Stat 89 Lecture 2

Commands used today. Words in italics can or should be substituted with your own.

I. System commands

log using *yourlogfilename*, text

log close

type *yourlogfilename*

log open

II. From last time:

```
use "http://web.stat.ucla.edu/~vlew/stat10/movies2003"
```

1. Which distributor(s) produced the most films in 2003?
   a. tabulate dist, sort
2. Which distributor's films were the highest US grossing, on average?
   a. tabulate dist, summarize(totusgross)
3. What are the mean and median number of theatres used at the "widest" openings?
   a. summarize widest, detail
4. What is the value of the 25^{th} percentile of total US gross? what about the 75^{th}?
   a. summarize totusgross, detail
5. What is the IQR of overseas gross?
   a. summarize overseasgross, detail
6. What is range of overseas gross? (same command as #5)

III. Making New Variables with the GENERATE command

example:

generate daysrun=closedate – opendate
generate totalgross = totusgross + overseasgross
generate ratio = totalgross/(prodcost * 1000000)

IV. Graphs

a. Boxplot

examples:

graph box totalgross, over(quarter)

```
graph box ratio, over(quarter) yline(1.6)
graph box weekend1 weekend2
```

b. Histogram

examples:

histogram opening

```
histogram opening, width(200)
histogram month, discrete xlabel(1 2 to 12)
```

VI. Normal curve related

histogram advert, normal

histogram prodcost, normal

display norm(Z)

display norm(1)
display norm(-1)
display norm(0)
display norm(-.67)
display norm(+1.65)
display norm(5)
display norm(6)

pnorm totalgross
pnorm advert
pnorm daysrun

Extra Stuff (not required) for you

```
use "http://web.stat.ucla.edu/~vlew/stat10/lacoschools"
```

This is a database of all public schools in Los Angeles County in 2003.

1. Among these variables, tell me which one is approximately normal, right skewed, and left skewed.

   wh_num api03 pct_hi

2. Create a box plot which shows the api scores of African Americans, Asians, Hispanics, and Whites on the same graph.
3. Create a box plot which shows the api scores by school type (i.e. Elementary, High and Middle Schools)

4. Choose an approximately normal variable (you can use the one in #1) and have Stata draw a normal curve over it. Approximately what original value would correspond with a Z score of +1.96?