mean(aus$too_much_aid[aus$treatment_group == 3], na.rm = T)
```

```
## [1] 0.0929299
```

```
mean(aus$too_much_aid[aus$treatment_group == 2], na.rm = T) -
  mean(aus$too_much_aid[aus$treatment_group == 3], na.rm = T)
```

```
## [1] 0.01398885
```

Foreign aid and public attitudes

Compare using regression models:

- control and measured conditions
- measured and forceful conditions

```
# Linear model coefficients == diff-in-means estimators
lm(too_much_aid ~ treatment_group, data = aus2)
```

```
##
## Call:
## lm(formula = too_much_aid ~ treatment_group, data = aus2)
##
## Coefficients:
##     (Intercept)  treatment_group
##         0.59671         -0.07894
```

```
lm(too_much_aid ~ treatment_group, data = aus3)
```

```
##
## Call:
## lm(formula = too_much_aid ~ treatment_group, data = aus3)
##
## Coefficients:
##     (Intercept)  treatment_group
##         0.46680         -0.01399
```

Foreign aid and public attitudes

More measures:

- More aid to Pacific region.
- Aid to promote Aussie strategic goals.
- Aid to help poor countries in region.

```
# Diff-in-mens estimators
mean(aus$more_to_pac[aus$treatment_group == 1], na.rm = T) -
  mean(aus$more_to_pac[aus$treatment_group == 2], na.rm = T)
```

```
## [1] -0.05192231
```

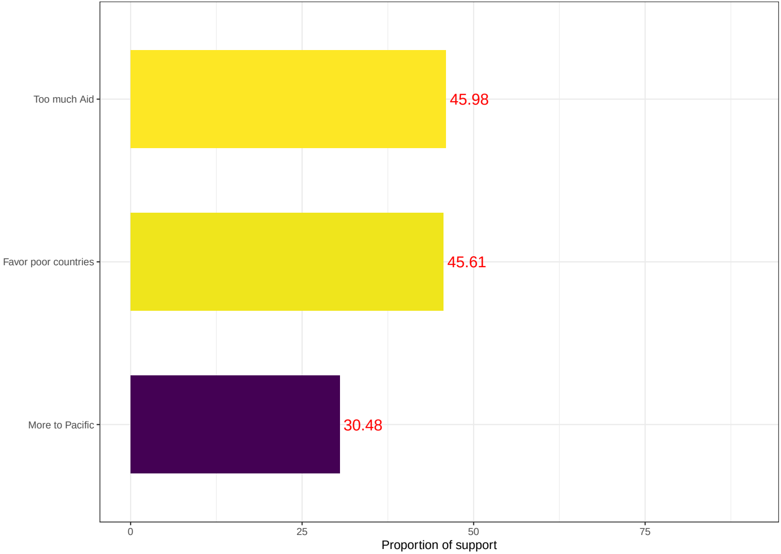
```
mean(aus$favour_aus[aus$treatment_group == 1], na.rm = T) -
  mean(aus$favour_aus[aus$treatment_group == 2], na.rm = T)
```

```
## [1] 0.06338742
```

```
mean(aus$favour_poor[aus$treatment_group == 1], na.rm = T) -
  mean(aus$favour_poor[aus$treatment_group == 2], na.rm = T)
```

