PS 0700
Basic Statistical Methods:
Crosstabulation Example, ANES 2020

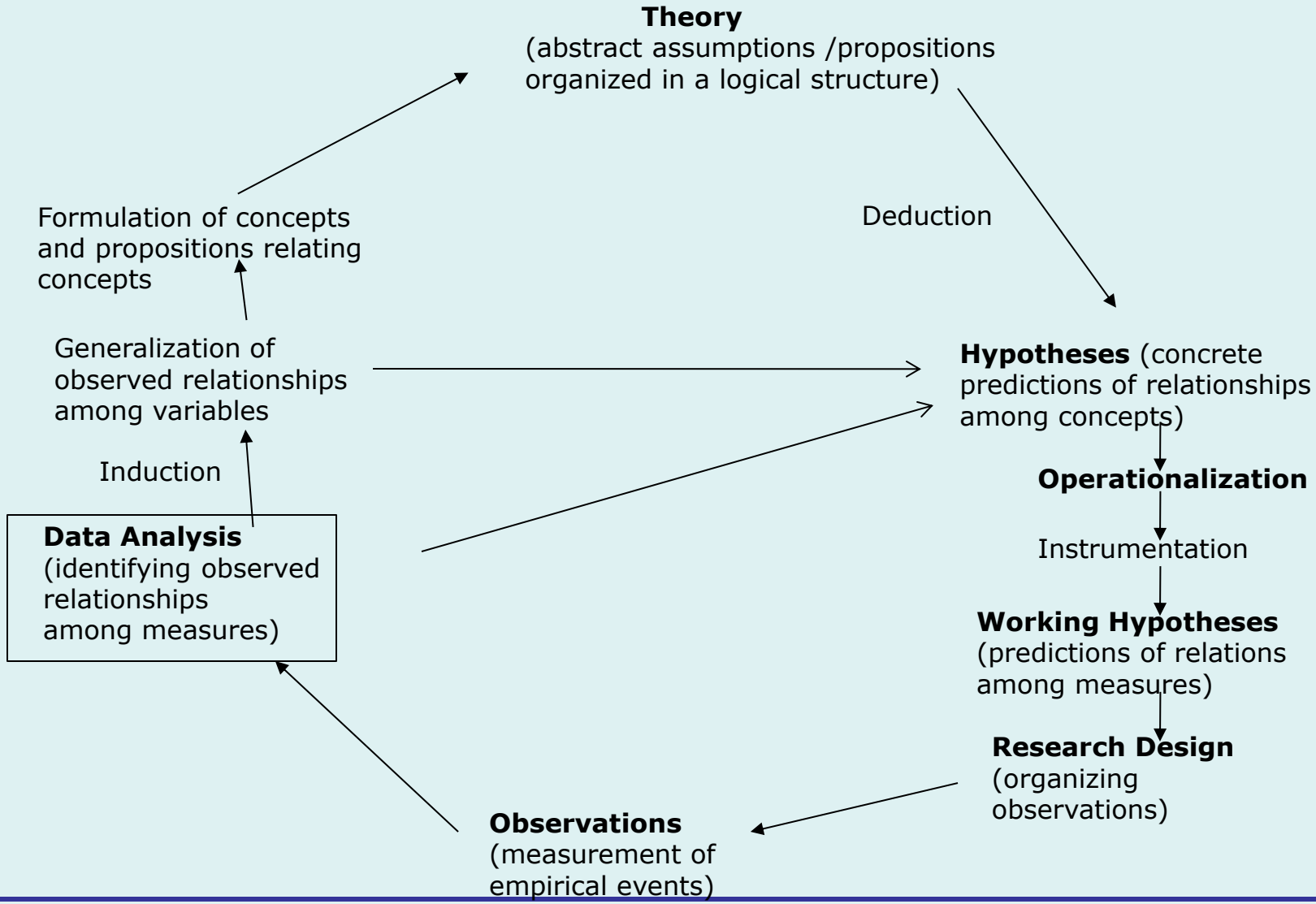
Political Science Research Methods

Professor Steven Finkel

Spring Semester 2022

Week 12

A Model of the Research Process



R Example: 2020 Election Data

- Question: Did vote intention for Donald Trump or Joe Biden depend on individuals' perceptions of the national economy?
- Steps:

