Two Sample Nonparametric Tests in SAS

1. The carapace lengths (in mm) of crayfish were recorded for samples from two sections of a steam in Kansas.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Section 1 | 5 | 11 | 16 | 8 | 12 |  |
| Section 2 | 17 | 14 | 15 | 21 | 19 | 13 |

The following code uses the Wilcoxon rank-sum test, the van der Waerden scores test, the Siegel-Tukey test, the Ansari-Bradley test, and the Kolmorgorov-Smirnov test to examine the differences.

**proc** **npar1way** data=scores wilcoxon VW Savage ST AB EDF;

class schtyp;

var write;

**run**;

Multi-Sample Data

Description:

Thirty magazines were ranked by educational level of their readers. The magazines were divided into three groups of 10 each, with the groups being determined by the educational level of their readers. Group 1 is the highest educational level, Group 2 a medium education level, and Group 3 the lowest educational level. Three magazines were randomly selected from the first, second, and third set of ten magazines. Six advertisements were randomly selected from each of the nine selected magazines. The magazines were Group 1 were 1. Scientific American 2. Fortune 3. The New Yorker. For Group 2 the magazines were 4. Sports Illustrated 5. Newsweek 6. People. For group 3 the magazines were 7. National Enquirer 8. Grit and 9. True Confessions. For each advertisement, the data below were observed. The variable names and descriptions are as follows:

WDS = number of words in advertisement copy

SEN = number of sentences in advertising copy

SYL3 = number of 3+ syllable words in advertising copy

MAG = magazine (1 through 9 as above)

GRP = educational level (as above)

We want to examine differences in the number of words for the various magazines.

Reference: F.K. Shuptrine and D.D. McVicker, "Readability Levels of Magazine Ads," Journal of Advertising Research, 21:5 (October 1981), p 47.

Here are the data and the code.

**data** words;

input WDS SEN SYL3 MAG GROUP @@;

datalines;

205 9 34 1 1 203 20 21 1 1

229 18 37 1 1 208 16 31 1 1

146 9 10 1 1 230 16 24 1 1

215 16 39 2 1 153 9 10 2 1

205 11 17 2 1 80 13 18 2 1

208 22 32 2 1 89 16 17 2 1

49 5 3 3 1 93 18 10 3 1

46 6 6 3 1 34 6 5 3 1

39 5 6 3 1 88 12 6 3 1

191 25 13 4 2 219 17 22 4 2

205 23 25 4 2 57 7 3 4 2

105 10 5 4 2 109 9 2 4 2

82 10 9 5 2 88 10 3 5 2

39 5 0 5 2 94 11 4 5 2

206 18 29 5 2 197 18 26 5 2

68 9 5 6 2 44 9 5 6 2

203 17 24 6 2 139 13 15 6 2

72 11 3 6 2 67 7 8 6 2

162 14 16 7 3 31 6 9 7 3

85 11 10 7 3 111 12 3 7 3

88 11 12 7 3 60 15 10 7 3

97 6 10 8 3 169 12 11 8 3

78 14 12 8 3 68 11 13 8 3

32 4 1 8 3 50 9 9 8 3

208 18 43 9 3 81 12 13 9 3

83 12 14 9 3 195 17 32 9 3

111 11 24 9 3 208 20 15 9 3

;

**run**;

**proc** **print** data=words; **run**;

Next, we would like to examine any differences in the number of words in advertising copy using nonparametric tests. The following code does a permutation test on the scores, a Kruksal-Wallis test, and a test using Van der Waerdan scores.

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

class group;

var mag;

exact wilcoxon / mc;

**run**;

Unfortunately, Proc Npar1way does not do multiple comparisons on ranks. We have to trick SAS. The following code assigns ranks to the data (proc rank) and then uses proc ANOVA to perform Fisher, Tukey, and Bonferroni procedures on the ranked data.

\* ranking the data ;

**proc** **rank** data = words out = rkwords ;

var wds ;

ranks rwds ;

**run** ;

\* ANOVA on both the raw and ranked data ;

**proc** **anova** data = rkwords ;

class mag ;

model rwds = mag ;

means mag / bon tukey lsd ;

**run** ;

Do fewer words in a magazine advertising copy necessarily mean that the advertisers are targeting an uneducated audience? Explain, using the results of this lab. For example, what might explain the differences in the number of words used between Scientific American (magazine 1) and The New Yorker (magazine 3)?