Paper Number ST-01

Analysis of Complex Survey Data

Varma Nadimpalli, Westat, Rockville, Maryland

ABSTRACT

Analysis of dose-response relationships has long been the main focus of clinical and epidemiological research. More recently, this type of analysis has been employed to evaluate public education campaigns. The data that are collected in such evaluations are likely to come from standard household survey designs with all the usual complexities of multiple stages, stratification, and variable selection probabilities. To meet these challenges, a new jackknifed Gamma test for a monotone dose-response relationship is proposed. The main focus of this paper is analysis of the data from the complex survey and computation of the jackknifed gamma test using SAS®.

KEY WORDS: Gamma, Jackknifed, Complex Survey

INTRODUCTION

The traditional realm for the application of dose-response relationships has been clinical and epidemiological research. In contrast, the context for our research is the evaluation of the effectiveness of the National Youth Anti-Drug Media Campaign, a Campaign that has been funded by the U. S. with the stated goal of reducing and preventing drug use among youth, both directly, and indirectly by influencing their parents and other significant adults in their lives. The primary tool for the evaluation of the Campaign's effectiveness in achieving its objectives is the National Survey of Parents and Youth (NSPY). NSPY is a national in-home survey of complex design. NSPY collects initial and followup data from nationally representative samples of youth and the parents of these youth.

The implementation of this strategy involves four components: (1) measuring the dose, that is the level of exposure to Campaign messages; (2) measuring the response, that is the set of outcomes the Campaign is supposed to affect; (3) controlling for potential confounders of the dose-response relationships; and (4) testing the hypothesis that a specific observed exposure-outcome relationship is in line with the hypothesis that exposure to Campaign messages is related to the outcome “the right way”, that is, in the desired direction.

The choice of statistic for analysis of data from this survey was the Jonckheere-Terpstra at first, but switched to Gamma as the chief analyst found it easier to interpret. We opted to work with the Gamma test, which is a nonparametric test designed to reject null hypotheses of no association against ordered alternatives. An additional advantage of the Gamma test was that it is available in SAS and could be made a part of an already very complex data

processing stream that needed to be programmed in SAS. Judkins, Zador and Nadimpalli (2002) reported on the performance of a jackknifed Jonckheere-Terpstra (JT) for testing dose-response relationship on data from a complex survey data. Gary Simon (1978) indicated that most of the association tests for data from a simple random sample are asymptotically equivalent (but he did not study the JT). Nadimpalli, Judkins and Zador (2003) extend the results on the jackknifed JT to tests based on other association measures mainly Gamma, Kendal's Tau and to the Cochran-Mantel-Haenzel (CMH) test and demonstrated that all these tests are equivalent even when the data are from a complex survey and the statistics are jackknifed.

COMPLEX SAMPLE DESIGN AND WEIGHTS

The National Youth Anti-Drug Media Campaign was funded by Congress to reduce and prevent drug use among young people 9- to 18- years of age, by addressing youth directly as well as indirectly, and by encouraging their parents and other adults to take actions known to affect youth drug use. The primary tool for the evaluation is the National Survey of Parents and Youth (NSPY). The NSPY is a household-based survey with a sample of over 25,000 youth and 18,000 parents. NSPY employs a stratified multi-stage sample design with unequal weights making it a complex survey. The selection of dwelling units and youth was done in stages using a stratified multistage probability sampling design. The sampling was done in such a manner as to provide an efficient and nearly unbiased cross-section of America's youth and their parents. At the first stage of selection, 90 primary sampling units (PSUs) were selected from 50 strata with probabilities proportionate to size (PPS). The PSUs were generally metropolitan statistical areas (MSAs) or groups of non-metropolitan counties. Within the selected PSUs, a sample of 2,800 second-stage units referred to as segments was selected. The segments were of two types: area segments consisting of Census-defined blocks or block groups, and new construction segments consisting of groups of building permits issued by building permit offices over a specified interval of time. Within the sampled segments, 81,000 dwelling units were selected and screened to identify eligible households (i.e., households with youth 9- to 18- years of age). All types of residential dwellings units were included in the sample. However, institutions, group homes and dormitories were excluded from the study. Within the households with multiple eligible youth, up to two youth were selected at random for the study. The sampling of youth was designed to obtain sufficient numbers of youth in each of the targeted ranges: 9 to 11 years, 12 to 13 years, and 14 to 18 years. These age ranges were judged to be important analytically for evaluating the impact of the campaign. For the NSPY, parents were defined to include natural parents, adoptive parents, foster parents who lived in the same household as the sampled youth, as well as stepparents and other relatives serving as parents provided they lived with the child for at least six months. If more than one parent or caregiver was present in the household, one was randomly selected with no preference given to selecting mothers over fathers. Over the course of the study, completed interviews were obtained for 25,000 youth and 18,000 parents associated with the sampled youth.

