EXAM

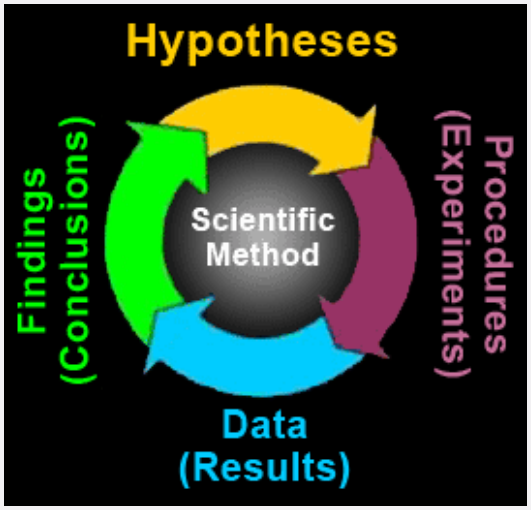
- Wednesday Dec 8: Exam #3
  - Bring a calculator (no phone etc.)
  - Allowed to bring one single-page letter-size (8.5x11) sheet with you. What you put on it is up to you, but it has to be your own sheet (we'll collect it)
- If you take exam at CDR, please sign up now!

EXAM

- Material covered
  - Everything from Nov 3 (More Bivariate Hypothesis Testing, Hypothesis Testing When Using a Sample) to Dec 6 (next Monday)

SCIENTIFIC PROCESS

- Formulate research question
- Propose explanation/theory, hypotheses
- Data collection process
- Use data to evaluate hypotheses
- Reassess explanation



LINEAR REGRESSION

- General form: y = a + b * x
  - y: dependent variable
  - a: intercept
  - b: slope
  - x: independent variable
- Interpretation
  - Slope: For every one unit increase in x, y changes by b units
  - Intercept: When x=0, y takes the value a
- Caveat
  - *Linear* relationships

R-SQUARE

- R^2 tells us how much variation of the dependent variable is explained by the independent variable (in a linear regression)

USING SAMPLES

- Bivariate relationship between two variables in sample
- Is this a real relationship that we would find in the population as well, or is it something that only shows up in our sample?

HYPOTHESIS

- H_0: In the population, there is *no relationship* between dependent and independent variable
  - If there is a difference in the sample, it is due to random sampling error
- H_A: There *is* a relationship between the independent and dependent variable in the population

ERRORS

	There Is A Relation In The Population	There Is No Relation In The Population
We Conclude There Is A Relation	✔	✘ Type I
We Conclude There Is No Relation	✘ Type II	✔

ERRORS

	There Is A Relation In The Population	There Is No Relation In The Population
We Conclude There Is A Relation	✔	✘ Type I in at most 5% of cases
We Conclude There Is No Relation	✘ Type II	✔

