**SCM 651 Business Analytics**

**Final Exam**

**Spring 2017**

**Academic Integrity: This is an individual exam and must represent your own work. Any sharing of information with other people is not allowed. Evidence of collaboration will result in a grade of zero.**

**Instructions:**

1. **Your camera must be turned on at all times, not paused.**
2. **Submit a copy of your completed final to Course page: Assignments and Deliverables: Final Exam: Submission, then upload the exam**
3. **Also send a copy via email to lflee100@syr.edu**

**Total possible points: 100**

**Part 1: Concepts – Short Answer (10 questions - 30 points)**

**Part 2: Tools – Multiple Choice (8 questions - 16 points)**

**Part 3: Techniques – Multiple Choice (6 questions - 12 points)**

**Part 4: Regression Assumptions – Multiple Choice (3 questions - 6 points)**

**Part 5: Interpretation – Short Answer (7 questions - 24 points)**

**Part 6: Business Issues from Articles – Short Answer (4 questions - 12 points)**

**Part 1: Concepts – Short Answer (section total: 30 points)**

1.1: Describe the difference between correlation and linear regression (3 points)

**Correlation is the measurement of the relationship between two or more variables. Linear regression is the fitting of the line between the inputs and the outputs- How an input(s) impacts the output.**

1.2: Describe the difference among slicers, filters and timelines in Excel (3 points)

**Slicers are generally used when looking at pivoted data (pivot tables), and work similar to filters. They are buttons used to query for specific categorical inputs that meet your criteria.**

**Filters are a way to query for specific sets of categorical or numerical data – essentially eliminating all results that don’t meet your query restraints. Filters are available at the column level of a dataset.**

**Timelines are a way of manipulating data in a time series manner in order to see trends/results over a period of time. Probably done easiest through the use of a pivot table and a pivot chart with the date range acting as the column.**

1.3: Describe the difference between dummy variables and moderating effects (3 points)

**Dummy variables are used to measure differences in intercepts between two groups – Prevalent when looking at categorical data. Brand 1 is absent, and acts the baseline. Brand 2 is X distance away from brand 1, etc.**

**Moderating effects (interactions terms) measure the difference in slope between two groups.**

1.4: Describe the differences between logit and probit (3 points)

**Logit, or logistic regression, uses the logistic distribution and is more sensitive at extreme values.**

**Probit uses the normal distribution and is more sensitive at values of variables near their means (averages).**

**They each use different formulas to derive probabilities.**

1.5: What is the difference between the T-statistic and its p-value and the F-statistic and its p-value in a linear regression? (3 points)

**T statistic measures the significance of one coefficient, while f statistic measures the significance of the entire equation.**

1.6: Describe an example of a business problem which would require an exponential regression (3 points)

**A business problem that would require exponential regression would generally be something related to sales – For example, book sales over a period of time, much like our homework assignment. Or something deeply impacted by seasonality, such as retail, where we might see sales skyrocket during the fourth quarter of a year as the holiday season approaches. Or perhaps, a more normal example would be the temperature of a shower as time elapses. The steep rise in temperature would be a good example of something need exponential fitting.**

1.7: Describe three ways to highlight data with conditional formatting (3 points)

You can highlight your data in a number of ways, most notably using color coding.

1. **Heat map with one color signifying the greatest values in a set, and the colors gradually decreasing intensity as the values shrink.**
2. **Reciprocal of (1). One color representing the lowest value, and that color increasing as the values get larger.**
3. **Data Scales which show a bar chart of a specific value against the greatest value in the set.**
4. **Icon sets which show general directionality amongst a data set.**

1.8: Identify three types of data formats which can be imported into Microsoft Access (3 points)

**Txt File. Excel File. Access. ODBC database. XML File. PDF**

1.9: What is the purpose of the criteria field in Microsoft Access? (3 points)

**Identification of a subset of data.**

1.10: Describe the difference between Goal Seek and Solver (3 points)

**Goal seek searches for a single goal (breakeven) and allows for one variable to be changed in that search. No constraints allowed.**

