**SCM 651 Business Analytics**

**Final Exam**

**Winter2020**

**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 (11 questions - 30 points)**

**Part 2: Tools – Multiple Choice (16 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 (6 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. Describe two similarities and two differences between logit and neural networks (4 points)

Answer:

\* Logit assumes that Y is a probability ranging from zero to one.

\* the regression determines the likelihood of something.

\* They are both used to predict using different parameters.

\* They both us coefficients as a representation of the prediction as a linear representation.

\* The neural networks offer more depth in the understanding of the prediction.

1. Describe one similarity and all of the differences between goal seek and solver (4 points)

Answer:

They both help determine the value needed in order to get the value desired. Goal seek only works for the input for one cell. On the other hand, solver can check the optimal result of many dependencies and constraints.

1. Identify three types of files which Tableau can connect to (3 points)

Answer:

\* Excel

\* Text

\* Database

1. In Google Analytics, identify three channels. (3 points)

Answer:

\* Referral

\* Email

\* Social

1. What are three types of constraints in Solver? (3 points)

Answer:

\*More Than, less than, equal to

1. Describe the similarities and differences between correlation and linear regression (3 points)

Answer:

\* Correlation and Linear regression are both based on a linear approach

\* If the correlation is negative, then the slope is negative and same if positive.

\* correlation explains how much of Y is explained by X which is can also be visualized in the linear regression.

\*Correlation is one value, while the regression is an equation.

1. A neural network might not find the optimal solution. Why? (2 points)

Answer:

\* Because the neural network is based on the training data. In order to find the optimal solution, the neural network would need a huge amount of data. Also, the neural network is trained iteratively, so sometimes it might not reach the optimal solution if it is stopped prematurely.

1. Describe a business problem that would require a moderating effect (2 points)

Answer:

\* If we are trying to figure out whether to give a loan to someone or not. We have different variables like income and how big their family size is. By using the moderating effect, we would be able to tell the interaction between the two variables to see if we get any deeper knowledge of the data.

1. How do you identify dirty data in Microsoft Access? (2 points)

Answer:

\* We can write a query to find empty values. Or we could use a left join in order to get rid of values that do not have a matching value in another column.

1. Describe the business explanation of the sensitivity analysis in the Titanic survivor problem (2 points)

Answer:

\* In the titanic survivor problem, we wanted to know whether a person was going to survive or not. Some of the parameters that were included was the age, class and the gender of the person. We wanted to figure out what effects do each of the parameters would have on the people surviving. Lets say if you were a woman, then you were given access to the emergency boats before men. This are the things we are trying to find out.

1. In Google Analytics, what is the difference between Recency and Frequency? (2 points)

Answer:

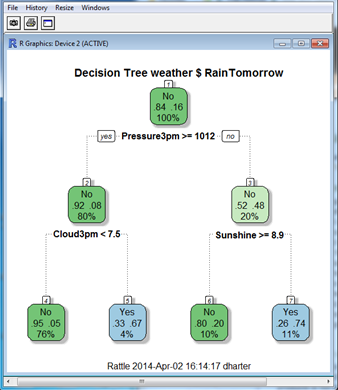
\* Frequency refers to the number of times a visitor returns while Recency refers to how long it took a user to come back to the webpage.

**Part 2: Tools – Multiple Choice (section total: 16 points; questions 1 point each)**

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

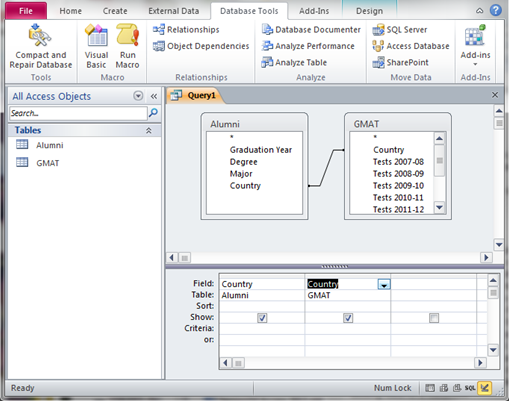
1. Is the following result produced by Excel, Access, Google Analytics, R, or Tableau? (1 point)

