Benjamin Epstein

Problem Set 3

Empirical Problem - Duranton and Turner Code

Urban Economics, Fall 2022

Question 2: Replicating and Interpreting Main Results [2 points]

Preliminary notes on variable names (remember that all variables are at the MSA level):

|  |  |
| --- | --- |
| Variable name | Variable description |
| vmt\_IH\_YR | Mean daily VKT for year YR (interstates) |
| ln\_km\_IH\_YR | Mean lane km for year YR (interstates) |
| popYR | Population for year YR (in MSA) |

In this question, you will need to *replicate* results. In this context, replicate means to produce code that prints a results table on the screen when I run it. Note that you need to replicate *both the standard errors and the point estimates* (except as noted). Also, note that the authors exclude from these tables any MSA for which they do not have data on KM of highway in 1983, 1993, or 2003. (A missing data point for the mean lane km variable is one where the data table has a zero for that MSA for that year).

1. Table 1: Generate a data frame/matrix with the average across MSAs for the three variables listed above (for all years) [you do not need standard deviations]. [Hint: You can use either the lapply or summary functions.]

Create subset, removing MSA = 0.

dtdata\_nozeros <- subset.data.frame(dtData, msa > 0)

A table of statistics with numbers and text

Description automatically generated

Create necessary variables:

> VKT\_urbanizedshare <- dtdata\_nozeros$vmt\_IHU\_83/dtdata\_nozeros$vmt\_IH\_83

> VKT\_urbanizedshare\_1993 <- dtdata\_nozeros$vmt\_IHU\_93/dtdata\_nozeros$vmt\_IH\_93

> VKT\_urbanizedshare\_2003 <- dtdata\_nozeros$vmt\_IHU\_03/dtdata\_nozeros$vmt\_IH\_03

> Lane\_km\_share\_1983 <- dtdata\_nozeros$ln\_km\_IHU\_83/dtdata\_nozeros$ln\_km\_IH\_83

> Lane\_km\_share\_1993 <- dtdata\_nozeros$ln\_km\_IHU\_93/dtdata\_nozeros$ln\_km\_IH\_93

> Lane\_km\_share\_2003 <- dtdata\_nozeros$ln\_km\_IHU\_03/dtdata\_nozeros$ln\_km\_IH\_03

> Lane\_km\_share\_10000pop <- dtdata\_nozeros$ln\_km\_IH\_83/10000

> S\_truck83\_10000pop <- dtdata\_nozeros$S\_truck83/10000

> S\_truck93\_10000pop <- dtdata\_nozeros$S\_truck93/10000

> S\_truck03\_10000pop <- dtdata\_nozeros$S\_truck03/10000

> peak1983 <-max(dtdata\_nozeros$l\_max\_84bus)

> peak1993 <-max(dtdata\_nozeros$l\_max\_94bus)

> peak2003 <-max(dtdata\_nozeros$l\_max\_04bus)

> peak1983\_10000pop <-peak1983/10000

> peak1993\_10000pop <-peak1993/10000

> peak2003\_10000pop <-peak2003/10000

## Assemble Table Parts for Each Year:

##1983:

> table1983 <- summary(dtdata\_nozeros$vmt\_IH\_83, dtdata\_nozeros$aadt\_IH\_83,

> dtdata\_nozeros$ln\_km\_IH\_83, dtdata\_nozeros$Lane\_km\_share\_10000pop,

> dtdata\_nozeros$vmt\_MRU\_83, dtdata\_nozeros$aadt\_MRU\_83)

> dtdata\_nozeros$ln\_km\_MRU\_83, dtdata\_nozeros$VKT\_urbanizedshare,

> Lane\_km\_share\_1983, s\_truck83, peak1983\_10000pop, peak1983)

##1993:

> table1993 <- summary(dtdata\_nozeros$vmt\_IH\_93, dtdata\_nozeros$aadt\_IH\_93,

> dtdata\_nozeros$ln\_km\_IH\_93, dtdata\_nozeros$Lane\_km\_share\_10000pop,

> dtdata\_nozeros$vmt\_MRU\_93, dtdata\_nozeros$aadt\_MRU\_93,

> dtdata\_nozeros$ln\_km\_MRU\_93, dtdata\_nozeros$VKT\_urbanizedshare,

> Lane\_km\_share\_1993, s\_truck93, peak1993\_10000pop, peak1993)

##2003:

Table2003 <- summary(dtdata\_nozeros$vmt\_IH\_03, dtdata\_nozeros$aadt\_IH\_03, dtdata\_nozeros$ln\_km\_IH\_03, dtdata\_nozeros$Lane\_km\_share\_10000pop, dtdata\_nozeros$vmt\_MRU\_03, dtdata\_nozeros$aadt\_MRU\_93, dtdata\_nozeros$ln\_km\_MRU\_03, dtdata\_nozeros$VKT\_urbanizedshare, Lane\_km\_share\_2003, s\_truck93, peak2003\_10000pop, peak2003)

Tab\_model(table1983, table1993, table2003)

##Table 3 Regressionss

##Column One

model1983\_1A <- lm(dtdata\_nozeros$l\_vmt\_IH\_83 ~ dtdata\_nozeros$ln\_km\_IH\_83 + dtdata\_nozeros$l\_pop80, dtdata\_nozeros, coeftest(regRes, vcov = vcovHC(regRes, "HC1"))”

model1993\_2A <- lm(dtdata\_nozeros$l\_vmt\_IH\_93 ~ dtdata\_nozeros$ln\_km\_IH\_93 + dtdata\_nozeros$l\_pop90, dtdata\_nozeros, coeftest(regRes, vcov = vcovHC(regRes, "HC1"))”

)

model2003\_3A <- lm(dtdata\_nozeros$l\_vmt\_IH\_03 ~ dtdata\_nozeros$ln\_km\_IH\_03 + dtdata\_nozeros$l\_pop00, dtdata\_nozeros, coeftest(regRes, vcov = vcovHC(regRes, "HC1"))”

)

##Column Four

##Restrict subset to only 1993 and 2003

<- ln\_km\_IH\_1993\_2003 <- dtdata\_nozeros$ln\_km\_IH\_1993 + dtdata\_nozeros$ln\_km\_IH\_2003, coeftest(regRes, vcov = vcovHC(regRes, "HC1"))”

<- ln\_vmu\_1993\_2003 <- dtdata\_nozeros$l\_vmt\_IH\_83 + dtdata\_nozeros$l\_vmt\_IH\_83, coeftest(regRes, vcov = vcovHC(regRes, "HC1"))”

<- model\_table5 <- lm(dtdata\_nozeros$ ln\_vmu\_1993\_2003 <- dtdata\_nozeros$l\_vmt\_IH\_83 + dtdata\_nozeros$l\_vmt\_IH\_83

~ ln\_km\_IH\_1993\_2003+ dtdata\_nozeros$D\_lpop, dtdata\_nozeros)