

Using SAS to Calculate Tests of Cliff's Delta

Kristine Y. Hogarty and Jeffrey D. Kromrey

Department of Educational Measurement and Research, University of South Florida

ABSTRACT

This paper discusses a program that calculates three estimates for making statistical inferences regarding the delta statistic. The delta statistic (Cliff, 1993) may be used when testing null hypotheses about group differences on ordinal level measurements. This statistic and the inferential methods associated with it are addressed by considering data arranged in a dominance matrix. The program calculates the marginal values of the dominance matrix that are used in the inferential statistics associated with d (sample estimate of the population parameter). The three methods of inference for d include two estimates of the variance of d , and a confidence band around the sample value of d . The paper provides a demonstration of the SAS/IML® code and examples of application of the code to simulation studies in statistics.

INTRODUCTION

A variety of options are available to researchers interested in examining response variables that are measured as ordered categories, such as Likert scale and other rating scale items. A Pearsonian chi-square test of homogeneity or a test for the equality of population means such as the independent means t-test are commonly employed when testing for the equality of two groups on such a response variable. Implicit in the former analysis is the treatment of the response variable as nominal-level measurement, while the latter analysis implies an assumption of interval-level data. Ordinal indices of association fall between these two extremes, but are less frequently seen in applied research (Cliff, 1996a).

Norman Cliff (1993, 1996a) has proposed the use of the delta statistic for testing null hypotheses about group differences on ordinal level measurements.

The delta statistic is used to test equivalence of probabilities of scores in each group being larger than scores in the other (the property that Cliff (1993) referred to as "dominance"). Cliff (1996a) states several reasons for preferring ordinal analyses over the more familiar methods. First, ordinal measures of association such as Tau and delta are useful both descriptively and inferentially because of their robustness properties when compared to traditional parametric tests such as the independent means t-test. Second, many variables in behavioral research can only be given ordinal scale status, and ordinal statistics are invariant under monotonic transformations, whereas those based on certain parametric analysis may not be. Lastly, ordinal statistics may provide answers more directly aligned with the research questions under investigation.

In a recent Monte Carlo study, Kromrey and Hogarty (1998) compared the Type I error control and statistical power of four tests of group differences on ordered categorical response data: a parametric test of mean differences (independent means t-test), the Pearsonian chi-square test of homogeneity, the cumulative logit model recommended by Agresti (1989, 1996), and the delta statistic recommended by Cliff (1993, 1996a). The number of categories of the response variable, sample size, population distribution shape, and effect size were examined. The findings suggest, that under certain conditions (e.g., skewed marginal distributions, 7-point response scale), the cumulative logit models and the tests of delta tend to be more powerful than the other tests. These initial findings suggest that for many data conditions, the choice of an appropriate test statistic is vitally important to the validity of research inferences.

AN EXAMPLE

This program was designed to build a dominance matrix; calculate the sample estimate of the delta statistic (d); calculate both a consistent and an unbiased estimate of the variance of d; and to construct three asymmetric confidence intervals (i.e., 90%, 95%, and 99%). These statistics will be presented in reference to the set of data presented in Table 1.

Table 1
Sample of Two Groups' Responses to a 5-Point Likert Item

Control Group	Experimental Group
1	1
1	2
2	3
2	4
2	4
3	5
3	
3	
4	
5	

These data, consisting of responses to a 5-point Likert item, were obtained from six members of an experimental group and ten members of a control group. The research question to be addressed is whether the two populations from which the samples were obtained differ in their response to this item.

Norman Cliff (1993, 1996a) has proposed the use of the delta statistic for testing null hypotheses about group differences on ordinal level measurements. The population parameter for which such tests are intended is the probability that a randomly selected member of one population has a higher response than a randomly selected member of the second population, minus the reverse probability. That is,

$$\text{delta} = \Pr(x_{i1} > x_{j2}) - \Pr(x_{i1} < x_{j2})$$

where x_{i1} is a member of population one and x_{j2} is a member of population two.

