Homework 2 - MATH 141

Due Date: Monday 09/13/2021, 11:59 PM

Instructions:

- Please provide complete answers/solutions for each question/problem.

- If it involves mathematical computations, please provide your reasoning and/or detailed solutions.

- There are two ways you can write your answers, a: by handwriting (either physically or digitally), or b: by typing on a template document with file type options, Word or RMarkdown, which can be downloaded from the course website.

- If you had handwritten your answers/solutions on a physical paper, make sure to label it properly and please scan your document using a scanner app for convenience. Suggestions: (1) "Tiny Scanner" for Android or (2) "Scanner App" for iOS.

- If you have questions or concerns, please feel free to ask the instructor.

- **Please save your work as pdf file(s), don't put your name in any part of the document, and submit it to the Gradescope page for this course. Your document upload will correspond to your name automatically in Gradescope.**

I. Exploring Categorical Data

I.A. Consider the contingency table for homeownership and loan status shown in Table 1.

1. Create a *row proportions* table using Table 1.

2. Create a *column proportions* table using Table 1.

3. What is the percentage of late (31-120 days) payments among renters?

4. What is the percentage of fully paid loans given that the homes are owned?

5. What is the percentage of mortgages given that they are in grace period?

Table 1: A contingency table for homeownership and loan status

homeownership	Loan status						Total
	charged off	current	fully paid	in grace period	late (16-30 days)	late (31-120 days)	
rent	2	3630	152	27	14	33	3858
mortgage	4	4496	225	28	16	20	4789
own	1	1249	70	12	8	13	1353
Total	7	9375	447	67	38	66	10000

I.B. Consider the barplots for homeownership and loan status shown in Fig. 1.

1. Identify each barplot according to type (stacked, dodged, standardized).

2. Which type of visualization is helpful in understanding the fraction of homeownership in each level of loan status?

3. Which type of visualization clearly displays that "current" loan status is the most common?

4. Which type of visualization shows that mortgage homeownership type is the most common among "current" loan status?

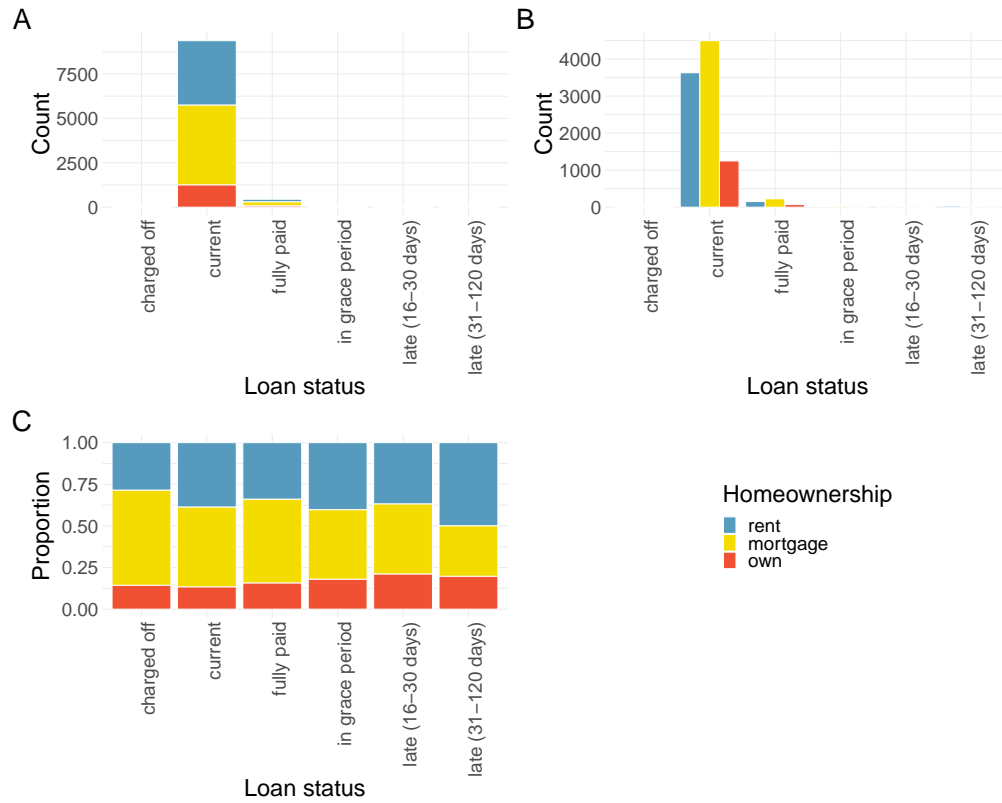5. Which type of visualization shows that approximately half of late payments are renters?

