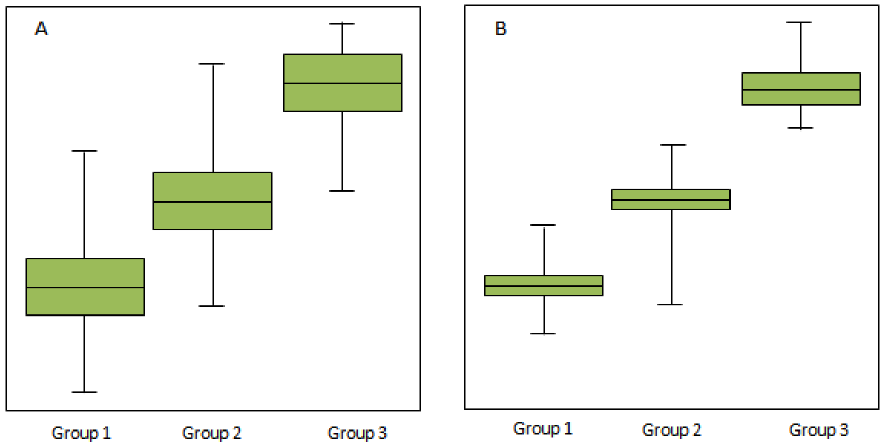
MET CS 555 - Data Analysis and Visualization
Quiz - 5

1. Assuming the same scale for the side-by-side box plots in A and B and that the means of Group 1 are the same, the means in Group 2 are the same, and that the means in Group 3 are the same across plots, which of the following statements are true. Select all that apply.



   A. A will result in a larger p value for the one way ANOVA
   B. B will result in a larger p value for the one way ANOVA
   C. A will result in a larger F-statistic for the one way ANOVA
   D. B will result in a larger F-statistic for the one way ANOVA
   E. Since means are the same, the F-statistic and p-value will be the same.

   Answer (A) and (D)

   Description. The F-statistic is the ratio between the between group variability to the within group variability. The between group variability is the same in each case (since the means are the same). The within group, variability, however, is smaller in B than in A. As the variability within each group goes down (as in B), the denominator of the F statistic gets smaller, and as a result, the F-statistic gets larger. Larger F-statistics are associated with smaller p-values. As such, B will have a larger F-statistic and a smaller p-value.

   Given this, the following selection are true:
   - A will result in a larger p value for the one way ANOVA
   - B will result in a larger F-statistic for the one way ANOVA

---

2. A random survey of 205 drivers measured each driver's "anger" score. Higher scores indicated higher propensities for road rage. The average anger scores were calculated for drivers aged 16-25, 26-35, 36-45, 46-55, and 56–65. A one-way analysis of variance was performed to assess whether or not there were differences between anger scores of drivers in the different age categories. The resulting partial ANOVA table is shown below. Calculate the F-statistic.

	SS (Sum of Squares)	df (degrees of freedom)	MS (Mean Square)	F-statistic	p-value
Between			50		
Within					
Total	5200				

A. 0.50
B. 1.93
C. 2.00
D. 2.02
E. 2.05

Answer (C)

Description. We need to use the relationships between value in the ANOVA table to answer this question. We can start by filling in pieces of information that we know. The df for the first row is equal to k-1 where k is the number of groups. Here, we have 5 groups (16-25, 26-35, 36-45, 46-55, and 56–65), so k-1 = 4. The df for the second row is equal to n-k = 205 - 5 = 200.

The SS between (SSB) divided by the associated df between equals the MS Between (MSB). That is, SSB/df between = MSB or SSB / 4 = 50, which means SSB = 200.

The SS within (SSW) + SSB = SS Total (SST). That is, SSW + SSB = SST or SSW + 200 = 5200 which means SSW = 5200.

The MS Within (MSW) is equal to the SSW divided by the associated df. Or, MSW= SSW/df within = 5000/200 = 25.

