



LUND UNIVERSITY

# Ensidig variansanalys - inlämningsuppgift 1

STAG24 - Variansanalys

VT 21

Kim Thurow

# Innehållsförteckning

<b>1</b>	<b>Sammanfattning</b>	<b>1</b>
<b>2</b>	<b>Draghållfasthet i cement</b>	<b>2</b>
2.1	Hypotesprövning F-test 3.11 a) . . . . .	2
2.2	Fisher's LSD test - 3.11 c) . . . . .	3
2.3	Residualanalys - 3.11 d samt e) . . . . .	4
2.4	Tukey's test HSD - uppgift 3.12 . . . . .	5
2.5	Tolka resultat med hjälp av konfidensintervall - uppgift 3.13 . . . . .	6
2.6	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 första av fem. Uppgifterna beräknas i onlineprogrammet SAS Studio och koden redovisas som bilaga.

## 2 Draghållfasthet i cement

Draghållfasthet i cement från en delstat i USA, Portland, har studerats i ett randomiserat experiment. Fyra olika blandningstekniker har använts och totalt finns det 16 observationer, fyra per grupp. Undersökningen är balanserad, eftersom grupperna innehåller lika många observationer.

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:  $\epsilon \sim N(0, \sigma^2)$ . Signifikansnivån  $\alpha = 0.05$  används genomgående genom hela rapporten.

### 2.1 Hypotesprövning F-test 3.11 a)

A-uppgiften går ut på att testa hypotesen om blandningsteknik påverkar styrkan i cement. Hypotesen är tvåsidig, vi ska testa om det finns en skillnad i medelvärden mellan grupperna. Nollhypotesen blir därmed att alla medelvärdena är lika samt mothypotesen att det finns skillnader mellan minst en av grupperna.

$H_0$ :  $\tau_i = 0$  för alla  $i$

$H_1$ :  $\tau_i \neq 0$  för något  $i$

I SAS får jag fram att det observerade F-värdet samt det kritiska f-värdet.

Table 1: Tabell F-värden

Observerat värde	12.73
Kritiskt värde	2.73

Det observerade värdet är noterbart större än det kritiska värdet och nollhypotesen kan förkastas. Skillnader kan därmed påvisas, vilket innebär att valet av blandningsteknik påverkar styrkan i cement.

## 2.2 Fisher's LSD test - 3.11 c)

När man sedan studerar gruppernas medelvärden med ett Fisher LSD test, kan man se att det minsta värdet som kan skilja mellan grupperna är 174,48, för att resultatet ska räknas som signifikant. De två grupper som därmed inte är signifikanta är grupp 1 och grupp 3.

The GLM Procedure				
t Tests (LSD) for strength				
Note: This test controls the Type I comparisonwise error rate, not the experimentwise error rate.				
Alpha	0.05			
Error Degrees of Freedom	12			
Error Mean Square	12825.69			
Critical Value of t	2.17681			
Least Significant Difference	174.48			

Comparisons significant at the 0.05 level are indicated by ***.				
Mixingtechnique Comparison	Difference Between Means	95% Confidence Limits		
2 - 1	185.25	10.77	359.73	***
2 - 3	222.50	48.02	396.98	***
2 - 4	490.00	315.52	664.48	***
1 - 2	-185.25	-359.73	-10.77	***
1 - 3	37.25	-137.23	211.73	
1 - 4	304.75	130.27	479.23	***
3 - 2	-222.50	-396.98	-48.02	***
3 - 1	-37.25	-211.73	137.23	
3 - 4	267.50	93.02	441.98	***
4 - 2	-490.00	-664.48	-315.52	***
4 - 1	-304.75	-479.23	-130.27	***
4 - 3	-267.50	-441.98	-93.02	***

Figure 1: Skillnader i medelvärden. \*\*\* innebär att skillnaden är signifikant.

## 2.3 Residualanalys - 3.11 d samt e)

Eftersom residualerna är det enda i modellen som inte är konstant, blir det den slumpmässiga faktorn. Med hjälp av residualerna kan man därför kontrollera att den modell som används samt de antaganden man gör är korrekta samt oberoende.

För att kontrollera att de antaganden som görs är oberoende bör residualerna vara normalfördelade. Nedan visas en QQPlot med residualerna på y-axeln och normalkvantiler på x-axeln. I och med residualerna, alltså avvikelsen från medelvärdet, stiger längs en rät linje verkar residualerna vara normalfördelade och därmed kan antagandet om slumpfaktorn  $e_{ij}$  oberoende bekräftas.

