Name: ___________  
UIN: ___________

Class time (Please Circle): 11:10am-12:25pm or 12:45pm-2:00pm

Instructions:

1. Please provide your name and UIN.
2. Circle the correct class time.
3. To get full credit on answers to this exam, be clear, rigorous, and thorough in your responses.
4. You cannot get full credit (full or partial) unless something is written.
5. Sign the Aggie Pledge:

“On my honor, as an Aggie, I have neither given nor received unauthorized aid on this academic work.”

Signature ___________ Date ___________
1. Fundamentals of Regression Analysis (29 points)

(1pt) 1. The sum of the residuals in any regression analysis where the model is linear in 
parameters and where the model includes an intercept is ________________.

(1pt) 2. Name the theorem that guarantees the OLS estimators are BLUE

\textit{Gauss-Markov Theorem}

(1pt) 3. The acronym OLS stands for ________________.

(1pt) 4. With the use of OLS, we choose parameter estimates so as to ________________ the 
error sum of squares.

(1pt) 5. The degrees-of-freedom for a regression analysis where the sample size is 153 and the 
number of parameters to be estimated is 6 is ________________.

6. If the SSE = 100 and the SSR = 400, then

\begin{align*}
\text{a. SST} &= 500 \\
\text{b. The } R^2 \text{ associated with this situation is: } 0.8 \
\end{align*}

7. If the AIC for one model (model A) is 26.79 and the AIC for another model 
(model B) is 27.89,

\begin{align*}
\text{a. which model is preferred? Is it model A or model B? Circle the correct answer.} \\
\text{b. The acronym AIC stands for ________________}. \\
\text{c. Besides AIC, what is another model selection criterion?} \\
\text{d. Model selection criteria depend on three factors. Name two of them.} \\
\end{align*}

8. Let \( Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + u_i \).

\begin{align*}
\text{a. If } SSE = 240 \text{ and if } n = 64, \text{ then the unbiased estimate of the residual variance} \\
\text{is } \frac{\hat{\sigma}^2 = SSE}{n-p} = \frac{240}{60}.
\end{align*}

\begin{align*}
\text{b. The standard error of the regression is } 2. \\
\text{c. Provide the data matrix } X. \text{ List the columns of the matrix and indicate the relevant} \\
\text{number of rows.} \\
\end{align*}

\[
X = \begin{bmatrix} \text{i} & x_{i1} & x_{i2} & x_{i3} \\
\vdots & \vdots & \vdots & \vdots \\
\text{i} & x_{641} & x_{642} & x_{643} \end{bmatrix}_{n \times 4}, n = 64
\]
(1pt) d. Write the expression of the OLS estimator for $\beta_0$, $\beta_1$, $\beta_2$, and $\beta_3$ in terms of the data matrix $X$ and the column vector $y$. 

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'y$$

(1pt) 9. True or False. $R^2 > R^2$. Circle the correct answer. 

(2pts) 10. The residual variance ($\sigma^2$) is assumed to be constant across all observations in regression analysis. The technical name of this assumption is **Homoscedasticity**. If this assumption is violated, the technical name of this condition is **Heteroscedasticity**.

(2pts) 11. Regression analysis consists of four components or phases. List TWO of these components.

a. **SPECIFICATION**

b. **ESTIMATION**

c. **VERIFICATION**

d. **PREDICTION**

12. Suppose that $e_i$ and $e_j$ represent two error terms at two different observations (observation $i$ and observation $j$, respectively).

(1pt) a. We assume that the $\text{corr}(e_i,e_j) = \bigcirc$.

(1pt) b. What is the technical name for this assumption? **INDEPENDENT ERROR TERMS**

(1pt) 13. Let $X$ represent any explanatory variable in a multiple regression, and let $e$ represent the error term. We assume that $\text{corr}(X,e) = \bigcirc$.

(1pt) 14. Other things the same, simple models generally are preferable to complex models, especially in forecasting. This statement refers to what fundamental principle? **SIMPLICITY**

(1pt) 15. Imposing additional information on a regression model often improves model performance. This statement refers to what fundamental principle? **SHRINKAGE**

(1pt) 16. By assumption, we may not write any explanatory variable as a perfect linear combination of other exogenous factors in a multiple regression model. What is the technical name associated with this assumption? **NO MULTICOLLINEARITY**

17. Demand estimation may rest on regression analysis as well as on two other quantitative approaches. Name these approaches.

(1pt) a. **CONSUMER INTERVIEWS**

(1pt) b. **MARKET EXPERIMENTS**
2. Data Considerations (12 points)

(3pts) a. Calculate the mean and median of the following sample of observations for a variable labeled EMPR. Show all work.
18, 10, 15, 13, 17, 15, 12, 15, 18, 16, 11

Mean = 14.545
Median = 15

The mean and median are both measures of **CENTRAL TENDENCY** for EMPR.

(1pt) b. Calculate the range of the data for variable EMPR.

\[
\text{Range} = \text{Maximum} - \text{Minimum} = 18 - 10 = 8
\]

(2pts) c. Suppose that skewness coefficient for the variable EMPR is equal to -0.30. Explain this situation.

\[\text{LEFT TAIL ON THE DISTRIBUTION OF EMPR} \]
\[\text{NOT SYMMETRIC} \]

(2pts) d. Suppose the kurtosis coefficient is equal to 1.89. Explain this situation.

\[\text{LESS PEAKED THAN A NORMAL DISTRIBUTION} \]
\[\text{KURTOSIS COEFFICIENT OF A NORMAL DISTRIBUTION IS 3} \]

(1pt) e. Name the statistic associated with testing whether the variable EMPR follows a normal distribution.

\[\text{JB OR JARQUE-BERA TEST STATISTIC} \]
\[\text{JB} \sim X^2 \]

(1pt) f. Suppose that the standard deviation for EMPR is 2.73. Calculate the coefficient of variation for EMPR.

\[CV = \frac{s}{\mu} \times 100\% = 18.77\% \]

(1pt) g. What SAS procedure (PROC) could be used to obtain information about descriptive statistics?

\[\text{PROC MEANS} \]

(1pt) h. Calculate the mode for EMPR.

\[\text{MOST FREQUENT OBSERVATION} = 15 \]
3. Simple Linear Regression (12 points)

Suppose we wish to estimate the simple linear regression model of personal consumption expenditures (PCE) as a function of disposable income (DPI):

\[ PCE_t = \beta_0 + \beta_1 DPI_t + u_t \quad t = 1, 2, ..., 60. \]

(1pt) a. What is the technical name of \( u_t \)?

\[ u_t \rightarrow \text{Error term, Disturbance term, Innovation} \]

(1pt) b. True or False. \( u_t \) explicitly relates that the relationship between \( PCE_t \) and \( DPI_t \) is not an identity. Circle the correct answer.

(1pt) c. \( u_t \) represents two elements. Name ONE of them.

- (1) Omission of other variables
- (2) Measurement error

(2pts) d. Provide a reasonable scatterplot associated with the relationship between PCE and DPI. Be sure to label your axes.

![Scatterplot](image)

(1pt) e. What is the technical name of \( PCE_t \), in the context of regression analysis?

- Dependent variable
- Endogenous variable
- Regressand

(1pt) f. What is the technical name of \( DPI_t \), in the context of regression analysis?

- Explanatory variable
- Independent variable
- Exogenous variable
- Predetermined variable
- Right-hand side variable
g. Suppose that \[ \sum_{t=1}^{60} (DPI_t - \overline{DPI})(PCE_t - \overline{PCE}) = 1,500 \]

and that \[ \sum_{t=1}^{60} (DPI_t - \overline{DPI})^2 = 2,000 \]

(1pt) i. Calculate \( \hat{\beta}_1 \). Show all work.

\[
\hat{\beta}_1 = \frac{1,500}{2,000} = 0.75
\]

(1pt) ii. Interpret this estimated coefficient.

A one unit change in DPI leads to a 0.75 unit change in PCE.

(1pt) h. Suppose that \( \overline{DPI} = $100,000 \) and \( \overline{PCE} = $80,000 \). Calculate \( \hat{\beta}_0 \). Show all work.

\[
\hat{\beta}_0 = \overline{PCE} - 0.75 \overline{DPI} = 80,000 - 0.75(100,000) \\
\hat{\beta}_0 = 80,000 - 75,000 = 5,000
\]

(2pts) i. On the basis of the information given in (g) and (h), what is the percentage change in PCE due to a one percent change in DPI? Show all work.

\[
\frac{\% \Delta PCE}{\% \Delta DPI} = \frac{0.75 \times DPI}{PCE} = \frac{0.75 \times 100,000}{80,000}
\]

\[
\frac{\% \Delta PCE}{\% \Delta DPI} = \frac{75,000}{80,000} = 0.9375
\]
4. Multiple Regression (16 points)

Consider the multiple regression model
\[ Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + u. \]

Our estimated model is given as:
\[ Y = -10.52 - 1.45X_1 + 0.35X_2 + 0.45X_3 \]
\[ (6.73) \quad (0.24) \quad (0.01) \quad (0.03) \]

Standard errors of the estimated coefficients are reported in parentheses. The number of the sample observations in this regression analysis is 54.