F = MSB/MSW = 50/25 = 2

---

3. A random survey of 205 drivers measured each driver's "anger" score. Higher scores indicated higher propensities for road rage. The average anger scores were calculated for drivers aged 16-25, 26-35, 36-45, 46-55, and 56–65. A one-way analysis of variance was performed to assess whether or not there were differences between anger scores of drivers in the different age categories. Using the F-distribution table, what is the appropriate decision rule that we should use for the global F-test if we are interested in maintaining a type I error rate that is less than 0.05?

   A. Reject if $F \leq 2.11$
   B. Reject if $F \leq 2.26$
   C. Reject if $F \leq 2.42$
   D. Reject if $F \leq 2.63$
   E. Reject if $F \geq 2.65$
   F. Reject if $F \geq 2.11$
   G. Reject if $F \geq 2.26$
   H. Reject if $F \geq 2.42$
   I. Reject if $F \geq 2.63$
   J. Reject if $F \geq 2.65$

   Answer (H)

   Description. Based on the above, the degrees of freedom are 4 (numerator) and 200 (denominator). Using the table (2nd page) we look for the df = 4 (numerator) column and the df = 200 (denominator) row. Since the alpha = 0.05, we look at that row to find the value 2.42. Our decision rule is that we will reject the null if our critical value is greater than or equal to 2.42.

---

4. A random survey of 205 drivers measured each driver's "anger" score. Higher scores indicated higher propensities for road rage. The average anger scores were calculated for drivers aged 16-25, 26-35, 36-45, 46-55, and 56–65. A one-way analysis of variance was performed to assess whether or not there were differences between anger scores of drivers in the different age categories. If we would reject the global null hypothesis and were interested in **all of the possible pairwise comparisons**, how many additional tests would we have to perform?

A. 4

B. 5

C. 8

D. 10

E. 20

F. 25

Answer (D)

Description. The number of possible pairwise comparisons is equal to k* (k-1) /2 where k = number of groups. Here, we have 5 groups, so k*(k-1)/2 = 5*4/2 = 20/2 = 10.

---

5. A random survey of 205 drivers measured each driver's "anger" score. Higher scores indicated higher propensities for road rage. The average anger scores were calculated for drivers aged 16-25, 26-35, 36-45, 46-55, and 56–65. If we were to create dummy variables for age group, how many dummy variables would be needed?

   A. 1

   B. 2

   C. 3

   D. 4

   E. 5

   F. 6

   Answer (D)

   Description. You always need one less than the number of groups. Here, we have 5 groups, so we need 4 dummy variables.

---

6. If we performed 20 different linear regressions at the 0.10 level, how many of these would we expect to show significance by chance?

   A. 0

   B. 2

   C. 5

   D. 10

   E. 20

   Answer (B)

   Description. If we performed 20 different linear regressions at the 0.10 level, then we'd expect that 20 * .10 = 2 would show significance by chance. This is not to be confused with the formula for the family wise error rate which is the probability of falsely rejecting at least one null hypothesis.

---

7. The resulting p-value from a one-way ANOVA 0.11 based on a sample of over 1000 patients. The level of significance was pre-defined to be 0.10. Which of the following is true? Select all that apply.

   A. We may have made a Type I Error

   B. We may have made a Type II Error

   C. We could have rejected the null hypothesis if we had chosen a smaller level of significance.

   D. We could have rejected the null hypothesis if we had chosen a larger level of significance.

E. We aren't sure whether or not there really is no difference between groups with respect to the mean outcome value or whether or not our sample size was too small.

F. We aren't sure whether or not there really is no difference between groups with respect to the mean outcome value or whether or not our sample size was too large.

G. We rejected the null hypothesis, but perhaps we should not have.

Answer (B), (D) and (E)

Description. Given the above, we would not have rejected the null hypothesis since our p value was greater than out level of significance. Given this, we can evaluate each statement:

(1) We did not make a Type I Error since a Type I Error is when we reject the null hypothesis when the null hypothesis is true. Since we failed to reject the null, we could not have made a Type I error.

(2) We may have made a Type II Error. A Type II Error happens when you fail to reject the null hypothesis and the null hypothesis is false. When we fail to reject the null hypothesis, is always possible that we have made a Type II Error.