```
## [1] -0.06338742
```

Aussies foreign aid views



Binary predictors

Linear model elements:

- *Slope* (β): the average chnage in Y when X increases by 1 unit.

When X is binary:

- Treatment: yes or no (no information or China focus).
- X change by 1 unit $\rightarrow$ no to yes.
- Y (support) changes as well (measured in percentages).

Regression model

Why sanctions fail?

	Likelihood of Success Versus Failure						
	Model 1	*Model 2*	*Model 3*	*Model 4*	*Model 5*	*Model 6*	*Model 7*
Hypothesized Variables							
All Busters	−0.24 (0.12)**	−0.46 (0.15)***					
Black Knight Allies			0.05 (0.25)	−0.08 (0.27)			
Black Knight Great Powers					−0.27 (0.24)	−0.44 (0.40)	
HSE Black Knight							0.03 (0.67)
Control Variables							
US Cooperation		−0.99 (0.57)*		−0.93 (0.57)		−0.84 (0.57)	
IO Support		−2.76 (1.41)*		−2.56 (1.46)*		−2.17 (1.49)	
IO × Coop		1.59 (0.60)***		1.54 (0.61)**		1.37 (0.62)**	
US Defensive Alliance		−0.70 (0.76)		−0.59 (0.72)		−0.73 (0.77)	
Target Defense Alliances		0.00 (0.02)		0.00 (0.02)		0.00 (0.02)	
Modest Goal		1.82 (0.68)***		1.77 (0.68)***		1.73 (0.66)***	
Prior Relations		1.38 (0.46)***		1.37 (0.45)***		1.34 (0.46)***	
Democracy		−0.58 (0.71)		−0.46 (0.68)		−0.31 (0.71)	
Post-Cold War		−0.79 (0.64)		−0.79 (0.61)		−0.74 (0.64)	
Time	−0.08 (−0.18)	0.04 (0.77)	−0.01 (0.69)	−0.11 (0.76)	−0.09 (0.18)	−0.08 (0.76)	−0.11 (0.18)
Time^2	0.00 (−0.01)	0.03 (0.14)	0.04 (0.13)	0.05 (0.14)	0.00 (0.01)	0.05 (0.15)	0.00 (0.01)
Time^3	−0.00 (0.00)	−0.00 (0.01)	−0.00 (0.01)	−0.00 (0.01)	−.00 (.00)	−0.00 (0.01)	0.00 (0.00)
Constant	0.40 (−0.63)	−1.88 (1.59)	−0.77 (1.02)	−3.08 (1.61)*	−.12 (.51)	−2.79 (1.60)*	−0.25 (0.51)
$\text{Prob} > X^2$	0.02	0.00	0.07	0.00	0.02	0.00	0.01
Observations	840	753	789	753	840	753	840

Regression model

Multiple predictors

$$Y = \alpha + \beta_1 * X_1 + \beta_2 * X_2 + ... + \beta_p * X_p + \epsilon$$

How to interpret β_j?

- Change in Y with 1-unit increase in X_j...
- As all other predictors are **held constant**.
- Independent effect of each β.

Least squared: Multiple predictors

Sum of Squared Residuals (SSR)

$$SSR = \sum_{i=1}^{n} \hat{\epsilon}^2 = \sum_{i=1}^{n} (Y_i - \hat{\alpha} - \hat{\beta}_1 * X_1 - \hat{\beta}_1 * X_1 - ... - \hat{\beta}_p * X_p)^2$$

- Estimate parameters: $\hat{\alpha}, \hat{\beta}_p$.
- Minimize SSR.

Foreign aid data

- Multiple predictors for aid support
- Using factor variables: binary outcome

```
### Generate a Factor variable for all groups

aus$grp <- NA
aus$grp[aus$treatment_group == 1] <- "Control"
aus$grp[aus$treatment_group == 2] <- "Measured"
aus$grp[aus$treatment_group == 3] <- "Forceful"

# Check levels of factor
levels(factor(aus$grp))
```

```
## [1] "Control"  "Forceful" "Measured"
```

Multiple binary predictors

$$Y(Support) = \alpha + \beta_1 * Control + \beta_2 * Measured + \beta_3 * Forceful + \epsilon$$

```
fit <- lm(favour_poor ~ factor(grp), data = aus)
fit
```

```
##
## Call:
## lm(formula = favour_poor ~ factor(grp), data = aus)
##
## Coefficients:
##         (Intercept)  factor(grp)Forceful  factor(grp)Measured
##             0.40230              0.09690              0.06339
```

```
mean(aus$favour_poor[aus$grp == "Measured"], na.rm = T) -
  mean(aus$favour_poor[aus$grp == "Control"], na.rm = T)
```

```
## [1] 0.06338742
```

Multiple binary predictors

Coefficients = diff-in-means??

