Exam 2

STA209-04: Applied Statistics

April 12, 2019

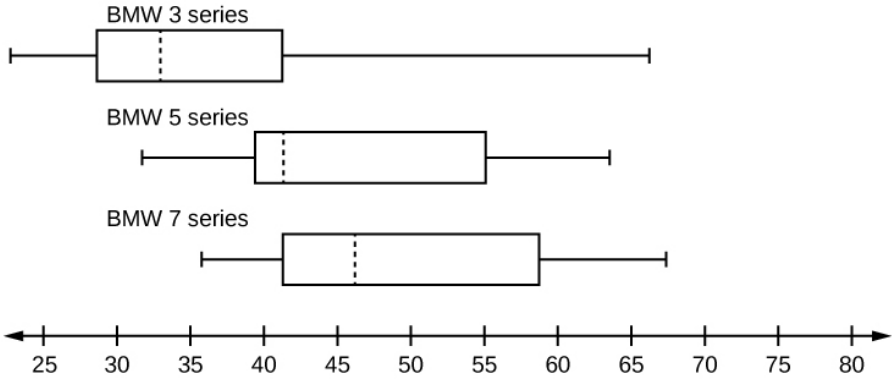
Please carefully read each question. You will have 80 minutes to complete this exam. Show all of your work. All short answers should be no more than three sentences in length. Each sentence beyond this limit will result in a one point penalty. Write your name on the upper right hand corner of your answer sheet.

1) [50 pts] In 2014, Parenti et al. published the results of a study comparing the efficacy of three methods of brewing. The three methods compared were the traditional Bar Machine method (BM), the Hyper Espresso method (HIP), and the I-Espresso System (IT). For each method, the foam index, or ratio between the foam and liquid volume was measured. Our interest is in determining whether there is a difference among these methods in the quality of espresso, as measured by the foam index. Table 1 provides basic summary statistics of the foam index for each method.

Table 1: Foam Index Summary Measures

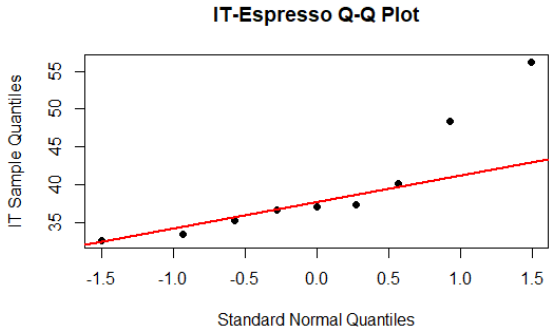
Method	**Min**	**Max**	25^{th} **(%)-ile**	**Median**	**Mean**	75^{th} **(%)-ile**	**Std. Dev.**
BM	21.02	39.65	24.81	35.96	32.40	37.74	7.30
HIP	46.68	73.19	54.26	62.53	61.30	70.84	10.10
IT	32.68	56.19	35.35	37.12	39.70	40.11	7.70

i) Using appropriate information from the table above, sketch boxplots of the foam index for each method. Your sketch should be based on a single horizontal axis ranging from 20 to 75, with each tick on this axis marking a 5-unit increment. Be sure to label each boxplot with its corresponding method. Specify which quantities from Table 1 are used in constructing each boxplot. Shown below is an example of the format your sketch should follow:



ii) For each group, decide whether it appears that the data are normally distributed. Use either your sketch from (i) or what you know about the mean and median to justify your response. (Hint: Recall that, unlike the median, the mean is not robust to outliers. In other words, the mean is affected by extreme data points.)

iii) One other popular method for visually assessing normality is through **quantile-quantile**, or **q-q plots**. Q-Q plots are generated by plotting the quantiles of your sample data distribution against the quantiles of a theoretical, standard normal distribution. If your sample data are normally distributed, the plotted points should form a roughly straight and diagonal line. Forming a perfectly straight and diagonal line would indicate a perfect match of the quantile of your sample data to the quantile of the standard normal data. The following q-q plot was generated by the IT espresso method data:



Based on your knowledge of the shape of the IT method data (i.e. symmetric, left-skewed, right-skewed), explain why the two right-most points are substantially above the red line. (Hint: **Quantiles** are points that cut up a distribution into sections of equal area, or probability. A familiar example of quantiles are the quartiles of a distribution, which define intervals each containing 25% of the data distribution.)

