

Assessing Hypotheses

Communication Research Methods

Jennifer Pan

Assistant Professor
Department of Communication
Stanford University

February 10, 2016

Announcements

Announcements

- ▶ Be sure to set aside time for psets, pset3 out

Announcements

- ▶ Be sure to set aside time for psets, pset3 out
- ▶ Midterm, Pset2: did well, Friday return

Announcements

- ▶ Be sure to set aside time for psets, pset3 out
- ▶ Midterm, Pset2: did well, Friday return
- ▶ Getting help: SECTION, office hours (Tuesday 10:30am-12:30pm, Friday 1:00-3:00pm)

Announcements

- ▶ Be sure to set aside time for psets, pset3 out
- ▶ Midterm, Pset2: did well, Friday return
- ▶ Getting help: SECTION, office hours (Tuesday 10:30am-12:30pm, Friday 1:00-3:00pm)
- ▶ Mid Course Survey

Where we are

Where we are

- ▶ Up to now:

Where we are

- ▶ Up to now:
 - ▶ What is research and its components

Where we are

- ▶ Up to now:
 - ▶ What is research and its components
 - ▶ Basics of R

Where we are

- ▶ Up to now:
 - ▶ What is research and its components
 - ▶ Basics of R
 - ▶ Causal inference

Where we are

- ▶ Up to now:
 - ▶ What is research and its components
 - ▶ Basics of R
 - ▶ Causal inference
 - ▶ Describing data

Where we are

- ▶ Up to now:
 - ▶ What is research and its components
 - ▶ Basics of R
 - ▶ Causal inference
 - ▶ Describing data
- ▶ This week:

Where we are

- ▶ Up to now:
 - ▶ What is research and its components
 - ▶ Basics of R
 - ▶ Causal inference
 - ▶ Describing data
- ▶ This week:
 - ▶ From what to *why?*: assessing **hypotheses**

Where we are

- ▶ Up to now:
 - ▶ What is research and its components
 - ▶ Basics of R
 - ▶ Causal inference
 - ▶ Describing data
- ▶ This week:
 - ▶ From what to *why*?: assessing **hypotheses**
 - ▶ We must compare: and match method to level of measurement, e.g., cross-tabs, mean comparisons

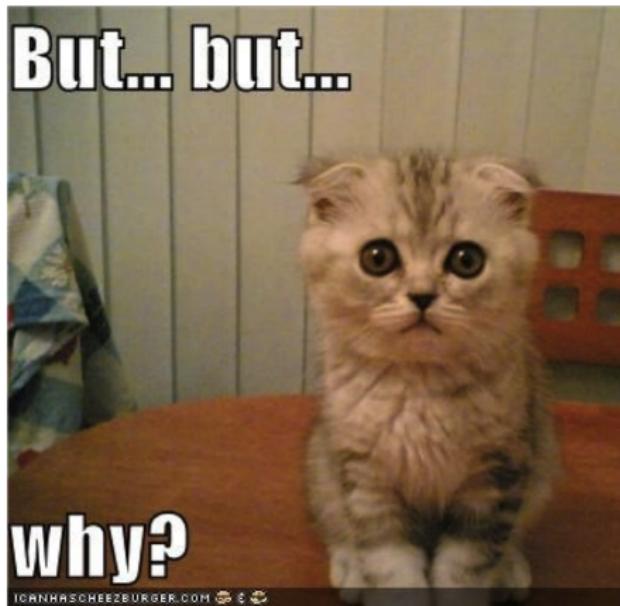
Where we are

- ▶ Up to now:
 - ▶ What is research and its components
 - ▶ Basics of R
 - ▶ Causal inference
 - ▶ Describing data
- ▶ This week:
 - ▶ From what to *why*?: assessing **hypotheses**
 - ▶ We must compare: and match method to level of measurement, e.g., cross-tabs, mean comparisons
 - ▶ A picture is worth a thousand words: graphs and relationships

Where we are

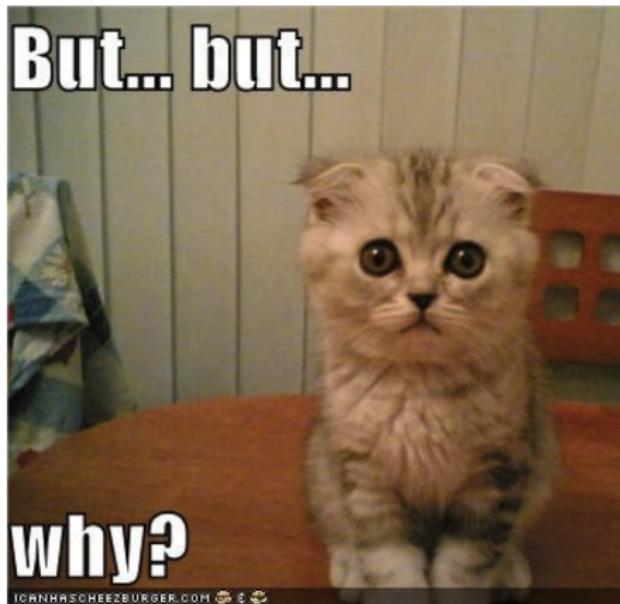
- ▶ Up to now:
 - ▶ What is research and its components
 - ▶ Basics of R
 - ▶ Causal inference
 - ▶ Describing data
- ▶ This week:
 - ▶ From what to *why*?: assessing **hypotheses**
 - ▶ We must compare: and match method to level of measurement, e.g., cross-tabs, mean comparisons
 - ▶ A picture is worth a thousand words: graphs and relationships
 - ▶ testing our claims: statistical significance, null hypothesis, measures of association

Why...