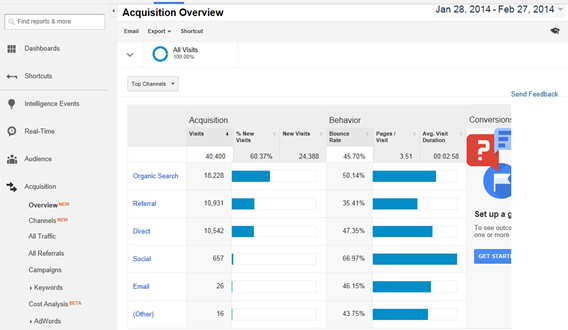
**Solver searches for multiple goals/solutions and allows for the use of constraints, such as demand must be greater than 50,000, etc..**

**Part 2: Tools –Multiple Choice (section total: 16 points)**

**In each of the following problems, a picture is presented from one of the following packages. Identify the package used to produce the result.**

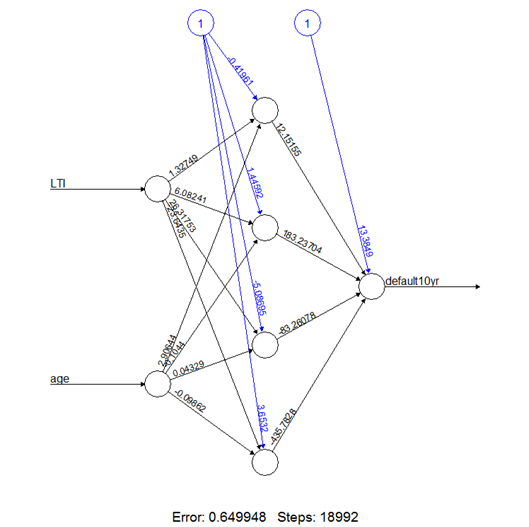
2.1: Is the following result produced by Excel, Access, Google Analytics, R, or Tableau? (2 points)

**Answer: Google Analytics**



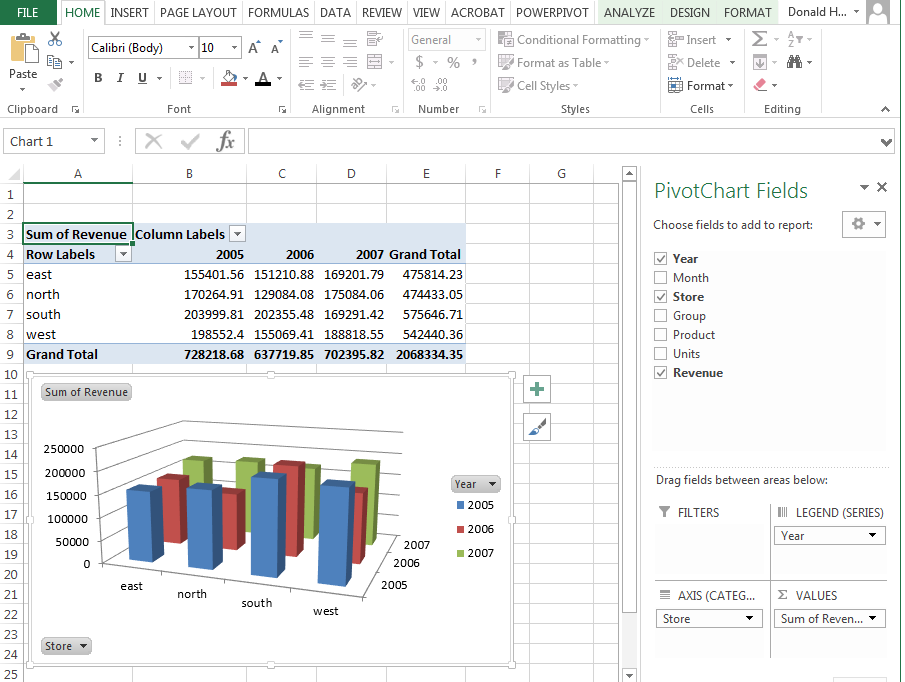
2.2: Is the following result produced by Excel, Access, Google Analytics, R, or Tableau? (2 points)

**Answer: R**



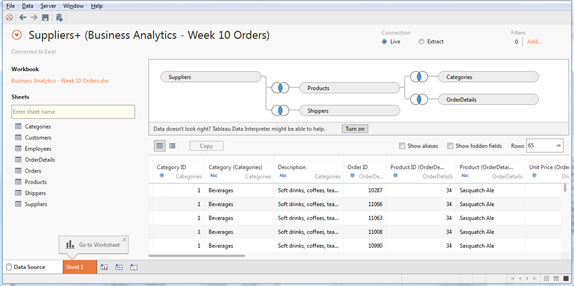
2.3: Is the following result produced by Excel, Access, Google Analytics, R, or Tableau? (2 points)

