Categorical Data Analysis
Lab material #3

Sometimes, both the row variable and the column variable are ordinally scaled. This is common when you are studying responses that are evaluated on an ordinal scale and what is being compared are different dosage levels, which are also ordinally scaled. Consider the data in Table 1. A water treatment company is studying water additives and investigating how they affect clothes washing. The treatments studied were no treatment (plain water), the standard treatment, and a double dose of the standard treatment, called super. Washability was measured as low, medium and high.

Table 1: Washability Data

	Washability			
Treatment	Low	Medium	High	Total
Water	27	14	5	46
Standard	10	17	26	53
Super	5	12	50	67

Measures of association are produced in the PROC FREQ output by specifying MEASURES as an option in the TABLES statement. The following statements produce measures of association for the washability data. Using SCORES=RANK on the second TABLES statement specifies that rank scores are to be used in calculating Pearson's correlation coefficient to produce a Spearman's rank correlation coefficient.

```
* example3_1.sas;
data wash;
  input treatment $ washblty $ count @@;
  cards;
  water low 27 water medium 14 water high 5
  standard low 10 standard medium 17 standard high 26
  super low 5 super medium 12 super high 50
  ;
proc freq order=data;
  weight count;
  tables treatment*washblty / measures noprint;
  tables treatment*washblty / measures scores=rank noprint;
run;
```

Output 1 contains the table produced by the first PROC FREQ invocation. All of the measures of ordinal association indicate a positive association. Note also that the Somer's D statistics, Kendall's tau-b, and Stuart's tau-c all have smaller values than gamma. Somer's D statistic has two forms: Somer's D C—R means that the column variable is considered the dependent, or response, variable and Somer's D R—C means that the row variable is considered the response variable.

Output 1 Measures of Association

```
    Statistics for Table of treatment by washblty

Statistic                              Value       ASE
------------------------------------------------------
Gamma                                 0.6974    0.0636
Kendall’s Tau-b                       0.4969    0.0553
Stuart’s Tau-c                        0.4803    0.0545

Somers’ D C|R                         0.4864    0.0542
Somers’ D R|C                         0.5077    0.0572

Pearson Correlation                   0.5538    0.0590
Spearman Correlation                  0.5479    0.0596

Lambda Asymmetric C|R                 0.2588    0.0573
Lambda Asymmetric R|C                 0.2727    0.0673
Lambda Symmetric                      0.2663    0.0559

Uncertainty Coefficient C|R           0.1668    0.0389
Uncertainty Coefficient R|C           0.1609    0.0372
Uncertainty Coefficient Symmetric     0.1638    0.0380

                  Sample Size = 166
```

Printed next to each statistics is the asymptotic standard error (ASE). Although the measure of association is always valid, these standard errors are only valid if the sample size is large. If the sample size is adequate, then the measure of association is approximately normally distributed and you can form the confidence intervals of interest. For example,

$$\text{measure} \pm 1.96 \times \text{ASE}$$

forms the bounds of a 95 percent confidence interval.

Output 2 contains the output produced by the second PROC FREQ invocation. The only difference is that rank scores were used in the calculation of Pearson’s correlation coefficient. When rank scores are used, Pearson’s correlation coefficient is equivalent to Spearman’s correlation, as illustrated in the output. (However, the asymptotic standard errors are not equivalent.)

Output 2 Rank Scores for Pearson’s Correlation

```
                  The FREQ Procedure

    Statistics for Table of treatment by washblty

Statistic                              Value       ASE
------------------------------------------------------
Gamma                                 0.6974    0.0636
Kendall’s Tau-b                       0.4969    0.0553
Stuart’s Tau-c                        0.4803    0.0545

Somers’ D C|R                         0.4864    0.0542
Somers’ D R|C                         0.5077    0.0572

Pearson Correlation (Rank Scores)     0.5479    0.0591
Spearman Correlation                  0.5479    0.0596

Lambda Asymmetric C|R                 0.2588    0.0573
Lambda Asymmetric R|C                 0.2727    0.0673
Lambda Symmetric                      0.2663    0.0559

Uncertainty Coefficient C|R           0.1668    0.0389
Uncertainty Coefficient R|C           0.1609    0.0372
Uncertainty Coefficient Symmetric     0.1638    0.0380

                  Sample Size = 166
```

If you don't specify SCORE=, then you get the default table scores. The column (row) numbers are the table scores for character data and the actual variable are used as scores for numeric variables. Other SCORE= values are RANK, MODRIDIT, and RIDIT.