**중간시험 대체 과제물**

**보건통계학과**

**2022540028**

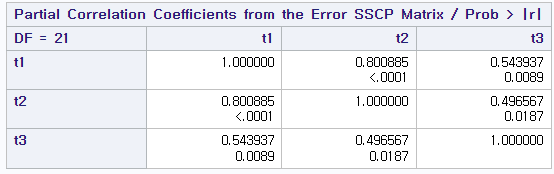
**허은정**

1. 반응변수를 세 번 반복측정한 우울증 점수를 (t1, t2, t3)로 두고 기저시점에서 측정된 우울증 점수 t0와 처리그룹(Group)을 독립변수로 다루어 이들 변수들에 의해 3번 측정된 우울증 점수가 어떻게 영향을 받는지 반복측정분산분석(RMANOVA)을 하고 다음 질문사항에 대해 답하시오.

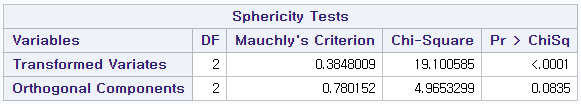
\* model : t1 t2 t3 = t0 Group으로 설정하는 경우

* 1. 구형성 가정을 만족하는지 확인하고 그 결과를 제시하시오.

시점(t1, t2, t3) 간 우울증 점수(score)의 분산-공분산 행렬을 상관계수로 표현한 결과는 아래와 같다.



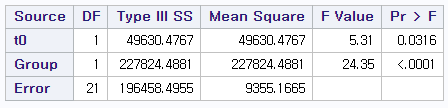
분산-공분산 행렬이 구형성 구조를 따르는지 검정한 결과는 아래와 같다.  
- 귀무가설 : 분산-공분산 행렬이 구형성 구조를 따름  
- 대립가설 : 분산-공분산 행렬이 구형성 구조를 따르지 않음



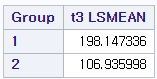
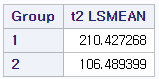
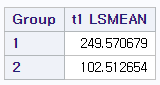
Orthogonal Component의 p-value(Pr > ChiSq)가 0.0835 > 0.05로 5% 유의수준에서 귀무가설을 기각할 수 없다.  
따라서 본 자료의 분산-공분산 행렬은 구형성 가정을 만족한다.

* 1. 처리그룹 간 차이는 통계적으로 유의한가? 만약 유의하다면 대조군에 비해서 실험군은 얼마나 영향을 미치고 있는지 설명하시오. (lsmeans 이용)

처리그룹 간 효과 검정 결과는 아래와 같다.  
- 귀무가설 : Group간 평균의 차이가 없음  
- 대립가설 : Group간 평균의 차이가 있음

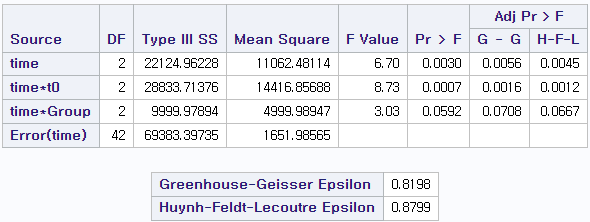


Group의 p-value(Pr > F)가 <.0001 < 0.05로 5% 유의수준에서 귀무가설을 기각하고 대립가설을 채택할 수 있다.  
따라서 본 자료를 통해 모집단의 Group간 우울증 점수의 평균 차이가 있다고 추론할 수 있다.