```
# Regression w/o the intercepts
fit3 <- lm(favour_poor ~ -1 + factor(grp), data = aus)
fit3
```

```
##
## Call:
## lm(formula = favour_poor ~ -1 + factor(grp), data = aus)
##
## Coefficients:
##  factor(grp)Control  factor(grp)Forceful  factor(grp)Measured
##              0.4023               0.4992               0.4657
```

Multiple binary predictors

Same with tapply()

```
tapply(aus$favour_poor, aus$grp, mean, na.rm = T)
```

```
##   Control  Forceful  Measured
## 0.4022989 0.4991974 0.4656863
```

Average treatment effect: Control vs. Measured conditions

```
# Using coef() function
coef(fit3)["factor(grp)Control"] - coef(fit3)["factor(grp)Measured"]
```

```
## factor(grp)Control
##        -0.06338742
```

Model fit: multiple predictors

R^2 with multiple predictors $\rightarrow$ Adjusted R^2

Degrees of freedom (DOF):

- How many observations vary ‘freely’?
- DOF: $(n - p - 1) = n - (p + 1)$
- Multiple predictors $\rightarrow$ larger R^2
- Large sample (data) $\rightarrow$ not much difference b-w R^2 and adjusted R^2

Model fit: multiple predictors

R^2 and adjusted R^2 in regression model

```r
# summary() model with multiple predictors
summary(lm(favour_poor ~ grp + urban + hhold_income + academic, data = aus))
```

```
##
## Call:
## lm(formula = favour_poor ~ grp + urban + hhold_income + academic,
##     data = aus)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.6335 -0.4465 -0.3319  0.5172  0.6962
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.998e-01  3.635e-02   8.248 3.23e-16 ***
## grpForceful  1.146e-01  2.929e-02   3.911 9.55e-05 ***
## grpMeasured  6.253e-02  2.942e-02   2.125   0.0337 *
## urban        2.812e-02  3.162e-02   0.889   0.3740
## hhold_income 1.984e-07  2.373e-07   0.836   0.4032
## academic     1.464e-01  2.564e-02   5.708 1.35e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4907 on 1673 degrees of freedom
##   (322 observations deleted due to missingness)
## Multiple R-squared:  0.03317,    Adjusted R-squared:  0.03028
## F-statistic: 11.48 on 5 and 1673 DF,  p-value: 6.477e-11
```

Heterogenous treatment effects

- Variation in effect of main predictor
- When?
- ATE vary among individuals: positive/negative
- Experiments: differences guide treatment assignment

Aussie foreign aid:

- Respondents' age and views of aid
- Do older respondents' support certain type of aid?

Heterogenous treatment effects

Aid to Pacific by respondents **age** category (over/under 50)

```
# Subset of over-50 respondents
aus.age <- subset(aus, over_fifty == 1)

# Diff-in-means: support for aid by groups
mean(aus.age$more_to_pac[aus.age$treatment_group == 1], na.rm = T) -
  mean(aus.age$more_to_pac[aus.age$treatment_group == 2], na.rm = T)
```

```
## [1] -0.04676688
```

```
# Subset of under-50 respondents
aus.age2 <- subset(aus, over_fifty == 0)

# Diff-in-means: support for aid by groups
mean(aus.age2$more_to_pac[aus.age2$treatment_group == 1], na.rm = T) -
  mean(aus.age2$more_to_pac[aus.age2$treatment_group == 2], na.rm = T)
```

```
## [1] -0.05992362
```

Estimated ATE

```r
# Estimated treatment effect for age (over/under 50) by group
(mean(aus.age$more_to_pac[aus.age$treatment_group == 1], na.rm = T) -
  mean(aus.age$more_to_pac[aus.age$treatment_group == 2], na.rm = T)) -
(mean(aus.age2$more_to_pac[aus.age2$treatment_group == 1], na.rm = T) -
  mean(aus.age2$more_to_pac[aus.age2$treatment_group == 2], na.rm = T))
```