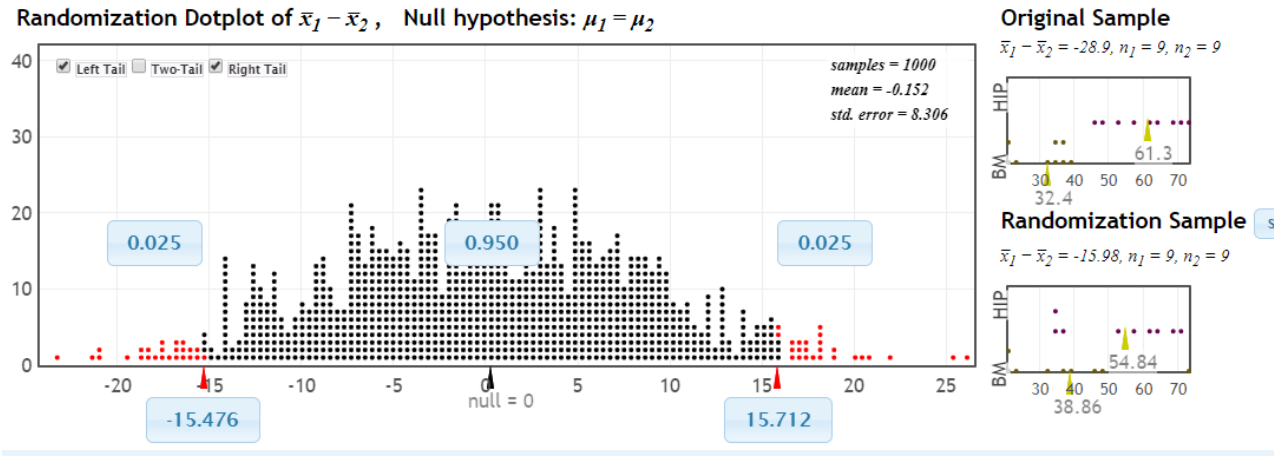
iv) Suppose that our data satisfy normality assumptions. Construct separate 95% confidence intervals of the average foam index for each brewing method. Note that there were 9 measurements obtained for each method, resulting in a total of 27 measurements. Additionally, the following are potentially useful cutoffs each corresponding to a confidence level of 95%: $t_{df=8} = 2.306$, and $t_{df=26} = 2.056$. State whether you believe there is a difference in espresso foam index across the three methods. Explain your reasoning.

v) Describe an approach you might use to test for differences across the three methods.

vi) Suppose that, after implementing the approach described in (v), you obtain a single statistically significant result. If the test(s) performed in this approach used $\alpha = 0.1$, would you suspect that you committed a type I error? Explain.

vii) Suppose that our data do not satisfy normality assumptions. Assuming our interest is in testing for a difference between the BM and HIP methods, describe how we would generate a randomization distribution that could be used to perform the appropriate statistical test.

viii) Use the randomization distribution of the mean differences to test whether the average foam index is different between the BM and HIP methods. State your conclusion in the context of the research question.



ix) Would it be reasonable to approximate the previous randomization distribution with a normal distribution? If so, explain why and state what the mean and standard deviation would be. If not, explain why not.

x) Suppose that you decided to approximate the randomization distribution using some normal distribution. Additionally, suppose you used this normal approximation to construct a 95% confidence interval for the mean foam index difference. How would this interval compare to a 95% confidence interval constructed using a t-distribution with eight degrees of freedom? Would it be wider, narrower, or about the same? Explain.

2) [20 pts] In 1975, Bickel et al. published a study on gender bias in graduate admissions using data from the University of California at Berkeley. The study was based on admissions data from the Fall of 1973, collected from the six largest departments in the university. Aside from recording whether an applicant was accepted or rejected for admission, data on the sex of each applicant was also obtained. Shown below is a two-way frequency table providing the counts admitted and rejected for males and females:

Table 2: Admissions by Sex

	Admitted	**Rejected**
Male	1198	1493
Female	557	1278

i) One of the motivations for the collection of these data was to investigate whether there was disproportionate admissions across sex. Conduct a statistical test to assess whether the proportion of applicants admitted differs by sex. State your conclusion in the context of the research question.

