Missing Data

Nicholas Mattei, *Tulane University*

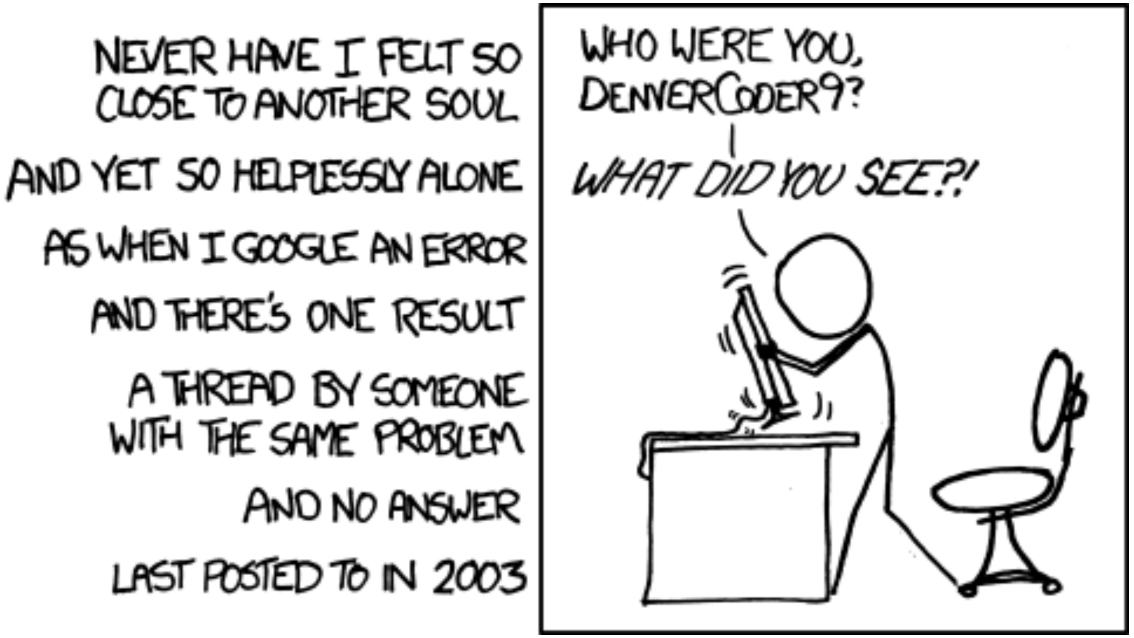
CMPS3660 – Introduction to Data Science – Fall 2019

https://rebrand.ly/TUDataScience

Many Thanks
Slides based off Introduction to Data Science from John P. Dickerson - https://cmsc320.github.io/

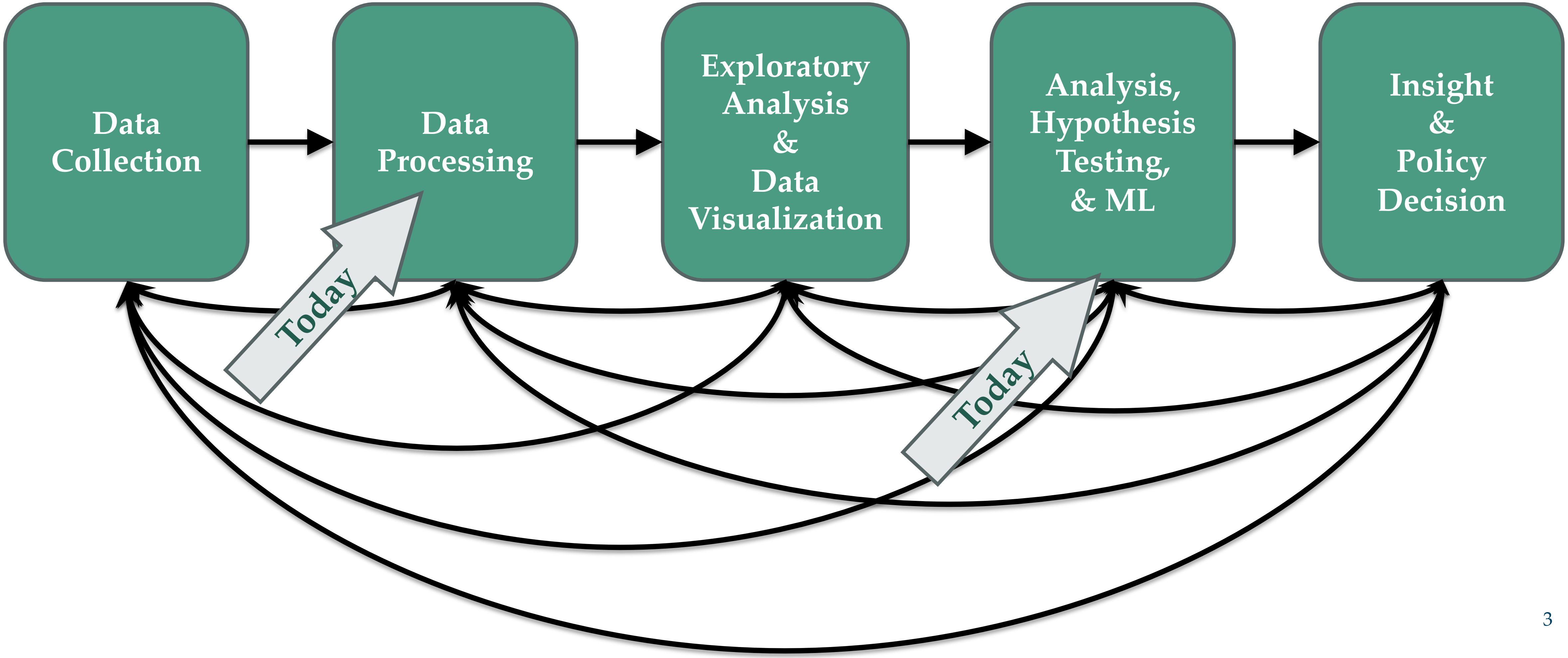
Announcements

- Peer Observation Tuesday 10/29
- Finish Lab 7
- Review Questions
- Today's Lecture



https://xkcd.com/979/

The Data LifeCycle



Today's Lecture

- Missing Data …
  - What is it?
  - Simple methods for **imputation.**
  - … with a tiny taste of Stats/ML lecturers to come.

Thanks to John Atwood and Wenjiang Fu

Missing Data

- Missing data is information that we want to know, but don't!

- It can come in many forms, e.g.:
  - People not answering questions on surveys
  - Inaccurate recordings of the height of plants that need to be discarded
  - Canceled runs in a driving experiment due to rain

- Could also consider missing columns (no collection at all) to be missing data …

[JA]

Key Question

- **Why is the data missing?**
  - What mechanism is it that contributes to, or is associated with, the probability of a data point being absent?
  - Can it be explained by our observed data or not?

- The answers drastically affect what we can ultimately do to compensate for the missing-ness

[JA]

Complete Case Analysis

- Delete all tuples with any missing values at all, so you are left only with observations with all variables observed

```
# Clean out rows with nil values
df = df.dropna()
```

- Default behavior for libraries for analysis (e.g., regression).
  - We'll talk about this much more during the Stats/ML lectures

- This is the simplest way to handle missing data. In some cases, will work fine; in others, ?????????????:
  - Loss of sample will lead to variance larger than reflected by the size of your data.
  - May bias your sample.

[JA]

Tulane University

Example

- **Dataset:** Body fat percentage in men, and the circumference of various body parts [Penrose et al., 1985]
- **Question:** Does the circumference of certain body parts predict body fat percentage?

- Given **complete** data, how would you answer this ?????????
- One way to answer is **regression analysis:**
  - One or more independent variables ("predictors”)
  - One dependent variables (“outcome”)

- What is the relationship between the predictors and the outcome?
- What is the conditional expectation of the dependent variable given fixed values for the independent variables?

- Generalized body composition prediction equation for men using simple measurement techniques", K.W. Penrose, A.G. Nelson, A.G. Fisher, FACSM, Human Performance research Center, Brigham Young University, Provo, Utah 84602 as listed in Medicine and Science in Sports and Exercise, vol. 17, no. 2, April 1985, p. 189.
- http://staff.pubhealth.ku.dk/~tag/Teaching/share/data/Bodyfat.html

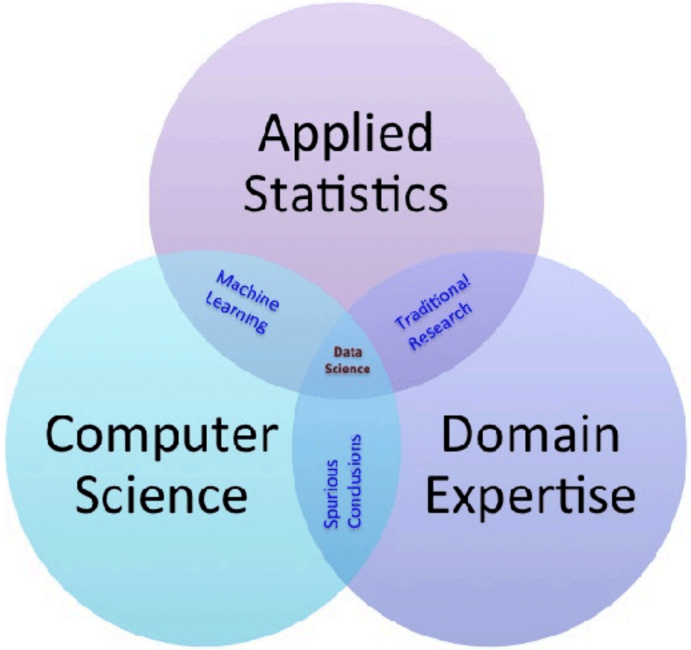
A Side Note On Terms

- For linear regression we have equations of where we want to know the relationship between an **outcome** given some **predictors**.

- If you have a ML background:
  - Get **target, outcome** given **predictors, observations.**

- If you have a Stats background:
  - Get **endogenous** variables given **exogenous** variables.

- If you are more of a Math person:
  - Get **dependent variable** (y-axis) given one or more **independent variables** (x-axis).



Linear Regression

- Assumption: relationship between variables is **linear:**
  - (We'll relax linearity, study in more depth later.)

Population Y-Intercept

Population Slope

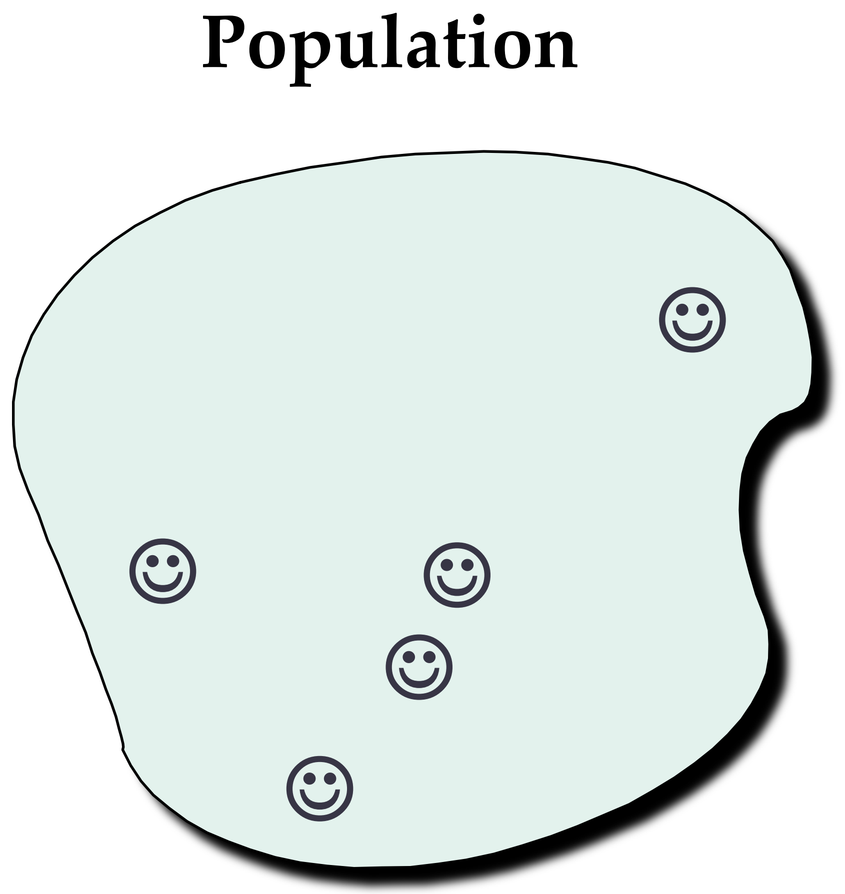
Random Error

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

Dependent Variable (e.g., ?????????)

Independent Variable(s) (e.g., ?????????)

