ESTIMATING TREATMENT EFFECTS IN SELECTED PAIRED DATA

Kazemnejad A.[1], Mohebbi M. [2], Sanagoo M. [2]

ABSTRACT

Consider a test-retest study in which subjects are selected to participate in an intervention because one baseline measurement of a variable is within a particular range. For example screening specimens from extreme of a population. This requires taking measurements on selected subjects, usually without a parallel control group and, hence, it is necessary to adjust the effect for regression to the mean. We propose a two-step procedure to estimate an additive regression to the mean model. First estimate the model parameters by least square or moments. Second use these parameter estimates as initial values fit into maximum likelihood procedure to handling several kinds of sampling procedures: truncated, censored, selected and complete sampling. This method also does not have the restriction of knowing the exact value of the population mean or proportion of truncated population. Tests based on this new approach are also discussed. An example from an educational study is illustrated.

Key Word: Bivariate normal, Regression to the mean, Wald's test, Likelihood ratio test.

[1]- Assistant Professor, Department of Biostatistics, Tarbiat Modares University, Tehran, I.R.Iran.

[2]- Mc. Student, Department of Biostatistics, Tarbiat Modares University, Tehran, I.R.Iran.

1. INTRODUCTION

Subjects often are selected for treatment because they have high (or low) values of some baseline measurement of interest. For example, suppose all students who failed an exam are given the opportunity to retake a similar exam. Because subjects are selected from and extreme portions of the distribution, subsequent observation on score of students will tend to regress toward the mean regardless of the presence or absence of a treatment effect, Since before and after treatment measurement on the variable of interest are usually correlated unperfectly, the assessment of treatment effects must be adjusted for regression to the mean. Similar problems occur in some medical studies where patients who exceed a critical point on the variable of interest are selected for trial due to have a marker level within a defined range at study baseline. A review of previous approaches and a discussion of techniques of parameters estimation from a model which is assumed to have a bivariate normal distribution was provided by James [6] that proposed a model with a multiplicative treatment effect and obtained moment estimators. Senn and Brown [10,11] considered the same problem and let the treatment effect be additive or multiplicative. Chen and Cox [1] studied this model in large screening program and proposed an approach that is asymptotically equivalent to the maximum likelihood method. Chen, Cox and Cui [2] again in screening program used a stratified regression to the mean model including both additive and multiplicative treatment effects. Mee and Chua [8] proposed a regression-based test as an alternative to paired t-test for truncated samples. Since the effect of any treatment is likely to be related to the initial level in a multiplicative model, it seems preferable to an additive model. However, this does not rule out an overall additive effect that may or may not be related to the treatment. Assume that the value x (the first measurement) and y (the second measurement) follow a bivariate

normal distribution. When there is no treatment effect, the two variates have the same mean (μ), standard deviation (σ), and correlation coefficient (ρ). We often expect $\rho > 0$, although this is not necessary. So under the null hypothesis the joint density of X and Y is:

$$f(X,Y) = \frac{\exp\left\{-\frac{1}{2(1-\rho^2)\sigma^2}\left[(X-\mu)^2 - 2\rho(X-\mu)(Y-\mu) + (Y-\mu)^2\right]\right\}}{2\pi\sigma^2\sqrt{1-\rho^2}} \qquad (1.1)$$

And the conditional density of Y is:

$$g(Y/X) = \frac{\exp\left\{\frac{-\frac{1}{2}[(Y-\mu)-\rho(X-\mu)]^2}{[\sigma^2(1-\rho^2)]}\right\}}{\sqrt{2\pi\sigma^2(1-\rho^2)}}$$

If we compare the treatment group with the control group, both groups will tend to regress toward the mean. Assuming that there is no unintentional treatment effect, the change in the control group can be attributed to the effect of regression toward the mean. When we take the difference between the two groups, the regression effect is adjusted and perhaps we can use the usual two-sample t-test to evaluate the treatment effect. Obtaining a control group especially from a high-risk group may not always be practical, ethnical or economical in many experiments. If we want to compare the before and after treatment measurements on the same objects, without a control group, any treatment effect that may be present will be confound by the regression effect, so the usual paired tests cannot be used. Suppose there is no treatment effect, the before and after measurements are related together by the equation:

$$(Y-\mu) = \rho(X-\mu) + \varepsilon \qquad (1.2)$$

Where ε is distributed normally and independently of X with mean 0 and standard deviation $\sqrt{\sigma^2(1-\rho^2)}$.