ICANHASCHEEZBURGER.COM

Why...



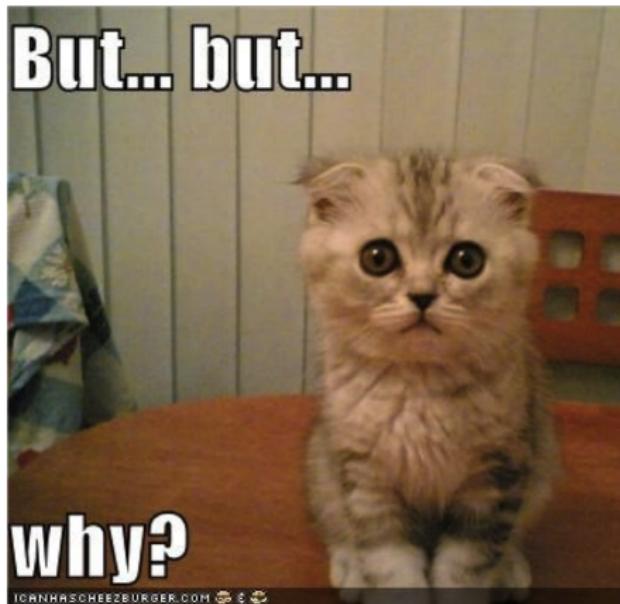
- ▶ ...do some people vote Democrat, while others vote Republican?

Why...



- ▶ ...do some people vote Democrat, while others vote Republican?
- ▶ ...do women earn less than men?

Why...



- ▶ ...do some people vote Democrat, while others vote Republican?
- ▶ ...do women earn less than men?
- ▶ ...do wars happen?

Notice

- ▶ ...do some people vote Democrat, while others vote Republican?
- ▶ ...do women earn less than men?
- ▶ ...do wars happen?
- ▶ we make an explicit observation about a characteristic that varies: vote choice, salary, war...this will become our dependent variable (Y)
- ▶ we need a causal explanation as an answer which must...
 1. describe a connection between the dependent variable and an independent variable (X)
 2. assert a direction for the difference
 3. be testable!

Causal Explanations: Gender and Earnings



Causal Explanations: Gender and Earnings

- ▶ Why do women **earn less** than men, even in the same jobs?
(professor v professor, doctor v doctor etc.)



Causal Explanations: Gender and Earnings

- ▶ Why do women **earn less** than men, even in the same jobs?
(professor v professor, doctor v doctor etc.)
- ▶ ...because of **sexism**



Causal Explanations: Gender and Earnings

- ▶ Why do women **earn less** than men, even in the same jobs?
(professor v professor, doctor v doctor etc.)
- ▶ ...because of **sexism**
- ▶ too vague!



Causal Explanations: Gender and Earnings

- ▶ Why do women **earn less** than men, even in the same jobs?
(professor v professor, doctor v doctor etc.)
- ▶ ...because of **sexism**
- ▶ too vague!
- ▶ Think about the causal process



Causal Explanation 1

intervening variables

Women have babies.



Women take care of those babies.



Women have to take time off work.



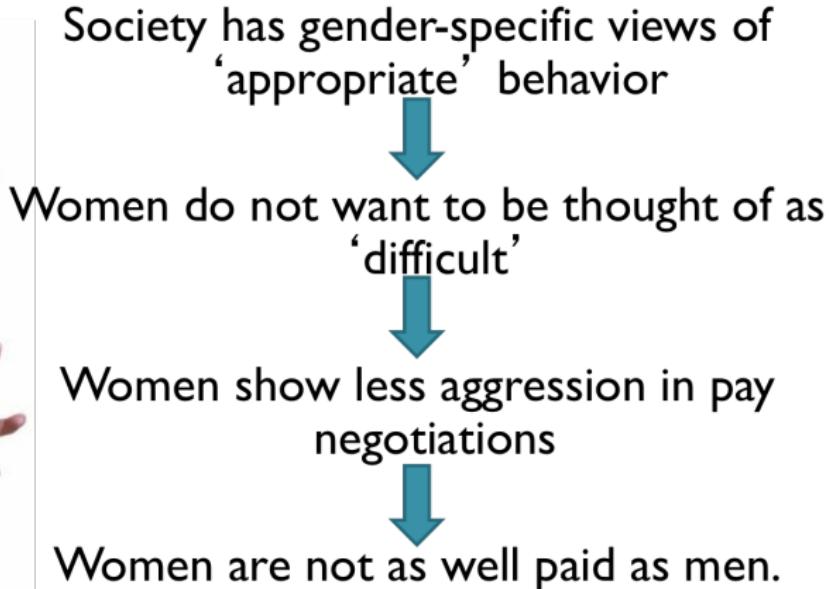
Women miss promotion opportunities.



