Bush 631-603: Quantitative Methods

Lecture 5 (02.14.2023): Measurement vol. II

Rotem Dvir

The Bush school of Government and Public Policy

Texas A&M University

Spring 2023

What is today's plan?

- More on measurement.
- Latent concepts.
- Visuals: scatterplots.
- Final project prep: data collection (Wendi Kasper).
- Correlation.
- R work: scatterplot, subset(), cor()

Measurement

Why?

- ▶ Social science: develop and test causal theories.
- ▶ Leader background and conflict behavior.
- ▶ Minimum wage and levels of full-time employment?
- ▶ Concepts: level of unemployment, leader background, public approval.

How?

Measures - the context of theoretical concepts

Complex measurement

Latent concepts:

- Hard to measure.
- Variation in definitions.
- Democracy: the polity debate.
- Ideology scale.

A new suspect:

- Terrorism: which violent events are terrorism?

What is terrorism?

Government $\rightarrow$ the objectives/outcomes of violence.

Researchers $\rightarrow$ objective measures:

- Identity: perpetrators and victims.
- Population-wide psychological effects.
- Clear political objective.

The Public?

You tell me

Public views of terrorism?

Huff and Kertzer (2018):

- ▶ Objective: 'facts on the ground'.
- ▶ Subjective: 'who and why?'

The Method: Conjoint experiment

- ▶ No control group.
- ▶ Multiple treatments.
- ▶ Outcome: is it terrorism? (yes/no)
- ▶ How each factor contributes to viewing an incident as terrorism?

Conjoint experiment: Terrorism

Scenario 1
The incident: shooting
The incident occurred in a church in a foreign democracy with a history of human rights violation
Two individuals died.
The shooting was carried by a Muslim individual with history of mental illness.
News suggest the individual had ongoing personal dispute with one of the targets

Scenario 2
The incident: bombing
The incident occurred in a police station in a foreign dictatorship.
No fatalities reported.
The bombing was carried by a Muslim organization.

News suggest the group was motivated by the goal of overthrowing the government.

Objective path: results

(a) Extremity of tactics

(b) Severity of violence
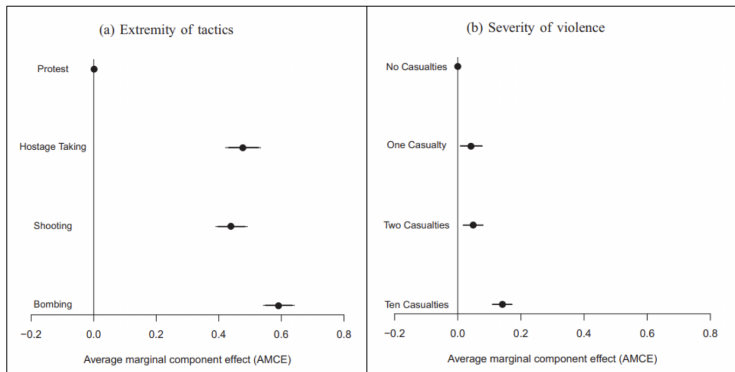
Subjective path: results

FIGURE 5 Social Categorization Effects

No Ideology
Christian
Muslim
Left-Wing
Right-Wing

Average marginal component effect (AMCE)
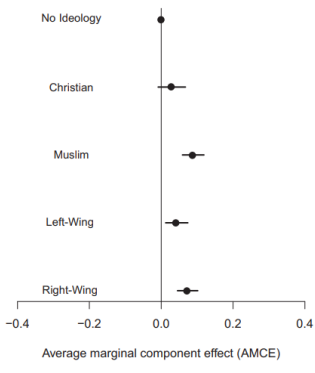
FIGURE 6 Motive Attribution Effects

Personal Dispute
Unclear Motivation
Hatred
Policy Change
Government Overthrow

Average marginal component effect (AMCE)
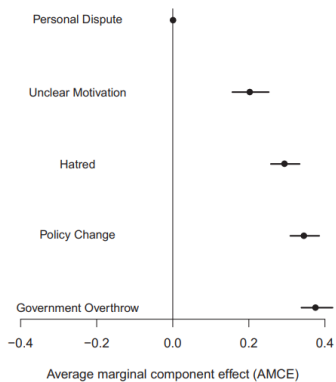
Terrorism data

Type: event data

A lot of resources:

- GTD - START (Maryland).
- Individuals radicalization (PRIUS) - START (Maryland).
- Episodes of political violence (1946-2017) (Vienna, Austria).
- Suicide terrorism - CPOST (Chicago)
- List (Link)

Terrorism data

Global Terrorism Database (GTD):

- ▶ Time frame: 1970-2019.
- ▶ Events: International & domestic terrorism.
- ▶ Scope: over 100,000 cases.
- ▶ Sources: open source media.

