NAME:_____

MATH 150 MIDTERM #1

February 11, 2016

INSTRUCTIONS: This is a closed book, closed notes exam. You are not to provide or receive help from any outside source during the exam.

- You may NOT use a calculator (except for the one provided).
- You may NOT use your smartphone for any purpose. Please silence it or turn it off.
- Show all of your work.

Question	Points	Score
1	10	
2	7	
3	6	
4	8	
5	3	
6	8	
7	6	
8	6	
9	6	
10	8	
11	3	
Total:	71	

1. A study published in the *Journal of Personality and Social Psychology* asked a group of 200 randomly sampled men and women to evaluate how they felt about various subjects, such as camping, health care, architecture, taxidermy, crossword puzzles, and Japan in order to measure their dispositional attitude towards mostly independent stimuli. Then, they presented the participants with information about a new product: a microwave oven. This microwave oven does not exist, but the participants didn't know this, and were given three positive and three negative fake reviews. People who reacted positively to the subjects on the dispositional attitude measurement also tended to react positively to the microwave oven, and those who reacted negatively also tended to react negatively to it. Researchers concluded that "some people tend to like things, whereas others tend to dislike things, and a more thorough understanding of this tendency will lead to a more thorough understanding of the psychology of attitudes."

   (a) [2 points] What are the observational units (cases)?

   200 randomly sampled men and women

   (b) [2 points] What is (are) the response variable(s) in this study?

   Reaction to the microwave oven.

   (c) [2 points] Is this an observational study or an experiment? Explain your reasoning.

   Observational study because we observe how the people react to stimuli. There isn't a control group.

   (d) [2 points] Can we establish a causal link between the explanatory and response variables?

   No because it is an observational study.

   (e) [2 points] Can the results of the study be generalized to the population at large?

   Yes, because they are randomly sampled.

2. A survey was conducted to study the smoking habits of UK residents. Below is a data matrix displaying a portion of the data collected in this survey. Note that "£" stands for British Pounds Sterling, "cig" stands for cigarettes, and "N/A" refers to a missing component of the data.

	sex	age	marital	grossIncome	smoke	amtWeekends	amtWeekdays
1	Female	42	Single	Under £2,600	Yes	12 cig/day	12 cig/day
2	Male	44	Single	£10,400 to £15,600	No	N/A	N/A
3	Male	53	Married	Above £36,400	Yes	6 cig/day	6 cig/day
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1691	Male	40	Single	£2,600 to £5,200	Yes	8 cig/day	8 cig/day

   (a) [2 points] What does each row of the data matrix represent?

   A UK resident sampled.

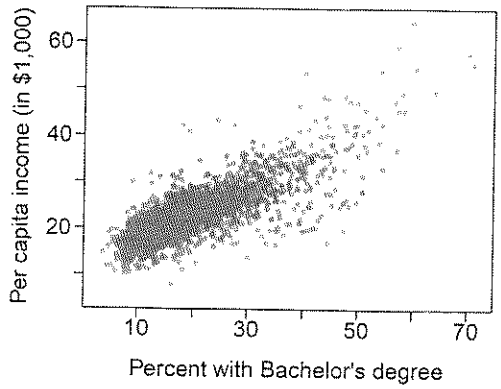
   (b) [2 points] How many participants were included in the survey?

   1691.

   (c) [3 points] 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.

   sex is categorical but not ordinal.
   age is numerical and discrete
   marital is categorical but not ordinal.
   gross Income is categorical and ordinal
   smoke is categorical and not ordinal.
   amt weekends is numerical and discrete.
   amt weekdays is numerical and discrete.

3. The scatterplot below shows the relationship between per capita income (in thousands of dollars) and percent of population with a bachelor's degree in 3,143 counties in the US in 2010.



(a) [2 points] What are the explanatory and response variables?

Explanatory: Percent with Bachelor's degree. Response: Per Capita income

(b) [2 points] Describe the relationship between the two variables. Make sure to discuss unusual observations, if any.

Positive association.

(c) [2 points] Can we conclude that having a bachelor's degree increases one's income?

No because it's an observational study. There might be confounding variables.

4. In order to assess the effectiveness of taking large doses of vitamin C in reducing the duration of the common cold, researchers recruited 400 healthy volunteers from staff and students at a university. A quarter of the patients were assigned a placebo, and the rest were evenly divided between 1g Vitamin C, 3g Vitamin C, or 3g Vitamin C plus additives to be taken at onset of a cold for the following two days. All tablets had identical appearance and packaging. The nurses who handed the prescribed pills to the patients knew which patient received which treatment, but the researchers assessing the patients when they were sick did not. No significant differences were observed in any measure of cold duration or severity between the four medication groups, and the placebo group had the shortest duration of symptoms.

(a) [2 points] Was this an experiment or an observational study? Why?

Experiment because they are split into groups by random assignment.

(b) [2 points] What are the explanatory and response variables in this study?

Explanatory: Vitamin C additives. Respore variable: Length of cold, severity of cold.

(c) [2 points] Were the patients blinded to their treatment?

Yes, since they don't know what pill they took.

(d) [2 points] Was this study double-blind?

Yes, because the assesment came from researchers not knowing what pills were given to a particular patient.

5. [3 points] Workers at a particular mining site receive an average of 35 days paid vacation, which is lower than the national average. The manager of this plant is under pressure from a local union to increase the amount of paid time off. However, he does not want to give more days off to the workers because that would be costly. Instead he decides he should fire 10 employees in such a way as to raise the average number of days off that are reported by his employees. In order to achieve this goal, should he fire employees who have the most number of days off, least number of days off, or those who have about the average number of days off?

He should fire the ten people that have the least number of days off.