Answer: R

****

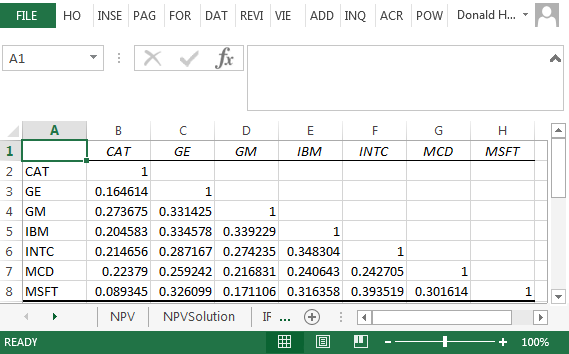
1. Is the following result produced by Excel, Access, Google Analytics, R, or Tableau? (1 point)

Answer:Access

****

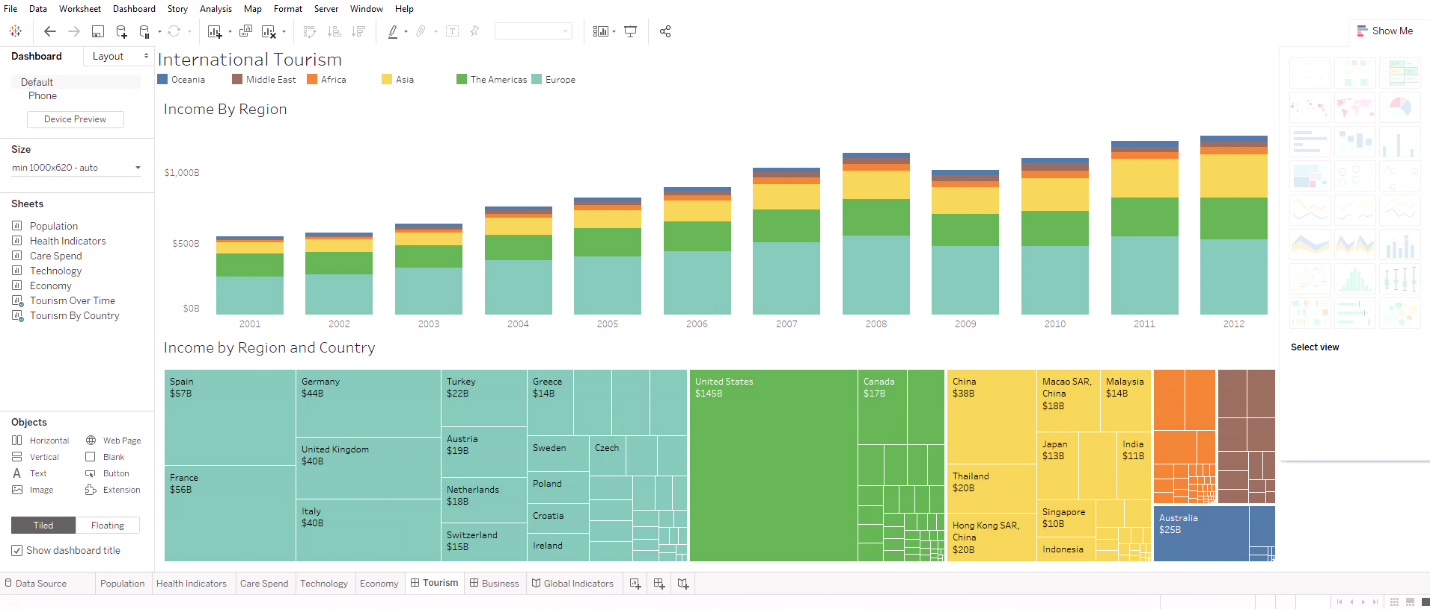
1. Is the following result produced by Excel, Access, Google Analytics, R, or Tableau? (1 point)

Answer:Excel

****

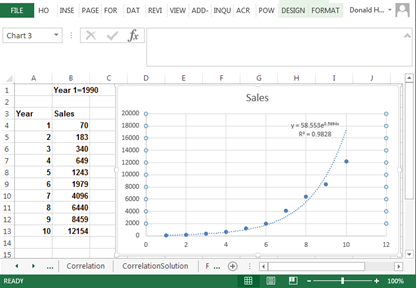
1. Is the following result produced by Excel, Access, Google Analytics, R, or Tableau? (1 point)

