A SAS® PROGRAM FOR THE GENERALIZATION OF THE PAIRED T-TEST

James Lee and H P Lee
Faculty of Medicine
National University of Singapore

1. ABSTRACT

The paired t-test is the standard statistical procedure for comparing the mean value of a quantitative response variable based on a pair-matched (1-to-1 matched) sample. Not infrequently, however, the research situation warrants N-to-1 or N_1-to-N_2 matching rather than pair - matching. Rosner proposed a generalization of the paired t-test for the analysis of quantitative data generated by these matching schemes. Herein we describe two SAS programs to carry out the generalization of the paired t-test. The first program (GPTTEST) performs the test in accordance to Rosner's procedure; the program handles data generated by the N-to-1 and N_1-to-N_2 matching schemes, and N_1 and N_2 may vary among matched-sets. The second program (GPTGRAF) provides a graphic display depicting the extend to which the stimulus-response relationship is modified by the quantitative matching variable. A numeric example was used to illustrate the application of the two SAS programs.

2. INTRODUCTION

The objective of many research investigations is to determine the effect of a stimulus on some quantitative response. This often entails comparing the mean value of the response variable between two groups of subjects, one exposed and one not exposed to the stimulus of interest. Thus to test the postulation that cigarette smoking adversely affects pulmonary function, a study might be carried out to compare the mean forced expiratory volume (FEV) between "smokers" and "nonsmokers". Comparing crude means may, however, give rise to a mistaken impression about the effect of stimulus (smoking) on the response (FEV), since such a comparison is potentially biased by confounding covariates such as age, sex, body height and so on. Similarly, a clinical trial might be carried out to compare the mean reduction in blood pressure in hypertensive subjects given either a new or established antihypertensive drug. Again, comparing the crude mean reduction in blood pressure is potentially biased by differences in baseline blood pressure, smoking habits and so on in the two groups of subjects. It is crucial, therefore, that confounding bias be minimized in a comparison study(1).

The general strategy to reduce confounding bias, especially in a nonrandomized comparison study, includes matching at the study design stage, and/or statistical adjustment, either via analysis of covariance (2) or the Z-score method (3), at the data analysis stage. The relative merits and implications of matching versus statistical adjustment as bias-reduction techniques are well documented(4-7).

Quantitative response data generated by 1-to-1 matching (pair-matching) of selected potential confounding covariates (matching variables) are appropriately analysed by the paired t-test(8). However, 1-to-1 matching may not be an optimal study design option if the population of one comparison group is much larger than that of the other. Consider an investigation to test the postulation that "long term" usage of oral contraceptives is associated with an increase in blood pressure in women. Since there are many more nonusers than are users, it would be advantageous in terms of statistical efficiency (9) by matching more than one nonusers to each user on selected covariates such as age and body weight. Similarly, in retrospective case-control studies, since the control population is typically many folds larger than the case population, one or more controls could be matched to each case to form a matched-set. This procedure is generally referred to as N-to-1 matching, where N may be constant or it may vary among matched-sets. Some research situations may warrant group-matching, wherein one or more subjects in one comparison group is matched to one or more subjects in the other group to form a matched-set. This is referred to as N_1-to-N_2 matching, where N_1 and N_2 may vary among matched-sets. A SAS program is available for the construction of matched-sets(10).

Rosner (11) proposed a statistical procedure which is a generalization of the paired t-test for the analysis of matched-set data generated by the N-to-1 or N_1-to-N_2 matching study design. Herein we describe two SAS programs. One to carry out Rosner's method, and the other to graphically display the extend to which the stimulus-response relationship is affected by the matching variable.

3. STATISTICAL MODEL AND NOTATION

Rosner's generalization of the paired t-test for matched set data stipulates that the within matched-set mean difference between the two comparison groups for the j-th matched-set, viz., $d_j = Y_{1j} - Y_{2j}$, follows a one-way classification random effects analysis of variance model,

$$d_j = \bar{Y}_{1j\cdot} - \bar{Y}_{2j\cdot} = \alpha + \Delta_j + e_j, \quad j = 1,2,\ldots,R.$$

The notations pertaining to the above model are :

R = number of matched-sets, indexed by j. N_{1j} and N_{2j} are the number of subjects in the j-th matched-set in comparison groups 1 and 2, respectively.

Y_{1jk} and Y_{2jk} = measured values of the response variable for the k-th subject from the j-th matched-set in comparison groups 1 and 2, respectively.

α = overall weighted average of d_j, $j = 1,2,\ldots,R$.

Δ_j = random effect representing a matched-set specific deviation from α for the j-th matched-set; Δ_j is assumed to be normally distributed around zero with variance σ_b^2 (between matched-set variance).

e_j = random effect between subjects within the j-th matched set; e_j is assumed to be normally distributed around zero with variance σ_j^2, and σ^2 and is overall average within matched-set variance, i.e., average of, σ_j^2, $j = 1,2,\ldots,R$.

The goal of Rosner's method is to test the statistical hypothesis that $\alpha = 0$. The algebraic procedure is described in detailed in Rosner's paper (11). Essentially, the test is based on normal distribution theory and is therefore asymptotically valid under certain conditions.

