**SCM 651 Business Analytics**

**Final Exam**

**Winter 2021 - B**

**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 dharter@syr.edu**

**Total possible points: 100**

**Part 1: Concepts – Short Answer (12 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)**

**4 point questions (total 8 points)**

1. Describe the similarities and differences between logit and neural networks (3 points)

Answer:

Similarities

• Logit - Relies on the logistic distribution/function.

o Is more accurate at and sensitive to extreme values. (very low or high incomes)

• Neural networks use (rely on) the logistic function to represent nonlinear behavior and hidden nodes

Neural networks work more in the way the human brain works

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

Answer:

Similarities

Alter a variable to achieve a specific objective

Differences

• Goal Seek: alter one variable to determine which value achieves a specific objective (e.g., how to maximize profit, how to minimize costs, product price)

• Solver (unconstrained): alter multiple variables to determine which combination results in a maximum or minimum result. Unconstrained optimization has no limits on the possibilities of the variables

• Solver (constrained): solve objective by altering multiple variables, subject to constraints

* Relies on linear, nonlinear, or integer programming techniques to find the solution, depending on the problem. (needs to be linear)

The goal seek command searches for a solution by varying a parameter until a solution is found

Solver has the ability to search for an optimal solution subject to constraint. In contrast, goal seek would search for a solution that would match a specific value, such as demand-supply=0.

Solver will find the maximum or minimum of a function, called the optimal feasible solution, if one exists. In some cases, particularly overly constrained problems, there will be no solution.

1. Identify three calculations which can be performed in a Microsoft Access query (3 points)

Answer:

Sums, averages, counts

1. What does ANOVA do? Describe a business example that could use ANOVA. (4 points)

Answer:

• Analysis of variance (ANOVA) is used to compare the averages for different categories or brands.

• ANOVA calculates the average of each item, combines with standard deviation/variance, and determines if they are statistically different from one another.

Example: Product placement

1. Describe the difference between dummy variables and moderating effects (3 points)

Answer:

* **Dummy variables** change the intercept of some items to be different from others – “we might actually change the intercept for some of our data”
* This allows you to look at price differentiation in different product markets.
* **Moderating effect (also called an interaction):** An interaction of variables that leverage each other
  + The moderating effect acts as a catalyst to accelerate/increase the effect of certain variables.
* Dummy variables can only change the intercepts.
* Moderating effects allow the slope of the line to change.

1. Describe the difference between correlation and linear regression (2 points)

Answer:

Correlation is measure of two or more independent variables (x, x1, x2, xn) and how the move with respect to each other.

Linear regression is the relationship between the variables (x, …) and the dependent variable (y). Linear regression depends unpon the independent variables are not serially correlated.

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

Answer:

• Frequency: how frequently people come to our site (oftem)

• Recency: how recently did people come to our site (time)

1. In R, what does a boxplot do? (2 points)

Answer:

* Boxplots identify
  + Maximum and minimum – whiskers
  + Quartiles
    - Top of the box is the 75 percentile
    - Bottom of the Box is the 25 percentile
  + Median – center bar (thick black line in the middle)
    - i.e. median for distribution – prices

1. Describe a business problem that would require a dummy variable (2 points)

Answer:

Orange juice sales variable to see whether one brand has a price premium over another

1. What does the T-statistic and its p-value mean in a linear regression? (2 points)

Answer:

• T-statistic: measures if we have confidence in the coefficient of a variable

o Should have p-value of less than 0.05 (95% confidence)

o if your p-value is less than 0.05 for a coefficient, then you're fairly confident that that coefficient is important and relevant. If your p-value is greater than 0.05, well, you don't have very much confidence in that coefficient itself.

o Associated with it is the p-value

 p-Values

• 90% confidence = 0.10 p-value

• 99% confidence = 0.01 p-value

• 95% confidence = 0.05 p-value

1. In Google Analytics, what is the difference between pages and pages/session? (2 points)

Answer:

* Pages/Session: average number of pages viewed in each session
* Pages or page views - how many pages are viewed by all of those users