Women are not as well paid as men.



Causal Explanation 2



Causal Explanations: Gender and Earnings

- ▶ Very different explanations, but both might be 'sexism'



Causal Explanations: Gender and Earnings

- ▶ Very different explanations, but both might be 'sexism'
- ▶ We don't know if the links in the chain (**intervening variables**) are correct, but we could test them (surveys, experiments etc):



Causal Explanations: Gender and Earnings

- ▶ Very different explanations, but both might be 'sexism'
- ▶ We don't know if the links in the chain (**intervening variables**) are correct, but we could test them (surveys, experiments etc):
 - ▶ do mothers do more child care than fathers?



Causal Explanations: Gender and Earnings

- ▶ Very different explanations, but both might be 'sexism'
- ▶ We don't know if the links in the chain (**intervening variables**) are correct, but we could test them (surveys, experiments etc):
 - ▶ do mothers do more child care than fathers?
 - ▶ does time off lead to a lack of promotion?



Causal Explanations: Gender and Earnings

- ▶ Very different explanations, but both might be 'sexism'
- ▶ We don't know if the links in the chain (**intervening variables**) are correct, but we could test them (surveys, experiments etc):
 - ▶ do mothers do more child care than fathers?
 - ▶ does time off lead to a lack of promotion?
 - ▶ do people have notions of appropriate gender behavior?



Causal Explanations: Gender and Earnings

- ▶ Very different explanations, but both might be 'sexism'
- ▶ We don't know if the links in the chain (**intervening variables**) are correct, but we could test them (surveys, experiments etc):
 - ▶ do mothers do more child care than fathers?
 - ▶ does time off lead to a lack of promotion?
 - ▶ do people have notions of appropriate gender behavior?
 - ▶ do women negotiate with less aggression?



Causal Explanations: Gender and Earnings: Child-care responsibilities



Causal Explanations: Gender and Earnings: Child-care responsibilities

- ▶ Want to know if child-care responsibilities are the cause of the difference.



Causal Explanations: Gender and Earnings: Child-care responsibilities

- ▶ Want to know if child-care responsibilities are the cause of the difference.
- ▶ We would:



Causal Explanations: Gender and Earnings: Child-care responsibilities

- ▶ Want to know if child-care responsibilities are the cause of the difference.
- ▶ We would:
 - ▶ separate units on the **independent variable**
- ▶ If child-care responsibility are the cause of the difference, we should find a **difference**



Causal Explanations: Gender and Earnings: Child-care responsibilities

- ▶ Want to know if child-care responsibilities are the cause of the difference.
- ▶ We would:
 - ▶ separate units on the **independent variable** [child-care responsibilities]...
- ▶ If child-care responsibility are the cause of the difference, we should find a **difference**
- ▶ Now we have a hypothesis



Causal Explanations: Gender and Earnings: Child-care responsibilities

- ▶ Want to know if child-care responsibilities are the cause of the difference.
- ▶ We would:
 - ▶ separate units on the **independent variable** [child-care responsibilities]...
 - ▶ ...and compare values on the **dependent variable**
- ▶ If child-care responsibility are the cause of the difference, we should find a **difference**
- ▶ Now we have a hypothesis



Causal Explanations: Gender and Earnings: Child-care responsibilities

- ▶ Want to know if child-care responsibilities are the cause of the difference.
- ▶ We would:
 - ▶ separate units on the **independent variable** [child-care responsibilities]...
 - ▶ ...and compare values on the **dependent variable** [salary]...
- ▶ If child-care responsibility are the cause of the difference, we should find a **difference**
- ▶ Now we have a hypothesis



Causal Explanations: Gender and Earnings: Child-care responsibility and earnings hypothesis

A hypothesis is a testable statement about the empirical relationship between an independent variable and a dependent variable

Causal Explanations: Gender and Earnings: Child-care responsibility and earnings hypothesis

A hypothesis is a testable statement about the empirical relationship between an independent variable and a dependent variable

|

In a comparison of [units of analysis] those having [one value on the independent variable] will be more likely to have [one value on the dependent variable] than will those having [a different value on the independent variable].

Causal Explanations: Gender and Earnings: Child-care responsibility and earnings hypothesis

A hypothesis is a testable statement about the empirical relationship between an independent variable and a dependent variable

|

In a comparison of [units of analysis] those having [one value on the independent variable] will be more likely to have [one value on the dependent variable] than will those having [a different value on the independent variable].

Get together in a group with people near you: write down the hypotheses for child-care responsibility and earnings (5min - then come up to board)

Examples Hypotheses

Examples Hypotheses

|

In a comparison of [workers] those having [fewer child-care responsibilities] will be more likely to have [higher salaries] than will those having [more child care responsibilities].

Examples Hypotheses

|

n a comparison of [workers] those having [fewer child-care responsibilities] will be more likely to have [higher salaries] than will those having [more child care responsibilities].

|

n a comparison of [nations] those having [a less democratic polity] will be more likely to have [wars] than will those having [a more democratic polity].

Hypotheses Recap

A hypothesis is a testable statement about the empirical relationship between an independent variable and a dependent variable

Hypotheses Recap

A hypothesis is a testable statement about the empirical relationship between an independent variable and a dependent variable

- ▶ we state a **relationship** between two variables: dependent and independent

Hypotheses Recap

A hypothesis is a testable statement about the empirical relationship between an independent variable and a dependent variable

- ▶ we state a **relationship** between two variables: dependent and independent
- ▶ we talk about **comparing** this type of unit with that type

Hypotheses Recap

A hypothesis is a testable statement about the empirical relationship between an independent variable and a dependent variable

- ▶ we state a **relationship** between two variables: dependent and independent
- ▶ we talk about **comparing** this type of unit with that type
- ▶ we are **specific** about what we are comparing and on what grounds

Hypotheses Recap

A hypothesis is a testable statement about the empirical relationship between an independent variable and a dependent variable

- ▶ we state a **relationship** between two variables: dependent and independent
- ▶ we talk about **comparing** this type of unit with that type
- ▶ we are **specific** about what we are comparing and on what grounds
- ▶ we assert **direction**: more of this leads to more of that

How to Make Comparisons?

How to Make Comparisons?

- ▶ We choose different methods, depending on the **type of measure** we have

How to Make Comparisons?

- ▶ We choose different methods, depending on the **type of measure** we have
- ▶ Suppose the **independent variable** is nominal or ordinal, then:

How to Make Comparisons?

- ▶ We choose different methods, depending on the **type of measure** we have
- ▶ Suppose the **independent variable** is nominal or ordinal, then:
 - ▶ **Nominal dependent variable:** cross-tabulation

How to Make Comparisons?

- ▶ We choose different methods, depending on the **type of measure** we have
- ▶ Suppose the **independent variable** is nominal or ordinal, then:
 - ▶ **Nominal dependent variable:** cross-tabulation
 - ▶ **Ordinal dependent variable:** cross-tabulation

How to Make Comparisons?

- ▶ We choose different methods, depending on the **type of measure** we have
- ▶ Suppose the **independent variable** is nominal or ordinal, then:
 - ▶ **Nominal dependent variable:** cross-tabulation
 - ▶ **Ordinal dependent variable:** cross-tabulation
 - ▶ **Interval dependent variable:** comparison of means

Our hypothesis



Our hypothesis



- ▶ Men shave their face more often than women

Our hypothesis



- ▶ Men shave their face more often than women
- ▶ Independent variable:

Our hypothesis



- ▶ Men shave their face more often than women
- ▶ Independent variable: Gender,

- ▶ Method of comparison:
cross-tabulation (cross-tabs)

Our hypothesis



- ▶ Men shave their face more often than women
- ▶ Independent variable: Gender, nominal

- ▶ Method of comparison: cross-tabulation (cross-tabs)

Our hypothesis



- ▶ Men shave their face more often than women
- ▶ Independent variable: Gender, nominal
- ▶ Dependent variable:
- ▶ Method of comparison: cross-tabulation (cross-tabs)

Our hypothesis



- ▶ Men shave their face more often than women
- ▶ Independent variable: Gender, nominal
- ▶ Dependent variable: Frequency of shaving face,
- ▶ Method of comparison: cross-tabulation (cross-tabs)

Our hypothesis



- ▶ Men shave their face more often than women
- ▶ Independent variable: Gender, nominal
- ▶ Dependent variable: Frequency of shaving face, ordinal
- ▶ Method of comparison: cross-tabulation (cross-tabs)

Cross Tabulation

Cross tabs show the distribution of cases **across the values of a dependent variable** for cases that have *different values* on the **independent variable**.

Cross Tabulation

- ▶ The independent variable as the columns

Cross tabs show the distribution of cases **across the values of a dependent variable** for cases that have *different values* on the **independent variable**.

Cross Tabulation

Cross tabs show the distribution of cases **across the values of a dependent variable** for cases that have *different values* on the **independent variable**.

- ▶ The independent variable as the columns
- ▶ The dependent variable as the rows

Cross Tabulation

Cross tabs show the distribution of cases **across the values of a dependent variable** for cases that have *different values* on the **independent variable**.

- ▶ The independent variable as the columns
- ▶ The dependent variable as the rows
- ▶ Calculate percentages of categories of the independent variable (we ask for “column percentages” [columns sum to 100%])

Cross Tabulation

Cross tabs show the distribution of cases **across the values of a dependent variable** for cases that have *different values* on the **independent variable**.