(3) We reject the null hypothesis when our p-value is less than our significance level. Our p-value was 0.11 and our significance level was 0.10. If our significance level was lower (like .05) then this would not have helped us, as we still would have failed to reject our null hypothesis. That is, we could not have rejected the null hypothesis if we had chosen a smaller level of significance.

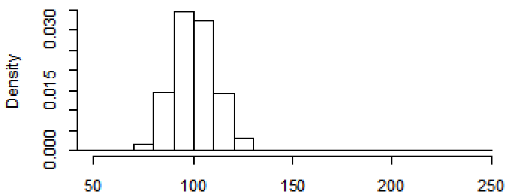
(4) On the other hand, if our significance level was higher (like .15) then we would have rejected our null hypothesis. That is, we could have rejected the null hypothesis if we had chosen a larger level of significance.

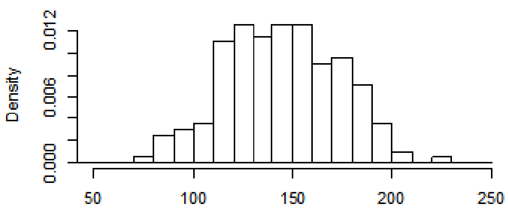
(5) When ever we fail to reject the null hypothesis we are never sure if the null hypothesis (no differences between groups) was true or because we lacked sufficient power (had too small of a sample size). That is, we aren't sure whether or not there really is no difference between groups with respect to the mean outcome value or whether or not our sample size was too small.

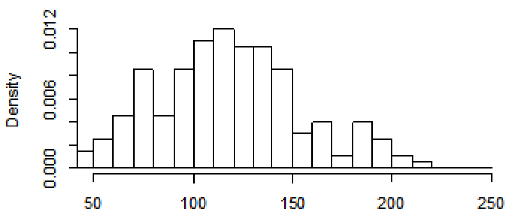
(6) Since we failed to reject the null hypothesis, the last one should not be selected. Also, our sample size is never too large!

---

8. The data for systolic blood pressure in three groups (A, B, and C, respectively) is shown below in the histograms (one per population/group). If one was to perform a one way ANOVA on this data, which of the following might be a concern?







A. The groups do not appear to be that different with respect to SBP so we may not reject the null hypothesis.

B. The distributions of SBP across groups do not appear to be relatively normally distributed.

C. The variability is much smaller in the first group.

D. The sample sizes may not be large enough in each group to detect differences.

Answer (C)

Description. In order for the inference for ANOVA to be valid, a number of conditions must be met:
(1) independent, random samples from each group/population
(2) outcome of interest in each group/population is normally distributed with unknown mean
(3) outcome of interest in each group/population has a similar unknown standard deviation

You aren't able to tell if the first of these is met by looking at the histograms (it is inherent in the data and depends on how it was collected). The second appears to be met as each histogram is roughly normally distributed. The last criteria is the answer in this case since it appears as though the variability of the first group is much higher than the variability in the other two groups. Since the histograms don't give any indication of sample size, the last choice should not be considered.

---

9. The stress levels of students during an exam were measured in three different testing environments: (a) complete silence, (b) with music playing, (c) while their instructor was talking on his cell phone. The mean stress levels were 73.5, 82.5, and 91.3, respectively. The overall F-statistic was calculated as 14.08 with 2 and 42 degrees of freedom with an associated p-value of ¡ 0.0001. What is the conclusion that we should make from this result?

   A. We do not have sufficient evidence that the mean stress levels differ in at least two of the groups.

B. We have sufficient evidence that the mean stress levels differ in at least two of the groups.

C. We have sufficient evidence that the mean stress levels differ across all groups.

D. All of the means are the same.

Answer (B)

Description. The global test for ANOVA is that all groups have the same underlying population mean. The alternative is that the underlying means are not all the same across groups. That is, the alternative is that there are at least two groups which have different underlying population means. Since the p-value is very very small, we reject the null hypothesis at any reasonable level of significance. As such, our conclusion is in favor of the alternative hypothesis. We have sufficient evidence that the mean stress levels differ in at least two of the groups

---

10. If we are interested in the following comparisons:

Silence versus Music

Cell phone talking versus Silence

Then, which group should we choose as the reference group?

