

### ECON 282 Assignment 5

(1) There are four datasets on the website. Please download the dataset which corresponds to your group number. The groups are:

Group 1	Group 2	Group 3	Group 4
Akepanidaworn,Korkrid	Findley,Morgan	Kellner,James	Postman,Andrew
Alatrash, Niveen	Gerner,Brian	Koshakow,Joseph	Rosenberg,Robert
Alvarado,Catherine	Gilson,Julian	Lee,Juho	
Areepipatkul, Paticha	Hardesty,Garrett	Li,Jeanne	Sica,Margaux
Branan,Steven	Holze,Dylan	Liu,Michael	Sneed,David
Chen,Hao	Horenstein,Samuel	Medina, Carlo	Tang,Jonathan
Chin,Taylor	Hoy, Dylan	Mejia,Michael	Tavarez,Gregory
Cohen,Daniel	Huang, Tiger	O'Sullivan,William	Weinhaus,Craig
Enriquez,Jehu		Oujevolk,Quincy	Yun, Kathy

(2) All the datasets have the same 3 variables, commute = total commute time, drive = 1 if person drives to work 0 otherwise, distance = distance from home to work, coffee = 1 if person was interviewed outside a coffee shop 0 otherwise.

(3) We'll break our rules and not use a testing/training setup. Just try running a regression

```
m1 <- lm(commute ~ drive + distance + coffee)
```

(4) Now, check for any selection on the observables drive and coffee. For example, do those who drive have a higher mean for coffee, and so forth. Use t-tests to see if the differences are statistically significant.

(5) Finally, do the results of (4) make you suspicious that there could be unobserved sources of selection bias?