Psyc 60: Intro to Statistics

Prof. Judith Fan

Spring 2021

Last Time

Mini-Review Session #2

Modeling data with the mean

Thinking about variability as model error

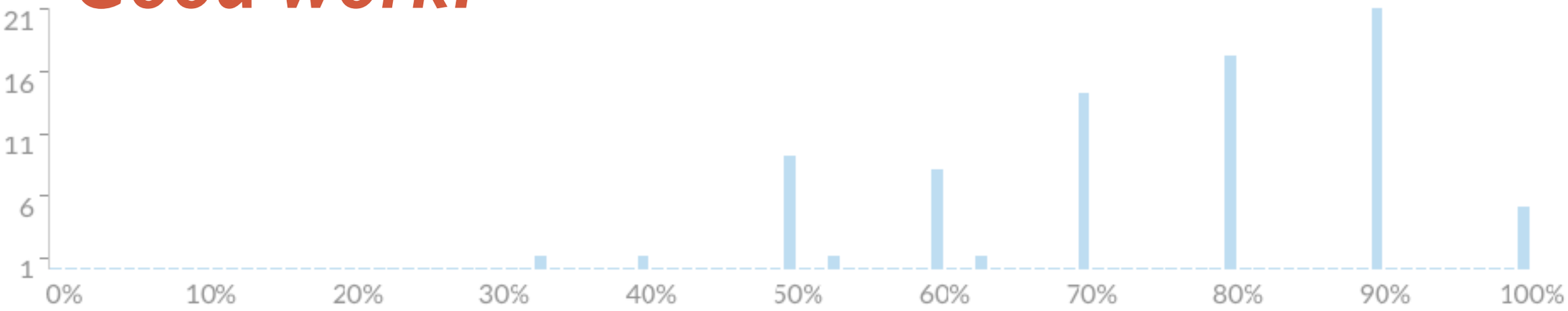
Weaving all these ideas together

Starring Jeffrey Xing!

Quiz 3

Average Score	High Score	Low Score	Standard Deviation	Average Time
75%	100%	33%	1.59	18:32

Good work!



Please review the questions you missed & post your questions to Slack and/or come talk to us about them in office hours.

Quiz 3: review

In the WellbeingStudy data frame, researchers are studying how stress level (**Stress**) predicts variation in the average hours of sleep per night (**AvgSleepHours**) undergraduate students at UCSD get through the Winter quarter of 2021. Given that the Stress data is continuous, choose the line of R code that divides the stress level into three groups -- **low**, **medium**, and **high,** and stores it as **Stress3Group** within the WellbeingStudy data frame.

WellbeingStudy$AvgSleepHours3Group <- ntile(WellbeingStudy$AvgSleepHours, 3)	1 respondent	1 %	
WellbeingStudy$Stress3Group <- ntile(WellbeingStudy$Stress, 3)	46 respondents	59 %	✓
WellbeingStudy$Stress3Group <- c("low", "medium", "high")	30 respondents	38 %	
ntile(WellbeingStudy$Stress, 3)	1 respondent	1 %	

Recall that the* ntile() *function will subdivide continuous variables into some number of bins, which you define.

Quiz 3: review

Consider a dataset named **BikeCommute**, where the column **Bike** stores the type of bike each biker is riding (either Steel or Carbon). If you run the following line of code, what information will it give you?

```
tally(~Bike==Steel, data = BikeCommute, format="proportion")
```

Number of Steel bikes in the dataset	4 respondents	5 %
Nothing! The code will return an error	19 respondents	24 % ✓
Total number of Bikes in the dataset		0 %
Proportion of Steel bikes in the dataset	55 respondents	71 %

Look carefully! If 'Steel' is the name of a material that a bike could be made out of, it will be wrapped in quotes in the dataframe. Without quotes, you'll get an error message something like:

```
Error in eval(x, data, env): object 'Steel' not found
Traceback:
```

Quiz 3: review

Consider the NutritionStudy Dataset. The mean **Cholesterol** consumed is 242.5 mg per day. A particular patient consumes 170.3 mg per day. Which of the following symbols would be used to represent the value 242.5 in the notation of the General Linear Model?

ei	2 respondents	3 %	
None of the above	16 respondents	21 %	
b0	46 respondents	**59 %**	✓
Yi	14 respondents	18 %	

"b0" is -- by convention -- shorthand for the "intercept" of the model. For the empty model, this intercept is the same thing as the mean value over the entire dataset.

Quiz 3: review

If we run the following line of R code

```
favstats(FatMice18$WgtGain4)
```

we get the following output

min	Q1	median	Q3	max	mean	sd	n	missing
3	6	8.5	10	17	8.4	3.3	18	0

What is the sample variance in the WgtGain4 values of the FatMice18 dataset?

10.9	50 respondents	64 %	✓
8.5	2 respondents	3 %	
8.4	4 respondents	5 %	
3.3	22 respondents	28 %	

Recall that the sample standard deviation ("sd") is the square root of the variance. To calculate the sample variance, take the square of the sample standard deviation.