For analysis purposes, separate sets of sampling weights were developed for youth, parents, and youth-parent dyads, where a dyad was defined to be a unique youth-parent combination. All of the weights were designed to reflect overall selection probabilities and to compensate for nonresponse and under coverage. Since only one parent was usually sampled per household while up to two youth could be sampled in the same household, a responding parent could be included in up to two distinct dyads. The weights for youth and dyads were developed using analogous procedures and involved a final post-stratification (raking) step to CPS-based estimates of person-level population counts. The derivation of the parent weights, however, required an intermediate step involving the post-stratification of the household weights to the corresponding CPS-based estimates of *household* counts. The goal of the raking was to reduce biases due to under-coverage and nonresponse, and to reduce the sampling error of the estimates. In the raking process, the weights were iteratively adjusted until the sum of the weights agreed with the corresponding population totals derived from the CPS.

For the first three waves of the NSPY (referred to as the 'recruitment" waves), the youth and dyad weights were raked to population counts (i.e., control totals) of youth 9- to 18- years of age. (For the subsequent follow-up waves, the weights were raked to counts of youth 12- to 18- years of age due to the aging of the sample.) The parent weights were not raked in the same fashion because no control totals exist for parents as defined for the NSPY. However, estimates of total households with youth in the relevant age ranges are available from the CPS and were used to rake the household weights from which the parent weights were derived.

ANALYSIS OF COMPLEX SURVEY

Although it might have been possible to develop a parametric model for each outcome in terms of exposure, factors with differential sampling rates, and confounders, the project schedule did not allow time to develop separate models for each outcome variable. A strong advantage of the propensity scoring method that we adopted instead is that it can be used to develop a set of counterfactual projection weights. With this single set of weights, it is possible to remove the confounding effects of all the variables that were used in the propensity scoring. The combination of propensity scoring and the Gamma test promised great speed in the production of a large number of tables that had been designed to look for Campaign effects on a variety of outcomes within a variety of population domains. In order to preserve this benefit of fast production of analytic tables while reflecting the non-ignorability of the sample with respect to family composition and age of housing, as well as incorporating adjustments for differential response rates and under-coverage, it was decided to base the counterfactual projection weights on the survey sampling weights and then to calculate the Gamma test using these counterfactual projection weights. The question then was how to adjust the Gamma for the complex sample design.

Survey practitioners had already made considerable progress in determining how to analyze contingency tables based on complex sample designs. Kish and Frankel (1974)

first established that although the impact of clustering on fixed parameters in models is smaller than on marginal means, it is non-negligible for high intraclass correlation. Holt and Scott (1981) and Scott and Holt (1982) confirmed and expanded upon on that work. Rao and Scott (1981) reviewed the early work and suggested a series of three alternate adjusted chi-square statistics for two-way tables and later generalized these to multi-way tables (Rao and Scott, 1984). Fay (1985) suggested a procedure for testing for independence and various forms of conditional independence in contingency tables using a jackknifed chi-square statistic. The Rao and Scott statistics have become standard features in the WesVar and SUDAAN Statistical software packages.

More recently, Wu, Holt and Holmes (1988) showed the seriousness of ignoring the clustering in determining an overall F statistic for clustered samples and how to correct it. Medical researchers have been slower to recognize these problems, but recent gains have been made in this field as well (c.f., Manda, 2002). We assumed that the problems identified for other types of analysis would also impact the Gamma unfavorably if we were to compute it from a weighted contingency table where the weights had been standardized. For that reason we wanted to do something for the Gamma similar to Fay's or Rao and Scott's corrections to chi-square tests for independence.

Since the Gamma has an asymptotic normal distribution under the null hypothesis of independence, it seemed like a straightforward procedure would be to replicate the Gamma on each of the replicated weights, then calculate a variance on the replicated Gamma, and finally use this in a z-test. More specifically, let J_0 be the standardized Gamma statistic formed on the contingency table of Y by Z using full-sample weights and let J_r be the standardized Gamma statistic formed on the contingency table of Y by Z using the *r*-th set of replicate weights. Let b_r be a factor associated with the *r*-th replicate and the method used to create the replicate weights. Then the "jackknifed" Gamma test is

$$JJT = \frac{J_0}{\sqrt{\sum_r b_r \left(J_r - J_0\right)^2}}$$

Note that we use "jackknifed" more broadly here than to imply that the replicate weights need to be created by a jackknife method. The replicate weights can be created by balanced repeated replications, a bootstrap, or any of a variety of re-sampling schemes.