**Answer: Excel**



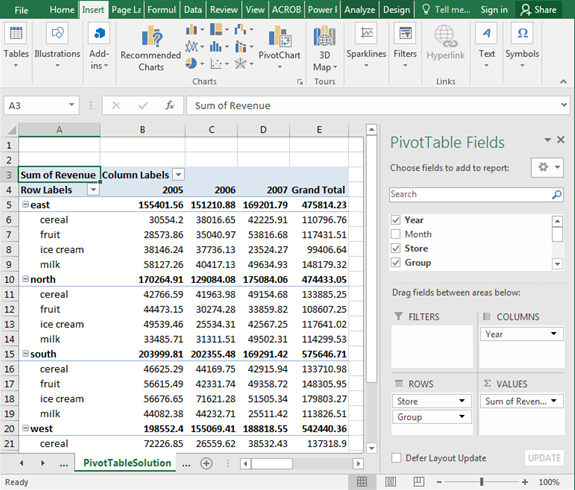
2.4: Is the following result produced by Excel, Access, Google Analytics, R, or Tableau? (2 points)

**Answer: Access**



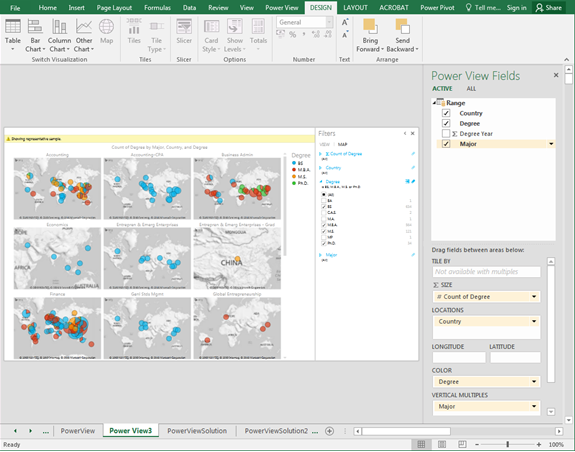
2.5: Is the following result produced by Excel, Access, Google Analytics, R, or Tableau? (2 points)

**Answer: Excel**



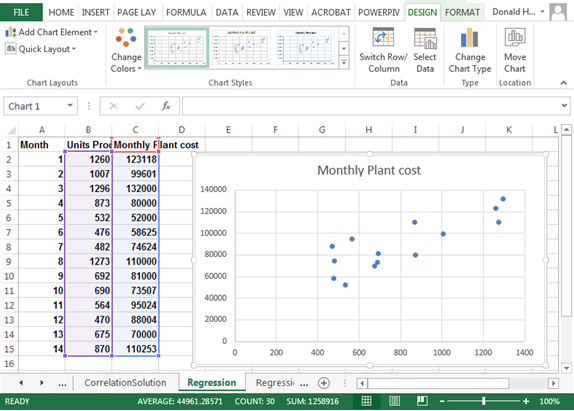
2.6: Is the following result produced by Excel, Access, Google Analytics, R, or Tableau? (2 points)

**Answer: Excel**



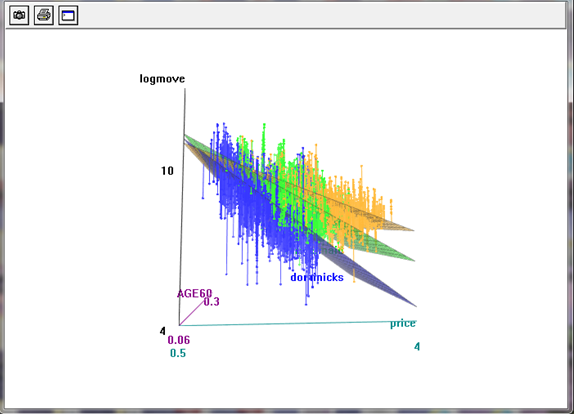
2.7: Is the following result produced by Excel, Access, Google Analytics, R, or Tableau? (2 points)

**Answer: Excel**

****

2.8: Is the following result produced by Excel, Access, Google Analytics, R, or Tableau? (2 points)