IDEA

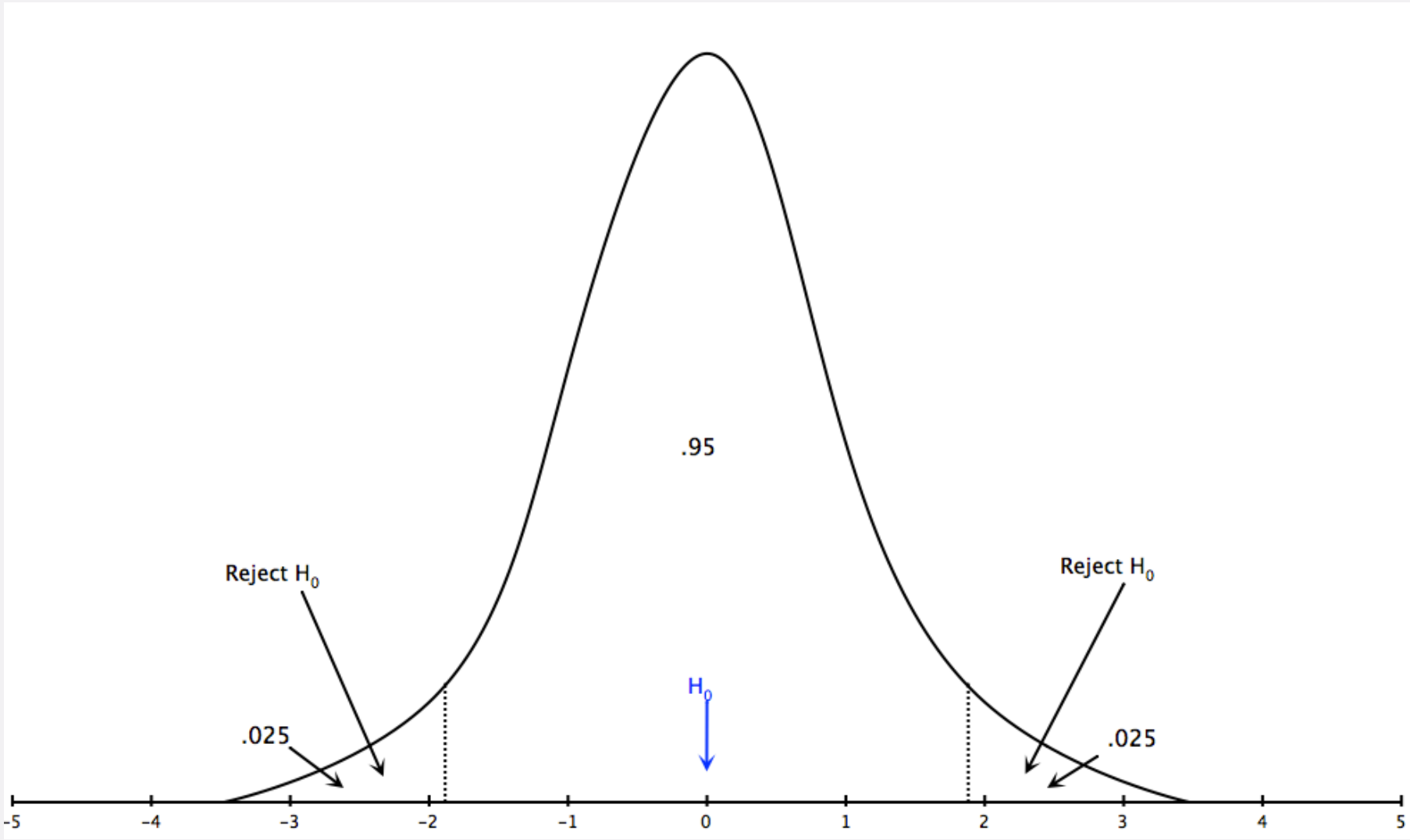
- We start out thinking H_0 is true
  - No relationship between X and Y in population
- We ask: If H_0 is true, how likely is it that a random sample would produce an effect as large (or larger) than the one we have observed?
  - If less than 5% ($p<0.05$): we reject H_0
  - If more than 5% ($p>0.05$): we don't reject H_0

T-STATISTIC

$$t = \frac{H_A - H_0}{\text{Standard Error}}$$

- **H_A: observed relation between X and Y in sample**
- **H_0: relation between X and Y if H_0 is true**

