공분산분석 들어가기 전 예제\*

일원배치완전랜덤화설계

기계(처리제곱합) : 무생물이라서 이상하다.

원사의 굵기(실의 굵기)도 데이터가 있다.

기계1,2,3 간의 성능 (+원사의굵기라는 설명변수를 추가함)

회귀분석 + 실험설계모형 혼합.

* 이것을 공분산분석 모형이다.

실험하는 사람이 통제할 수 있으면 (원사의 굵기를 통제할 수 있다고 하면 같은 굵기를 가지고 세 수를 삼등분해서 똑 같은 굵기로 건다음에 직물 짜서 비교해도 됨=> 랜덤화 블록설계 : 블록의 개념을 이용해서 오차를 줄이는 행동임. 3개원리 체크하기.)

하지만, 통제할 수 없는 상황이니까.(굵기를 관측할 수 밖에 없지, 블록을 잡을 수 없음) 그래서 공분산분석을 함.

일원배치법에 공변수 하나를 추가한게 다임.

완전모형: 회귀분석에서 부분집합에 대한 걸 검정할 때, 그 개념을 이용한 것임

* 1. 처리 효과가 있냐 없냐(이게 최종목적)
* 2. 공변수의 효과가 있냐 없냐(처리효과를 보고 그 남은 찌꺼기로 해석해보자)

축소모형은 2번째 단순선형회귀모형이야.

우린 일단 처리효과가 없다고 가정한게 단순선형회귀모형 (축소모형)

처리효과가 있다고 보는게 완전모형

잔차제곱합 계산해서 두개의 차이가 두 모형의 설명력 -> 추가제곱합 사용해서 검정하는 것임(회귀분석에선 추가 회귀 제곱합 같은 것임. 물론 추가회귀제곱합은 아님^^)

예

Proc sort

그림을 그리는데 기계3대 ->각 기계별로 5번 반복, 실의 원사의 굵기에 섬유의 강도 plot하면 기계마다 점 찍히니까 => 한꺼번에 15개씩 그리는데 5개 구분시킬려고 쓰는 것.

(공변수를 추가해서 효과를 보정해주는거, 이게 공분산분석모형)

By machine; 기계별로. Y랑x의 산점도를 그려라 뭐 그런거임.

3대의 기계 데이터

y\*x=기계 (색깔별루 구분함)

처리1

처리2

처리3

다 단순선형모형인데

타오가1일땐느 절편이 mu` + tao1

기울기는 변하는게 아니야, 공변수랑 y랑x 사이의 기울기는 똑같은데, 절편이 자꾸 변하는 모습임.

1. 축소모형은 직선 하나만 적합시키면된다. 차이가 상관없다. ( 단순선형회귀모형)
2. 그런데, 완전모형 차이가 있따면 회귀모형에 기계마다 y절편이 다 다름.

* 예제에서는 축소모형을 쓰는게 더 맞을 것이다. (scatter plot으로 확인)
* 그리고 처리가 다른지 아닌지를 한번 체크해보자.(가설검정을 해 봐야 함)

1. 공분산분석에서 기울기의 동질성으로 봐도 될까?
2. 공변수 자체가 정말 필요할까??? (회귀의 유의성 검정)

* 이게 만족이 되었다면 공분산분석을 해도 괜찮다.

1. 본론: 공분산분석에서 처리에 대한 검정.

완전모형 안에서 모형제곱합 잔차제곱합

축소모형 안에서의 모형제곱합 잔차제곱합

두 모형의 자유도와 나눠서 완전모형의 MSE랑 축소모형의 MSE 차이를 보고

처리효과가있는지 없는지를 체크해야함.

\*\* 완전모형하에서 최소제곱추정.

일원배치모형을 기억해내자.

공변수에 조정된 것 만큼 처리제곱합에 보정을 해야함.

