1. Which of the following scenarios represents a contradiction in Multiple Regression?
   a) The residual plot shows a random pattern but the ANOVA p-value is large. ×
   b) The residual plot shows a random pattern but the ANOVA p-value is small. ×
   c) The ANOVA p-value is large but one of the t-tests for the predictors has a small p-value. ✓
   d) The ANOVA p-value is small but one of the t-tests for the predictors has a large p-value. ×
   e) None of the above are contradictions.

2. Suppose there is a correlation of r = 0.9 between number of hours per day students study and GPAs. Which of the following is a reasonable conclusion?
   a) For each extra hour of study per day GPA increases by 0.9 points on average. ×
   b) The Mean Squared Error is 0.81 ✓
   c) 90% of students who study receive high grades. ×
   d) The standard deviation of the points around the regression line is 0.9 ✓
   e) 81% of the variation in GPAs can be explained by variation in number of study hours per day

\[ R^2 = \left( \frac{r}{\sqrt{2}} \right)^2 \]

3. We are interested in finding the linear relation between the number of widgets purchased at one time and the cost per widget. The following data has been obtained:
   
<table>
<thead>
<tr>
<th>X: Number of widgets purchased</th>
<th>1</th>
<th>3</th>
<th>6</th>
<th>10</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y: Cost per widget (in dollars)</td>
<td>55</td>
<td>52</td>
<td>46</td>
<td>32</td>
<td>25</td>
</tr>
</tbody>
</table>

Using the regression line computed as \( \hat{y} = -2.5X + 60 \), compute \( \hat{y} \) when \( x = 30 \) and interpret.
   a) \( \hat{y} = 15 \) dollars, so each widget costs $15, on average, when 30 are purchased.
   b) \( \hat{y} = -15 \) dollars, but because we have a negative slope, \( \hat{y} \) is actually +15 dollars per widget.
   c) \( \hat{y} = -15 \) dollars, which is obvious nonsense - the regression line must be incorrect.
   d) \( \hat{y} = -15 \) dollars, which if we try to interpret, constitutes extrapolation.
   e) \( \hat{y} = -15 \) dollars, which would be considered an influential point in this data set.

4. The statistical significance of the relationship between two quantitative variables can be determined by:
   a) the sign of the slope
   b) the sign of the correlation
   c) the sign of the intercept
   d) the p-value of the slope
   e) the p-value of the intercept
Data was collected to predict a student's GPA based on where they usually seat in the classroom (1=Front, 2=Middle, 3=Back) and their gender (coded as Male=0, Female=1). Graphs and partial output for several regression analyses appear on the last page of this exam. Use that page to answer the following questions.

5. For the model in output 1, the margin of error to construct a 95% confidence interval for the slope is:
   a) (3.53) (0.09017)  
   b) (3.53) (0.40323)  
   c) (2.048) (0.09017)  
   d) (2.045) (0.40323)  
   e) none of the above

6. Find the test statistic for the ANOVA test for the model in output 1.
   a) -3.53  
   b) 18.96  
   c) 12.47  
   d) 30.80  
   e) 69.19

7. Find the correlation between Seating and GPA.
   a) 0.308  
   b) 0.283  
   c) -0.555  
   d) -0.318  
   e) 0.403

8. Interpret the coefficient of Seating for the model in output 1.
   a) As GPA increases by 1 point, students seat, on average, 0.318 feet closer to the front.
   b) As Seating increases by 1, student's GPA decreases, on average, by 0.318 points.
   c) As GPA increases by 1 point, students seat, on average, 0.318 feet further from the front.
   d) As Seating increases by 1, student's GPA increases, on average, by 0.318 points.
   e) We should not interpret the coefficient of Seating for this model.

9. Interpret the intercept for the model in output 1.
   a) The average GPA of students who seat in the front of the room is 3.69.
   b) The average GPA of students who seat in the back of the room is 3.69.
   c) Students whose GPA is zero sit, on average, 3.69 feet from the front of the room.
   d) Students whose GPA is zero sit, on average, 3.69 feet from the back of the room.
   e) We should not interpret the intercept for this model.

10. Looking at the outputs labeled 1 and 2, which single variable is better at predicting GPA?
    a) Seating because it has a lower p-value.
    b) Seating because it has a higher constant.
    c) Gender because it has a higher p-value.
    d) Gender because it has a lower constant.
    e) They are both equally good because the Sums of Squares Total is the same.

11. The models presented have:
    a) the same number of observations and the same number of predictor variables.
    b) the same number of observations but different number of predictor variables.
    c) the same number of predictor variables, observations and responses.
    d) different number of observations and different number of predictor variables.
    e) different number of observations and different number of response variables.