Problem(s)?

- ▶ Events data → news sources.
- ▶ Temporal: less work prior to 1970.
- ▶ Biased and Selective reporting: strategic, sensational events.
- ▶ Errors in measurement.
- ▶ Measures matter - democracy and frequency of incidents (polity, strategic reporting).

Measuring ideology
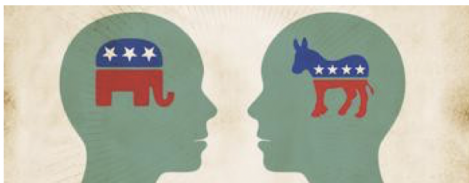
On a scale from 1 to 7, where 1 is extremely liberal, 7 is extremely conservative, and 4 is exactly in the middle, where would you place yourself?

Extremely liberal 1	2	3	In the middle 4	5	6	Extremely conservative 7
○	○	○	○	○	○	○

Measurement models:

- ▶ Summarize data.
- ▶ Learn about human behavior.

Measuring ideology

Legislators measurement model: congress roll-call votes

Voting $\rightarrow$ political orientation.

Complex concepts & measurement

What's the bottom-line?

- ► Latent concepts: democracy, ideology, terrorism.
- ► Tricky measurement: conjoint experiment, measurement models.

How to improve measures?

- ► Theoretical grounding.
- ► Replications.

Bivariate Relationships

Summarize relationship b-w 2 variables

Liberal-conservative ideology: Economy & Race

```
head(congress)
```

```
##   congress district   state   party       name dwnom1 dwnom2
## 1       80        0     USA Democrat     TRUMAN -0.276  0.016
## 2       80        1 ALABAMA Democrat  BOYKIN F. -0.026  0.796
## 3       80        2 ALABAMA Democrat   GRANT G. -0.042  0.999
## 4       80        3 ALABAMA Democrat ANDREWS G. -0.008  1.005
## 5       80        4 ALABAMA Democrat   HOBBS S. -0.082  1.066
## 6       80        5 ALABAMA Democrat   RAINS A. -0.170  0.870
```

Back to visuals

- ▶ Visualize relationship between 2 variables.
- ▶ Numeric/continuous values.

The 80th Congress
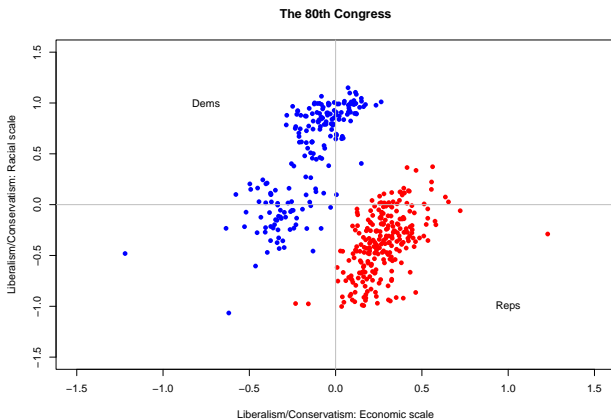
Congress ideology in the 21st century

