



LUND UNIVERSITY

Inlämningsuppgift 3 - Randomiserade block

STAG24 - Variansanalys

VT 21

Kim Thurow

Innehållsförteckning

1	Sammanfattning	1
2	4.9 Multipla jämförelser	2
2.1	a) Hypotesprövning	2
2.2	b) Tukey's test av skillnader	2
2.3	c) Residualanalys	3
3	4.28 Latinsk kvadrat	4
3.1	a) Hypotesprövning	4
3.2	b) Tukey's test av skillnader	4
3.3	c) Residualanalys	5
4	4.41 Grekisk-latinsk kvadrat	5
4.1	a) Hypotesprövning	5
4.2	b) Residualanalys	5
5	Bilaga: SAS-kod	7

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 tredje av fem. Uppgifterna beräknas i onlineprogrammet SAS Studio och koden redovisas som bilaga.

2 4.9 Multipla jämförelser

En kemist vill studera om olika blandningar av rengöringsvätska påverkar påväxten av bakterier i mjölkbehållare. Analysen är gjord i laboratorie och endast tre försök kan göras per dag. För att minska risken att de olika dagarna skulle öka variationen, görs ett randomiserat blockförsök. Observationerna är tagna under fyra dagar och tre försök per dag har gjorts, vilket resulterat i 12 observationer. Eftersom det är lika många observationer i varje grupp blir experimentet balanserat.

Modellen som används är $y = \mu + \tau_i + \beta_j + \epsilon_{ij}$. Jag gör antagandet att slumpfaktorn (det vill säga residualerna eftersom övriga variabler är konstanter) är oberoende och normalfördelade: $\epsilon \sim N(0, \sigma^2)$. Signifikansnivån $\alpha = 0.05$ används genomgående genom hela rapporten.

2.1 a) Hypotesprövning

Hypotesen om valet av blandning av rengöringsvätska påverkar mängden bakterier inuti en mjölkkartong är tvåsidig.

$H_0: \tau_i = 0$ för alla i

$H_1: \tau_i \neq 0$ för något i

Genom en PROC GLM-funktionen i SAS fås p-värdet på 0.0003 fram och nollhypotesen kan därmed förkastas. Skillnader kan ses mellan de olika blandningarna i experimentet.

2.2 b) Tukey's test av skillnader

Eftersom hypotesprövningen visade på signifikans görs även Tukey's test för att se mellan vilka vätskor som skillnaden ligger. Resultatet, som kan ses på skärmdumpen nedan från SAS, visade att det enbart finns skillnader mellan blandning 3 och övriga.

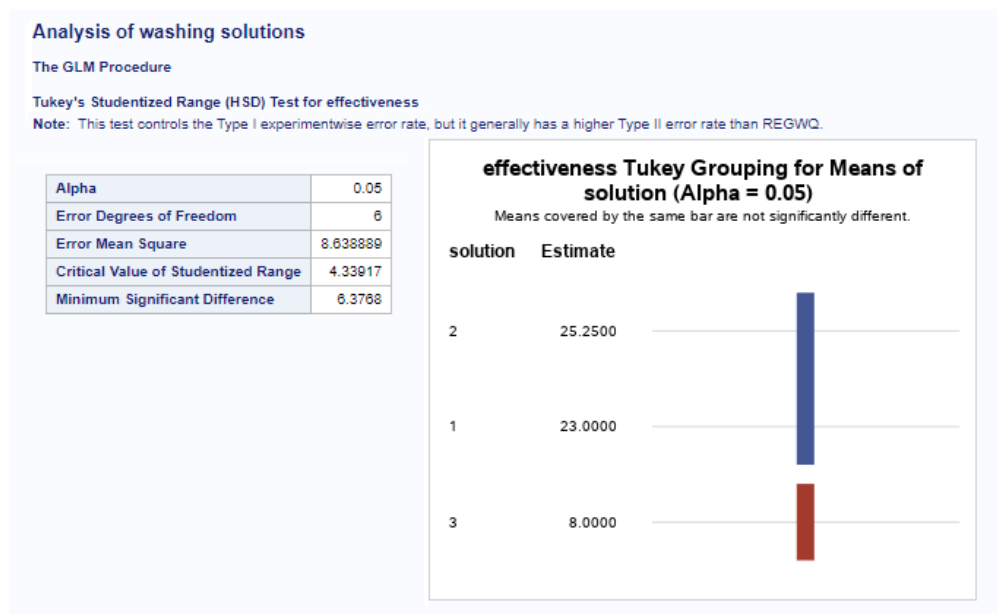


Figure 1: Tukey's test av skillnader

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 e_{ij} oberoende.