Answer:Tableu

****

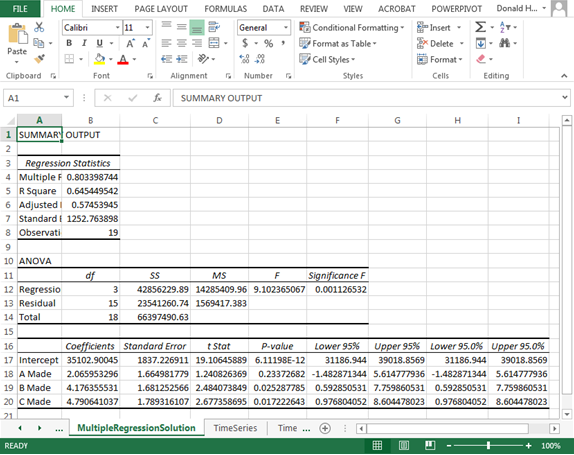
1. Is the following result produced by Excel, Access, Google Analytics, R, or Tableau? (1 point)

Answer:Excel

****

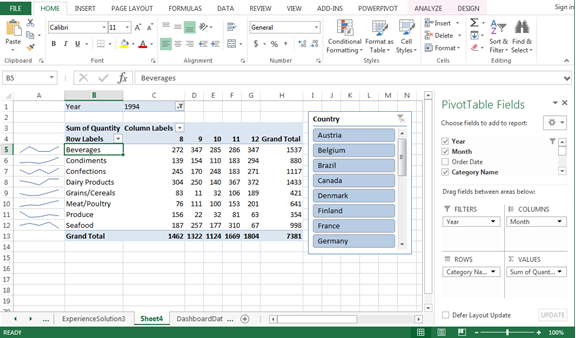
1. Is the following result produced by Excel, Access, Google Analytics, R, or Tableau? (1 point)

Answer:Excel

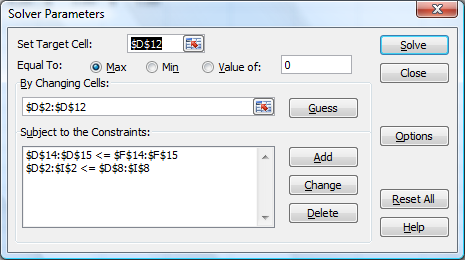
****

1. Is the following result produced by Excel, Access, Google Analytics, R, or Tableau? (1 point)

Answer:Excel

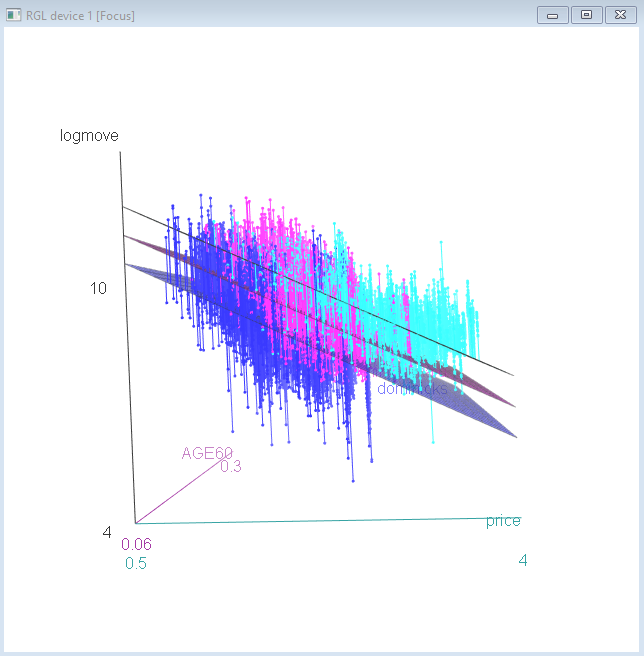
****

1. Is the following result produced by Excel, Access, Google Analytics, R, or Tableau? (1 point)

Answer:Excel

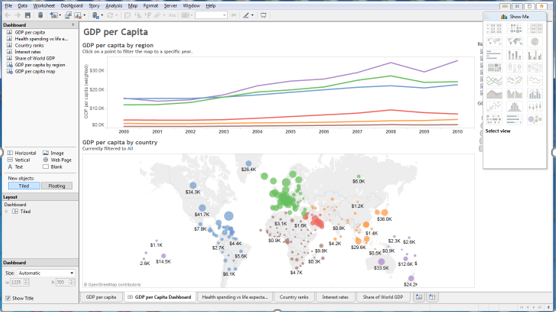
1. Is the following result produced by Excel, Access, Google Analytics, R, or Tableau? (1 point)