6. A 2010 SurveyUSA poll asked 500 Los Angeles residents, "What is the best hamburger place in Southern California? Five Guys Burgers? In-N-Out Burger? Fat Burger? Tommy's Hamburgers? Umami Burger? Or somewhere else?" The distribution of responses by gender is shown below.

		Gender		
		Male	Female	Total
	Five Guys Burgers	5	6	11
	In-N-Out Burger	162	181	343
Best	Fat Burger	10	12	22
hamburger	Tommy's Hamburgers	27	27	54
place	Umami Burger	5	1	6
	Other	26	20	46
	Not Sure	13	5	18
	Total	248	252	500

(a) [2 points] Are being female and liking Five Guys Burgers mutually exclusive?

No. 6 people are female and like Five Guys Burgers.

(b) [2 points] What is the probability that a randomly chosen male likes In-N-Out the best?

$162/248$

(c) [2 points] What is the probability that a randomly chosen In-N-Out fan is female?

$181/343$

(d) [2 points] What is the probability that a randomly chosen person is female or likes Umami best? (Note: In statistics, when we say "or" we include the possibility of being both)

$\frac{6+252-1}{500} = \frac{257}{500}$

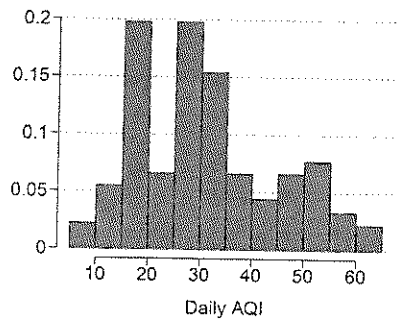
7. If you roll two dice:

(a) [2 points] What is the probability that the sum is 1? 0

(b) [2 points] What is the probability that the sum is 11? $2/36$

(c) [2 points] What is the probability that the sum is 7? $6/36$

8. Daily air quality is measured by the air quality index (AQI) reported by the Environmental Protection Agency. This index reports the pollution level and what associated health effects might be a concern. The index is calculated for five major air pollutants regulated by the Clean Air Act and takes values from 0 to 300, where a higher value indicates lower air quality. AQI was reported for a sample of 91 days in 2011 in Durham, NC. The relative frequency histogram below shows the distribution of the AQI values on these days.



(a) [2 points] Estimate the median AQI value of this sample.

≈ 27.5

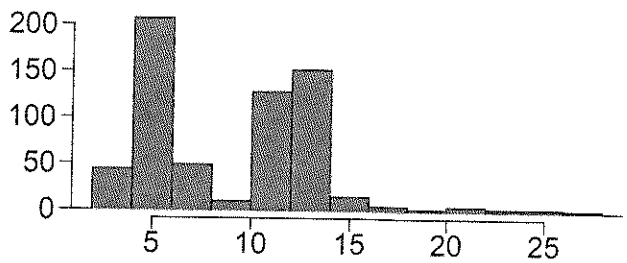
(b) [2 points] Would you expect the mean AQI value of this sample to be higher or lower than the median? Explain your reasoning.

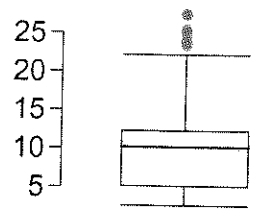
Mean to be higher.

(c) [2 points] Estimate Q1, Q3, and IQR for the distribution.

$Q1 \approx 18$, $Q3 \approx 39$ $IQR \approx 21$

9. Compare the two plots below.





(a) [2 points] What characteristics of the distribution are apparent in the histogram and not in the box plot? It has 2 peaks.

(b) [2 points] What characteristics are apparent in the box plot but not in the histogram?

It has several outliers to the right, yet Q3 is closer to the median than Q1

(c) [2 points] Is the data right-skewed or left-skewed?

Right-skewed (outliers to the right)

10. A 2012 Pew Research survey asked 2,373 randomly sampled registered voters their political affiliation (Republican, Democrat, or Independent) and whether or not they identify as swing voters. 35% of respondents identified as Independent, 23% identified as swing voters, and 11% identified as both.

    (a) [2 points] What percent of voters are Independent but not swing voters?

    24%

    (b) [2 points] What percent of voters are Independent or swing voters?

    47%

Indep. Swing
24 11 12

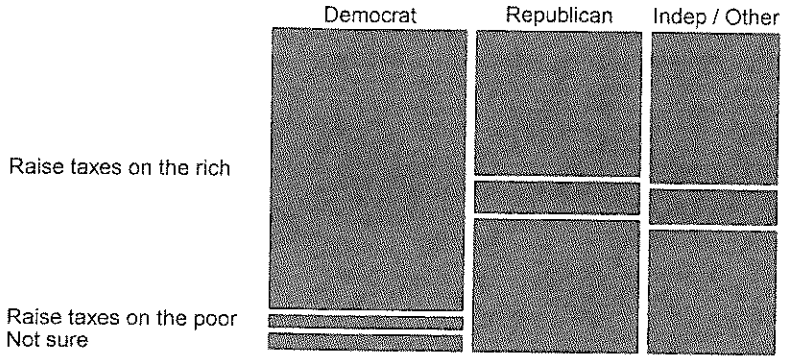
    (c) [2 points] What percent of voters are neither Independent nor swing voters?

    53%

    (d) [2 points] Is the event that someone is a swing voter independent of the event that someone is a political Independent?

    No. Otherwise $.35 \times .23 = .11$

11. [3 points] A random sample of registered voters nationally were asked whether they think it's better to raise taxes on the rich or raise taxes on the poor. The survey also collected information on the political party affiliation of the respondents. Based on the mosaic plot shown below, do views on raising taxes and political affiliation appear to be independent? Explain your reasoning.



They don't appear to be independent. Democrats are much more likely to want to raise taxes on the rich than Republicans or Independents.