(2pts) a. Suppose that \( X_2 \) is measured in dollars. The impact of a $100 increase in \( X_2 \) is equal to what kind of change in \( Y \)? Show all work.

\[ \frac{dy}{dx_2} = 0.35 \quad \text{so a $100 increase in } X_2 \text{ leads to a 35 unit change in } Y \]

(2pts) b. Consider \( H_0: \beta_1 = 0 \). What is the purpose of this test?

To determine if \( X_1 \) is a "driver" of \( Y \) or alternatively if \( X_1 \) is statistically associated with \( Y \).

(3pts) c. Provide a test statistic associated with the test in (b). Be sure to indicate the degrees of freedom. Would you likely reject or fail to reject this null hypothesis? Why?

\[ df = n-p = 54-4=50 \]
\[ t_{50} = \frac{\hat{\beta}_1}{\text{se(\hat{\beta}_1)}} = \frac{-1.45}{0.24} = -6.04 \]

\[ \text{reject } H_0 \text{ since the } p\text{-value of this test statistic is less than 0.05 or 0.10.} \]

(3pts) d. Test \( H_0: \beta_3 = 0.50 \). Provide the test statistic associated with this test. Be sure to indicate the degrees of freedom. Would you likely reject or fail to reject this null hypothesis? Why?

\[ t_{50} = \frac{\hat{\beta}_3 - 0.50}{\text{se(\hat{\beta}_3)}} = \frac{0.45 - 0.50}{0.03} = \frac{-0.05}{0.03} = -1.67 \]

\[ \text{likely to fail to reject } H_0 \text{ since the } p\text{-value of this statistic is greater than 0.01, 0.05, or 0.10.} \]

e. Suppose we wish to test \( \beta_2 + \beta_3 = 1 \).

(2pts) i. What test statistic do we use?

\[ F\text{-statistic} \]

(2pts) ii. What are the appropriate degrees-of-freedom associated with this test statistic?

\[ F_{1, 50} \]

(1pt) f. Suppose that this sample corresponds to monthly data from January 2007 to June 2011. We may characterize this data set as a ________.

(1pt) g. Suppose that this sample corresponds to a set of observations concerning the behavior of 54 firms. We may characterize this data set as a ________.
5. Multiple Regression/SAS Output (31 points)

The following regression was fit to determine the demand for fajita entrees sold at On the Border:

\[ \ln Q_t = b_0 + b_1 \ln P_t + b_2 \ln A_t + b_3 \ln DPI_t + u_t, \]

where \( Q_t \) is the quantity of On the Border fajita entrees sold during the \( t \)th quarter; \( P_t \) is the retail price in dollars of On the Border fajita entrees during the \( t \)th quarter; \( A_t \) represents the dollars spent for advertising during the \( t \)th quarter; and \( DPI_t \) is disposable income for the U.S. during the \( t \)th quarter. The summary of output of the regression analysis conducted in SAS is given in the attached page.

(4pts) a. Provide four program statements in SAS to obtain this output.

```sas
DATA ON_THE_BORDER;
  INPUT Q1 P A DPI;
  lnQ1 = log(Q1); lnP = log(P); lnA = log(A); ln_DPI = log(DPI);
  PROC REG DATA = ON_THE_BORDER;
  MODEL lnQ = lnP lnA ln_DPI;
```

(2pts) b. How much variability in the quantity of On the Border fajita entrees sold is explained by this regression analysis? Does this model provide a good fit to the data?

\[ R^2 = 0.9414, \quad 94.14\% \text{ of the variability in the quantity of On the Border fajita entrees sold is explained by the model.} \]

The model provides a good fit to the data.

