QMB 6315: Python for Business Analytics
College of Business
University of Central Florida
Spring 2025

Final Examination

Due Tuesday, April 29, 2025 at 11:59 PM
in digital form in your GitHub repository.

Instructions:

Complete this examination within the space on your *private* GitHub repo and save any scripts or output in a folder called `final_exam`. In this folder, save your answers to Questions 1 to 4 in a script called `purchases.py`, Store any printed output by writing or pasting into the `README.md` file provided. When you are finished, submit your code and responses by pushing your changes to your GitHub repository. Complete these exercises individually 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 a database `customers.db` with three tables: `Applications`, `CreditBureau`, and `Demographic`. The first table `Applications` contains the following variables.

`app_id`	=	a unique key for each customer who applied for credit
`ssn`	=	the social security number
`zip_code`	=	the the zip code in which the applicant resides
`income`	=	the applicant's reported income
`homeownership`	=	a categorical variable that indicates whether an applicant owns or rents a home
`purchases`	=	the monthly value of purchases on the account
`credit_limit`	=	the maximum amount that an applicant is approved to spend

Use the first table, `Applications`, to estimate a regression model to predict the prices of airplanes.

a) Wrte a query that will obtain the data from the `Applications` table.

b) Obtain the data from the `Applications` table and store it in a data frame called `purchase_app` in your workspace.

c) Estimate a regression model to predict `purchases` as a function of the other variables in the dataset. Ignore the variables `app_id`, `ssn` and `zip_code`, which are keys for databases. Print the printed estimation output from the command `print(reg_model_app.summary())` to see the results.

Question 2:

The next table `CreditBureau` contains the following variables.

`ssn`	=	the consumers unique social security number
`zip_code`	=	the zip code in which the consumer resides
`fico`	=	the consumer's credit score
`num_late`	=	the number of number of times a consumer has made a payment after the due date
`past_def`	=	the number of number of times a consumer has defaulted on a line of credit
`num_bankruptcy`	=	the number of number of times a consumer has filed for bankruptcy

Use both tables, `Applications` and `CreditBureau`, to estimate a better regression model to predict the prices of airplanes.

a) Wrte a query that will obtain the data from the `Applications` table and join it to the data from the `CreditBureau` table.

b) Obtain the data from the above query and store it in a data frame called `purch_app_bureau` in your workspace.

c) Estimate a regression model to predict `purchases` as a function of the other variables in the dataset. Ignore the variables `app_id`, `ssn` and `zip_code`, which are keys for databases. Print the printed estimation output from the command `print(reg_model_app_bureau.summary())` to see the results.

Question 3:

The next table `Demographic` contains the following variables.

`zip_code`	=	the zip code to indicate each geographic region
`avg_income`	=	the average income in each zip code
`density`	=	the population density in each zip code

Use all three tables to estimate an even better regression model to predict the prices of airplanes.

a) Wrte a query that will obtain the data from the `Applications` table and join it to the data from the `CreditBureau` table and the the `Demographic` table.

b) Obtain the data from the above query and store it in a data frame called `purchase_full` in your workspace.

c) Estimate a regression model to predict `purchases` as a function of the other variables in the dataset. Ignore the variables `app_id`, `ssn` and `zip_code`, which are keys for databases. Print the printed estimation output from the command
`print(reg_model_full.summary())` to see the results.

Part B: Advanced Regression Modelling

Question 4:

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

a) Create a new variable `utilization`, which is defined as the ratio of `purchases` to the consumer's `credit_limit`.

b) Calculate and copy the printed output from `describe`-ing the new variables. Use this to get familiar with the nature of this variable.

c) Estimate a regression model to predict `utilization` as a function of the other relevant variables in the dataset. *[Of course, don't use the variables used to calculate* `utilization`*. If you observed those, you could predict* `utilization` *exactly.]* Print the estimation output from the `summary` command to see the results.

d) Create a new variable `log_odds_util`, the log-odds ratio, which is defined as the logarithm of this ratio: `utilization` divided by one minus `utilization`. Use the logarithm function `math.log()` in `Python`.

e) Inspect the new variable and estimate a regression model to predict `log_odds_util` as a function of the other relevant variables in the dataset. *[As mentioned in part c, above, don't use the variables used to calculate* `utilization`*.]* Print the estimation output from the `summary` command to see the results.

f) Of all the regression models you estimated, recommend one as the best model. Print the output from a `summary` of the regression results of your final regression model and copy the output to paste it in the `README.md` file.

Part C: Version Control

Question 5:

Push your completed files to the `final_exam` folder in your GitHub repository following these steps. If you need a reminder, see the `README.md` in the folder `demo_02_version_control` in the QMB6315S25 course repository, but the process is the same as we have been following for the other assignments in this course.

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.