Answer:R



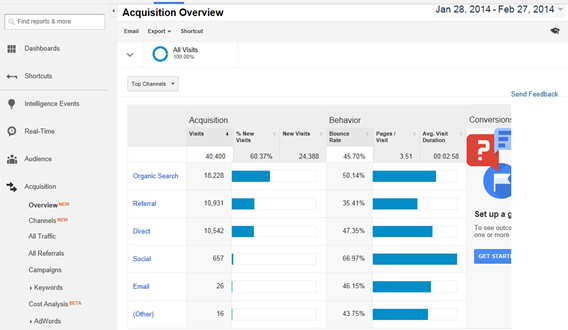
1. Is the following result produced by Excel, Access, Google Analytics, R, or Tableau? (1 point)

Answer:Tableu

****

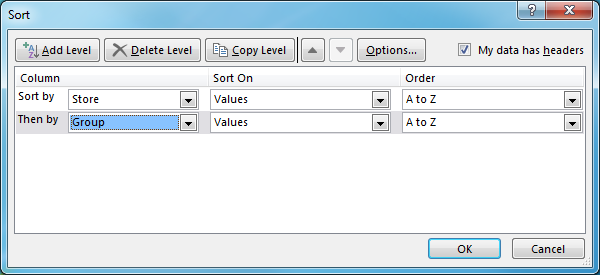
1. Is the following result produced by Excel, Access, Google Analytics, R, or Tableau? (1 point)

Answer:Google Analytics



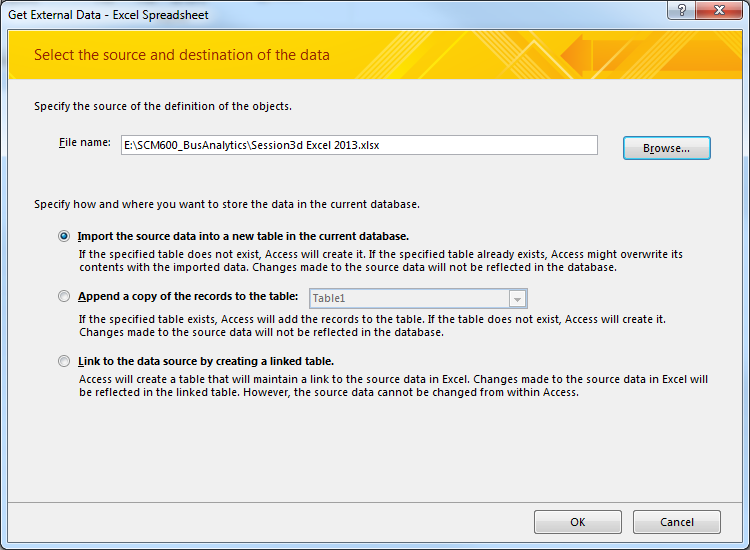
1. Is the following result produced by Excel, Access, Google Analytics, R, or Tableau? (1 point)

Answer:Excel



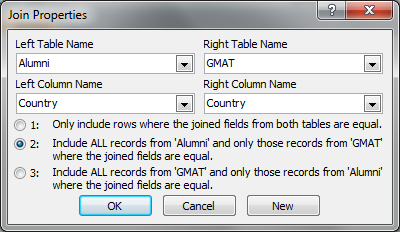
1. Is the following result produced by Excel, Access, Google Analytics, R, or Tableau? (1 point)

Answer:Access



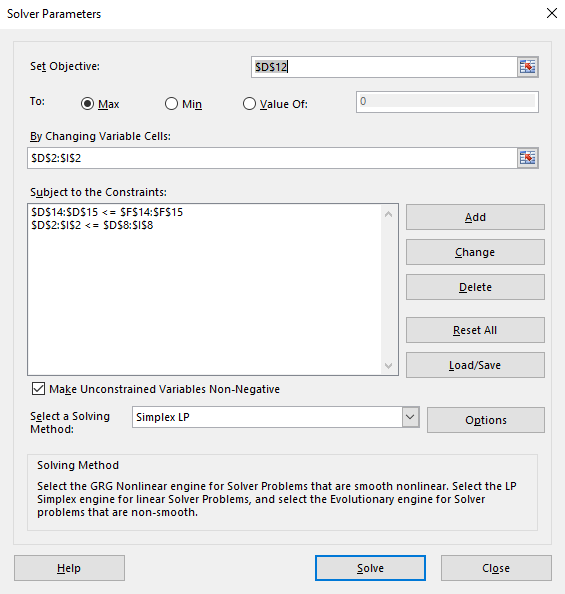
1. Is the following result produced by Excel, Access, Google Analytics, R, or Tableau? (1 point)