CONCLUSION

The jackknifed tests have been shown to be reasonable tests for monotone dose-response relationships on clustered data as might be expected in a complex sample survey, repeated measures design or randomized cluster design. The advantages of the jackknifed tests are most apparent at levels of intraclass correlation that may not often be achieved. However,

since the power loss is minimal in situations where the correction is unnecessary, we recommend that the procedure always be used on clustered data, regardless of the level of intraclass correlation expected. SAS was very fast and reliable in computing the jackknifed statistics. We tested the same statistic using other software but programming and computation took longer.

SAS CODE

These pieces of code are built within a macro to compute a gamma statistic for the trend and for the difference between two samples or time points, using a full sample weight and 100 replicate weights.

Gamma statistic for full sample weight:

```
   proc freq data=subset;
     tables eindex * outcome / measures;
     weight fullsampwt;
     test gamma;
     ods output gamma=outstat1(rename = (nvalue1 = GAMMA));
   run;
```

Gamma statistic using 100 replicate weights:

```
%do i=1 %to 100;
  proc freq data=subset;
   tables eindex * outcome / measures noprint;
   weight weight&i; test gamma;
   ods output gamma=outstat&i(rename = (nvalue1 = GAMMA)) ;
  run;
%end;
```

Computation of Jackknifed Gamma statistic for Trend:

```
data v;
 if _n_=1 then do;
  set outstat0 (rename=(gamma=gamma0));
 end;
 set outstat end=lr;
 v + ((gamma - gamma0) ** 2);
 llimit=gamma0-1.98*sqrt(v);
 ulimit=gamma0+1.98*sqrt(v);
 P=(1-probt(abs(gamma0/sqrt(v)),100))*2;
 if ((llimit>0 and ulimit>0) or (llimit<0 and ulimit<0)) then MDE = '*';
     else                                                  MDE = '';
 CI = mde||"("||lower||","||(upper)||(")");
 INTERVAL = compress(ci);
run;
```

Sample Output:

OUTCOME	SUBCAT	GAMMA0	LLIMIT	ULIMIT	INTERVAL
OUTCOME1	ALL	0.006621	-0.01496	0.028202	(-0.01,0.03)
OUTCOME1	1	0.014258	-0.00741	0.035927	(-0.01,0.04)
OUTCOME1	2	-0.11219	-0.15229	-0.07209	*(0.15,0.07)
OUTCOME1	3	0.126449	0.084271	0.168627	*(0.08,0.17)
OUTCOME2	ALL	0.020986	0.000705	0.041267	*(0.00,0.04)
OUTCOME2	1	0.008297	-0.01314	0.029736	(-0.01,0.03)
OUTCOME2	2	0.039642	-0.01733	0.096613	(-0.02,0.10)
OUTCOME2	3	-0.03134	-0.09264	0.029952	(-0.09,0.03)
OUTCOME3	ALL	0.060218	0.024634	0.095801	*(0.02,0.10)
OUTCOME3	1	0.065186	0.025619	0.104753	*(0.03,0.10)
OUTCOME3	2	0.002514	-0.09226	0.097292	(-0.09,0.10)
OUTCOME3	3	0.062671	-0.04165	0.166988	(-0.04,0.17)

Computation of Jackknife Gamma for Difference:

If Gamma1 is the Gamma computed using the full sample weight for Sample 1 and Gamma2 is the Gamma Computed using full sample weight for Sample 2 then the Gamma Difference is computed as

```
Gamma_diff = gamma1 - gamma2;
```

If Gammar1 is Gamma computed using the r^{th} replicate weight for sample 1 and Gammar2 is the Gamma computed using the r^{th} replicate weight for sample 2 then compute

```
R = ((gammar1 - gamma1) - (gammar2 - gamma2)) ** 2;
```

Compute SumR:

```
proc summary data = temp nway sum;
 var R;
 class variable;
 output out = temp1a sum=sumr;
run;
```

Compute the Upper and Lower Limits:

```
Llimit=gamma_diff-1.98*sqrt(sumr);
Ulimit=gamma_diff+1.98.*sqrt(sumr);
```

Variance on difference between actual and counterfactual estimate for Categorical Variables:

Compute the following, where cfpr0 is the full sample counter factual projection weight and cfpr1-cfpr100 are 1-100 replicate counter factual projection weights

```
proc summary data = cfpr;
 class variables;
 var cfpr0-cfpr100;
 output out=counter sum=;
run;
```

Compute the following, where “Weight” is the full sample weight and repwt1-repwt100 are 1-100 replicate weights

```
proc summary data = repwt;
  class variables;
  var weight repwt1-wt100;
  output out=actual sum=;
run;


data both;
  merge counter actual;
  by _type_;
run;

%macro denom;
data denom (drop= weight repwt1-repwt100 cfpr0 cfpr1-cfpr100);
  set both;
  if _type_ = 0;
  den_a=weight;
  den_c=cfpr0;
  %do _i_=1 %to 100;
    dena&_i_=repwt&_i_;
    denc&_i_=cfpr&_i_;
  %end;
run;
%mend denom;
%denom;

%macro numer;
data numer (drop= weight repwt1-repwt100 cfpr0 cfpr1-cfpr100);
 set both;
 if (_type_=1 and Var=1);
 num_a=weight;
 num_c=cfpr0;
 %do _i_=1 %to 100;
   numa&_i_=repwt&_i_;
   numc&_i_=cfpr&_i_;
 %end;
run;
%mend numer;
%numer;
```

```
%macro calc;
data calc (keep=esta estc vara varc covar vardiff sddiff sda sdc
diff uplimit limit);
  merge denom numer;
  esta= num_a/den_a;
  estc= num_c/den_c;
  covar=0; vara=0; varc=0;
  %do _i_=1 %to 60; (The first 60 replicates were designed to
measure the between PSU variance)
    covar+2.567700*(numa&_i_/dena&_i_-esta)*(numc&_i_/denc&_i_-
estc);
    vara+2.567700*(numa&_i_/dena&_i_-esta)**2;
    varc+2.567700*(numc&_i_/denc&_i_-estc)**2;
  %end;
  %do _i_=61 %to 100; (61-100 replicates were designed to measure
within PSU variance, so they have a different factor)
    covar+0.064200*(numa&_i_/dena&_i_-esta)*(numc&_i_/denc&_i_-
estc);
    vara+0.064200*(numa&_i_/dena&_i_-esta)**2;
    varc+0.064200*(numc&_i_/denc&_i_-estc)**2;
  %end;
  vardiff=vara+varc-2*covar;
  sddiff = sqrt(vardiff);
  sda = sqrt(vara);
  sdc = sqrt(varc);
  diff = esta - estc;
  llimit = diff - (1.98*sddiff);
  uplimit = diff + (1.98*sddiff);
run;
%mend calc;
%calc;
```

Variance on difference between actual and counterfactual estimate for Continuous Variables:

```
%macro cfp;
data cfp;
 set counter;
cfpwt = var*cfpr0;
 %do _i_ = 1 %to 100;
  cfpwt&_i_ = var*cfpr&_i_;
 %end;
run;
%mend cfp;
%cfp;
```

```
proc summary data=cfp;
  var cfpr0-cfpr100 cfpwt cfpwt1-cfpwt100;
  output out=counter sum=;
run;

%macro wt;
data rep;
 set actual;
 recdwt = weight*var;
 %do _i_ = 1 %to 100;
  recwt&_i_ = var*repwt&_i_;
 %end;
run;
%mend wt;
%wt;

proc summary data=rep;
  var weight repwt1-repwt100 recdwt recwt1-recwt100;
  output out=actual sum=;
run;

%macro both;
data both(drop= weight repwt1-repwt100 cfpr0 cfpr1-cfpr100);
 merge counter actual;
  by _type_;
 recratio = recdwt / weight;
 cfpratio = cfpwt  / cfpr0;
 %do _i_ = 1 %to 100;
  recra&_i_ = recwt&_i_ / repwt&_i_;
  cfpra&_i_ = cfpwt&_i_ / cfpr&_i_;
 %end;
run;
%mend both;
%both;

%macro calc;
data calc;
 set both;
 esta= recratio;
 estc= cfpratio;
 covar=0; vara=0; varc=0;
 %do _i_=1 %to 60;
   covar+2.567700*(recra&_i_-esta)*(cfpra&_i_-estc);
   vara+2.567700*(recra&_i_-esta)**2;
   varc+2.567700*(cfpra&_i_-estc)**2;
 %end;
 %do _i_=61 %to 100;
   covar+0.064200*(recra&_i_-esta)*(cfpra&_i_-estc);
   vara+0.064200*(recra&_i_-esta)**2;
```

```
   varc+0.064200*(cfpra&_i_-estc)**2;
 %end;
 vardiff=vara+varc-2*covar;
 sddiff = sqrt(vardiff);
 sda = sqrt(vara);
 sdc = sqrt(varc);
 diff=esta-estc;
 llimit  = diff - (1.98*sddiff);
 uplimit = diff + (1.98*sddiff);
run;
%mend calc;
%calc;
```

ACKNOWLEDGMENTS

The author thanks David Judkins, Paul Zador and Mike Rhoads for all the help he received.

DISCLAIMER

The contents of this paper is the work of the author and do not necessarily represent the opinions, recommendations, or practices of Westat.

REFERENCES

Brewer, K.,(2002) Combined Survey Sampling Inference, Arnold Publications, London.

Fay, R. E. (1985), “A jackknifed chi-squared test for complex samples”, *Journal of the American Statistical Association*, 80, pp. 148-157.

Holt, D., and Scott, A. J. (1981), “Regression analysis using survey data”, The Statistician, 30, pp. 169-178.

Jonckheere, A. R. (1954), “Distribution-free k-sample test against ordered alternatives”, *Biometrika,* 7, pp. 93-100.

Judkins, D., Zador, P., Nadimpalli, V. (2002), *Analysis of Dose-Response Relationships on Complex Survey Data, Proceedings of Statistics Canada Symposium.*

Nadimpalli, V., Judkins, D., Zador, P., (2003), Tests of Monotone Dose-Response in Complex Surveys, *Proceedings of American Statistical Association.*

Kish, L., and Frankel, M. R. (1974), “Inference from complex samples (with discussion)”, *Journal of the Royal Statistical Society*, Series B, 36, pp. 1-37.

Lê, T.N. and Verma, V.K. (1997). An Analysis of Sample Designs and Sampling Errors of the Demographic and Health Surveys. Demographic and Health Surveys, Analytic Reports No. 3. Calverton, MD. Macro International.

Lucas, W.F. (1983), "Gamma distribution", in Katz, S. and Johnson, N.L,(eds) Encyclopedia of Statistical Sciences, New York: Wiley, Vol. 3, pp. 292-298.

Mantel N. (1963), "Chi-square tests with one degree of freedom: extensions of the Mantel-Haenszel procedure", *Journal of the American Statistical Association*, 58, pp. 690-700.

Pirie, W. (1983), "Jonckheere tests for ordered alternatives", in Kotz, S., and Johnson, N. L (eds.) *Encyclopedia of Statistical Sciences*, New York:Wiley, vol. 4, pp. 315-318.

Rao, J. N. K., and Scott, A. J. (1981), "The analysis of categorical data from complex sample surveys: Chi-squared tests for goodness of fit and independence in two-way tables", *Journal of the American Statistical Association*, 76, pp. 221-230.

Rao, J. N. K., and Scott, A. J. (1984), "On chi-squared tests for multiway contingency tables with cell proportions estimated from survey data", *The Annals of Statistics*, 12, pp. 46-60.

Rust, K.F., and Ross, K.N. (1993), Multinational Survey of Educational Achievement, presented at the 49th Session of the International Statistical Institute, Florence Italy.

Scott, A. J., and Holt, D. (1982), "The effects of two-stage sampling on ordinary least squares methods", *Journal of the American Statistical Association*, 77, pp. 848-854.

Simon, G. (1978), *Effects of Measures of Association for Ordinal Contingency Tables, Journal of the American Statistical Association, September 1978, Volume73.*

Terpstra, T. J. (1952), "The asymptotic normality and consistency of Kendall's test against trend when ties are present in one ranking", *Indag. Mat*., 14, pp. 327-333.

Wesvar 4.0 User's Guide,(2000) Westat Inc. Rockville, MD.

Sudaan 8.02 Software (2003): RTI, Research Triangle Park, NC.

SAS Procedures Guide, Version 8 (1999): SAS Institute, Inc., Carey, NC.

Wu, C. F. J., Holt, D., and Holmes, D. J. (1988), "The effect of two-stage sampling on the F statistic", *Journal of the American Statistical Association*, 83, pp. 150-159.

CONTACT INFORMATION

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

Varma Nadimpalli
Westat
1650 Research Boulevard, WB272
Rockville, MD 20850
240-453-2799
VarmaNadimpalli@westat.com

TRADEMARKS