Ensidig variansanalys - inlämningsuppgift 2

STAG24 - Variansanalys

VT 21

Kim Thurow

Innehållsförteckning

1 Sammanfattning

Som deltagare på kursen Variansanalys förväntas var kursdeltagare lämna in beräkningsuppgifter från kursboken Design and analysis of experiments av D. C. Montgomery. Den här rapporten är den andra av fem. Uppgifterna beräknas i onlineprogrammet SAS Studio och koden redovisas som bilaga.

2 Effektivitet hos isolerande vätskor

Effektiviteten hos fyra olika isoleringsvätskor har studerats. Fyra olika vätskor har studerats i ett randomiserat experiment, vilket resulterat i 24 observationer, sex per grupp. Eftersom det är lika många observationer i varje grupp är undersökningen balanserad.
Modellen som används är y = $\mu + \tau_i + \epsilon_{ij}$. Jag gör antagandet att slumpfaktorn (det vill säga residualerna eftersom övriga variabler är konstanter) är oberoende och normalfördelade: ϵ N(0,σ^2). Signifikansnivån α = 0.05 används genomgående genom hela rapporten.

2.1 a) Hypotesprövning

Hypotesen om valet av vätska påverkar isoleringseffekten är tvåsidig.

H_0: τi= 0 för alla i

H_1: τi $\neq$ 0 för något i

Genom PROC GLM-funktionen i SAS visar p-värdet på 0.0525 och nollhypotesen kan därmed inte nollhypotesen förkastas. Inga skillnader kan ses mellan de olika vätskorna i experimentet.

2.2 b) Bäst effektivitet

Uppgiften består i att välja den vätska som ger längst effekt i h, vilket bör vara den vätska som har högst medelvärde. Nedan ses ett t-konfidensintervall, där även medelvärdet återgetts. Med tanke på att grupp 3 har högst medelvärde samt blev signfikant mellan grupp 1 och grupp 2 bör det vara det logiska valet.

The GLM Procedure

t Confidence Intervals for effect

Alpha	0.05
Error Degrees of Freedom	20
Error Mean Square	3.299667
Critical Value of t	2.08596
Half Width of Confidence Interval	1.546914

fluid	N	Mean	95% Confidence Limits	
3	6	20.9500	19.4031	22.4969
4	6	18.8167	17.2698	20.3636
1	6	18.6500	17.1031	20.1969
2	6	17.9500	16.4031	19.4969

Figure 1: t-konfidensintervall

2.3 c) Residualanalys

För att kontrollera att experimentets antaganden är oberoende bör residualerna vara normalfördelade. Nedan visas en QQPlot med residualerna på y-axeln och normalkvantiler på x-axeln. I och med residualerna stiger längs en rät linje bekräftas antagandet om att e_{ij} oberoende bekräftas.

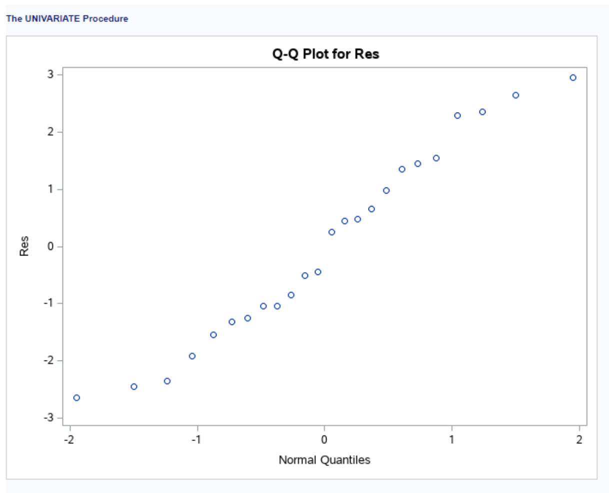

Figure 2: QQPlot residualer på y-axeln samt normalkvantiler på x-axeln

Nedan testas om modellen verkar vara korrekt. Inga residualer verkar avvika från sitt förväntade värde och man kan därmed konstantera att modellen som använts verkar stämma.

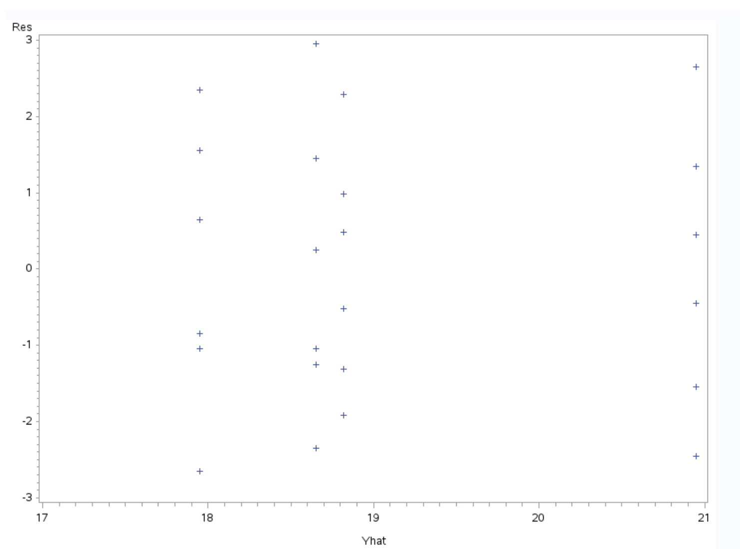

Figure 3: Kontroll av modell m.h.a residualer

2.4 3.59 Kruskal-Wallis test

I uppgift 3.59 ska datamaterialet från uppgift 3.27 återanvändas till ett rangsummatest, Kruskal-Wallis test. P-värdet (10,15%) överstiger det kritiska värdet 5% och nollhypotesen kan därmed inte förkastas.