In the presence of treatment effect two possible model are proposed:

1- Additive models (Mee and Chua[8], Senn and Brown[11]) that assumes the treatment effect is not related to the baseline measurement:

$$(Y-\mu)=\rho(X-\mu)+\tau+\varepsilon \qquad (1.3)$$

Where $\tau>0$ is a treatment parameter.

2- Multiplicative model (James[6], Senn and Brown[10]) that assumes the treatment effect is a multiplicative factor of the baseline measurement:

$$Y-\mu=\gamma\rho(X-\mu)+\varepsilon \qquad (1.4)$$

Where $|\gamma|<1$ is a treatment parameter.

Pocock[9] observed that there might be reasons for believing that the simple additive model provides a more straightforward description of patient improvement. Especially in our problem from the unconditional distribution of Y for model 1:

$$E(Y)=\mu+\tau \text{ And } var(Y)=\sigma^2$$

And for model 2:

$$E(Y)=\mu, var(Y)=\sigma^2[1+\sigma^2(\gamma^2-1)]$$

It means that the additive model alters the mean but not the variance of Y where the reverse is true for the multiplicative model. We restrict the discussion to the additive model.

In the next sections, we propose a two-step procedure in which the initial parameter estimates are obtained by the method of the least squares or method moments. Then fit the model into the maximum likelihood procedure to alter different sampling scheme.

2. METHODS

2.1. NOTATION

Cohen [3] described four kinds of samples from the population of baseline values with varying degrees of missing information from bivariate normal distributions: truncated, censored, selected, and complete sample. For a truncated sample, the sampling procedure is continued until n specimens for certain range of x (e. g, $x \leq k$ or $x \geq k$) are recorded. It is not possible to observe the eliminated x values. Corresponded y measurements are made of the n chosen individuals. In the case of a censored sample, both x and y are recorded if x is in the interval. However, account is kept of censored specimens for which x is not in the interval, although neither the x nor the y values of such specimens are recorded. For selected samples, full measurement is made and recorded for all the values of x whether it is in the interval or not. However, only the y values corresponded to the x in the interval are recorded. Finally a complete sample is one in which all the values of x and y is recorded whether they are in the interval or not. For convenience of notation, we assume that the first n subjects are chosen to receive treatment among a total of n+m specimens in the sample. Therefore we will have such a data set:

$$(x_1, y_1), \ldots (x_n, y_n), x_{n+1}, \ldots x_{n+m}$$

2.2. LIKELIHOOD FUNCTION

The joint density of X and Y can be written as the product of the marginal frequency function of X and the conditional frequency function of Y. thus (1.1) becomes:

$$f(x,y) = \left\{ \frac{\exp\left[-\frac{1}{2}\left(\frac{X-\mu}{\sigma} \right)^2 \right]}{\sqrt{2\pi\sigma^2}} \right\} \times \left\{ \frac{\exp\left[-\frac{1}{2}\left[\frac{(Y-\mu)-\rho(X-\mu)}{\sigma\sqrt{1-\rho^2}} \right]^2 \right]}{\sqrt{2\pi\sigma^2(1-\rho^2)}} \right\} \qquad (2.2.1)$$

Each of the four kinds of sample will include the same information regarding the X values in the treatment group so this part of function remains unchanged in each problem. Presence of any treatment effect will change the conditional part of the function. Now consider the selection of sample specimens from a population distributed according to (1.1) there are four possible kinds of sampling. The likelihood function requires a modification to reflect the way in which the normal density function is affected by the kind of sampling. Using (2.2.1) the likelihood function is:

$$L(\tau,\rho,\mu,\sigma) = G \times \frac{\exp\left[-\frac{1}{2}\sum_{1}^{n}\frac{(x_i-\mu)^2}{\sigma^2}\right]}{(2\pi\sigma^2)^{\frac{n}{2}}} \times \frac{\exp\left[\frac{-\frac{1}{2}\sum_{1}^{n}[(y_i-\mu)-\tau-\rho(x_i-\mu)]^2}{\sigma^2(1-\rho^2)}\right]}{(2\pi\sigma^2(1-\rho^2))^{\frac{n}{2}}} \quad (2.2.2)$$

Where G, is a restriction function; depending on the type of sample.