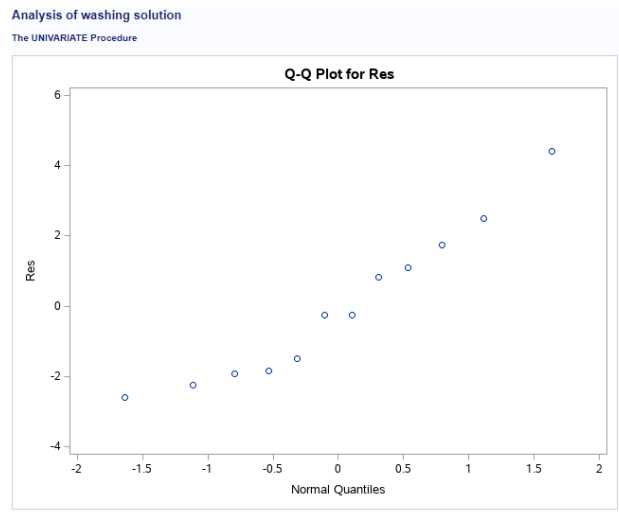


Figure 2: QQPlot residualer på y-axeln samt normalkvantiler på x-axeln

Nedan testas om modellen verkar vara korrekt. Residualerna har plottats mot variabeln 'solution'. Inga residualer verkar avvika från sitt förväntade värde och man kan därmed konstantera att modellen som använts verkar stämma.

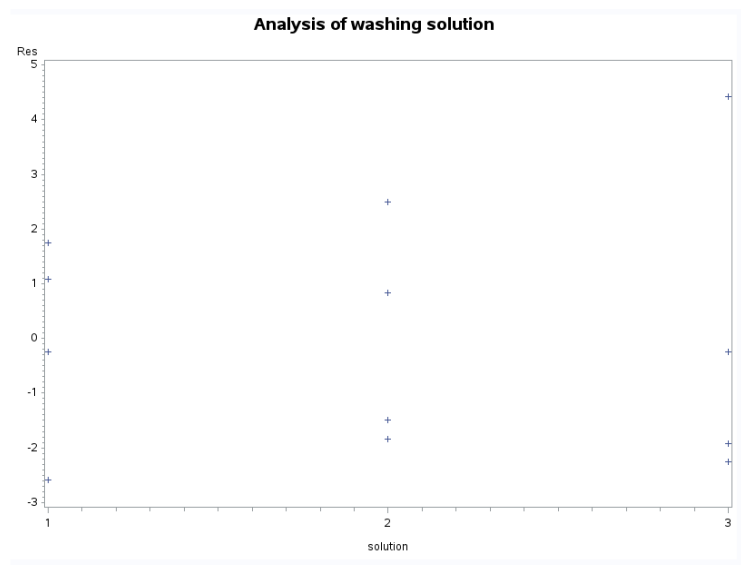


Figure 3: Kontroll av modell m.h.a residualer

3 4.28 Latinsk kvadrat

En ingenjör vill studera effekten av fyra ihopsättningsmetoder (A, B, C och D) på ihopsättningstiden för en viss komponent. Fyra operatörer väljs ut för en undersökning. Ingenjören är medveten om att operatören blir mer trött desto fler ihopsättningar personen gör, vilket påverkar tiden. För att motverka variationen i tid används en latinsk kvadratisk design.

Modellen som används är $y_{ijk} = \mu + \alpha_i + \tau_j + \beta_k + \epsilon_{ijk}$.

Ovan med antagandet att ϵ är oberoende och normalfördelad; $\epsilon_{ijk} \sim N(0, \sigma^2)$. Signifikansnivån $\alpha = 0.05$ används genomgående genom hela rapporten.