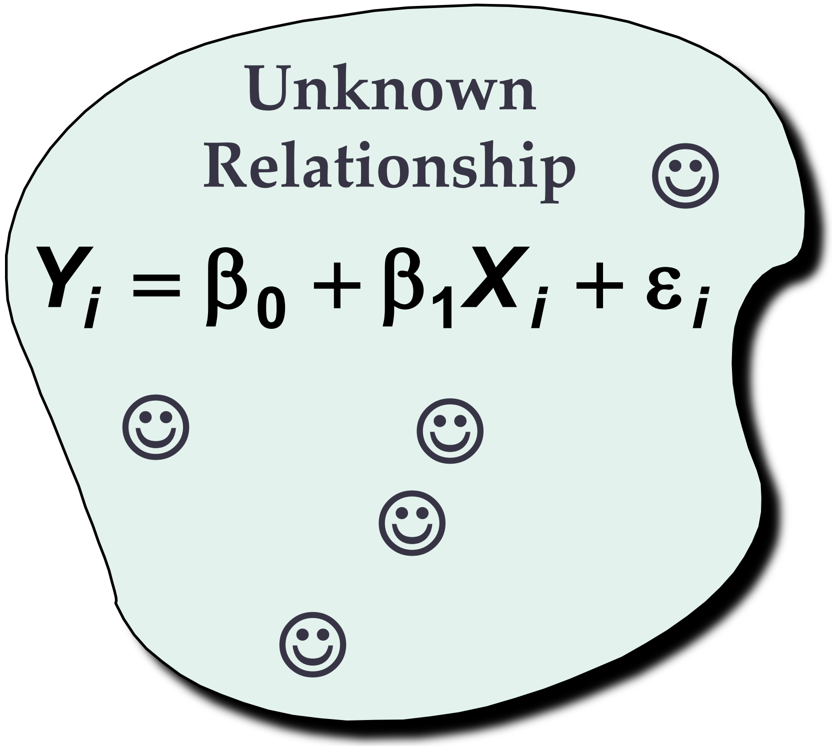
Population & Sample Regression Models



[WF]

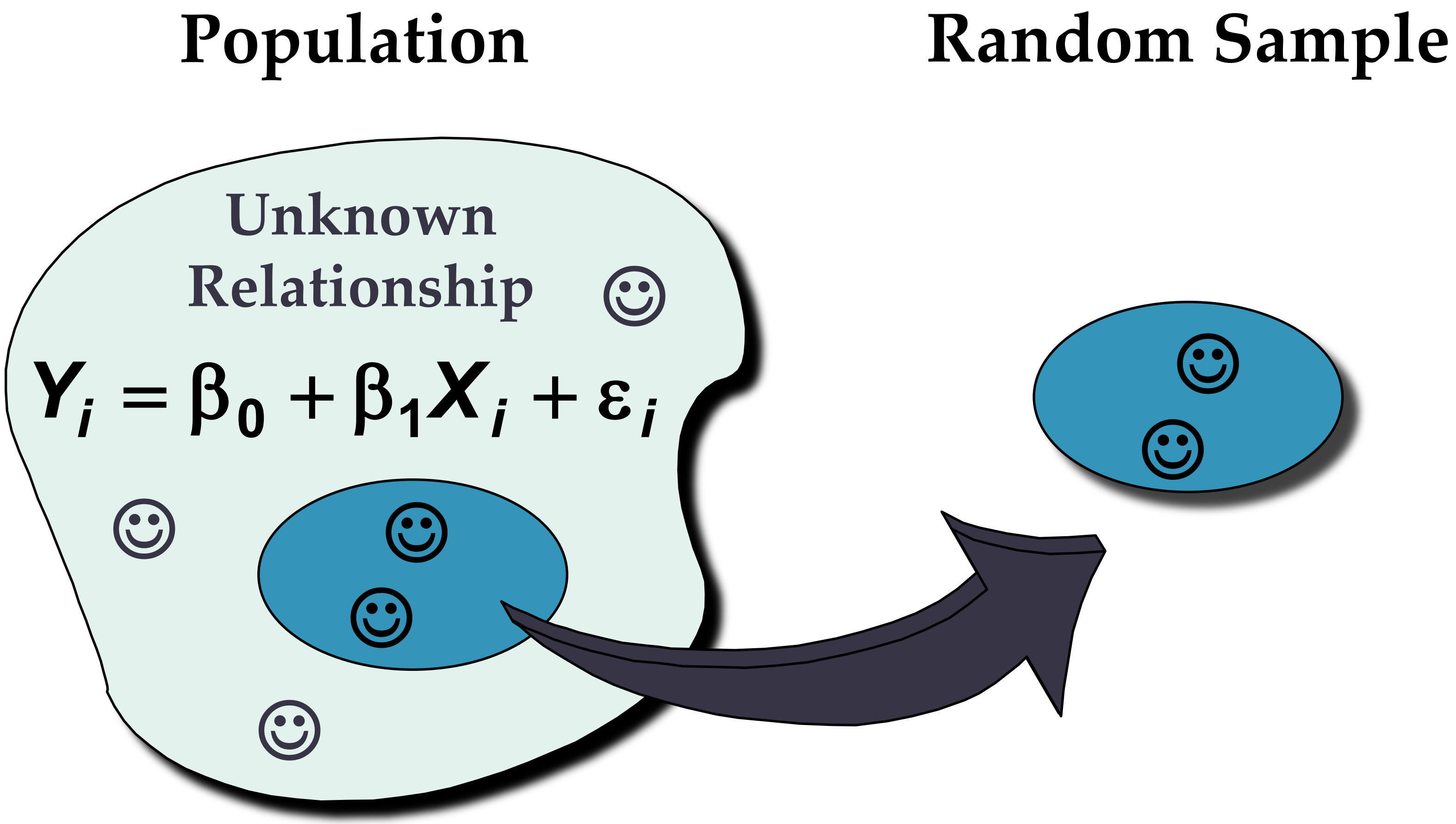
Population & Sample Regression Models

Population



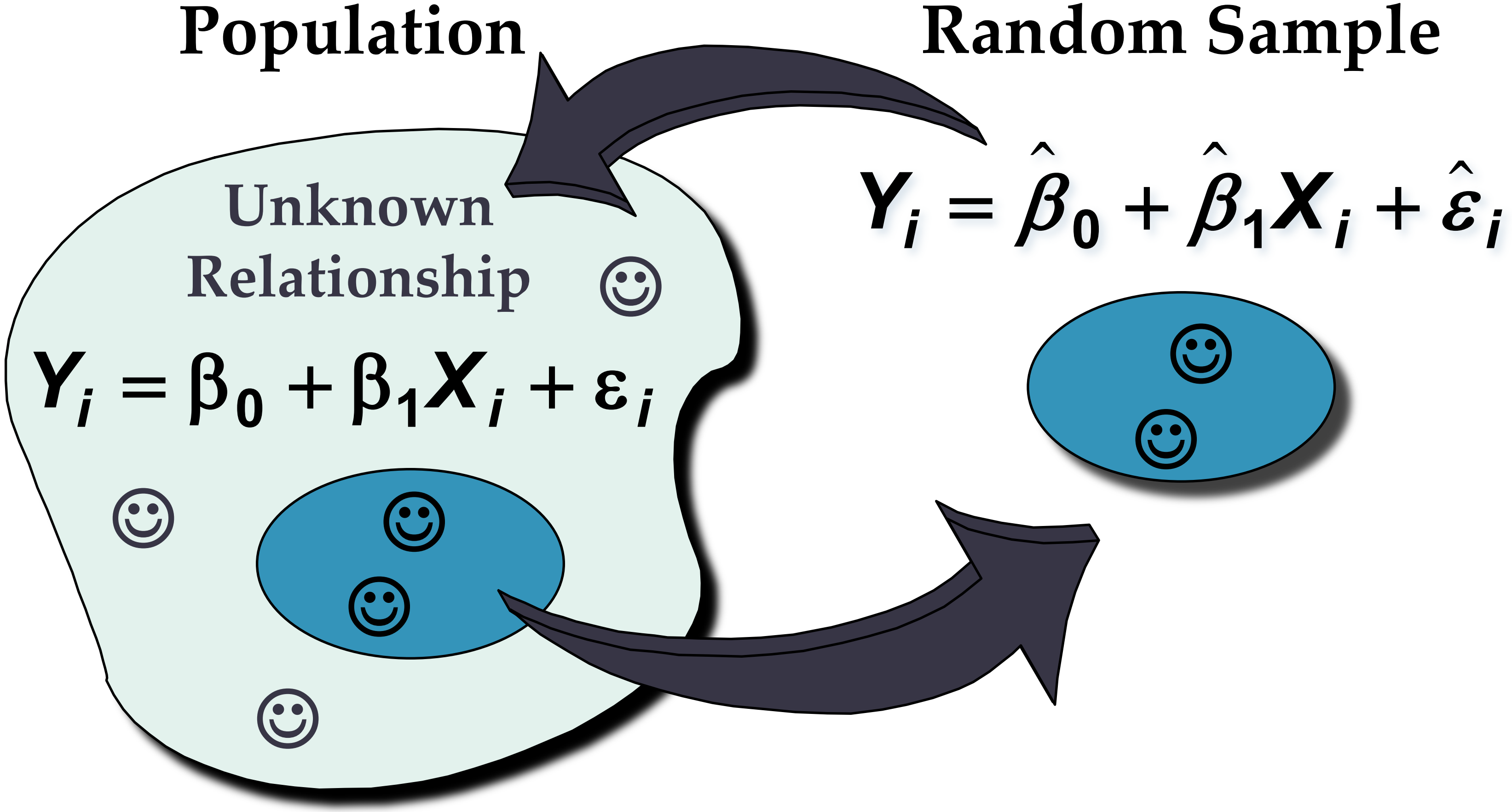
[WF]