The following restriction functions are appropriate

Truncated sample

$$G = \{\frac{1}{[1-\phi(zj)]}\}^n \quad (2.2.3)$$

Censored sample

$$G = [\phi(zj)]^m \quad (2.2.4)$$

Selected sample

$$G = \left[\frac{1}{2\pi\sigma^2}\right]^{\frac{m}{2}} \exp\left[\frac{-\frac{1}{2}\sum_{n+1}^{n+m}(x_i-\mu)^2}{\sigma^2}\right] \quad (2.2.5)$$

And complete sample

$$G = \left[\frac{1}{2\pi\sigma^2}\right]^{\frac{m}{2}} \exp\left[\frac{-\frac{1}{2}\sum_{n+1}^{n+m}(x_i-\mu)^2}{\sigma^2}\right] \times \frac{1}{(2\pi\sigma^2(1-\rho^2))^{\frac{m}{2}}} \exp\{\frac{-\frac{1}{2}\sum_{n+1}^{n+m}[y_i-\rho x_i+(1-\rho)\mu]^2}{\sigma^2(1-\rho^2)}\} \quad (2.2.6)$$

Where $z = \frac{k-\mu}{\sigma}$, j=1 if the treatment group is above the truncation point, and j=-1 in the case where the treatment group is below the truncation point. $\phi(zj)$ Is the standard normal distribution function.

2.3. ESTIMATION

The maximum likelihood estimation can be derived from solution of the score equation system of any given type of problem. From log of equation (2.2.2):

$$S_{\mu} = -\frac{\partial}{\partial\mu} G(\mu, \sigma^2, \rho, j) + \sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma^2}\right) + \sum_{i=1}^{n}\frac{[y_i - \mu - \tau - \rho(x_i - \mu)][1-\rho]}{\sigma^2(1-\rho)^2}$$

$$S_{\sigma^2} = -\frac{\partial}{\partial\sigma^2} G(\mu, \sigma^2, \rho, j) + \sum_{i=1}^{n}\left(\frac{(x_i - \mu)^2}{2\sigma^2}\right) + \sum_{i=1}^{n}\left(\frac{[y_i - \mu - \tau - \rho(x_i - \mu)]^2}{\sigma^2(1-\rho)^2}\right) - \frac{n}{\sigma^2} \qquad (2.3.1)$$

$$S_{\rho} = -\frac{\partial}{\partial\rho} G(\mu, \sigma^2, \rho, j) + \sum_{i=1}^{n}\left(\frac{[y_i - \mu - \tau - \rho(x_i - \mu)][x_i - \mu]}{\sigma^2(1-\rho)^2} - \frac{[y_i - \mu - \tau - \rho(x_i - \mu)]^2}{\sigma^2(1-\rho)^3}\right) + \frac{np}{1-\rho^2}$$

$$S_{\tau} = \sum_{i=1}^{n}\frac{y_i - \mu - \tau - \rho(x_i - \mu)}{\sigma^2(1-\rho)^2}$$

This will involve the use of some numerical method such as Newton-Raphson.

2.4. TESTING THE PARAMETER EFFECT

The test involve regressing Y on X in the case of null hypothesis of no effect of treatment between before and after groups is:

$$H_0 : Y = \mu + \rho X + \varepsilon \qquad (2.4.1)$$

The regression coefficient for X in (2.4.1) equals the correlation ρ because we assume that the standard deviations for X and Y are equal. Assuming an additive treatment effect the alternative hypothesis is:

$$H_a : Y = \mu + \tau + \rho X + \varepsilon \qquad (2.4.2)$$

The information matrix $I_{[\mu,\sigma^2,\tau,\rho]}\left(\underset{\sim}{\theta}\right)$ is a 4×4 matrix with following array:

Where $\underset{\sim}{\theta}$ is the vectore of parameter values

$$I_{ij_{[\mu,\sigma^2,\tau,\rho]}}\left(\underset{\sim}{\theta}\right)=-E\left(\frac{\partial^2 \ln L}{\partial\theta_i\partial\theta_k}\right)$$

Where L is the likelihood function from (2.2.2).