그래서 그 부분에 대해 taoi 검정을 해야함.(마치 일원배치마냥 함)

* 이제 완전모형에서랑 축소모형에서 조정된 함수가 달라지니까 beta.hat full
* 그래서 그 공변수 앞에 계수에 맞춰서 너무 과대평가 과소평가 된 부분을 taoi에 적용시켜서 검정을 하면 된다.

축소모형과 완전모형의 절편이 당연히다르잖아.

축소모형의 경우에는 일원배치 사라지는거니까 xij-x.bar yij-y.bar 이렇게 계산하는건데

완전모형의 경우 xij-xibar yij-yi.bar 이걸 계산해야함. 그 집단 안의 편차를 구해야함.

* 처리집단에 맞춰서 합쳐야함. (집단 3개 안에서 편차를 가지고 계산해야한다.)

(축소모형 하에서의 잔차제곱합)

SSE = Syy – Sxy^2/Sxx

SSR = Sxy^2/Sxx

* 노트에 맞게 SSE, SSR 로 변화함.(완전모형하에서 잔차제곱합)

두개의 차를 보고 검정하는 것임

모든 taoi 검정하기전에 공변수 필요한지 봐야하거든

One-way ANOVA SSE = Eyy

여기서 두개로 갈라짐 (회귀제곱합과 / 공변수에서의 잔차제곱합으로 쪼개짐)

Mse는 완전모형에서나누거든

(자유도 = 처리개수t개\*반복갯수n – 모수만큼 자유도가 줄어듬(2개) – (절편,처리효과 t개인데, 처리효과들의 합이 0이니까 t-1개를 뺌)) = t(n-1)-1 만큼 나눠줌

\*\* 가설검증

* 추가제곱합을 처리제곱합으로 함(추가모형제곱합)
* SS(Trt) / 두 자유도의 차, t개의 모수가 차이나지만 - 1 (왜냐면 타오의 합이 0이라는 제약식이 있어서
* 여기에 MSE를 나눈게 F통계량.
* 두번째화살표랑 세번째 화살표를 저울질 해보는 것임. (두개가 독립적이다)

SST = SSR + SStrt + SSE 가 일반적으로는 절대 성립안함.

표10.4 참고해서 보면 알 수 있음.

예제 다시 설명을 들어가보면\*

하나의 직선으로 설명할수 있나,(축소모형)

* 이거는 좀 간단해

세개의 직선을 따로해야하나(물론 기울기는 같다는 가정임. 여기선 그러함) =>완전모형

* 여기선 각 처리집단별로 평균을 구해서 빼야하거든.
* 공통의 기울기를 가정했으니까 다 풀링해서 공통의 기울기를 구함(beta hat full)

공통의 기울기 체크 => 제일 먼저 해야할 것**. (상호작용이 있는지를 체크해야함).**

Note1)

rB(x-x.bar)가 없는 모형인데, 똑같지 않으면 위에 rB(x-xbar)가 들어가야함.(상호작용이랑 같은 거임)

* 처리효과랑 기울기의 상호작용을 넣어줌.
* Tao=1일 때 tao=2일 때, tao=3일 때….
* 다시 표시해보면

Y(ij) = mu + tao(i) + [Beta+ taobeta(i)](xij – x.bar) + eps.(ij)

* 선형식 적합 기울기, 일반적으로 안같음. Taobeta(i)들이 다 같으면서 0일 때
* 왜냐면 세게 합이 0이라고 가정 주거든….ㅎㅎ
* 공통의 기울기를 가정한 공분산분석모형이 됨.

보통 귀무가설이 기각되길 바라지만, 이 경우에는 모든 taoBeta(i)=0이라고 기각이 되야 공분산분석이 넘어감. 그러니까 **귀무가설이 채택이 계속 되길바람.**

(공분산분석모형을 옹호하는 입장)

Note2)

그 전에 검정해야할 것. 과연 공변수가 들어가도 괜찮은건지 봐야함!!!!!

(x무시해도 되는지 안되는지를 체크)

보정된처리평균cjfl

Y(ij) = mu + tao(i) + eps(ij) 처리평균 처리

여기서 공변수 부분에 해당하는 크기만큼 보정해줘야햠.