SIGNIFICANCE TEST



- We reject H_0 if t<-1.96 or t>1.96
- This is equivalent to p<0.05

HURDLES TO CAUSALITY

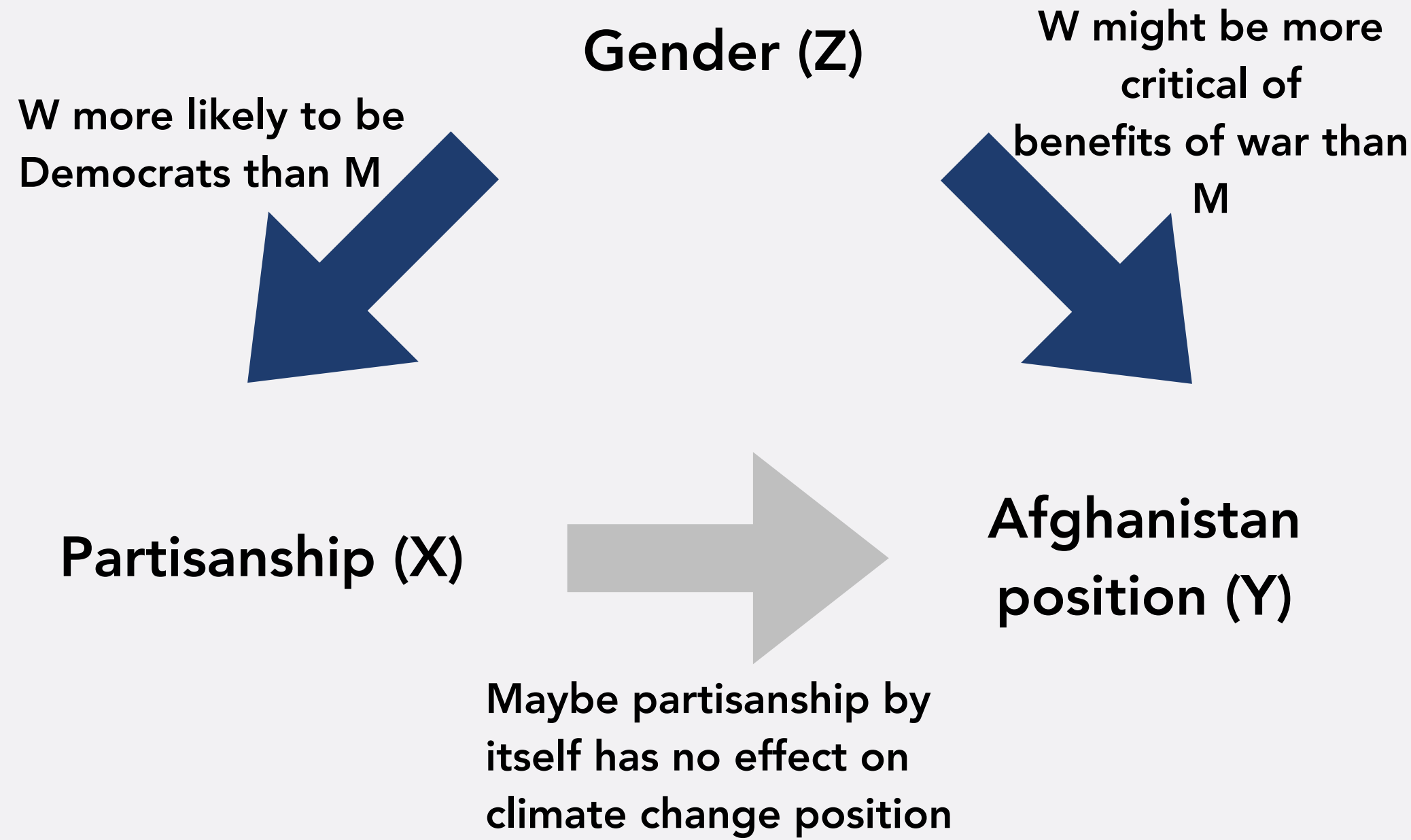
- Is there a credible causal mechanism that connects X to Y?
- Can we rule out the possibility that Y could cause X?
- Is there covariation between X and Y?
- Have we controlled for all confounding variables (Z) that might make the association between X and Y spurious?

BIVARIATE RELATIONSHIP

Partisanship → Evaluation of Afghanistan involvement

- Zero-order effect: Non-Democrats are 22.2% more likely to agree that Afghanistan involvement was beneficial than Democrats

MAYBE THIS IS GOING ON?