Population & Sample Regression Models



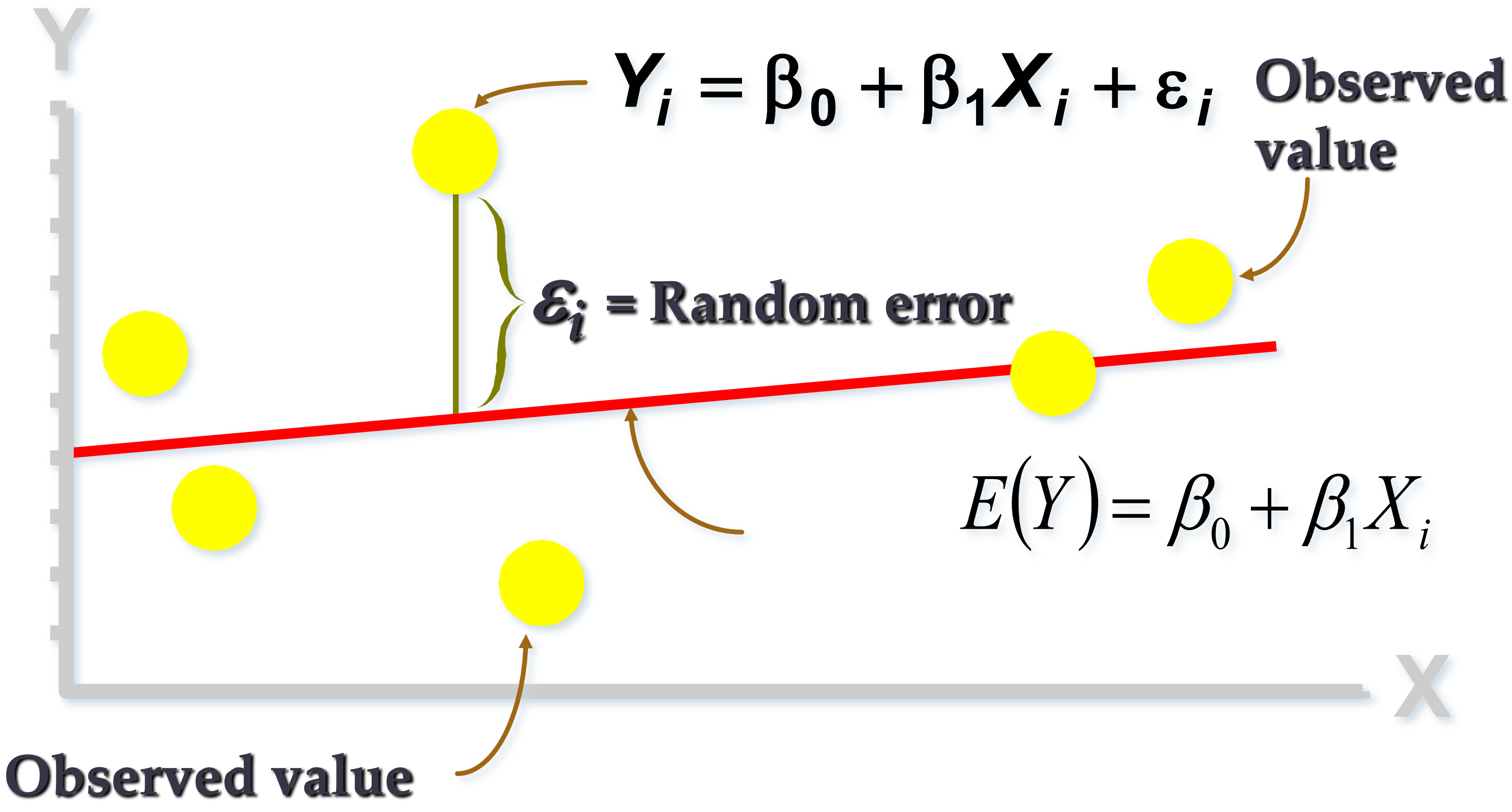
[WF

Population & Sample Regression Models



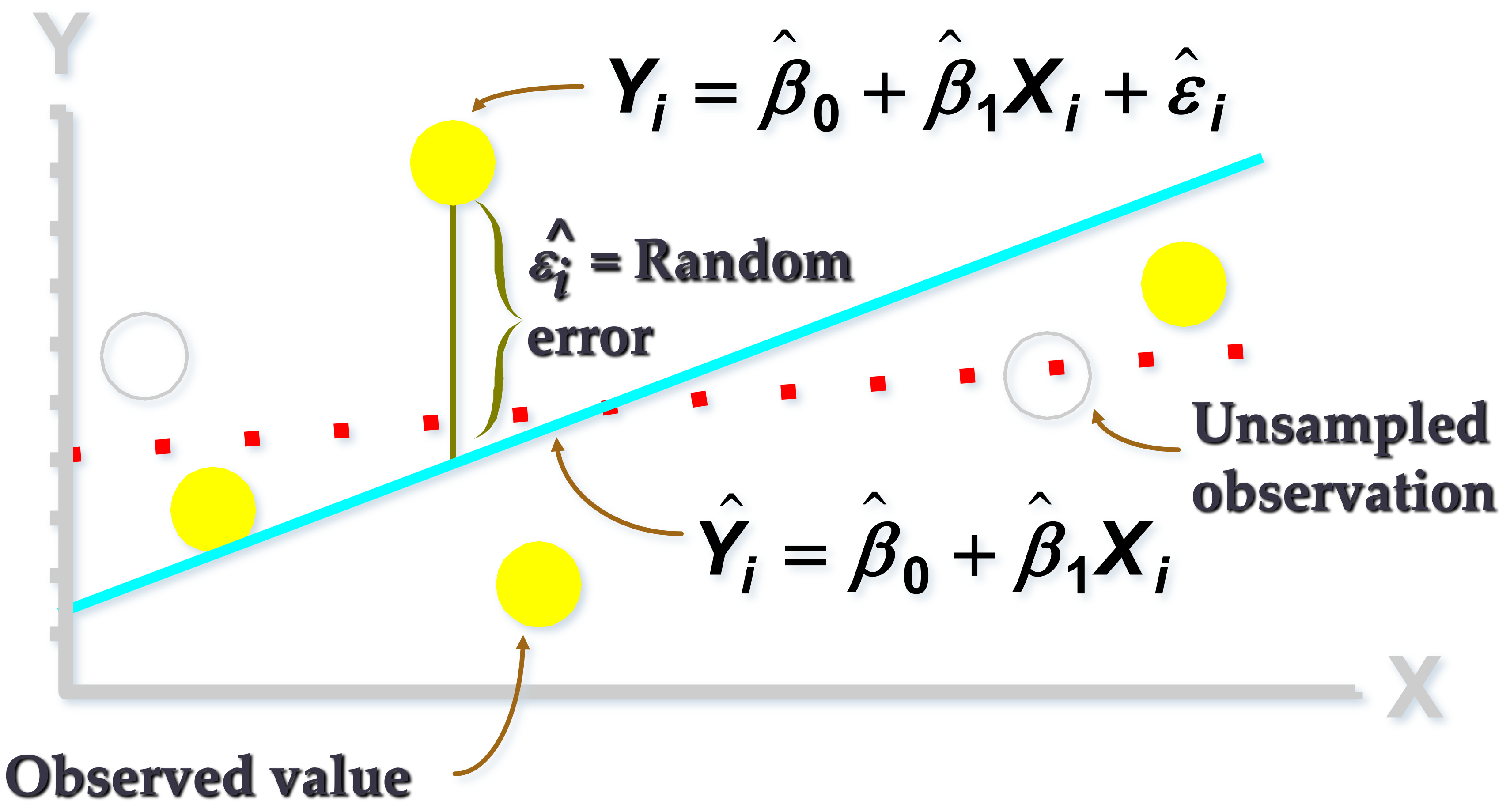
[WF]

Linear Regression



[WF]

Sample Linear Regression Model

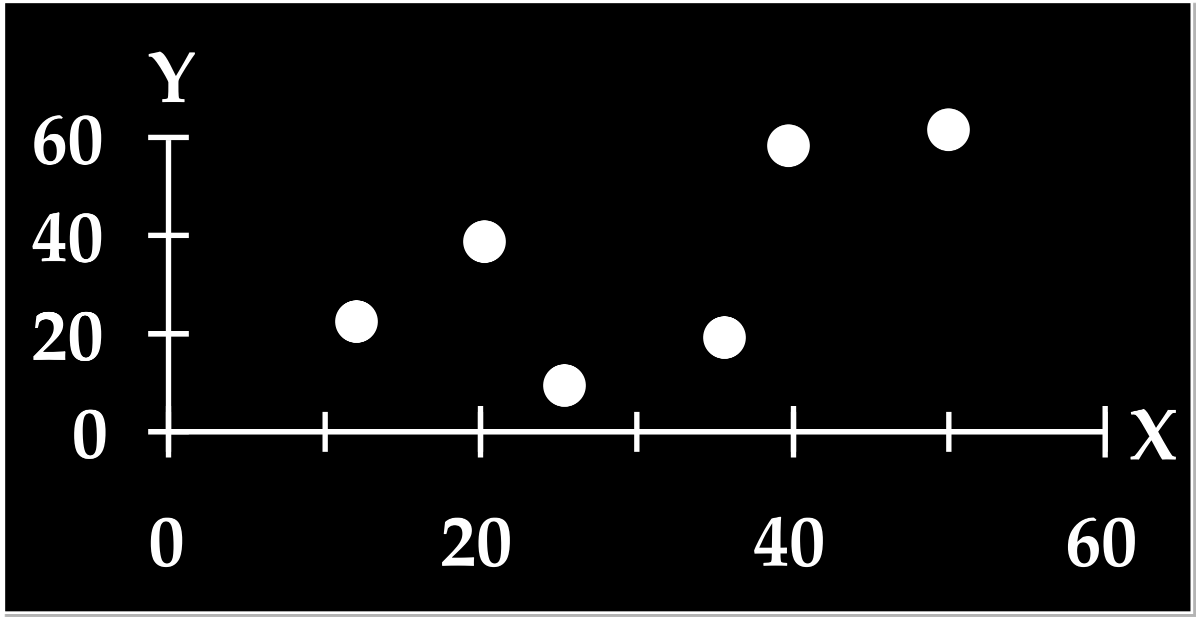


[WF]

Estimating Parameters: Least Squares Method

Scatter Plot

- Plot all (X_i, Y_i) pairs, and plot your learned model
- If you squint, suggests how well the model fits the data

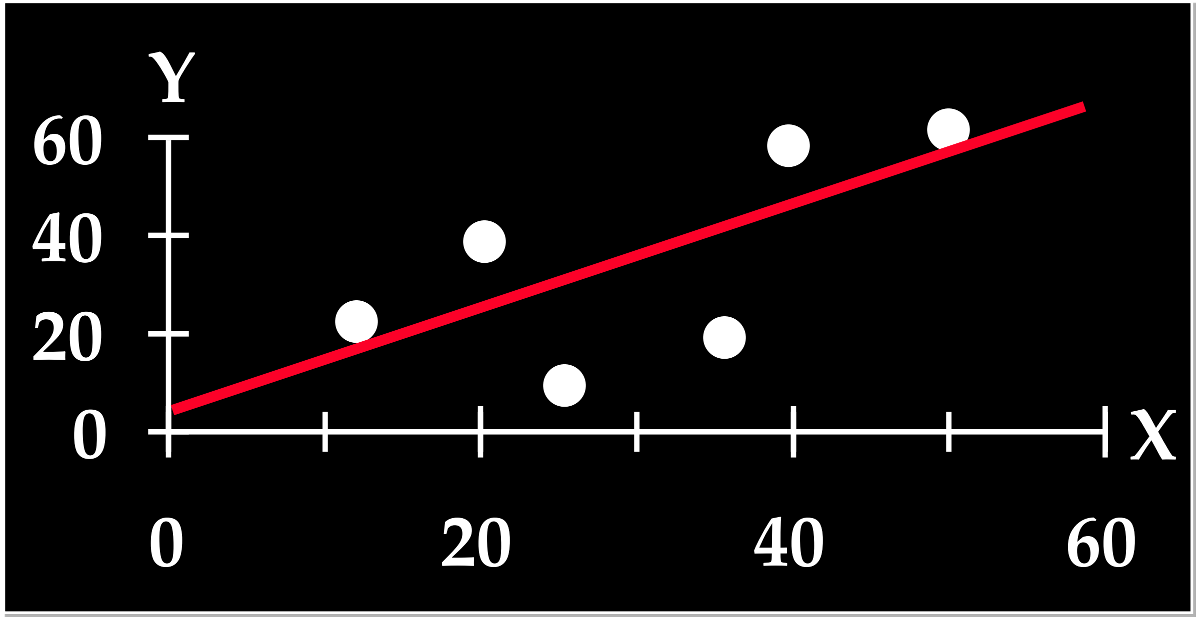


[WF]

Question

- How would you draw a line through the points?
- How do you determine which line "fits the best" …?

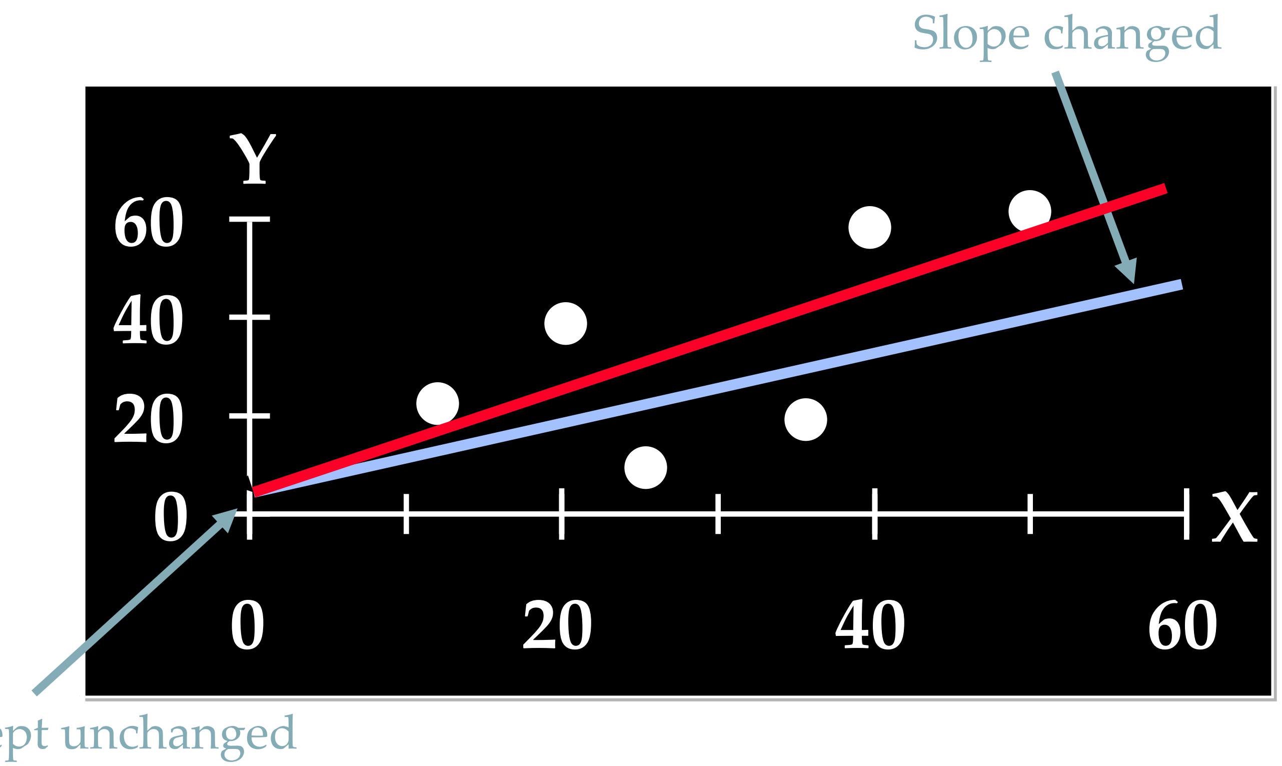
?????????



