Unbiased Estimates if Treatment Effect in Randomized Experiments with Logistic Model and Omited Covariates

Ana Marí a Sfer
Universidad Nacional de Tucumán
Ayacuho 591 PB B
Tucumán Argentina
Asfer@herrera.unt.edu .ar

1. Introduction:

The problems of deciding whether a proposed innovation constitutes an improvement over some standard procedure arise in many different contexts. XXXX

In this paper, we study the role of randomization for the logistic model with omitted covariates. We suppose a model with randomization as a random variable (r.v.). We use the likelihood equations to write the parameter of this model as functions of parameters of the same model without randomization. We find that the model with randomization is the same as a reduce model without randomization and without covariates.

2. The complete model with randomization:

Suppose that Y is the response variable, X is the covariate taking two values, i.e.,

$$X = \begin{cases} x_{01} \\ x_{02} \end{cases} \qquad Y = \begin{cases} 1 \\ 0 \end{cases}$$

and T is the treatment variable

$$Y = \begin{cases} 1 \text{ if treatment is assigned} \\ 0 \text{ if control is assigned.} \end{cases}$$

Finally, we let R be a random a r.v. corresponding to the randomization

$$Y = \begin{cases} 1 \text{ if there is a random assignment} \\ 0 \text{ if there is not random assignment.} \end{cases}$$

Then T || X | R and Y || X | R, T.

Considering the following model, complete with randomization as r.v.

Y | X, T, R = Y | T, R ~ B(Π(X, T, R)) where

$$\Pi(X,T,R) = \frac{e^{\mu+\alpha T+\beta(1-R)X}}{1+e^{\mu+\alpha T+\beta(1-R)X}}$$

Following Gail et al., by the law of the large numbers the likelihood equations are:

$$E(Y) = E_{XTR}\left[E(Y|X;T;R)\right] \tag{1}$$

$$E(TY) = E_{XTR}\left[T\,E(Y|X;T;R)\right] \tag{2}$$

$$E(XY) = E_{XTR}\left[X\,E(Y|X;T;R)\right] \tag{3}$$

Then we can write

$$\begin{aligned} E(Y) &= E_{XTR}\left[E(Y|X;T;R)\right] \\ &= E_{XT}\left[E(Y|X;T;R=1)\right]1 + E_{XT}\left[E(Y|X;T;R=0)\right]0 \\ &= E(Y|T=1;R=1)\,P(T=1) + E(Y|T=-1;R=1)\,P(T=-1) \\ &= \frac{e^{\mu+\alpha}}{1+e^{\mu+\alpha}}P(T=1) + \frac{e^{\mu-\alpha}}{1+e^{\mu-\alpha}}P(T=-1) \end{aligned} \tag{4}$$

Analogously

$$E(TY) = \frac{e^{\mu+\alpha}}{1+e^{\mu+\alpha}}P(T=1) - \frac{e^{\mu-\alpha}}{1+e^{\mu-\alpha}}P(T=-1) \tag{5}$$

and

$$E(XY) = E_X\left[X\,\frac{e^{\mu+\alpha}}{1+e^{\mu+\alpha}}\right] P(T=1) + E_X\left[X\,\frac{e^{\mu-\alpha}}{1+e^{\mu-\alpha}}\right] P(T=-1)$$

Here is necessary to observe that this model, complete with randomization, is the same as the model without the covariate X, i.e., the reduce model.

3. The complete model without randomization

Suppose the same model as a section 2, but without the r. v. randomization. Then $Y \mid X, T \sim B(\Pi(X,T))$ where

$$\Pi(X,T,R) = \frac{e^{\mu+\alpha T+\beta X}}{1+e^{\mu+\alpha T+\beta X}}$$

Following Gail et al., the likelihood equations are:

$$\begin{aligned} E(Y) &= E_{XT}\left[E(Y|X;T)\right] \\ &= E_X(Y|T=1)\; P(T=1) + E_X(Y|T=-1)\; P(T=-1) \qquad (7) \\ &= E_X\left(\frac{e^{\mu+\alpha+\beta X}}{1+e^{\mu+\alpha+\beta X}}\; P(T=1)\right) + E_X\left(\frac{e^{\mu-\alpha+\beta X}}{1+e^{\mu-\alpha+\beta X}}\; P(T=-1)\right) \end{aligned}$$

If we set $h(z) = \dfrac{e^z}{1+e^z}$ then

$$E(Y) = E_X\,(h(\mu+\alpha+\beta X) P(T=1) + E_X(h(\mu-\alpha+\beta X)\; P(T=-1) \qquad (8)$$

and analogously

$$\begin{aligned} E(TY) &= E_{XT}\left[T\,E(Y|X;T)\right] \\ &= E_X(Y|T=1)\; P(T=1) - E_X(Y|T=-1)\; P(T=-1) \qquad (9) \\ &= E_X\,(h(\mu+\alpha+\beta X))\; P(T=1) - E_X\,(h(\mu-\alpha+\beta X))\; P(T=-1) \end{aligned}$$

4. Comparing of the models:

Now we write the parameters of the complete model with randomization as functions of parameters of the same model without randomization.