TERMINOLOGY

- Controlled effect: relationship between an independent variable (X) and a dependent variable (Y) within one value of another independent variable (Z)

CONTROLLED COMPARISON TABLE

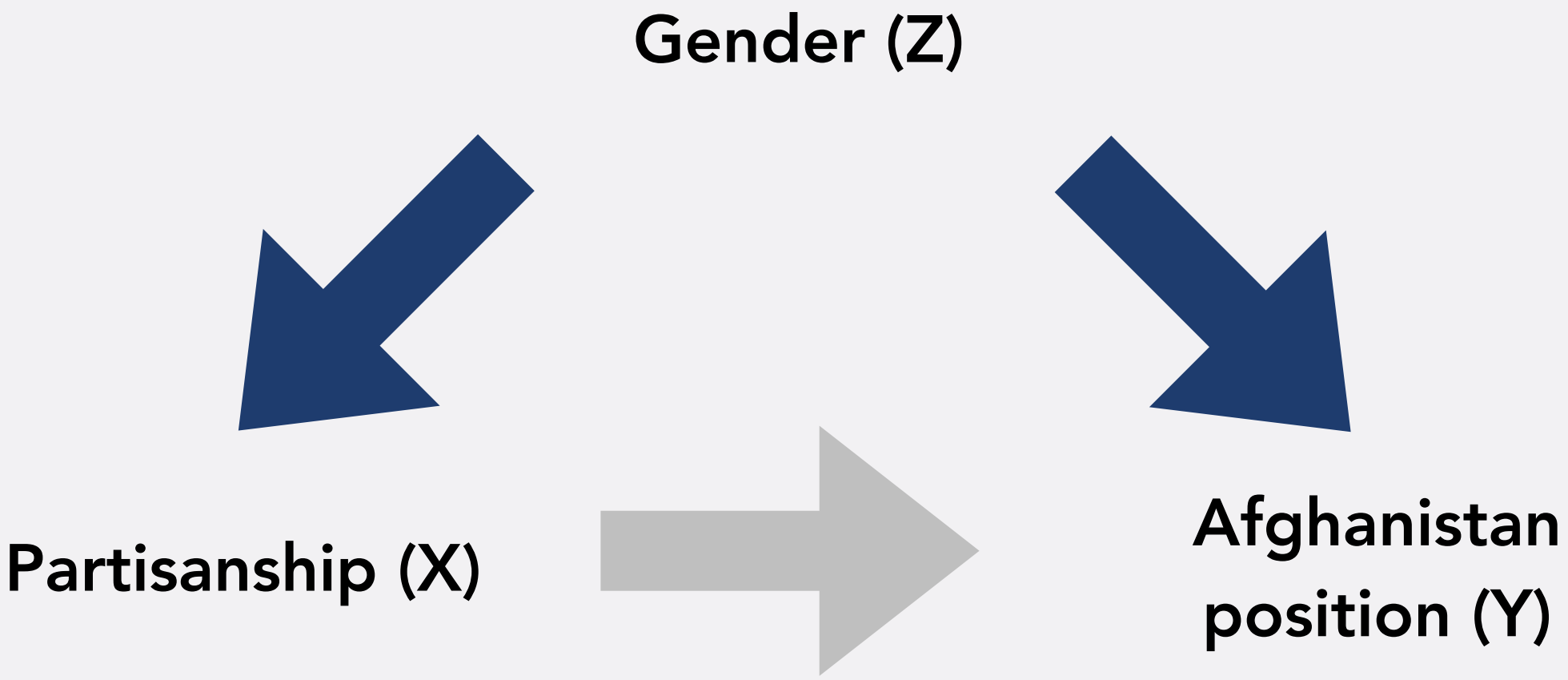
Afghanistan war was beneficial

	Female			Male		
	Dem	Non-Dem	Total	Dem	Non-Dem	Total
	22.3%			21.7%		
Agree	18.9% (7)	41.2% (7)	25.9% (14)	20.0% (3)	41.7% (5)	29.6% (8)
Disagree	81.1% (30)	58.8% (10)	74.1% (40)	80.0% (12)	58.3% (7)	70.4% (19)
Total	100% (37)	100% (17)	100% (54)	100% (15)	100% (12)	100% (27)