For a random sample $R_1...R_n$ where $R_t=(x_i,y_i)$ from a bivariate density function $f(R,\theta_i)$, $\theta_i\in\Omega$ we seek a test of $H_0:\theta_i\in\Omega_0$ versus $H_a:\theta_i\in\Omega_1=\Omega-\Omega_0$. The likelihood-ratio denoted by λ, is defined to be:

$$\lambda=\frac{\sup_{\theta\in\Omega_0} L(\theta_i,R_1,...,R_n)}{\sup_{\theta\in\Omega} L(\theta_i,R_1,...,R_n)}$$

Under quite general regularity conditions $\Lambda=-2\ln\lambda$ is approximately distributed as a chi-squared distribution of parameters in $(\Omega-\Omega_0)$.

In our case, for testing the hypothesis of no treatment effect i, e., $H_0:\tau=0$ we have

$$\Omega=\left\{(\mu,\sigma^2,\rho,\tau):-\infty<\tau<\infty,\sigma^2>0,-1<\rho<1,-\infty<\mu<\infty\right\}\text{And}$$

$$\Omega_0\left\{(\mu,\sigma^2,\rho,\tau),\tau=0,\sigma^2>0,-1<\rho<1,-\infty<\mu<\infty\right\}$$

So the statistics Λ is distributed asymptotically as chi-square with degree of freedom equal to the number of parameters. After obtaining the information matrix the Wald's test for testing the null hypothesis $H_0:\tau=0$ is also available:

$$W=\frac{\hat{\tau}}{\hat{\text{var}}(\hat{\tau})}$$

George et.al[5] in the case of truncation sampling show that Wald's test does not appear to be reasonable for testing treatment effect in the presence of regression to the mean for moderate sample sizes and the LR test appears to be the appropriate choice.

2.5. INITIAL VALUES

The maximum likelihood estimates can be obtained by setting (2.3.1) equal to 0 and solving the system of any given type of problem by numerical methods. Caution should be taken because different choices of initial estimates as well as different methods of maximization may lead to local rather than global maxima. To avoid this possibility, we suggest two methods for choosing initial values. (The global maximum likelihood would be the maximum of these methods if both of them performed). Application of each method is depended on the additional in formation of population.

1- Moment estimation (James [6])

James proposed the moment method of estimation from a bivariate normal distribution in the case of truncation sampling. A principal assumption of this method is that the truncation proportion is assumed known. It is difficult to imagine the situation where the truncation proportion would be known. However if any value near the exact value of truncation point is available, then following solution is provided.

Let x_0 be the truncation point from baseline measurement define $z_0 = \frac{(x_0 - \mu)}{\sigma}$

and $k_0 = \frac{\phi(z_0 j)}{1 - \Phi(z_0 j)}$ be the truncation point (It is taken to be known). Then:

1) $\hat{\sigma}^2 = \frac{s^2{}_x}{k_0(x_0 - k_0) + 1}$

2) $\hat{\mu} = \bar{x} - k_0\hat{\sigma}$ ($\bar{x}$ and s_x^2 are sampling mean and variance respectively)

3) $\hat{\rho} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$

4) $\hat{\tau} = \bar{y} - \rho\bar{x}(1-\hat{\rho})\hat{\mu}$

2- Least square estimation (Mee and Chua [8])

Assuming μ is known or any value near the population mean is available, define:

$x_i' = x_i - \mu$ Then:

1) $\hat{\rho} = \hat{\beta} = \dfrac{\sum(x_i' - \bar{x}')(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$

2) $\hat{\tau} + \mu = \beta_{\hat{0}} = \bar{y} - \hat{\rho}\bar{x}'$ And $\hat{\sigma}^2 = \dfrac{\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i'\right)^2}{n-2}$.