(2pts) c. The \( R^2 \) of the regression analysis is given as 0.9255. Derive this measure from the information given.

\[ R^2 = 1 - \left( \frac{SSE(n-1)}{SSY(n-1)} \right) = 1 - \left( \frac{0.0384}{0.6533} \right) = 0.9255 \]

\[ \text{also } R^2 = 1 - \left( 1 - R^2 \right) \frac{n-1}{n-p} = 1 - (0.0586) \frac{11}{11} = 0.9255 \]

(4pts) d. From the information given, test \( H_0: b_1 = b_2 = b_3 = 0 \) at the 0.05 level of significance. Please give the appropriate test statistic and p-value. Also specify your conclusion concerning the test.

\[ F_{3,11} = 58.95, \quad \text{p-value} < 0.0001 \]

reject \( H_0 \) since \( p \)-value < 0.05
Fall 2013
Test 2 - November 7, 2013

(1pt) e. True or False. The demand for On the Border fajita entrees is elastic. Circle the correct answer.

(1pt) f. True or False. On the Border fajita entrees are a luxury good. Circle the correct answer.

(2pts) g. If On the Border increases advertising expenditures by 5 percent, what is the percentage change in quantity of fajita entrees sold, ceteris paribus? Show all work.

\[ 1,309.23 \times 5 = 6,546 \]

(6pts) h. Which variable(s)—\( P, A, \) or \( DPI \)—significantly influence the quantity of On the Border fajita entrees sold? Assume a level of significance of 0.05. Be sure to fully explain your answer.

\[ H_0: b_1 = 0 \quad H_0: b_2 = 0 \quad H_0: b_3 = 0 \]

\[ t_{-1} = \frac{-0.67516}{0.21074} = -3.20 \quad t_{+1} = \frac{1,309.23}{0.14719} = 8.90 \quad t_{+2} = \frac{1,138.82}{0.40237} = 2.83 \]

\[ p-value < 0.001 \quad p-value < 0.001 \quad p-value 0.0164 \]

reject \( H_0 \) \quad reject \( H_0 \) \quad reject \( H_0 \)

all variables—\( P, A, \) and \( DPI \) significantly influence the quantity of On the Border fajita entrees sold.

(4pts) i. The 95% confidence interval for \( b_1 \), the coefficient associated with \( \ln P \), is given by \([-1.139, -0.211]\). Provide enough information to demonstrate how this confidence interval is derived.

\[ \hat{b}_1 \pm t_{0.025} \times s.e. (\hat{b}_1) \]

\[ -0.67516 \pm 2(0.21074) \quad c \approx 2 \]

\[ -0.62916 \pm 0.42148 \]

(3pts) j. Suppose that \( P = 10, A = 50,000, \) and \( DPI = 42,000 \). Predict the quantity of On the Border fajita entrees sold. Show all work.

Predicted \( = 62,059 \)

\[ -13.6518 + 1,309.23(50,000) + 1,138.82(42,000) - 0.67516 \ln(10) = 11,036 \]

(2pts) k. List TWO additional potential explanatory factors associated with the demand for fajita entrees sold at On the Border. Be specific.

(1) Seasonality
(2) Competitor price
The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: lquantity_sold

Number of Observations Read: 16
Number of Observations Used: 15
Number of Observations with Missing Values: 1

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>3</td>
<td>0.64049</td>
<td>0.21350</td>
<td>58.95</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>11</td>
<td>0.03984</td>
<td>0.00362</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>14</td>
<td>0.68033</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Root MSE: 0.06018
R-Square: 0.9414
Dependent Mean: 11.04092
Adj R-Sq: 0.9255
Coeff Var: 0.54505

Parameter Estimates

<table>
<thead>
<tr>
<th>Variable</th>
<th>DF</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>t</th>
<th></th>
<th>95% Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-13.69818</td>
<td>3.50543</td>
<td>-3.91</td>
<td>0.0024</td>
<td></td>
<td>-21.41359</td>
<td>-5.98277</td>
</tr>
<tr>
<td>lprice</td>
<td>1</td>
<td>-0.67516</td>
<td>0.21074</td>
<td>-3.20</td>
<td>0.0084</td>
<td></td>
<td>-1.13900</td>
<td>-0.21131</td>
</tr>
<tr>
<td>ladvertising_exp</td>
<td>1</td>
<td>1.30923</td>
<td>0.14719</td>
<td>8.90</td>
<td>&lt;.0001</td>
<td></td>
<td>0.98528</td>
<td>1.63318</td>
</tr>
<tr>
<td>ldisp_income</td>
<td>1</td>
<td>1.13882</td>
<td>0.40237</td>
<td>2.83</td>
<td>0.0184</td>
<td></td>
<td>0.25322</td>
<td>2.02443</td>
</tr>
</tbody>
</table>