The 112th Congress
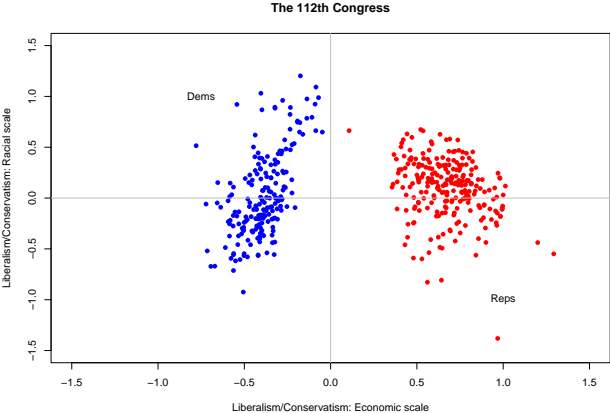
Congress ideology: time trend

```
median_dw1 <- congress %>%
  filter(party %in% c("Republican","Democrat")) %>%
  group_by(party,congress) %>%
  summarise(median_dw1 = median(dwnom1))

ggplot(median_dw1, aes(x=congress,y=median_dw1,color = party)) +
  geom_line(size = 2.2) + xlab("Congress") +ylab("DW-NOMINATE score") +
  scale_color_manual(values = c("blue","red")) + theme_bw()
```
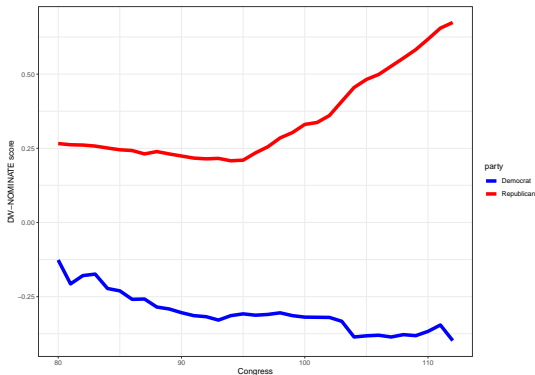
'International' Ideology

UN $\rightarrow$ International institution.

Voting patterns $\rightarrow$ countries orientation/ideology.

UN voting data (1946-2012)

```
dim(mydata)
```

```
## [1] 9120    6
```

```
summary(mydata)
```

```
##      Year       CountryAbb         CountryName         idealpoint
## Min.   :1946   Length:9120        Length:9120        Min.   :-2.6552
## 1st Qu.:1972   Class :character   Class :character   1st Qu.:-0.6406
## Median :1987   Mode  :character   Mode  :character   Median :-0.1644
## Mean   :1985                                         Mean   : 0.0000
## 3rd Qu.:2001                                         3rd Qu.: 0.7968
## Max.   :2012                                         Max.   : 3.0144
##
##    PctAgreeUS      PctAgreeRUSSIA
## Min.   :0.0000   Min.   :0.0000
## 1st Qu.:0.1395   1st Qu.:0.5053
## Median :0.2400   Median :0.6567
## Mean   :0.2960   Mean   :0.6219
## 3rd Qu.:0.3902   3rd Qu.:0.7424
## Max.   :1.0000   Max.   :1.0000
## NA's   :1        NA's   :5
```

Global ideologies

Voting with US $\rightarrow$ measure of foreign policy similarity.

Similar FP $\rightarrow$ similar global orientation.

```r
# Tidyverse approach to data management
# Arrange by year, calculate mean for US / Russia voting
annual.agree <- mydata %>%
  group_by(Year) %>%
  summarize(us.agree = mean(PctAgreeUS, na.rm = T),
            ru.agree = mean(PctAgreeRUSSIA, na.rm = T))

head(annual.agree)
```

```
## # A tibble: 6 x 3
##    Year us.agree ru.agree
##   <int>    <dbl>    <dbl>
## 1  1946    0.585    0.362
## 2  1947    0.621    0.383
## 3  1948    0.578    0.279
## 4  1949    0.541    0.377
## 5  1950    0.635    0.312
## 6  1951    0.487    0.402
```

Trends in global ideology

```
ggplot(data = annual.agree) +
  geom_line(mapping = aes(x = Year, y = us.agree), color = "blue", size = 1.1) +
  geom_line(mapping = aes(x = Year, y = ru.agree), color = "red", size = 1.1) +
  geom_text(aes(x = 2000, y = 0, label = "Voting with US"), color = "blue") +
  geom_text(aes(x = 2000, y = 1, label = "Voting with Russia"), color = "red") +
  geom_vline(aes(xintercept = 1989), linetype = "dotdash", color = "black") +
  geom_text(aes(x = 1993, y = 0.5, label = "Cold War Ends"), color = "black") +
  ylab("Proportion voting with Superpower") + theme_classic()
```
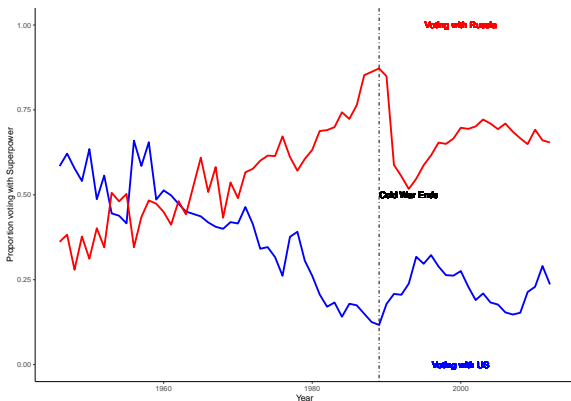
Grouping observations