The NPAR1WAY Procedure

Wilcoxon Scores (Rank Sums) for Variable effect Classified by Variable fluid					
fluid	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
1	6	67.00	75.0	14.996739	11.166667
2	6	53.50	75.0	14.996739	8.916667
3	6	111.00	75.0	14.996739	18.500000
4	6	68.50	75.0	14.996739	11.416667
Average scores were used for ties.					

Kruskal-Wallis Test		
Chi-Square	DF	Pr > ChiSq
6.2177	3	0.1015

Figure 4: Kruskal-Wallis rangsummatest

3 3.29 Skillnad mellan kemister

I uppgift 3.29 tillfrågas fyra kemister att avgöra procenthalten av methylalkohol i en viss kemisk förening.

3.1 a) Hypotesprövning

I a-uppgiften undersöks om valet av kemist påverkar resultatet av studien, vilket är en tvåsidig hypotes. Det vill säga, skiljer sig medelvärdet tillräckligt mycket mellan grupperna för att vara en signifikant skillnad.

H_0: $\tau i = 0$ för alla i

H_1: $\tau i \neq 0$ för något i

P-värdet ligger på cirka 0,08, väl över 5 procent. Inga variationer mellan kemister kunde alltså påvisas.

The GLM Procedure

Class Level Information		
Class	Levels	Values
chemist	4	1 2 3 4

Number of Observations Read	12
Number of Observations Used	12

The GLM Procedure

Dependent Variable: percentage

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1.04456667	0.34818889	3.25	0.0813
Error	8	0.85820000	0.10727500		
Corrected Total	11	1.90276667			

Figure 5: Test för signifikant skillnad mellan olika operatörer

3.2 b) Residualanalays

Som tidigare nämnts, bör residualerna vara normalfördelade för att experimentets antaganden om oberoende ska bekräftas. Nedan visas en QQPlot med residualerna på y-axeln och normalkvantiler på x-axeln. I och med residualerna stiger längs en rät linje bekräftas antagandet om e_{ij} oberoende. Även modellen verkar vara korrekt. Inga residualer avviker från sitt förväntade värde och man kan därmed konstantera att modellen som använts verkar stämma.

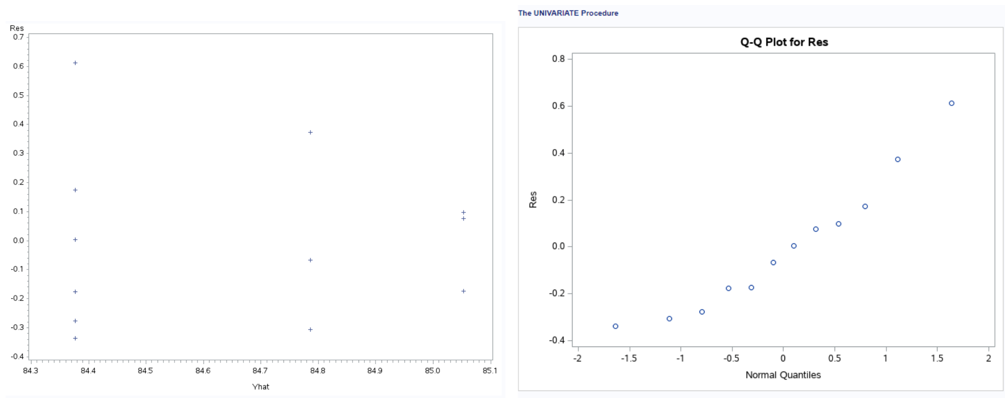

Figure 6: Kontroll av antaganden

3.3 c) Ortogonala kontraster

I c-uppgiften ska ortogonala kontraster tas fram som hade varit hjälpsamma vid starten av experimentet. Första kontrasten ämnar undersöka om det är någon skillnad mellan kemist nummer två och övriga kemister. Hypotesen ställs med kemist no 2 som kontrollgrupp: H_0: μ_2=(μ_1+μ_3+μ_4)/3 samt C_1: -μ_1+ 3μ_2-μ_3-μ_4.
Nästa kontrast ämnar undersöka om det finns skillnader mellan kemist no 1 och kemist 3 och 4.
H_0: μ_1=(μ_3+μ_4)/2 samt C_2: 2μ_1- μ_3-μ_4.
Den sista kontrasten undersöker eventuella skillnader mellan kemist 3 och 4. H_0: μ_3=μ_4 samt C_3: μ_3-μ_4.