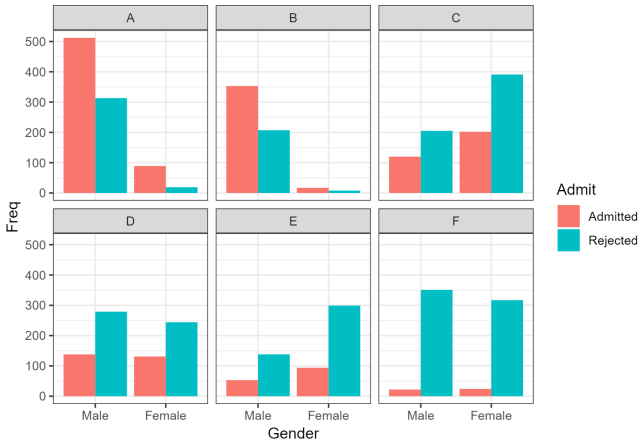
ii) Do the results of the test performed in (i) indicate that the differences in admissions rates are due to gender discrimination? Explain.

iii) These data can be stratified by department. The output below summarizes the results of six different difference in proportions tests done separately within each of the six departments:

	$p_{\text{adm}\mid\text{male}}$	$p_{\text{adm}\mid\text{female}}$	**p-value**
Dept A	0.621	0.824	0.000
Dept B	0.630	0.680	0.771
Dept C	0.369	0.341	0.426
Dept D	0.331	0.349	0.638
Dept E	0.277	0.239	0.369
Dept F	0.059	0.070	0.640

Interpret the results of these six tests holistically rather than individually in the context of evaluating possible gender bias in admissions at UC Berkeley.

iv) The figure below shows the frequencies of admitted/rejected students by department and gender. Using this figure, explain the seemingly different results of the test in (i) versus the results shown in (iii).



3) [30 pts] Data on all field goal attempts in 2008 were collected from the website *advancedfootballanalytics.com*, which has repositories of detailed datasets for the National Football League (NFL). A field goal attempt occurs when the offensive team attempts to kick the football through the upright goalposts. The offensive team may attempt a field goal at any time, but it is typically done when the team is close to the goalposts and cannot run any additional plays without yielding ball control to the other team.

Tables 3 and 4 display summary information of successful (GOOD = 1) and failed (GOOD = 0) field goal attempts. These tables also characterize the distributions of the kicking distance (i.e. how far away the kick was attempted) and the score differential at the time of the kick. Note that for the score differential, negative values are indicative of the team trailing in score whereas positive values indicate the team is leading.

Table 3: Kick Distance Summary Measures by Field Goal Attempt Result

GOOD	**Q1**	**Median**	**Q3**	**Mean**	**Std. Dev**	**N**
0	40.5	47	50	45.3	8.3	139
1	27.0	35	43	35.3	9.4	900

Table 4: Score Differential Summary Measures by Field Goal Attempt Result

GOOD	Q1	Median	Q3	Mean	Std. Dev	N
0	-3.5	0	6	0.7	8.8	139
1	-6.0	0	6	0.3	9.8	900

i) Assuming these data are representative of all modern NFL seasons, construct a 95% confidence interval for the overall proportion of successful field goal attempts.

ii) Inspection of Table 3 seems to indicate that failed field goal attempts are, on average, made from further away. Conduct a statistical test to determine if the observed difference in the distances of successful and failed field goal attempts could be due to random chance. State your conclusion in the context of the research question.

iii) Suppose we suspect that field goal attempts are more difficult in the fourth quarter when the game is tied or the kicking team is trailing by three or fewer points. Use the following tables to construct a two-way frequency table relating this situation and whether the field goal attempt was successful or not.

Table 5: Fourth Quarter Kicks

GOOD	N
0	29
1	226

Table 6: Fourth Quarter Kicks when Game is Tied or Trailing by 3 or Fewer Points

GOOD	N
0	10
1	66

iv) Assuming there is no association between the previously described situation (i.e. fourth quarter and tied or trailing by 3 or fewer points) and whether an attempt is successful, construct a two-way table of expected counts.

v) Conduct a statistical test to determine whether there is an association between the situation of interest and the success of a field goal attempt.

vi) Describe an alternate approach to evaluating whether the situation of interest is associated with the success of an attempt. How would you expect the results of this alternate approach to compare to the results of the test from (v)?