Which side are you on?

Grouping countries: FP Similarity measures

```
# Table for voting close to US
# USA
mydata %>%
  group_by(CountryName) %>%
  summarise(mean.pctUS = mean(PctAgreeUS)) %>%
  arrange(desc(mean.pctUS)) %>%
  head(n = 11) %>%
  filter(CountryName != "United States of America")
```

```
## # A tibble: 10 x 2
##    CountryName                     mean.pctUS
##    <chr>                           <dbl>
##  1 Palau                           0.736
##  2 United Kingdom                  0.652
##  3 Taiwan                          0.643
##  4 Israel                          0.640
##  5 Federated States of Micronesia  0.594
##  6 Canada                          0.586
##  7 Luxembourg                      0.571
##  8 Netherlands                     0.562
##  9 Belgium                         0.562
## 10 France                          0.549
```

Visualizing distributions

QUNATILE QUNATILE PLOT

Scatter-plot of quantiles

```
### Q-Q plot
qqplot(mydata$PctAgreeUS, mydata$PctAgreeRUSSIA, xlab = "UN voting with US",
       ylab = "UN voting with Russia",
       main = "UN voting with superpower: trend over time")
abline(0,1)
```
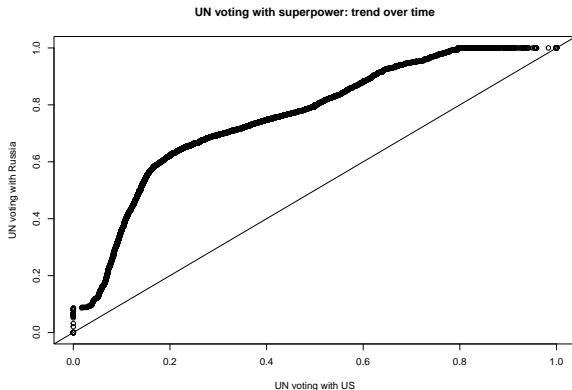
Political polarization: QSS textbook

Income inequality $\rightarrow$ political polarization.

The *Gini coefficient*

Cumulative share of people from lowest to highest incomes
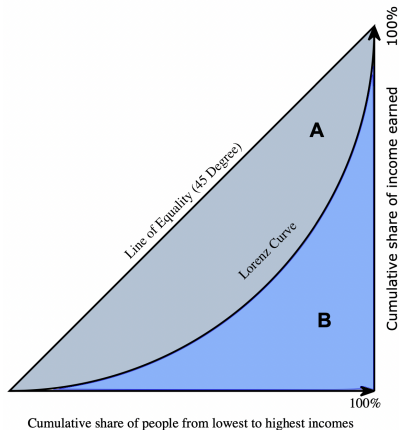
Income inequality measures

Gini coefficient: 2018-2021 data (OECD website)
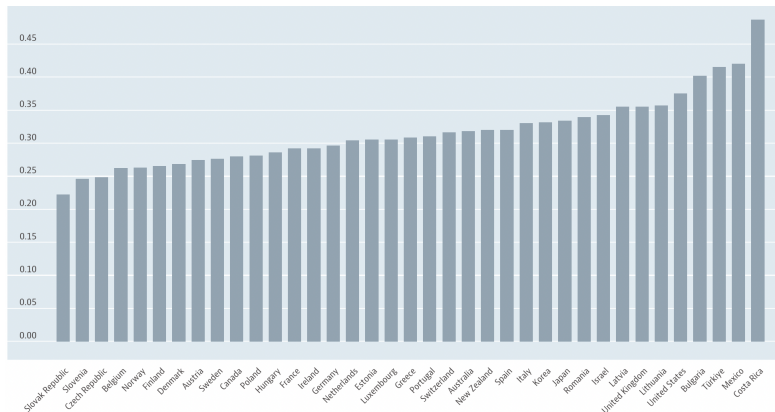
US test case

