MSCI:9110

Advanced Analytics

Instructor: Mark Bennett

Spring 2019

**Final Exam**

**Instructions:**

1. This is an open book/note exam.
2. You are allowed to use calculator, printed course materials and ICON.
3. You are not allowed to use any communication tools including cellphone, email, and instance messaging software and so on.
4. You should submit the answer sheet. Electronic files used during this exam can be submitted to the dropbox in ICON if you want.
5. Save your electronic files often! You may want to create multiple copies of the files to avoid changing back and forth.
6. Your solution for each question must be written down on the answer sheet. Otherwise, you will lose some points even if the solutions can be found in your electronic submissions.
7. Maximum score is 60 points.
8. You have two and one quarter hours to finish this exam.

*Good Luck!*

**At the conclusion of the exam, please signify your agreement to the Tippie College of Business Honor Code.**

**“I have neither given nor received assistance on this exam.”**

**Signature: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Printed Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_­­\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Problem 1 (6 points)**

In financial regulation, "politically exposed person" (PEP) is a term describing someone who has been entrusted with a prominent public function. A PEP generally presents a higher risk for potential involvement in bribery and corruption by virtue of their position and the influence that they may hold. In this problem, using the file *pep.jmp,* use appropriate independent variables as inputs to a 20 x 20 two layer Neural model with 5-Fold Cross Validation to predict whether the individual is a PEP. The *positive* case is when the person is a PEP. Report your Traing and Validation Accuracy, Precision, and Recall below with calculation steps included:

Training

Accuracy = **(246+197)/480 = 92.3%** Precision = **197/(197+15) = 92.9%** Recall = **197/(197+22) = 89.9%**

Validation

Accuracy = **(61+50)/120 = 92.5%**  Precision = **50/54= 92.6%** Recall = **50/55 = 90.9%**

**Problem 2 (9 points)**

Each week in 2019 General Motors ships a 2380 automobile supply of three types known as the Lambda Platform (SUVs) to the West Coast by rail through three intermediate points: Council Bluffs, Kansas City, and St. Louis from Detroit. Chicago and Decatur are intermediate points along the way. Suppose the automobiles are transported through the following distribution network with the unit shipping cost given next to each arc. A network optimization model will minimize the shipping cost, which related directly to the distance in miles shipped. It uses the variables to represent the number of automobiles shipped from city to city .

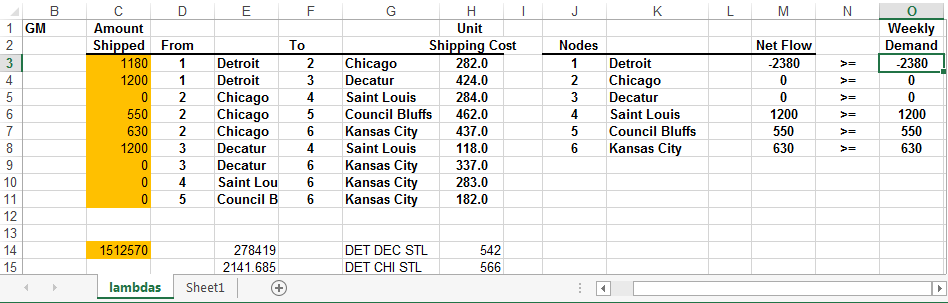
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 282  337  283  462  182  284   |  | | --- | |  | |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  | 424 |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  | 437 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  | 118 |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |

Model this network using the preliminary Excel spreadsheet called *shipSUVsByRail*.*xls*. Setup the dependent spreadsheet cells to match the variables in the network diagram. The flow of traffic begins at Detroit and ends up with demands of 1200 automobiles in Saint Louis, 550 automobiles in Council Bluffs, and 630 in Kansas City. The demand at Chicago and Decatur is 0 as these are intermediate points.

1. (3 points) Determine the overall objective function to minimize. Write that here in math notation using variables as a sum of products of the number of automobile units shipped and shipping cost per unit.
2. (2 points) Write the flow-balance constraint at the point Decatur. Use variables in your formula.
3. (2 points) Write the flow-balance constraint at the point Kansas City. Use variables in your formula.
4. (3 points) Construct the required spread-sheet entry for the objective function in Cell C14, and demand constraint constants in Cells O3:O8. Write or copy those spreadsheet cell values for Cells O3:O8 here below. **C14: sumproduct(C3:C11,H3:H11)**

|  |
| --- |
| **-2380** |
| **0** |
| **0** |
| **1200** |
| **550** |
| **630** |