If we let α^* and μ^* be the parameters of the complete model with randomization and we set (4) = (8) and (5) = (9) the result is:

$$h(\mu^*+\alpha^*) P(T=1) + h(\mu^*-\alpha^*) P(T=-1) = E_X(h(\mu+\alpha+\beta X)) P(T=1) + E_X(h(\mu-\alpha+\beta X)) P(T=-1) \qquad (10)$$

and

$$h(\mu^*+\alpha^*) P(T=1) + h(\mu^*-\alpha^*) P(T=-1) = E_X(h(\mu+\alpha+\beta X)) P(T=1) + E_X(h(\mu-\alpha+\beta X)) P(T=-1) \qquad (11)$$

Then

$$h(\mu^*+\alpha^*) = E_X(h(\mu+\alpha+\beta X)) \quad \text{and} \quad h(\mu^*-\alpha^*) = E_X(h(\mu-\alpha+\beta X))$$

As we know, X takes on two values. Then we can write:

$$h(\mu^*+\alpha^*) = h(\mu+\alpha+\beta x_{01}) \qquad (12)$$

$$h(\mu^*+\alpha^*) = h(\mu+\alpha+\beta x_{01}) \qquad (13)$$

and it follows that

$$\mu^*+\alpha^* = \mu+\alpha+\beta x_{01} \qquad (14)$$

$$\mu^*+\alpha^* = \mu+\alpha+\beta x_{01} \qquad (15)$$

Then $\mu^* = \mu + \beta x_{01}$ and $\alpha^* = \alpha$ for X =x_{01} and analogously, $\mu^* = \mu + \beta x_{02}$ and $\alpha^* = \alpha$ for X =x_{02}. This means that, for any value of X, we have $\alpha^* = \alpha$.

5. Simulations

We do 9 simulation with 1000 cases, with and without randomization and we fit complete and reduced model. Then we calculate the difference between the treatment effect in each model, i.e.

Effect Treatment in complete model - Effect Treatment in reduced model.

We found that these differences are approximately zero when there is random assignment and the opposite occurs when there is not random assignment.

We considering $\Pi < 0.5$, $\Pi = 0.5$ and $\Pi > 0.5$, the effect of covariate mayor, equal and minor than treatment effect. The results are showed in Table 1and a descriptive statistics are in Table 2.

Randomization	
With	**Without**
0.0015307	0.41578748
-0.00096822	0.1970432
-0.00357731	0.66810431
0.0012976	-0.15867904
0.0054882	0.12638548
0.002788	0.1899339
0.00328511	-0.39469908
-0.00376429	0.41485533
6.0417E-05	0.18034691

Tabla 1: Difference between treatment effect, complete model - reduce model, when there is and there is not random assignment.

Statistics	**Randomization**	
	With	**Without**
Mean	0.00068225	0.18211983
Mediana	0.0012976	0.1899339
Estándar deviation	0.00309093	0.31567476
Varianza	9.5539E-06	0.09965056
Mí nimum	-0.00376429	-0.39469908
Máximum	0.0054882	0.66810431
Confidence Interval(95.0%)	0.0023759	0.24264926

Table 2: Descriptive statistics for the difference if treatment effects, complete model – reduced model, when there is and there is not random assignment.

REFERENCES

Fisher, R.(1966). The design of experiemtns, 8th edition. Edinburgh: oliver and Boyd.

Gail, M. H. Wiend, S. And Piantadosi, S. (1984). Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omited covariates. Biometrika, 71, 3, pp. 431-44.

Hosmer, D. W. And Lemeshow, S. (1989). Applied Logistic Regression. Jhon Wiley & Sons.

Kempthorne, O. (1977). Why randomize?. J. Statis. Plan. Infer. 1, 1-25.

Lehman, E. L. (1975). Nonparametrics: Statistical Methods based in ranks. San Francisco: Holden – Day.

Cox, D. R. (1958). Planning of Experiments. New York. Wiley.

RESUME

The problems of deciding whether a proposed innovation constitutes an improvement over some standard procedure arise in many different contexts. For example: Does a new cure prolong the life of cancer patients? Does a new expensive gasoline additive increase the mileage? Is televised instruction less effective than live classroom teaching?

In many cases we have experimental units to assign the new procedure or to the standard one. If this assignment is done using randomization, the responses to different treatment can typically be shown to be independent of other initial characteristics of the units. This is a major reason why randomization is widely accepted as a critical element in the design of experiment (Hill, 1960; Byar et al., 1970)

Randomization produces comparable or balanced groups with a large number of experimental units and provides a valid bases to estimates and evaluate the treatment effect (Fisher, 1966, pp. 19-21; Kempthorne, 1977; Lehman, 1975, Chapter 1).In the context of the normal linear model, randomization also leads to unbiased estimates of an additive treatment effect, even when important covariates are unknown or unmeasured (Cox, 1958, Chapter 5). In the context of the logistic model it is no clear from the literature whether randomization plays a similar role (c.f., Gail et al., 1984).

In this paper, we study the role of randomization for the logistic model with omitted covariates. We introduce the randomization via a dummy variable and we take advantage of conditional independence to remove the effect of the covariates in the maximum likelihood estimates for the treatment effect. We also present simulations that confirm the role of randomization in removing the effect of the covariates.