Ex.직물강도 실험에서, 공변수 영향없다면

공변수 영향이있고, 들어오면 i번째 처리평균은 y(i).bar에서 원사의 굵기만큼 조정을 해줘야 함.

* 그 평균이 진짜 기계의 따른 성능차이를 반영한 평균이다.

Type1 SS, Type3 SS

Oneway랑 가정했을 때 ANOVA의 처리제곱합 Type1 SS = 순차제곱합. Sequential sum of squares

처리교화가 있는데 공분산이 추가될 때 어느정도 증분인가????(x에 있는 type1 SS)

두개 교호작용은 마지막으로 들어가는거(얘가 순수한 기여도 그 외엔 아님)

Type3 SS = 편제곱합 partial sum of squares. (순수한제곱합)

두개의 항이 들어있는 상태에서 machine이 들어갈 때 SS가 얼마나 늘어나는가

순수한 기여도를 따지는 것.. 얼마나 모형제곱합이 늘어나냐

여기에는 순서는 machine… 편의 개념임. 자기만의 순수한 그런 개념에 해당 2.664 이게 그거임.

기계효과, 상호작용이 있는 상태에서 x의 모형의 추가제곱합.(모형이 얼마나 늘어나는지)

X\*machine도 마찬가지임.

일반적으로 Type1, Typ3는 항상 같을 수 없지만

마지막 output SS 는 같게 나와야함. 그래서 상호작용term은 같은 SS값이 나옴.

직교한 경우에(독립은 아님) type1, type3가 같게 나옴.(지난학기 기초실험계획에서 한 내용)

* 직교분할만 같게 나옴.(서로간에 영향을 안미치는 경우, 모형제곱합이 누가있든 말든 항상 같은거 이게 직교분할, 이 케이스만 같다는 의미임)
* 이때는 proc anova 쓰면됨, glm 안해도됨.

(균형불완전블락설계가 직교분할이 안됨)

* 블락에서 빠지는 처리, 거기서 직교분할.\*\*\*\*( 지난학기 노트 참고 )

Glm: 일반선형모형, 회귀분석.(선형회귀모형), 비선형회귀모형은 적합이 막막함.

분산분석, 공분산분석 다해당이 됨.

# 직교대비\*\*[Orthogonal contrasts](https://stats.stackexchange.com/questions/47148/orthogonal-contrasts-in-r)

여기서도 proc anova를쓰면 암됨.

여기서는 typ1,3든 상호작용만 보면 되고 두개 p-value를 보니 공통의 상호작용이 의미가없으니

기울기 따로따로 할 필요 없겟구나! 생각하면 될 듯.

* 공분산분석모형으로 바로 넘어감.
* 보정된처리평균 lsmeans (SAS)를 계산해서 한번 비교해서 봄.
* Tdiff(차이값의 크기에 대한 t, 차이값, 유의확률 계산해줌)

일단 교호작용 빠진것의 결과에선 Type3 SS를 읽어야함.

순수한 machine의 SS를 봐야하거든. (직물처리의 효과를 다 제거하고 기계의 효과만 봐야하니까 그런 것임)

* 기계가 유의하지않음. 1원배치모형은 유의한것처럼 보이지만, 실제로 type3는 유의하지않다
* 즉, 기계간의 직물강도의 차이가있다고 뜨는데 이건 원사의 차이인데 기계간의 차이가 아니다.

세개의 직선을 평행하게끔 적합시킨 모양

하나직선으로하는게 나을까 세개로 하는게 나을까.

그림을 보니 절편이 크게 차이가없어서 하나로 적합시켜도 괜찮겟다..