[WF]

Question

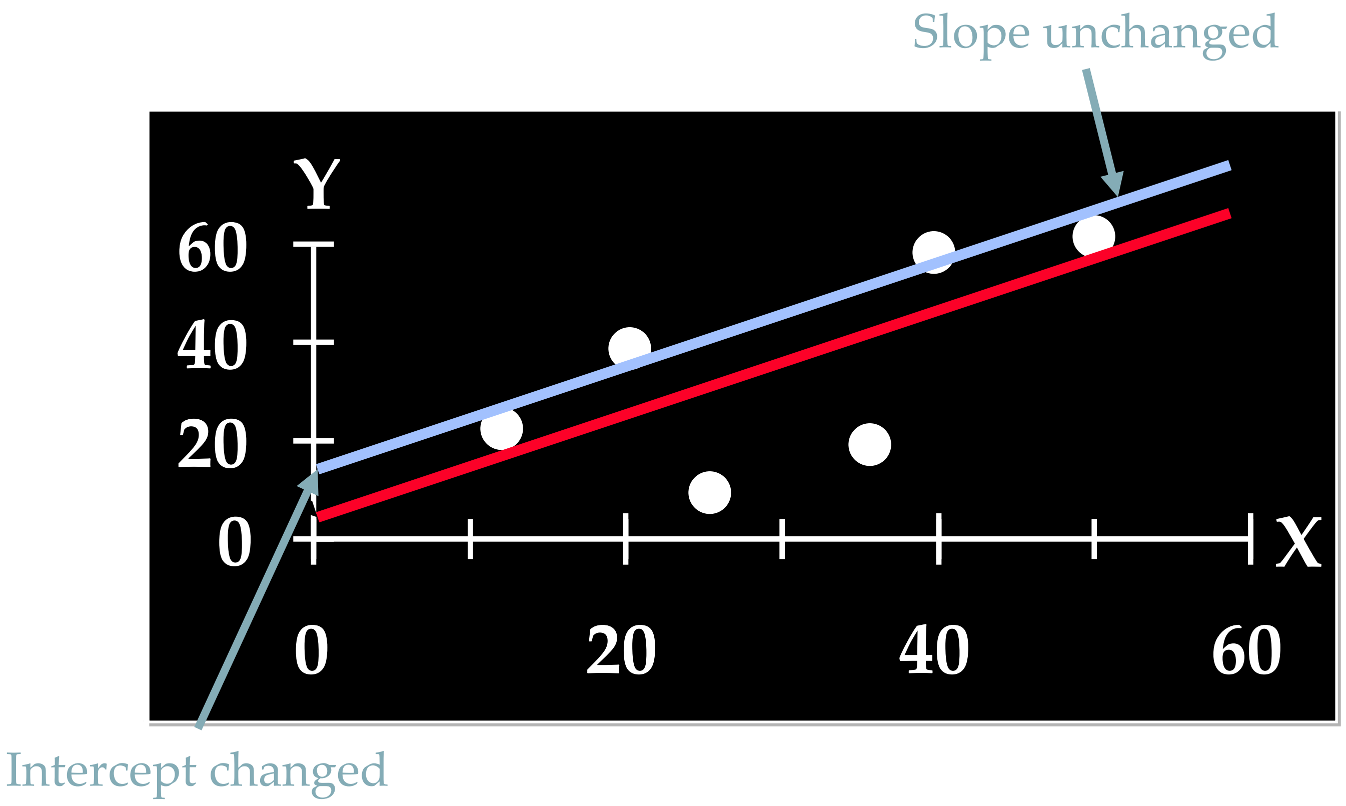
- How would you draw a line through the points?
- How do you determine which line “fits the best” ?????????



[WF]

Question

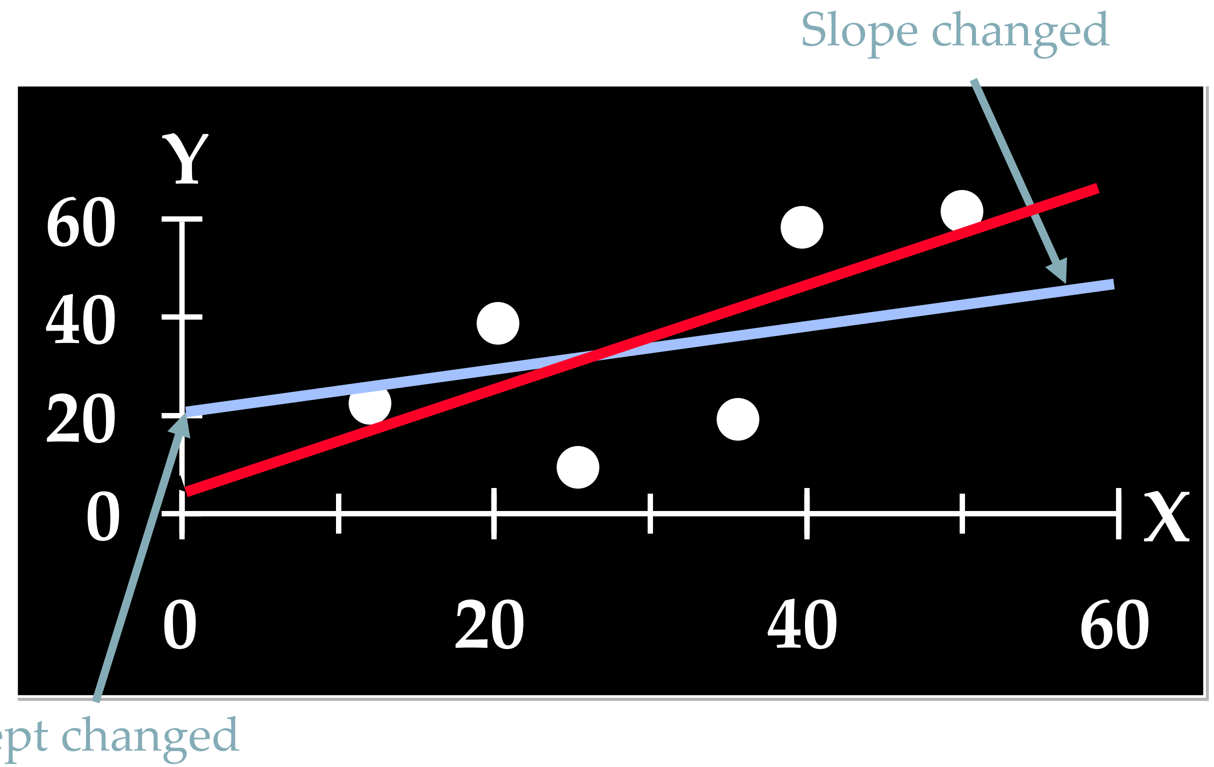
- How would you draw a line through the points?
- How do you determine which line “fits the best” ?????????



[WF]

Question

- How would you draw a line through the points?
- How do you determine which line “fits the best” ?????????



[WF]

Least Squares

- **Best Fit:** difference between the true Y-values and the estimated Y-values is minimized:
  - Positive errors offset negative errors …
  - … square the error!

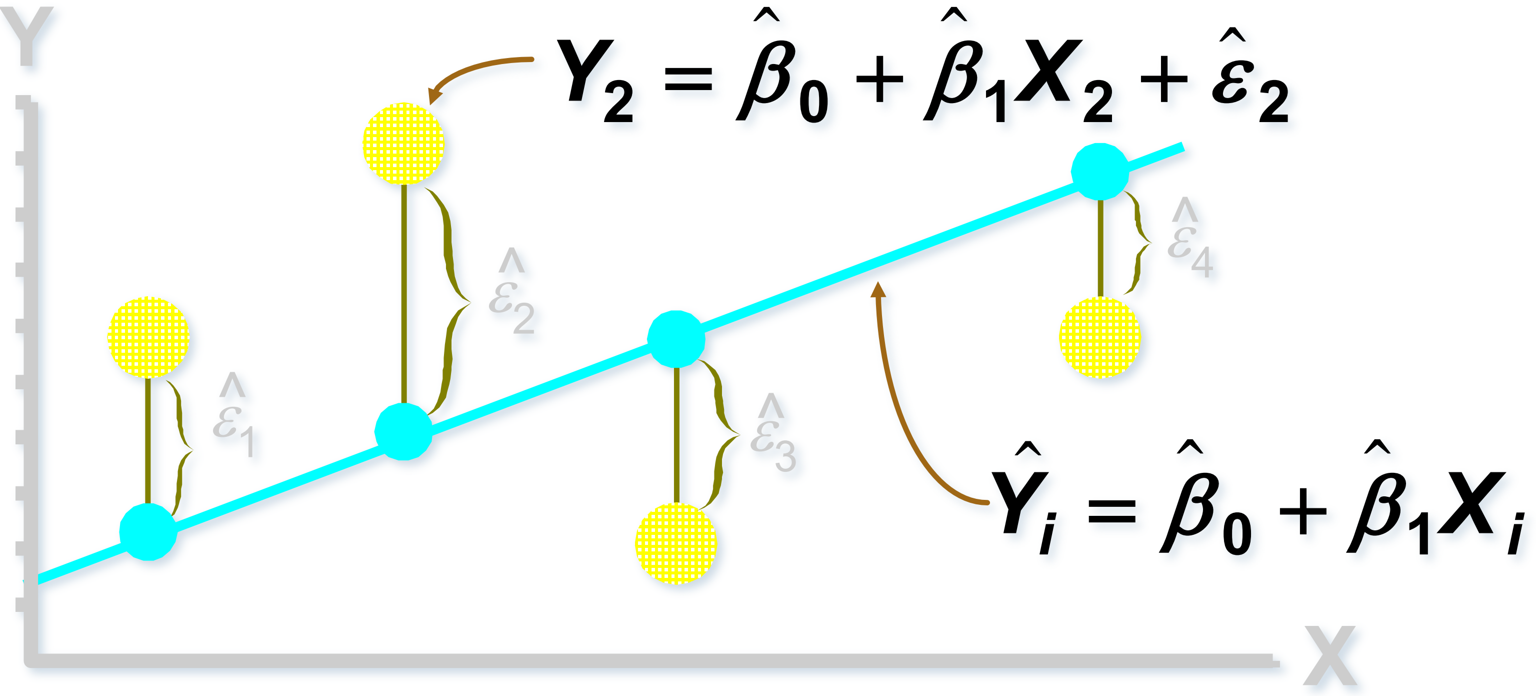
$$\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 = \sum_{i=1}^{n}\hat{\varepsilon}_i^2$$

- Least squares minimizes the sum of the squared errors
  - Why squared? We'll cover this in more depth in a few weeks.
  - Until then: http://www.benkuhn.net/squared

[WF]

Least Squares, Graphically

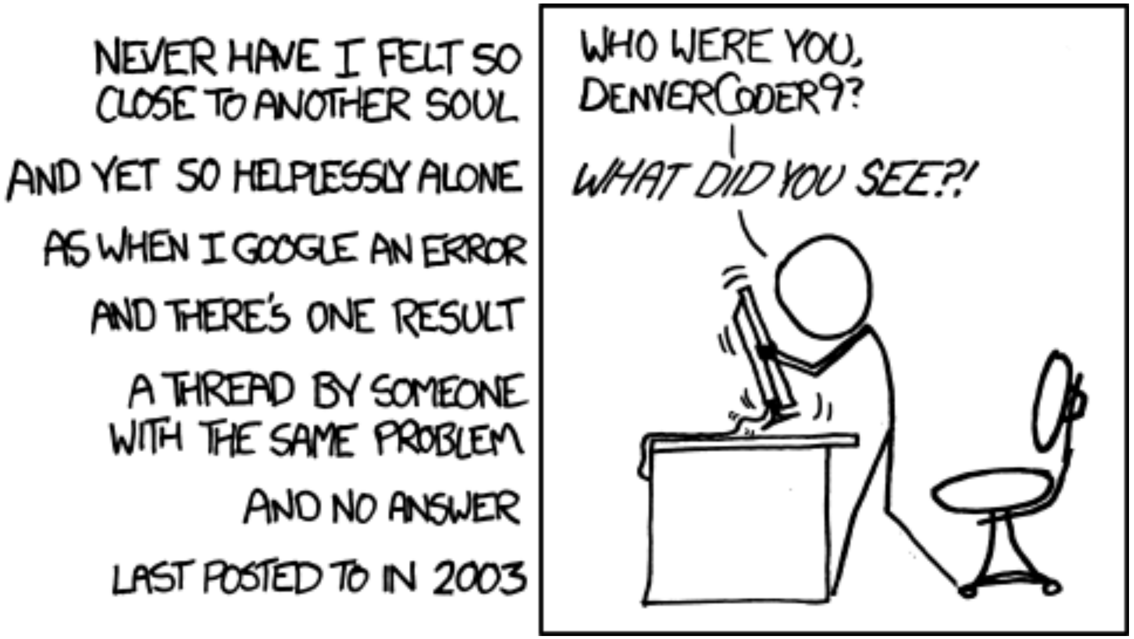
$$\textbf{LS minimizes} \sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \hat{\varepsilon}_1^2 + \hat{\varepsilon}_2^2 + \hat{\varepsilon}_3^2 + \hat{\varepsilon}_4^2$$