- ▶ The independent variable as the columns
- ▶ The dependent variable as the rows
- ▶ Calculate percentages of categories of the independent variable (we ask for “column percentages” [columns sum to 100%])
- ▶ Compare percentages across columns for the same value of the dependent variable

Our hypothesis



- ▶ Men shave their face more often than women
- ▶ Independent variable: Gender (nominal)
- ▶ Dependent variable: Frequency of shaving face (ordinal)
- ▶ Method of comparison: cross-tabulation (cross-tabs)

Cross Tabulation: Shaving and Gender

		F	M	Total
Shaving	Frequently Count	0	4	4
	% within F/M	.0%	15.4%	7.3%
	Sometimes Count	0	11	11
	% within F/M	.0%	42.3%	20%
	Occasionally Count	0	8	8
	% within F/M	.0%	30.7%	14.5%
	Rarely Count	0	2	2
	% within F/M	.0%	7.7%	3.6%
	Never Count	29	1	30
	% within F/M	100.0%	3.8%	54.5%
	Total	29	26	55
	% within F/M	100.0%	100.0%	100.0%

Cross Tabulation: Shaving and Gender

independent variable

		F	M	Total
Shaving	Frequently Count	0	4	4
	% within F/M	.0%	15.4%	7.3%
Sometimes Count	0	11	11	
	% within F/M	.0%	42.3%	20%
Occasionally Count	0	8	8	
	% within F/M	.0%	30.7%	14.5%
Rarely Count	0	2	2	
	% within F/M	.0%	7.7%	3.6%
Never Count	29	1	30	
	% within F/M	100.0%	3.8%	54.5%
Total	29	26	55	
within F/M	100.0%	100.0%	100.0%	

Percent of ind variable

dependent variable

column sums to 100%

Cross Tabulation in R

Cross Tabulation in R

- ▶ Data: shave.csv

Cross Tabulation in R

- ▶ Data: shave.csv
- ▶ Cross tab function: table()

Cross Tabulation in R

- ▶ Data: shave.csv
- ▶ Cross tab function: table()
- ▶ Recode “shave” variable: ifelse()

Cross Tabulation in R

- ▶ Data: shave.csv
- ▶ Cross tab function: table()
- ▶ Recode “shave” variable: ifelse()
- ▶ Column percentages: prop.table()

Cross Tabulation in R

- ▶ Data: shave.csv
- ▶ Cross tab function: table()
- ▶ Recode “shave” variable: ifelse()
- ▶ Column percentages: prop.table()

Men shave their face more often than women: 29 women never shave, but only 1 man never shaves (100% of women never shave, but only 3.8% of men never shave)

How to Make Comparisons?

- ▶ We choose different methods, depending on the **type of measure** we have
- ▶ Suppose the **independent variable** is nominal or ordinal, then:
 - ▶ **Nominal dependent variable:** cross-tabulation
 - ▶ **Ordinal dependent variable:** cross-tabulation
 - ▶ **Interval dependent variable:** comparison of means

Comparison of Means



Comparison of Means

- ▶ Use when we have an interval-level dependent variable



Comparison of Means

- ▶ Use when we have an interval-level dependent variable
- ▶ Examples:



Comparison of Means

- ▶ Use when we have an interval-level dependent variable
- ▶ Examples:
 - ▶ income of blacks vs. whites



Comparison of Means

- ▶ Use when we have an interval-level dependent variable
- ▶ Examples:
 - ▶ income of blacks vs. whites
 - ▶ height of boys vs. girls



Comparison of Means

- ▶ Use when we have an interval-level dependent variable
- ▶ Examples:
 - ▶ income of blacks vs. whites
 - ▶ height of boys vs. girls
 - ▶ feelings towards Trump for voters in different states



Comparison of Means: Examples

Income by Education Level	2010	2012	% Change
Overall Average	\$48,016	\$42,380	-11.7%
No High School Diploma	\$28,016	\$24,492	-12.6%
Some College	\$43,854	\$37,804	-13.8%
Associates Degree	\$49,970	\$40,820	-18.3%
Bachelor's Degree	\$68,728	\$55,432	-19.3%
Master's Degree	\$78,541	\$67,600	-13.9%
Professional Degree	\$99,130	\$90,220	-9.0%
Doctoral Degree	\$96,830	\$84,448	-12.8%

Comparison of Means: Examples

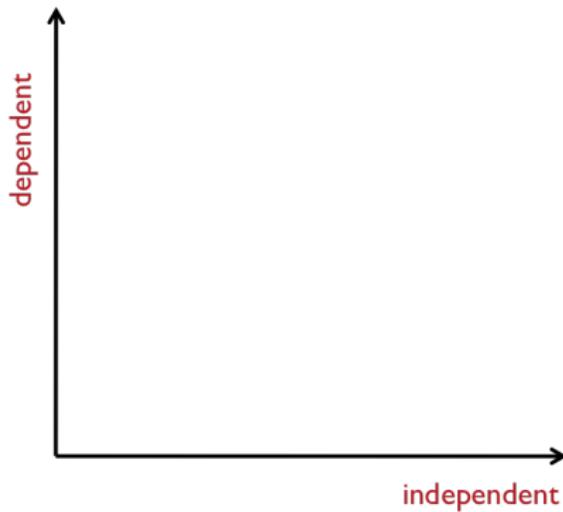
Table 1: Per capita milk consumption, milk (milk equivalents)