A sample estimate of this parameter can be obtained by enumerating the number of occurrences of a sample one member having a higher response value than a sample two member, and the number of occurrences of the reverse. This gives the sample statistic

$$d = \frac{\#(x_{i1} > x_{j2}) - \#(x_{i1} < x_{j2})}{n_1 n_2}$$

This statistic, and inferential methods associated with it, are readily addressed by considering the data in an arrangement called a dominance matrix. This n_1 by n_2 matrix has elements taking the value of 1 if the row response is larger than the column response, -1 if the row response is less than the column response, and 0 if the two responses are identical. The sample value of d is simply the average value of the elements in the dominance matrix. The dominance matrix for the Table 1 data is presented in Table 2.

Table 2
Dominance Matrix for the Sample Data

	1	2	3	4	4	5	di.
1	0	-1	-1	-1	-1	-1	-0.833
1	0	-1	-1	-1	-1	-1	-0.833
2	1	0	-1	-1	-1	-1	-0.0500
2	1	0	-1	-1	-1	-1	-0.0500
2	1	0	-1	-1	-1	-1	-0.0500
3	1	1	0	-1	-1	-1	-0.167
3	1	1	0	-1	-1	-1	-0.167
3	1	1	0	-1	-1	-1	-0.167
4	1	1	1	0	0	-1	0.333
5	1	1	1	1	1	0	0.833
dj	0.8	0.3	-0.3	-0.7	-0.7	-0.9	-0.250

The row and column marginals of this table provide mean values of the elements in the respective rows and columns of the matrix. These marginals are used in the inferential statistics associated with d. The null hypothesis tested in such inferential statistics (representing no relationship between the grouping variable and the response variable) is that delta is equal to zero.

Cliff (1996b) presented three methods of inference for d . The first method uses an "unbiased" estimate of the variance of d . This estimate is given by

$$S_d^2 = \frac{n_2^2 \sum (d_i - d)^2 + n_1^2 \sum (d_j - d)^2 - \sum \sum (d_{ij} - d)^2}{n_2 n_1 (n_2 - 1)(n_1 - 1)}$$

where d_i is the marginal value of row i , d_j is the column marginal of column j , and d_{ij} is the value of element ij in the matrix. For the sample data in Table 1, the value of d is -0.25 and the value of S_d^2 is 0.098. The square root of this variance is used as the denominator of a z statistic: $z_{\text{unbiased}} = d / S_d$

For the sample data, the value of z is -0.798, yielding a probability under the null hypothesis of 0.425. The unbiased test fails to reject the null hypothesis of $\delta = 0$.

The second method of inference for d uses a "consistent" estimate of the variance:

$$S_{dc}^2 = \frac{(n_2 - 1)S_{di}^2 + (n_1 - 1)S_{dj}^2 + S_{dij}^2}{n_1 n_2}$$

where $S_{di}^2 = \sum (d_i - d)^2 / (n_1 - 1)$,
 $S_{dj}^2 = \sum (d_j - d)^2 / (n_2 - 1)$, and
 $S_{dij}^2 = \sum \sum (d_{ij} - d)^2 / [(n_1 - 1)(n_2 - 1)]$.

As with the "unbiased" estimate of variance, the square root of this "consistent" estimate of the variance of d can be used as the denominator of a z statistic:

$$z_{\text{consistent}} = d / S_{dc}$$

For the Table 1 data, the value of S_{dc}^2 is 0.106, yielding a value for $z_{\text{consistent}}$ of -0.768, with a probability under the null hypothesis of 0.443. The conclusion with this sample is the same as that reached with the unbiased test, that is, a failure to reject the null hypothesis of $\delta = 0$. The final method of inference regarding d uses S_{dc} to construct an asymmetric confidence interval around the sample value of d . When such an interval does not include the value of zero, the null hypothesis of $\delta = 0$ can be

rejected. The limits of this asymmetric confidence interval are given by

$$\frac{d - d^3 \pm Z_{\alpha/2} S_{dc} [(1-d^2)^2 + Z_{\alpha/2}^2 S_{dc}^2]^{1/2}}{1 - d^2 + Z_{\alpha/2}^2 S_{dc}^2}$$

where $Z_{\alpha/2}$ is the normal deviate corresponding to the $(1 - \alpha/2)^{\text{th}}$ percentile of the normal distribution.