1. Get 2020 American National Data, and read the data into R
2. Determine the independent and dependent variables from the codebook
   - Independent Variable: Perceptions of National Economy (V20135)
   - Dependent Variable: Vote Intention (V201033)
3. Run frequencies on each variable to see the range of valid responses, and whether you need to assign "missing values" and/or recode the variable to make the analysis meaningful
4. Recode and/or generate new variables to prepare for analysis
5. Generate crosstabulation and chi-square test and interpret the results

```
library(haven)
anes20 <- read_dta(file = "/Users.../anes_timeseries_2020_stata_20210324.dta")
```

```
str(anes20$V201325) # Economic better in the last year?
```

```
##  dbl+lbl [1:8280] 2, 2, 3, 3, 3, 2, 2, 2, 2, 3, 2, 2, 2, 3, 2, 3, 3, 3, 3, ...
##  @ label       : chr "PRE: National economy better or worse in last year"
##  @ format.stata: chr "%12.0g"
##  @ labels      : Named num [1:5] -9 -8 1 2 3
##   ..- attr(*, "names")= chr [1:5] "-9. Refused" "-8. Don't know" "1. Gotten better" "2. Stayed about the same"
...
```

```
table(anes20$V201325)
```

```
##
##   -9   -8    1    2    3
##   27    8 1552 1704 4989
```

```
anes20 <- anes20 %>%
  mutate(natecon = ifelse((V201325 == 1|V201325 == 2), 1,
                          ifelse(V201325 == 3, 0, NA)))
```

```
anes20$natecon[anes20$V201325 == 1|anes20$V201325 == 2] <- 1
anes20$natecon[anes20$V201325 == 3] <- 0
anes20$natecon[anes20$V201325 == -8|anes20$V201325 == -9] <- NA
```

This is the perceptions of national economy variable – **"natecon"**

I created a two-category variable from this: value 1 will correspond to perceptions of the economy that are "better in the last year" or "stayed about the same", and value 0 will correspond to perceptions that are "worse". Values -9 and -8 will be excluded from the analysis.

```r
library(expss)
cross_cases(anes20, V201033)
```

		#Total
PRE: For whom does R intend to vote for President	-9. Refused	101
	-8. Don't know	33
	-1. Inapplicable	1008
	1. Joe Biden	3759
	2. Donald Trump	3016
	3. Jo Jorgensen	133
	4. Howie Hawkins	55
	5. Other candidate {SPECIFY}	175
	#Total cases	8280

```{r}
anes20$vote <- anes20$V201033
anes20$vote[anes20$vote == 1]<- 0
anes20$vote[anes20$vote == 2] <- 1
anes20$vote[anes20$vote == -1|
              anes20$vote==3|
              anes20$vote==4|
              anes20$vote==5|
              anes20$vote == -8|
              anes20$vote == -9] <- NA
str(anes20$vote)
```

This is the self-reported vote intention in the pre-election interview

The main responses I want to analyze are "1" Joe Biden" and "2" Donald Trump". Minor candidates and other responses could be interesting but not for what I want in this analysis.

So I will treat all responses other than 1 or 2 as "NA", and create a new variable called "vote" which is coded as "0" if the person intends to vote Biden, and "1" if the person intends to vote for Trump

```
library(gmodels)
CrossTable(anes20$vote, anes20$natecon, expected = TRUE, prop.c=TRUE, prop.r = FALSE,
prop.t = FALSE, prop.chisq = FALSE, chisq = TRUE)
```

```
              | anes20$natecon
 anes20$vote  |         0 |         1 | Row Total |
--------------|-----------|-----------|-----------|
            0 |      3245 |       871 |      4116 |
              |  2495.737 |  1620.263 |           |
              |     0.752 |     0.311 |           |
--------------|-----------|-----------|-----------|
            1 |      1071 |      1931 |      3002 |
              |  1820.263 |  1181.737 |           |
              |     0.248 |     0.689 |           |
--------------|-----------|-----------|-----------|
Column Total  |      4316 |      2802 |      7118 |
              |     0.606 |     0.394 |           |
--------------|-----------|-----------|-----------|


Statistics for All Table Factors


Pearson's Chi-squared test
------------------------------------------------------------
Chi^2 =  1354.899     d.f. =  1     p =  1.32771e-296
```

Relationship: Whereas .752 or 75.2 % of people who think the economy is worse (Group 0) intend to vote for Biden, only .31 or 31.1% of people who who think the economy is better or the same intend to vote for Biden. That's a 44 percentage point difference: very large effect!!!

This is the obtained value of chi-square: 1354.9

This is the probability of obtaining a chi-square value this size or greater *if the null hypothesis of independence between the two variables were true*. It is less than 1 in 10000.

Conclusion: **There is a statistically significant relationship between individuals' perception of the state of the national economy and their intended vote choice in 2020 at the .05 significance level.** And it is a big effect substantively!

Next steps: Compare to other variables' effects, and control for possible Z variables in multivariate analyses