1. In R, what does a histogram do? (2 points)

Answer:

* Histograms show frequency of data within intervals or bins
* Bars represent bins which are for different ranges of values, such as age groups

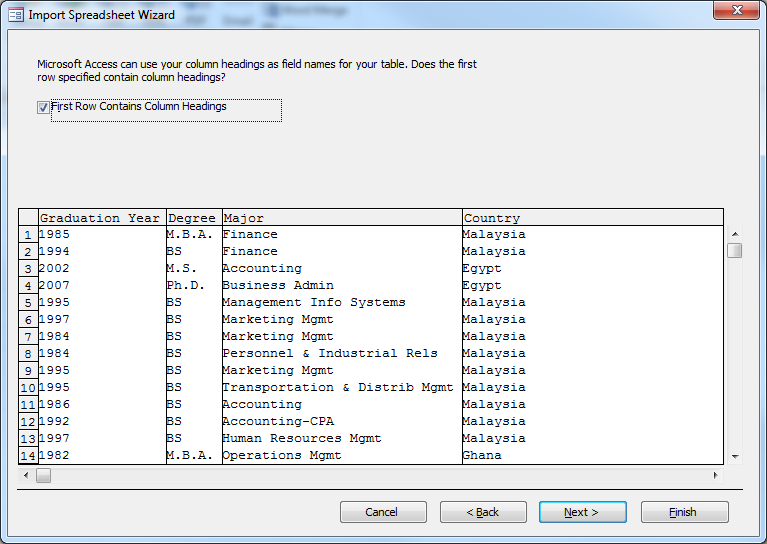
**Part 2: Tools – Multiple Choice (section total: 16 points; questions 1 point each)**

**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:

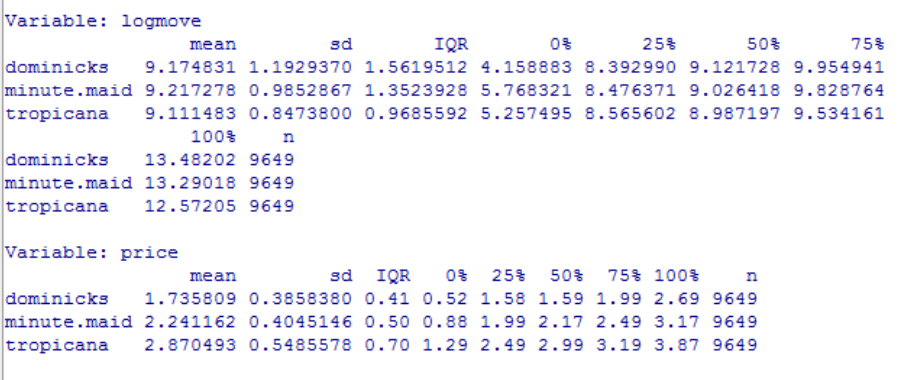
Access



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

Answer:

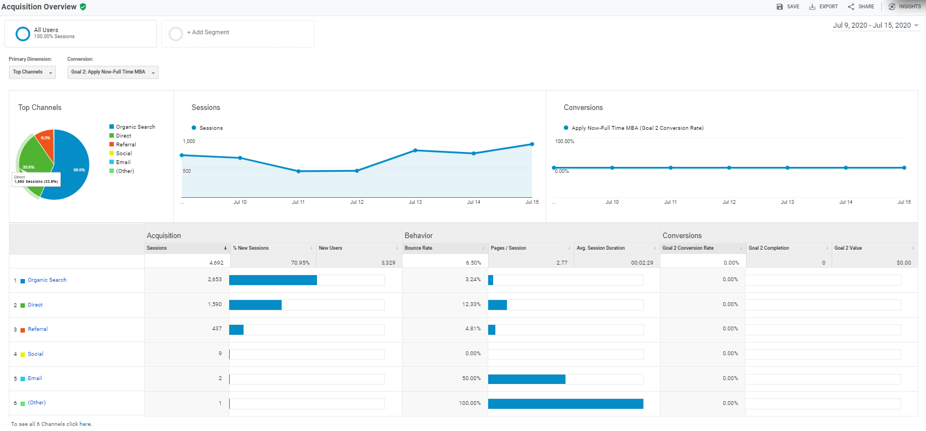
R



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

Answer:

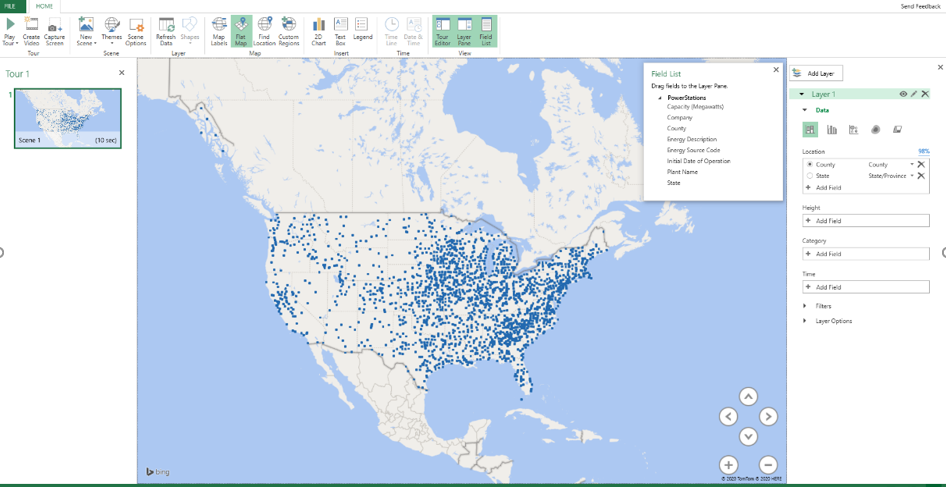
Google analytics



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

Answer:

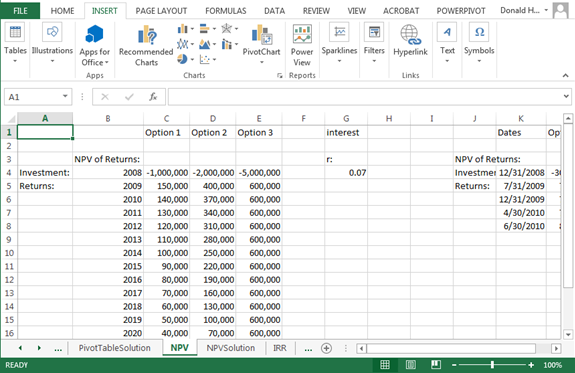
Excel 3D Maps



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

Answer:

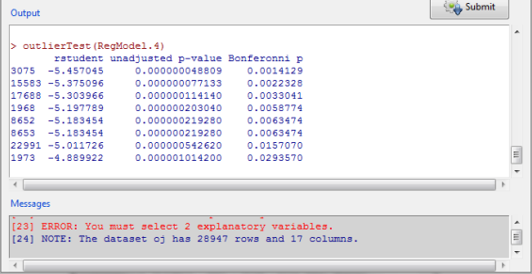
Excel



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

Answer:

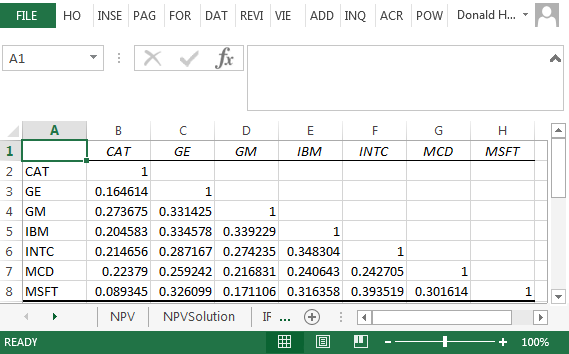
R

****

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

Answer:

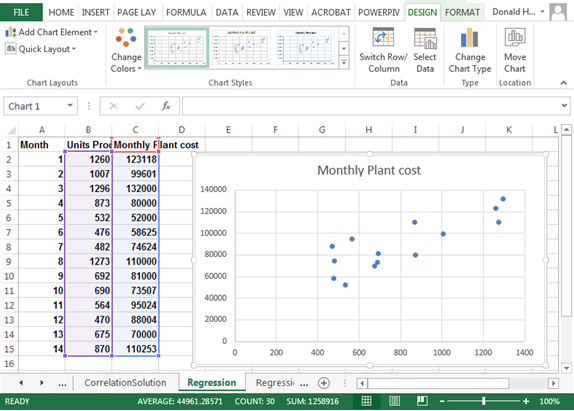
Excel

****

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

Answer:

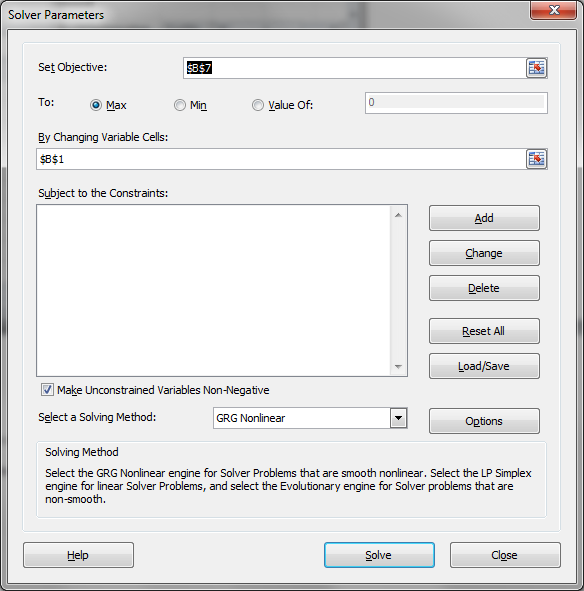
Excel

****

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

Answer:

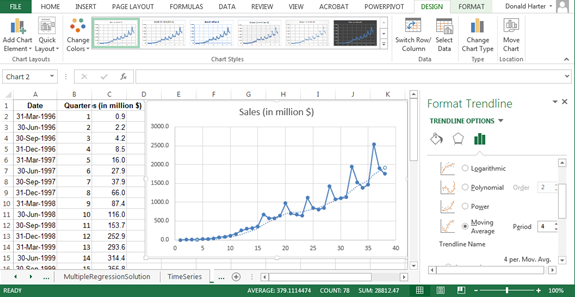
Excel



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

Answer:

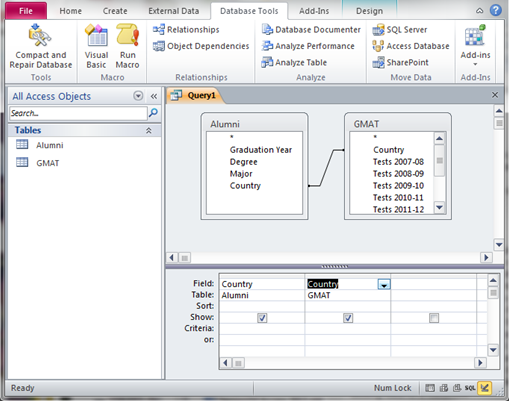
Excel

****

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

Answer:

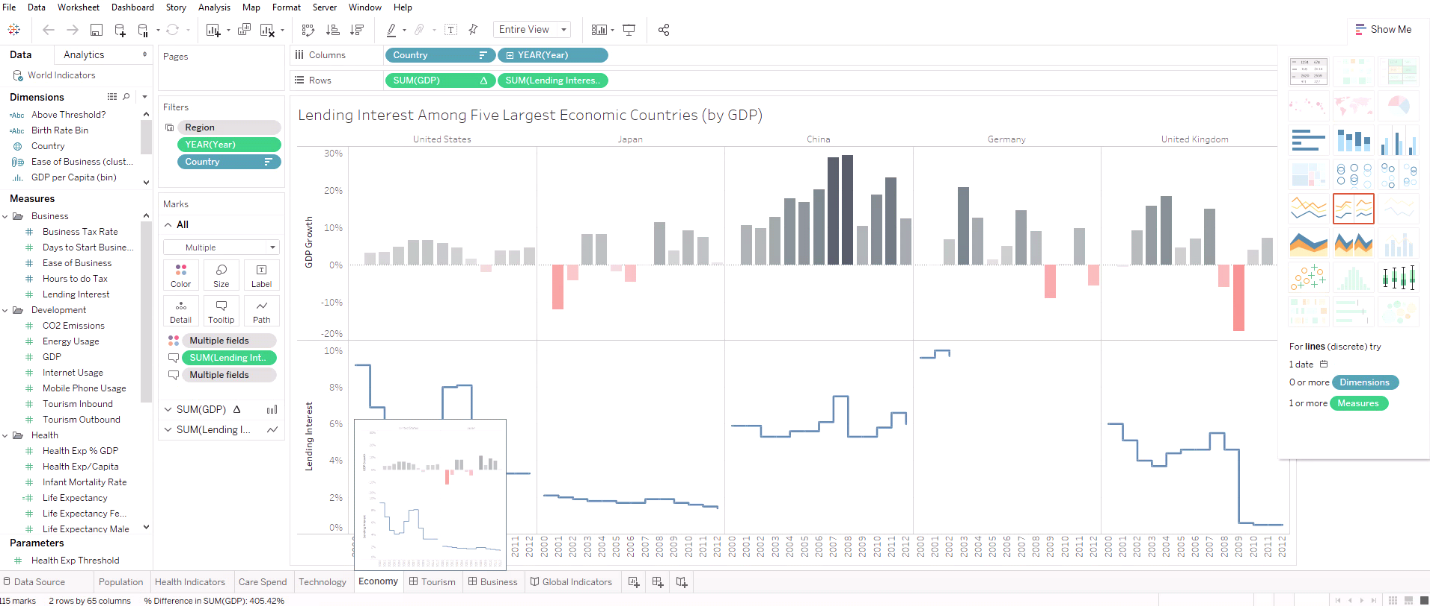
Access

****

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

Answer:

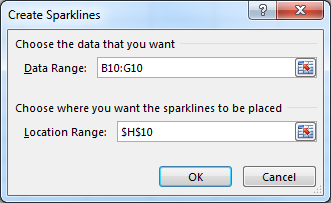
Tableau

****

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

Answer:

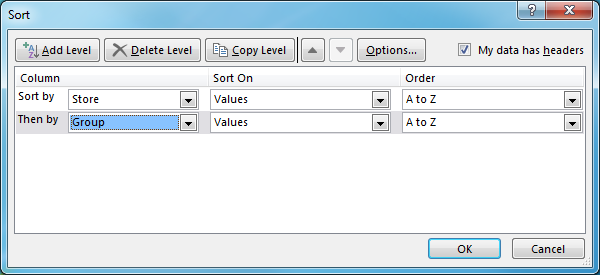
Excel



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

Answer:

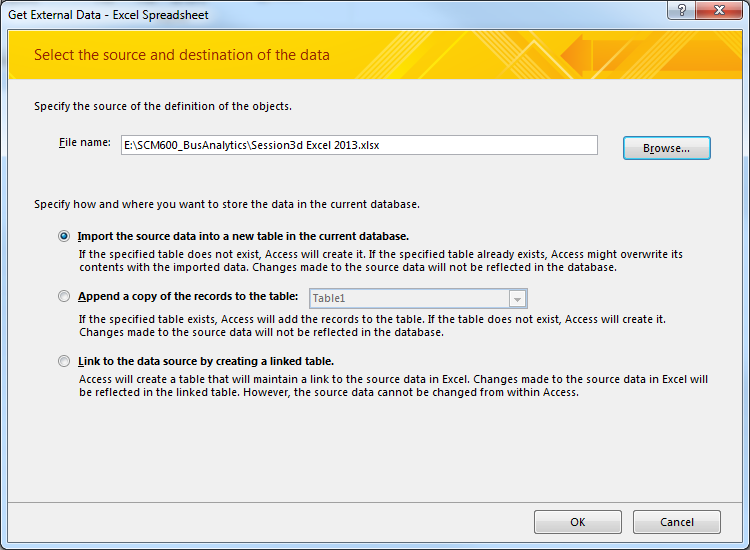
Excel



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

Answer:

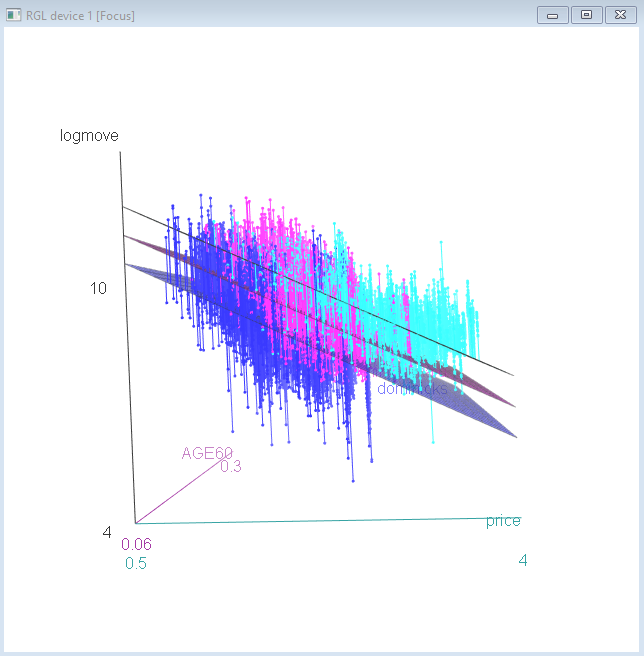
Access



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

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.**

1. Does the following result represent linear regression, exponential regression, power regression, moving average, logit, correlation, 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, correlation, or neural network? (2 points)

Answer:

Logit

P(Loan)

Ndist(Σ)

Credit

Card

Age

1

0.011

-9.636

Income

Family

0.043

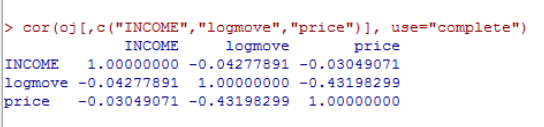
0.850

0.074

1. Does the following result represent linear regression, exponential regression, power regression, moving average, logit, correlation, or neural network? (2 points)

Answer:

correlation



1. Does the following result represent linear regression, exponential regression, power regression, moving average, logit, correlation, or neural network? (2 points)