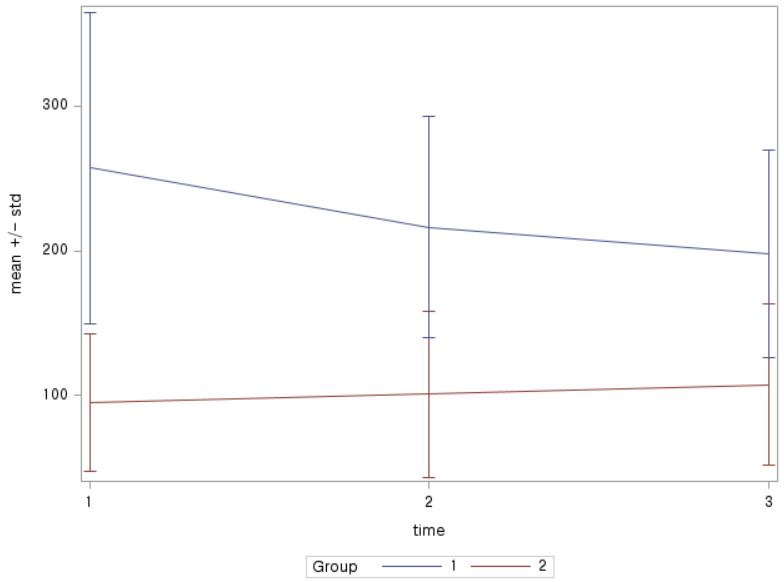


대조군(Group 1)에 비해 실험군(Group 2)은 더 작은 LSMEAN값을 가지며,  
t1에서 실험군과 대조군의 LSMEAN 격차가 가장 크고 t2, t3으로 갈수록 그 격차가 줄어드는 경향을 보인다.

* 1. 구형성 가정을 만족할 경우 보정된 일변량 분산분석 결과에 의해 각 요인의 효과에 대한 p-value를 제시하고 그 결과를 해석하시오.



Epsilon > 0.75이기 때문에 p-value로 H-F-L Adj Pr > F를 참조하면,  
time의 효과는 p-value가 0.0045 < 0.05로 5% 유의수준에서 유의하다.  
time과 t0의 상호작용 효과는 p-value가 0.0012 < 0.05로 5% 유의수준에서 유의하다.  
time과 Group의 상호작용 효과는 p-valuer가 0.0667 > 0.05로 5% 유의수준에서 유의하지 않다.  
따라서 본 자료를 통해 모집단에서 Group 내 time(시점)의 단독 효과와 time(시점)과 t0(기저시점의 우울증 점수)의 상호작용 효과는 우울증 점수에 유의미한 영향을 미치지 못한다고 추론해볼 수 있다.

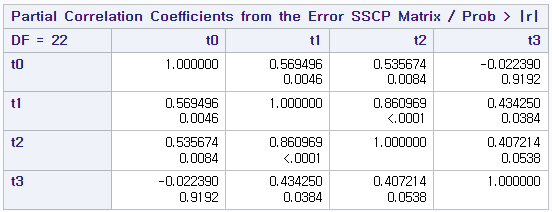
* 1. Group\*Time 간 상호작용이 통계적으로 유의하다면 표본평균을 이용하여 평균반응 프로파일 플롯을 작성하고 그 결과를 해석하시오. 

Group과 Time 간 상호작용은 5% 유의수준에서는 통계적으로 유의하지 않다.  
평균반응 프로파일 플롯을 그려본 결과 Group 1은 t1->t2->t3으로 갈수록 우울증 점수가 감소하나, Group 2는 증가하는 경향을 보인다.  
하지만 이러한 차이가 통계적으로 유의미한 수준은 아닌 것으로 판단된다.

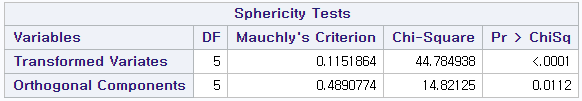
\* model : t0 t1 t2 t3 = Group으로 설정하는 경우

* 1. 구형성 가정을 만족하는지 확인하고 그 결과를 제시하시오.

시점(t0, t1, t2, t3) 간 우울증 점수(score)의 분산-공분산 행렬을 상관계수로 표현한 결과는 아래와 같다.



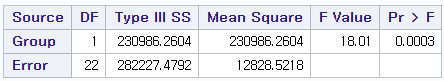
분산-공분산 행렬이 구형성 구조를 따르는지 검정한 결과는 아래와 같다.  
- 귀무가설 : 분산-공분산 행렬이 구형성 구조를 따름  
- 대립가설 : 분산-공분산 행렬이 구형성 구조를 따르지 않음



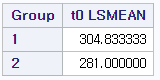
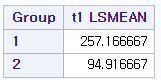
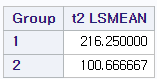
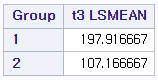
Orthogonal Component의 p-value(Pr > ChiSq)가 0.0112 < 0.05로 5% 유의수준에서 귀무가설을 기각할 수 있다.  
따라서 본 자료의 분산-공분산 행렬은 구형성 가정을 만족하지 않는다.