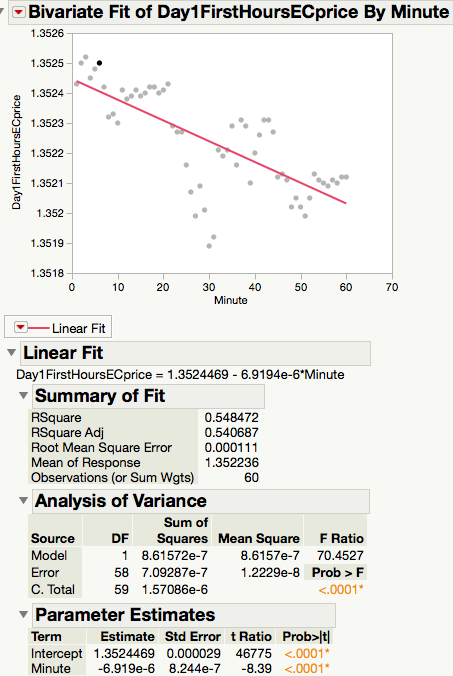
1. (3 points) Solve for the optimal shipping amounts to each city (Cells C3:C11). What is the total cost of shipping these 2380 automobiles? **1,512,570** How many automobiles are shipped from Detroit to Chicago? **1180** Detroit to Decatur? **1200**



**Problem 3 (12 points, 2 points for each)**

Your team trades currencies for a global manufacturing company in order to get the best prices in the E.U. market. The Euro Currency (EC) is quoted through the month on a continuous basis. A dataset with one month of the price of the EC in U.S. Dollars for October 2013 for each minute is in the file *ECprices201310.jmp.* Examine this JMP file.

1. Use the 60 points called Day1FirstHoursECprice to determine the linear regression trend line for this small subset of the dataset. Save this to the data Table.



1. What is the slope of the trend line? **-6.9194e-6**
2. Does this slope represent an upward or downward overall price trend? **Downward**
3. Save the residuals from the linear regression to your .jmp file (data table). Use a Normal Goodness of Fit test to determine the normality of the residuals. Save this report to the .jmp file (data table). Are the residuals normal? **Need to run Goodness of Fit n JMP.** **No p-value < .0001**
4. Does the autocorrelation (AC) indicate a stationary series? Yes or no and briefly why?

**Not stationary. ACF function trails off slowly so trending and non-stationary.**

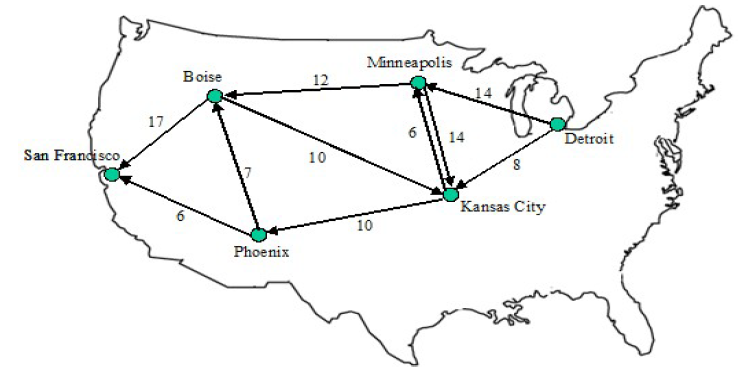
1. Apply an ARIMA(p,d,q) model set to the series individually with d=1 and p=0,1, q=0,1. You may run these individually or as a group. Save to the Data Table.

What is the worst performing model of ARIMA(0,1,0), ARIMA(1,1,0), ARIMA(0,1,1), ARIMA(1,1,1)? **ARIMA(0,1,0) = I(1) is worst and this makes sense because there is no moving average component to smooth the noise.**

**Problem 4 (9 points, 3 points for each)**

Circle the correct answer A through B or D – there is only one correct answer.

(a) Suppose there is a demand of 1000 car engines in San Francisco and the warehouses in Detroit and Kansas City both have 600 engines to supply the demand. Suppose the engines are transported through the following distribution network with the unit shipping cost given next to each arc. A network optimization model is built to minimize the shipping cost, which uses the variables to represent the number of engines shipped from to . What is the flow-balance constraint associated to Kansas City?



D

K

M

P

B

S

D

K

M

P

B

S

|  |  |
| --- | --- |
| A |  |
| B |  |
| C |  |
| D |  |

(b) In a capital budgeting problem, a decision maker having a limited amount of budget is considering to give financial support to research project A, B, C, D and E. Suppose is a binary variable, which equals 1 if project is chosen and 0 otherwise. Which of the following logical constraints correctly models the requirement that *“If Project C is not chosen, Project A and B cannot be chosen at the same time”*? **We can reason that A is not correct because when C is not chosen then it is preventing both A and B from occurring too. B allows for this case.**

|  |  |
| --- | --- |
| A |  |
| B | (c) Which of the following statements about time series is **correct**?   |  |  | | --- | --- | | A | Linear exponential smoothing is a technique to analyze a time series with strong seasonality. | | B | A seasonal series will have positive autocorrelations up to a number of lags. | | C | ARIMA(0,0,0)(1,1,0) model is essentially applying ARIMA(1,1,0) to each sub-series that consists of data points from the same season. 🡨 | | D | Centered simple moving average method will delay the turning point in the time series it forecasts. | |