	1981	1990	2000	2007e
	Kg/capita			
World	86	80	104	113
Developed	222	180	235	248
Developing	35	40	56	68
Asia	33	39	50	64
China	3	6	11	na
Mongolia	141	144	176	na
SE Asia	6	10	16	32
Thailand	8	14	26	na
Philippines	14	19	22	na
Malaysia	48	41	52	na
Viet Nam	1	1	8	na
South Asia	543	66	84	93
India	50	63	79	na
Pakistan	113	134	180	na
Bangladesh	16	17	18	na
Sri Lanka	24	28	33	na
Nepal	49	50	50	na
US	271	274	287	295
EU	na	363	496	382

Source: FAO estimates

Making Comparisons by Graphing

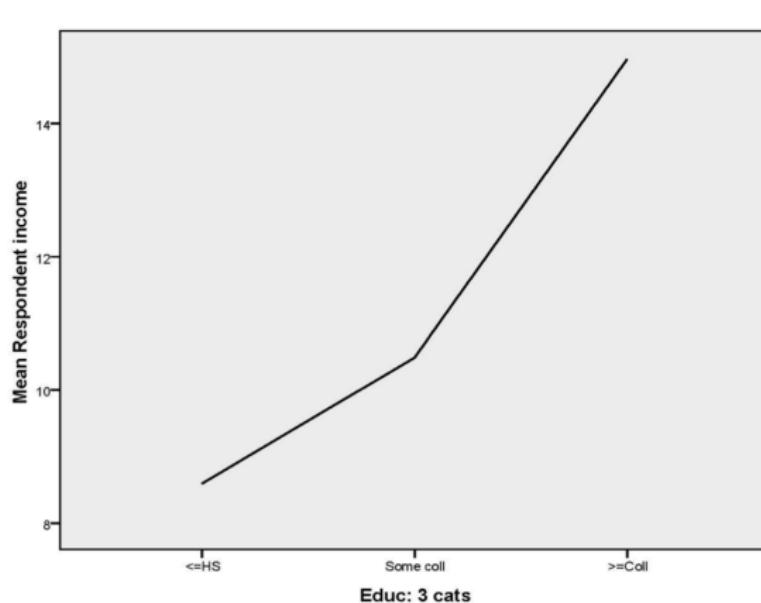
Put the *independent* variable on the x-axis (horizontal axis),
put the *dependent* variable on the y-axis (vertical axis)



Making Comparisons by Graphing: Positive Relationship

positive relationship

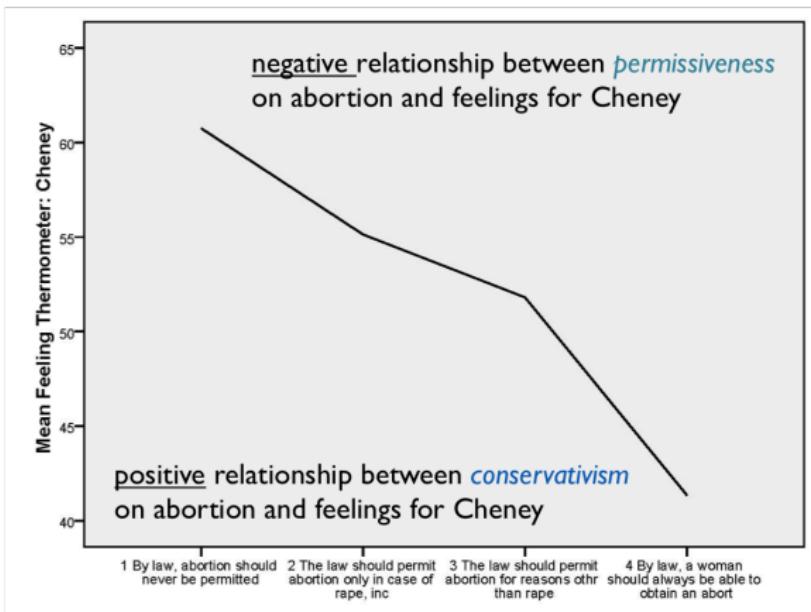
- both variables **increase together** (increase X, increase Y)
- both variables **decrease together** (decrease X, decrease Y)



