العلوم الاقتصادية العدد(٢٧) المجلد السابع كانون الأول ٢٠١١

Design Of Analysis Covariance Table for Tata Designed by Balanced Incomplete Block Design when Relation Exists Between Block and Covariate variable

تصميم جدول تحليل التغاير لبيانات تجربة مصممة بأسلوب القطاعات غير الكاملة المتزنة عند و جود علاقة بين القطاع و المتغير المرافق

م . م علي ناصر حسين
قسم الإحصاء / كلية الإدارة والاقتصاد
جامعة البصرة

المستخلص

يسعى أي لباحث عند تصميم التجربة إلى تقليل الخطاء العشوائي للتجربة و يتم ذلك بتجميع الوحدات المتشابه في قطاعات قبل إجراء التجربة ، و لإعطاء نتائج أكثر دقة يتم استخدام تحليل التغاير للتخلص من تأثير المتغيرات المرافقة أو المصاحبة و يستخدم هذا الإجراء بعد تنفيذ التجربة. في هذه الدراسة يتم الجمع بين الأسلوبين لتقليل الخطاء العشوائي.

Abstract

When an experiment is designed, random error is preferred to be reduced as much as possible. This is achieved by collecting similar plots into blocks before the experiment is executed. For more precise results, Analyses of covariance is used to get rid of variable covariate effect. This method is used after achieving the experiment .In this study

تم استلام البحث بتاريخ 2010/١١/١١ تم قبول البحث بتاريخ 2011/1/16

1- *Objective of Research*

The aim of this study is to construct the analyses of covariance table for data of an experiment designed by balanced incomplete block design when relation between the covariate variable and block takes place.

2- *Importance of Research*

The importance of this research comes through

a) Reducing the experimental random error by two methods

- *Treating the problem of small block size using balanced incomplete block design, this design is one of several designs which can be used for treating small sizes of blocks which cannot contain the treatments.*
- *Using analysis of covariance method to riddance of covariate variable effect.*

b) *Analysis of data experiment designed by balanced incomplete block when relationship between the covariate variable and block is existed. This helps to reduce experimental error.*

3- *Research hypothesis*

Suppose there is an experiment designed by balanced incomplete block, and suppose there is relationship between the covariate variable and block, where mathematical model of this design is given by::

$$1y_{ij} = \mu + t_i + b_j + B_j x_{ij} + e_{ij}$$

This study requires the following

$\sum_{i=1}^{t} t_i = 0$ $\sum_{j=1}^{k} b_k = 0$ $e_{ij} \sim N(\mu, \sigma^2)$

4- *Introduction*

The balanced incomplete block design is used for solving uncountable treatments within a block , such treatments are needed to compare their means among each other . The reason behind uncountable treatments within a block belongs to the following

1- *Small size of the block such that the plots in this block cannot contain all treatments of experiment.*

2- *The treatments of the experiment are big such that there is no block can be found to contain them in its homogeneous plots.*

A design is called balanced incomplete block on the following condition

1- *Each block has the same number of plots(k).*
2- *Each treatment occurs for the same number of times (r).*
3- *Each pair of treatments occurs together in all blocks for the same number of times (λ).*

The following relationship must be met for a balanced incomplete block design to be possible:

$$\boldsymbol{t \times r = b \times k = N}$$

Balanced concept comes from the following relationship:

$$\lambda = \frac{r(k-1)}{(t-1)}$$

Many of researchers study balanced incomplete block design because this design is one of the most important designs. Previous studies do not mention the relationship between regression coefficient and block.

5- *Analysis of covariance*

The analysis of covariance (generally known as ANCOVA) is a technique that sits between analysis of variance and regression analysis .it has a number of purposes but the two that are, perhaps ,of most importance are:

a) *To increase the precision of comparisons between groups by accounting to variation on important prognostic variables.*
b) *to "adjust" comparisons between groups for imbalance in important prognostic variable between these groups.*

6- *Theoretical Analysis*

The mathematical model for balanced incomplete block design in the case of existence of covariate variable and relation between the regression coefficient and block is :

$$y_{ij} = \mu + t_i + b_j + B_j x_{ij} + e_{ij} \qquad ٢$$

where

y_{ij}: *Independent variable observations in (j) block under (i) treatment.*

x_{ij} : *covariate variable observation in (j) block under (i) treatment.*

μ: *general average,* t_i: *treatment effect (i) ,* b_j : *block effect (j)*

B_j: *regression coefficient which is affected by block (j).*

The total sum of squares error is:

$$\sum_{ij} n_{ij} e_{ij}^2 = \sum_{ij} n_{ij} \left(y_{ij} - \mu - t_i - b_j - B_j x_{ij}\right)^2$$

The normal equations for the model of equation (1) are:

$$\sum_{ij} n_{ij}\, y_{ij} = bk\mu + \sum_{ij} n_{ij}\, B_j x_{ij} \qquad 3$$

$$y_{i.} = r\mu - rt_i + \sum_j n_{ij}\, b_j + \sum_j n_{ij}\, B_j x_{ij} \qquad 4$$

$$y_{.j} = k\mu + \sum_i n_{ij}\, t_i + kb_j + B_j x_{.j} \qquad 5$$

$$\sum_i n_{ij}\, y_{ij} x_{ij} = \mu x_{.j} + \sum_i n_{ij} t_i\, x_{ij} + b_j\, x_{.j} + \sum_i n_{ij}\, B_j x_{ij}^2 \qquad 6$$

From equations (4) & (5)

$$y_i - \frac{1}{k}\sum_j n_{ij} y_{.j} = \sum_j n_{ij} B_j x_{ij} - \frac{1}{k}\sum_j n_{ij} B_j x_{.j} + rt_i - \frac{1}{k}\sum_{ij} n_{ij} n_{is} t_{is} \qquad 7$$

Simplifying the quantity $\sum_{ij} n_{ij} n_{is} t_{is}$ *depending on idiosyncrasy of balanced incomplete block design yields*

$$rt_i - \sum_{ij} n_{ij} n_{is} t_{is} = \hat{t}_i \left(\frac{r-\lambda}{k}\right)$$

Substituting $\hat{t}_i \left(\frac{r-\lambda}{k}\right)$ *in equation (7) gives*

$$y_{i.} - \frac{1}{k}\sum_j n_{ij} y_{.j} = t_i \left(\frac{\lambda t}{k}\right) + \sum_j n_{ij} B_j x_{ij} - \sum_j n_{ij} B_j x_{.j} \qquad 8$$

To estimate regression coefficient, equations (5) &(6) are use as follows

$$\sum_j n_{ij} y_{ij} x_{ij} - \frac{1}{k} y_{.j} x_{.j} = \sum_i n_{ij}\, t_i \left(x_{ij} - \frac{1}{k} x_{.j}\right) + B_j \left(\sum_i n_{ij}\, x_{ij}^2 - \frac{1}{k}\left(x_{.j}\right)^2\right) \qquad 9$$

The treatment effect can be divided by using the following equation[4]

$$t_i = t_{iy} - B_j t_{ix} \qquad 10$$

where :

t_i: *total treatment effect (i)*

t_{iy}: *treatment effect (i) in case of ignoring covariate variable (* $B_j = 0$*)*

t_{ix} *: treatment effect (i) in case of exchanging covariate variable by accompaniment variable.*

Substituting equation (10) in equation (9) yields

$$\sum_j n_{ij} y_{ij} x_{ij} - \frac{1}{k} y_{.j} x_{.j} - \sum_i n_{ij} t_{iy} \left(x_{ij} - \frac{1}{k} x_{.j} \right) = B_j \left(\left(\sum_i n_{ij} x_{ij}^2 - \frac{1}{k} (x_{.j})^2 \right) - \left(\sum_i n_{ij} t_i \left(x_{ij} - \frac{1}{k} x_{.j} \right) \right) \right) \qquad 11$$

The sum of squares of error for the model in equation (1) is calculated as follows

$$\sum_{ij} n_{ij} e_{ij}^2 = \sum_{ij} n_{ij} \left(y_{ij} - \mu - t_i - b_j - B_j x_{ij} \right)^2 \qquad 12$$

Assuming $\sum_{ij} n_{ij} e_{ij}^2 = E_{yy}$ *, and substituting it in equation (12) gives*

$$E_{yy} = \left(\sum_{ij} n_{ij} y_{ij}^2 - \frac{y_{ij}^2}{bk} \right) - \left(\frac{1}{k} \sum_i (y_{.j})^2 - \frac{y_{ij}^2}{bk} \right) w - \sum_i t_i \left(y_{i.} - \frac{1}{k} \sum_j n_{ij} y_{.j} \right) - \sum_j B_j \left(\left(\sum_i n_{ij} x_{ij} y_{ij} - \frac{y_{ij}^2}{bk} \right) - \left(\frac{1}{k} y_{.j} x_{.j} - \frac{y_{ij}^2}{bk} \right) \right) \qquad 13$$

where other terms are zeros,

From equations (10) & (12)

$$E_{yy} = \left(\sum_{ij} n_{ij} y_{ij}^2 - \frac{y_{ij}^2}{bk} \right) - \left(\frac{1}{k} \sum_i (y_{.j})^2 - \frac{y_{ij}^2}{bk} \right) - \sum_i t_{iy} \left(y_{i.} - \frac{1}{k} \sum_j n_{ij} y_{.j} \right) - \sum_j B_j \left(\left(\sum_i n_{ij} x_{ij} y_{ij} - \frac{y_{ij}^2}{bk} \right) - \left(\frac{1}{k} y_{.j} x_{.j} - \frac{y_{ij}^2}{bk} \right) \right) + \sum_i B_j t_{ix} \left(y_{i.} - \frac{1}{k} \sum_j n_{ij} y_{.j} \right) \qquad 14$$

The equation above represents the sum of squares of error which is the sum of total squares of variable observation (y) minus the sum of squares of a block for variable observation (y) and sum of total squares of results of multiplying variable observation (x) and variable observation (y) in addition to the sum of squares of plots for results multiplication minus covariate variable error.

To complete the analysis of covariance table the null hypothesis

$H_0 = t_1 = t_2 = \cdots = t_t = 0$

is put against the following alternative hypothesis

$H_1 = t_1 \neq t_2 \neq \cdots \neq t_t \neq 0$

The model in equation (1) according to the null hypothesis becomes.

$$y_{ij} = \bar{x} + \alpha_j + \beta_j x_{ij} + \epsilon_{ij} \qquad 15$$

where :

$\bar{x}$*: is the general average reduced model*

α_j *: effect of block (j) in reduced model* β_j *: regression coffecint for reduced model.*

The normal equations for reduced model will be

$$y_{..} = bk\bar{x} + \textstyle\sum_j \beta_j x_{.j} \qquad 16$$

$$y_{.j} = k\bar{x} + k\alpha_j + \beta_j x_{.j} \qquad 17$$

$$\textstyle\sum_i n_{ij} y_{ij} x_{ij} = \bar{x}x_{.j} + \alpha_j x_{.j} + \sum_i n_{ij} \beta_j x_{ij}^2 \qquad 18$$

Regression coefficient for reduced model is estimated from equations (17) &(18) as

$$\textstyle\sum_j n_{ij} y_{ij} x_{ij} - \frac{1}{k} y_{.j} x_{.j} = \beta_j \left(\sum_i n_{ij} x_{ij}^2 - \frac{1}{k}(x_{.j})^2\right) \qquad 19$$

The sum of squares of error for reduced model is

$$\textstyle\sum_{ij} n_{ij} e_{ij}^2 = \sum_{ij} n_{ij} \left(y_{ij} - \bar{x} - \alpha_j - \beta_j x_{ij}\right)^2 \qquad 20$$

$$e_{yy} = \left(\textstyle\sum_{ij} n_{ij} y_{ij}^2 - \frac{y_{ij}^2}{bk}\right) - \left(\frac{1}{k}\sum_i (y_{.j})^2 - \frac{y_{ij}^2}{bk}\right) - \sum_j \beta_j \left(\left(\sum_i n_{ij} x_{ij} y_{ij} - \frac{y_{ij}x_{ij}}{bk}\right) - \left(\frac{1}{k} y_{.j} x_{.j} - \frac{y_{ij}x_{ij}}{bk}\right)\right) \qquad 21$$

where $\sum_{ij} n_{ij} e_{ij}^2 = e_{yy}$

The adjusted sum of squares of error effect of blocks & covariate variable is

$$\tilde{\bar{E}}_{yy} = e_{yy} - E_{yy} \qquad 22$$

where $\tilde{E}_{yy}$ *represents adjusted sum of squares error from effect blocks & covariate variable*

Therefore the adjusted sum of squares error from effect blocks & effect covariate variable can be written as

$$\tilde{E}_{yy} =$$

$$\textstyle\sum_i t_i \left(y_{i.} - \frac{1}{k}\sum_j n_{ij} y_{.j}\right) - \sum_j \dfrac{\left(\left(\sum_i n_{ij} y_{ij} x_{ij} - \frac{y_{ij}x_{ij}}{bk}\right) - \left(\frac{1}{k} y_{.j} x_{.j} - \frac{y_{ij}x_{ij}}{bk}\right)\right)^2}{\left(\sum_i n_{ij} x_{ij}^2 - \frac{x_{ij}^2}{bk}\right) - \left(\frac{1}{k}(x_{.j})^2 - \frac{x_{ij}^2}{bk}\right)} +$$

$$\textstyle\sum_j \dfrac{\left(\left(\sum_i n_{ij} y_{ij} x_{ij} - \frac{y_{ij}x_{ij}}{bk}\right) - \left(\frac{1}{k} y_{.j} x_{.j} - \frac{y_{ij}x_{ij}}{bk}\right)\right)^2 - \sum_i t_{iy}\left(x_{ij} - \frac{1}{k}x_{.j}\right)\left(\sum_i n_{ij} x_{ij} y_{ij} - \frac{1}{k} x_{.j} y_{.j}\right)}{\left(\left(\sum_i n_{ij} x_{ij}^2 - \frac{x_{ij}^2}{bk}\right) - \left(\frac{1}{k} x_{.j}^2 - \frac{x_{ij}^2}{bk}\right)\right) - \left(\sum_i t_{ix}\left(x_{ij} - \frac{1}{k}\sum_j x_{.j}\right)\right)} \qquad 23$$

So, the table of analysis will be as follows

Table number (1): Analysis of covariance table when relationship between covariate variable and regression coefficient is found

Source of variation	*D.f*	*Unadjusted sum of*	*D.f*	*Adjusted sum of*	*m.s*	*f*

		squares				squares		
		xx	xy	yy				
Block(ignoring treatment)	b-1	φ_{xx}	φ_{xy}	φ_{yy}				$\dfrac{\widehat{T_{yy}}\big/_{t-1}}{\widehat{E_{yy}}\big/_{N-b-}}$
Treatments(eliminating block)	t-1	$\emptyset_{xx}$	$\emptyset_{xy}$	$\emptyset_{yy}$	t-1	$\widehat{T_{yy}}$	$\widehat{T_{yy}}\big/_{t-1}$	
Intra-block error	N-b-t+1	e_{xx}	e_{xy}	e_{yy}	N-b-t	$\widehat{E_{yy}}$	$\widehat{E_{yy}}\big/_{N-b-}$	
Total	N-1	A_{xx}	A_{xy}	A_{yy}				
Treatment + error		$\widetilde{T}_{xx}$	$\widetilde{T}_{xy}$	$\widetilde{T}_{yy}$		ω_{yy}		

Note :This table is prepared by the researcher.

$$\varphi_{xx} = \frac{1}{k}\sum_j (x_j)^2 - \frac{x_{ij}^2}{bk}, \quad \varphi_{xy} = \frac{1}{k}\sum_j (x_j y_j)^2 - \frac{x_{ij}\,y_{ij}}{bk}, \quad \varphi_{yy} = \frac{1}{k}\sum_j (y_{.j})^2 - \frac{y_{ij}^2}{bk}$$

$$\emptyset_{yy} = \sum_i t_i \left(y_{i.} - \frac{1}{k}\sum_j n_{ij} y_{.j}\right), \qquad \emptyset_{xy} = \sum_i t_{ix} \left(y_{i.} - \frac{1}{k}\sum_j n_{ij} y_{.j}\right)$$

$$\emptyset_{xx} = \sum_i t_i \left(x_{i.} - \frac{1}{k}\sum_j n_{ij} x_{.j}\right)$$

$$e_{yy} = \left(\sum_{ij} n_{ij} y_{ij}^2 - \frac{y_{ij}^2}{bk}\right) - \varphi_{yy} - \emptyset_{yy}\ ,\ e_{xx} = \left(\sum_{ij} n_{ij} x_{ij}^2 - \frac{x_{ij}^2}{bk}\right) - \varphi_{xx} - \emptyset_{xx}$$

$$e_{xy} = \left(\sum_{ij} n_{ij} y_{ij}\, x_{ij} - \frac{y_{ij}\, x_{ij}}{bk}\right) - \varphi_{xy} - \emptyset_{xy}$$

$$A_{yy} = \left(\sum_{ij} n_{ij} y_{ij}^2 - \frac{y_{ij}^2}{bk}\right) \qquad A_{xx} = \left(\sum_{ij} n_{ij} x_{ij}^2 - \frac{x_{ij}^2}{bk}\right)$$

$$A_{xy} = \left(\sum_{ij} n_{ij} y_{ij} x_{ij} - \frac{y_{ij}x_{ij}}{bk}\right)$$

$$\widetilde{T}_{yy} = \emptyset_{yy} + \varphi_{yy} \qquad \widetilde{T}_{xx} = \emptyset_{xx} + \varphi_{xx} \qquad \widetilde{T}_{xy} = \emptyset_{xy} + \varphi_{xy}$$

$$\omega_{yy} = \widetilde{T}_{yy} - \frac{\widetilde{T}_{xy}^{\,2}}{\widetilde{T}_{xx}}$$

$$\widehat{E_{yy}} = \sum_j \frac{\left(\left(\sum_i n_{ij} y_{ij} x_{ij} - \frac{y_{ij}x_{ij}}{bk}\right) - \left(\frac{1}{k} y_{.j} x_{.j} - \frac{y_{ij}x_{ij}}{bk}\right)\right)^2 - \sum_i t_{iy}\left(x_{ij} - \frac{1}{k}x_{.j}\right)\left(\sum_i n_{ij} x_{ij} y_{ij} - \frac{1}{k}x_{.j} y_{.j}\right)}{\left(\left(\sum_i n_{ij} x_{ij}^2 - \frac{x_{ij}^2}{bk}\right) - \left(\frac{1}{k}x_{.j}^2 - \frac{x_{ij}^2}{bk}\right)\right) - \left(\sum_i t_{ix}\left(x_{ij} - \frac{1}{k}\sum_j x_{.j}\right)\right)}$$

7-Application side :

Application side of this study includes an experiment designed by balanced incomplete block where the researcher poplared an experiment executed by college of agriculture at bagdad university . To compare the effects of proper generation of subjective inoculation (t1 ,t2 ,t3 , t4, t5) through testing the pure bloodlines of corn. The recorded data represent vegetable height (y) and length of outgrowth (x). Table (2) show these measurements (in cm) where the number between brackets represents the dependent variable (height of vegetable) and the number on the left represents the covariate variable (length of outgrowth).

Table number (2) represented covariate variable observation & depended variable observation

		treatments				
		1	*2*	*3*	*4*	*5*
blocks	*1*	*74 (190*	*77 (195*	*75 (185*	-	-
	2	*76 (175*	*69 (180*	-	*73 (190*	-
	3	*75 (200*	*67 (150*	-	-	*75 (195*
	4	*75 (195*	-	*76 (165*	*80 (165*	-
	5	*75 (170*	-	*80 (165*	-	*79 (170*
	6	*79 (225*	-	-	*74 (160*	*71 (200*
	7	-	*77 (175*	*69 (180*	*69 (190*	-
	8	-	*74 (145*	*73 (170*	-	*78 (190*
	9	-	*70 (165*	-	*79 (200*	*74 (185*
	10	-	-	*79 (180*	*73 (190*	

8-Statistical analysis

Since the experiment of this study is designed by balanced incomplete block and the number of treatments in the experiment are (t=5), and each treatment is repeated for six times (r=6) in ten blocks (b=10) where each block capacity is three plots (k=3), the appearance of any pair of treatments is three ($\lambda = 3$) , then the condition of balance is achieved through

1- $tr = bk = 30$

2- $\lambda(t-1) = r(k-1) = 12$

Using equations in theoretical analysis to analyze the data in table(2), results in table (3) are obtained.

Table (3) Analysis of covariance table assuming existence of relationship between regression coefficient and block

Source of variation		*Unadjusted sum of squares*				*Adjusted sum of squares*		
	d.f	*xx*	*xy*	*yy*	*d.f*		*m.s*	*F*
Block(ignoring treatment)	9	108.8	-180.333	1917.5				
Treatments(eliminating block)	*5*	*18.3111*	*143.556*	*2267.778*	*5*	*981.09*	*196.218*	*19.0912*
Intra-block error	*15*	*262.356*	*686.886*	*4132.23*	*14*	*143.891*	*10.27793*	
Total	*29*	*389.467*	*363*	*8317.5*				
Treatment + error		*56.424*	*543.33*	*6356.93*	*19*	*1124.981*		

9- Conclusion

1. *From application side, the use of covariance analysis leads to reduce random error of experiment because it removes the effect of two error sources, the first is the effect of difference among blocks, while the other source is the effect of covariate variable related to block.*
2. *From analysis of experimental data taking into account the relationship of covariate variable with the block gives sum of squares of error less then what it is on ignoring this relationship. So, from table (3) we find that the sum of squares of error reaches to (143.891) otherwise, it will be (*2333.86*) by ignoring this relation and analyzing the data.*

10 - Recommendation :

1. *From measuring the degree of correlation between the blocks and regression coefficient, if correlation is strong, analysis of experimental data must be achieved using table (1), otherwise if this relationship is weak, analysis of data is achieved without taking into account the existence of this relationship.*
2. *Studying another incomplete designs like partially balanced design, and constructing analysis of covariance table when relationship is existed.*

Foreign References

1- *Jerome, L, Myers.(1979)."Fundamental of Experiment design" Third edition.*
2- *Kempthorne , O.(1952)."The Design & Analysis of Experiment" Wily ,New York.*
3- *Margan, J.P.(1996)"Nested design " S .Ghosh & C.R.Rao.Eds.,handbook of stat.vol.13.*
4- *Zelen, Marvin,(1957)."The analysis of covariance for incomplete block designs" Biometrics، vol.13، no.2.pp503-24.*

Arabic References

١ - *حسين ، علي ناصر "دراسة احصائية في تحليل التغاير لبعض تصاميم التجارب غير الكاملة المتزنة "رسالة ماجستير مقدمة الى كلية الادارة و الاقتصاد ، جامعة بغداد ٢٠٠٢،*

٢- الحمزاوي ، رحيم جبار ظاهر "دراسة في تصاميم القطاعات غير الكاملة المتزنة جزئيا مع تطبيق عملي " رسالة ماجستير مقدمة الى كلية الادارة و الاقتصاد ،جامعة بغداد،١٩٨٨ .

٣- الراوي ،د.خاشع محمود ، د.عبد العزيز محمد " تصميم و تحليل التجارب الزراعية"مؤسسة دار الكتب للطباعة و النشر،جامعة الموصل ،١٩٨٠ .

٤- المشهداني ،د.محمود حسن ، كمال علوان خلف "تصميم و تحليل التجارب"بيت الحكمة للنشر و الترجمة و التوزيع ،جامعة بغداد ، ١٩٨٩ .

The international information network (internet)

1- http://ansc.umd.edu/wwwfaculty/Douglass/Lab%20Notes/Covariance%20Notes%20-%20student%20version.pdf

2- http://www.for.gov.bc.ca/hre/biopamph/pamp53.pdf

3- http://www.mas.ncl.ac.uk/~njnsm/medfac/docs/ancova.pdf

4- http://media.wiley.com/product_data/excerpt/75/04715517/0471551775.pdf

5- -http://www.stat.purdue.edu/~bacraig/notes1/topic13.pdf

6- http://www.stat.yale.edu/Courses/1997-98/312/Solution5.pdf