QMB 6316: R for Business Analytics
College of Business
University of Central Florida
Fall 2024

Final Examination

Due Friday, December 6, 2024 at 11:59 PM
in your GitHub repository.

Instructions:

Complete this examination within the space on your GitHub repo in a folder called `final_exam`. Store any printed output by writing or pasting into the `README.md` file provided and save all of your code in the `final_exam` folder. When you are finished, submit your code and responses by pushing your changes to your GitHub repository. Complete these exercises individualy and push your own work.

Part A: Data Handling and Preliminary Regression Modelling

Estimate the best regression model you can by solving Questions 1 to 4.

Question 1:

The folder `final_exam` contains three `.csv` files: `airplane_sales.csv`, `airplane_specs.csv`, and `airplane_perf.csv`. The first dataset `airplane_sales.csv` contains the following variables.

`SALE_ID`	=	a unique key for each airplane sold
`price`	=	price of an airplane
`age`	=	age of the aircraft, in years

Use this dataset to estimate a regression model to predict the prices of airplanes.

a) Read in the `airplane_sales.csv` dataset and store it in a data frame called `airplane_sales` in your workspace.

b) Calculate and store the printed output from a `summary` of the data. Use this to get familiar with the contents of the dataset.

c) Estimate a regression model to predict `price` as a function of `age`. Store the printed estimation output with the `summary` command.

Question 2:

Now use two files `airplane_sales.csv` and `airplane_specs.csv` in the folder `final_exam`. The dataset `airplane_specs.csv` contains the following variables.

`SALE_ID`	=	a unique key for each airplane sold
`pass`	=	the number of passengers an airplane can accommodate
`wtop`	=	an indicator that the wings are above the fuselage
`fixgear`	=	an indicator for fixed landing gear (i.e. wheels are not retractable)
`tdrag`	=	an indicator that a wheel is on the tail (a tail-dragger)

Use the variables from both datasets to estimate a better regression model to predict the prices of airplanes.

a) Form a dataset `airplane_sales_specs.csv` by `merge`ing these datasets.

b) Store the new dataset in a data frame called `airplane_sales_specs` in your workspace.

c) Calculate and store the printed output from a `summary` of the data. Use this to get familiar with the contents of the dataset.

d) Estimate a regression model to predict `price` as a function of `age`, `pass`, `wtop`, `fixgear`, and `tdrag`. Store the printed estimation output with the `summary` command.

Question 3:

Now use all three files `airplane_sales.csv`, `airplane_specs.csv`, and `airplane_perf.csv` in the folder `final_exam`. The dataset `airplane_perf.csv` contains the following variables.

`SALE_ID`	=	a unique key for each airplane sold
`horse`	=	the horsepower of the engine
`fuel`	=	the volume of the fuel tank, in gallons
`ceiling`	=	the maximum flying height of an airplane, in feet
`cruise`	=	the cruising speed, in MPH

Use the variables from these datasets to estimate an even better regression model to predict the prices of airplanes.

a) Form a dataset `airplane_full.csv` `merge`ing the datasets.

b) Store the new dataset in a data frame called `airplane_full` in your workspace.

c) Calculate and store the printed output from a `summary` of the new variables. Use this to get familiar with the contents of the dataset.

d) Estimate a regression model to predict `price` as a function of `age`, `pass`, `wtop`, `fixgear`, and `tdrag`, as well as `horse`, `fuel`, `ceiling`, and `cruise`. Store the printed estimation output with the `summary` command.

Part B: Advanced Regression Modelling

Question 4:

Now calculate new variables to estimate a model for airplane prices using a different functional form.

a) Create new variables `log_price`, `log_age`, `log_horse`, `log_fuel`, `log_ceiling`, and `log_cruise` from the variables `price`, `age`, `horse`, `fuel`, `ceiling`, and `cruise`, using the logarithm function `log()` in `R` to create these new variables.

b) Calculate and store the printed output from a `summary` of the new variables. Use this to get familiar with the contents of the dataset.

c) Estimate a regression model to predict `log_price` as a function of `log_age`, `pass`, `wtop`, `fixgear`, and `tdrag`, as well as `log_horse`, `log_fuel`, `log_ceiling`, and `log_cruise`. Store the printed estimation output with the `summary` command.

d) If you notice that any coefficients are statistically insignificant, estimate the model by removing them one at a time. For each variable removed, determine whether the variable should be removed by considering the four specification criteria: statistically significant t-statistics, an increase in $\bar{R}^2$, a good theoretical justification, and no large change in the other coefficients.

Part C: Version Control

Question 5:

Push your completed files to the `final_exam` folder in your GitHub repository following these steps. See the `README.md` in the folder `demo_02_version_control` in the QMB6316F24 course repository for more instructions.

1. Make sure to save all of your examination materials to the folder `final_exam` in your private, personal GitHub repository.
2. Use GitHub Desktop to add and commit your files to your repository. Include an informative message to indicate that the submission includes your final examination results.
3. Push your changes up to the online repository. You can do this by pressing the blue "Push" button in GitHub Desktop. After this step, the changes should be visible on a browser, after refreshing the page.
4. As a last resort, you could upload your files individually through your internet browser.
5. Most importantly, verify that all files in your submission appear in the online repository in your browser window. Only the contents of your repository will be graded.