Statistics 104 – Autumn, 2004 – Assignment 1 Addendum

Due Friday, October 1. Homework is handed in at the **Friday's** lecture. Please include your name, your TF's name, and your section time at the top of your assignment.

Computer Assignment (STATA)

As mentioned in class, there are problems with the some of the original Stata code listed in the original version of the assignment. The following code should work with version 8 of Stata. Note that the instruction discussing the use of menus discussed in the original version of the assignment should be fine.

4 To start a log file of your Stata session type

```
log using hw1out
```

in the Stata Command window. To close log file type

```
log close
```

5 Basic summary statistics for all variables can be calculated using the command

```
summarize
```

A summary of the variable player can be given with the command

```
tabulate player
```

To list all the date for observations 1 to 10, type the command

```
list in 1/10
```

Changing the numbers of course will list a different subset of the data.

6 Make graphs of the data. Briefly describe the distribution of homeruns and batting attempts and the relation between the variables. The plots can be generated by

```
histogram homeruns, bin(8) name(histhr)
histogram atbats, bin(8) name(histbat)
scatter homeruns atbats, name(plothrbat)
```

Note: If you omit the `name(`*`filename`*`)` modifier above the plot will still be displayed on the screen, but not saved in a file for later printing. You should omit this modifier when you are exploring data sets, and use it only when you find a plot you wish to print later.

8 To create the homerun rate variable and display it, you could run the commands

```
generate hrrate=homerun/atbats
hist hrrate, bin(8) name(histrate)
```

Note that for many commands only the first few letters need to be given. See help command to what the abbreviation is.

Here is a trick for reducing the amount of paper you use in printing. Graphs can be combined on the same page. If you print the graph named plothw1, you will get the four previous plots together.

```
graph combine histhr histbat plothrbat histrate, name(plothw1)
```

9 To compare the players, the following commands can be used

```
sort player
tabulate player, summ(homerun)
tabulate player, summ(hrrate)
graph box hrrate, over(player)
```

Note that instead the final command `graph box hrrate, over(player)` could be replace by `graph box hrrate, by(player)`, though it gives a different plot and one that probably isn't as appropriate for the desired comparison.

11 Exit Stata and print output.

12 Organize your output and make comments.

A transcript of your session is in the file '`hw1out`', or the file name you chose. Remove any mistakes and other extraneous text. You can do this electronically by editing the file, or by cutting, with scissors, and taping the worthwhile output to standard sized paper. On the graphs, comment on any interesting features you see. (Note that there is no need to include the log file as part of the answer, though you can if desired.)