CONTROLLED EFFECT

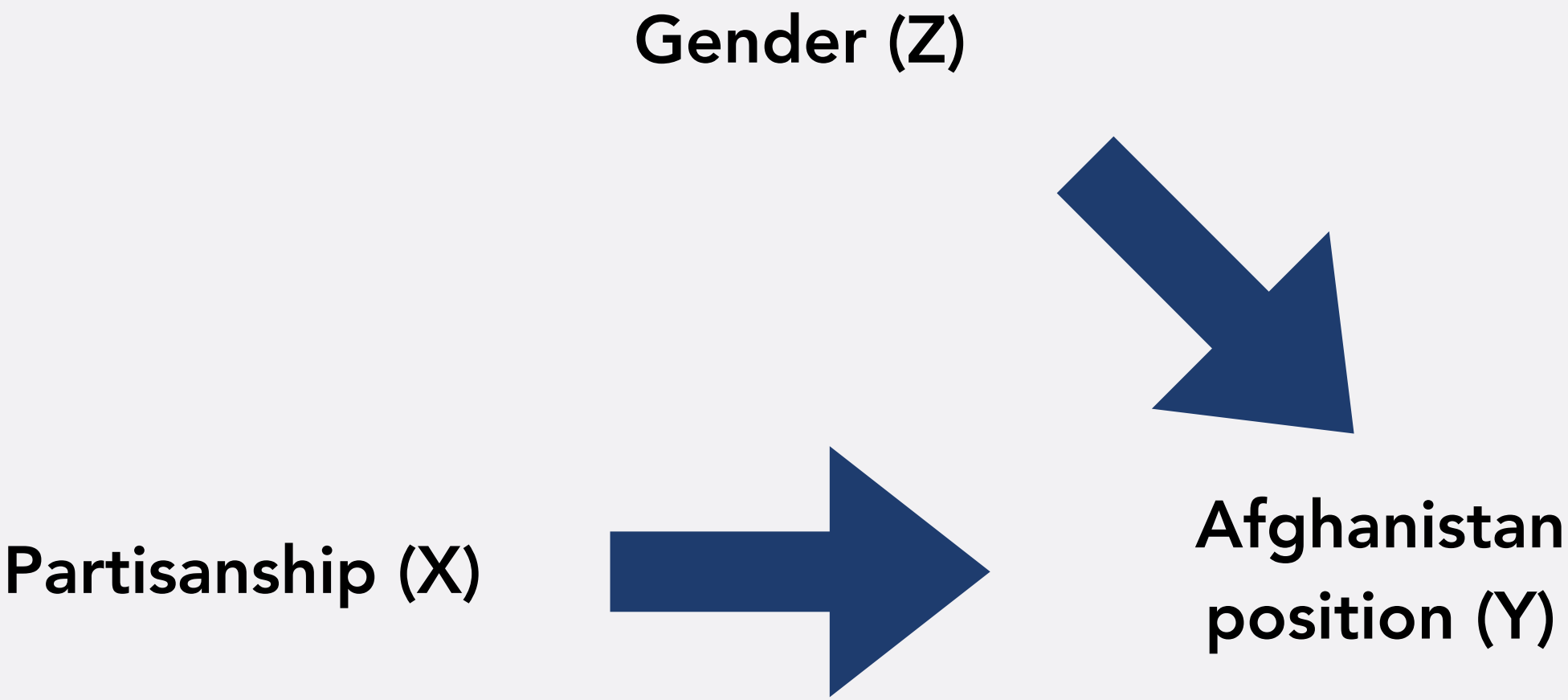
- Even when looking just among men, and just among women, partisanship still has an effect on Afghanistan evaluation
- Effect of partisanship holds when "controlling for" gender