4. SAS PROGRAMS

Two SAS programs were written to carry out the generalization of the paired t-test for the analysis of matched-set quantitative data. The first program (GPTTEST) performs the test in accordance to Rosner's procedure (11); the program handles data generated by the N-to-1 as well as the N_1-to-N_2 matching schemes. The second program (GPTGRAF) provides a graphic display depicting the extend to which the effect of stimulus on response is modified by the quantitative matching variable.

5. A NUMERIC EXAMPLE

We will consider a numeric example to illustrate the application of the two SAS programs. Supposing we were to investigate the postulation that smoking affects pulmonary function, as measured by forced expiratory volume (FEV), in a certain population. One possible study design would be to compare FEV between "smokers" and age-matched "nonsmokers", where one or more nonsmokers are matched to one or more smokers according to age ($\pm$ 1 year), i.e., N_1-to-N_2 matching. The hypothetical data from this study (Table 1) were analysed by GPTTEST and the output is reproduced in Table 2. It can be seen that the mean FEV for nonsmokers (3.3622 liters) is higher than that for smokers (3.3574), and that the weighted mean difference in FEV between the two groups (-0.0196) is statistically significant at the 0.05 probability level, for the 2-sided test (p=0.045336) and for the 1-sided test (p=0.022668). Note that "sigma-squared W" (0.00022) is the weighted average of the two group-specific within matched-set variances (0.00026 and 0.00019). In an N-to-1 matching situation, the within matched-set variance for the comparison group with one subject per matched-set cannot be estimated, and "sigma-squared W" will simply be the within matched-set variance of the comparison group with more than one subject per matched-set.

The program GPTGRAF outputs FIGURE 1. This figure shows the relationship between FEV difference (Smoker - Nonsmoker) and age. From this figure, we see that the effect of smoking on FEV is highly dependent on age. Specifically, smoking adversely affects pulmonary function to a much greater extend in the older age group than it does in the younger age group. Clearly, these descriptive information can provide a very useful insight on the stimulus-response relationship, an insight which a simple comparison of means, viz., statistical information outputted by GPTTEST, cannot possibly provide.

GPTGRAF is of course applicable only if the matching variable is quantitative (e.g., age, height). If the matching variable were categorical (e.g., sex, race, body weight discretized into 'normal weight' and 'overweight'), one would simply compare the mean difference in FEV between smokers and nonsmokers among each category of the matching variable to assess whether the effect of smoking on FEV is modified by that matching variable.

6. DISTRIBUTION OF THE PROGRAMS

We have two versions of the SAS programs (GPTTEST and GPTGRAF), one version runs on the IBM mainframe under CMS (SAS Release 5) and the other version runs on the microcomputer (SAS Release 6.03). A listing of these programs, including the input data and outputs for the numeric example we described in this article, are available upon request to :

James Lee, Associate Professor, Department of Community, Occupational and Family Medicine, NUH, National University of Singapore, Lower Kent Ridge Road, SINGAPORE 0511.

The material will be sent by sea-mail.

7. REFERENCES

1. Lavori P W, Louis T A, Bailer J C, et al. Design for experiments : parallel comparisons. N. Engl. J. Med. 1983 ; 1291-1299.

2. Lee J. Analysis of Covariance by the SAS GLM procedure. Comp. Biol. Med. 1987 ; 17:221-225.

3. Lee J, Koh D. Adjustment of bias in the comparison of means. Brit. J. Indust. Med. 1987; 44:430.

4. Rubin D B. The use of matched sampling and regression adjustment to remove bias in observational studies. Biometrics 1973; 29:185-203.

5. McKinlay S M. The design and analysis of the observational study : a review. J. Amer. Stat. Assoc. 1975; 70:503-520.

6. Anderson S, Auquier A, Hauck W W, et al. Statistical Methods for Comparative Studies : Techniques for Bias Reduction. Wiley, New York, 1980.

7. Lee J, Lee H P. A situation in which age-adjustment by covariance analysis maybe inappropriate. Int. J. Epidemiol. 1989; 18:000 (Letter).

8. Rosner B. Fundamentals of Biostatistics. Duxbury Press, Boston, 1982.

9. Ury H K. Efficiency of case-control studies with multiple controls per case : continuous or dichotomous data. Biometrics 1975; 31:643-649.

10. Hanrahan L P, Olson J, Anderson H A, et al : STRATRAN : Stratified random selection of controls for matched case-control data. Proc. 13th Ann. Users Group Int. Conf., SAS Institute, Inc., Cary, N C (1988), pp. 1006-1010.