[WF]

Announcements

- Prof. Daniele is here!
- Questions 6 Fixed
- Milestone 1 Decompress (Updated Project 2) – Professionalism.
- Lab Friday – Be here with Laptop – new setup!
- More Missing Data!



https://xkcd.com/979/

Interpretation of Coefficients

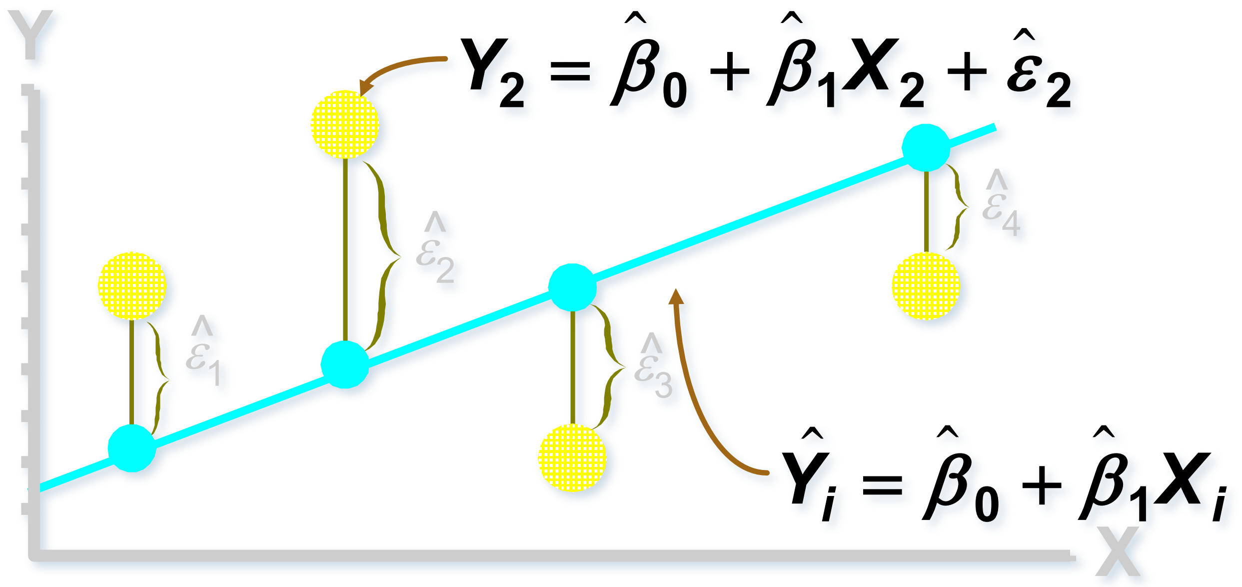
- Slope ($\hat{\beta}_1$):
  - Estimated Y changes by $\hat{\beta}_1$ for each unit increase in X.
  - If $\hat{\beta}_1 = 2$, then Y Is expected to increase by 2 for each 1 unit increase in X.

- Y-Intercept ($\hat{\beta}_0$):
  - Average value of Y when $X = 0$.
  - If $\hat{\beta}_0 = 4$, then average Y is expected to be 4 when X Is 0.

In-depth analysis to come!

[WF]

LS minimizes $$\sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \hat{\varepsilon}_1^2 + \hat{\varepsilon}_2^2 + \hat{\varepsilon}_3^2 + \hat{\varepsilon}_4^2$$



For an in depth derivation in matrix form:
https://www.stat.cmu.edu/~cshalizi/mreg/15/lectures/13/lecture-13.pdf

Now, Back to Missing Data …

Example

- **Question:** Does the circumference of certain body parts predict BF%?
- **Assumption:** BF% is a linear function of measurements of various body parts and other features …
- **Analysis:** Results from a regression model with BF% …

(Interpretation ???????????)

Predictor	Estimate	S.E.	p-value
Age	0.0626	0.0313	0.0463
Neck	-0.4728	0.2294	0.0403
Forearm	0.45315	0.1979	0.0229
Wrist	-1.6181	0.5323	0.0026

If you want to jump ahead on stats:
http://rpubs.com/nicholas_dirienzo/linear_regression_fall2019

[JA]

Hypothesis Testing

- One of the core ideas of Data Science – **Should be at the core of all the analysis that you do.**

- Define a **test statistic** - a quantity (numerical summary of a data-set that reduces the data to one value) derived from the sample used to test a hypothesis.

- **The Null Hypothesis:** *H0* = The feature of interest has no effect on the target.
- **The Alternative Hypothesis:** *H1* = The feature of interest does have an effect on the target.

- You're testing if H0 is true or not. If it is, then you say there's no relationship between our features. If it's not, then you say '*we reject H0 and accept H1*' – i.e., the features are related!

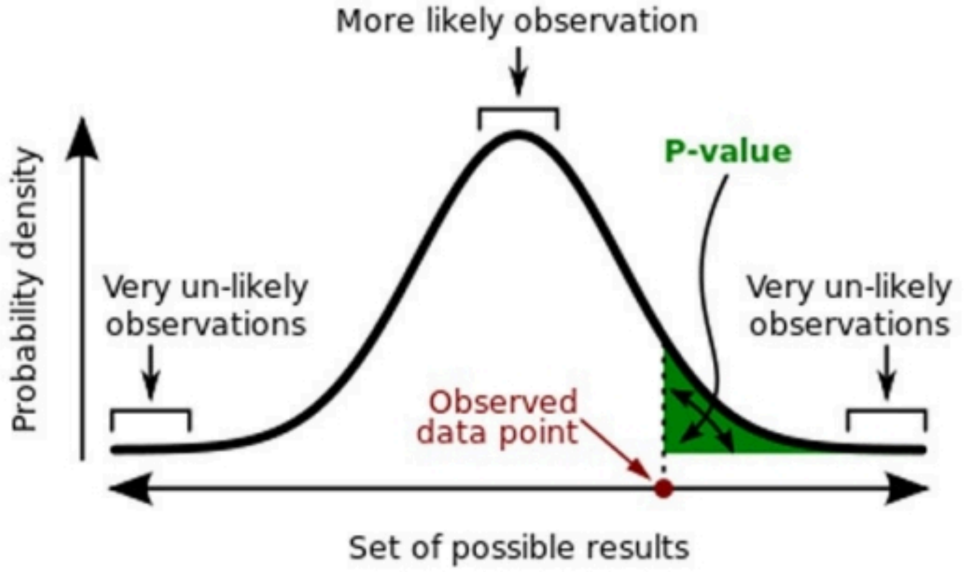
- This gets into some tricky logic that statisticians argue over as you actually aren't testing H1, but we're going to avoid that discussion here. Book Link (On Webpage): Introduction to Statistical Learning http://faculty.marshall.usc.edu/gareth-james/ISL/

P-values

- A **p-value** is is the probability of obtaining a result equal to or "more extreme" than what was actually observed, when the null hypothesis is actually true.

- **Null Hypothesis:** There is no significant difference between the specified populations. The observed difference is due to sampling or experiment error.

- Typically, when we have a p-value below a certain threshold, say 0.05, then we can reject the null hypothesis and say that there is an effect.



A **p-value** (shaded green area) is the probability of an observed (or more extreme) result assuming that the null hypothesis is true.

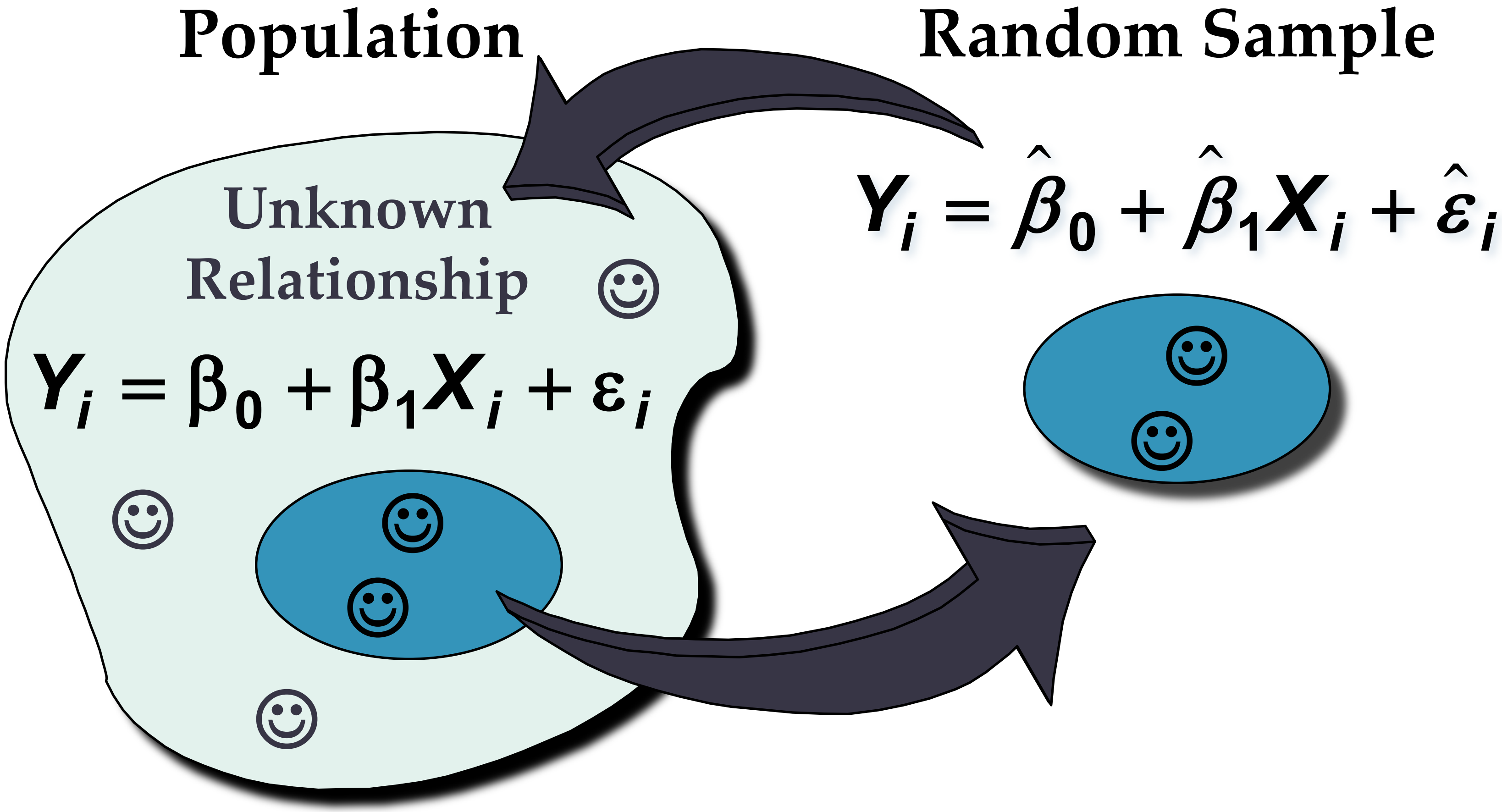
What If Data Were Missing?

- In this case, the dataset is complete:
  - But what if 5 percent of the participants had missing values? 10 percent? 20 percent?
- What if we performed complete case analysis and removed those who had missing values?

