

STOR 151 SECTION 1 MIDTERM 1
SEPTEMBER 19, 2019

This is an open book exam. Course text, personal notes and calculator are permitted. You have 75 minutes to complete the test. Personal computers and cellphones are not to be used during the exam. If you have any queries about the meaning of the questions, or if you think there is an error, ask the proctor for assistance. Answers are to be written in a blue book.

Before you begin:

- (a) Write your name and PID on the front of your blue book.
- (b) Sign the “pledge”. If this is not preprinted, copy out the following statement and sign it: *On my honor, I have neither given nor received unauthorized aid in this exam.*

SHOW ALL WORKING — even correct answers will not get full credit if it’s not clear how they were obtained. Incorrect answers will gain substantial credit if the method of working is substantially correct.

Answer all questions. The score for each part is indicated at the end of the question (total 100).

1. The following data represent the 50 largest 5-day rainfalls (in inches, in increasing order) in the six states bordering the Gulf of Mexico over 1949–2018:

9.21, 9.23, 9.23, 9.24, 9.27, 9.28, 9.31, 9.35, 9.36, 9.36, 9.38, 9.42, 9.48, 9.5, 9.5, 9.54, 9.56, 9.63, 9.68, 9.9, 9.91, 9.91, 9.99, 10.06, 10.09, 10.18, 10.37, 10.38, 10.44, 10.67, 10.87, 10.94, 10.99, 11.02, 11.09, 11.51, 11.54, 11.56, 11.62, 11.91, 12.1, 12.71, 12.73, 12.87, 13.63, 14.19, 15.07, 15.26, 16.48, 27.99.

The last value (27.99) represents Hurricane Harvey, the huge storm that hit Houston in August 2017.

- (a) Calculate the median. [5]
- (b) The first and third quartiles (Q1 and Q3) are 9.48 and 11.56. What is the IQR? [5]
- (c) Draw a boxplot of these data, indicating the locations of the box and whiskers. [10]
- (d) Are there any outliers in this dataset? Explain how you determine whether there are or are not. [6]
- (e) The mean of these values is 11.13, and the standard deviation is 3.00. Is there a discrepancy between the mean and the median, or between the standard deviation and the IQR? Which would you regard as the more meaningful measures and why? [6]

CONTINUED ON OTHER SIDE — PLEASE TURN OVER

2. A survey asked people “When is premarital sex wrong?” Respondents were allowed to give one of the following answers: always, almost always, sometimes or never. They were also asked their religion. The responses were as follows:

	Always	Almost Always	Sometimes	Never	Total
Protestant	355	117	227	384	1083
Catholic	62	37	120	226	445
Jewish	0	3	14	34	51
None	20	13	45	147	225
Other	15	13	23	40	91
Total	452	183	429	831	1895

- (a) What proportion of the sample are Protestants who believe that premarital sex is always wrong? [7]
- (b) What proportion of Jewish respondent believe that premarital sex is never wrong? [7]
- (c) Among those who believe that premarital sex is sometimes wrong, what proportion are Catholics? [7]
- (d) Would you say that the events “The responder is Protestant” and “The respondent believes that premarital sex is always wrong” are independent? Explain why or why not. [11]
3. In a restaurant near the UNC campus, 45% of the customers identify as male and the rest as female. Menu items are classified as steak (S), chicken (C) or vegetarian (V). Among those identifying as male, 40% order steak, 40% order chicken and the rest order vegetarian. Among those identifying as female, 20% order steak, 45% order chicken and the rest order vegetarian. You visit the restaurant and sample a randomly selected customer.
- (a) Draw a tree diagram to represent this scenario. [10]
- (b) What is the probability that the randomly selected customer orders steak and identifies as male? [8]
- (c) What is the probability that the randomly selected customer orders vegetarian? [8]
- (d) Given that the randomly selected customer orders vegetarian, what is the probability that the customer identifies as female? [10]

STOR 151 SECTION 1 MIDTERM 1
SEPTEMBER 19, 2019
SKETCH SOLUTIONS

1. (a) The middle two values are 10.09 and 10.18, so the median is $\frac{10.09+10.18}{2} = 10.35$.
- (b) The IQR is $11.56-9.48=2.08$.
- (c) 1.5 times the IQR is 3.12, so the boundary points for outliers are $11.56+3.12=14.68$ and $9.48-3.12=6.36$. There are no points beyond the lower boundary so the bottom whisker is at the smallest observation, 9.21. For the upper boundary, the closest points are 14.19 (below), 15.07, 15.26, 16.48, 27.99 (above). So according to this rule, there are four outliers at the top end, and the top whisker is drawn at 14.19 (largest number inside the boundary). The boxplot is drawn below (it would also be fine if this was drawn vertically instead of horizontally).

