BOSTON COLLEGE
Department of Economics
EC 327 Financial Econometrics
Spring 2014, Prof. Baum, Mr. Zago (grader)
Problem Set 1
Due Wednesday 3 February 2014
Total Points Possible: 135 points

1. (0 pts) Load data.

2. (5 pts) There are 4596 observations in the data set, 16 variables, two of which are string variables (origin and destin). To verify, use the **describe** command.

3. (5 pts) There are 188 observations containing flights which originate in Boston, MA. (Use **count if origin == "BOSTON, MA"**). There are also 20 flights whose destination is Boston.

4. (10 pts) The shortest flight is 95 miles; it is from Cleveland, OH to Detroit, MI. The longest flight is 2724 miles, from Miami, FL to Seattle, WA. To get the minimum and maximum distances you can use **summarize dist**. Then use **list origin destin dist year if dist==95** and **list origin destin dist year if dist==2724** to get the route information.

5. (5 pts) Using **describe bmktshr**, you can see that it gives the market share of the largest carrier in a market. Since it is hopefully a proportion between 0 and 1, use **summarize bmktshr** to check. There is no difference between the two variables "concen" and "bmktshr". You could use **describe concen** to find this information.

6. (10 pts) Use **list origin destin year if bmktshr == 1** to get the following five monopolies: Atlantic City, NJ and Cleveland, OH (2000); Atlantic City, NJ and Myrtle Beach, SC (1997, 1999); and Kansas City, MO and Tulsa, OK (1997, 2000)

7. (5 pts) Use **correlate concen dist ldist** to see that there is a correlation of -0.5283 between concentration and length, as well as a correlation between -0.5319 between concentration and log length, which seems to suggest the airline carrier market is more competitive on longer routes.

8. (10 pts) Using **tabstat fare, by(year)** you can see that the mean fares for 1997-2000 were $173.75, $175.44, $177.97, and $188.02 respectively.

9. (15 pts) To create this new variable rfare, I used **generate rfare=fare/1.615** since to get the actual CPI we need to divide by 100. Then use **replace rfare=fare/1.64 if year == 1998** and similar commands for 1999 and 2000 to get rfare. Finally, use **tabstat rfare, by(year)** to see that real fares fell between 1997 and 1999, but rose slightly from 1997-2000.

10. (10 pts) To create the real cost-per-mile variable, cpm, use **generate cpm=rfare/dist**, then use **tabstat cpm, stat(p25 p50 p75)** to get the quartile values of cpm: 8.36 cents per mile, 10.97 cents per mile, and 16.28 cents per mile.

11. (5 pts) Use the correlation command to see there is a positive correlation between cost per mile and concentration. This does agree with the priors that as markets become more competitive, prices fall.

12. (10 pts) To find the cheapest and most expensive fares in terms of cpm, use **sort cpm** to arrange flights by their cpm. Then use **list cpm origin destin year dist in 1/5** to see that the cheapest flight was between Los Angeles and Providence, RI in 1999 with a cpm of $0.0405. Then use **list cpm origin destin year dist in -5/$\ell$** to see that the most expensive flight was between Cleveland and Detroit in 2000 with a cpm of $1.2274.

13. (5 pts) Use **list origin destin year dist passen if passen > 7500** to see that 3 flights meet these criteria: Ft. Lauderdale to New York (2000) and Los Angeles to New York (1997,1998).

14. (10 pts) Use **regress lpassen year** which yields a coefficient on year of .0295, or 2.95 percent growth annually. This is significant at the 98 percent level.

15. (10 pts) Regress cost per mile on passenger volume, distance, and market concentration. All of the variables are significant at the 99.9 percent level. As average passengers increase, cost per mile decreases. As distance increases, cost decreases. As market concentration increases, cost increases. All of these variables have the expected signs.

16. (10 pts) Run the previous regression with the factor variable `c.dist##c.dist` which will introduce both distance and distance$^2$ in the model. The distance$^2$ variable is significant at the 99.9% level. Adding the distance$^2$ variable increases the $R^2$ considerably, from .3176 to .4485. To interpret the sign on distance, as distance rises, the cost per mile falls. To interpret the sign on distance$^2$, as distance rises, the cost per mile falls more slowly.

17. (5 pts) To compute the average marginal effects, use **margins, dydx(all)**. Then you will get average marginal effects of -.0000132 on passengers, -.0001313 on distance, and .04522 on concentration. All of these average marginal effects are significantly different from zero.

18. (5 pts) Adding an interaction term between passengers and market concentration (`c.passen##c.concen`) and then computing marginal effects as before, we see that the magnitude of the marginal effects for `passen` and `dist` increased while that for `concen` decreased. The interaction term is significant at the 99.99% level.