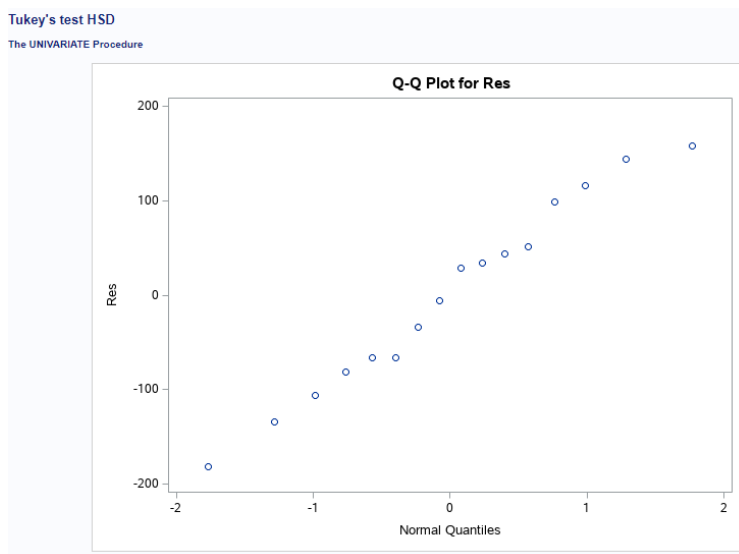


Figure 2: QQPlot över residualerna

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.

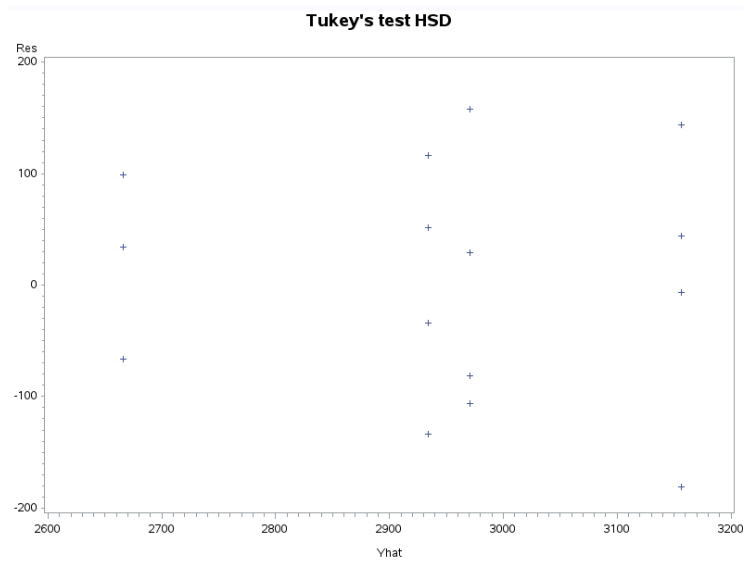


Figure 3: Kontroll av modell m.h.a residualer

## 2.4 Tukey's test HSD - uppgift 3.12

När jag jämför gruppernas medelvärden med Tukeys test istället, blir värdet på den minsta signifikanta summan högre. Detta medför att skillnaden mellan tre av fyra grupper inte blir signifikant. Den enda grupp som kan sägas skiljas från övriga grupper är grupp 4, gruppen med lägst medelvärde.