3. AN EXAMPLE REVISITED

Florida high school students had to pass a literacy test before they receive a high school diploma. The students who failed the test could take a refresher course and then retake an equivalent test. McClave and Dietrich [7] gave paired sample data for eight students to evaluate the effectiveness of the refresher course (see table 1). They did not report the scores for the students or the number of students who passed. So it was a truncated sampling.

The usual paired-sample t-test has the following results:

$H_0 : \mu_2 - \mu_1 = 0$

$H_a : \mu_2 - \mu_1 > 0$

T=2; d.f=7

And the p-value for one-sided alternative hypothesis is P (t>2)=0.042. Although there was a significant increase in mean score of the subpopulation participated the refresh course, we want to know whether there is evidence of an intervention effect or the increase caused by regression to the mean.

Suppose that $\mu = 75$ then we can compute initial values from second method.

In terms of log scores Maximum likelihood estimation of treatment effect are obtained from numerical method:

$$\hat{\tau} = 0.066, \text{S.E} = 0.23$$

The Wald's test statistics for the null hypothesis $H_0 : \tau = 0$ were nonsignifficant ($p \cong 1$) and also the likelihood ratio statistics ($p \cong 1$).

We therefore conclude that there is no significant increase in score levels and beyond increasing is attributed to regression to the mean.

4. DISCUSSION

We first checked whether the distribution of x and y values were approximately normal. In fact the distributions were slightly skewed and the distribution of natural logarithmically transformed values was more nearly normal. Therefore we based the analysis on the log-transformed data. This had the disadvantage that the magnitude of the effect was more difficult to interpret because of the change of scale. In any case, we believed the transformation was important if we were to have meaningful results.

We were careful about making distinctions between the method of moment, least square and maximum likelihood, but in some cases the difference between approaches had been the handling of constraints. For example, an old paper of Fisher[4] shows that the ML solution for moments of the truncated univariate normal is the same as the method of moments. It is the use of constraints, for example, that forms the difference between this method and method of Mee and Chua[8].

Initial values proposed in this paper are suitable for a truncated sample. These values are capable for all other kinds of sampling although it will ignore some available data in doing so.

REFRENCES

1. Chen, S. and Cox, C. (1992). "Use of baseline data for estimation of treatment effects in the presence of regression to the mean," *Biometrics,* 48, 593 – 598.

2. Chen, S. and Cox, C. and Cui, L. (1998). "A more flexible regression-to- the mean model with possible stratification," *Biometrics,* 54, 939 – 947**.**

3. Cohen, A. C. (1955). "Restriction and selection in samples from bivariate normal distributions," *Journal of the American Statistical Association,* 50, 884 – 893**.**

4. Fisher, RA. (1931). "The sampling error of estimated deviates together with other illustration of the properties and applications of the integrals and derivatives of the normal error function," in *collected Papers of R.A. fisher*, vol.2, bennet, J. H., Ed., The University of Adelaide, Adelaide; pp. 517-526.

5. George, V., Johnson, W.D., Shahane, A. and Nick, T.G. (1997). "Testing for treatment effect in the presence of regression toward the mean," *Biometrics,* **53**, 49-59.

6. James, K. E. (1973). "Regression Toward the mean in uncontrolled clinical studies," *Biometrics,* 29 , 121 – 130**.**

7. McClave, J. T. and Dietrich, F. H. (1988). Statistics (4th ed.), San Francisco: Dellen,

8. Mee, R. W. and Chua, T.C. (1991). "Regression to the mean and the paired sample t-test," *The American Statistician,* **45***,* 39-42**.**

9. Pocock, S. J. (1983). *Clinical Trials*, Chichester: Wiley.

10. Senn, S. J. and Brown, R. A. (1985). "Estimating treatment effects in clinical trials subject to regression to the mean," *Biometrics,* 41, 555 – 560.

11. Senn, S. J. and Brown, R. A. (1989). "Maximum likelihood estimation of treatment effects for samples subject to regration to the mean," *Communications in Statistics-Theory and Methods,* **18**, 3389-3406.

Table 1. Educational.data

Student	Before	After	difference
1	45	49	4
2	52	50	-2
3	63	70	7
4	68	71	3
5	57	53	-4
6	55	61	6
7	60	62	2
8	59	67	8

Source: McClave and Dietrich