Chapter 9 The Analysis of Covariance

- 9.1 Single factor Covariance Model
  单因素协方差模型
- 9.2 Multifactor Covariance Mode
  多因素协方差模型

1

- Analysis of covariance (ANOCOVA) is a technique that combines features of analysis of variance and regression.
- Covariance is used when the response variable y, in addition to being affected by the treatments, but also related to another variable x.
- It augment the analysis of variance model containing the factor effects with one or more additional quantitative variables that are related to the dependent variable.

2

减少残差的效应.

- It intended to reduce the variance of the error terms in the model.
- ANOCOVA utilizes the relationship between the dependent variable and one or more independent quantitative variables for which observations are available in order to reduce the error term variability and make the study a more powerful one for comparing treatment effects.

3

- Each independent quantitative variable added to the study is called a concomitant variable (相伴变量). 协变量
- The choice of concomitant variables is an important one.
- If such variables have no relation to the dependent variable, nothing is to be gained by covariance analysis.
- Concomitant variables frequently used with human subjects include attitudes, age, IQ, and pre-study aptitude.

4

①协变量与依变量是线性 ②不同组间残差方差相同.且独立.正态. ③不同组间回归系数一致.

9.1 Single-factor Covariance Model

Case Studied (1)

- A company wished to study the effects o: three different types of promotions on sales of its crackers. 不同促销机制对销量的影响.
  - Treatment I: Sampling of product by customers in store and regular shelf space
  - Treatment 2: Additional shelf space in regular location
  - Treatment 3: Special display shelves at ends of aisle in addition to regular shelf space

5

- Fifteen stores were selected for the study.
- Each store was randomly assigned one of the promotion types, with five stores assigned to each type of promotion.
- Other relevant conditions under the control of the company, such as price and advertising, were kept the same for all stores in the study.
- Dependent Variable Y: sales of the product during the promotional period
- Concomitant variable X: sales of the product in the preceding period

6

促销前的销售量.

	Study Unit (j)									
Treatment	1		2		3		4		5	
i	Y	X	Y	X	Y	X	Y	X	Y	X
1	38	21	39	26	36	22	45	28	33	19
2	43	34	38	26	38	29	27	18	34	25
3	24	23	32	29	31	30	21	16	28	29

$$Y_{ij} = \mu + \beta X_{ij} + \overset{\text{Treatment}}{\alpha_i} + \varepsilon_{ij}$$
$$i = 1, 2, \ldots, a; j = 1, 2, \ldots, r;$$

7

```
Data Cracker;
Input Treat @;
  Do R = 1 to 5;
    Input Y X @;
    Output;
    End;
Datalines;
1 38 21 39 26 36 22 45 28 33 19
2 43 34 38 26 38 29 27 18 34 25
3 24 23 32 29 31 30 21 16 28 29
;
PROC GLM;
Class Treat;
MODEL Y = Treat;
LsMeans Treat/STDERR PDIFF Adjust=Tukey;
RUN;
PROC GLM;
Class Treat;
MODEL Y = X Treat;
LsMeans Treat/STDERR PDIFF Adjust=Tukey;
RUN;
```

Note: the data is input in the do...end loop, where the output means that an observation will be output during each loop

ANOVA

ANCOVA 若放入X，就变成ANOVA了.

8

Table of ANOVA

ANOVA Table for Model without Covariate (Type III)

Source	df	SS(Adj)	MS(Adj)	F value	P-value
$R(\alpha\|\mu)$	$t-1=2$	338.8000000	169.4000000	6.61	0.0116
$\text{Error}(\mu,\alpha)$	$t(r-1)=12$	307.6000000	25.6333333		

ANOVA Table for Model with One Covariate (Type III)

Source	df	SS(Adj)	MS(Adj)	F value	P-value
$R(\beta\|\mu,\alpha)$	1	269.0286915	269.0286915	76.72	<.0001
$R(\alpha\|\mu,\beta)$	$t-1=2$	417.1509137	208.5754568	59.48	<.0001
$\text{Error}(\mu,\alpha,\beta)$	$t(r-1)-1=11$	38.5713085	3.5064826		

X与Y有回归关系

三种处理机制之间有显著差异

残差MS减少.

9

Analysis without Covariate

Treat	Y LSMEAN	SE	Pr>|t|	Number
1	38.2000000	2.2642144	<.0001	1
2	36.0000000	2.2642144	<.0001	2
3	27.2000000	2.2642144	<.0001	3

i/j	1	2	3
1		0.7753	0.0127
2	0.7753		0.0434
3	0.0127	0.0434	

ANOVA

不显著

Analysis with One Covariate

Treat	Y LSMEAN	SE	Pr > |t|	Number
1	39.8174070	0.8575507	<.0001	1
2	34.7420168	0.8496605	<.0001	2
3	26.8405762	0.8384392	<.0001	3

i/j	1	2	3
1		0.0044	<.0001
2	0.0044		<.0001
3	<.0001	<.0001	

ANCOVA

不同促销方法之间有显著差异.