SPURIOUS RELATIONSHIP



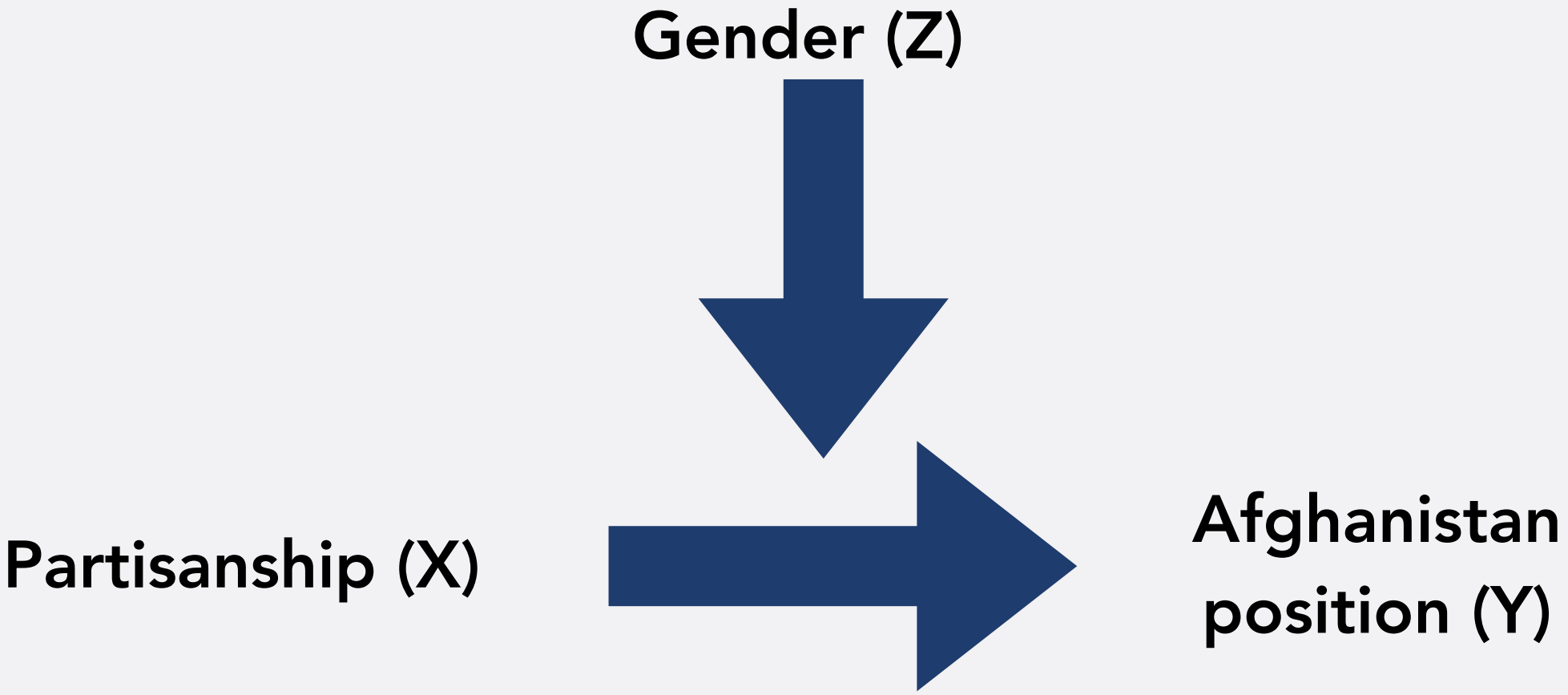
- Once we control for gender, no independent effect of partisanship
- All controlled effects zero or close to zero

ADDITIVE RELATIONSHIP



- Both partisanship *and* gender determine Y
- Controlled effects not zero and of roughly same size

INTERACTIVE RELATIONSHIP



- Gender determines how much partisanship affects Y
- Controlled effects not zero and of different size

