

Course: **CVE 593 – Special Topics in Civil Engr**

Task: **Homework #2**

Due Date: **Oct 23, 2022 at 11:59 PM**

**Problem 1 (20 Points)**

Data provided on BB for this problem are estimated trips generated from 120 traffic analysis zones. These data were collected to understand the mobility pattern during off-peak hours. The definition of the attributes in the dataset is as follows:

- Average zonal income in thousands (income),
- Zonal population in thousands (pop),
- Average number of vehicles in the household (veh),
- Average number of children in a household (children)
- Average trip rates during off-peak period (Trips\_rates).

Task: Use Python particularly *statsmodels* package to calibrate linear regression models that estimate trip rates.

e.g.,

```
import statsmodels.formula.api as smf
lm_m = smf.ols(formula="...", ...)
```

- Fit a regression between trip rate and income **(2 Points)**
- Fit a regression between trip rate and pop **(2 Points)**
- Fit a multiple linear regression between trip rate with all the attributes in a table. Identify which attributes are significant at 95% confidence interval **(2 Points)**
- Use the calibrated model in part c to develop a prediction function for the average trip rates. Test the function for prediction using a traffic analysis zone with the number of pop = 20,000, income = 65, 000, average children = 2, and vehicle per household = 2 **(5 Points)**

e.g.,

```
def Trip_Rates(...):
    .....
    return pred_trip_rates
```

- In each of the three models, comment on the quality of the fitted model and identify all the significant factors in the models. **(5 Points)**
- Among the three fitted models, which regression equation would you recommend for use for planning purposes? Why? **(5 Points)**

## Problem 2 (30 Points)

Data for this problem are bicycle counts that were collected using a loop detector at one specific location in Cleveland area. The data were cleaned and integrated with weather data. The attributes for analysis include:

- bike counts,
- days of the week (Mon through Sun),
- a day if was a holiday (holiday),
- day light hours holiday (DayLightHrs),
- average temperature (AvgTempInC),
- precipitation in inches (PRCP\_IN),
- if the day was dry or not (Dry\_Day)

Task: Use Python *statsmodels* particularly a *GLM* package to calibrate Poisson and Negative Binomial regression models that estimate bike counts.

```
import statsmodels.api as sm
from patsy import dmatrices

# Fit GLM in statsmodels using Poisson link function
sm.GLM(y, X, family = sm.families.Poisson()).fit().summary()
```

- Fit a Poisson regression model with all the attributes, identify which attributes are significant at 95% confidence interval, and comment on each variable how it influences the bike counts **(10 Points)**
- Fit a Negative Binomial regression model with all the attributes, identify which attributes are significant at 95% confidence interval, and comment on each variable how it influences the bike counts **(10 Points)**
- Compare and contrast Poisson and Negative Binomial estimates. Assume that you work for the city of Cleveland, which model will you recommend to be used for planning purposes? Why? **(10 Points)**

**Submit your solution together with your code to receive full points**