**Answer: R**



**Part 3: Techniques – Multiple Choice (section total: 12 points)**

**In each of the following problems, a picture is presented using one of the following techniques. Identify the technique used to produce the result.**

3.1: Does the following result represent linear regression, exponential regression, power regression, moving average, logit, probit, or neural network? (2 points)

**Answer: (multiple) Linear Regression**

Units A

Produced

Σ

1

\* β1

\* β0

Units C

Produced

\* β2

\* β3

Units B

Produced

Monthly

Plant Cost

3.2: Does the following result represent linear regression, exponential regression, power regression, moving average, logit, probit, or neural network? (2 points)

**Answer: Probit**

Ndist(Σ)

P(Loan)

Credit

Card

Age

1

0.011

-9.636

Income

Family

0.043

0.850

0.074

3.3: Does the following result represent linear regression, exponential regression, power regression, moving average, logit, probit, or neural network? (2 points)

**Answer: Linear Regression**

Monthly

Plant Cost

Units

Produced

Σ

1

\* 64.269

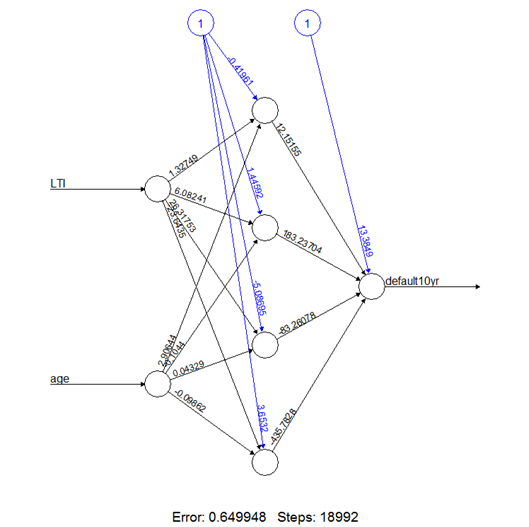
\* 37894

3.4: Does the following result represent linear regression, exponential regression, power regression, moving average, logit, probit, or neural network? (2 points)

**Answer: Moving Average**

3.5: Does the following result represent linear regression, exponential regression, power regression, moving average, logit, probit, or neural network? (2 points)

**Answer: Neural Network**



3.6: Does the following result represent linear regression, exponential regression, power regression, moving average, logit, probit, or neural network? (2 points)

**Answer: Exponential Regression**

**Part 4: Regression Assumptions – Multiple Choice (section total: 6 points)**

**In each of the following problems, a picture or a description of a violation is presented describing a linear regression assumption. Select the best answer.**

4.1: The diagram below show the relationship between X variables. Which regression assumption does this violate: linearity, multi-collinearity, heteroscedasticity, serial correlation, or outliers? (2 points)

**Answer: Multi-collinearity – X variables should not be correlated. Drop or combine variables.**

****

4.2: When the X variables are correlated (multicollinearity), what is the solution (2 points):

a. take the logarithm, square, inverse, or other transformation

**b. drop or combine variables**

c. rho differencing

d. drop data points

Answer:  **b. drop or combine variables**

4.3: When the error terms (residuals) do not have constant variance (heteroscedasticity), what is the solution (2 points):

**a. take the logarithm, square, inverse, or other transformation**

b. drop or combine variables

c. rho differencing

d. drop data points

Answer:  **a. take the logarithm, square, inverse, or other transformation**

**Part 5: Interpretation – Short Answer (section total: 24 points)**

**In each of the following problems, provide a short answer to the question.**

5.1: The regression results below represent the cost of factory production of automobile headlights for Toyota cars at one factory. Answer the following questions. (12 points)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Regression Statistics* | |  |  |  |  |
| Multiple R | 0.749591807 |  |  |  |  |
| R Square | 0.661887878 |  |  |  |  |
| Adjusted R Square | 0.474265453 |  |  |  |  |
| Standard Error | 42861.35168 |  |  |  |  |
| Observations | 19 |  |  |  |  |
|  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |
| Regression | 3 | 35341695233 | 11780565078 | 6.412603636 | 0.005203127 |
| Residual | 15 | 27556432019 | 1837095468 |  |  |
| Total | 18 | 62898127252 |  |  |  |
|  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* |  |
| Intercept | 81275.15 | 77244.49507 | 13.99808684 | 5.13021E-10 |  |
| Corolla | 9.2850886 | 13.03175458 | 1.479853575 | 0.09607524 |  |
| Camry | 18.72528603 | 7.689781205 | 2.695172396 | 0.01661886 |  |
| Avalon | 23.23260325 | 32.71775573 | 1.015736028 | 0.25859682 |  |