Due This Week

7	May 10	**Models with categorical explanatory variable** *Before:* Chapter 7 *During:* Lab 4A	**Models with quantitative explanatory variable** *Before:* Chapter 8 *During:* Lab 4B	Practice Quiz 4	Lab 3 Due

Note: There are two CourseKata modules due this week. Chapter 7 is due today. Chapter 8 is due Wednesday.

Due This Week

7	May 10	**Models with categorical explanatory variable** *Before:* Chapter 7 *During:*Lab 4A	**Models with quantitative explanatory variable** *Before:* Chapter 8 *During:* Lab 4B	Practice Quiz 4	Lab 3 Due

Released Thursday at 5PM & due by 4:59PM on Friday

Due This Week

7	May 10	**Models with categorical explanatory variable** *Before:* Chapter 7 *During:*Lab 4A	**Models with quantitative explanatory variable** *Before:* Chapter 8 *During:* Lab 4B	Practice Quiz 4	Lab 3 Due

Lab 3 must be submitted via DataHub by 11:59pm PT on Friday, May 14. Feel free to submit early!

In Section This Week

- **Last week:** For **Project Milestone 3: Preregistration,** your group settled on which two variables you planned to focus on in your final project. Only the corresponding author submitted this worksheet!

- **This week:** In **Project Milestone 4A: Data preprocessing (due May 21)** you will be getting practice preparing your data to be fit with a linear model by applying any needed preprocessing (e.g., filtering out obvious outliers, recoding variables)

- **Next week: Project Milestone 4B: Model fitting (due May 21)** next week, you will use what you learned in Chapters 7&8 (and practiced in Lab 4A & 4B) to actually fit a linear model to your data & interpret your results.

Today

Lab 4a: Analysis of Variance (ANOVA)

General announcements

Break out into lab groups

Want real-time help?
(1) Post to #lab-assignments, mention both your TA & Room
(2) "Asking for help" in Zoom
(#3 is a bit less reliable!)

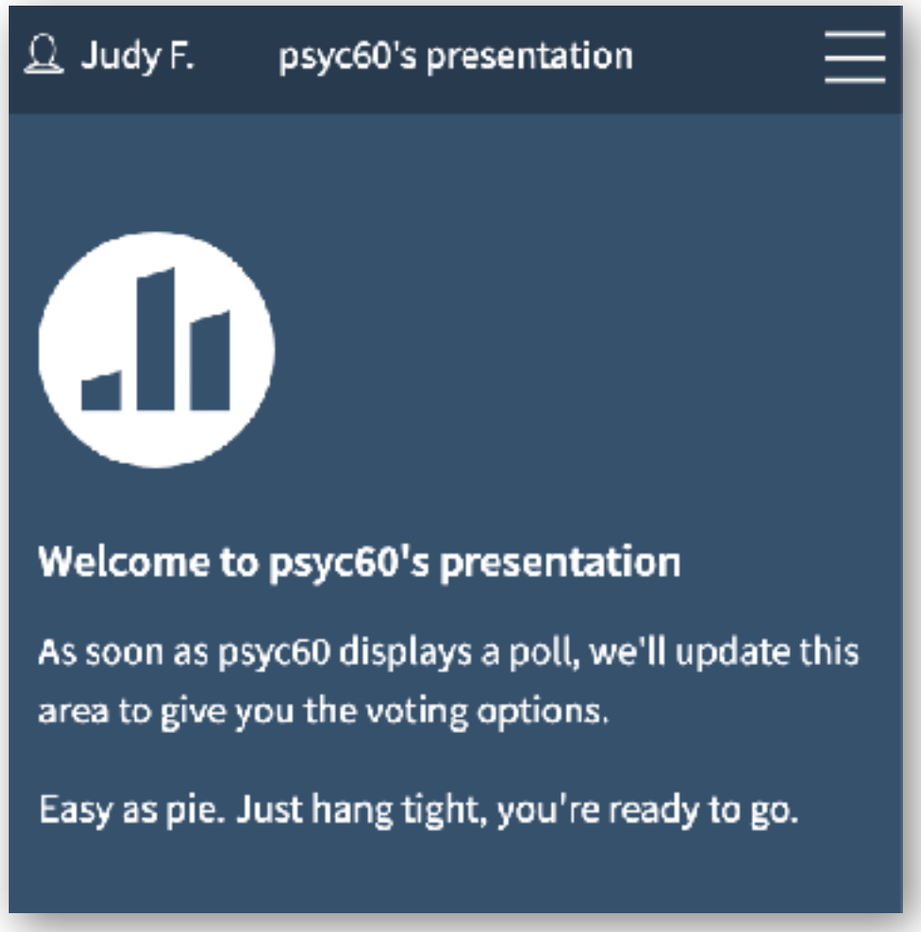
Return to main room and debrief

Everyone come back at 2:10pm PT

Debrief

1. Take your phone or laptop out.
2. Make sure you are connected to the internet.
3. Open any web browser & type in this URL:
   PollEv.com/psyc60
4. Make sure to log in to your account using your UCSD email & name as it appears in Canvas.

You should see something like this



PSYC 60: Lab 4A | General Impressions

When survey is active, respond at pollev.com/psyc60

0 done

0 underway