Table 1: Linjära ortogonala kontraster

Grupp	Γ_1	Γ_2	Γ_3
1	-1	2	0
2	3	0	0
3	-1	-1	1
4	-1	-1	-1
Summa:	0	0	0

Alla kolumnerna Γ summerar till noll, vilket innebär att kontrasterna är ortogonala. I tabellen nedan utreds kontrasternas eventuella oberoende. Eftersom alla paren summerar till noll är kontrasterna oberoende.

Tidigare hypotesprövningar har inte visat på signifikant skillnad, men här blir p-värdet 3,85 % i första

Table 2: Beräkning av kontrasternas eventuella oberoende

Grupp	Γ_1*Γ_2 $c_i d_i$	Γ_1*Γ_3 $c_i d_i$	Γ_2*Γ_3 $c_i d_i$
1	-1*2 = -2	-1*0= 0	2*0= 0
2	3*0= 0	2*0= 0	0*0= 0
3	-1*-1= 1	-1*1= -1	-1*1= -1
4	-1*-1= 1	-1*-1= 1	-1*-1= 1
Summa:	0	0	0

kontrasten, vilket innebär att den nyanställdes resultat skiljer sig åt från kollegorna. Det visar på att ortogonala kontraster höjer styrkan jämfört med exempelvis F-test där alla kemister ställdes mot varandra. Övriga kontraster där kemist 1 ställdes mot övriga kemister, förutom kemist no 2 blev skillnaderna inte signifikanta. Inte heller när kemist 3 ställdes mot kemist 4. Därmed kan det sägas finnas en skillnad mellan den nyanställde och övriga men inte mellan dem som varit anställda en längre tid.

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Kemist 2 som kontrollgrupp	1	0.65610000	0.65610000	6.12	0.0385
Kemist 1 som kontrollgrupp	1	0.00845000	0.00845000	0.08	0.7861
Skillnad mellan kemist 3 och 4	1	0.38001667	0.38001667	3.54	0.0966

Figure 7: Tabell test ortogonala kontraster.

4 Bilaga: SAS-kod)

```
/*  INL MNING 2 */

OPTIONS LS=80 PS=60 NODATE NOCENTER ;


/*  Uppgift 3.27, 3.59 */

DATA uppgift327 ;
INPUT fluid obs effect ;
LABEL fluid='Fluid␣type'
obs = "Observation"
effect = "Life␣(in␣h)␣at␣35␣kV␣Load"
;

DATALINES ;
1 1 17.6
1 2 18.9
1 3 16.3
1 4 17.4
1 5 20.1
1 6 21.6
2 1 16.9
2 2 15.3
2 3 18.6
2 4 17.1
2 5 19.5
2 6 20.3
3 1 21.4
3 2 23.6
3 3 19.4
3 4 18.5
3 5 20.5
3 6 22.3
4 1 19.3
4 2 21.1
4 3 16.9
4 4 17.5
4 5 18.3
4 6 19.8
;;
```

```
RUN ;



PROC GLM DATA=uppgift327 PLOT=ALL ;
CLASS fluid ;
MODEL effect=fluid ;
MEANS fluid / CLM CLDIFF ;
OUTPUT OUT=new PREDICTED=Yhat RESIDUAL=Res ;
RUN ;

/* Residualanalys */

PROC GPLOT DATA=new ;
PLOT RES*YHAT ;
RUN ;


PROC UNIVARIATE DATA=new NORMAL ;
VAR RES ;
QQPLOT RES ;
RUN ;

/* Kruskal-Wallis test */
PROC NPAR1WAY DATA=uppgift327 WILCOXON ;
CLASS fluid ;
VAR effect ;
RUN ;


/* INLMNING 2 */

OPTIONS LS=80 PS=60 NODATE NOCENTER ;

/* Uppgift 3.29 */


DATA uppgift329 ;
DO chemist = 1 to 4 ;
DO OBS = 1 TO 3 ;
```

```
INPUT percentage @@ ;
OUTPUT ;
END ;
END ;
LINES ;
84.99 84.04 84.38 85.15 85.13 84.88 84.72 84.48 85.16 84.20 84.10 84.55
;
RUN ;


/* Test om valet av operat r p verkar */


PROC GLM DATA=uppgift329 PLOT=ALL ;
CLASS chemist ;
MODEL percentage=chemist ;
OUTPUT OUT=new PREDICTED=Yhat RESIDUAL=Res ;
RUN ;

/* Residualanalys */

PROC GPLOT DATA=new ;
PLOT RES*YHAT ;
RUN ;


PROC UNIVARIATE DATA=new NORMAL ;
VAR RES ;
QQPLOT RES ;
RUN ;


PROC GLM DATA=uppgift329 ;
CLASS chemist ;
MODEL percentage=chemist ;
CONTRAST 'Kemist 2 som kontrollgrupp' chemist -1 3 -1 -1 ;
CONTRAST 'Kemist 1 som kontrollgrupp' chemist 2 0 -1 -1 ;
CONTRAST 'Skillnad mellan kemist 3 och 4' chemist 0 0 1 -1 ;
RUN ;
```