

Multiple Comparisons on 2xc proportions

John Horne, Dave Plaehn
InsightsNow, Inc. Corvallis, OR (USA)

ABSTRACT

The Freeman-Tukey double-arcsine transformation is used to transform binomial or Poisson data so that they correspond with probabilities under the standard normal. This permits use of the normal distribution for significance testing on proportions. Tukey's HSD and other "post hoc" multiple comparison tests have traditionally been used to test differences among groups of means. The double-arcsine transformation allows for the use of these to test differences among proportions in 2xc contingency tables. This paper presents a Base SAS® macro that will run Tukey's HSD, LSD, Bonferroni, Dunn-Sidak and Scheffé tests on groups of proportions having a 2xc structure. Simulation testing showed that each of these tests produced more conservative results and protected family-wise error rates much closer to alpha risk than pairwise χ^2 tests of association on the same proportions. The macro permits inputs as a 2xc counts table or as proportions and their associated sample sizes.

INTRODUCTION

Multiple comparisons (MC) tests are used to make pairwise tests of significance among means or proportions in experiments containing more than two groups. A general test of significance, such as ANOVA when comparing means or a χ^2 test of association when comparing proportions, is useful for determining whether there are differences between *any* means or proportions. These tests are not useful for making determinations of significance between individual means or proportions. MC tests are employed for the latter.

A critical feature of MC tests is that they manage *family-wise* error rates. Sometimes called experiment-wise error rates, these are the probabilities of concluding one or more significant differences between means or proportions in an experiment when no such differences exist. Family-wise error rates are a function of Type I error or *comparison-wise* error rates (α_c) and the number of pairwise comparisons in an experiment. Each of c means or proportions can be used in up to $c-1$ pairwise comparisons. The maximum number of pairwise comparisons that may be made in an experiment is therefore $c(c-1)/2$ and family-wise error rates may be estimated by

$$\alpha_f = 1 - (1 - \alpha_c)^k,$$

where α_f is the family-wise error rate, α_c is the comparison-wise error rate, and $1 \leq k \leq c(c-1)/2$ is the number of comparisons made in the experiment.

Suppose we ran 100 experiments, each having six groups; we could make up to 1500 pairwise comparisons among these groups. If we protected comparison-wise error risk at .05, we would expect to find 75 significant differences distributed across about 54 of the 100 experiments if the null hypothesis of no differences was always true, all comparisons were independent of all others and all possible pairwise comparisons were made. Finding differences in this many experiments when no differences actually exist is likely an unacceptable risk. MC tests minimize this risk by protecting family-wise error rates instead of comparison-wise error rates. For instance, if we set α_c in the above example at .0034, we would expect to find only about five significant differences among the 1500 pairwise comparisons, and we would expect these differences to be distributed across five different experiments, making α_f .05.

Most MC tests rely on normal probabilities. Means tend to be normally distributed and normal probabilities may accurately be applied when testing mean differences. Proportions on the other hand tend to follow a binomial. The normal distribution $N[np, np(1-p)]$ is a good approximation of the binomial $B[n, p]$ if n is large enough and p is not too close to the margins. Ross, 1989, for instance, recommended using the normal approximation to the binomial any time $np(1-p) \geq 10$. Based on the central limit theorem and the null hypothesis (H_0) of no difference between proportions, we can use the following standard normal random variable as an approximation to the binomial $p_i - p_j$,

$$z = \frac{p_i - p_j}{\sqrt{\text{Var}(p_i - p_j)}}, \text{ with } E(p_i - p_j) = 0 \text{ from } H_0.$$

The denominator of this statistic, also called the standard error (SE) is often estimated as

$$SE = \sqrt{\frac{p_i(1-p_i)}{n_i} + \frac{p_j(1-p_j)}{n_j}}.$$

A second approach is to estimate SE using a pooled proportion p_{ij} , where,

$$p_{ij} = \frac{n_i p_i + n_j p_j}{n_i + n_j} \quad \text{and} \quad SE = \sqrt{p_{ij}(1-p_{ij}) \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}.$$

This latter construction of SE makes the z -test equivalent to a pairwise χ^2 test of association from a 2×2 contingency table. Because the random variable z has a mean of 0 and a variance of 1, normal probabilities may be applied, and the reasoning behind MC tests ought to apply as well. However, it may still be inappropriate to expect this random variable to be normally distributed due to skewness of $p_i - p_j$ when n is small and/or p is near the margins. Skewness of $p_i - p_j$ can be estimated as the difference of the 3rd central moments divided by the sum of the variances to the 3/2 power, i.e.,

$$\gamma = \left[\frac{p_i(1-p_i)(1-2p_i)}{n_i^2} - \frac{p_j(1-p_j)(1-2p_j)}{n_j^2} \right] / \left[\frac{p_i(1-p_i)}{n_i} + \frac{p_j(1-p_j)}{n_j} \right]^{3/2}$$

As both n_i and $n_j \rightarrow \infty$, skewness of $p_i - p_j \rightarrow 0$, but for a finite n_i or n_j , skewness becomes an issue as $p_{ij} \rightarrow 1$ or $p_{ij} \rightarrow 0$. Skew at the margins calls into question the applicability of normal probabilities to tests of differences between proportions and has led to development of methods intended to correct for it. Sokal and Rohlf, 1995 recommend an angular transformation when proportions fall outside of the interval [.3 .7]. Freeman and Tukey, 1950 presented an averaged double arcsine transformation (see below), similar to earlier transformations given by Bartlett, 1936 and Curtiss, 1943. The Freeman-Tukey (FT) transformation was claimed to stabilize variance of the transformed binomial within $\pm 6\%$ of

$$\frac{1}{n+.5} \text{ (angles in radians),} \quad \frac{1}{n+.5} \left(\frac{180}{2\pi} \right)^2 \text{ (angles in degrees),}$$

for almost all cases where the smaller of np , $n(1-p) \geq 1$. These observations suggest that such a transformation might be useful in making MC tests on proportions. Zar, 1999 did just that, using the Freeman-Tukey transformation and variance estimate to apply Tukey's HSD test to a group of proportions. We are aware of other, more recent uses of the same method. Elliott and Reisch, 2006, for example, published a Base SAS macro that applied Tukey's HSD test to proportions using the FT transformation and variance estimate.

This paper presents a new Base SAS macro that extends this body of work to an unrestricted number of groups and finite degrees of freedom. The macro presented here also applies the FT transformation and variance estimate to four other MC tests in addition to Tukey's HSD test. Results from simulations, employing the macro on several different random variables, are presented to show the efficacy of using the FT transformation and variance estimate for hypothesis testing of differences among proportions.

METHOD

Consider the data in Table 1. These represent counts on a binary variable ($bin=1$, $bin=0$) for six different treatment groups.

Treatment	1	2	3	4	5	6
Count ($bin=1$)	34	39	43	32	19	23
Count ($bin=0$)	66	66	51	68	84	75
N	100	105	94	100	103	98
Proportions ($bin=1$)	0.34	0.3714	0.4574	0.32	0.1845	0.2347

Table 1. Hypothetical proportions data.

The SAS macro named MCPNPN (full code below) expects either a columnar SAS data set containing counts or a columnar SAS data set containing proportions and associated sample sizes. The counts data set may be created with the following code:

```
data counts;
do trt=1 to 6;
do bin=1 to 0 by -1;
input ct @@; output;
end;
end;
datalines;
34 66 39 66 43 51
32 68 19 84 23 75
;
```

```
run;
```

and the data set containing proportions may be created with this code:

```
data proportions;
do trt=1 to 6; input n prop @@; output; end;
datalines;
100 .34 105 .3714 94 .4574
100 .32 103 .1845 98 .2347
;
run;
```

If data are entered as proportions, the first step in mcpnp is to convert the data back to a counts table. Rounding is used to ensure that the counts table contains only integers. Care must be taken when entering data as proportions that there is sufficient precision to resolve to the correct integer. Once the macro has a counts table, it uses the FREQ procedure to organize the data into a 2x6 structure and calculate an overall χ^2 test of association.

Output from this step on the data in Table 1, which is suppressed by the 'noprint' option in mcpnp, is shown in Fig. 1. The overall χ^2 test of association from this table resulted in a χ^2 statistic of 21.6835, which has a probability of .0006 when associated with 5 degrees of freedom. There are clearly differences to be found among these proportions. The purpose of the remaining calculations is to determine which are different from which.

bin		trt						Total
Frequency	Col Pct	1	2	3	4	5	6	
1		34 34.00	39 37.14	43 45.74	32 32.00	19 18.45	23 23.47	190
0		66 66.00	66 62.86	51 54.26	68 68.00	84 81.55	75 76.53	410
Total		100	105	94	100	103	98	600

Statistics for Table of bin by trt				
Statistic	DF	Value	Prob	
Chi-Square	5	21.6835	0.0006	
Likelihood Ratio Chi-Square	5	22.1612	0.0005	
Mantel-Haenszel Chi-Square	1	9.9219	0.0016	
Phi Coefficient		0.1901		
Contingency Coefficient		0.1868		
Cramer's V		0.1901		

Sample Size = 600

Figure 1. 2x6 table from proc FREQ on the data shown in Table 1.

The next calculation step in mcpnp is to transform each proportion to degrees using the FT transformation,

$$p_i = \frac{180}{2\pi} \left[\sin^{-1} \sqrt{\frac{x_i}{n_i + 1}} + \sin^{-1} \sqrt{\frac{x_i + 1}{n_i + 1}} \right],$$

where p_i is the transformed proportion in degrees, x_i is the count and n_i the sample size, all for the i^{th} category. Applying this transformation to the proportion in treatment group 1 from the data shown in Table 1,

$$p_1 = \frac{180}{2\pi} \left[\sin^{-1} \sqrt{\frac{34}{100+1}} + \sin^{-1} \sqrt{\frac{34+1}{100+1}} \right] = 35.7638.$$

Each of the six FT transformed proportions from the data in Table 1 are shown in Table 2. Because this transformation is in degrees, the effective range of this double-arcsine transformation is about 2, when $x_i / n_i = 0$, to about 88, when $x_i / n_i = 1$, for sample sizes in the range of about 100 to 500. This range contracts with smaller sample sizes and expands with larger ones, but it cannot fall outside the values of 0 and 90.

Treatment	1	2	3	4	5	6
Proportions (bin=1)	0.34	0.3714	0.4574	0.32	0.1845	0.2347
FT transformation	35.7638	37.6212	42.5846	34.5587	25.6572	29.1561

Table 2. Proportions and Freeman-Tukey transformed proportions from the data in Table 1.

Once the FT transformations are calculated, they are rank ordered from smallest to largest. The macro then places all relevant data into a single observation in a single SAS data set and creates arrays in order to loop through each of the pairwise comparisons. The critical value from either the Studentized range distribution or the Scheffé modified F-distribution is calculated based on which MC test was requested. The code used by mcpnp for each of these critical values is summarized below.

Tukey's HSD: `probmcc('range', ., 1-&alpha, dfe, &trt);`
LSD: `probmcc('range', ., 1-&alpha, dfe, 2);`
Bonferroni *P*-value adjustment: `probmcc('range', ., 1-(&alpha/&cpw), dfe, 2);`
Dunn-Sidak *P*-value adjustment: `probmcc('range', ., (1-&alpha)**(1/&cpw), dfe, 2);`
Scheffé modified *F*-test: `sqrt(2 * (&trt-1) * finv(1-&alpha, &trt-1, dfe));`

The logic involved in each of these calculation steps goes beyond the scope of this paper. See Westfall, et al., 1999 for more information about how these critical values are calculated.

Next, `mcprpn` uses the 'file print' function and a series of 'put' statements as part of a DATA step to send the results from each pairwise comparison to the OUTPUT window. The following statistics are calculated and tabulated for each *i* vs. *j* comparison:

Diff: difference between the two FT transformed proportions;

SE: FT estimate of pooled standard error (in degrees);

$$SE_{ij} = \sqrt{\frac{a}{n_i + .5} + \frac{a}{n_j + .5}}, \text{ where } a = \frac{1}{2} \left(\frac{180}{2\pi} \right)^2 = 410.3508;$$

$q(\text{obs})$: Diff / SE.

Calculating $q(\text{obs})$ for the comparison involving the largest FT transformed proportion and the smallest FT transformed proportion (Treatment 3 vs. Treatment 5), we find:

$$\text{Diff} = 42.5846 - 25.6572 = 16.9274$$

$$SE_{3vs5} = \sqrt{\frac{410.3508}{94 + .5} + \frac{410.3508}{103 + .5}} = 2.8822$$

$$q(\text{obs}) = \frac{16.9274}{2.8822} = 5.8731.$$

Using the `PROBMC` function in SAS, we find that the critical quantile used in Tukey's HSD test, given the above set of six FT transformed proportions (Table 2) is 4.0433.

`probmcc('range', ., .95, 594, 6);` → 4.0433

Since 5.8731 is larger than 4.0433, a significant difference is found between Treatments 3 and 5. In response, `mcprpn` outputs a line identifying the above comparison with two asterisks (**) along with the three above statistics. Finally, once the macro has output lines for each pairwise comparison and indicated whether or not each difference is statistically significant, it displays the rank-ordered un-transformed proportions for reference.

Tukey HSD Multiple Comparisons on Proportions			
Number of proportions:		6	
Alpha:		.05	
Overall p-Chi Square		0.0006	
DF (error):		594	
Critical Studentized Quantile:		4.0433	
Compare	Diff	SE	q(obs)
3 vs 5 **	16.927	2.882	5.873
3 vs 6 **	13.428	2.917	4.604
3 vs 4	8.026	2.903	2.765
3 vs 1	6.821	2.903	2.350
3 vs 2	4.963	2.869	1.730
2 vs 5 **	11.964	2.803	4.269
2 vs 6	8.465	2.838	2.983
2 vs 4	3.063	2.824	1.085
2 vs 1	1.857	2.824	0.658
1 vs 5	10.107	2.837	3.563
1 vs 6	6.608	2.872	2.301
1 vs 4	1.205	2.858	0.422
4 vs 5	8.901	2.837	3.138
4 vs 6	5.403	2.872	1.881
6 vs 5	3.499	2.851	1.227
Ordered Proportions			
5	0.184		
6	0.235		
4	0.320		
1	0.340		
2	0.371		
3	0.457		

Figure 2. Output from `MCPRPN`

Output from the program, configured to run Tukey's HSD tests on the transformed proportions in Table 2 is shown in Fig. 2. Significant differences were found between treatments 3 and 5, 3 and 6, and 2 and 5. None of the other 12 observed quantiles exceeded the critical quantile of 4.0433.

