Two Sample Nonparametric Tests in SAS

**The Data:** The data file hsb2.csv contains 200 observations from a sample of high school students with demographic information about the students, such as their gender (**female**), socio-economic status (**ses**) and ethnic background (**race**). It also contains a number of scores on standardized tests, including tests of reading (**read**), writing (**write**), mathematics (**math**) and social studies (**socst**).

**The Problem:** You are the head of the state department of education. These 200 observations are a random sample of high school students from your state. Now we want to test whether or not students from private schools have different writing test scores than do students from public schools. The variables are

obs: 200 highschool and beyond (200 cases)

vars: 12

variable variable

name type about the variable

-----------------------------------------------------------------------------

id scale student id

female nominal (0 if male, 1 if female)

race nominal ethnicity (1=hispanic 2=asian 3=african-amer 4=white) ses ordinal (1=low 2=middle 3=high)

schtyp nominal type of school (1=public 2=private)

prog nominal type of program (1=general 2=academic 3=vocational) read scale standardized reading score

write scale standardized writing score

math scale standardized math score

science scale standardized science score

socst scale standardized social studies score

**Enter the Data:**

**data** scores;

infile 'C:Desktop\hsb2.csv' dlm = ',' firstobs=**2**;

input id female race ses schtyp prog read write math science socst;

/\* We are inputting all of the variables from the CSV file \*/

**run**;

**proc** **sort** data=scores; by schtyp; **run**;

/\* We need to sort the data by school type in order to \*/

/\* obtain descriptive statistics for each type using Proc Univariate \*/

**proc** **print** data=scores; **run**;

/\* Always, always, print your data to make sure it is in SAS correctly!\*/

Once you have your data into SAS, answer the following questions.

1. What is the population from which the sample was drawn?
2. What are the null and alternative hypotheses for the test you will be conducting?

**Graphics in SAS**

The first thing you want to do is get some descriptive statistics in order to determine what test is appropriate. We will use PROC UNIVARIATE and the ODS feature to get nice looking plots and tables. The ODS procedure can take a while to run. Be patient.

ODS graphics on;

ODS rtf;

**proc** **univariate** data=scores /\* No semicolon yet ... \*/

plot /\* Produces stem-and-leaf and boxplots in line printer format \*/

var write;

by schtyp;

histogram; /\* Produces a histogram in ODS format \*/

qqplot; /\*Produces a QQ plot in ODS format \*/

**run**;

ODS rtf close;

ODS graphics off;

**quit**;

1. Examine the histogram, boxplot, and Normal QQplot. What do they tell you about the distribution of the writing scores for each group (males and females)?
2. What are the values for the mean, median, and mode? What does their relationship tell you about the distribution of the writing scores for each group?

Our main objective is to test the differences in writing scores between private and public schools. The following code uses the Wilcoxon rank-sum test to examine the differences.

**proc** **npar1way** data=scores wilcoxon;

class schtyp;

var write;

**run**;

1. Write a sentence or two explaining the results. In your explanation, you should mention the null and alternative hypothesis of the test, the p-value, the decision, and what that decision means in plain language.
2. Do you have any concerns about the internal or external validity of the analysis?
3. Repeat questions 3-6 for math scores.

Now look at differences in math and writing scores by race.

**proc** **npar1way** data=words wilcoxon;

class race;

var write;

exact Wilcoxon/mc;

**run**;