3.1 a) Hypotesprövning

$H_0: \tau_j = 0$ för alla j

$H_1: \tau_j \neq 0$ för något j

Den intressanta variabeln 'assembly' studeras genom ett F-test och i SAS fås resultatet att det finns en signifikant skillnad, med p-värdet 0,0042.

3.2 b) Tukey's test av skillnader

Eftersom hypotesprövningen visade på signifikans görs även Tukey's test för att se vart skillnaden ligger. Minsta signifikanta skillnad måste vara 3,2381 vilket innebär att det finns skillnader mellan metod 1 och 3, 1 och 4 samt 2 och 3.

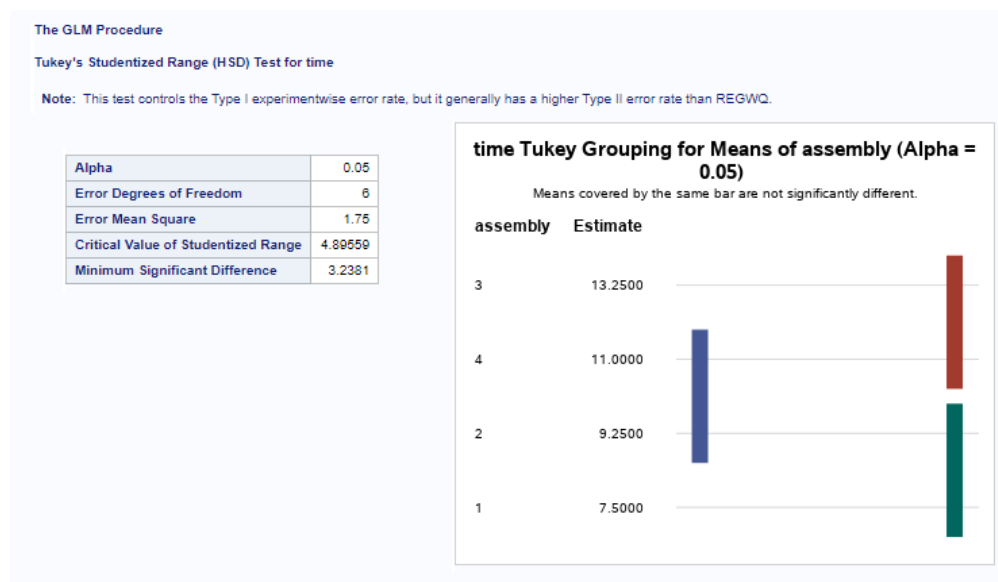


Figure 4: Tukey's test av skillnader

3.3 c) Residualanalys

För att kontrollera att experimentets antaganden är oberoende bör residualerna vara normalfördelade. Nedan visas en Q-QPlot med residualerna på y-axeln och normalkvantiler på x-axeln. I och med residualerna stiger längs en någorlunda rät linje verkar antagandet om e_{ijk} vara korrekt. Residualerna har plottats mot variabeln 'assembly', för att se om residualerna avviker från sitt förväntade värde. Eftersom det inte verkar så, antas modellen stämma.