Answer: Access



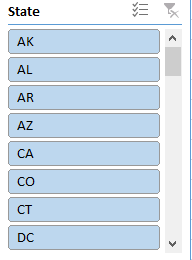
1. Is the following result produced by Excel, Access, Google Analytics, R, or Tableau? (1 point)

Answer:Excel



1. Is the following result produced by Excel, Access, Google Analytics, R, or Tableau? (1 point)

Answer:Tableu

****

**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.**

1. Does the following result represent linear regression, exponential regression, power regression, moving average, logit, ordered logit, multinomial logit, k-means clustering, conjoint analysis, or neural network? (2 points)

Answer: linear regression

Units A

Produced

Σ

1

\* β1

\* β0

Units C

Produced

\* β2

\* β3

Units B

Produced

Monthly

Plant Cost

1. Does the following result represent linear regression, exponential regression, power regression, moving average, logit, ordered logit, multinomial logit, k-means clustering, conjoint analysis, or neural network? (2 points)

Answer:Multinomial logit

P(Loan)

exp(Σ)

(1+exp(Σ))

Credit

Card

Age

1

0.011

-9.636

Income

Family

0.043

0.850

0.074

1. Does the following result represent linear regression, exponential regression, power regression, moving average, logit, ordered logit, multinomial logit, k-means clustering, conjoint analysis, or neural network? (2 points)

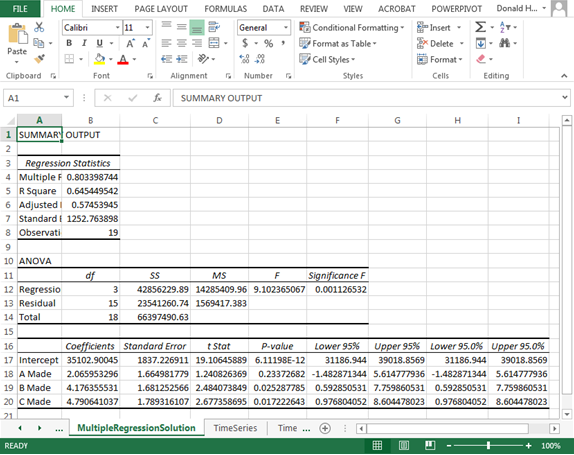
Answer: Exponential Regression

1. Does the following result represent linear regression, exponential regression, power regression, moving average, logit, ordered logit, multinomial logit, k-means clustering, conjoint analysis, or neural network? (2 points)

Answer: Moving Average

1. When you have quarterly seasonality in data, which technique would you use: linear regression, exponential regression, power regression, moving average, logit, ordered logit, multinomial logit, k-means clustering, conjoint analysis, or neural network? (2 points)

Answer: Moving Average

****

1. Does the following result represent linear regression, exponential regression, power regression, moving average, logit, ordered logit, multinomial logit, k-means clustering, conjoint analysis, or neural network? (2 points)

Answer: Power regression

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

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

1. Describe how to correct your regression model when the multi-collinearity assumption is violated. (2 points)

Answer:

\* Combine variables or drop one

1. When the linearity assumption of a regression is violated, what is the solution (2 points):

Answer:

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

b. drop or combine variables

c. rho differencing

d. drop data points

1. When the error terms (residuals) are correlated (serial correlation), what is the solution (2 points):

Answer:

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

b. drop or combine variables

c. rho differencing

d. drop data points

**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 | 0.561887878 |  |  |  |  |
| 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 | 1081275.15 | 77244.49507 | 13.99808684 | 5.13021E-10 |  |
| Corolla | 19.2850886 | 13.03175458 | 1.479853575 | 0.015960754 |  |
| Camry | 20.72528603 | 7.689781205 | 2.695172396 | 0.001661886 |  |
| Avalon | 33.23260325 | 32.71775573 | 1.015736028 | 0.325859682 |  |

* + 1. What does the R2 mean in this specific example? (2 points)

Answer:

It means that 56% of the cost of the Toyota cars at a factory is explained by the Corolla, Camry and Avalon variables.