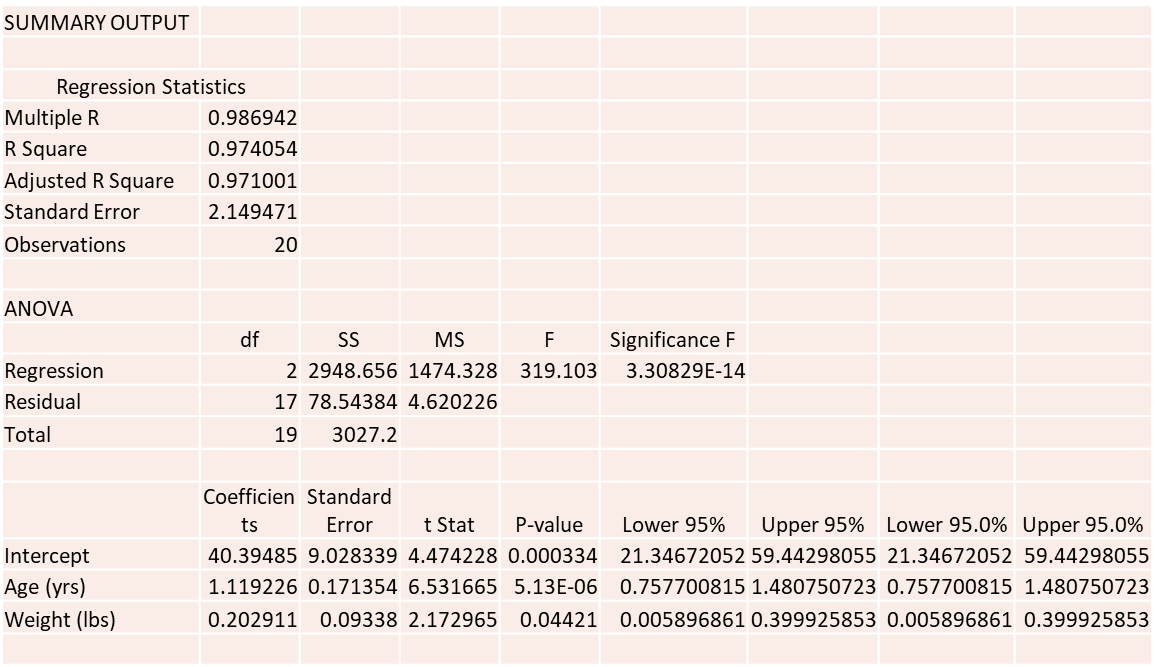
Answer:

power regression

1. Does the following result represent linear regression, exponential regression, power regression, moving average, logit, correlation, or neural network? (2 points)

Answer:

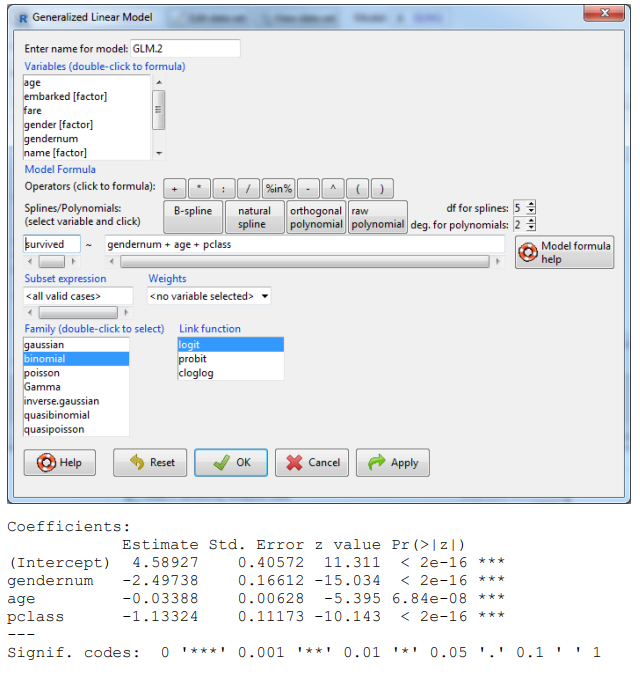
linear regression



1. Does the following result represent linear regression, exponential regression, power regression, moving average, logit, correlation, or neural network? (2 points)

Answer:

Logit



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

**In each of the following problems, either a picture or a question is presented describing a linear regression assumption. Select the best answer or respond to the question provided.**

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

Answer:

you change the type of data so that it is no longer violating the linear assumption—by transforming it using or taking the logarithm or square or square root or other transformation

This should make it so that the linearity assumption is satisfied and, at the same time, heteroscedasticity disappears.

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

Answer:

outliers

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

Answer:

a

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: 20 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 | 0.159607524 |  |
| Camry | 20.72528603 | 7.689781205 | 2.695172396 | 0.01661886 |  |
| Avalon | 33.23260325 | 32.71775573 | 1.015736028 | 0.325859682 |  |

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

Answer:

.562 of the result can be attributed to the variables provided.