5.1.1: What does the R2 mean in this specific example? (2 points)

**How much of the variance in the cost of factory production is related to the different models’ headlights. 66% of that variance is accounted for by the defined variables in this equation.**

5.1.2: What does the F statistic mean in this specific example? (2 points)

**The F Statistic measures the significance of the entire equation. This means we should be highly confident in the soundness of our particular equation. Significance F = .005 falls well below an arbitrary alpha of 0.05.**

5.1.3: Interpret the intercept and all coefficients. What do they mean in this specific example? (8 points)

**Y Intercept = $81275.15 is a fixed cost. Even if we don’t produce any parts at all, our cost will never go below this amount.**

**X1 – Coefficient. 9.2850886. For every output of Corolla Headlights, our cost increases by $9.28**

**X2- Coefficient. 18.72528603. For every output of Camry Headlights, our cost increases by $18.73**

**X3- Coefficient. 23.23260325. For every output of Avalon Headlights, our cost increases by $23.23.**

**Y= 81275.15 + 9.28(x1) + 18.72(x2) + 23.23(x3)**

5.2: The results below represent the correlation matrix of the number of vehicles manufactured for Corolla, Camry, Avalon automobiles and the overall production cost of the factory. Answer the following questions. (12 points)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *Cost* | *Corolla* | *Camry* | *Avalon* |
| Cost | 1 |  |  |  |
| Corolla | 0.647197343 | 1 |  |  |
| Camry | 0.600006014 | 0.251683223 | 1 |  |
| Avalon | 0.353741011 | -0.097464383 | 0.105464325 | 1 |

5.2.1 Which two cars (production quantities) have the strongest correlation to each other (largest magnitude)? What is the correlation? (3 points)

**Corolla and Camry have the strongest correlation to eachother, as defined above, at 25%. In terms of cost, Corolla has the strongest relationship with cost.**

5.2.2 What is a business (intuitive) explanation of the strongest correlation in this specific example? (3 points)

**If we look at model to model, our relation examined above between corolla and camry suggests a weak positive correlation between the production outputs of these two models. This could be due to increased marketing for both of these makes, or perhaps a trickle down effect in brand loyalty- Or maybe even moreso that the cars look relatively similar to one another.**

5.2.3 Which two cars (production quantities) have the negative correlation? (3 points)

**Corolla and Avalon, at -10%. As the Corolla is produced more, the Avalon is produced less. This could be due to constraints in budget, or unit output per model.**

5.2.4 What is a business (intuitive) explanation of the negative correlation in this specific example? (3 points)

**As the Corolla is produced more, the Avalon is produced less, as shown by the weak negative correlation of -10%. This could be due to constraints in budget, or unit output per model. Perhaps some of the money that was to go towards Avalon production is shifted due to increased sales of the Corolla. Also could have something to do with marketing costs as they relate to vehicle production.**

**Part 6: Business Issues from Articles – Short Answer (section total: 12 points)**

**In each of the following problems, provide a short answer to the question.**

6.1: How are companies using analytics? (Answer just three of the following) (3 points)

* Banks

**Banks are using analytics to detect and deter fraud.**

* Manufacturing

**Production planning, inventory management**

* Retail

**Product recommendations (amazon), Marketing.**

* Pharmaceuticals

**Getting drugs to the market more quickly.**

6.2: What are some business applications of web analytics? (3 points)

**Understanding the user journey/preferences. Knowing which landing page to send a user to. Knowing which products/promotions someone might be interested in. Monetizing the use of that data to big retailers (IE the fashion app article we read).**

6.3: What are some barriers to analytics in sports? (3 points)

**Difficulty in collecting data – some sports don’t have sensory movements tracked.**

**Coaches rely on experience/insight rather than data.**

**Hesitancy to change.**

**Not wanting to make the games seem robotic. Keeping the human element.**

6.4: Why did Caesars use analytics? (3 points)

**Creating a customer experience. Understanding the consumer buying behavior cycle, and truly understanding their users’ preferences. Marketing for actions, not just preferences. Gaming analytics (wins/losses), rev management, finance, marketing, labor, etc..**