12. Find $R^2$ for the model in output 3:
    a) 31.0%  
    b) 25.7%  
    c) 41.0%  
    d) 26.9%  
    e) 33.3%

13. Comparing model 1 and model 3 we can say that:
    a) Model 3 is better because it has a higher p-value for Seating.
    b) Model 1 is better because it has a higher $R^2$ adjusted.
    c) Model 3 is better because it has more predictor variables.
    d) Model 1 is better because it has a lower MSRes.
    e) They are both equally good because their MSEs are almost the same.
14. Model 3:
   a) Agrees with Model 2 in determining that Gender is not a significant predictor of GPA. ✔
   b) Agrees with Model 1 in determining that Seating is not a significant predictor of GPA. ✗
   c) Agrees with Model 2 in determining that Gender is not a significant predictor of Seating. ✗
   d) Agrees with Model 1 in determining that Seating is not a significant predictor of Gender. ✗
   e) Disagrees with both Models 1 and 2 about which predictors are significant.

15. Model 4:
   a) Should be refined by eliminating Seating because it has a high p-value.
   b) Should be refined by eliminating Gender because it has a high p-value in other models.
   c) Should be refined by eliminating the Constant, since it has a p-value of zero.
   d) Does not need to be refined because the only p-value we interpret here is the ANOVA.
   e) Does not need to be refined because the interaction had a low p-value.

16. The p-value for Seating changed dramatically from Models 1 and 3 to Model 4. This can be explained by:
   a) Multicollinearity  b) Outliers  c) Dummy Variable  d) Influential Points  e) Extrapolation

17. Which of the four models is the best at predicting GPA?
   a) Model 1  b) Model 2  c) Model 3  d) Model 4  e) several of them are equally good

18. The assumptions can be written as $\varepsilon \sim N(0, \sigma)$ for which model?
   a) all of them  b) Model 1  c) Model 2  d) Model 3  e) Model 4

19. Is it reasonable to say the assumption of Normal distribution is satisfied here?
   a) Yes -- the scatterplot of GPA vs Seating shows a pretty good diagonal pattern.
   b) Yes -- the histogram of the residuals is fairly bell-shaped.
   c) Yes -- the plot of Residuals vs Order shows a random pattern.
   d) No -- the plot of Residuals vs Fit does not show a linear pattern.
   e) No -- there is not enough information given to determine this.

20. Is it reasonable to say the assumption of random samples is satisfied here?
   a) Yes -- the scatterplot of GPA vs Seating shows a pretty good diagonal pattern.
   b) Yes -- the histogram of the residuals is fairly bell-shaped.
   c) Yes -- the plot of Residuals vs Order shows a random pattern.
   d) No -- the plot of Residuals vs Fit does not show a linear pattern.
   e) No -- there is not enough information given to determine this.

Match the model with the output:
   a) $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$
   b) $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$
   c) $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon$
   d) $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \varepsilon$
   e) $y = \alpha + \beta x + \varepsilon$

For Model 4, identify the following:
   a) intercept for females
   b) slope for females
   c) intercept for males
   d) slope for males
   e) none of the above

\[
\begin{array}{ccc}
\text{male} & \text{int} & \text{slope} \\
\text{gender=0} & 3.15 & -0.433 \\
\text{female} & 4.23 & 0.59 \\
\end{array}
\]
### Predictor Coef SE Coef T  P
1. Constant 3.6935 0.1948 18.96 0.000
   Seating -0.31835 0.09017 -3.53 0.001
   S = 0.403231 R-Sq = 30.8% R-Sq(adj) = 28.3%

### Analysis of Variance
Source DF SS MS F  P
Regression 1 2.0269 2.0269 0.001
Residual Error 28 4.5527 0.1626
Total 29 6.5796

### Predictor Coef SE Coef T  P
2. Constant 3.0679 0.1251 24.52 0.000
   Gender -0.0221 0.1770 -0.12 0.902
   S = 0.404618 R-Sq = 0.1% R-Sq(adj) = 0.0%

### Analysis of Variance
Source DF SS MS F  P
Regression 1 0.0037 0.0037 0.02 0.902
Residual Error 28 6.5759 0.2349
Total 29 6.5796

3. Predictor Coef SE Coef T  P
   Constant 3.7046 0.2120 17.48 0.000
   Seating -0.31835 0.09178 -3.47 0.002
   Gender -0.0221 0.1499 -0.15 0.884
   S = 0.410465 R-Sq(adj) = 25.7%

### Analysis of Variance
Source DF SS MS F  P
Regression 2 2.0305 1.0153 6.03 0.007
Residual Error 27 4.5490 0.1698
Total 29 6.5796

4. Predictor Coef SE Coef T  P
   Constant 3.1545 0.2334 13.51 0.000
   Seating -0.0433 0.1081 -0.40 0.692
   Gender 1.0761 0.3301 3.27 0.003
   Seating*Gender -0.5501 0.1528 -3.60 0.001
   S = 0.341713 R-Sq = 53.9% R-Sq(adj) = 48.5%

### Analysis of Variance
Source DF SS MS F  P
Regression 3 3.5436 1.1812 10.12 0.000
Residual Error 26 3.0360 0.1168
Total 29 6.5796

**Residual Plots for GPA**
- Normal Probability Plot
- Residual vs. Fits
- Histogram
- Versus Order