For the Table 1 data, the lower limit of the 95% confidence interval is -0.713, and the upper limit is 0.364. Because this interval contains the value of zero, the null hypothesis is not rejected at the .05 level. Cliff (1996a) has pointed out that the well-known Mann-Whitney Wilcoxon statistic can also provide a test of $\delta = 0$ (because d and U are related by $d = 2U/[n_1 n_2 - 1]$). However, the rank test is not recommended by Cliff because it is actually testing for the equivalence of the two groups' distributions rather than focusing on the parameter δ .

MACRO

```
%macro cliffd(testdata,grpvar,dvscore);
```

```
proc iml;
  use &testdata;
  read all var {&dvscore} where (&grpvar=1) into A;
  read all var {&dvscore} where (&grpvar=2) into B;
  NA=NROW(A);
  NB=NROW(B);
```

```
START CLIFF(A,B,NA,NB);
  SCORE={&dvscore};
  GRP = {&grpvar};
```

```
DOM_MTRX = J(NA,NB,0);
  do i = 1 to NA;
    do j = 1 to NB;
      if A[i,1] > B[j,1] then DOM_MTRX[i,j] = 1;
      if A[i,1] < B[j,1] then DOM_MTRX[i,j] = -1;
    end;
  end;
```

```
*print DOM_MTRX;
```

```
di = DOM_MTRX[+,+] # (1/NB);
dj = DOM_MTRX[+,+] # (1/NA);
d = DOM_MTRX[+,+] # (1/(NA#NB));
```

```

di = (di - d # J(NA,1,1)) ##2;
sdi = di[+,]#(1/(NA-1));

dj = (dj - d # J(1,NB,1)) ##2;
sdj = dj[+,]#(1/(NB-1));

dij = (DOM_MTRX - d # J(NA,NB,1))##2;
sdij = dij[+,+] # (1/((NA-1)#(NB-1)));

* +-----+
  "unbiased" estimate
+-----+;
sd2 = (NB##2#di[+,] + NA##2#dj[+,] - dij[+,]) /
      (NA#NB#(NA-1)#(NB-1));

sd2 = ((NB#sdi)/(NA#(NB-1))) + ((NA#sdj)/
      (NB#(NA-1))) - (sdij/(NA#NB));
if sd2 < ((1 - d##2)/(NA#NB-1)) then
  sd2 = ((1 - d##2)/(NA#NB-1));

z1 = d / sqrt(sd2);
prob_z1 = (1 - probnorm(abs(z1)))#2;

* +-----+
  "consistent" estimate
+-----+;
sd2c = ((NB-1)#sdi + (NA-1)#sdj + sdij) / (NA#NB);

z2 = d / sqrt(sd2c);
prob_z2 = (1 - probnorm(abs(z2)))#2;

* +-----+
  Asymmetric confidence intervals
+-----+;
CI_95U = (d - d##3 + 1.96#sqrt(sd2c) #
  sqrt((1 - d##2)##2 + 1.96##2#sd2c)) /
  (1 - d##2 + 1.96##2#sd2c);
CI_95L = (d - d##3 - 1.96#sqrt(sd2c) #
  sqrt((1 - d##2)##2 + 1.96##2#sd2c)) /
  (1 - d##2 + 1.96##2#sd2c);
CI_90U = (d - d##3 + 1.645#sqrt(sd2c) #
  sqrt((1 - d##2)##2 + 1.645##2#sd2c)) /
  (1 - d##2 + 1.645##2#sd2c);
CI_90L = (d - d##3 - 1.645#sqrt(sd2c) #
  sqrt((1 - d##2)##2 + 1.645##2#sd2c)) /
  (1 - d##2 + 1.645##2#sd2c);
CI_99U = (d - d##3 + 2.576#sqrt(sd2c) #
  sqrt((1 - d##2)##2 + 2.576##2#sd2c)) /
  (1 - d##2 + 2.576##2#sd2c);
CI_99L = (d - d##3 - 2.576#sqrt(sd2c) #
  sqrt((1 - d##2)##2 + 2.576##2#sd2c)) /
  (1 - d##2 + 2.576##2#sd2c);

file print;
put @20 'Tests of d'/