10

Case Studied (2)

- A study on the effects of two treatments and control to the learning-disable children.

Groups					
Treatment 1		Treatment 2		Control	
Pre	Post	Pre	Post	Pre	Post
85	100	86	92	90	95
80	98	82	99	87	80
92	105	95	108	78	82
257	303	263	299	255	257

11

一开始不同组的IQ可能就不一样

```
Data Child;
Input Treat $ Pre Post @@;
Datalines;
Trt1  85   100  Trt1  80   98
Trt1  92   105  Trt2  86   92
Trt2  82   99   Trt2  95   108
Trt3  90   95   Trt3  87   80
Trt3  78   82
;
PROC GLM;
Class Treat;
MODEL Post = Treat;
LsMeans Treat/STDERR PDIFF Adjust=Tukey;
RUN;
PROC GLM;
class Treat;
MODEL Post = Pre Treat;
LsMeans Treat/STDERR PDIFF Adjust=Tukey;
RUN;
```

ANOVA

ANCOVA 协变量

12

Table of ANOVA

ANOVA Table for Model without Covariate (Type III)

Source	df	SS(Adj)	MS(Adj)	F value	P-value
T|μ	2	432.8888889	216.4444444	4.52	0.0635 不显著?
Error	6	287.3333333	47.8888889		

ANOVA Table for Model with One Covariate (Type III)

Source	df	SS(Adj)	MS(Adj)	F value	P-value
X|μ, T	1	137.8946147	137.8946147	4.61	0.0845
T|μ, X	2	366.2012282	183.1006141	6.13	0.0452 显著.
Error	5	149.4387187	29.8877437		

是有疗效的.

13

Analysis without Covariate

Treat	Y LSMEAN	SE	Pr>|t|	Number
1	101.000000	3.995368	<.0001	1
2	99.666667	3.995368	<.0001	2
3	85.666667	3.995368	<.0001	3

i/j	1	2	3
1		0.9699	0.0778
2	0.9699		0.1050
3	0.0778	0.1050	

Analysis with One Covariate

Treat	Y LSMEAN	SE	Pr > |t|	Number
1	101.337357	3.160261	<.0001	1
2	98.485918	3.203866	<.0001	2
3	86.510059	3.180684	<.0001	3

i/j	1	2	3
1		0.8105	0.0467
2	0.8105		0.1004
3	0.0467	0.1004	

排除X的效应.

14

第一种与第二种差别不大.
第一种与Control有显著差异. 有治疗效果.

9.2 Multifactor Covariance Model

- Two-factor Factorial in a Randomized Complete Block Design with One Covariable

$$Y_{ijk}=\boxed{\mu+\beta \overset{\text{covariance}}{X_{ijk}}}+\boxed{\alpha_i+\gamma_j+\overset{\text{interaction}}{\alpha\gamma_{ij}}+\overset{\text{block}}{\rho_k}+\ \varepsilon_{ijk}}$$

$$i=1,2,\ldots,a;\ j=1,2,\ldots,b;\ k=1,2,\ldots,r;$$

ANOVA Table for Model with One Covariate (Type III)

Source	df	SS(adj)
R(β|μ,α,γ,αγ,ρ)	1	SSX
R(α|μ,γ,αγ,ρ,β)	$a-1$	SSA
R(γ|μ,α,αγ,ρ,β)	$b-1$	SSB
R(αγ|μ,α,γ,ρ,β)	$(a-1)(b-1)$	SSAB
R(ρ|μ,α,γ,αγ,β)	$r-1$	SSR
Error	$(ab-1)(r-1)-1$	SSE

15

Case Study (3)

- A horticulturist conducted an experiment to study the effects of flower variety (factor A: LP, WB) →不同品种鲜花. and moisture level (factor B: low, high) →湿度. on yield of salable flowers (Y).
- Because the plots were not of the same size, the horticulturist wished to use plot size (X) as the concomitant variable. 排除种植面积.
- Six replications were made for each treatment.