```
## [1] 0.01315674
```

```r
# Estimated treatment effect for age (across groups)
mean(aus$more_to_pac[aus$over_fifty == 1], na.rm = T) -
  mean(aus$more_to_pac[aus$over_fifty == 0], na.rm = T)
```

```
## [1] 0.0884818
```

- Older respondents are more supportive of aid to pacific (8% overall, 1% by experimental groups)

Regression model: conditional effects

- Add predictor to the model

$$Y(Support) = \alpha + \beta_1 * Treatment + \beta_2 * RespondentGender + \epsilon$$

- However, *conditional effect* $\rightarrow$ Interaction model

$$Y(Support) = \alpha + \beta_1 * Treatment + \beta_2 * RespondentGender + \beta_3 * Treatment * RespondentGender + \epsilon$$

Interaction models

$$Y = \alpha + \beta_1 * X_1 + \beta_2 * X_2 + \beta_3 * X_1 * X_2 + \epsilon$$

- Coefficient β_3: How X_1 depends on X_2.
- Average effect of men respondents (and experimental group): $\beta_2 + \beta_3$.
- Average effect of women respondents: β_2.

Interaction model in R

Syntax: use the (*) or (:) between factors

```r
# Interaction model: gender and treatment group
summary(lm(favour_poor ~ treatment_group * male, data = aus2))
```

```
##
## Call:
## lm(formula = favour_poor ~ treatment_group * male, data = aus2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.4937 -0.4358 -0.3973  0.5642  0.6027
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)           0.32021    0.06219   5.149 3.05e-07 ***
## treatment_group       0.08673    0.03935   2.204   0.0277 *
## male                  0.03850    0.08973   0.429   0.6679
## treatment_group:male -0.04818    0.05670  -0.850   0.3957
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.495 on 1217 degrees of freedom
##   (112 observations deleted due to missingness)
## Multiple R-squared:  0.005842,   Adjusted R-squared:  0.003392
## F-statistic: 2.384 on 3 and 1217 DF,  p-value: 0.06775
```

Interaction model: continous predictors

- How the average treatment effect varies along age scale?
- Linearity assumption: one-unit increase in predictor $\rightarrow$ similar increase in outcome.

- Data: ICB (observational).
- Variables:
  - International crises: 1918-2015.
  - Y: Crisis management technique (how to respond).
  - X_1: Trigger event severity/type
  - X_2: Leaders' age.
  - Model: how response varies based on tirgger event (and leader's age).

Interaction model: ICB data

$$CrisisAction = \alpha + \beta_1 * Trigger + \beta_2 * Age + \beta_3 * Trigger * Age + \epsilon$$

	cracid	actor	systrgyr	systrgda	crisname	leader	cris_date	triggr	crismg	lead_age
1	2	USA	1937	12	PANAY INCIDENT	Roosevelt, F.	12/12/37	9	1	55
2	2	USA	1946	7	TURKISH STRAITS	Truman	8/7/46	2	4	62
3	2	USA	1947	21	TRUMAN DOCTRINE	Truman	2/21/47	2	4	63
4	2	USA	1948	24	BERLIN BLOCKADE	Truman	6/24/48	3	4	64
5	2	USA	1948	23	CHINA CIVIL WAR	Truman	9/23/48	8	1	64
6	2	USA	1950	25	KOREAN WAR I	Truman	6/25/50	8	8	66
7	2	USA	1950	30	KOREAN WAR II	Truman	9/30/50	9	8	66
8	2	USA	1953	16	KOREAN WAR III	Eisenhower	4/16/53	9	7	63
9	2	USA	1953	12	GUATEMALA	Eisenhower	12/12/53	7	4	63
10	2	USA	1954	13	DIEN BIEN PHU	Eisenhower	3/13/54	2	1	64
11	2	USA	1954	3	TAIWAN STRAIT I	Eisenhower	9/3/54	8	4	64
12	2	USA	1956	29	SUEZ NATN.-WAR	Eisenhower	10/29/56	5	6	66
13	2	USA	1957	18	SYRIA/TURKEY CONFRNT.	Eisenhower	8/18/57	2	4	67
14	2	USA	1958	8	IRAQ/LEB. UPHEAVAL	Eisenhower	5/8/58	2	6	68
15	2	USA	1958	17	TAIWAN STRAIT II	Eisenhower	7/17/58	8	1	68
16	2	USA	1958	27	BERLIN DEADLINE	Eisenhower	11/27/58	2	1	68
17	2	USA	1961	9	PATHET LAO OFFENSIVE	Kennedy	3/9/61	8	1	44
18	2	USA	1961	15	BAY OF PIGS	Kennedy	4/15/61	2	5	44

Interaction model: ICB data

Outcome - crisis management method:

- Negotiation, mediation
- Non-military pressure (economic)
- Non-violent military
- Violence

Predictor - triggering event: Verbal/political act, violent act.