MCPRPN CALLS

The complete code of the mcprpn macro is below:

```
%macro mcprpn(data=, trt=, alpha=.05, freq=0, row=, col=, prop=, num=, mctype=0, k=0, xcomp=0);
/*****
* ARGUMENTS:
*
* data: name of a SAS dataset containing either raw count data having two rows where
*        proportions associated with the first row (must be identified as row 1) are to
*        be evaluated, or proportions data in rows with associated total n's
* trt: number of proportions to be evaluated
* alpha: alpha level for significance testing, must be constrained by [0 1]
* freq: 0 (default) use a raw counts table as input, 1 use proportions and n's as input
* row: identify the variable representing rows in the dataset
*       (ignored when freq=1)
* col: identify the variable representing columns to be analyzed
*       NOTE: this argument represents columns in the dataset if freq=0 and rows in
*       the dataset if freq=1
* prop: identify the variable representing proportions in the dataset
*       (ignored when freq=0)
* num: identify the variable representing raw counts if freq=0 or the variable
*       representing total n if freq=1
* mctype: multiple comparisons test type, 0 (default) is Tukey's HSD, must be constrained
*         by integers [0 4]
* k: identify the number of comparisons to make for Bonferroni and Dunn-Sidak
*     planned comparisons, must be constrained by [0 &cpw], ignored for tests other
*     than Bonferroni and Dunn-Sidak, 0 (default) indicates all possible pairwise
*     comparisons
* xcomp: identify the name of a data set containing [0 1] data in logical order of
*         numbered comparisons [e.g., 1 2, 1 3, 1 4, 2 3, 2 4, 3 4] where a 1 indicates
*         a planned comparison will be made and a 0 indicates a planned comparison
*         will not be made.
*****/

%let raddeg2 = (180 / (2 * 3.1415926));
%let cpw = (&trt * (&trt - 1) / 2);
%let dist = Studentized;
%let kFlag = 0;

%if &k < 0 %then %do;
    %put ERROR: k must be equal or greater than 0. Macro MCPRPN stopped.;
%return;
%end;
%if &k > &cpw %then %do;
    %put ERROR: k must be less than or equal to &cpw.. Macro MCPRPN stopped.;
%return;
%end;
%if &mctype < 0 or &mctype > 4 %then %do;
    %put ERROR: mctype must be a valid number between 0 and 4. Macro MCPRPN stopped.;
%return;
%end;

%if &k > 0 and &k < &cpw and (&mctype = 2 or &mctype = 3) %then %do;
    %let cpw = &k;
    %if &xcomp = 0 %then %do;
        %put ERROR: You must specify a planned comparisons data set using xcomp under these
        conditions. Macro MCPRPN stopped.;
    %return;
    %end;
    proc means data=&xcomp noprint sum; var xc; output out=xcomps sum=; run; quit;
    data xcompz; set xcomps; if _type_=0; if xc ne &k then call symput('kFlag',1); run;
    %if &kFlag %then %do;
        %put ERROR: Planned comparisons number mis-match. Macro MCPRPN stopped.;
    %return;
    %end;
    proc transpose data=&xcomp out=xcompt prefix=comp; var xc; run;
%end;

%if &xcomp ne 0 and &mctype ne 2 and &mctype ne 3 %then %do;
    %put ERROR: Planned comparisons may only be used in Bonferroni and Dunn-Sidak tests.
    Macro MCPRPN stopped.;
%return;
%end
```

```

%if &mctype = 0 %then %let type=Tukey HSD;
%else %if &mctype = 1 %then %let type=LSD;
%else %if &mctype = 2 %then %let type=Bonferroni;
%else %if &mctype = 3 %then %let type=Dunn-Sidak;
%else %if &mctype = 4 %then %do; %let type=Scheffe; %let dist=Scheffe-F; %end;

%if &freq %then %do;
  data propmc; set &data;
  count1 = round(&num * &prop, 1);
  count0 = &num - count1;
  run;

  proc transpose data=propmc out=px0; var count1 count0; by &col; run;
  data cdat; set px0; rename col1=count;
  if _name_='count1' then a=1; else a=0; drop _name_; run;
%end;
%else %do;
  data cdat; set &data; rename &row=a &num=count; run;
%end;

proc freq data=cdat noprint; weight count;
  tables a*&col / chisq out=freqout outpct; output out=stat chisq; run;

data p0; set freqout; if a=1;
  prop = pct_col / 100;
  n = count / prop;
  p = (arsin(sqrt(count / (n + 1))) + arsin(sqrt((count + 1) / (n + 1)))) * &raddeg2; run;

proc sort data=p0; by p; run;
proc transpose data=p0 out=p1 prefix=p; var n p &col prop; run;

data tmp1; set p1; if _n_=1; sum=0; %do i=1 %to &trt; sum=sum+p&i; rename p&i=n&i; %end; run;
data tmp2; set p1; if _n_=2; run;
data tmp3; set p1; if _n_=3; %do i=1 %to &trt; rename p&i=b&i; %end; run;
data tmp4; set p1; if _n_=4; %do i=1 %to &trt; rename p&i=prop&i; %end; run;

%if &xcomp = 0 %then %let mset = tmp1 tmp2 tmp3 tmp4 stat;
%else %let mset = tmp1 tmp2 tmp3 tmp4 stat xcomp;

data p2; merge &mset;
  title "&type Tests on Proportions";
  array p(&trt); array na(&trt) n1-n&trt; array b(&trt); array prop(&trt);
  array comp(%eval(&trt * (&trt - 1) / 2));
  dfe = sum - 1 - (&trt - 1);

  if &mctype=0 then qcrit = probmc('range', ., 1-&alpha, dfe, &trt);
  else if &mctype=1 then qcrit = probmc('range', ., 1-&alpha, dfe, 2);
  else if &mctype=2 then qcrit = probmc('range', ., 1-(&alpha/&cpw), dfe, 2);
  else if &mctype=3 then qcrit = probmc('range', ., (1-&alpha)**(1/&cpw), dfe, 2);
  else if &mctype=4 then qcrit = sqrt(2 * (&trt-1) * finv(1-&alpha, &trt-1, dfe));

  file print;
  put " ";
  put " &type Multiple Comparisons on Proportions";
  put " ";
  put " Number of proportions: &trt";
  put " Alpha: &alpha";
  put " Overall p-Chi Square " @41 p_pchi 7.4;
  put " DF (error): " @42 dfe;
  put " Critical &dist Quantile: " @41 qcrit 7.4;

  if &mctype = 2 or &mctype = 3 then
    put " Valid Num Planned Comp.: " @42 "%eval(&cpw)";

  put " ";
  put " ";
  put " Compare" @22 "Diff" @34 "SE" @42 "q(obs)";
  put " -----";

  if &mctype=1 then do;
    put " NOTE: this test does not protect family-wise error";
    if p_pchi > &alpha then do;
      put " p-Chi Square is greater than &alpha, no MC tests calculated";
      put " ";
    end;
  end;

  if &mctype ne 1 or p_pchi le &alpha then do;
    do i=&trt to 1 by -1;
      do j=1 to i;
        if i ne j then do;
          mnb = min(b(i),b(j));
          mxb = max(b(i),b(j));
          arr = mxb - mnb + ((&trt * (mnb-1)) - (mnb * (mnb-1)/2));

