Hanna Syed Answers in BLUE

DAT 411 – Exam 1

Practice Exam 1 has three sections and worth 200 points. The first section is worth 30 points, while the second part is worth 60 and third part is worth 110 points, respectively. This is a take-home exam, and such all class materials can be utilized. However, this exam must be completed independently. Do not share work with other students. ChatGPT can be only used for two reasons: editing your already written work and helping with code. DO NOT USE IT TO HELP YOU ANSWER THE QUESTION. The exam is due by 11:59 PM on Tuesday, October 8th.

Section 1: Short Answer

A graph with blue dots

Description automatically generated

Source (Joe Mahoney (SB Nation) March 22 2022 – X(Formerly Twitter))

1. Say you saw this tweet from Joe Mahoney, a writer from SB Nation, which shows the connection between Hand Size and Fumbles per game started for Quarterbacks in the NFL. Joe Mahoney claims that this shows there is no relationship. Critique this analysis. You do not need to know anything about Football to critique this.

R^2 has no context. By looking at solely at R^2, it does not tell you anything as you can only use r^2 for comparison across different regressions with the same dependent variable. Often when people show the r^2 and a graph, the people who publish it don’t know what r^2 means, or they are trying to fool their audience. Joe Mahoney may have intent on swaying people into believing that hand size and fumbles are not related. R^2 can also be driven by the variance of the error term. Hand size is not the only factor that can impact the number of fumbles, causing there to be a lot of noise in the r^2 and can be driven by the error term.

1. Say you are trying to study the relationship between the change in housing prices and mortgage rates. As such you estimate a simple linear model. You find a positive value of the coefficient on mortgage rates. Further you conduct a hypothesis test on the coefficient on reject the null hypothesis. Does this result make sense? Think about what possible classical assumptions that might be violated (hint: does GDP growth matter?)? Explain.

Typically when mortgage rates increase, it is more expensive to borrow money to buy homes, causing a decrease in demand for homes, and making the price of housing decrease (Quicken Loans). This result does not make sense as the mortgage rate coefficient is positive when we should expect it to be negative. The classical assumption that may have been violated is “no relationship between regressors and error term making the price of housing decrease (Quicken Loans). This result does not make sense as the mortgage rate coefficient is positive when we should expect it to be negative. The classical assumption that may have been violated is “no relationship between regressors and the error term”. In this context, housing prices and mortgage rates. The GDP growth rate is one variable that could have been in the error term but should have been a part of the linear model. P directly impacts mortgage rates, when GDP grows, mortgage rates increase. When GDP growth is not included in the linear model in creates a bias in the coefficients of the regression model. Our expected values of the estimated beta is not equal to the true value of beta due to biases that violates the classical assumptions.

Source: https://www.quickenloans.com/learn/interest-rates-vs-house-prices

Section 2: Written Analysis

1. Say that we ran the following regression from the NBA Salary data set. Salary is salary in dollars and GS is games started.

A computer screen shot of a number

Description automatically generated

1. Write down the regression equation.

Salary = B 0 + B1(GS) + e

Salary = 3025135 + 238024(GS)

1. Interpret the coefficients.

Intercept: If an NBA player's games started equal 0 it is associated with a salary of 3025135 dollars.

GS: For every one unit increase in games started for an NBA player it is associated with a salary increase of 238024 dollars

1. Interpret and from the regression output.

Our model can explain 36.27% of variance of NBA player salary

Accounting for the adjusted loss of degree of freedom our model can explain 36.14% of variance of NBA player salary

1. Conduct the following hypothesis test:
   * + 1. Calculate the t-stat by hand.

238024/14631 = 16.269

* + - 1. Interpret the results of the hypothesis test.

We reject the null hypothesis at 5% because we find statistical evidence that a change in games started (GS) is associated with a change in salary.

This means we have strong evidence that starting in more NBA games is associated with a higher salary

* 1. Say that we ran the following regression from the NBA Salary data set. Salary is salary in dollars, GS is games started, Age is the players age, Total.Minutes is the total minutes played and FG. is field goal percentage.

A screenshot of a computer program

Description automatically generated

1. Write down the regression equation.

Salary = B0 + B2(GS) + B3(Age) + B4(Total Minutes) + B5(FG) + e

1. Interpret the coefficients.

Intercept: If an NBA player's games started, age, total minutes and field goal percentage all equal 0, it is associated with a salary of 21070000 dollars.

GS: For every one unit increase in games started for an NBA player, the salary is expected to increase by 167600 dollars, holding all else constant.

Age: For every one unit year increase in age, the salary is expected to increase by 901400 dollars holding all else constant.

Total Minutes: For every one unit increase in minutes played in total, the salary is expected to increase by 2098 dollars, holding all else constant.

FG: For every one unit increase in percentage point for Field Goal Percentage, the salary is expected to increase by 194400 dollars, holding all else constant.

1. Interpret and and compare with the previous result.

Our model can explain 49.64% of the variance of NBA player salary.

Accounting for the adjusted loss of degree of freedom our model can explain 49.21% of variance of NBA player salary

1. Conduct the following hypothesis test and interpret the results:

167600/ 27300= 6.14

We reject the null hypothesis at 5% because we find statistical evidence that a change in games started is associated with a change in salary.

This means we have strong evidence that starting in more NBA games is associated with a higher salary.

901400/83200 = 10.834

We reject the null hypothesis at 5% because we find statistical evidence that a change in age is associated with a change in salary.

This means we have strong evidence that a player's age is associated with a higher salary.

2098/893.7 = 2.348

We reject the null hypothesis at 5% because we find statistical evidence that a change in total minutes played is associated with a change in salary.

This means we have evidence that playing more total minutes is associated with a higher salary.

Section 3: R Portion

1. Use the following “DAT 411 – Exam 1- Fall 2024.csv” data set and complete the following section in R. Make sure to include your script file. This file shows the excess returns (ER) from an industry portfolios and factors that are used in evaluating portfolio returns. Find your specific industry portfolio in the table below.

|  |  |
| --- | --- |
| Initial | Industry |
| MC | Autos |
| JE | Books |
| GF | Steel |
| PR | Games |
| HS | Coal |

* + 1. Import the data
    2. Get summary statistics of ER and ERM and discuss them.
    3. Conduct a hypothesis test of the population mean for ER and ERM
    4. Run the following Regression (CAPM):
       1. Discuss R-squared, and Adjusted R-squared
       2. Interpreted the coefficient, and conduct the following individual hypothesis test:
       3. Save the predicted values. Create a scatter plot of the predicted values and ERM and a line graph of predicted values and ERM in the same graph. Discuss this graph.
    5. Run the following Regression (FF3):
       1. Compare Rsquared and Adj Rsquared with the previous model.
       2. Interpreted the coefficient, and conduct the following individual hypothesis test: