The SAS contrastTest Macro

Molin Wang, Ruifeng Li, and Donna Spiegelman

September 17, 2014

Abstract

The %contrastTest macro conducts heterogeneity test for comparing the exposure-disease associations obtained from separate subtype-specific analysis based on the cohort or nested case-control studies. Specifically, the user runs separate Cox (for cohort studies) or conditional logistic models (for nested case-control studies) for each subtype, and then tests the heterogeneity hypothesis using the outputs from the separate models, or the user takes the estimates (and standard errors) from the literature and test the heterogeneity hypothesis. In the subtype-specific analysis, the confounders-disease associations are allowed to be different among the subtypes.

Keywords: SAS, macro, heterogeneity test, subtype, contrast

Contents

1 Description

%contrastTest is a SAS macro that performs heterogeneity test for comparing exposure-disease associations across disease subtypes, in the cohort or nested case-control studies. The exposure-disease subtype associations can be obtained from separate subtype-specific analysis or literature. In the subtype-specific analysis, the confounders-disease associations are allowed to be different among the subtypes, which is the so-called unconstrained analysis (Wang, Spiegelman, *et al.*, 2014).

2 Invocation and Details

In order to run this macro, your program must know where to look for it. You can tell SAS where to look for macros by using the options:

```
options mautosource sasautos=<directories macro is located>;
```

In the Channing servers, the option statements might be

```
options mautosource sasautos='/usr/local/channing/sasautos';
```

Below is a description of the input parameter.

```
%contrastTest(data=);

The input dataset must contain the following two variables:
         ESTIMATE STDERR
which, in cohort studies, can be obtained from separate standard Cox
regression model analysis, using SAS PROC PHREG, for each subtype.
For the colon cancer example below, the data set used in the
PROC PHREG procedure for the jth subtype is the Anderson-Gill
data structure where the censoring variable is 1 or 0, depending on
whether the jth subtype occurs or not, and time is the months
```

```
from the start of the questionnaire cycle until colon cancer incidence
of any subtype, next questionnaire return or end of follow up,
whichever happens first. For each subtype, we typically stratify the
Cox model analysis jointly by age in months at start of follow-up and
calendar year of the current questionnaire cycle to control as
finely as possible for confounding by age, calendar time and any
possible two-way interactions between these two time scales.
The first observation in the input dataset of the macro is the log RR and
standard errors from the model for the first subtype; the second is from
the model for the second subtype, and so on. This macro will test the
null hypotheses that these relative risks are the same for all subtypes,
versus the alternative that the relative risk for at least one subtype is
different from at least one of the others. For nested case-control studies,
ESTIMATE STDERR can be obtained from separate conditional logistic
analysis, using SAS PROC PHREG or PROC LOGISTIC, for each subtype.
```

One way to create this dataset is:

```
data DATANAME;
 input ESTIMATE STDERR;
 datalines;
 (insert log relative risk and its standard error FROM the FIRST MODEL)
 (insert log relative risk and its standard error FROM the SECOND MODEL)
 (etc.)
 ;
```

3 Examples

Below is an example illustrating the use of this macro, based on a study of the association of alcohol consumption with colon cancer subtypes defined by LINE-1 methylation level (3 categories, low, medium and high).

We first create the dataset containing the subtype-specific $log(\hat{RR})$, representing the exposure-subtype associations, and their standard errors.

```
data colon;
```

```
input subtype $1-11 ESTIMATE    STDERR;
datalines;
LINE-1-low          0.30950       0.13467
LINE-1-med          0.44929       0.10814
LINE-1-high      -0.0007371       0.11743
;
```

The macro call is:

```
%contrastTest(data=colon);
```

The macro printed out the following output.

```
Global test of differences between k subtypes
=============================================
Test statistic for H0: beta_type1 = ... = beta_typek   8.1291
p-value                                                0.0172
```

4 How should I describe this in my Methods section?

The estimates of the subtype-specific relative risks are based on a fully unconstrained approach in which the confounder effects are allowed to be difference among the subtypes. To test whether the exposure-disease association differs among the subtypes, we used the contrast test method (Wang, Spiegelman, *et al.*, 2014).

5 Correspondence

Questions should be addressed to Molin Wang via email stmow@channing.harvard.edu.

6 References

Molin Wang, Donna Spiegelman, Aya Kuchiba, Paul Lochhead, Sehee Kim,

Andrew T. Chan, Elizabeth M. Poole, Rulla Tamimi, Shelley S. Tworoger, Edward Giovannucci, Bernard Rosner, Shuji Ogino.
Conceptual and Statistical Frameworks in Epidemiology for Studying Disease Heterogeneity and Pathogenesis. Submitted. 2014.

Anderson TW. Introduction to multivariate statistics. New York, NY: John Wiley and Sons; 1984.