처리평균이 같다는 모습이 볼 수 있는데,

* Tao 효과에 대해서 약간 다른 결과가 나옴, 세개 이상 처리평균 할 때
* 쌍별비교가 3개나오면 전체비교해서 유의수준 5% 시각한거랑
* 쌍별비교, 전체비교랑 제1종오류확률이 14.몇퍼센트로 나옴.
* 그래서 유의수준에 3정도 나눠야 원하는 유의수준에 도달 할 수 있따.
* LSmean(i),(j)여기선 쌍별비교에서 유의수준 5퍼센트로 하면 안됨!! 개별 비교할 때 주의해야함!!!!!!!!!!! (잊지말아줘)

95% 신뢰구간 귀무가설 주장하는 값을 포함한다 포함하지않는다 그 그래프 그림 모양

Y comparisons for machine

Note4)

가설검증에 SSTrt=SSE`-SSE, 사용해서 F검정한 거 체크

완전모형하에서 SSE 체크, Exx, Eyy,Exy 사용.

여기서 저렇게 계산 안하고 Note4처럼 계산 가능함.

일원배치법

SST = *SSTrt* + SSE(잔차제곱합)

위에서 아래의 노테이션으로 바꿈. 그래서 간접계산함.

Syy = *Tyy* + Eyy (직교분할가능)

위에껄 가지고 아래도성립합.

Sxx = Txx + Exx (직교분할가능)

Sxy = Txy + Exy (직접 풀어보면 직교분할 가능) -3번째 케이스

이 관계식을 이용하면 Exx, Eyy, Exy를 간접계산할 수 있다.

3번째케이스 잘 보면 두개 따져보면 직교함.

책 이상의 내용\*\*

랜덤화블록설계 + 공변수를 추가한 모형임.

Yj = beta0 + beta1x + eps

Yj = beta0 + beta1(x-x.bar) + eps

공변수 들어가면,

블락효과에 대응되는 제곱합을 넣어줌 (Ryy)

* 랜덤화블록설계에서 총제곱합은 블락제곱합+처리+잔차 = SST
* 여섯개의 끼리끼리 곱한건 다 사라지는 것이 되거든

Sxy = Txy + Rxy + Exy 꼴로 나옴. ( 이 식이 성립함 )ab-a

\*\*\* 기타 공분산분석모형

랜덤화블록설계에 공변수 도입해서 만든 것.

완전모형 하에서의 추정치

자유도= 처리의수a, 블락의수b(실험회수 ab) - mu의추정치 – 회귀계수추정치-처리효과추정치-블락효과의추정치 = ab-a-b

축소모형하에서의 추정치

처리효과가 0인 모형이 축소모형.

블록효과는 여전히 남아있는 것.

모든처리효과가 0이다 : 처리모형**=> (랜덤화블록설계 축소모형)**

일원배치법 공분산모형.

완전모형 써서 한번 봐봐. **=>공분산분석모형**

두개를 겹쳐씀,

BetaFull(추정치)= Exy/Exx

SSE = Eyy-Exy^2/Exx (Exx = double summation (xij – x(i)bar)^2…)

두개 겹쳐쓰는데,

일단 유사하잖아? betaj, taoi만 차이가 있음.

이 둘의 차이지만 i번째 처리집단에서 절편이 달라지는거랑 j번째 블록에 대해 절편이 달라지는 것에 대한 유사성이 있음.

처리집단에서 달라지는게 일원배치ㅗ안전모형 공분산분석모형

블락별로 절편이 달라지는게 랜덤화블록설계 공분산분석의 축소모형.

* 여기에 맞춰서 직선을 적합시키는 거잖아. 다만 이제 처리집단별이냐 블락별이냐. 그 차이지.

이때 계산하기에 이 두개가 유사하기 때문에 최소제곱추정법 계산에서

하나만 받아들이면 나머진 계산에서 이끌어내서 계산할 수 있다.

그러니까 축소모형에서는

beta\_red\_hat =double summation(xij-x.(j)bar)(yij-y.j.bar)/ double summation(xij-x(j)bar)^2

첨자를 i닷에서 닷j로 바꾸면 끝임.

SSE’(축소모형) = double summation(yij-yjbar)^2 - double summation

위에꺼 그대로 유도해서 쓰면 됨. 그래서 Txx + Exx 이런식으로 나오는 그런 서메이션 결합으로 나타남.