Making Comparisons by Graphing: Negative Relationship

negative relationship

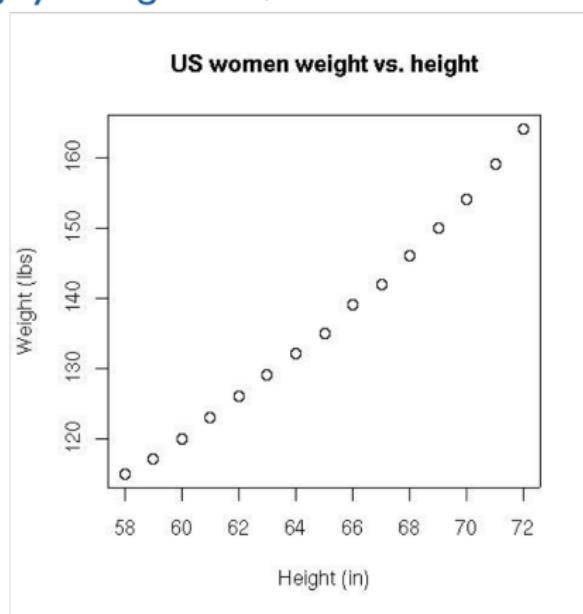
- as value of one variable increases, other decreases
- be specific in describing increases and decreases (in what sense?)



Making Comparisons by Graphing: Linear Relationships

If we can join the points on the graph (or average the points) with a **roughly straight line**, we have a **linear relationship**

fitting a straight line through
points is what we do in "linear
regression"



Making Comparisons by Graphing: Non-linear Relationships

If we **cannot** draw a roughly straight line through the points....

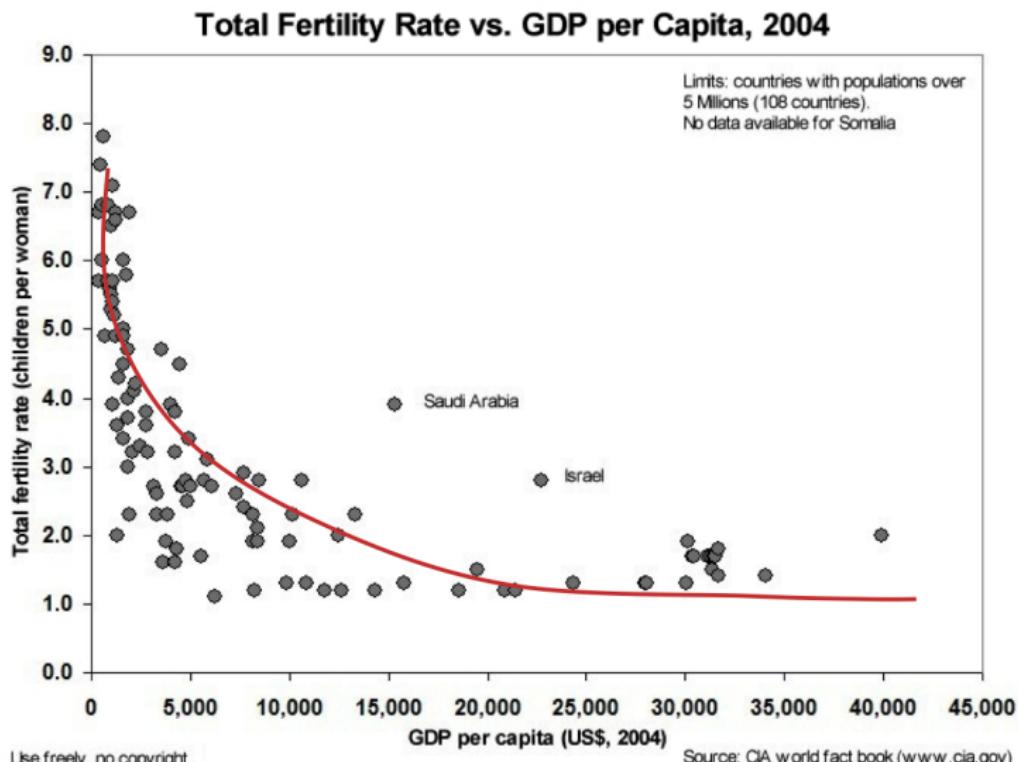
...relationship between X and Y depends on **which range** of independent variable is being examined.

If the line is generally increasing, or generally decreasing, but not linear...