- Let's examine the effect if we do this if when the data is **missing completely at random (MCAR)**
  - Removed cases at random, reran analysis, stored the p-values
  - *p-value:* probability of getting at least as extreme a result as what we observed given that there is no relationship
  - Repeat 1000 times, plot *p-values of the hypothesis test on slope=0 (why?).*

[JA]

Population & Sample Regression Models



[WF]

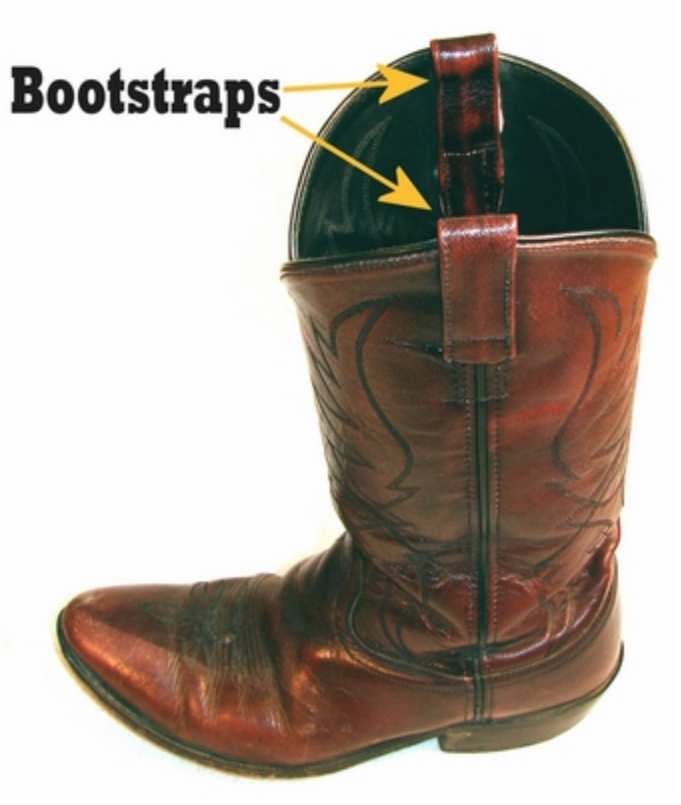
The Bootstrap

- What happens if we only have one sample? Why might we need more?

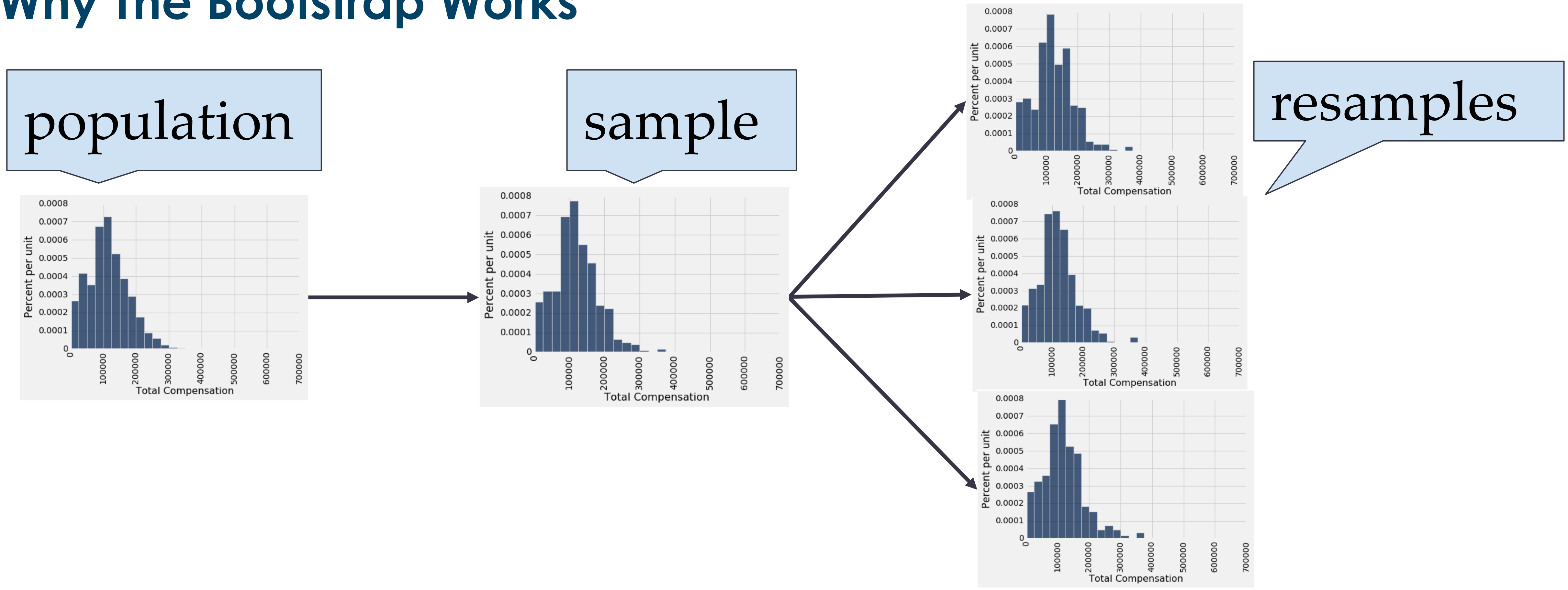
- All that we have is the original sample.
  - … which is large and random.
  - Therefore, it probably resembles the population.

- So we sample at random from the original sample!

- We'll dive more into this later but it's a powerful tool – can be used to calculate values you might not otherwise be able to estimate.



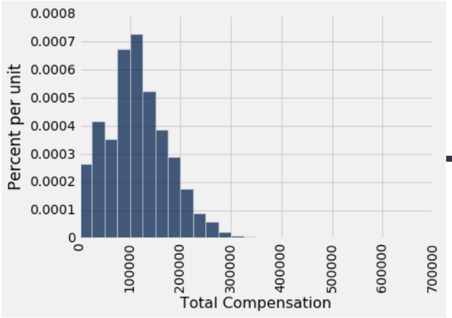
Why The Bootstrap Works



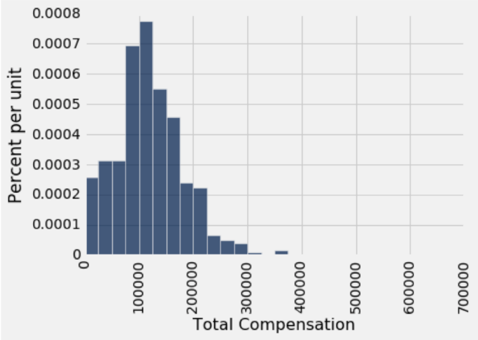
- All of these look similar… most likely…

Why We Need The Bootstrap

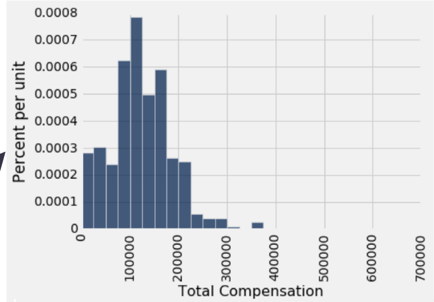
population

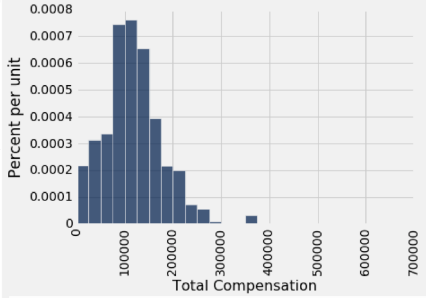


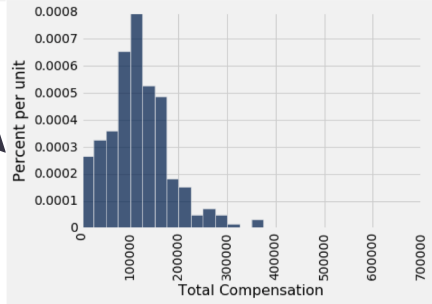
sample



resamples







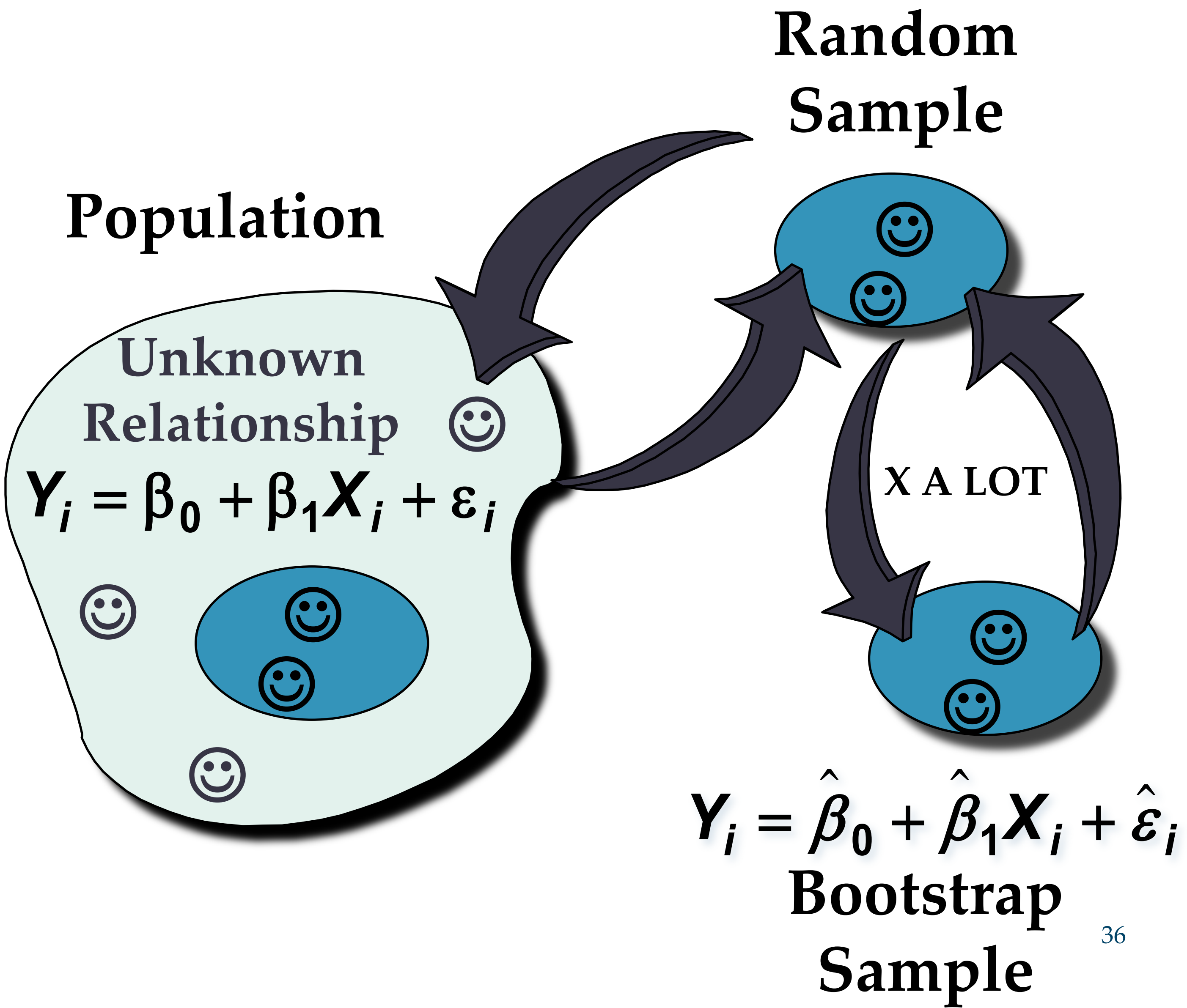
What we wish we could get

What we really get

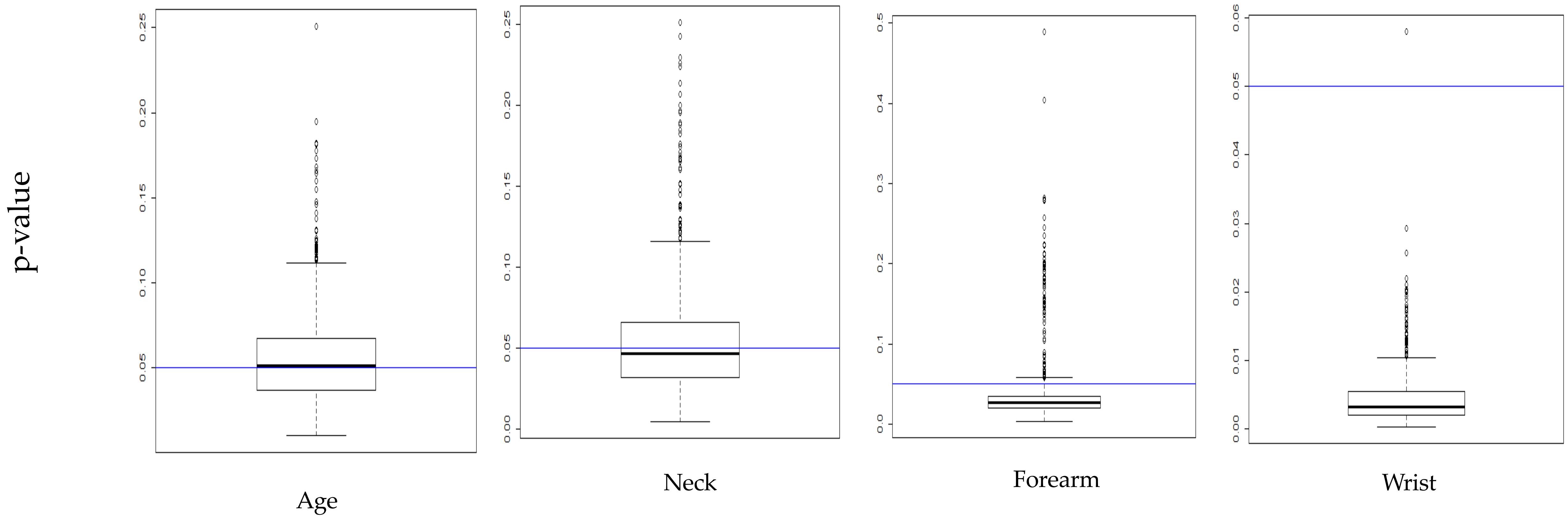
Key to Resampling

- From the original sample,
  - draw at random
  - with replacement
  - as many values as the original sample contained

- The size of the new sample has to be the same as the original one, so that the two estimates are comparable.

