

한국 道別 인구수 변천에 대한 共分散分析

연세의대의대 예방의학교실

申 敏 雄

—Abstract—

The Analysis of Covariance of Do(province) Population Variability

Min Wong Shin

Dept. Preventive Medicine and Public Health Yonsei Medical College

The Mechanism for sorting out the covariance effect is known as the covariance analysis. The sorting out of regression and correlation effect is an obvious application of the covariance analysis.

The result of Do population by age groups (15 ages interval) from 1966 Census and from 1970 Census has been applied to analyzing covariability by the analysis of covariance.

The results are as follows.

- (1) The signigance of the regression of 1970 population on 1966 population is assured as $F=116.5$
- (2) There is a significant difference between mean of each age group. ($F=88.1$)
- (3) There is very little evidence of significant heterogeneity of regression between age groups. ($F=0.72$)

1. 서 론

共分散의 효과를 분류하는 메카니즘이 共分散分析 (Analysis of Covariance)으로 알려져있다. 회귀와 상관(Regression and Correlation)의 효과를 분류하는 것도 共分散分析의 응용의 하나이다.

이러한 공분산분석을 하여 1966년 센서스 및 1970년 센서스에 나타난 각 도(道)에 인구수에 대한 변이를 분석하였다.

2. 연령간의 유의성과 回歸의 異質性檢定

각 道의 연령별 인구는 1966년 및 1970년 센서스에 의해 다음과 같다. (표 1) 여기에서 통계분석을 위하여, 道를 Replicate로 연령을 Treatment로 놓았다.

공분산분석을 위한 계산과정은 다음과 같다.

$$\text{전체 } \sum_1^{nr} (xy) = \sum_1^{nr} (XY) - \frac{G_x G_y}{nr}$$

$$\text{연령 I } 14,951,012$$

$$\text{연령 II } 4,194,425$$

$$\text{연령 III } 2,111,881$$

$$\text{연령 IV } 1,998,516$$

$$23,245,834$$

$$23,245,834 - \frac{23,603 \times 23,668}{32} = 5,779,614$$

$\sum xy = (X - \bar{X})(Y - \bar{Y})$ 를 계산하는 과정은

$$\text{Replicates } \frac{(3,102 \times 3,353) + \dots + (4049 \times 4005)}{4}$$

$$-17,466,220 = 1,792,341$$

$$\text{Treatments } \frac{(10,522 \times (10,378) + \dots + (3,676 \times 3,938))}{8}$$

$$-17,466,220 = 3,636,625$$

$$\text{Error } 5,779,614 - 1,792,341 - 3,636,625 = 350,648$$

$\sum x^2$ 를 계산하는 과정은

$$\text{Treatments } x^2 = \frac{10,522^2 + \dots + 3,676^2}{8} - \frac{23,603^2}{32}$$

$$= 3,843,679$$

$$\text{Replicates } x^2 = \frac{3102^2 + \dots + 4,049^2}{4} - 17,971,019$$

$$= 1,188,974$$

표 1.

각 도의 연령별 인구수, 1966=X 및 1970=Y

(단위 1,000)

도	연령 연도	연령				전 체
		I 0-14	II 15-29	III 30-44	IV 45-	
경기	X	1,355	749	530	469	3,102
	Y	1,401	816	609	527	3,353
강원	X	838	414	309	270	1,831
	Y	841	407	331	287	1,865
충북	X	714	339	243	252	1,549
	Y	672	300	245	263	1,480
충남	X	1,308	675	466	453	2,903
	Y	1,273	616	482	487	2,858
경북	X	1,951	1,099	717	706	4,473
	Y	1,932	1,088	771	765	4,556
경남	X	1,397	748	519	512	3,175
	Y	1,341	709	528	541	3,119
전북	X	1,139	587	401	394	2,521
	Y	1,102	525	394	411	2,432
전남	X	1,820	976	636	620	4,049
	Y	1,816	874	657	658	4,005

※ 한국 1966 및 1970 센서스 자료임.

표 2.

공분산 분석, 표 1의 자료

	DF	x ²	xy	y ²	byx	byxΣxy	Σy ² _d	DF
Replicates	7	1,188,974	1,792,341	2,064,818				
Treatments	3	3,843,679	3,636,625	3,470,946	0.9461	3,440,611	30,335	2
Error	21	959,383	350,643	150,167	0.3655	128,161	22,006	20
Treat+Error	24	4,803,062	3,986,273	3,621,113	0.8299	3,308,208	312,905	23

Error x²=Total-Replicates-Treatments=959,383
y²를 구하는 과정도 마찬가지이다.

표 2에서 회귀의 유의성을 검정하면, F=117로서 X

표 3.

	SS	DF	MS	F
Regression	128,161	1	128,161	116.5
Error	22,006	20	1,000	

표 4. 연령별 차이의 유의성 검정

	DF	SS	MS	F
Treat+Error	23	312,905		
Error	20	2,006	1100.3	88.1
Difference	3	290,899	96,966	
Treatment(Adzusted)	2	30,335		
b _c -b _t	1	260,564	260,564	

에 대한 Y의 회귀는 유의성이 있다.

다음으로 연령별 평균들 사이에 차이의 유의성을 검

정하면 F=88로서 유의성이 있음을 알 수 있다.(표 4.)

조정(adjusted)된 각 연령별 평균들의 쌍사이에 유의성을 검정하기 위한, 두 평균들 사이에 차이의 분산을 공식

$$\frac{2S^2}{r} \left[1 + \frac{A_1}{A_2(t-1)} \right] \text{에서}$$

(r은 도의 수)

$$\frac{2 \times 1100}{8} \left(1 + \frac{3,843,679}{959,383 \times 3} \right) = 642$$

평균 차의 표준오차=25.34

전후의 조정의 결과에 대한 상대 정밀성은

$$\frac{150,167}{21} \times \frac{22,006}{20} = 0.65$$

회귀의 異質性에 대한 검정은 다음과 같다.

F=0.72로서 유의한 異質性(significant heterogeneity)은 없었다.

이상으로 각 도에 대한 인구구조에 대하여 다음과 같

표 5.

연령별사이에 회귀의 이질성 검정

연 령	DF	x^2	xy	y^2	byx	$byx \sum xy$	$\sum y^2_d$	DF
0—14	7	1,287,601	1,301,349	1,322,500	1.0107	1,315,273	7,227	6
15—29	7	465,572	457,896	467,775	0.9835	450,341	17,434	6
30—44	7	178,239	194,743	216,869	1.0926	212,776	4,093	6
45—	7	172,668	189,005	207,379	1.0946	206,885	494	6
전 체	28	2,104,080	2,142,993	2,214,523	1.0185	2,182,638	31,885	27

	DF	SS	MS	F
Total	27	31,885		
연 령	24	29,248	1,219	
차 이	3	2,637	879	0.72

은 결론을 얻었다.

① 1966년 인구에 대한 1970년의 인구의 회귀(regression)는 유의성이 있다.

② 연령별 회귀의 이질성은 유의한 차가 없다.

③ 30세 이하 인가와 30세 이상의 인구수의 차이는 적어지는 방향으로 나타나고 있다.