```

```

        if &xcomp = 0 or comp(arr) then do;
            c = (&raddeg2**2) / 2;
            se = sqrt(c / (na(i) + .5) + c / (na(j) + .5));
            diff = p(i) - p(j);
            q = diff / se;

            if q > qcrit then
                put @5 b(i) "vs" @10 b(j) "***" @20 diff 7.3 @30 se 7.3 @40 q 7.3;
            else put @5 b(i) "vs" @10 b(j) @20 diff 7.3 @30 se 7.3 @40 q 7.3;
            end;
        end;
    end;
end;

put " ";
put " ";
put " Ordered Proportions";
put " -----";

do m=1 to &trt; put @10 b(m) @20 prop(m) 7.3; end;

put " ";
put " ";

run;

%mend mcprpn;

```

The macro has 11 arguments in its call. Not all are required, depending on the structure of the data set, and five have default values. The call to mcprpn can therefore be streamlined considerably. Two generic calls are permitted, depending on the type of data (i.e., a counts table or proportions with associated sample sizes). Examples of both are shown below:

For counts data, the data set must contain at least three variables. One must identify groups and may have an unlimited number of levels; a second variable must identify bins within groups, may have only two levels, with the bin associated with the proportions to be analyzed identified as 1; a third variable must contain the counts associated with each group x bin combination. Given the below DATA step as an example, the following call may be made to mcprpn:

```

%include "mcprpn.sas";

data counts;
    do trt=1 to 6;
        do bin=1 to 0 by -1;
            input ct @@; output;
        end;
    end;
datalines;
34 66 39 66 43 51
32 68 19 84 23 75
;
run;

%mcprpn(data=counts, trt=6, row=bin, col=trt, num=ct);

```

This call leverages all five default values and the values specified in the call to run Tukey's HSD tests on the proportions. It results in the output shown in Fig. 2. It is also possible to use mcprpn to run LSD (conditioned), Bonferroni, Dunn-Sidak or Scheffé tests on the same proportions by changing the value associated with the 'mctype' argument (see macro for documentation).

The Bonferroni and Dunn-Sidak *P*-value adjustment techniques both permit planned comparisons on a subset of the possible pairwise comparisons. If one of these two tests is requested, mcprpn uses the 'k' argument to control the number of valid comparisons. The default ('k'=0) indicates that all possible pairwise comparisons may be made. Setting 'k' to a positive integer permits only that number of valid planned comparisons. When 'k' is a positive integer between 0 and the number of possible pairwise comparisons, a separate data set is required and must be identified using the 'xcomp' argument. This data set must contain a single vector of zeroes and ones in a variable called 'xc' logically ordered as the paired comparisons would be made from the group labels (e.g., 1 2, 1 3, 1 4, 2 3, 2 4, 3 4 if there are four groups). This vector must contain values up to the last logical comparison that will be made.

A Bonferroni test on the data in the 'counts' data set that makes the five planned comparisons of treatment group 2 against all others may be run given the following data set and call:

```

data comparisons; input xc @@; datalines;
1 0 0 0 0 1 1 1 1
; run;

%mcprpn(data=counts, trt=6, row=bin, col=trt, num=ct, k=5, mctype=2, xcomp=comparisons);

```

This call results in the output shown in Fig.3.