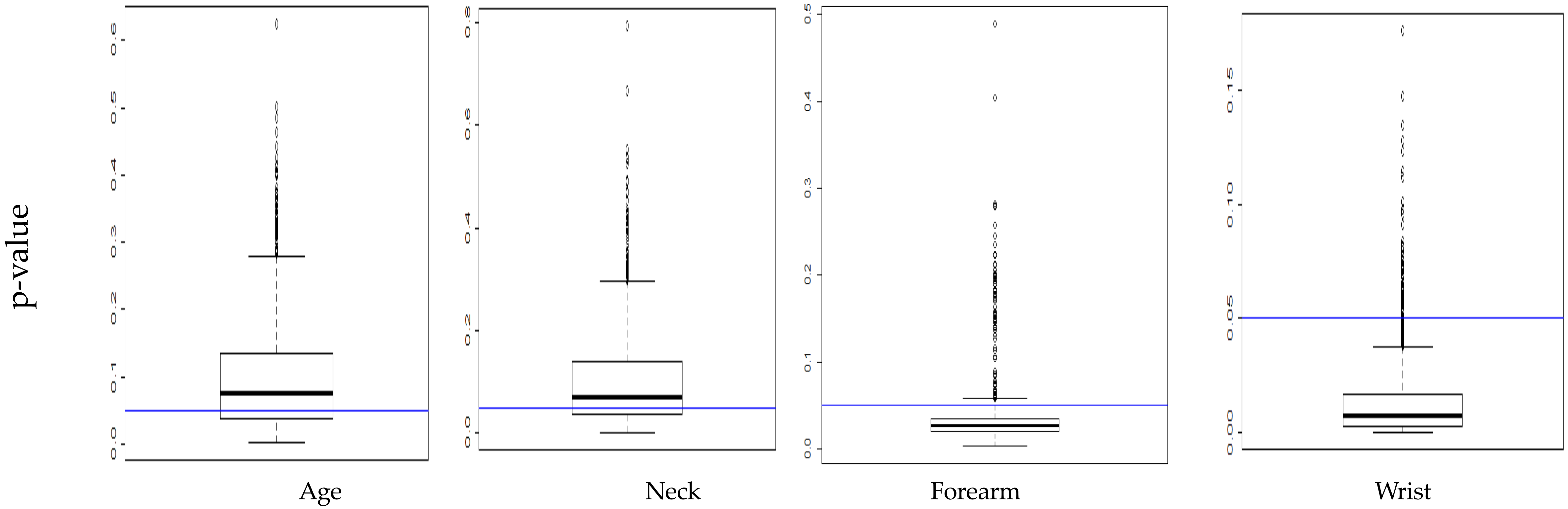


~5% Deleted (N=13)



[JA]

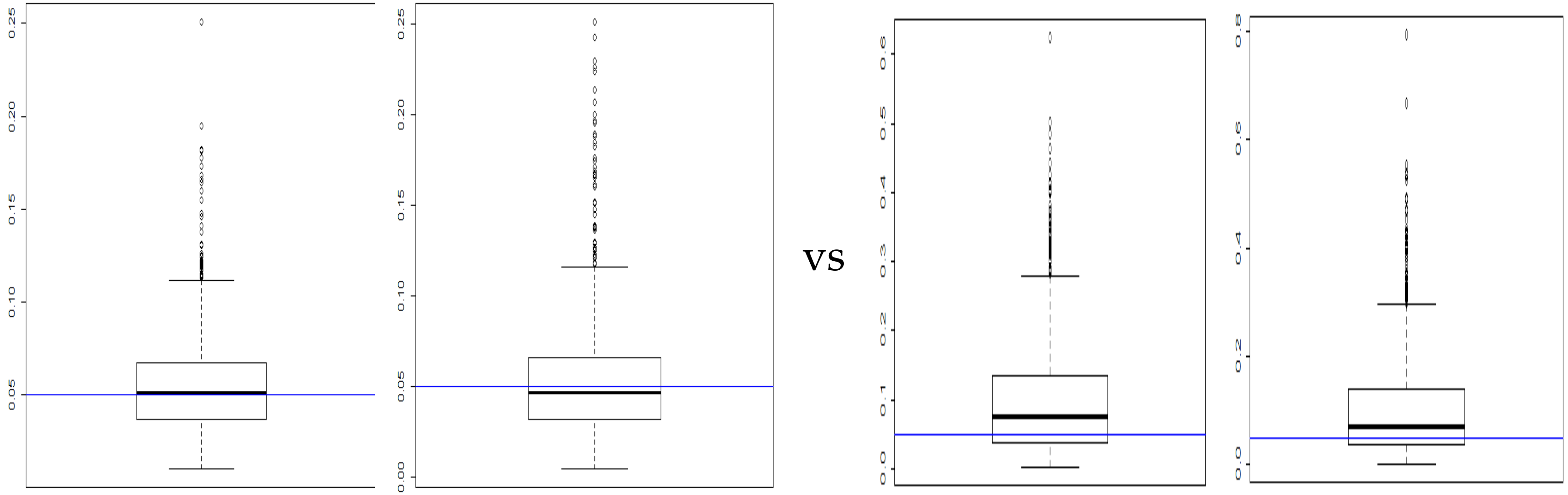
~20% Deleted (N=50)



[JA]

Conclusions seem to change …

- Age/Neck: fail to reject the null hypothesis usually?



Age (5%) Neck (5%) Age (20%) Neck (20%)

Still reject Forearm/Wrist most of the time

This is assuming the missing subjects' distribution does not differ from the non-missing. This would cause bias …

[JA]

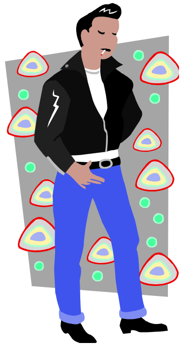
Types Of Missing-ness

- Missing Completely at Random (MCAR)
- Missing at Random (MAR)
- Missing Not at Random (MNAR)

[JA]

What Distinguishes Each Type of Missing-ness?

- Suppose you're loitering outside of STH one day …



Students just received their mid-semester grades

You start asking passing students their CMPS3660 grades

- **You don't force them to tell you or anything**
- **You also write down their gender and hair color**

[JA]

Your Sample

Hair Color	Gender	Grade
Red	M	A
Brown	F	A
Black	F	B
Black	M	A
Brown	M	
Brown	M	
Brown	F	
Black	M	B
Black	M	B
Brown	F	A
Black	F	
Brown	F	C
Red	M	
Red	F	A
Brown	M	A
Black	M	A

Summary:
- 7 students received As
- 3 students received Bs
- 1 student received a C

Nobody is failing!
- But 5 students did not reveal their grade …

[JA]

What Influences a Data Point's Presence?

- Same dataset, but the values are replaced with a "0" if the data point is observed and "1" if it is not

- **Question:** for any one of these data points, what is the probability that the point is equal to "1" …?
  - P(R) i.e., the probability that we didn't see it.

- What type of missing-ness do the grades exhibit?

Hair Color	Gender	Grade
0	0	0
0	0	0
0	0	0
0	0	0
0	0	1
0	0	1
0	0	1
0	0	0
0	0	0
0	0	0
0	0	1
0	0	0
0	0	1
0	0	0
0	0	0
0	0	0

[JA]

Experiment and Sample Space

- Experiment: a procedure that yields one of a given set of possible outcomes
  - Ex: flip a coin, roll two dice, draw five cards from a deck, etc.
- Sample space Ω: the set of possible outcomes
  - We focus on countable sample space: Ω is finite or countably infinite
  - In many applications, Ω is uncountable (e.g., a subset of $\mathbb{R}$)
- Event: a subset of the sample space
  - Probability is assigned to events
  - For an event $A \subseteq \Omega$, its probability is denoted by $\mathrm{P}(A)$
    - Describes beliefs about likelihood of outcomes

Examples

- Ex. 2: consider rolling a 6-sided biased (loaded) die
  - Sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$
  - Assume $P(3) = \frac{2}{7}, P(1) = P(2) = P(4) = P(5) = P(6) = \frac{1}{7}$
  - What is the probability of getting an odd number?

Let B denote the event of getting an odd number

$$B = \{1, 3, 5\}$$

$$P(B) = \frac{1}{7} + \frac{2}{7} + \frac{1}{7} = \frac{4}{7}$$

Independence

- Two events A and B are independent if and only if $\mathrm{P}(A \cap B) = \mathrm{P}(A)\,\mathrm{P}(B)$
- Ex. 4: Consider an experiment involving two successive rolls of a 4-sided die in which all 16 possible outcomes are equally likely and have probability $1/16$. Are the following pair of events independent?

(a) $A = \{\text{ 1st roll is 1 }\}$, $B = \{\text{ sum of two rolls is 5 }\}$ Yes

(b) $A = \{\text{ 1st roll is 4 }\}$, $B = \{\text{ sum of two rolls is 4 }\}$ No

Conditional Probability

Definition: Let E and F be events with $P(F) > 0$. The conditional probability of E given F, denoted by $P(E|F)$, is defined as:

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

Example: A bit string of length four is generated at random so that each of the 16 bit strings of length 4 is equally likely. What is the probability that it contains at least two consecutive 0s, given that its first bit is a 0?

Solution: Let E be the event that the bit string contains at least two consecutive 0s, and Let F be the event that the first bit is a 0.

- Since $E \cap F = \{0000, 0001, 0010, 0011, 0100\}$, $P(E \cap F) = 5/16$.
- Because 8 bit strings of length 4 start with a 0, $P(F) = 8/16 = ½$.

Hence,

$$P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{5/16}{1/2} = \frac{5}{8}$$

MCAR: Missing Completely at Random

- If this probability is not dependent on any of the data, observed or unobserved, then the data is **Missing Completely at Random (MCAR).**
- Suppose that X is the observed data (hair and gender) and Y is the unobserved data (grade).
- Call our "missing matrix" R.
- Then, if the data are MCAR, P(R|X,Y) = ??????????

$$P(R|X,Y) = P(R)$$

- Probability of those rows missing is **independent of the observed and unobserved data.**
- **I.e.,** the probability of that any given datapoint is missing is equal over the whole dataset. Each datum that is present had the same probability of being missing as each datum that is absent. Implies that ignoring the missing data will not bias inference.

Hair Color	Gender	Grade
0	0	0
0	0	0
0	0	0
0	0	0
0	0	1
0	0	1
0	0	1
0	0	0
0	0	0
0	0	0
0	0	1
0	0	0
0	0	1
0	0	0
0	0	0
0	0	0

[JA]

Totally Realistic MCAR Example

- You are running an experiment on plants grown in pots, when suddenly you have a nervous breakdown working on Project 1 and smash some of the pots
- You will probably not choose the plants to smash in a well-defined pattern, such as height age, etc.
- Hence, the missing values generated from your act of madness will likely fall into the MCAR category

[JA]

Applicability of MCAR

- A completely random mechanism for generating missing-ness in your data set just isn't very realistic
- Usually, missing data is missing for a reason:
  - Maybe older people are less likely to answer web-delivered questions on surveys.
  - In longitudinal studies people may die before they have completed the entire study.
  - Companies may be reluctant to reveal financial information.

MAR: Missing at Random

- **Missing at Random (MAR):** probability of missing data is dependent on the observed data but not the unobserved data.
- Suppose that X is the observed data and Y is the unobserved data. Call our "missing matrix" R.
- Then, if the data are MAR, $P(R|X,Y)$ = ??????????

$$P(R|X,Y) = P(R|X)$$

- Not exactly random (in the vernacular sense).
  - There is a probabilistic mechanism that is associated with whether the data is missing.
  - Mechanism takes the observed data as input.

Hair Color	Gender	Grade
0	0	0
0	0	0
0	0	0
0	0	0
0	0	<u>1</u>
0	0	<u>1</u>
0	0	<u>1</u>
0	0	0
0	0	0
0	0	0
0	0	<u>1</u>
0	0	0
0	0	<u>1</u>
0	0	0
0	0	0
0	0	0

Examples?

