Lab 7

1. Use the GSS84 data to get separate means and standard deviations for males' and females' incomes. These can be obtained using the following SPSS program:

```
frequencies general = rincome,sex / statistics = mean,variance.
recode rincome (1=500)(2=2000)(3=3500)(4=4500)(5=5500)(6=6500)
temporary.
select if (sex eq 1).
frequencies general = rincome / statistics = mean,variance.
temporary.
select if (sex eq 2).
frequencies general = rincome / statistics = mean,variance.
```

Note that the RECODE statement in the above program is needed to convert the codes in the data file to an interval measure. This can be verified by noting labels printed on output associated with the first "frequencies" command in the above program.

a. Calculate a 99% confidence interval for the difference in mean incomes (males minus females). Is there evidence that the difference in men's and women's incomes is statistically different from zero? (Use the .01 significance level for any hypothesis tests you may conduct while obtaining the confidence interval.) Some help:

\[ F_{477,.01} = 1.24 \]

b. The difference in mean incomes (found in part a) comprises how many standard errors? (Hint: You will need to calculate a value of z.)

c. Test the null hypothesis (at the .001 level) that men's and women's incomes are the same. Do you reject this null hypothesis? (Some help: There is no reason to redo the F-test performed in part a.)

2. You have gained the confidence of a Palestinian terrorist group and are doing an in-depth study of their motivations to perform terrorist acts. Around half of the terrorists in the group had relatives who were killed in the massacre of Palestinian refugees by Christian militia during the Israeli invasion of Lebanon. You also have records on how often each terrorist participated in terrorist acts during the three
years of your study. Consistent with the "Revenge Theory" of terrorism, you believe that terrorism is motivated by an attempt to avenge personal harm that has been done to oneself or to loved ones. Your data are as follows:

Table 1: Mean Number of Terrorist Acts with Variance and Sample Size for Terrorists Who Did versus those Who Did Not Have Relatives Killed in the Massacre.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Variance</th>
<th>Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relatives killed</td>
<td>11.3</td>
<td>12.9</td>
<td>28</td>
</tr>
<tr>
<td>No relatives killed</td>
<td>9.8</td>
<td>15.1</td>
<td>25</td>
</tr>
</tbody>
</table>

a. Do the data provide statistically significant support for your theory? (Use the .05 level of significance throughout.)

b. Assume that the theory is correct and that on average terrorists whose relatives were killed in the massacre take part in one more terrorist act per year than other terrorists. What is the probability of your having made a Type II error? Use the .05 level of significance and the variance estimate(s) from part "a." (Hint: Do not forget that your study lasted three years, not one.)

3. Imagine that you have data on four variables from a large sample (except where n=15 as indicated below) of Ames residents:

ANXIETY--an anxiety measure that takes two values: 1 = High anxiety and 2 = Low anxiety.

ULCER--an indicator of whether or not someone has had medical treatment for a stomach ulcer during the past year: 1 = Had ulcer treatment and 2 = No ulcer treatment.

ASSETS--a measure of how rich the respondent is (in thousands of dollars that the resident has in the bank).

GROUCHY--a measure of how grouchy the resident is (in "grumbles-per-hour" while a one-hour interview was taking place).

You proceed to analyze these data two-variables-at-a-time. In each analysis you wish to use the most appropriate from among the following test statistics:

(1) logistic regression       (5) Wilcoxon test
(2) bivariate regression      (6) Fisher's exact test
(3) t-test for a difference in means (7) binomial distribution
(4) t-test for a difference in proportions
For each of the following research questions write (in the blanks provided) the name of the dependent variable and the number of the appropriate statistical procedure for answering the research question.

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Statistical procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. _____________</td>
<td>Are rich people grouchy?</td>
</tr>
<tr>
<td>b. _____________</td>
<td>Do anxious people get ulcers?</td>
</tr>
<tr>
<td>c. _____________</td>
<td>Are anxious people grouchy?</td>
</tr>
<tr>
<td>d. _____________</td>
<td>Are rich people less anxious than poor people?</td>
</tr>
<tr>
<td>e. _____________</td>
<td>Assuming a sample of n=15, are anxious people grouchy?</td>
</tr>
<tr>
<td>f. _____________</td>
<td>Assuming a sample of n=15, do anxious people get ulcers?</td>
</tr>
</tbody>
</table>

Below please find R and SAS code for problem 1:

```r
# R
# Directions:
# Copy the below R code into the "R Editor" window (accessed by
# selecting "New script" under the "File" pull-down menu), swipe the
# code, and press F5.

# Code:
# read lab5data.txt into "gss"
gss<-read.table('http://www.public.iastate.edu/~carlos/401/labs/lab5data.txt')

# assign new values to rincome so that the data are in dollar units.
gss[gss[,2]==1,2]=500
```
gss[ gss[, 2] == 2, 2] = 2000
  gss[ gss[, 2] == 3, 2] = 3500
  gss[ gss[, 2] == 4, 2] = 4500
  gss[ gss[, 2] == 5, 2] = 5500
  gss[ gss[, 2] == 6, 2] = 6500
  gss[ gss[, 2] == 7, 2] = 7500
  gss[ gss[, 2] == 8, 2] = 9000
  gss[ gss[, 2] == 9, 2] = 12500
  gss[ gss[, 2] == 10, 2] = 17500
  gss[ gss[, 2] == 11, 2] = 22500
  gss[ gss[, 2] == 12, 2] = 35000

# read gss into gssnew without the missing data codes for rincome
# (var2=13 (refused) or 99)

# read gssnew into two new data sets:
# male incomes (i.e., var2 when var3=1 for male)
# female incomes (i.e., var2 when var3=2 for female)
mincome<-gssnew[ gssnew[, 3] == 1, 2]
fincome<-gssnew[ gssnew[, 3] == 2, 2]

# Results
# for males find sample size, mean, and variance on income
length(mincome)
mean(mincome)
var(mincome)

# for females find sample size, mean, and variance on income
length(fincome)
mean(fincome)
var(fincome)

* SAS

* Directions:
  * Copy lab5data.txt into the C-drive's root (i.e., into "C:").
  * Copy the below SAS code into the "Editor" window,
    * and press the button with the figure of a little guy running.

* Code:
  * read lab5data.txt into "gss";
data gss;
infile 'C:\lab5data.txt';
input age rincome sex fear papres16 prestige educ agewed xnorcsiz;
run;

* remove missing data while copying "gss" into "gssnew";
data gssnew;
  set gss;
  if (rincome=13) then delete;
  if (rincome=99) then delete;
run;

* assign values for a new variable called income;
data gssnew;
  set gss;
  if (rincome=1) then income=500;
  if (rincome=2) then income=2000;
  if (rincome=3) then income=3500;
  if (rincome=4) then income=4500;
  if (rincome=5) then income=5500;
  if (rincome=6) then income=6500;
  if (rincome=7) then income=7500;
  if (rincome=8) then income=9000;
  if (rincome=9) then income=12500;
  if (rincome=10) then income=17500;
  if (rincome=11) then income=22500;
  if (rincome=12) then income=35000;
run;

* read gssnew into 2 new data sets for male & female incomes;
data gssmale;
  set gssnew;
  if (sex=2) then delete;
run;
data gssfemale;
  set gssnew;
  if (sex=1) then delete;
run;

* Results (the "univariate" procedure generates mean, variance, and much, much more for income on each data set);
proc univariate data=gssmale;
var income;
run;
proc univariate data=gssfemale;
var income;
run;