11. Rosner B. A generalization of the paired t-test. Appl. Statist. 1982 ; 31:9-13.

Table 1. Forced expiratory volume in liters (FEV) in "smokers" and their age-matched "nonsmokers"

Matched set number	Smokers						Age-matched Nonsmokers					
	Person 1		Person 2		Person 3		Person 1		Person 2		Person 3	
	Age	FEV	Age	FEV	Age	FEV	Age	FEV	Age	FEV	Age	FEV
1	24	4.08	23	4.09	23	4.06	23	4.07	23	4.06	24	4.09
2	25	4.20	26	4.16			25	4.14	26	4.18	25	4.16
3	27	3.90					27	3.92				
4	29	3.77	28	3.74			28	3.70				
5	30	3.70	31	3.72	39	3.69	30	3.65				
6	32	3.70	33	3.68			33	3.72	32	3.70		
7	35	3.49					35	3.48	34	3.46	35	3.47
8	36	3.65					36	3.63	36	3.65		
9	37	3.38					38	3.32				
10	40	3.32	40	3.34			39	3.36				
11	41	3.39					41	3.38	41	3.39		
12	42	3.33					42	3.35				
13	43	3.30					43	3.33				
14	44	3.33	44	3.35	44	3.37	44	3.36				
15	45	3.32					45	3.32	45	3.34		
16	47	3.12					47	3.15	48	3.13		
17	49	2.98	49	3.01			49	3.07	49	3.06		
18	50	3.08					50	3.16				
19	51	2.86	51	2.87			51	3.00	51	3.02		
20	52	2.78	52	2.80			52	2.86	52	2.85		
21	53	2.90	54	2.90			53	2.97	53	2.95		
22	57	2.66	58	2.67			58	2.74	58	2.73		
23	60	2.62	60	2.63			61	2.64	61	2.66	60	2.63

Table 2. Statistical Results Output by Program GPTTEST

GENERALIZATION OF THE PAIRED-T TEST FOR MATCHED-SETS DATA
(CMS INPUT DATAFILE = FEV DATA)

COMPARISON GROUP	NO. OF SUBJECTS	MEAN OF FEV(L.)	WITHIN MATCHED-SET VARIANCE OF FEV(L.)
SMOKER	39	3.3574	0.00026
NONSMOKER	41	3.3622	0.00019

R = NO. OF MATCHED-SETS = 23

SIGMA-SQUARED D = BETWEEN MATCHED-SET VARIANCE
OF FEV(L.) = 0.00191

SIGMA-SQUARED W = WITHIN MATCHED-SET VARIANCE
OF FEV(L.) = 0.00022

ALPHA = WEIGHTED MEAN DIFFERENCE IN FEV(L.) = -0.0196

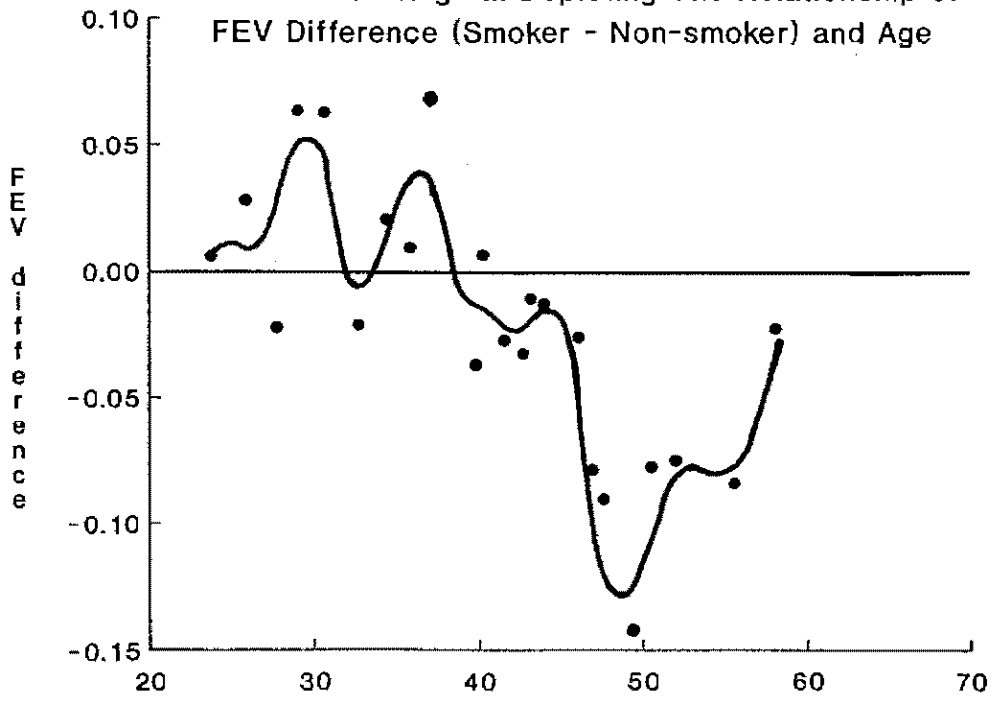
S_ALPHA = STANDARD ERROR OF ALPHA = 0.00980

LAMBDA = NORMAL DEVIATE (Z) = 2.00152

2-SIDED SIGNIFICANCE PROBABILITY = 0.045336

1-SIDED SIGNIFICANCE PROBABILITY = 0.022668

Figure 1. Graphic Display Output by Program Gptgraf
Bivariate Scattergram Depicting The Relationship of
FEV Difference (Smoker - Non-smoker) and Age



Interpolation Line Obtained by Cubic Spline Smoothing