```
summary(mydata$lead_age)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   18.00   48.00   56.00   55.84   64.00   91.00       2
```

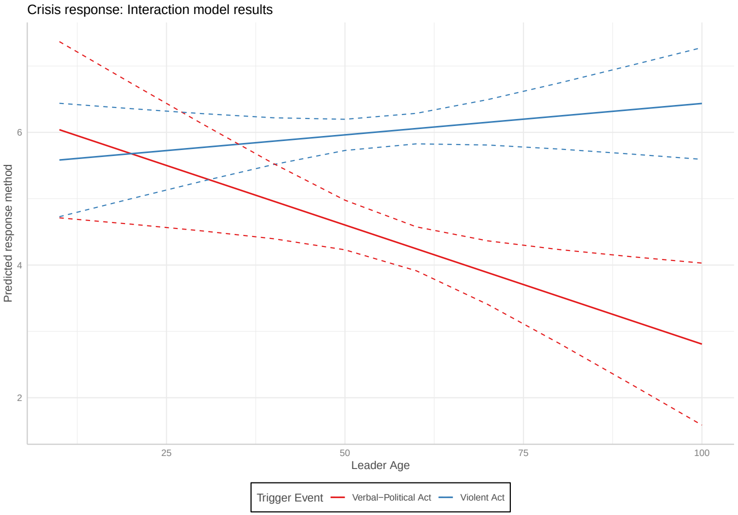
Interaction model: ICB data