Figure 1: Three bar plots displaying homeownership and loan status variables.

II. Exploring Numerical Data

II.A. Consider the scatterplot shown in Fig. 2.

1. What is the relationship between unemployment rate and poverty rate - is it linear or non-linear? Explain why.

2. Would you consider the trend to be positively or negatively associated? Explain why.

3. Are there any potential outliers in the scatterplot and which variable do the outliers belong to? Explain how these outliers might affect the statistical model fitted into the data (shown in dashed line).
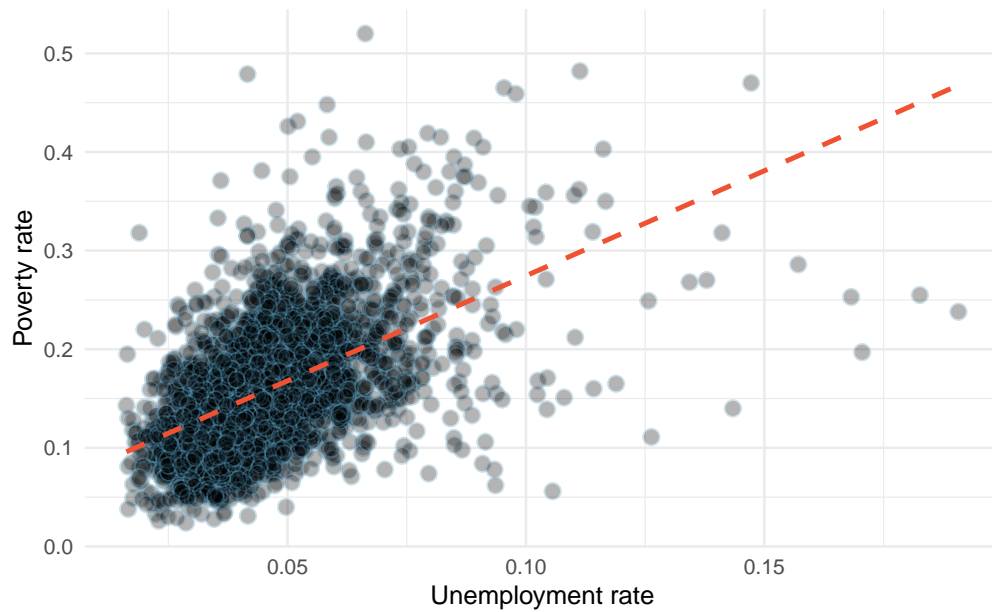

Figure 2: A scatterplot of the unemployment rate against the poverty rate for the county dataset. A statistical model has also been fit to the data and is shown as a dashed line.

II.B. Consider the number of observations per education level shown in Table 2, the summary statistics shown in Table 3, and the histograms and boxplots shown in Fig. 3.

1. Compute the interquartile Range (IQR) for each education level. What does this number represent and where does it show in the boxplots? Explain why we can't compute the IQR of the below_hs variable.

2. Using the mean and median, describe the skewness of the distributions, and which distribution is the most skewed. How many modes does each distribution have?