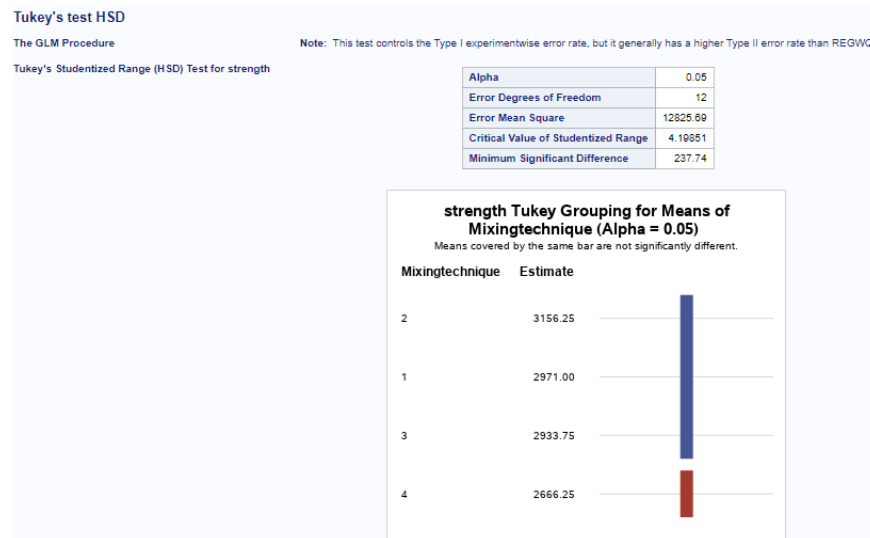


Figure 4: Skillnader i medelvärden. Tukey's test

Skillnaden mellan Tukey's test och Fisher's test LSD ligger i att Fisher's test egentligen är ett t-test med en konstant som multipliceras med standardavvikelsen. Tukey's test är en korrigering med ett tabellvärde som bygger på frihetsgrader och antal observationer, vilket medför att Tukey's test blir mer korrekt.

## 2.5 Tolka resultat med hjälp av konfidensintervall - uppgift 3.13

Hur kan man med hjälp av konfidensintervall tolka experimentets resultat? Om man bortser från att det i SAS tydligt markeras ut vilka grupper som är signifikanta eller ej, så kan man med hjälp av LSD-värdet ta reda på vilka som faktiskt skiljer sig åt. I detta material är värdet 174.48 och det enda fall där skillnaden är mindre än värdet är mellan grupp 1 och grupp 3.

The GLM Procedure	
t Confidence Intervals for strength	
Alpha	0.05
Error Degrees of Freedom	12
Error Mean Square	12825.69
Critical Value of t	2.17881
Half Width of Confidence Interval	123.3759

Mixingtechnique	N	Mean	95% Confidence Limits	
2	4	3156.25	3032.87	3279.63
1	4	2971.00	2847.62	3094.38
3	4	2933.75	2810.37	3057.13
4	4	2666.25	2542.87	2789.63

Alpha	0.05
Error Degrees of Freedom	12
Error Mean Square	12825.69
Critical Value of t	2.17881
Least Significant Difference	174.48

Comparisons significant at the 0.05 level are indicated by ***.				
Mixingtechnique Comparison	Difference Between Means	95% Confidence Limits		
2 - 1	185.25	10.77	359.73	***
2 - 3	222.50	48.02	396.98	***
2 - 4	490.00	315.52	664.48	***
1 - 2	-185.25	-359.73	-10.77	***
1 - 3	37.25	-137.23	211.73	
1 - 4	304.75	130.27	479.23	***
3 - 2	-222.50	-396.98	-48.02	***
3 - 1	-37.25	-211.73	137.23	
3 - 4	267.50	93.02	441.98	***
4 - 2	-490.00	-664.48	-315.52	***
4 - 1	-304.75	-479.23	-130.27	***
4 - 3	-267.50	-441.98	-93.02	***

Figure 5: t-konfidensintervall, medelvärden (t.v) samt skillnader i medelvärden (t.h)

## 2.6 Bilaga: SAS-kod)

```
/* INLMNING 1 */

OPTIONS LS=80 PS=60 NODATE NOCENTER ;

/* Uppgift 3.11 */

DATA uppgift311 ;
INPUT Mixingtechnique obs strength ;
LABEL Mixingtechnique='Mixing_technique'
      obs = "Observation"
      strength = "Tensile_strength_(lb/sq.inch)"
;

DATALINES ;
1 1 3129
1 2 3000
1 3 2865
1 4 2890
2 1 3200
2 2 3300
2 3 2975
2 4 3150
3 1 2800
3 2 2900
3 3 2985
3 4 3050
4 1 2600
4 2 2700
4 3 2600
4 4 2765
;;
RUN ;

PROC GLM DATA=uppgift311 PLOT=ALL ;
CLASS Mixingtechnique ;
MODEL strength=Mixingtechnique ;
```



```

MEANS Mixingtechnique / CLM CLDIFF LSD ;
OUTPUT OUT=new PREDICTED=Yhat RESIDUAL=Res ;
RUN ;

```

```

/* Uppgift 3.13 Tukeys test */

```

```

PROC GLM DATA=uppgift311 ;
TITLE "Tukey's test LHSD" ;
CLASS Mixingtechnique ;
MODEL strength=Mixingtechnique ;
MEANS Mixingtechnique / TUKEY ;
OUTPUT OUT=tukey ;
RUN ;

```

```

/* Residualanalys */

```

```

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

```

```

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

```