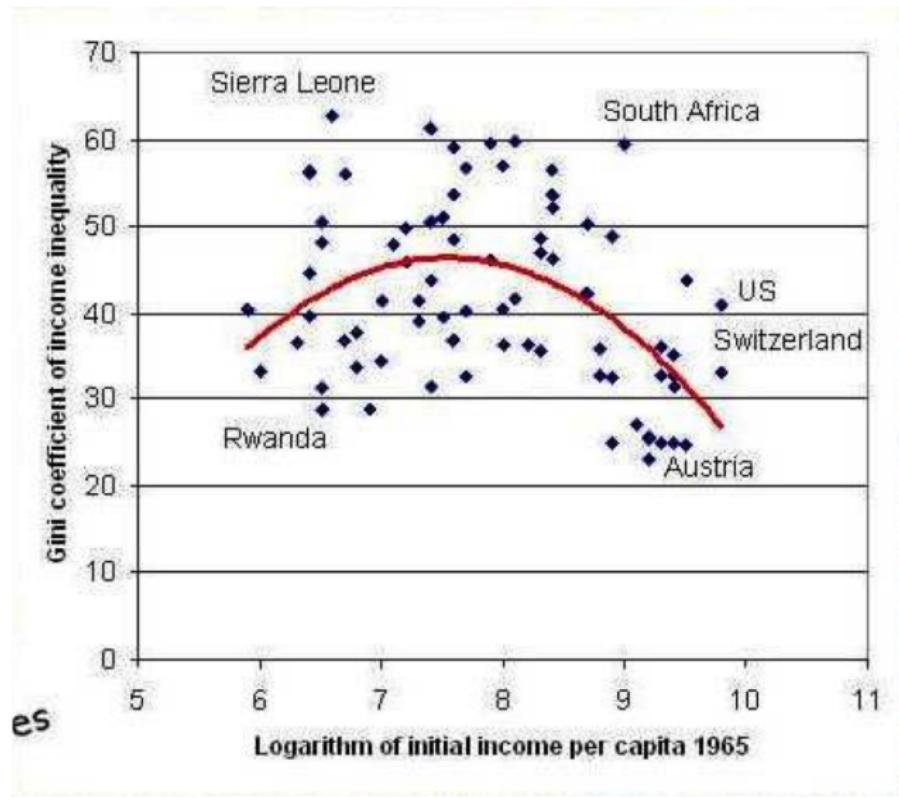
...we probably have a **curvilinear relationship**

It might also be **“U-shaped”** or **“inverted-U-shaped”**

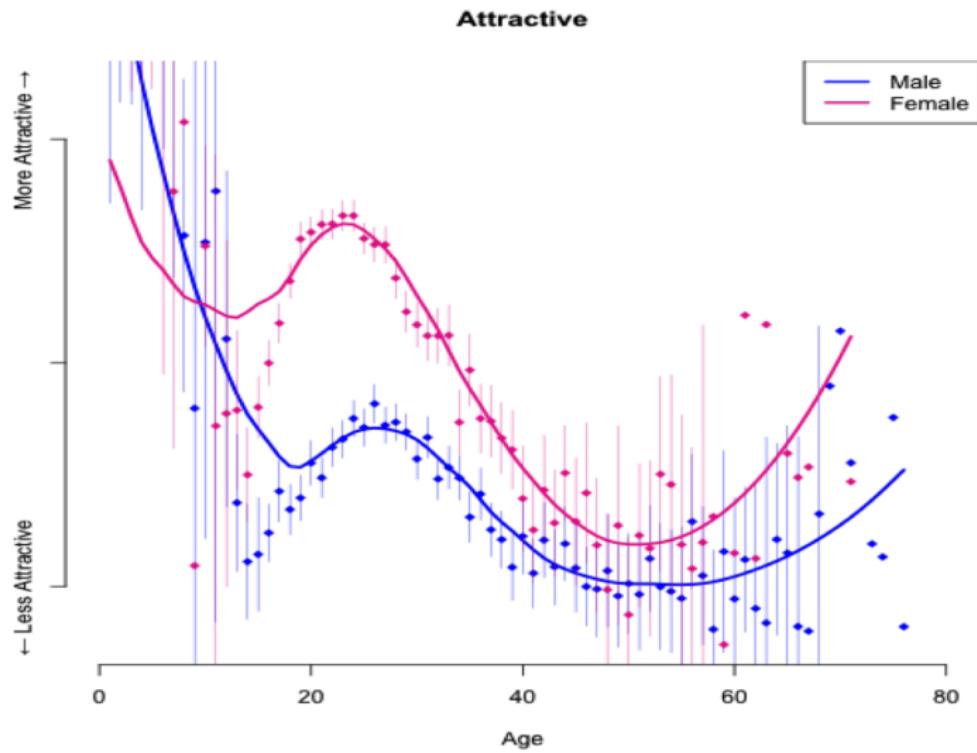
Making Comparisons by Graphing: Non-linear Relationships



Making Comparisons by Graphing: Non-linear Relationships



Making Comparisons by Graphing: Non-linear Relationships



Making Comparisons by Graphing: R

`plot()`

Summing Up

Summing Up

- ▶ Answering questions with hypotheses

Summing Up

- ▶ Answering questions with hypotheses
- ▶ Making comparison to answer questions, and matching method to type of variable, e.g., cross-tabs, mean comparisons

Summing Up

- ▶ Answering questions with hypotheses
- ▶ Making comparison to answer questions, and matching method to type of variable, e.g., cross-tabs, mean comparisons
- ▶ Making comparisons by graphing

Summing Up

- ▶ Answering questions with hypotheses
- ▶ Making comparison to answer questions, and matching method to type of variable, e.g., cross-tabs, mean comparisons
- ▶ Making comparisons by graphing
- ▶ Next time: testing our claims