3. What information does the box plots provide that the histograms does not - and vice versa? Give a general answer and a specific answer related to the data showed.

Table 2: Number of observations per education level

Education level				
below_hs	hs_diploma	some_college	bachelors	Total
1	1187	1335	37	2559

Table 3: A summary statistics table by education level

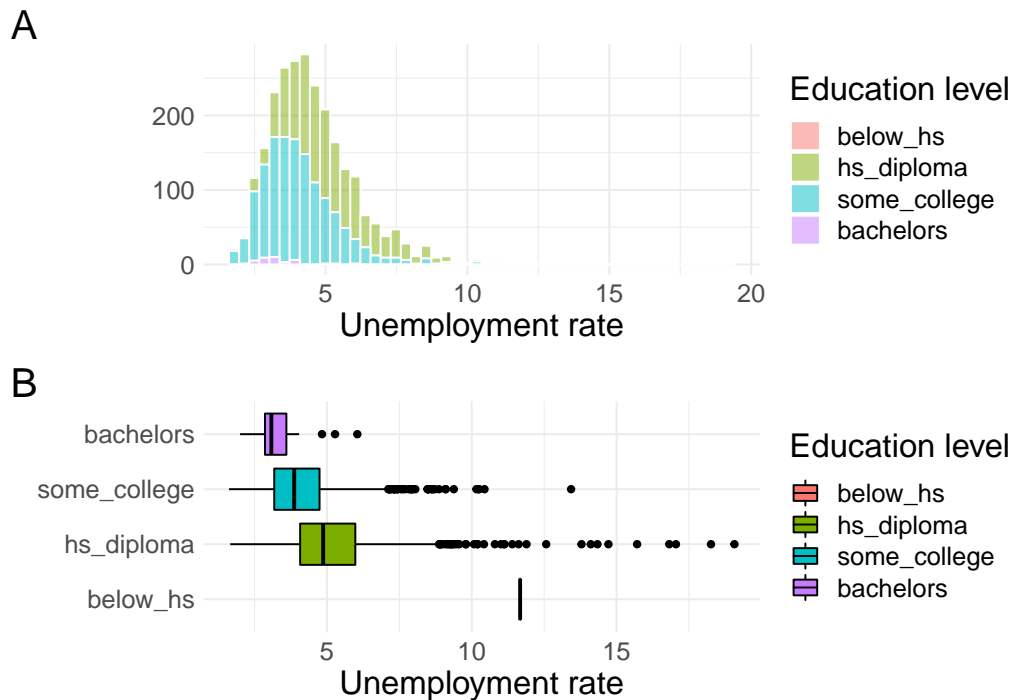| Education level	Summary statistics						
median_edu	min	q1	mode	median	mean	q3	max
below_hs	11.67	11.67	11.67	11.67	11.67	11.67	11.67
hs_diploma	1.66	4.08	4.37	4.87	5.22	5.98	19.07
some_college	1.62	3.18	3.66	3.87	4.08	4.75	13.43
bachelors	2.00	2.86	3.09	3.08	3.28	3.60	6.05

Figure 3: Histograms (Plot A) and side by-side box plots (Plot B) for unemployment rates split by education level

III. Textbook Exercises

Note: To view the selected exercises below, please refer to the textbook, *OpenIntro: Introduction to Modern Statistics (2021) by Mine Çetinkaya-Rundel and Johanna Hardin, First Edition.*

Section 4.8: Choose any 2 exercise problems below to answer.

2. Views on immigration.

4. Raise taxes.

6. Shipping holiday gifts data display.

Section 5.10: Choose any 6 exercise problems below to answer.

8. Medians and IQRs.

10. Histograms and box plots.

12. Median vs. mean.

14. Facebook friends.

16. Distributions and appropriate statistics.

18. Exam scores.

20. Oscar winners.

22. Income at the coffee shop.

24. Commute times.

26. NYC marathon winners.