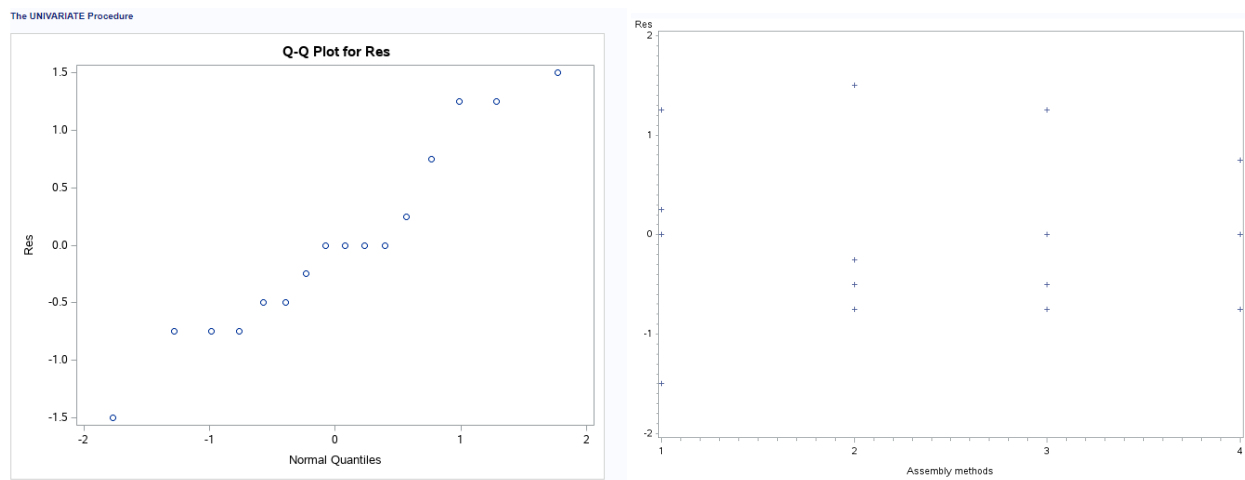


Figure 5: Plot residualer t.h samt Q-QPlot t.v

4 4.41 Grekisk-latinsk kvadrat

Fortsättning på föregående uppgift. Ingenjören vill nu pröva om arbetsplatsen kan påverka tiden för ihopsättningen. Därför läggs en fjärde variabel till, 'workplace'.

Modellen som används är $y_{ijkl} = \mu + \theta_i + \tau_j + \omega_k + \psi_l + \epsilon_{ijkl}$.

Ovan med antagandet att ϵ är oberoende och normalfördelad; $\epsilon_{ijkl} \sim N(0, \sigma^2)$.

4.1 a) Hypotesprövning

$H_0: \tau_j = 0$ för alla j

$H_1: \tau_j \neq 0$ för något j

Den intressanta variabeln 'assembly' och 'workplace' studeras genom ett F-test i SAS. Resultatet blev att det inte finns signifikanta skillnader, med p-värden på 0,17 för 'assembly' och 0,82 för 'workplace'. Det vill säga, arbetsplatsen påverkade inte tiden för personalen att sätta ihop en komponent.

4.2 b) Residualanalys

Nedan kontrolleras residualerna. I och med residualerna stiger, om än ganska så knackigt, längs en någorlunda rät linje verkar antagandet om e_{ijk} vara korrekt. Residualerna har även plottats mot variabeln 'assembly',

för att se om residualerna avviker från sitt förväntade värde. Eftersom det inte verkar så, antas modellen stämma.

