Final Exam - Part 2 (Programming)

Started: Apr 23 at 8:07am

Quiz Instructions

Use the following for Q1-2

Please use the data set '**direct_marketing.csv
(https://www.dropbox.com/s/ostavtpp6grm60d/direct_marketing.csv?dl=0)** ' to answer the
following questions.

Question 1	1 pts

We are interested in the effects of categorical variable 'Gender' and numerical variable
'Salary' on 'AmountSpent'. After running the linear regression of 'AmountSpent' on
'Gender' and 'Salary', which of the following is correct?

○ A. The intercept is significant on 95% confidence level

○ B. Male customers spend less than female customers

○ C. The 'Salary' effect is not significant on 95% confidence level

○ D. For one certain customer, higher salary generally means higher amount spent

---

Question 2	1 pts

Convert the Married column into a dummy variable with 'Married' being 1 and 'Single'
being 0. Now create another model for AmountSpent regressed against Gender, Salary
and Married. Based off your model, how much more (or less) would a married person
spend than a single person? (Assume arbitrary units)

○ A) 1.577

○ B) 0.36

○ C) -2.356

○ D) 3.970

Use the following for Q3-5

We are interested in predicting which houses have median value greater than $30,000 using the variables in the Boston housing dataset. The dataset Boston will be loaded into the environment once the following command is run:

library(MASS)

Please run the following code in R to learn more about the variables in the dataset:

?Boston

Create a new binary variable Result with a value of 1 if the medv (median value of owner-occupied homes in $1000s) variable is greater than $30k and 0 otherwise. Create a logistic regression to model this question using all the variables in the Boston dataset. Please do not forget to remove the medv variable while building the model. Use the information from the model to answer the following questions. Select the closest answer.

Question 3   1 pts

How should one interpret the coefficient of rm?

○ A. If rm increases by 1 unit, the natural log of the odds of the house median value being greater than $30,000 increases by 2.3549.

○ B. If rm increases by 1 unit, the odds of the house median value being greater than $30,000 increase by 2.021.

○ C. If rm increases by 1 unit, the odds of the house median value being greater than $30,000 increase by exp(2.854).

○ D. All of the above.

Question 4   1 pts

What is the Sensitivity of the model?

○ A. 0.650

○ B. 0.921

○ C. 0.857

○ D. 0.781

Question 5   1 pts

What is the AUC (Area Under the Curve) for this model?

○ A. 0.8549

○ B. 0.73

○ C. 0.5523

○ D. 0.9167

Use the following for Q6-7

Use "**AMZN.csv**   **(https://www.dropbox.com/s/wdhivoevziv06rn/AMZN.csv?dl=0)**" for the following two questions.

AMZN.csv file includes daily closing price for Amazon.com, Inc from 10/22 to 11/22. The risk-free rate is 0.19%.

The formula for Sharpe Ratio is as follows:

$$\text{Sharpe Ratio} = \frac{R_p - R_f}{\sigma_p}$$

R_p	Return of portfolio
R_f	Risk-free rate
σ_p	Standard deviation of portfolio's excess return

Question 6

1 pts

What is the Sharpe Ratio across the period 10/22/2019 to 11/22/2019 for Amazon.com? (Please use arithmetic return and omit 10/22 when you calculate Sharpe Ratio, i.e.: only 23 data points are used to calculate Sharpe Ratio)

(Hint: Use the PerformanceAnalytics library to solve this question)

○ A. 0.34

○ B. -0.34

○ C. 1.23

○ D. -1.23

Question 7

1 pts

Calculate the holding period return for the Amazon stock from 10/22/2019 to 11/22/2019 (include the beginning and the end dates).

○ A. -1.27%

○ B. -0.56%

○ C. -2.22%

○ D. +2.22%

Use the following for Q8-10

Please use the data set **UPS_KO.csv**
(https://www.dropbox.com/s/vqil143rbd2b55m/UPS_KO.csv?dl=0) to answer the following questions. For each column:

Date: This column represents date from 09/2014 to 08/2019.

Mkt_RF: This column represents market premium (i.e., Market return – risk_free rate).

SMB: This column represents size factor.

HML: This column represents value factor.

RF: This column represents risk free rate.

UPS: This column represents return of UPS.

KO: This column represents return of KO.

Question 8

1 pts

Estimate a three-factor model by regressing return in excess of the risk free rate on Mkt_rf; SMB; and HML for both UPS and KO.

The coefficient of SMB for the three factor model for KO suggests that:

○ A. KO is tilted towards small cap stocks

○ B. KO is tilted towards large cap stocks

○ C. KO is tilted towards value stocks

○ D. KO is tilted towards growth stocks

Question 9

1 pts

Use excess returns of UPS and KO and run three-factor (i.e., Mkt_RF, SMB, HML) models for both UPS and KO. Which of the following factor is statistically significant at the 0.01 significance level for both models?

