Name: ______ Key________________________
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
(25pts) **Fundamentals of Regression Analysis**

(1 pt) 1. What is the name of the statistical technique associated with estimating parameters of a multiple regression model? **Ordinary Least Squares (OLS)**

(1 pt) 2. If the SST = 1,000 and SSE = 100, the SSR = __900__________.

(1 pt) 3. Calculate the $R^2$ statistic associated with Question 2. __.9__________

(3 pts) 4. In a particular regression analysis, the sample size is 128, the number of parameters to be estimated is 8, and SSE = 1,920.

(a) Degrees of freedom = ____120__________.

(b) Residual variance = ____1,920/120=16__________.

(c) Standard error of the regression = ____4__________.

(1 pt) 5. **True** or **False**. The residual variance is assumed to be constant across all observations in regression analysis. The technical name of this assumption is heteroscedasticity. **False**

(1 pt) 6. **True** or **False**. $R^2$ is always between 0 and 1. **False**

(2 pts) 7. Suppose that $e_i$ and $e_j$ represent two error terms at two different observations (observation $i$ and observation $j$, respectively).

(a) We assume that corr($e_i$, $e_j$) = ____0______.

(b) What is the technical name for this assumption? __no autocorrelation or independence of error terms__________

(3 pts) 8. The ____Gauss-Markov__________ theorem guarantees the parameter estimates associated with any regression model have desirable statistical properties. These properties have been referred to by the acronym BLUE. The B in this acronym stands for ____Best__________, and the U in the acronym stands for ____Unbiased__________.

(1 pt) 9. **True** or **False**. In a linear model with the inclusion of an intercept, it is possible that all the residuals can be positive. **False**
(2 pts) 10. Regression analysis consists of four components or phases. List TWO of these components.

(a) Model Specification  
(b) Estimation  
(c) Verification/Hypothesis Testing  
(d) Prediction/Forecasting

(1 pt) 11. Other things the same, simple models generally are preferable to complex models, especially in forecasting. This statement refers to what fundamental principle? Principle of Parsimony

(1 pt) 12. Imposing restrictions either on estimated parameters or on forecasts often improves model performance. This statement refers to what fundamental principle? Shrinkage Principle

(1 pt) 13. Regression analysis________________ is a statistical technique that describes relations among dependent and explanatory variables.

(1 pt) 14. A deterministic relation is an association among variables that is known with certainty. Provide one example of such a relation in managerial economics.  
Total Revenue=Price*Quantity  
Total Cost=Total Fixed Cost +Total Variable Cost  
Profit=Total Revenue-Total Cost

(1 pt) 15. A __time series____________ is a daily, weekly, monthly, or annual sequence of data on various economic variables.

(1 pt) 16. A ____cross section____________ is a group of observations on various economic variables at any given point in time.

(1 pt) 17. By assumption, we may not write any explanatory variable as a perfect linear combination of other predetermined variables in a multiple regression model. What is the technical name associated with this assumption?  
_____No perfect collinearity____________

(1 pt) 18. The _consumer interview (or survey)__ method requires questioning customers or potential customers to estimate the relation behavior demand and a variety of underlying functions.

(1 pt) 19. **True or False.** Market experiments are expensive and are usually undertaken on a scale too small to allow high levels of confidence in the results. Circle the correct answer. **True**
(12 pts) **Key Issues**

(5pts) 20. A local firm offered $2.00 off the $12.00 regular price of its product and sales rose from 3,200 to 4,000 units sold. On the basis of this information, piece together the simple linear demand curve. \( Q = a + bP \). That is, find \( a \) and \( b \).

\[ \begin{align*}
P &= $12, \ Q = 3,200; & \quad P &= $10, \ Q = 4,000 \\
Q &= 8,000 - 400P; & \quad \text{so } a &= 8,000 \text{ and } b = -400
\end{align*} \]

(7 pts) 21. A common form of demand functions is the multiplicative model given by \( Q_t = b_0 P_t^{b_1} A_t^{b_2} I_t^{b_3} \), where \( Q_t \) represents sales (in units) of a particular commodity, \( P_t \) represents the price of the commodity ($/unit) in the time period \( t \), \( A_t \) represents advertising and promotion expenditures in the period \( t \) ($), and \( I_t \) represents disposable personal income in time period \( t \) ($).  