Bonferroni Multiple Comparisons on Proportions			
Number of proportions:	6		
Alpha:	.05		
Overall p-Chi Square	0.0006		
DF (error):	594		
Critical Studentized Quantile:	3.6549		
Valid Num Planned Comp.:	5		
Compare	Diff	SE	q(obs)
3 vs 2	4.963	2.869	1.730
2 vs 5 **	11.964	2.803	4.269
2 vs 6	8.465	2.838	2.983
2 vs 4	3.063	2.824	1.085
2 vs 1	1.857	2.824	0.658
Ordered Proportions			
5	0.184		
6	0.235		
4	0.320		
1	0.340		
2	0.371		
3	0.457		

Figure 3. Output from MCPRPN demonstrating planned comparisons

For proportions data, the data set must also contain at least three variables. One variable must identify groups, a second variable must identify the sample size associated with each group, and a third variable must identify the proportions to be tested. Sufficient precision must be used in entering proportions that the product of sample size and proportion rounded to the nearest integer resolves to the correct count associated with that proportion. A generic call to mcprpn using proportions data is shown below:

```
%include "mcprpn.sas";

data proportions;
do trt=1 to 6; input n prop @@; output; end;
datalines;
100 .34 105 .3714 94 .4574
100 .32 103 .1845 98 .2347
;
run;

%mcprpn(data=proportions, trt=6, freq=1, col=trt, prop=prop, num=n);
```

This call also results in the output shown in Fig. 2. The same rules apply with respect to the 'mctype', 'k' and 'xcomp' arguments in this call as in the call involving counts data.

SIMULATIONS

Efficacy of using FT transformations and variance estimates to apply common MC tests to proportions was tested using simulations. Integer data (0-4) were generated to resemble 5-point rating scale data commonly used in marketing research for six underlying binomial distributions ($n = 100$, $p = .5, .6, .7, .8, .9, .95$) and for a uniform distribution ($n = 100$). A binary variable was generated from each distribution, where categories 0, 1 and 2 were placed in one bin ($bin=0$) and categories 3 and 4 were placed in the other ($bin=1$). This was repeated for 2000 samples each containing six groups for a total of $12,000 \times 100 = 1.2$ million observations. An example of the code used to generate these data is shown below:

```
data sims;
do sample = 1 to 2000;
do group = 1 to 6;
do i = 1 to 100;
binomial = ranbin(0,4,p);
uniform = int(5*ranuni(0));
if binomial in (3,4) then bt2b=1; else bt2b=0;
if uniform in (3,4) then ut2b=1; else ut2b=0; output;
end;
end;
end;
run;
```

Distributions of categorical variables and $bin=1$ (observed and tested) for each underlying distribution are shown in Table 3. Proportions were calculated for $bin=1$ for each sample and group. The mcprpn program was modified to calculate differences among FT transformed proportions using all five MC tests at once and to calculate differences among un-transformed proportions using pairwise χ^2 tests of association. Numbers of significant differences by each method for each sample were calculated and output. A different macro (not shown) was written to pass each sample from each underlying distribution individually into the modified mcprpn program.

All statistical testing was done at $\alpha = .05$. Because each proportion in a given sample was drawn from the same underlying distribution, no significant differences were expected and any that were found were attributed to Type I error.

Distribution	p	category					bin=1	
		0	1	2	3	4	Obs.	Test
Binomial	0.5	0.0623	0.2504	0.3751	0.2499	0.0624	0.3122	0.3125
Binomial	0.6	0.0256	0.1532	0.3457	0.3461	0.1294	0.4755	0.4752
Binomial	0.7	0.0080	0.0757	0.2645	0.4112	0.2407	0.6519	0.6517
Binomial	0.8	0.0016	0.0259	0.1536	0.4094	0.4096	0.8190	0.8192
Binomial	0.9	0.0001	0.0036	0.0489	0.2913	0.6560	0.9473	0.9477
Binomial	0.95	7.5E-07	0.0005	0.0136	0.1718	0.8142	0.9859	0.9860
Uniform	n/a	0.1996	0.1990	0.2000	0.2012	0.2001	0.4013	0.4000

Table 3. Proportions from simulated data sets

Family-wise and comparison-wise error rates were calculated for each test and each underlying distribution as the number of samples where at least one significant difference was found divided by the total number of samples (i.e., 2000), and the number of significant differences divided by the product of the number of samples and the number of pairwise comparisons made in each sample (i.e., $2000 \times 15 = 30,000$). For LSD and pairwise χ^2 tests of association, conditional error rates were calculated by ignoring all differences in samples where the overall χ^2 test of association resulted in a P -value $> \alpha$.

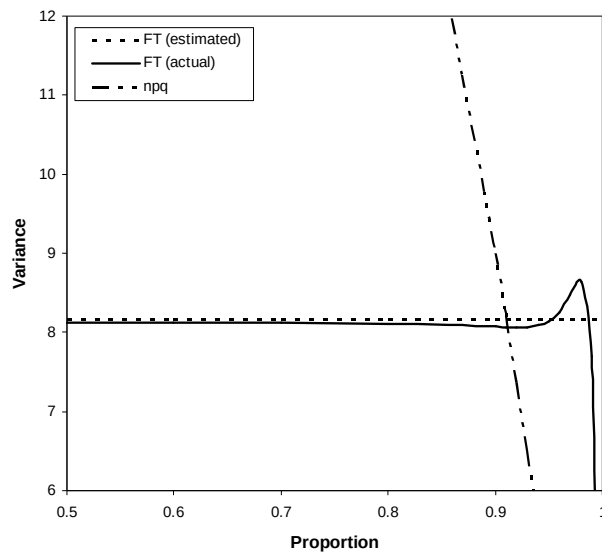


Figure 4. Relationship actual FT, estimated FT and binomial variances as a function of binomial parameter p , $n = 100$