A. Music

B. Silence

C. Cell phone talking

Answer (B)

Description. Since both of the comparisons of interest involve the Silence group, we should choose this as the reference group. This would mean creating a dummy variable for the other two groups if we were performing this analysis using the regression framework.

---

11. Dummy variables were created for Silence and for Cell Phone Talking. What should be the value of Silence and Cell Phone Chatting for ID = 5 below?

ID	**Group**
1	Silence
2	Cell phone talking
3	Music
4	Silence
5	Music

A. 0 and 0, respectively

B. 0 and 1, respectively

C. 1 and 0, respectively

D. 1 and 1, respectively

Answer (A)

Description.

ID number 5 is the music group. As such, their value for dummy variables for Silence and Cell Phone Talking would be 0 and 0.

Here's what the others would look like:

ID	Group	Dummy Variable:	Silence Dummy Variable; Cell Phone Talking
1	Silence	1	0
2	Cell Phone Talking	0	1
3	Music	0	0
4	Silence	1	0
5	Music	0	0

---

12. If we reject the null hypothesis from a one-way ANOVA model, then we. . . .

    A. can conclude that the underlying means are all the same

    B. need to perform pair-wise comparisons in order to determine where the differences lie

    C. need to re-run ANOVA as linear model to confirm results

    Answer (B)

    Description. The global test for ANOVA is that all groups have the same underlying population mean. The alternative is that the underlying means are not all the same across groups. That is, the alternative is that there are at least two groups which have different underlying population means. If we reject the null hypothesis, then the next logical question is which two of the groups have different underlying population means. To address this, we need to perform pair-wise comparisons in order to determine where the differences lie.

---

13. If we were to adjust for multiple comparisons by changing the significance level for each test, what would be the result on our "alpha"? (Remember that when we do adjust for multiple comparisons that we try to make it harder for ourselves to reject the null hypothesis).

    A. It remains the same

    B. It increases

    C. It decreases

    Answer (C)

    Description. When we adjust for multiple comparisons, we want to make it harder to reject the null hypothesis (in order to reduce the overall type I error rate). To do this, we can either:
    (1) Increase the critical value (since we reject when F, for example, is > a particular value).
    (2) Decrease our significance level (which goes hand in hand with the above since decreasing the significance level will lead to a larger critical value)
    (3) Increase the p-value. This is the way adjusted p-values work. If you use this option, your level of significance stays the same and you compare your adjusted p-values to that value.

---

14. If we were to adjust for multiple comparisons using software (which generally provide adjusted p-values), how would the adjusted value compare with the regular (or unadjusted) p-value?

    A. It remains the same

    B. It increases

    C. It decreases

    Answer (B)

    Description. When we adjust for multiple comparisons, we want to make it harder to reject the null hypothesis (in order to reduce the overall type I error rate). To do this, we can either:
    (1) Increase the critical value (since we reject when F, for example, is > a particular value).

(2) Decrease our significance level (which goes hand in hand with the above since decreasing the significance level will lead to a larger critical value)
(3) Increase the p-value. This is the way adjusted p-values work. If you use this option, your level of significance stays the same and you compare your adjusted p-values to that value.

---

15. A study was conducted to investigate the effects of sleeping 8 hours per night and caffeine use on perceived energy levels. Subjects were categorized by their sleeping habits and by whether or not they used caffeine. Average energy score (as rated on a 10 point scale) by each group were calculated and are summarized in the table below:

Slept 8 Hours?	Caffeine Use?	Average Energy Score
Yes	Yes	9.2
	No	8.1
No	Yes	8.7
	No	5.4

Do the means suggest that an interaction may be present?

A. Yes, it appears as though there is a quantitative interaction

B. Yes, it appears as though there is a qualitative interaction

C. No, it does not appear as though there is any evidence of an interaction

Answer (A)

Description.