○ A. Market premium

○ B. SMB

○ C. HML

○ D. None of above

Question 10

1 pts

Which firm(s) has statistically significant alpha (95% confidence level) according to the models from Question 9?

○ A. UPS

○ B. KO

○ C. Both of UPS and KO

○ D. Neither UPS nor KO

Use the following for Q11-15

Please use the Facebook Ad dataset **KAG_conversion_data_wrangled.csv (https://www.dropbox.com/s/xzznyebxmz04vm0/KAG_conversion_data_wrangled.csv?dl=0)** for the next set of questions. We advise to solve these questions using R (preferably using *dplyr* library wherever applicable) after reviewing the code provided for Week 11 and other resources provided for learning *dplyr* in R Learning Guide

Load the dataset as:

data <- read.csv("KAG_conversion_data_wrangled.csv",stringsAsFactors = FALSE)

Question 11

1 pts

Which ad (ad_id) has the highest cost per click?

○ A. 1121223

○ B. 1121129

○ C. 1121413

○ D. 1121229

Question 12

1 pts

What percentage of ads (ad_id) have a cost per click (CPC) of at least $1? (>=1).

NOTE: Please round the answer to 2 decimal places.

- A. 59.49%

- B. 76.29%

- C. 78.83%

- D. 85.40%

Question 13

1 pts

Which age group did *Campaign 936* (campaign_id = 936) reach out most to according to mean number of impressions?

- A. 32

- B. 37

- C. 42

- D. 47

Question 14

1 pts

Which 'age' group has highest number of total 'impressions' (cumulative)?

- A. 32

- B. 37

- C. 42

- D. 47

Question 15   1 pts

Which campaign_id has the lowest value of the mean 'CPC'?

○ A. 916

○ B. 936

○ C. 1178

Use the following for Q16-17

In the dataset "**Queue** **(https://www.dropbox.com/s/3h6vcy4mrpy4fj6/Queue.csv?dl=0)** ",
you are given the data of the first three days of the week for a Bank of America outlet
based in Downtown Atlanta. All times are in minutes.

Question 16   1 pts

The arrival rate per hour on Monday is 125.71 customers/hour calculated as – numbers of
customers arrived divided by the total number of hours they arrived in (880/7 = 125.71).
On the basis of the above, calculate the arrival rate (in hours) on Tuesday and
Wednesday.

Note: Consider total working hours on Tuesday and Wednesday as 8 hours and 7 hours
respectively.

Now calculate the utilization rate on Monday, Tuesday and Wednesday using the above
calculated arrival rates and the following service rates: Monday is 140 customer/hour,
Tuesday is 120 customers/hour and on Wednesday is 200 customers/hour.

○ A. utilization Monday = 0.69, utilization Tuesday = 0.65, utilization Wednesday = 0.73

○ B. utilization Monday = 0.89, utilization Tuesday = 0.75, utilization Wednesday = 0.83

○ C. utilization Monday = 0.69, utilization Tuesday = 0.65, utilization Wednesday = 0.83

○ D. utilization Monday = 0.89, utilization Tuesday = 0.75, utilization Wednesday = 0.73

Question 17   1 pts

Using the queuing package in R, develop a MM1 model and generate a report of the queue characteristics **for Tuesday** using the arrival rates and service rates mentioned in the previous question. (take n = 20) Hint: Refer to the code in the "practice questions" file for week 12.

What is the average number of customers in the queue and the average time a customer waits in the queue?

○ A. Average time = 0.025, Average number of customers = 2.25

○ B. Average time = 0.015, Average number of customers = 2.05

○ C. Average time = 0.035, Average number of customers = 3.15

○ D. Average time = 0.05, Average number of customers = 3.25

Use the following for Q18-20

You need **Daily_Demand.csv (https://www.dropbox.com/s/z6k411yff8zdp3x/Daily_Demand.csv?dl=0)** for this and using simple exponential smoothing to forecast demand. Convert the csv to an xts object. (Make sure you only have 2 columns, Date and total while converting)

Question 18   1 pts

Model the data with alpha = 0.35 and h =5. What is the RMSE? (Hint: Use accuracy())

○ A. 95

○ B. 200

○ C. 35

○ D. 70

Question 19   1 pts

What is the MAE?

○ A. 32

○ B. 74

○ C. 124

○ D. 102

Question 20   1 pts

Tune the value of alpha, i.e. try to minimize RMSE across different values of alpha (from 0.01 to 0.99). What value of alpha do you get and what does this imply?

○ A. 0.95, there is a low reaction of forecasts to the difference in the previous forecasted and actual demand.

○ B. 0.95, there is a high reaction of forecasts to the difference in the previous forecasted and actual demand.

○ C. 0.01, there is a high reaction of forecasts to the difference in the previous forecasted and actual demand.

○ D. 0.01, there is a low reaction of forecasts to the difference in the previous forecasted and actual demand.

Quiz saved at 8:07am   Submit Quiz