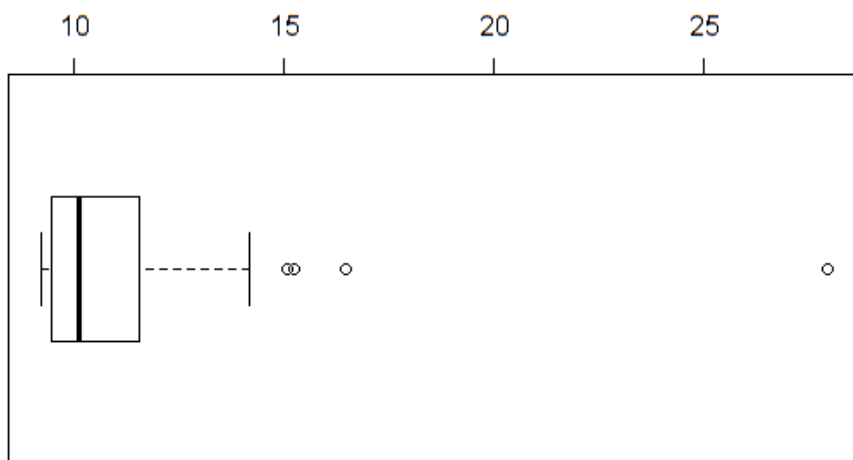


Figure 1: Box plot for Question 1(b).

- (d) The four largest values are outliers as explained in the answer to the previous question.
 - (e) The mean and standard are both substantially larger than the median and IQR respectively. This is because of the extreme outlier due to Hurricane Harvey. It's better to use the median and IQR because they are not influenced by the extreme outlier
2. (a) $\frac{355}{1895} = 0.187$
 - (b) $\frac{34}{51} = 0.667$
 - (c) $\frac{120}{429} = 0.280$
 - (d) If A is the event "Responder is Protestant" and B is the event "Responder believes that premarital sex is always wrong," then $P(A) = \frac{1083}{1895} = 0.572$, and $P(B) = \frac{452}{1895} = 0.239$. But then $P(A) \times P(B) = 0.572 \times 0.239 = 0.137$, substantially less than the observed probability which is 0.187. Therefore, the two events are not independent.

(a) See Figure 2.

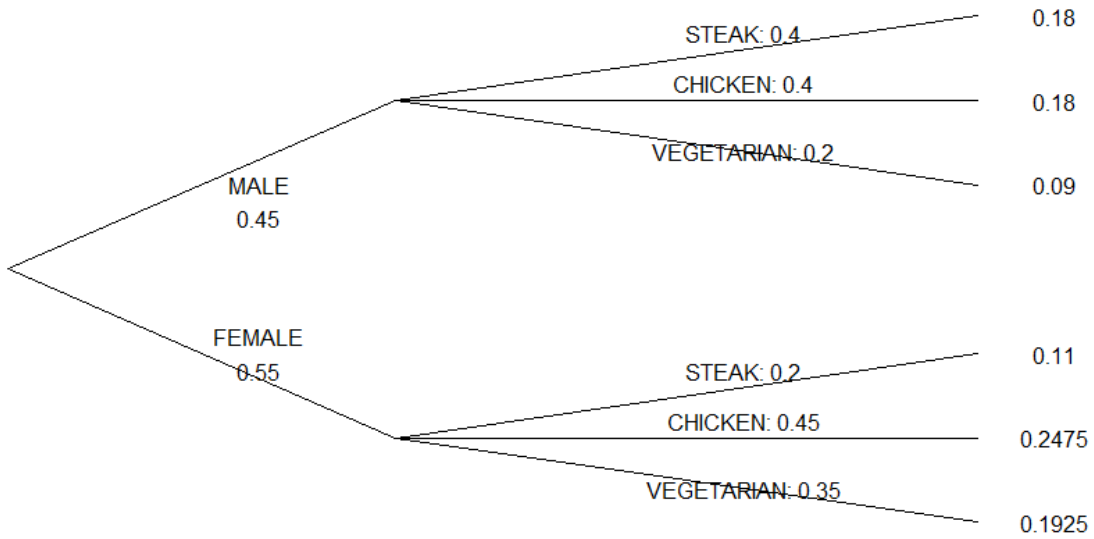


Figure 2: Tree for Question 3(a).

(b) 0.18

(c) $0.09 + 0.1925 = 0.2825$.

(d) $\frac{0.1925}{0.2825} = 0.6814$ (or 68%).