Gini coefficient - Political Polarization
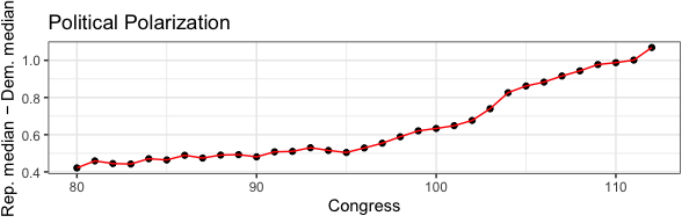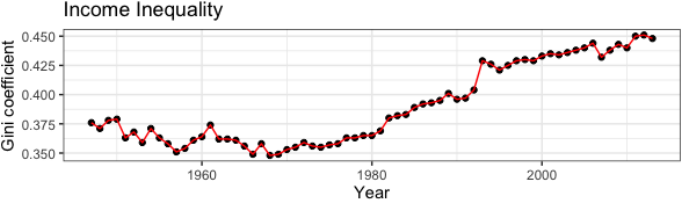
Association b-w variables: Correlation

Income inequality $\rightarrow$ Political polarization?

Correlation does not mean causation

Correlation & causality
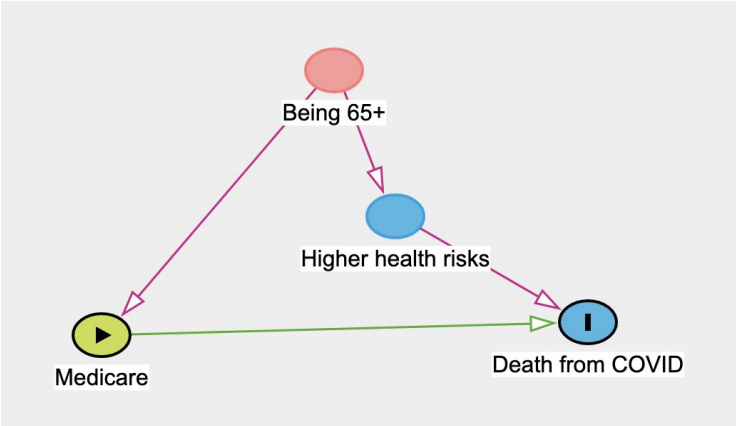
Association b-w variables

Correlation:

- Summary of bivariate relationship.
- How two factors 'move together' on average.
- Always relative to mean value.

Product of z-scores:

$$cor(x, y) = \frac{1}{n} \sum_{i=1}^{n} (Z - x_i * Z - y_i)$$

Z-scores

- A measure for the deviation from the mean (in SD terms)
- Standardize variable
- Allows comparison with *common units*

$$Zscore(X_i) = \frac{x_i - \bar{x}}{SD(X_i)}$$

Z score $> 0 \rightarrow$ unit larger than mean
Z score $< 0 \rightarrow$ unit smaller than mean

z-score example: Test scores

Where do we stand versus our cohort?

- Total of 500 students
- Mean grade ($\bar{X} = 85$)
- SD ($\sigma = 6$)

```
# Our grades = 81, 90, 65
z1 <- (81-85)/6
z1
```

```
## [1] -0.6666667
```

```
z2 <- (90-85)/6
z2
```

```
## [1] 0.8333333
```

```
z3 <- (65-85)/6
z3
```

```
## [1] -3.333333
```

Correlation

- Average product of z-scores:
  - Positive correlation: when x is bigger than its mean, so is y
  - Negative correlation: when x is bigger than its mean, y is smaller
- z-score: not sensitive to unit used
- Correlation is identical even for different measuring units of variable

Correlation - how do the data look?