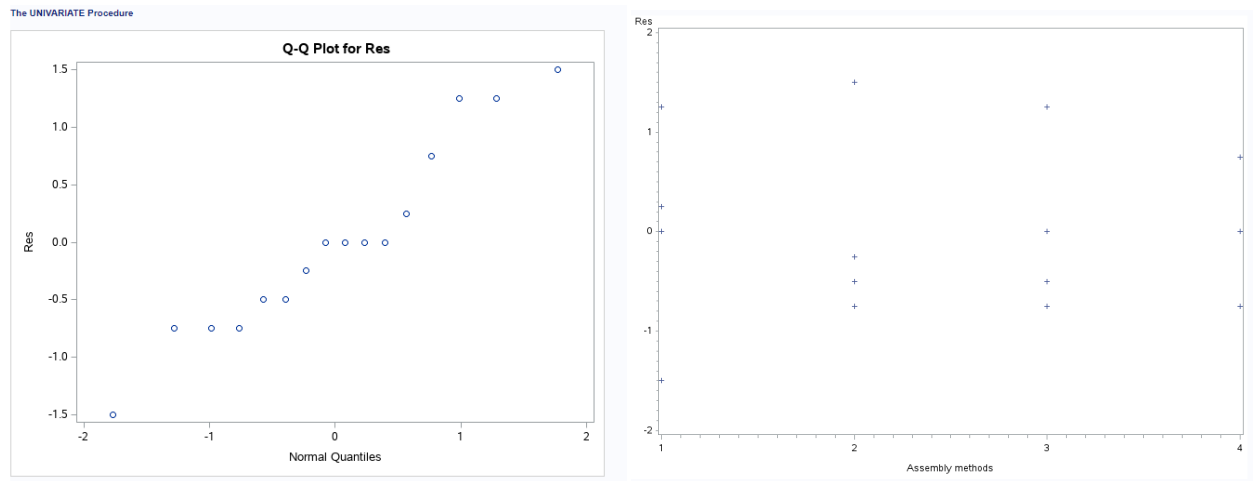


Figure 6: Plot residualer t.h samt QQPlot t.v

5 Bilaga: SAS-kod

```
OPTIONS LS=80 PS=60 NODATE NOCENTER ;
```

```
/* Uppgift 4.9 */
```

```
DATA uppgift49 ;  
DO solution = 1 to 3 ;  
DO days = 1 TO 4 ;  
INPUT effectiveness @@ ;  
OUTPUT ;  
END ;  
END ;  
LINES ;  
13 22 18 39  
16 24 17 44  
5 4 1 22  
;  
RUN ;
```

```
PROC GLM DATA=uppgift49 ;  
TITLE 'Analysis of washing solutions' ;  
CLASS solution days ;  
MODEL effectiveness=solution days ;  
MEANS solution / TUKEY ;  
RUN ;
```

```
/* Residualanalys */
```

```
PROC GLM DATA=uppgift49 PLOT=ALL ;  
CLASS solution days ;  
MODEL effectiveness=solution days ;  
OUTPUT OUT=new PREDICTED=Yhat RESIDUAL=Res ;  
RUN ;
```

```
PROC GPLOT DATA=new ;  
PLOT RES*solution ;  
RUN ;
```



```

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

OPTIONS LS=80 PS=60 NODATE NOCENTER ;

/*  Uppgift 4.28 */

DATA uppgift428 ;
INPUT order operator assembly time ;
LABEL order='Order_of_assembly'
operator = 'Operat r'
assembly = 'Assembly_methods'
time = "time_minutes"
;

DATALINES ;
1 1 3 10
1 2 4 14
1 3 1 7
1 4 2 8
2 1 2 7
2 2 3 18
2 3 4 11
2 4 1 8
3 1 1 5
3 2 2 10
3 3 3 11
3 4 4 9
4 1 4 10
4 2 1 10
4 3 2 12
4 4 3 14
;
RUN ;

PROC GLM DATA=uppgift428 ;
CLASS order operator assembly ;
MODEL time=assembly order operator ;

```

```

MEANS assembly / Tukey ;
RUN ;

/* Residualanalys */

PROC GLM DATA=uppgift428 PLOT=ALL ;
CLASS order operator assembly ;
MODEL time=assembly order operator ;
OUTPUT OUT=new PREDICTED=Yhat RESIDUAL=Res ;
RUN ;

```

```

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

```

```

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

```

```

OPTIONS LS=80 PS=60 NODATE NOCENTER ;

```

```

/* Uppgift 4.41 */

```

```

DATA uppgift441 ;
INPUT order operator assembly time workplace ;
LABEL order= 'Order_of_assembly'
operator = 'Operat r'
assembly = 'Assembly_methods'
time = 'time_in_minutes'
workplace = 'Workplace'
;

```

```

DATALINES ;
1 1 3 11 2
1 2 2 10 3
1 3 4 14 4
1 4 1 8 1

```

```

2 1 2 8 1
2 2 3 12 4
2 3 1 10 3
2 4 4 12 2
3 1 1 9 4
3 2 4 11 1
3 3 2 7 2
3 4 3 15 3
4 1 4 9 3
4 2 1 8 2
4 3 3 18 1
4 4 2 6 4
;
RUN ;

```

```

PROC GLM DATA=uppgift441 ;
CLASS order operator assembly workplace ;
MODEL time=assembly order operator workplace ;
RUN ;

```

```

/* Residualanalys */

```

```

PROC GLM DATA=uppgift441 PLOT=ALL ;
CLASS order operator assembly workplace ;
MODEL time=assembly order operator workplace ;
OUTPUT OUT=new PREDICTED=Yhat RESIDUAL=Res ;
RUN ;

```

```

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

```

```

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

```