(3pts) (a) Demonstrate how this relationship can be transformed into a multiple regression model.

\[ \ln Q_t = \ln b_0 + b_1 \ln P_t + b_2 \ln A_t + b_3 \ln I_t + e_t \]

(1pt) (b) What is the expected sign of \( b_1 \)? <0, negative

(1pt) (c) What is the expected sign of \( b_3 \) for most commodities? >0, positive

(2 pts) (d) Provide a reasonable scatter plot associated with the relationship for \( Q_t \) and \( A_t \)?

Diagram such that \( Q \) is on the y-axis and \( A \) is on the x-axis with a set of observations that yield a positive slope.
(12 pts) **Data Considerations**

22. Suppose that the following information was obtained from a SAS output for a variable labeled INTR. The number of observations recorded was 96.

   Mean = 3,175  
   Median = 3,200  
   Standard Deviation = 409.27  
   Minimum = 2,600  
   Maximum = 3,800  
   Skewness = 0.11  
   Kurtosis = -1.28  
   Mode = 3,200

   (1 pt) (a) What SAS procedure (PROC) is used to obtain this information?  
   **Proc Means**

   (1 pt) (b) What is the 50th percentile of INTR?  **Median = 3,200**

   (1 pt) (c) What is the spread or dispersion about the mean for INTR?  
   **Standard Deviation = 409.27**

   (2 pts) (d) What statement can you make about the symmetry or lack of symmetry of the distribution of INTR? Provide supporting evidence.  
   *Because the skewness coefficient equals 0.11, the distribution is not symmetric. This distribution has a tail on the right.*

   (2 pts) (e) What statement can you make about the peakedness of the distribution of INTR? Provide supporting evidence.  
   *Because the kurtosis coefficient equals -1.28, the distribution is less peaked than a standard normal distribution.*

   (1 pt) (f) Calculate the data range for INTR. **Range = Maximum - Minimum = 1,200**

   (1 pt) (g) Name the statistic associated with testing whether the variable INTR follows a normal distribution.  **Jarque-Bera (JB)**

   (1 pt) (h) What is the coefficient of variation for INTR?  
   *The coefficient of variation = (standard deviation/mean)*100*  
   *(409.27/3,175)*100 = 12.89*

   (1 pt) (i) What is the most frequent observation for INTR? **3,200 = mode**

   (1 pt) (j) **True** or **False**. The mean and the median are both measures of central tendency for INTR. Circle the correct answer. **True**
(51 pts) **Problems of Regression Analysis**

23. Suppose we wish to consider the relationship between food away from home expenditures (FAFH) and disposable income (DPI). We express this relationship as:

\[ FAFH_t = \beta_0 + \beta_1 DPI_t + u_t \quad t = 1, 2, \ldots, 90 \]

(1 pt) (a) What is the technical name of this relationship?

Simple Linear Regression Model

(1 pt) (b) What is the technical name of FAFH\(_t\), in the context of regression analysis?

Dependent variable or Endogenous variable or regressand

(1 pt) (c) What is the technical name of DPI\(_t\), in the context of regression analysis?

Explanatory Variable, Exogenous Variable, Predetermined Variable, RHS variable, Independent Variable, Regressor

(2 pts) (d) Suppose that this model is **NOT** able to explain 19% of the variation in FAFH. What is the correlation coefficient between FAFH and DPI?

\[ R^2 = .81 \quad r = \text{correlation coefficient} = \sqrt{.81} = .9 \]

(3 pts) (d) Suppose that \[ \sum_{t=1}^{90} (DPI_t - \overline{DPI})(FAFH_t - \overline{FAFH}) = 1,200 \]

and that \[ \sum_{t=1}^{90} (DPI_t - \overline{DPI})^2 = 1,600 \]

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

\[ \hat{\beta}_1 = 1,200/1,600 = 3/4 = .75 \]

(2 pts) (e) Suppose that \( \overline{DPI} = $50,000 \) and \( \overline{FAFH} = $40,000 \). Calculate \( \hat{\beta}_0 \).

