STAT 101
Assignment 1
Due date: 19 January 2012 at 4:30 in the box outside the Stat Lab.

1. From the text: # 1.30 on page 29. ZZ

2. From the text: # 1.33 on page 31-32.

3. From the text: # 2.25 on page 59.

4. From the text: # 2.30 on page 60.

5. From the text: # 3.41 on page 90.

   NOTE: the notation N(64,2.7) means the mean is 64 and the SD is 2.7.

6. From the text: # 3.36 on page 89.

7. From the text: # 3.37 on page 89.

8. For adults aged 15 to 24 with income in BC in 2000 the census shows an average income of about $10,200 with a standard deviation of approximately $10,100. Make a normal approximation to estimate the proportion of such adults whose income is negative. Is a normal approximation wise here? Explain.

9. Is the median income likely higher or lower than $10,000 in the group in the previous question? Explain your answer.

10. The standard deviation for heights in the data set I collected for a 2004 introductory class is 3.7 inches. Do you expect the standard deviation of height for women to be more or less than 3.7 inches? Explain.

11. The mean height for the same class is 66.8 inches. Make normal approximations to estimate the 90th percentile of heights and the interquartile range.

12. This last question is intended to get you started on doing computing. It uses data from

   http://lib.stat.cmu.edu/datasets/CPS_85_Wages

The file contains data on a sample of 534 people drawn from the 1985
Current Population Survey in the United States. For each person 11
variables were measured:

```
EDUCATION: Number of years of education.
SOUTH: Indicator variable for Southern Region (1=Person lives in  i
     South, 0=Person lives elsewhere).
SEX: Indicator variable for sex (1=Female, 0=Male).
EXPERIENCE: Number of years of work experience.
UNION: Indicator variable for union membership (1=Union member,
     0=Not union member).
WAGE: Wage (dollars per hour).
AGE: Age (years).
RACE: Race (1=Other, 2=Hispanic, 3=White).
OCCUPATION: Occupational category (1=Management,2=Sales, 3=Clerical,
  4=Service, 5=Professional, 6=Other).
SECTOR: Sector (0=Other, 1=Manufacturing, 2=Construction).
MARR: Marital Status (0=Unmarried,  1=Married)
```

In this problem I want you to read in the data, make a histogrm of
EDUCATION and of WAGE, a bar plot of Occupation and a scatterplot
of Education on the x axis against Wage on the y axis.

Here are the steps involved to do this using JMP:

(a) If you have not installed JMP on your computer go to
    http://www.sfu.ca/itservices/technical/software.html
    to download JMP 9. It is free to students and available for MAC or
    Windows. Download and install the software and the license file.
    Then start JMP and browse for the licence file you downloaded.

(b) Download the data in the file CPS.csv to your computer.

(c) In JMP there is a window called "JMP Starter". If you can't see
    it select "JMP Starter" from the menu "Window".

(d) Click on `Open Data Table`, navigate to find the file "CPS.csv"
    which you downloaded and click on `Open`. (When you save the
    file its name *might* get changed to "CPS.txt" but this should not
    matter to JMP.) The file should open up in a window title `CPS`.

Click on `Import as Data`. You should now see a spreadsheet format with columns labelled EDUCATION and so on.

(e) To make the histograms select `Distribution` under `Analyze`. In the window which comes up click on say `EDUCATION` to highlight it then on `Y, Columns` and then OK. By click on various places in hte resulting window you can adjust labels, make the histogram horizontal instead of the vertical default and so on but for this assignment it is enough to just print out the result (under the `File` menu).

(f) To make the bar plot select `Chart` under `Graph`. Click on `Occupation` then select `N` under the `Statistics` drop down list. Then click `OK`. Just print the result.

(g) To make the scatterplot select `Scatterplot Matrix` under `Graph`. Highlight `Education` and click on `Y, Columns` then do the same for `WAGE`. Click `OK` and print the resulting scatterplot.