Assuming a binomial distribution, where n and p are known, the actual variance of FT transformed proportions can be calculated. Actual FT variance (in degrees, $n = 100$, $p = .5 : 1$) is compared to the FT estimate of the same variance and the actual variance from the binomial $B[n, p]$ in Fig. 4. The actual variance of the FT transformed proportions differs from the FT estimate by no more than about 1.25% until p reaches about 0.92 or $n(1-p) = 8$. From about this point, there is an anomaly where actual FT variance reaches a maximum at about $p = .98$, followed by a decline to its limit at 0, where $p = 1$. The difference between FT actual and estimated variance exceeded 6% only at $p \geq .99$, where $n(1-p) < 1$. This anomaly shifts left as sample size is decreased and $n(1-p)$ passes 1 at smaller values of p . Bartlett, 1936 found similar and symmetric anomalies in calculated variance of a transformed binomial for marginal values of p . These observations are also in keeping with the earlier claims made by Freeman and Tukey, 1950. Actual variance of the binomial has the same limit as the actual FT variance, but the former follows a parabolic function, reaching a maximum of 25 in this case where $p = .5$. From these calculations, the FT variance estimate appears to be a valid approximation of the actual variance of the FT transformed proportions and may be used in statistical hypothesis testing.

Data from each of the first three central moments of the distributions of proportions and FT transformed proportions in the simulated data sets are shown in Table 4. Columns headed by 'calc' are calculated from the hypothetical underlying distributions; those headed by 'simul' are calculated from the simulated data sets. Means, variances and skew of the proportions all conformed to expected values. Variance was maximized and the magnitude of skew was minimized as the mean (np) $\rightarrow 50$. Variance of the FT transformed proportions on the other hand was much more stable across underlying distributions and was near the FT variance estimate of 8.1662 (see Fig. 4). Likewise,

skewness of the FT transformed proportions was much more stable across underlying distributions and always smaller in magnitude than skewness of the un-transformed proportions. This trend was especially prevalent among marginal proportions. The FT transformation therefore not only provides a reliable estimate of variance for hypothesis testing, but it appears to stabilize skew as well, which would extend the range of p where the normal approximation is applicable.

Underlying distribution	p	Proportions						FT transformed proportions		
		Mean		Variance		Skew		Mean	Variance	Skew
		calc	simul	calc	simul	calc	simul	simul	simul	simul
Binomial	0.5	31.25	31.2223	21.4844	21.6407	0.0809	0.0702	34.0272	8.187	- 0.0504
Binomial	0.6	47.52	47.5486	24.9385	24.8374	0.0099	0.0098	43.6018	8.0934	- 0.0051
Binomial	0.7	65.17	65.1912	22.6987	22.3598	- 0.0637	- 0.0832	53.798	7.9963	0.01
Binomial	0.8	81.92	81.8962	14.8111	14.6838	- 0.1659	- 0.1805	64.7011	8.0134	0.0679
Binomial	0.9	94.77	94.7341	4.9565	4.9421	- 0.4022	- 0.4028	76.4667	8.0709	0.282
Binomial	0.95	98.5981	98.5934	1.3822	1.4081	- 0.8267	- 0.834	82.8388	8.4802	0.1465
Uniform	n/a	40.	40.1298	24.	23.5708	0.0408	0.0338	39.3363	7.9698	- 0.0253

Table 4. Central moments of the simulated data sets

What remains to be seen is whether the use of the FT transformation and variance estimate results in expected comparison-wise and family-wise error rates when applied to common MC tests. The nominally expected comparison-wise error rate (α_c) to protect α_f at .05 is .0034, i.e.,

$$\alpha_c = 1 - e^{-\frac{\ln(1-\alpha_f)}{k}} = 1 - e^{-\frac{\ln(.95)}{15}} = .0034,$$

if all pairwise comparisons among a set of proportions are independent. The nominally expected comparison-wise error rates for each un-conditioned MC test may also be found from the probabilities associated with critical quantiles under the Studentized range distribution, assuming only two proportions are compared. The following use of the PROBMC function can be used to find these probabilities:

$$\alpha_c = \text{probmc}(\text{'range'}, q, ., 594, 2);.$$

These values are shown for the tests and parameters used in the simulations in Table 5.

MC Test	q	$\alpha(c)$
HSD	4.0433	0.0044
Bonferroni	4.1682	0.0033
Dunn-Sidak	4.1576	0.0034
Scheffé	4.7214	0.0009

Table 5. Critical quantiles and expected comparison-wise error rates.

Fig. 5 shows calculated comparison-wise and family-wise error rates for each of five MC tests on FT transformed proportions and for χ^2 tests of association on un-transformed proportions in the simulated data sets. Results from the LSD and χ^2 tests are conditioned on the P -value of the overall χ^2 test in each sample being $\leq .05$.