Show all work.

\[ \hat{\beta}_0 = FAFH - \hat{\beta}_1 \overline{DPI} = 40,000 - .75*50,000 = 2,500 \]

(3 pts) (f) Suppose that the SIC of this model (call it model A) is 18.74. Now you estimate an alternative version of the model, namely, \( FAFH_t = 1,000 + .12 \ln DPI_t \).

You find that the SIC for this model (call it model B) is 17.89.

(i) What does the acronym SIC stand for? **Schwarz Information Criterion**
(ii) Which version of the relationship between FAFH and DPI do you choose? That is, do you choose model A or model B? **Model B**

The SIC for model B is lower than the SIC for model A.

(iii) Name another model selection criterion besides SIC.

- Akaike Information Criterion
- Adjusted R2

24. Consider the multiple regression model \( Y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + a_3 x_{3t} + a_4 x_{4t} + v_t \).

The estimated model is given as:

\[
Y = 5.12 + 1.72x_1 + 1.18x_2 + 0.57x_3 -0.75x_4.
\]

(6.73)  (0.72)    (0.47)    (1.25)    (0.31)

Standard errors of the estimated coefficients are reported in parentheses. The number of the sample observations in this regression analysis is 65.

(2 pts) (a) Suppose that \( x_4 \) is measured in dollars. The impact of a $20 increase in \( x_4 \) is equal to a ___-15_________ unit change in \( y \). Show all work.

(2 pts) (b) If \( x_2 = 10 \) and \( y = 50 \), calculate the percentage change in \( y \) due to a 1 percent change in \( x_2 \). Show all work.

Elasticity=1.18* \( x_2 \) / \( y \)=1.18*10/50=0.236

(3 pts) (c) Demonstrate how to test \( H_0 : a_1 = 0 \). Specify the t-statistic and the degrees-of-freedom. Would you likely reject or fail to reject this null hypothesis? Why?

\[
t\text{-statistic with 60 df; } t = \frac{1.72}{0.72} = 2.39; \text{ we reject this null hypothesis at the 5 percent level of significance (t-value > 2)}
\]

(3 pts) (d) Demonstrate how to test \( H_0 : a_3 = 0 \). Specify the t-statistic and the degrees-of-freedom. Would you likely reject or fail to reject this null hypothesis? Why?

\[
t\text{-statistic with 60 df; } t = \frac{0.57}{1.25} = 0.456; \text{ we fail to reject this null hypothesis at the 5 percent level of significance (t-value < 2)}
\]

(3 pts) (e) Demonstrate how to test \( H_0 : a_2 = 1 \). Specify the t-statistic and the degrees-of-freedom. Would you likely reject or fail to reject this null hypothesis? Why?

\[
t\text{-statistic with 60 df; } t = \frac{(1.18-1)}{0.47} = 0.38; \text{ we fail to reject this null hypothesis at the 5 percent level of significance (t-value < 2)}
\]
25. Electronic Data Processing (EDP), Inc. is a small but rapidly growing firm that provides data services. Potential influences of contract sales (Q) are: (1) personal selling expenses (PSE); (2) advertising expenditures (AD); and (3) contract price (P). You hypothesize a linear relation between contract sales, contract price, advertising expenditures and personal selling expenses as follows:

\[ Q_t = b_0 + b_1 P_t + b_2 A_D_t + b_3 PSE_t + u_t, \]

where \( Q_t \) is the quantity of contracts sold during the \( t^{th} \) month; \( P_t \) is the contract price in dollars in month \( t \); \( A_D_t \) represents the dollars spent for advertising in month \( t \); and \( PSE_t \) corresponds to personal selling expenses in month \( t \). The summary of the output of the regression analysis conducted in SAS is given in the attached pages.