축소모형에서 분모모형이 Txx + Exx로 갈라질 수 있다.

Sxx’ = Txx + Exx

Beta.reduced – hat = (Txy + Exy) / (Txx + Exx)

SSE’ = Tyy + Eyy – (Txy + Exy)^2 / ( Txx + Exx )

공식들 체크해두기.

모든 처리가설이 0이란 검정을 하려면(모든처리효과가 0이라고 둔거가 a개고 그 합이 0이니까 자유도의 차가 a-1 그걸 나누면 F통계량

1. 회귀직선의 기울기의 동질성검정
2. 상호작용 여부
3. 회귀의 유의성검정
4. 그리고 모든 처리효과가 0이냐의 검정(본론부분)

* 여기서도 이 과정을 따라야 함.

회귀유의성 검정할 때 필요한거

SSR = Exy^2/Exx 회귀공변수 제곱합 파트. 이거 아에랑 뒤에랑 원리가 같음.

일원배치법 테이블의 원리를 매우매우 잘 알아야함(이 부분의 내용이 시험에 백프로 낼것임)

테이블 주고 내가 그 테이블 내용 계산할 수 있으면 됨.

Step1.

테이블 기계적으로 계산 가능함.

랜덤화블록설계의 경우

* 이런 원리를 이용하면 다른 모형도 똑같이 적용 가능하니까 직접 한번 풀어보기.

SSE’ 구하는 공통원리가 있음.

오차항에 검정하고 싶은 것을 더하는 것이다.\*\*

(테이블 요소 더해서 제곱해서 빼면되는것임, 원리가 간단함)

ㅇ기서 공변수 한 개인데

실제로 공변수 두개 세개 가능함.(여러 개 가능-> 이때 sas가 계산함)

이원배치법

반복이잇는 이원배치는 첨자 세개

표현이 어떤 제곱합인지 체크하기.

축소모형하에서의 잔차제곱합.

* 지난학기 배웟던 모형을 다 여기에 적용해 볼 수 있음.
* (과거모형 복습하기)

F = SSA/(a-1) / MSE

**(시험칠 때 주의해)**

이때 자유도는 df1 = a-1 df2 =

회귀계수에 대한 자유도 하나 빼는거 너네 진짜 잘 틀리는거 중요해

(회귀계수 중요행. 분산분석표의 자유도를 쓰면 안되고 회귀계수의 자유도를 써야행)

ANOVA는 이원배치법의 자유도고

완전모형은 공변수 하나 들어갔잖아, 그래서 한 개 빼는거 잊어버리더라고 하나 더 빼야함

(예 10.1a) = 이원배치법자료

조합이 총 4가지가 나오는데, 각각의 조건 하에서 6번씩 재배한 수확량과 재배면적 데이터

여기서 A,B,AB상호작용이 유의한지 공변수를 넣고 한번 체크해보자!

계산을 내지는 않지만, 원리를 알고 계산할 줄은 알아야 함.

모형에서 ()+()+()+() 여기서

4가지 평균, 혹은 부분합을 알아야 함

T ( ij )dat 이런 계산을 알아야 함.

Y에 대한 총합,

써메이션 파트에

Ayy = summation yi.. ^2/ br

T 안쓰는게 헷깔릴까봐 y로 쓴거임.

일원화시킨 총제곱합이니까. Eyy= Syy – AByy

여기 Homework에서 **df 값 체크 조심해야해.**

잔차의 자유도 총 24번 실험에서 24-(a효과1)-(b의효과1)-(상호작용에대한 것1)-(평균에대한 자유도1)-(공변수1개)=19

그래서 공변수 한 개 진짜 잊으면 안되 이거 시험에서 너네 틀려 조심해

음,

CTxy 계산방법

Correction term의 계산 규칙

CTy = {GTy(총합)}^2 / N(총 실험횟수)

CTx = {GTx (총합)}^2 / N(총 실험횟수)

CTxy = {GTx (총합)}\* {GTy (총합)} / N(총 실험횟수)