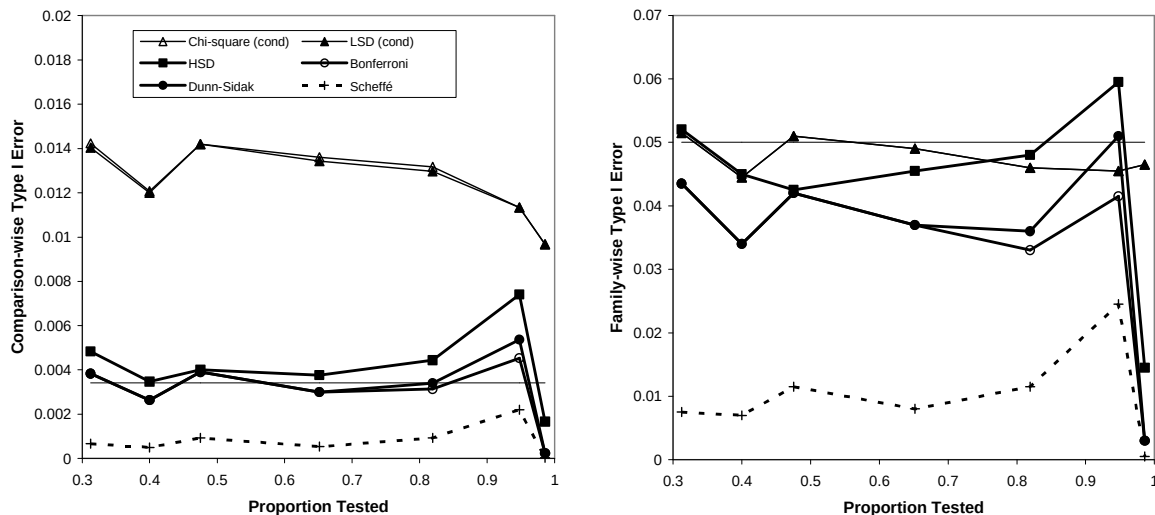


Figure 5. Actual comparison-wise and family-wise Type I errors by MC test and underlying distribution

The anomalies observed at marginal values of p for the other four MC tests resulted in part from deviations between actual and estimated FT variance in the same range (see Fig. 4). When $p \approx .95$ there is an increase in Type I errors as the FT variance estimate (used in significance testing) under-estimates actual variance. As $p \rightarrow 1$ beyond this point, there is a decrease in Type I errors as the FT variance estimate over-estimates the actual. These error rates could presumably be stabilized further by using actual instead of estimated variance in tests of significance. The former however requires either knowledge of p , which under ordinary circumstances we do not have, or its calculation from the binomial probability density function. This calculation could consume an unacceptable amount of computing resources and time, especially as sample sizes become large.

CONCLUSIONS

A Base SAS macro is given that performs several common multiple comparisons tests on proportions. The macro permits the user to specify test type along with several parameters of the statistical hypothesis tests (e.g., alpha, the number of comparisons to make within a given set). It also permits data to be entered in one of two formats: as counts or as proportions.

The macro relies on an angular transformation of the proportions. This transformation stabilizes variance very near an estimate that is used in statistical hypothesis testing and reduces skewness, especially among proportions near the margins. The latter has the effect of extending the range of proportions where a normal substitution can be made for binomial probabilities. Tests that rely on normal probabilities (e.g., Tukey's HSD) can consequently be applied to those proportions. Data simulations using several underlying distributions demonstrated that use of the transformation and variance estimate led to expected error rates among all but the most extreme proportions tested.

REFERENCES

- Bartlett, M.S. 1936. The square root transformation in analysis of variance. *Supplement to the Journal of the Royal Statistical Society*, 3(1):68-78.
- Curtiss, J.H. 1943. On transformations used in the analysis of variance. *Annals of Mathematical Statistics*, 14:107-122.
- Elliott, A.C. and Reisch, J.S. 2006. Implementing a multiple comparison test for proportions in a 2xc crosstabulation in SAS®. *Proceedings of the 31st Annual SAS Users Group International*, San Francisco, CA, Paper 204-31.
- Freeman, M.F., and Tukey, J.W. 1950. Transformations related to the angular and the square root. *Annals of Mathematical Statistics*, 21:607-611.
- Ross, S. 1989. *Introduction to Probability Models, Fourth Edition*. San Diego: Academic Press.
- Sokal, R.R., and Rohlf, F.J. 1995. *Biometry, Third Edition*. New York: W.H. Freeman, pp.419-422.
- Westfall, P., Tobias, R., Rom, D., Wolfinger, R. and Hochberg, Y. 1999. *Multiple Comparisons and Multiple Tests Using SAS*. Cary, NC: SAS Institute, Inc.
- Zar, J.H. 1999. *Biostatistical Analysis, Fourth Edition*. New York: Prentice Hall, p.564.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

John Horne
InsightsNow, Inc.
1600 SW Western Blvd., Suite 350
Corvallis, OR 97333 (USA)
Ph. +1 541 757 1404 ext. 212
FAX: +1 541 757 1402
john.horne@insightsnow.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.
Other brand and product names are trademarks of their respective companies.