(a) How much variability in the contract sales is explained by the regression analysis?

\[ R^2 = 0.9697 \]

(b) What is the standard error of the regression?

\[ \text{Sqrt}(15357) = 123.92 \]

(c) 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, degrees of freedom, and p-value. Would you reject or fail to reject this null hypothesis?

\[ F \text{ statistic} = 85.40 \text{ with 3 and 8 degrees of freedom ; } p\text{-value} < 0.0001 \]

We reject this null hypothesis at any reasonable level of significance.

(d) The t-statistic associated with the estimated coefficient for \( P_t \) is -2.91. Derive this measure from the information given.

\[ t = \frac{-0.29648}{0.10195} = -2.91 \]

(d) The \( R^2 \) of the regression analysis is given as 0.9584. Derive this measure from the information given.

\[ R^2 = 1 - \frac{(SSE/8)/(SST/11)} = 1 - \frac{15357}{368844.73} = 1 - 0.0416354 = 0.95836 \]
(6pts) (e) Which variable(s)—$P_t$, $AD_t$, or $PSE_t$ are significantly significant drivers of the quantity of Mrs. Smyth’s pies sold? Assume a level of significance of 0.01. Provide supporting documentation.

At the 0.01 level, $P_t$ and $PSE_t$ are significantly significant drivers of the quantity of Mrs. Smyth’s pies sold but not $AD_t$. The p-value of $P_t$ is $0.0196/2 = 0.0098$ which is < 0.01. The p-value of $PSE_t$ is $0.0017/2 = 0.00085$ which is < 0.01. The p-value of $AD_t$ is $0.0327/2 = 0.01635$ which is > 0.01.

(3 pts) (f) The 95% confidence interval for $b_2$, the coefficient associated with $PSE_t$, is given by $[0.03306, 0.09935]$. Derive this confidence interval from the information given. Assume a value of 2.306 for the t-statistic. Show all work.

Confidence interval equal to estimated coefficient plus or minus 2.306*standard error of the estimated coefficient. So,

$0.06621 \pm (2.306)(0.01437) = 0.06621 \pm 0.03314$

$0.06621 - 0.03314 = 0.03307$; $0.06621 + 0.03314 = 0.09935$

(2 pts) (g) Predict the number of contracts sold with $P_t = 3,200$, $AD_t = 20,000$, and $PSE_t = 40,000$. Show all work. Round your answer to the nearest integer.

2,302 is the prediction

(3 pts) (h) Based on your answer for (g) calculate the own-price elasticity of demand for services provided by EDP. Show all work. What statement can you make about the demand for services provided by EDP?

Own-price elasticity of demand=$-0.29648*3,200/2,302=-0.412$; therefore, the demand for these services is inelastic.

(1 pt) (i) What SAS procedure (PROC) was used to produce the output in this regression analysis?

Proc Reg or Proc Autoreg or Proc Model

(1 pt) (j) The residual associated with the first observation is -66.88. Explain how this residual is derived using the information given in the SAS output.

Any residual is calculated as the difference between the actual value and the predicted value. For the first observation, the actual value is 2500, and the predicted value is 2567. So the difference is -67.
The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: units_sold

Number of Observations Read  12
Number of Observations Used  12

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>3934438</td>
<td>1311479</td>
<td>85.40</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>8</td>
<td>122854</td>
<td>15357</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>11</td>
<td>4057292</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Root MSE            123.92241    R-Square     0.9697
Dependent Mean     2020.83333    Adj R-Sq     0.9584
Coeff Var             6.13224