```
summary(fit.age <- lm(crismg ~ triggr * lead_age, data = mydata))
```

```
##
## Call:
## lm(formula = crismg ~ triggr * lead_age, data = mydata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.2086 -1.6012  0.9619  1.8246  4.0730
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      6.512835   0.935138   6.965 6.24e-12 ***
## triggr          -0.113761   0.134857  -0.844  0.39913
## lead_age        -0.041579   0.016074  -2.587  0.00984 **
## triggr:lead_age  0.005672   0.002337   2.427  0.01541 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.387 on 927 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.06487,    Adjusted R-squared:  0.06184
## F-statistic: 21.44 on 3 and 927 DF,  p-value: 1.984e-13
```

Interaction model: ICB data

Heterogeneous treatment effects: trigger over age



Writing professional docx

THE 5PS

- **People**: who is my audience?
- Purpose: what is the goal of my product?
- **Problem**: what is the topic / issue at-stake.
- **Product**: what product am I preparing?
- Process: what's the plan?

The 5Ps: people

Who are my readers / audience?

- Guiding questions:
  - Who are they?
  - How much they know about the topic?
  - How open are they to the message?
  - How can I provide value?

The 5Ps: people

Example: study water access in Chad

- Audience: state dept. officials
  - General background.
  - Wide range of solutions.
  - Local offices that manage project.
- Audience: NGO group ("Grant Water in Chad")
  - Know about the issues.
  - Focused solutions.
  - Local people that can implement effectively.

The 5Ps: problem

- Clear definition of the issue and scope.
- Offer problem statement.
- Background.
- Diverse and high-quality supportive data.
- Systematic analysis of data.
- Draw useful conclusions that address the problem.

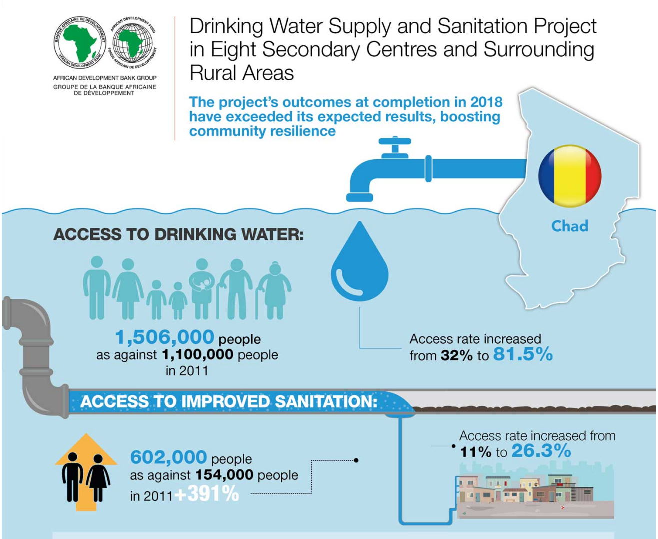
Example:Chad

The 5Ps: Product

- Clear message.
- Integrate high-quality, reliable data.
- BLUF.
- Organization and logic of content.
- Visuals for emphasis.

The 5Ps: Product

Infographic to present project



Causality with observational data

- The problem of *free riding*

Leaders and alliance contribution

Business experience and military alliances (Fuhrmann 2020):

- Leader experience explain variations.
- Business: executive level.
- Smaller contributions (defense expenditures), Why?
- Egoistic tendencies.
- Belief in self-efficacy and power.

Our goals:

1. Evaluate casual effect with linear regression (Δ spending per year).
2. Run *placebo test*: strengthen the proposed causal links.

Alliance contribution

NATO Defense spending data (1949-2020)

```
head(matt1)
```

```
## # A tibble: 6 x 74
##   Country ccode  `1949`  `1950`  `1951`  `1952`  `1953`  `1954`  `1955`  `19
##   <chr>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <d
## 1 Canada     20      NA   3809.   7718.  12405.  14234.  13242.  13113.  133
## 2 USA         2 147593. 158620. 339387. 478080. 492223. 424699. 402015. 4072
## 3 Czechia   316      NA      NA      NA      NA      NA      NA      NA
## 4 Hungary   310      NA      NA      NA      NA      NA      NA      NA
## 5 Poland    290      NA      NA      NA      NA      NA      NA      NA
## 6 Belgium   211   2074.   2092.   3095.   4574.   4554.   4698.   3891.   37
## # ... with 64 more variables: 1957 <dbl>, 1958 <dbl>, 1959 <dbl>, 1960 <dbl>
## #   1961 <dbl>, 1962 <dbl>, 1963 <dbl>, 1964 <dbl>, 1965 <dbl>, 1966 <dbl>,
## #   1967 <dbl>, 1968 <dbl>, 1969 <dbl>, 1970 <dbl>, 1971 <dbl>, 1972 <dbl>,
## #   1973 <dbl>, 1974 <dbl>, 1975 <dbl>, 1976 <dbl>, 1977 <dbl>, 1978 <dbl>,
## #   1979 <dbl>, 1980 <dbl>, 1981 <dbl>, 1982 <dbl>, 1983 <dbl>, 1984 <dbl>,
## #   1985 <dbl>, 1986 <dbl>, 1987 <dbl>, 1988 <dbl>, 1989 <dbl>, 1990 <dbl>,
## #   1991 <dbl>, 1992 <dbl>, 1993 <dbl>, 1994 <dbl>, 1995 <dbl>, 1996 <dbl>,
```

Leaders and military alliances expenditures

NATO leaders and defense spending data

ccode \textsc{COW numeric country code}	year \textsc{year}	leadername \textsc{leader name}	business \textsc{business experience}	Country	def.exp	def.delta
2	2003	G.W. Bush	1	USA	612232.612	13.81651492
2	2004	G.W. Bush	1	USA	667284.639	8.99201159
2	2005	G.W. Bush	1	USA	698019.039	4.60589054
2	2006	G.W. Bush	1	USA	708077.303	1.44097276
2	2007	G.W. Bush	1	USA	726971.529	2.66838457
2	2008	G.W. Bush	1	USA	779854.123	7.27436936
2	2009	Obama	0	USA	841220.473	7.86895241
2	2010	Obama	0	USA	865268.025	2.85865034
2	2011	Obama	0	USA	855022.313	-1.18410840
2	2012	Obama	0	USA	807530.267	-5.55448034
2	2013	Obama	0	USA	745415.975	-7.69188406
2	2014	Obama	0	USA	699563.842	-6.15121420
20	1949	St. Laurent	0	Canada	*NA*	*NA*
20	1950	St. Laurent	0	Canada	3808.656	*NA*
20	1951	St. Laurent	0	Canada	7718.028	102.64439720
20	1952	St. Laurent	0	Canada	12404.681	60.72344453
20	1953	St. Laurent	0	Canada	14234.412	14.75032982

Testing a causal mechanism

Does business experience matter?