* + 1. What does the significance of F mean in this specific example? (2 points)

Answer:

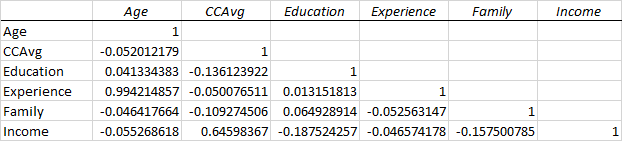
\* It means that the probability of error is less than .0052. Therefore, the regression has a .52% chance of being erroneous.

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

Answer:

It means that even if you are not producing a car, then the fixed cost is going to be at least 1081275.15. Therefore, you would want to produce a quantity of cars that would profit much bigger than this fixed cost.

**5.2** The results below are the correlations of variables which affect the likelihood of taking out a loan. (6 points)



* + 1. Which two variables would be the most likely to cause multi-collinearity if included together in a linear regression? Why? (3 points)

Answer: Experience and Age since they are very well correlated. I.e., one of the variables can explain the other. Therefore, there is no need of using both variables in the linear regression, so I would recommend to drop one.

* + 1. Which two variables would be the least likely to cause multi-collinearity if included together in a linear regression? (3 points)

Answer: I would say education and experience since they have the smallest correlation factor. Therefore, they would bring unique insights into the model instead of repeating the same information.

**5.3.1** The results below represent the logit regression on the probability of taking out a loan. (6 points)

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -11.9761754 1.5197691 -7.880 3.27e-15 \*\*\*

Age -0.0709939 0.0569870 -1.246 0.212841

Education 1.6895305 0.1065229 15.861 < 2e-16 \*\*\*

Experience 0.0831009 0.0566942 1.466 0.142710

Family 0.6869306 0.0687963 9.985 < 2e-16 \*\*\*

Income 0.0539414 0.0023995 22.480 < 2e-16 \*\*\*

Mortgage 0.0006643 0.0005226 1.271 0.203658

Interpret all coefficients in business terms in this specific example. What is the effect of each variable on the probability of taking out a loan? (6 points)

Answer:

Intercept: If you have none of the above, the likeliness of you getting a loan is very low. Therefore, you need to gain as much of the other significant variables as possible in order to get a loan.

Age: Of not very high importance, and the older you get the least likely you are to get the loan.

Education: Very important variable. The more education you get, the more likely you are to get the loan.

Experience: Very little contribution to the likeliness of you getting a loan, but it will likely increase your chances.

Family: The more family members the better chance you have of taking out a loan.

Income: Tied for the most significant parameter in the logit regression as it can increase your chances of getting a loan.

Mortgage: Very unlikely to contribute to your chances of getting a loan, but they will increase you changes of getting the loan by a small amount.

**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: Define Analytics. (3 points)

Answer:

Analytics is the use of all kinds of information to gather insight into a process or dataset. Analytics can be broken down into three kinds: predictive, prescriptive, and descriptive. They all serve to either find out what is happening in the data, where the data is heading and how to change the process to you benefit.

6.2: Why are web analytics better than surveys? (3 points)

Answer:

Because in surveys you are only getting the insight from a select few people that consent to their information being used. This also brings in some statistical error because of how the person was feeling that day or because the question was worded in a vague way. Web analytics takes out the human error from the equation. You are able to see exactly what the user wants, how they navigate and where they are likely to go next. This information is more useful to analysts than answers to questions likely to have human error added to them.

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

Answer:

Caesar hotel used analytics to improve their business. It ranges from marketing to finance. Almost all of their departments use analytics to help bring more revenue. Building the website, the gaming floor and even hotel pricing are all products of analytics now.

6.4: How should you deal with non-perfect data? (3 points)

Answer:

In short, you should extract the most information possible from the non-perfect dataset to make it as perfect as possible. The thing about non-perfect data, is that it is also data as well. Having an empty value can actually be a pattern in the data. On the other hand, you can also fill in the values or replaced them by the mean or median. Other two popular ways is either by interpolation or extrapolation which means filling in the outside or inside numbers using the current population. In the end, one will most likely spend most of the time in analytics trying to make sense of dirty data, but is our job to clean it up as much as possible.