Parameter Estimates

| Variable                | DF | Estimate  | Standard Error  | t Value | Pr > |t| | 95% Confidence Limits |
|-------------------------|----|-----------|-----------------|---------|------||  |                |
| Intercept               | 1  | -117.51347| 333.52587       | -0.35   | 0.7337 | | -886.62552    651.59857 |
| price                   | 1  | -0.29648  | 0.10195         | -2.91   | 0.0196 | | -0.53158     -0.06138 |
| advertising_exp         | 1  | 0.03598   | 0.01395         | 2.58    | 0.0327 | | 0.00381      0.06814 |
| personal_selling_exp    | 1  | 0.06621   | 0.01437         | 4.61    | 0.0017 | | 0.03306      0.09935 |

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: units_sold

Output Statistics

<table>
<thead>
<tr>
<th>Obs</th>
<th>Dependnt Variable</th>
<th>Predicted Value</th>
<th>Mean</th>
<th>Std Error</th>
<th>Std Error</th>
<th>Student Residual</th>
<th>Residual Standard</th>
<th>Cook's Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2500</td>
<td>2567</td>
<td>78.3647</td>
<td>-66.8800</td>
<td>95.999</td>
<td>-0.697</td>
<td>0.285</td>
<td>0.081</td>
</tr>
<tr>
<td>2</td>
<td>2250</td>
<td>2213</td>
<td>66.0123</td>
<td>37.0164</td>
<td>104.9</td>
<td>0.353</td>
<td>0.019</td>
<td>0.012</td>
</tr>
<tr>
<td>3</td>
<td>1750</td>
<td>1758</td>
<td>55.7005</td>
<td>-8.3549</td>
<td>110.7</td>
<td>-0.0753</td>
<td>0.019</td>
<td>0.000</td>
</tr>
<tr>
<td>4</td>
<td>1500</td>
<td>1646</td>
<td>58.7320</td>
<td>-146.2475</td>
<td>109.1</td>
<td>-1.340</td>
<td>0.019</td>
<td>0.010</td>
</tr>
<tr>
<td>5</td>
<td>1000</td>
<td>1029</td>
<td>71.5874</td>
<td>-29.2614</td>
<td>101.2</td>
<td>-0.289</td>
<td>0.019</td>
<td>0.013</td>
</tr>
<tr>
<td>6</td>
<td>2500</td>
<td>2310</td>
<td>102.1521</td>
<td>189.8421</td>
<td>70.155</td>
<td>2.706</td>
<td>0.019</td>
<td>0.130</td>
</tr>
<tr>
<td>7</td>
<td>2750</td>
<td>2597</td>
<td>74.6980</td>
<td>153.4850</td>
<td>98.879</td>
<td>1.552</td>
<td>0.019</td>
<td>0.344</td>
</tr>
<tr>
<td>8</td>
<td>1750</td>
<td>1804</td>
<td>41.9742</td>
<td>-53.8297</td>
<td>116.6</td>
<td>-0.462</td>
<td>0.019</td>
<td>0.007</td>
</tr>
<tr>
<td>9</td>
<td>1250</td>
<td>1187</td>
<td>70.1725</td>
<td>63.4411</td>
<td>102.1</td>
<td>0.621</td>
<td>0.019</td>
<td>0.046</td>
</tr>
</tbody>
</table>

10
<table>
<thead>
<tr>
<th>Year</th>
<th>Age</th>
<th>Difference</th>
<th>Residual</th>
<th>Predicted</th>
<th>Press Residual</th>
<th>Marker</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>3000</td>
<td>3133</td>
<td>89.8087</td>
<td>-133.3452</td>
<td>85.388</td>
<td>-1.562</td>
<td>***</td>
</tr>
<tr>
<td>11</td>
<td>2000</td>
<td>1931</td>
<td>66.5651</td>
<td>69.1030</td>
<td>104.5</td>
<td>0.661</td>
<td>*</td>
</tr>
<tr>
<td>12</td>
<td>2000</td>
<td>2075</td>
<td>63.6735</td>
<td>-74.9689</td>
<td>106.3</td>
<td>-0.705</td>
<td>*</td>
</tr>
</tbody>
</table>

Sum of Residuals: 0
Sum of Squared Residuals: 122854
Predicted Residual SS (PRESS): 572804