```

```

@10 '_____ '//
@10 'Dependent Variable=' @35 SCORE //
@10 'Grouping Variable=' @35 GRP //
@10 'Sample Sizes=' @34 NA 3. @40 NB 3. //
@10 'Sample Value of d=' @ 35 d 8.6 //
@10 'Confidence Bands' @30 'Lower Limit'
@50 'Upper Limit' /
@30 '_____ ' @50 '_____ ' /
@15 '90%' @31 CI_90L 8.6 @51 CI_90U 8.6 /
@15 '95%' @31 CI_95L 8.6 @51 CI_95U 8.6 /
@15 '99%' @31 CI_99L 8.6 @51 CI_99U 8.6 //
@10
'_____ '//
@10 'Tests of delta=0:' @37 'Z'
@48 'Two-tailed probability' /
@10 '_____ ' @34 '_____ '
@47 '_____ '//
@12 'Consistent' @34 z2 6.3 @52 prob_z2 8.6 /
@12 'Unbiased' @34 z1 6.3 @52 prob_z1 8.6 /;

FINISH;
run cliff(A,B,NA,NB);
quit;
%mend cliffd;
data one;
input group score;
cards;
1 1
1 1
1 2
1 2
1 2
1 2
1 3
1 3
1 3
1 4
1 5
2 1
2 2
2 3
2 4
2 4
2 5
;
%cliffd(one,group,score)
run;

```

OUTPUT

Table 3
Tests of delta

Tests of d		
Dependent Variable=	SCORE	
Grouping Variable=	GROUP	
Sample Sizes=	10 6	
Sample Value of d=	-.250000	
Confidence Bands	Lower Limit	Upper Limit
90%	-.663806	0.280995
95%	-.712540	0.364009
99%	-.785832	0.500230
Tests of delta=0:	Z	Two-tailed probability
Consistent	-0.768	0.442688
Unbiased	-0.798	0.424656

CONCLUSION

As stated earlier, under certain circumstances, tests of delta have been found to be more powerful than some of the more commonly used tests such as the independent means t-test or the Pearsonian chi-square. This macro is provided to facilitate the calculation of estimates of the delta statistic (d), both the unbiased and consistent estimates of the variance of d; and to construct asymmetric confidence intervals.

REFERENCES

Agresti, A. (1989). Tutorial on modeling ordered categorical response data. Psychological Bulletin, *105*, 290-301.

Agresti, A. (1996). An introduction to categorical data analysis. New York: Wiley.

Cliff, N. (1993). Dominance statistics: Ordinal analyses to answer ordinal questions. Psychological Bulletin, *114*, 494-509.

Cliff, N. (1996a). Answering ordinal questions with ordinal data using ordinal statistics. Multivariate Behavioral Research, *31*, 331-350.

Cliff, N. (1996b). Ordinal methods for behavioral data analysis. Hillsdale, NJ: Erlbaum.

Kromrey, J. D., & Hogarty, K. Y. (1998). Analysis Options for Testing Group Differences on Ordered Categorical Variables: An Empirical Investigation of Type I Error Control and Statistical Power. Multiple Linear Regression Viewpoints, *25*, 70-82.

SAS/IML is a registered trademark of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

CONTACT INFORMATION

The authors can be contacted at the University of South Florida, Department of Educational Measurement and Research, FAO 100U, 4202 E. Fowler Ave, Tampa, FL 33620, by telephone at (813) 973-3220, or by e-mail: khogarty@luna.cas.usf.edu or kromrey@typhoon.coedu.usf.edu