* 1. 처리그룹 간 차이는 통계적으로 유의한가? 만약 유의하다면 대조군에 비해서 실험군은 얼마나 영향을 미치고 있는지 설명하시오. (lsmeans 이용)

처리그룹 간 효과 검정 결과는 아래와 같다.  
- 귀무가설 : Group간 평균의 차이가 없음  
- 대립가설 : Group간 평균의 차이가 있음

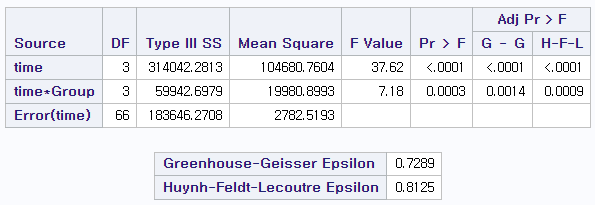


Group의 p-value(Pr > F)가 0.0003 < 0.05로 5% 유의수준에서 귀무가설을 기각하고 대립가설을 채택할 수 있다.  
따라서 본 자료를 통해 모집단의 Group간 우울증 점수의 평균 차이가 있다고 추론할 수 있다.

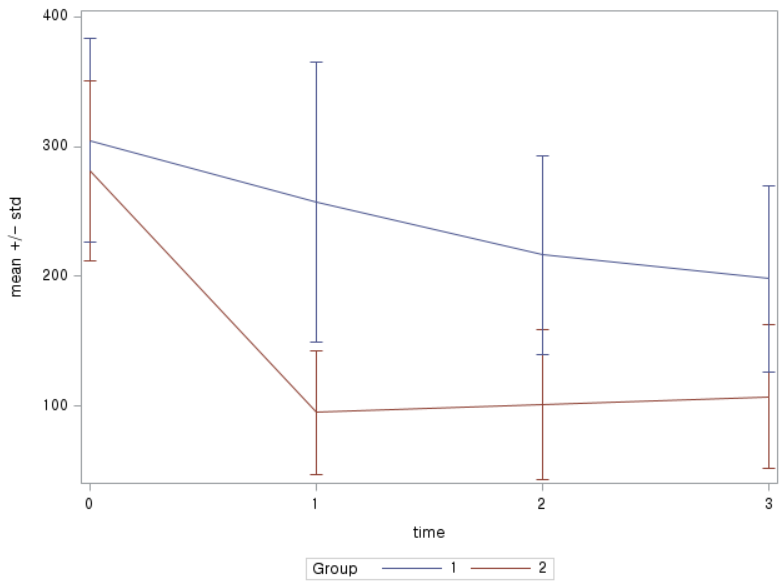
대조군(Group 1)에 비해 실험군(Group 2)은 모든 시점에서 더 작은 LSMEAN값을 가진다.  
그런데 시점 t1에서 실험군과 대조군의 LSMEAN격차가 커지고 t2, t3으로 갈수록 그 격차가 줄어드는 경향을 보인다.

* 1. 구형성 가정을 만족할 경우 보정된 일변량 분산분석 결과에 의해 각 요인의 효과에 대한 p-value를 제시하고 그 결과를 해석하시오.



Huynh-Feldt-Lecoutre Epsilon > 0.75이기 때문에 p-value로 H-F-L Adj Pr > F를 참조하면,  
time의 효과는 p-value가 <.0001 < 0.05로 5% 유의수준에서 유의하다.  
time과 Group의 상호작용 효과는 p-valuer가 0.0009 < 0.05로 5% 유의수준에서 유의하다.  
따라서 본 자료를 통해 모집단에서 Group 내 time(시점)의 단독 효과와 time과 Group의 상호작용 효과는 우울증 점수에 영향을 미친다고 추론해볼 수 있다.

* 1. Group\*Time 간 상호작용이 통계적으로 유의하다면 표본평균을 이용하여 평균반응 프로파일 플롯을 작성하고 그 결과를 해석하시오.



기저시점의 우울증점수는 Group 1과 Group 2에 큰 차이가 없다.   
그런데 t1시점이 되면 Group 1과 Group2의 우울증 점수에 격차가 생기고, Group 1이 Group 2에 비해 유의미하게 높은 우울증 점수 분포를 갖는다.  
Group 1은 t1->t2->t3으로 갈수록 우울증 점수가 감소하는 경향을 보이는 반면, Group 2는 t1->t2->t3으로 갈수록 우울증 점수가 같거나 상승하는 경향을 보인다.