16

Factor A (flower variety) *i*	Factor B (moisture level) *j*: B_1 (low) Y	B_1 (low) X	B_2 (high) Y	B_2 (high) X
A_1 (variety LP)	98	15	71	10
	80	4	80	12
	77	7	86	14
	80	9	82	13
	95	14	46	2
	64	5	55	3
A_2 (variety WB)	55	4	76	11
	60	5	68	10
	75	8	43	2
	65	7	47	3
	87	13	62	7
	78	11	70	9

17

SAS Program

```
Data Flower;
Input A $ B $ Y X @@;
Datalines;
LP Low   98  15   LP  High  71  10
LP Low   60  4    LP  High  80  12
LP Low   77  7    LP  High  86  14
LP Low   80  9    LP  High  82  13
LP Low   95  14   LP  High  46  2
LP Low   64  5    LP  High  55  3
WB Low   55  4    WB  High  76  11
WB Low   60  5    WB  High  68  10
WB Low   75  8    WB  High  43  2
WB Low   65  7    WB  High  47  3
WB Low   87  13   WB  High  62  7
WB Low   78  11   WB  High  70  9
;
PROC GLM;
Class A B;
MODEL Y = X A|B;
LsMeans A B/STDERR PDIFF Adjust=Tukey ETYPE=3;
LSMeans A*B/STDERR PDIFF Adjust=Tukey ETYPE=3;
RUN;
```

The ETYPE= option specifies the type (1, 2, 3, or 4, corresponding to Type I, II, III, and IV tests, respectively) of the E= effect. If you specify the E= option but not the ETYPE= option, the highest type computed in the analysis is used. If you omit the E= option, the ETYPE= option has no effect.

18

Table of ANOVA

ANOVA Table for Model without Covariate (Type III)

Source	df	SS	MS	F value	P-value
A|μ, B,AB	1	486.0000000	486.0000000	2.36	0.1399
B|μ, A,AB	1	486.0000000	486.0000000	2.36	0.1399
AB|μ, A,B	1	0.0000000	0.0000000	0.00	1.0000
Error	20	4114.000000	205.700000		

都不显著.

ANOVA Table for Model with One Covariate (Type III)

Source	df	$SS_{(Adj)}$	$MS_{(Adj)}$	F value	P-value
X|μ,A,B,AB	1	3994.518817	3994.518817	635.21	<.0001
A|μ, B,AB, X	1	96.601826	96.601826	15.36	0.0009
B|μ, A,AB,X	1	323.849473	323.849473	51.50	<.0001
AB|μ, A,B,X	1	16.042244	16.042244	2.55	0.1267
Error	19	119.481183	6.288483		

显著.

显著.

使残差大大减小

19

Standard			H0:LSMEAN=0	H0: LSMean1=LSMean2
A	Y LSMEAN	Error	Pr > |t|	Pr > |t|
LP	74.5000000	4.1402496	<.0001	0.1399
WB	65.5000000	4.1402496	<.0001	

Standard			H0:LSMEAN=0	H0:LSMEAN1=LSMean2
B	Y LSMEAN	Error	Pr > |t|	Pr > |t|
High	65.5000000	4.1402496	<.0001	0.1399
Low	74.5000000	4.1402496	<.0001	

Adjusted by One Covariable

Standard			H0:LSMEAN=0	H0:LSMEAN1=LSMean2
A	Y LSMEAN	Error	Pr > |t|	Pr > |t|
LP	72.0423387	0.7304444	<.0001	0.0009
WB	67.9576613	0.7304444	<.0001	

Standard			H0:LSMEAN=0	H0: LSMean1= LSMean2
B	Y LSMEAN	Error	Pr > |t|	Pr > |t|
High	66.3192204	0.7246356	<.0001	<.0001
Low	73.6807796	0.7246356	<.0001	

20

- ☐ The covariance analysis with two covariates

$$Y_{ijk} = \mu + \beta_1 X_{ijk(1)} + \beta_2 X_{ijk(2)} + \alpha_i + \gamma_j + \alpha\gamma_{ij} + \rho_k + \varepsilon_{ijk}$$

$$i = 1,2,\ldots,a;\ j = 1,2,\ldots,b;\ k = 1,2,\ldots,r;$$

ANOVA Table for Model with One Covariate (Type III)

Source	df	SS(adj)
$R(\beta_1\|\mu,\alpha,\gamma,\alpha\gamma,\rho,\beta_2)$	1	$SSX_{(1)}$
$R(\beta_2\|\mu,\alpha,\gamma,\alpha\gamma,\rho,\beta_1)$	1	$SSX_{(2)}$
$R(\alpha\|\mu,\gamma,\alpha\gamma,\rho,\beta_1,\beta_2)$	$a-1$	SSA
$R(\gamma\|\mu,\alpha,\alpha\gamma,\rho,\beta_1,\beta_2)$	$b-1$	SSB
$R(\alpha\gamma\|\mu,\alpha,\gamma,\rho,\beta_1,\beta_2)$	$(a-1)(b-1)$	SSAB
$R(\rho\|\mu,\alpha,\gamma,\alpha\gamma,\beta_1,\beta_2)$	$r-1$	SSR
Error	$(ab-1)(r-1)-2$	SSE

21