Lab 3
Trip Generation (2)

Yufeng Zhang
zhan4879@umn.edu

CEGE-3201: Transportation Engineering

February 7, 2019

University of Minnesota

Announcements

- Report 1 will be due Feb 13 (Wednesday) at 23:59. Your answers for worksheet 1, 2 and 3 should all be included in Report 1.
- Name your .pdf file with your last name (both if work in pairs); Report done in pairs needs only one submission.

Objectives

- Building trip generation linear regression models
- Checking linear model validity
- Creating plots using R

Data sets

- **TAZ2010_truncated.xlsx** is the main data set for this lab
- It contains aggregated TAZ info and trip data(from TBI!) and demographic data from the the US Census
- Use **TAZ2010-MetaData.pdf** as a reference for explanation of variables in the data set.

Recap

A summary of a linear model in R

```
> summary(model1)

Call:
lm(formula = Otrips ~ POPOVER18 + HHTYPE3 + RENTEROCC, data = TAZ2010)

Residuals:
   Min     1Q Median     3Q    Max
-10991  -2250  -1254   1158  48781

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1441.5332   117.7682  12.240   <2e-16 ***
POPOVER18      2.8151     0.1398  20.134   <2e-16 ***
HHTYPE3       -3.4089     2.8158  -1.211   0.2261
RENTEROCC      1.3449     0.6499   2.069   0.0386 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4335 on 3026 degrees of freedom
Multiple R-squared:  0.2187,    Adjusted R-squared:  0.2179
F-statistic: 282.3 on 3 and 3026 DF,  p-value: < 2.2e-16
```

If t-value is larger, it is more likely to reject the null hypothesis.

If the p-value of t test is smaller, it is more likely to reject the null hypothesis of the t test.
(0.05 and 0.1 are commonly used thresholds)

If the p-value of F-test is smaller, it is more likely to reject the null hypothesis of the F-test.

F-test:

- Null hypothesis: Otrips = the average value of Otrips
- Alternative hypothesis: Otrips = $b_0 + b_1$*POPOVER18+b_2*HHTYPE3+b_3*RENTEROCC

T-test:

- Null hypothesis: the coefficient is zero
- Alternative hypothesis: the coefficient is not zero

Figure: Summary of a linear model in R

Linear regression model diagnostics

Linear regression model assumptions:

- **Linearity of the data**: the relationship between the predictor (x) and the outcome (y) is assumed to be linear
- **Normality of residuals**: the residual errors are assumed to be normally distributed
- **Homogeneity of residuals variance**: the residuals are assumed to have a constant variance
- **Independence of residuals error terms**

Some commands we will use for today

```
plot(xvalues, yvalues)
abline(LinearModel)
pairs(DataColumns)
plot(xvalues, yvalues, main = "title", xlab = "
   x axis title", ylab ="y axis name")
```

Practice in R...

Refer to the document "Linear model diagnostics.pdf".