POSITIVE CORRELATION
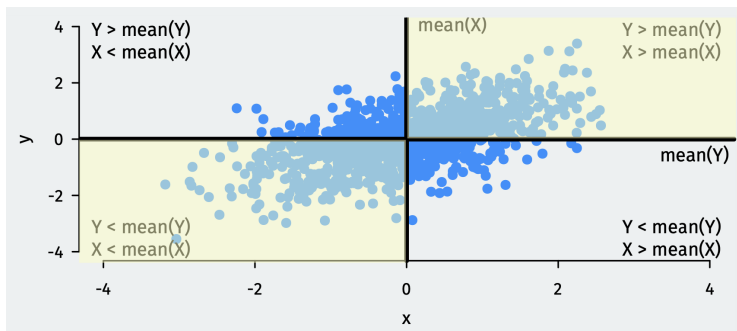
Correlation - how do the data look?

NEGATIVE CORRELATION
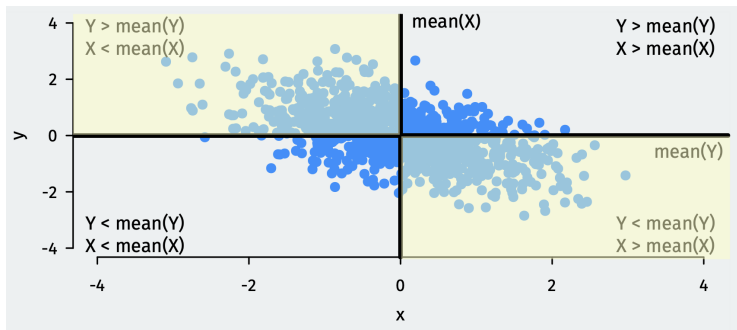
Correlation

- Measures **linear** association

- Order does not matter: cor(x,y) = cor(y,x)

- Interpretation:
    - Values range between (-1) to 1.
    - Close to 'edges' $\rightarrow$ stronger association.
    - Value of zero $\rightarrow$ no association.
    - Positive correlation $\rightarrow$ positive association.
    - Negative correlation $\rightarrow$ negative association.

Correlation in R

UN Voting: association b-w ideal point & liberal FP approach

```
# Voting with US
cor(mydata$idealpoint, mydata$PctAgreeUS, use = "pairwise")

## [1] 0.7498446

# Voting with Russia
cor(mydata$idealpoint, mydata$PctAgreeRUSSIA, use = "pairwise")

## [1] -0.7050107
```

Visualize Correlations: FP Similarity measures

```
ggplot(cor_dat, aes(x=mn1,y=mn2)) +
  geom_point() + xlab("Liberal FP measure - Avg.") + ylab("Mean voting with US") +
  geom_label(aes(label = CountryName), size = 3, fill = "yellow") +
  scale_y_continuous(labels = scales::percent_format(accuracy = 1)) +
  theme_classic() + ggtitle("Liberal Foreign Policy and UN voting correlation - top 20 countries") +
  theme(plot.title = element_text(size = 18, face = "bold", hjust = 0.5))
```


Liberal Foreign Policy and UN voting correlation – top 20 countries
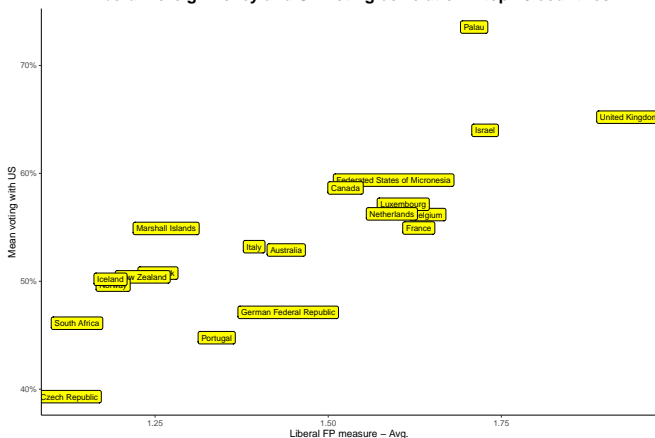
Wrapping up week 5

Summary:

- Measuring complex (latent) concepts: terrorism, ideology.
- Visualize bivariate relations: scatter plot, QQplot.
- z-scores and standardizing units.
- Correlation: how two factors 'move together'.
- R work: scatterplots, cor(), qqplot().