**Problem 5 (10 Points)**

A trust officer at the Blacksburg National Bank needs to determine how to invest $100,000 in the following collection of bonds to maximize the total annual return (before tax).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Bond** | **Annual Return** | **Maturity** | **Risk** | **Tax-Free** |
| A | 9.5% | Long | High | Yes |
| B | 8.0% | Short | Low | Yes |
| C | 9.0% | Long | Low | No |
| D | 9.0% | Long | High | Yes |
| E | 9.0% | Short | High | No |

The officer wants to invest as least 50% of the money in short-term issues and no more than 50% in high-risk issues. At least 30% of the funds should go in tax-free investments, and at least 40% of the total annual return should be tax free.

Suppose the decision variable represents the amount of money invested in bond for . Formulate a linear programming (LP) model to solve the optimal strategy.

(a) (2 points) Write down the constraint requiring “*invest as least 50% of the money in short-term issues*”.

(b) (2 points) Write down the constraint requiring “*at least 30% of the funds should go in tax-free investments*”.

(c) (2 points) Write down the constraint requiring “*at least 40% of the total annual return should be tax free*”.

(d) (4 points) Solve the optimal strategy using the preliminary spreadsheet “*Investment.xls*”. Write down the optimal solution and the optimal objective value:

=**20339** =**20339**

=**29661** =**0**

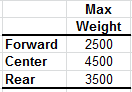
=**29661** Objective function value=**8898.3**

**Problem 6 (8 Points)**

Paul Bergey is in charge of loading cargo ships for International Cargo Company (ICC) at the port in Newport News, Virginia. Paul is preparing a loading plan for an ICC freighter destined for Ghana. An agricultural commodities dealer would like to transport the following four products aboard this ship.



Paul can elect to load any and/or all of the available commodities. However, the ship has three cargo holds with the following weight restrictions.



Moreover, only one type of commodity can be placed into the forward and the rear cargos and at most two types can be placed into the center cargo. Paul would like to load the commodities to maximize the total profit.

We need the following two groups of decision variables

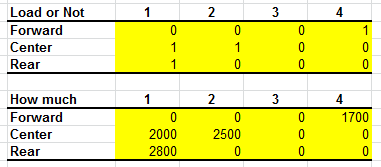
* if commodity is placed in cargo and if commodity is not placed in cargo , for and ;
* the weight of commodity placed in cargo , for and ;

(a) (2 points) Use the variables above to write down the constraint that represents “*the total weight of the commodities placed in the forward cargo cannot exceed its capacity*”.

(b) (2 points) Use the variables above to write down the constraint that represents “*the total weight of commodity placed in the whole ship cannot exceed its amount available*”.

(c) (2 points) Use the variables above to write down the constraint that represents “*at most two types of commodities can be placed in the center cargo*”.

(d) (4 points) Use preliminary Excel spreadsheet “*Shiploading.xlsx*” to solve the optimal loading plan. Which commodity is loaded in each cargo? How much (in tons) of each commodity is loaded in each cargo? **Objective function value: 597,000**



**Problem 7 (6 points)**

You work for a Des Moines financial firm doing a Pairs Trading study for two industrial stocks. Pairs Trading involves buying one stock while selling another.

Two time series are supplied: one for John Deere (DE) and one for Caterpillar (CAT).

A plot for these two series is supplied in the first tab. The difference of the two time series is plotted and saved at the second tab. Use the file “DEandCAT.jmp”

A time series ACF of DE appears at the next tab. A time series ACF of CAT appears at the next tab.

(a) If you were to model the DE time series with nonseasonal ARIMA(p,d,q), what would

be your best choice of d: 0 or 1? \_\_\_\_\_**1**\_\_\_\_\_

(b) If you were to model the CAT time series with nonseasonal ARIMA(p,d,q), what would

be your best choice of d: 0 or 1? \_\_\_\_\_**1**\_\_\_\_\_

(c) Perform the following analysis.

1. Create a new column called "DE AdjClose Lag 1" which is the Lag(.,1) of "DE AdjClose"
2. Create a new column called "DE AdjClose Log Ret" which is the difference of the Log(.) of "DE AdjClose" and the Log(.) of "DE AdjClose Lag 1". (In Finance this is known as the "log return".)
3. Create a new column called "CAT AdjClose Lag 1" which is the Lag(.,1) of "CAT AdjClose"
4. Create a new column called "CAT AdjClose Log Ret" which is the difference of the Log(.) of "CAT AdjClose" and the Log(.) of "CAT AdjClose Lag 1".
5. Using Analyze->Multivariate find the scatterplot and *correlation* of these two new series and report that figure to 3 digits behind the decimal point here: \_\_\_**.73**\_\_\_ and Save to the Data Table
6. Would you deem the log returns of these two stocks (DE and CAT) as having high, low, or negative correlation for the sample given? \_\_**high\_\_**