

Stat 185 -Exam 1

Name: _____

Please write all your solutions (answers with organized supporting work) on separate paper, put your name on this sheet, and staple it all together when you turn it in.
I don't plan to grade anything written on the exam sheet itself.

1. A survey was conducted to study the smoking habits of NC residents. Below is a data matrix displaying a portion of the data collected in this survey. Note that "N/A" refers to a missing component of the data.

	sex	age	marital	grossIncome	smoke	amtWeekends	amtWeekdays
1	Female	42	Single	\$2600	Yes	12 cig/day	12 cig/day
2	Male	44	Single	\$9500	No	N/A	N/A
3	Male	53	Married	\$7200	Yes	6 cig/day	6 cig/day
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
200	Male	40	Single	\$5,200	Yes	8 cig/day	8 cig/day

- (a) What does each row of the data matrix represent?
 - (b) How many participants were included in the survey?
 - (c) Indicate whether each variable in the study is numerical or categorical. If numerical, identify as continuous or discrete. If categorical, indicate if the variable is ordinal.
 - (d) What is the population under study?
 - (e) What is the sample?
2. A 2010 survey asked 827 randomly sampled registered voters in California "Do you support? Or do you oppose? Drilling for oil and natural gas off the Coast of California? Or do you not know enough to say?" Below is the distribution of responses, separated based on whether or not the respondent graduated from college.

(a) What percent of college graduates and what percent of the non-college graduates in this sample oppose drilling for oil and natural gas off the Coast of California?	<i>College Grad</i>	
	Yes	No
Support	154	132
Oppose	180	126
Do not know	104	131
Total	438	389

- (b) Does support vs opposition for drilling appear to be independent of education?

3. Given the data set $\{3, -2, 5, 8, 3, 8, 4, 0, 1, -1\}$,
 - (a) Compute the corresponding five point summary
 - (b) Find the inter-quartile range
 - (c) Draw the corresponding box-plot

4. Let X be the data set $\{3, 4, 6, 7, 10\}$.
- Write down the formula showing that the mean of X is 6.
 - Compute the standard deviation of X , again showing your work.
5. On average, the 127 college football teams in Division I gained 5322 yards throughout the 2016 season with a standard deviation of 887 yards.
- Use a normal model to estimate the number of teams that rushed more than 6000 yards.
 - A histogram for this data is shown in figure 1. Based on the histogram, do you see any issues with using a normal model for this estimate?
6. Continuing with the previous example on college football teams, suppose that the average Win/Loss percentage for teams was 52%, with a standard deviation of 20%. Suppose also that the correlation between total yards gained and Win/Loss percentage is 0.69.
- What is the regression line estimating Win/Loss percentage in terms of total yards gained?
 - What Win/Loss percentage does the regression line predict if the total yards gained is 6000?

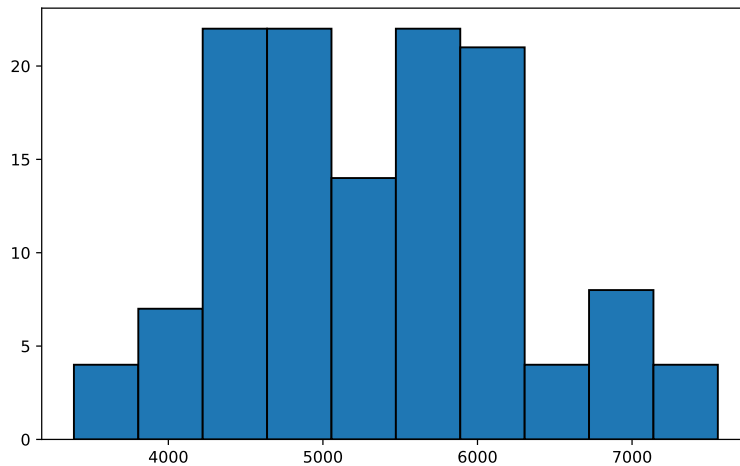


Figure 1: A histogram of total yards gained by each team for the 2016 season