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

Answer:

The accuracy of the regression

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

Answer:

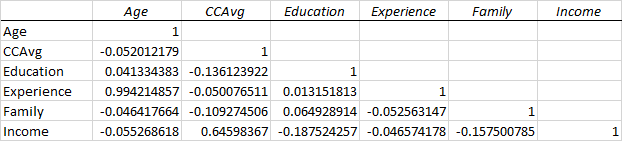
Intercept 1081275.15

Coefficients

|  |  |
| --- | --- |
| Corolla | 19.2850886 |
| Camry | 20.72528603 |
| Avalon | 33.23260325 |

The cost of production = 1081275.15 + 19.2850886\* (#Corollas) + 20.72528603\*(#Camrys) + 33.23260325\* (#Avalons)

**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? What is their correlation? (3 points)

Answer:

Experience and Age. Highest to |1| on correlation scale. The x variables must not be correlated

1. Which two variables would be the least likely to cause multi-collinearity if included together in a linear regression? Why? What is their correlation? (3 points)

Answer:

Education and Experience closest to 0 on correlation scale

**5.3** The results below represent a linear regression with WeeklyVolume as the dependent variable and Age60 through WorkWomen as independent variables. (6 points)

Coefficients:

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

(Intercept) -6163.0 2368.5 -2.602 0.01119 \*

Age60 970.3 1152.7 0.842 0.40263

Education -731.9 271.6 -2.695 0.00870 \*\*

Ethnicity 154.9 156.4 0.990 0.32516

Income 523.9 188.6 2.779 0.00692 \*\*

Mortgage -281.9 225.2 -1.251 0.21471

Poverty 920.8 1413.2 0.652 0.51671

Retired -428.5 1693.3 -0.253 0.80091

Single -192.2 495.5 -0.388 0.69911

Unemployed 4167.0 2564.7 1.625 0.10847

WorkWomen 1528.6 1187.0 1.288 0.20182

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 123.5 on 74 degrees of freedom

Multiple R-squared: 0.1838, Adjusted R-squared: 0.07351

F-statistic: 1.666 on 10 and 74 DF, p-value: 0.105

Interpret all coefficients in business terms in this specific example. (6 points)

Answer:

Age60 970.3 positively affects model

Education -731.9 negatively affects model and has a .01 significance (the \*\*)

Ethnicity 154.9 Positively affects model

Income 523.9 positively affects modes and has a .01 significance (the \*\*)

Mortgage -281.9 negatively affects the model. And depending upon the scale, may greatly affect the model (is it in the hundreds or thousands)

Poverty 920.8 positively affects the model

Retired -428.5 negatively affects the model (and is likely a binary yes or no answer)

Single -192.2 negatively affects the model (and is likely a binary yes or no answer

Unemployed 4167.0 positively affects the model (and is likely a binary yes or no answer)

WorkWomen 1528.6 positively affects the model

The variables and their coefficients and intercept sum together to calculate WeeklyVolume.

With a low R-squared, all of the variables affect the model but the unknown factors are large

**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:

“a process of transforming data into actions through analysis and insights in the context of organizational decision making and problem solving”

◦ “the use of data, information technology, statistical analysis, quantitative methods, and mathematical or computer-based models to help managers gain improved insight about their business operations and make better, fact-based decisions”

6.2: What is an A/B test and what is its purpose? (3 points)

Answer:

A/B tests change a variable or variables to determine what changes occur.

A/B test are common in understanding user response to site or feature changes, and policy changes that come from different forms (user behavior data, transactional data, and customer service data).

Empirically compare two experiments to determine optimal change.

6.3: Describe or give an example for three of the following (3 points)

a. Data visualization – dashboards and scorecards

b. Text mining - extract structure from unstructured text files

c. Social network analysis - identify networks of calling circles, influencers

d. Contact optimization

Answer:

a. Data visualization – dashboards and scorecards

b. Text mining - extract structure from unstructured text files

c. Social network analysis - identify networks of calling circles, influencers

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

Answer:

Don’t wait for 100% perfect data

◦ Identify and explain data limitations as part of analysis