MULTIPLE REGRESSION

- Another way to control for potential confounding variables: multiple regression
  - Allows us to control for many potential confounders

DV: APPROVAL OF J. BIDEN

	Coefficient	Standard Error	T-Value
Intercept	101.8	60.8	1.68
Liberal-Conservative	-0.44	0.15	-2.93
Age	-1.89	3.08	-0.61
Gender (Male)	11.66	6.29	1.85

R^2: 0.15

EFFECT OF LIB/CONS

- Coefficient: -0.44
- Interpretation: For every one point increase on the liberal-conservative scale, the evaluation of J. Biden decreases by 0.44 points, *holding all other independent variables constant*

EFFECT OF LIB/CONS

$$t = \frac{H_A - H_0}{\text{Standard Error}}$$

$$t = \frac{-0.44 - 0.00}{0.15} = -2.93$$

- **We reject H_0, so effect of liberal-conservative on evaluation is significant at the 5% level**

DV: APPROVAL OF J. BIDEN

	Coefficient	Standard Error	T-Value
Intercept	101.8	60.8	1.68
Liberal-Conservative	-0.44	0.15	-2.93
Age	-1.89	3.08	-0.61
Gender (Male)	11.66	6.29	1.85

R^2: 0.15

EFFECT OF GENDER

- **Coefficient: -11.66**
- **Interpretation: If someone is male, their evaluation of J. Biden is expected to be 11.66 points higher than if someone is female, *holding all other indpendent variables constant***

EFFECT OF GENDER

- t-value: 1.85
- We do not reject H_0, so effect of gender on evaluation is not significant at the 5% level

DV: APPROVAL OF J. BIDEN

	Coefficient	Standard Error	T-Value
Intercept	101.8	60.8	1.68
Liberal-Conservative	-0.44	0.15	-2.93
Age	-1.89	3.08	-0.61
Gender (Male)	11.66	6.29	1.85

R^2: 0.15

PREDICTED VALUE

- Evaluation = 101.8 - 0.44*Lib/Cons - 1.89*Age + 11.66*Gender (Male)
- Expected approval for someone who is:
  - 50 on Lib/Cons scale
  - 22 years old
  - Male

PREDICTED VALUE

- Evaluation = 101.8 - 0.44*Lib/Cons - 1.89*Age + 11.66*Gender (Male)
- Expected approval for someone who is:
  - 50 on Lib/Cons scale
  - 22 years old
  - Male
- Evaluation = 101.8 - 0.44*50 - 1.89*22 + 11.66*1 = 49.88

DV: APPROVAL OF J. BIDEN

	Coefficient	Standard Error	T-Value
Intercept	101.8	60.8	1.68
Liberal-Conservative	-0.44	0.15	-2.93
Age	-1.89	3.08	-0.61
Gender (Male)	11.66	6.29	1.85

R^2: 0.15

OBSERVATIONAL RESEARCH DESIGN

- Linear regression is (usually) used in observational research design
  - Takes data as we find it in the world
  - Regression isolates the independent effect of X on Y, controlling for other variables (=potential alternative explanations)

OBSERVATIONAL RESEARCH DESIGN

- Can never be sure we controlled for all potential alternative explanations
  - Potentially low internal validity

EXPERIMENTAL RESEARCH DESIGN

- Researchers *actively decide* assignment of the independent variable
- Treatment and control groups
  - Subjects randomly allocated

EXPERIMENTAL RESEARCH DESIGN

- On average, treatment and control group are the same on *every* variable we can think of
  - *Except* on the independent variable of interest, where researcher assigns treatment and control
  - Unlikely that differences in Y between treatment and control groups caused by other variables
  - High internal validity

TYPES OF EXPERIMENTS

- Field Experiment
- Lab Experiment
- Survey Experiment

ISSUES WITH EXPERIMENTS

- May lack external validity
- Ethics issues
- Cannot study many things we are interested in experimentally