---

16. A study was conducted to assess car safety among child passengers in car accidents. The outcome was a severity of injury score. The study considered two main factors—booster seat use (yes, no) and position in car (front versus back seat). Before conducting a two-way ANOVA, the investigators assessed whether there was an interaction between these factors. They found that there was a qualitative interaction. Which are the following results are possible based on the information given? Select all that apply.

A. When used in the front seat, booster seats seemed to increase injury. However, booster seat use in the back seat was protective of injury.

B. When used in the front seat, booster seats seemed to decrease injury. However, booster seat use in the back seat increased risk of severe injury.

C. Booster seat use and position in car are not predictive of injury.

D. Booster seat use is predictive of injury but there is no effect of position in car.

E. Position in car is predictive of injury but there is no effect of booster seat use.

F. Booster seats are protective of injury, regardless of position. However, use of booster seats in the back seat was more protective of injury than use of these seats in the front seat.

G. Booster seats are protective of injury, regardless of position. However, use of booster seats in the front seat was more protective of injury than use of these seats in the back seat.

Answer (A) and (B)

Description.

---

17. A study was conducted to assess car safety among child passengers in car accidents. The outcome was a severity of injury score. The study considered two main factors—booster seat use (yes, no) and position in car (front versus back seat). Before conducting a two-way ANOVA, the investigators assessed whether there was an interaction between these factors. They found that there was a qualitative interaction. Given that there was a qualitative interaction, which of the following next steps are appropriate? Select all that apply.

    A. We should run a one way ANOVA for position and another one way ANOVA for booster seat use.

    B. We should ignore the effect of booster seats and just focus on a one way ANOVA for position.

    C. We should ignore the effect of position and just focus on a one way ANOVA for booster seat use.

    D. We should look at the effect of booster seat use in the front seat first (via a one way ANOVA for booster seat use in this subset) and then look at the effect of booster seat use in the back seat (via a one-way ANOVA for booster seat use in this subset).

    E. We should look at the effect of position among those with a booster seat first (via a one way ANOVA for position in this subset) and then look at the effect of position among those without a booster seat (via a one-way ANOVA for position use in this subset).

    F. The analysis work is complete. No additional analyses are needed.

    Answer (D) and (E)

    Description. Anytime there is an interaction, this means that the effect of one factor depends on the level of the other. That is, the effect of using a booster seat on injury depends on where it was positioned. Or, in other words, the effect of position in the car on injury depends on whether or not a booster seat was used. Given any interaction, we have to stratify by one of the factors and then perform a one way ANOVAs on the other factor so as to understand the nature of each factor by the level of the other. As such, either of these two responses would be appropriate:

    - We should look at the effect of booster seat use in the front seat first (via a one way ANOVA for booster seat use in this subset) and then look at the effect of booster seat use in the back seat (via a one-way ANOVA for booster seat use in this subset).
    - We should look at the effect of position among those with a booster seat first (via a one way ANOVA for position in this subset) and then look at the effect of position among those without a booster seat (via a one-way ANOVA for position use in this subset).

    When there is an interaction, there is not a consistent effect of either factor, so running separate one way ANOVAs for each factor would not be appropriate. The interaction model is just that - a way to evaluate the interaction. The stratified analysis mentioned above is also needed to assess the associations and the directions of the associations - so simply finding that there is an interaction is not enough.

---

18. An investigator is interested in looking at price of gasoline in five different areas of the country. They are concerned that the comparison may not be fair if there are state/local taxes that are added to the price per gallon (continuous/quantitative variable). What type of analysis would you recommend?

    A. One-way ANOVA with price as the response variable and area of the country as the factor/-grouping (explanatory) variable

    B. One-way ANCOVA with price as the response variable and area of the country as the factor/-grouping (explanatory) variable adjusting for taxes

    C. Two-way ANOVA with price as the response variable and area of the country and taxes as the factor/grouping (explanatory) variables

Answer (B)

Description. We'd like an analysis that compares the price across the five areas after adjusting for taxes. As such, we want to perform a one-way ANCOVA with price as the response variable and area of the country as the factor/grouping (explanatory) variable adjusting for taxes.

A two-way ANOVA is used when there are two grouping factors (like gender and age group) and you want to understand the effect of each factor on the response.