1. 동일자료에 대하여 다양한 공분산 패턴 모형 하에 선형혼합모형에 적합하여 다음 질문에 답하시오.
   1. 처리그룹과 3번의 시점을 이용하여 6개의 가변수를 생성한 후 우울증 점수가 이들 가변수와 기저시점의 우울증 점수에 의해 어떻게 영향을 받는지 PROC MIXED 프로시저를 이용해서 선형혼합모형에 다음과 같이 적합한 후 평균 프로파일(mean profile) 구조를 이용한 모형에서 추정된 평균을 제시하시오.

(ML 추정법 이용)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | 공분산 모형 | | | | | |
| UN | TOEP(1) | AR(1) | CS | UN(1) | simple |
| 대조군 | C11 | 257.17 | 257.17 | 257.17 | 257.17 | 257.17 | 257.17 |
| C12 | 216.25 | 216.25 | 216.25 | 216.25 | 216.25 | 216.25 |
| C13 | 197.92 | 197.92 | 197.92 | 197.92 | 197.92 | 197.92 |
| 실험군 | T21 | 94.92 | 94.92 | 94.92 | 94.92 | 94.92 | 94.92 |
| T22 | 100.67 | 100.67 | 100.67 | 100.67 | 100.67 | 100.67 |
| T23 | 107.17 | 107.17 | 107.17 | 107.17 | 107.17 | 107.17 |
|  | t0 | 0 | 0 | 0 | 0 | 0 | 0 |

(REML 추정법 이용)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | 공분산 모형 | | | | | |
| UN | TOEP(1) | AR(1) | CS | UN(1) | simple |
| 대조군 | C11 | 231.39 | 154.99 | 178.86 | 165.91 | 150.90 | 154.99 |
| C12 | 190.47 | 114.07 | 137.94 | 124.99 | 109.99 | 114.07 |
| C13 | 172.14 | 95.74 | 119.61 | 106.66 | 91.65 | 95.74 |
| 실험군 | T21 | 71.15 | 0.73 | 22.73 | 10.80 | -3.04 | 0.73 |
| T22 | 76.90 | 6.48 | 28.48 | 16.55 | 2.71 | 6.48 |
| T23 | 83.40 | 12.98 | 34.98 | 23.05 | 9.21 | 12.98 |
|  | t0 | 0.08 | 0.34 | 0.26 | 0.30 | 0.35 | 0.34 |

ML추정법에 비해 공분산 모형 간 추정값들의 편차가 크다.

* 1. 평균 프로파일(mean profile) 구조를 이용한 모형에서, ML 추정법에 의한 결과에 기초하여 비구조화 공분산 모형과 비교하여 데이터를 잘 적합시키는 공분산 모형은 무엇인지 가능도비 검정(likelihood ratio test)을 실시하고 그 결과를 제시하시오.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 공분산 모형 | | | | | |
| UN | TOEP(1) | AR(1) | CS | UN(1) | simple |
| -2 Log Likelihood | 774.9 | 814.4 | 787.4 | 791.9 | 812.5 | 814.4 |
| # of cov params | 7 | 2 | 3 | 3 | 7 | 2 |
| UN과 비교 | | | | | | |
| -2LL diff |  | 39.5 | 12.5 | 17.0 | 37.9 | 39.5 |
| cov params diff |  | 5 | 4 | 4 | 0 | 5 |
| p-value in chisq dist. |  | <.005 | <.025 | <.005 |  | <.005 |

공분산 모형들이 비구조화 공분산 모형과 유의미한 차이가 없을 때 해당 공분산 모형을 택하는 것이 좋다.  
그런데 본 자료를 적합한 결과 공분산 모형들이 비구조화 공분산 모형과 유의미한 차이가 있기 때문에, 비구조화 공분산 모형을 사용하는 것이 적절하다.

* 1. 평균 프로파일(mean profile) 구조를 이용한 모형에서 ML과 REML 추정법에 기초하여 동일한 평균모형에서 AIC와 BIC를 이용하여 최적 모형을 선택한다면 어떤 공분산 모형을 선택할지 그 결과를 제시하고 설명하시오.

