DO- directly operated by property, PT- service contracted to private operator.

MB-bus, LR-rail, DR-direct response, FB-ferry boat, VP- vanpool

Big Question:

Trips\_per\_person vs average fares -graphs

(Also logged version- graphs)

Elasticity derivation

-Differentiate by mode

-Differentiate by TOS

-Relationship between fares and operating expenses

-Relationship between fares-expenses and unlinked trips

-Control of population with total trips and fares

-Average trip length effect control?

-Fare per mile effect?

-Distribution of main variables (justify log from histogram and elasticity value)

-Plot lines for each mode/TOS?

Regression lists:

reg trips\_per\_person average\_fare, robust

reg log\_trips\_per\_person log\_average\_fare, robust

reg log\_trips\_per\_person log\_average\_fare if tos=="PT", robust

reg log\_trips\_per\_person log\_average\_fare if tos=="DO", robust

reg log\_trips\_per\_person log\_average\_fare if mode=="LR", robust

reg log\_trips\_per\_person log\_average\_fare if mode=="MB", robust

reg log\_trips\_per\_person log\_average\_fare if mode=="DR", robust

From this dataset, I wanted to answer the question of the relationship between transit fares (average fare per trip) and unlinked passenger trips. This is an interesting question to me, as we may be able to see if fares impact whether passengers choose to ride transit (sort of an elasticity question).

<https://www.transit.dot.gov/sites/fta.dot.gov/files/docs/2017%20NTD%20Policy%20Manual.pdf>

<https://www.transit.dot.gov/ntd/data-product/monthly-module-raw-data-release>

d. I plan to ultimately explore the relationship between fares and transit trips, while respecting that other variables should not be omitted in that their impact on trips may be significant. By including relevant controls, we can try to isolate the relationship between fares and trips. To explore the specification of the model, I plan to explore the distribution of the variables to determine if a log-transformation is necessary. Of course, I hope to estimate elasticity by using log transformations of both the dependent variable (transit trips) and average fares. I also will ensure to test for significance of results, conduct heteroskedasticity tests (and potentially deal with heteroskedasticity), and conduct meaningful analysis on independent variables (including summary statistics).

**Planning:**

* Abstract
* Intro- describe question (i) with economic formulas and context, discuss differentiation.
* Data (iii)- describe data source, relevant columns, summary statistics, cleaning process, importance of differentiation and controls, distribution of main variables (logging need), explain robust choice, need to normalize per person.
* Analysis- first construct initial regression and specification (ii), show graphs/fits/results (iv), explain new effects (by TOS/Mode).