?

- MAR allows for data to be missing according to a random process, but is more general than MCAR -- all units do not have equal probabilities of being missing.
- Missingness may only depend on information that is fully observed!
- For example, the reporting of income on surveys may vary according to some measured factor, such as age, race or sex. We can thus account for heterogeneity in the probability of reporting income by controlling for the measured covariate in whatever model is used for infrence.

MAR: Key Point

- We can **model** that latent mechanism and compensate for it!
- **Imputation**: replacing missing data with substituted values.
  - Models today will assume MAR.
  - Example: if age is known, you can model missing-ness as a function of age.

- Whether or not missing data is MAR or the next type, Missing Not at Random (MNAR), is not* testable.
  - Requires you to "understand" your data!

Hair Color	Gender	Grade
0	0	0
0	0	0
0	0	0
0	0	0
0	0	<u>1</u>
0	0	<u>1</u>
0	0	<u>1</u>
0	0	0
0	0	0
0	0	0
0	0	<u>1</u>
0	0	0
0	0	<u>1</u>
0	0	0
0	0	0
0	0	0

*unless you can get the missing data (e.g., post-study phone calls).

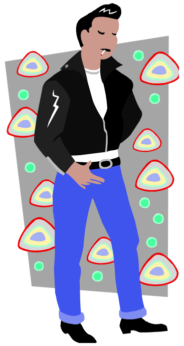
MNAR: Missing Not at Random

- **MNAR:** missing-ness has something to do with the missing data itself!
- Examples: ??????????
  - Do you binge drink? Do you have a trust fund? Do you use illegal drugs? What is your sexuality? Are you depressed?

- Said to be "non-ignorable":
  - Missing data mechanism must be considered as you deal with the missing data.
  - Must include model for why the data are missing, and best guesses as to what the data might be.

Back to STH ...

- Is the the missing data:
- MCAR;
- MAR; or
- MNAR?
- ???????????



Hair Color	Gender	Grade
Red	M	A
Brown	F	A
Black	F	B
Black	M	A
Brown	M	
Brown	M	
Brown	F	
Black	M	B
Black	M	B
Brown	F	A
Black	F	
Brown	F	C
Red	M	
Red	F	A
Brown	M	A
Black	M	A

Add a Variable

- Bring in the GPA:
- Does this change anything?

Hair Color	GPA	Gender	Grade
Red	3.4	M	A
Brown	3.6	F	A
Black	3.7	F	B
Black	3.9	M	A
Brown	2.5	M	
Brown	3.2	M	
Brown	3.0	F	
Black	2.9	M	B
Black	3.3	M	B
Brown	4.0	F	A
Black	3.65	F	
Brown	3.4	F	C
Red	2.2	M	
Red	3.8	F	A
Brown	3.8	M	A
Black	3.67	M	A

To Recap

- Assume that we have a matrix where X is the observed data, Y is the unobserved data, and R is Call our "missing matrix" R.
- **Missing Completely at Random (MCAR).**
  - Missingness does not depend on observed or unobserved data.
  - $P(R|X,Y) = P(R)$
- **Missing at Random (MAR).**
  - Missingness depends *only* on observed data.
  - $P(R|X,Y) = P(R|X)$
- **Missing Not At Random (MNAR).**
  - Neither MCAR or MAR.

Hair Color	GPA	Gender	Grade
Red	3.4	M	A
Brown	3.6	F	A
Black	3.7	F	B
Black	3.9	M	A
Brown	2.5	M	
Brown	3.2	M	
Brown	3.0	F	
Black	2.9	M	B
Black	3.3	M	B
Brown	4.0	F	A
Black	3.65	F	
Brown	3.4	F	C
Red	2.2	M	
Red	3.8	F	A
Brown	3.8	M	A
Black	3.67	M	A

Handling Missing Data …

Single Imputation

- **Mean Imputation:** imputing the average from observed cases for all missing values of a variable.

- **Hot-deck Imputation:** imputing a value from another subject, or "donor," that is most like the subject in terms of observed variables.
  - Last observation carried forward (LOCF): order the dataset somehow and then fill in a missing value with its neighbor.
  - E.g., In evaluations of interventions where pre-treatment measures of the outcome variable are also recorded, a strategy that is sometimes used is to replace missing outcome values with the pre-treatment measure.

- **Cold-deck Imputation:** bring in other datasets.
- Old and Busted:
  - All fundamentally impose too much precision.
  - Have uncertainty over what unobserved values actually are.
  - Developed before cheap computation.

Multiple Imputation

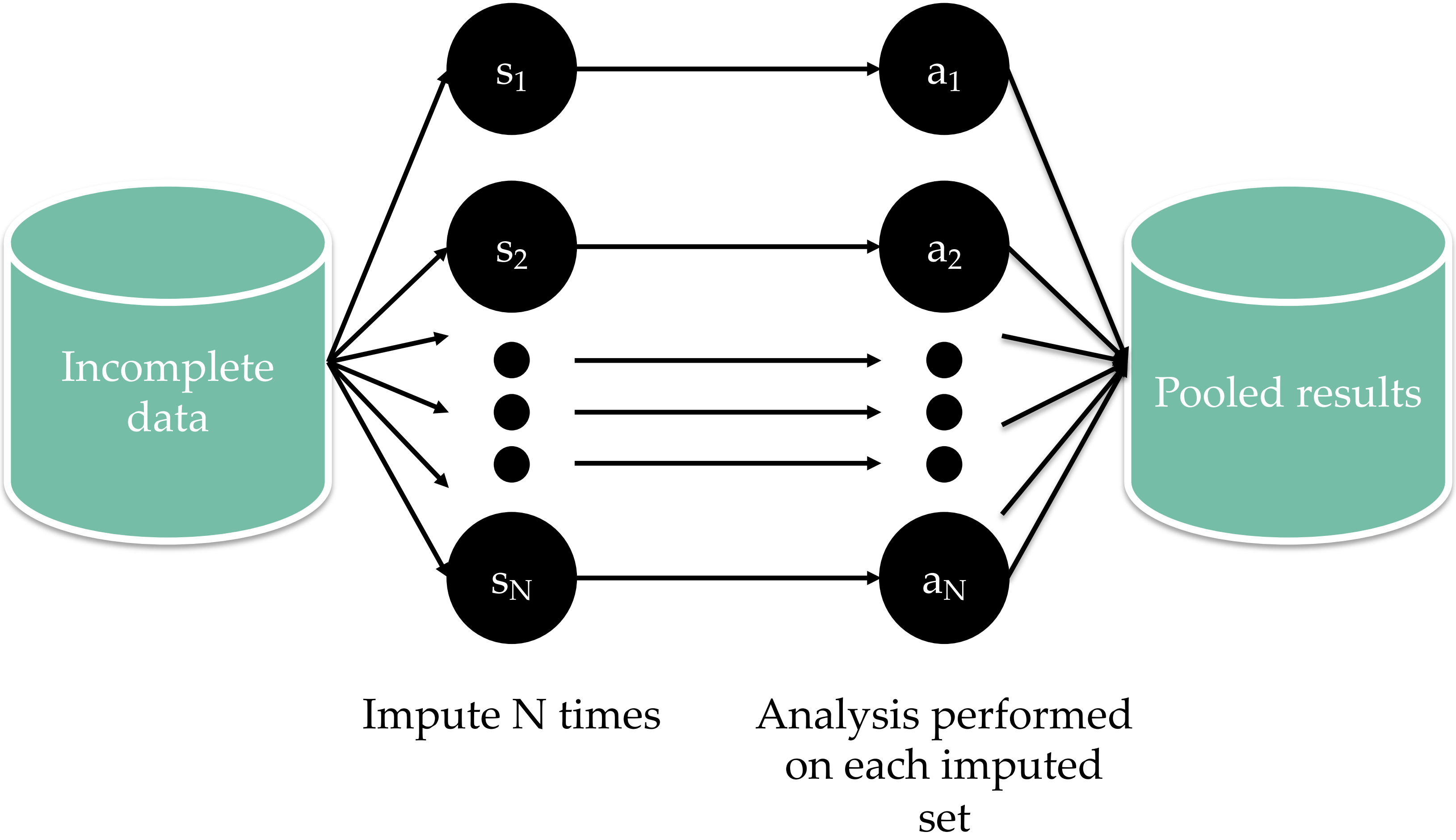
- Developed to deal with noise during imputation!
  - Impute once → treats imputed value as observed.

- We have uncertainty over what the observed value would have been

- **Multiple Imputation:** generate several random values for each missing data point during imputation and pool the results!

Multiple Imputation

- Developed to deal with noise during imputation!
  - Impute once → treats imputed value as observed.
- We have uncertainty over what the observed value would have been
- **Multiple Imputation:** generate several random values for each missing data point during imputation and pool the results!



Tiny Example

X	Y
32	2
43	?
56	6
25	?
84	5

Independent variable: X
Dependent variable: Y
We assume Y has a linear relationship with X

Let’s Impute Some Data!

- Use a predictive distribution of the missing values (how??):
  - Given the observed values, make random draws of the observed values and fill them in.
  - Do this N times and make N imputed datasets.

X	Y
32	2
43	5.5
56	6
25	8
84	5

X	Y
32	2
43	7.2
56	6
25	1.1
84	5

For very large values of N=2 …

Inference with Multiple Imputation

- Now that we have our imputed data sets, how do we make use of them? ???????????
  - Analyze each of the **separately.**

X	Y
32	2
43	5.5
56	6
25	8
84	5

Slope	-0.8245
Standard error	6.1845

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

X	Y
32	2
43	7.2
56	6
25	1.1
84	5

Slope	4.932
Standard error	4.287

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

Pooling analyses

- **Pooled Slope Estimate** is the average of the N imputed estimates.
- Our example, $\beta_{1p} = \frac{\beta 11+\beta 12}{2}$ = (4.932-.8245) x 0.5 = 2.0538

- The pooled slope **variance** is given by:
  - $s = \frac{\sum Zi}{N} + (1 + \frac{1}{N})$ x $\frac{1}{\mathrm{N}-1} * \sum(\beta 1i \; - \beta_{1p} \,)^2$

- Where Z_i is the standard error of the imputed slopes.

- Our example: (4.287 + 6.1845)/2 + (3/2)*(16.569) = 30.08925

- **Standard Error:** take the square root, and we get 5.485.

Predicting Missing Data Given the Observed Data

- Comprehensive treatment: http://www.bias-project.org.uk/Missing2012/Lectures.pdf and http://www.stat.columbia.edu/~gelman/arm/missing.pdf and https://www4.stat.ncsu.edu/~post/suchit/bayesian-methods-incomplete.pdf

- Given events A, B; and P(A) > 0 …
- Bayes' Theorem:

$$P(B|A) = \frac{P(A|B) * P(B)}{P(A)}$$

Probability of seeing evidence given the hypothesis

- In our case:

$$P(\mathbf{H}|\mathbf{E}) = \frac{P(\mathbf{E}|\mathbf{H}) * P(\mathbf{H})}{P(\mathbf{E})}$$

Prior probability of hypotheses

Posterior probability of the hypothesis given the evidence

Prior over the evidence

Bayesian Imputation

- Establish a **prior** distribution:
  - Some distribution of parameters of interest θ before considering the data, *P(θ).*
  - We want to estimate θ.

- Given θ, can establish a distribution $P(X_{obs}/\theta)$

- Use Bayes Theorem to establish $P(\theta/X_{obs})$ …
  - Make random draws for θ.
  - Use these draws to make predictions of $Y_{miss.}$

$$P(\theta|X_{obs}) = \frac{P(X_{obs}|\theta) * P(\theta)}{P(X_{obs})}$$

How Big Should N Be?

- Number of imputations N depends on:
  - Size of dataset
  - Amount of missing data in the dataset
- Some previous research indicated that a small N is sufficient for efficiency of the estimates, based on:
  - $(1 + \frac{\lambda}{N}) - 1$
  - N is the number of imputations and λ is the fraction of missing information for the term being estimated [Schaffer 1999]
- More recent research claims that a good N is actually higher in order to achieve higher power [Graham et al. 2007]

More Advanced Methods

- Interested? Further reading:
- Regression-based MI methods
- Multiple Imputation Chained Equations (MICE) or Fully Conditional Specification (FCS)
  - Readable summary from JHU School of Public Health: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3074241/
- Markov Chain Monte Carlo (MCMC)
  - A bit more complicated - http://stronginference.com/missing-data-imputation.html

- Comprehensive Tutorials (Grad level Stats):
  - http://www.bias-project.org.uk/Missing2012/Lectures.pdf
  - http://www.stat.columbia.edu/~gelman/arm/missing.pdf
  - https://www4.stat.ncsu.edu/~post/suchit/bayesian-methods-incomplete.pdf