(ML 추정법 이용)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 공분산 모형 | | | | | |
| UN | TOEP(1) | AR(1) | CS | UN(1) | simple |
| AIC | 798.9 | 828.4 | 803.4 | 807.9 | 830.5 | 828.4 |
| BIC | 774.9 | 814.4 | 787.4 | 791.9 | 812.5 | 814.4 |

AIC와 BIC 모두 값이 작을수록 최적의 모형에 해당한다. 따라서 AIC와 BIC의 값이 모두 가장 작은 UN(비구조화) 공분산 모형을 선택하는 것이 적절하다.

(REML 추정법 이용)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 공분산 모형 | | | | | |
| UN | TOEP(1) | AR(1) | CS | UN(1) | simple |
| AIC | 744.9 | 764.3 | 746.9 | 750.3 | 767.2 | 764.3 |
| BIC | 730.9 | 760.3 | 740.9 | 744.3 | 759.2 | 760.3 |

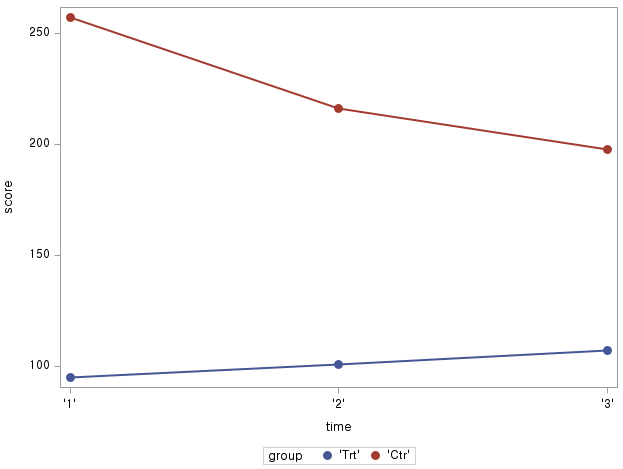
AIC와 BIC 모두 값이 작을수록 최적의 모형에 해당한다. 따라서 AIC와 BIC의 값이 모두 가장 작은 UN(비구조화) 공분산 모형을 선택하는 것이 적절하다.

* 1. Group, Time, Group\*Time, t0를 독립변수로 두고 ML 방법에 의해 각 효과를 추정하고 그 결과를 비교 설명하시오.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | 공분산 모형 | | | | | |
| UN | TOEP(1) | AR(1) | CS | UN | simple |
| 절편 |  |  | 197.92 | 197.92 | 197.92 | 197.92 | 197.92 | 197.92 |
| Group | Trt |  | -90.75 | -90.75 | -90.75 | -90.75 | -90.75 | -90.75 |
| Ctr |  | 0 | 0 | 0 | 0 | 0 | 0 |
| Time |  | 1 | 59.25 | 59.25 | 59.25 | 59.25 | 59.25 | 59.25 |
|  | 2 | 18.33 | 18.33 | 18.33 | 18.33 | 18.33 | 18.33 |
|  | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| Group  \*  Time | Trt | 1 | -71.50 | -71.50 | -71.50 | -71.50 | -71.50 | -71.50 |
| Ctr | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Trt | 2 | -24.83 | -24.83 | -24.83 | -24.83 | -24.83 | -24.83 |
| Ctr | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| Trt | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ctr | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| t0 |  |  | 0 | 0 | 0 | 0 | 0 | 0 |

각 효과의 추정값은 공분산 모형과 무관하게 유사하다. 따라서 모형에 대한 가정을 최소화한 UN(비구조화) 모형을 선택하는 것이 적절할 것이다.  
SAS 분석 결과 산출되는 모형별 Standard Error, -2LL, AIC, BIC 값들을 통해 UN모형을 선택하는 것이 적절하다는 것을 뒷받침할 수 있다.

* 1. 4)의 결과를 기반으로 상호작용 효과에 대하여 추정된 평균값을 이용하여 플롯을 제시하고 그 결과를 해석하시오.



상호작용 효과까지 고려하여 모델을 적합하고   
Group별, Time별 추정된 평균값을 그래프로 나타내어 보면,   
Group별로 Time에 대한 효과가 반대로 나타나는 것을 볼 수 있다.   
문제 1-4)에서 살펴보았던 Group별, Time별 표본평균 그래프와   
그 형태가 유사한 것으로 보아, 모델이 잘 적합된 것으로 보인다.