```
# subsets by business experience
no.business <- subset(def.matt, subset = (business == 0))
business <- subset(def.matt, subset = (business == 1))

## Diff-in-means estimator
mean(business$def.delta, na.rm = T) -
  mean(no.business$def.delta, na.rm = T)
```

```
## [1] -2.134511
```

```
# Regression model
lm(def.delta ~ business, data = def.matt)
```

```
##
## Call:
## lm(formula = def.delta ~ business, data = def.matt)
##
## Coefficients:
## (Intercept)     business
##       2.847       -2.135
```

The Placebo test

- Data: non-defense related expenses
- Business experience matters → not on other issues.

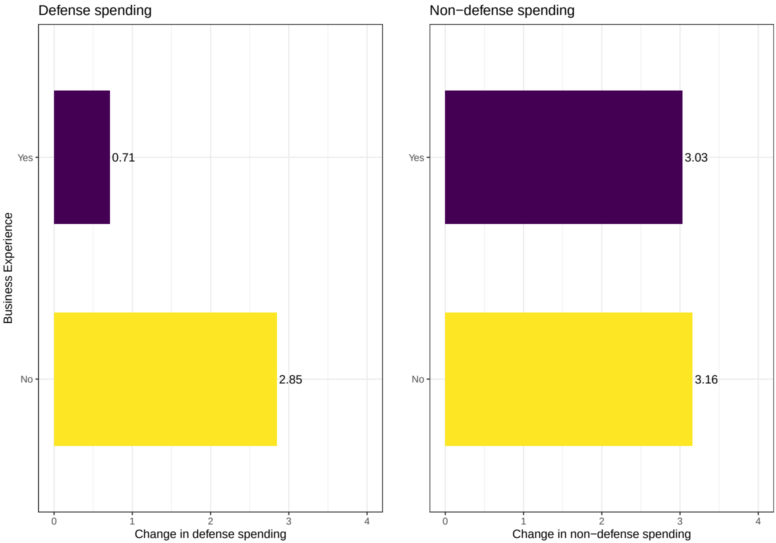
```r
## Diff-in-means estimator: non-defense spending
mean(business$nondefspend_ch, na.rm = T) -
  mean(no.business$nondefspend_ch, na.rm = T)
```

```
## [1] -0.1239881
```

```r
# Regression model
lm(nondefspend_ch ~ business, data = def.matt)
```

```
##
## Call:
## lm(formula = nondefspend_ch ~ business, data = def.matt)
##
## Coefficients:
## (Intercept)     business
##       3.164       -0.124
```

Businessmen, politicians and spending



Wrapping up week 8

Summary:

- Prediction and causal inference.
- Binary predictors and linear regression models.
- Multiple predictors.
- Heterogeneous effects: interaction models.
- Causal inference with observational data.

Final project

- Data: choose one (7 total).
- Proposal: single document with study objectives and plan.
- Data report: focus on data set you selected.