Labbrapport i Statistik

Laboration 13

732G46

Michael Debebe

Avdelningen för Statistik och maskininlärning
Institutionen för datavetenskap
Linköpings universitet

2021-12-01

Innehåll

Introduktion

I denna laboration kommmer tre dataset att analyseras.

Det första datasetet behandlar fyra olika limsystem att analyseras i form av att jämföra deras skalstyrka.

I det andra datasetet jämförs kostnaderna för att framföra två typer av däck i hastigheter mellan 10 och 70 hastighetsenhter.

I det tredje och sista datatsetet analyseras fyra olika typer av behandlingar på tre olika typer av land där koärtor skördas.

Lärande mål med denna laboration är att jag som enskild student ska besluta vilken typ av modell och tillvägagångssätt som är det lämpligaste för min typ av datamaterial.

Databehandling

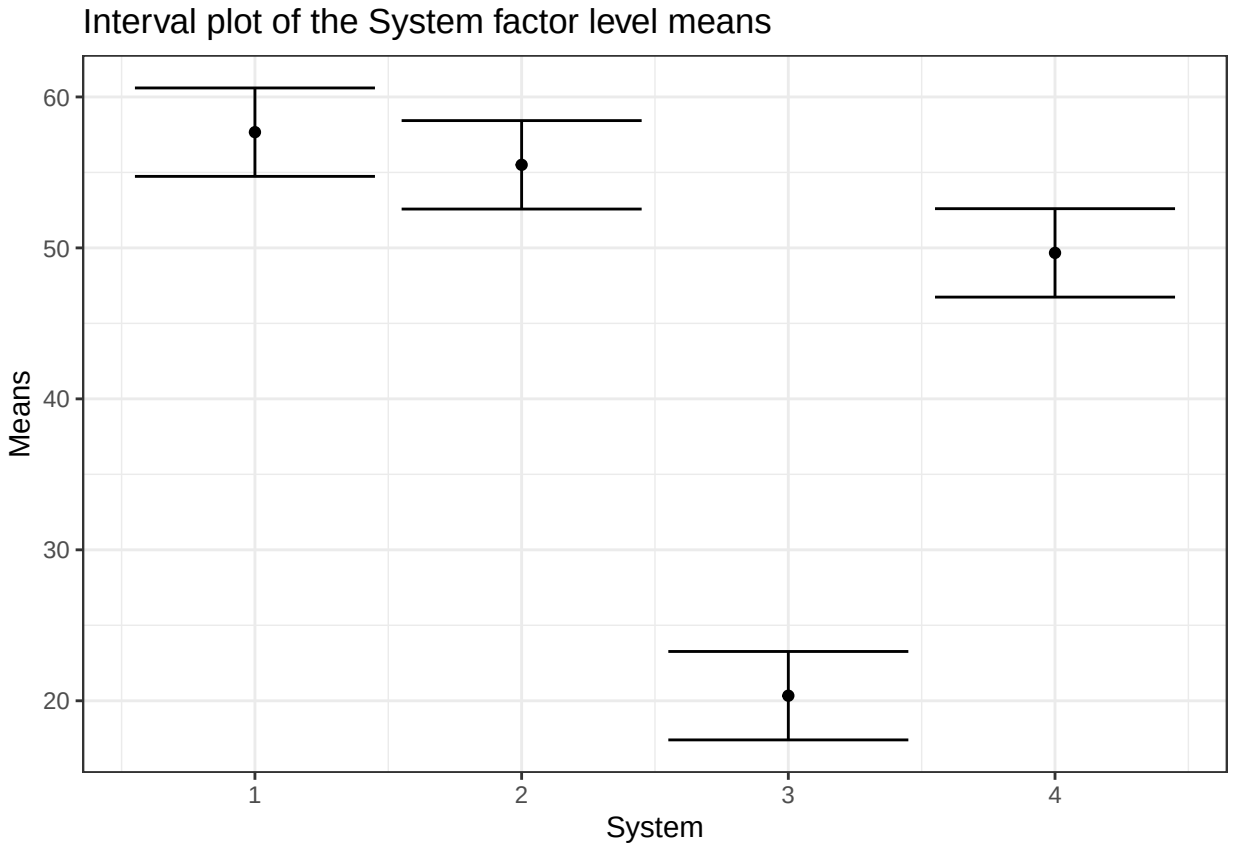
```
Peelstrength <-c(60,63,57,53,56,57,57,52,55,59,56,54,20,20,20,22,21,19,52,53,44,48,48,53)
System <-c(1,1,1,1,1,1,2,2,2,2,2,2,3,3,3,3,3,3,4,4,4,4,4,4)
System <-as.factor(System)
Adhesive <-as.data.frame(cbind(Peelstrength,System))

Speed <-c(10,20,20,30,40,40,50,60,60,70)
Cost <-c(9.8, 12.5, 14.2, 14.9, 19.0, 16.5, 20.9, 22.4, 24.1, 25.8,
15.0, 14.5, 16.1, 16.5, 16.4, 19.1, 20.9, 22.3, 19.8, 21.4)
Type <-c("A","A","A","A","A","A","A","A","A","A","B","B","B","B","B","B","B","B","B","B")
df <-as.data.frame(cbind(Speed,Cost,Type))

Yields <-c(45,43,46,50,45,48,61,60,63,58,56,60)
Block <-c(1,2,3,1,2,3,1,2,3,1,2,3)
Treatment <-c(1,1,1,2,2,2,3,3,3,4,4,4)
Cowpeas <-as.data.frame(cbind(Yields,Block,Treatment))
Cowpeas$Block <-as.factor(Cowpeas$Block)
Cowpeas$Treatment <-as.factor(Cowpeas$Treatment)
```

Uppgift 1

An experiment resulted in data to determine the effectiveness of four adhesive systems for bonding insulation in a chamber. The data are a measure of the peelstrength of the adhesives and are showed below. Adhesive system Observed peelstrength

a) Gör ett lämpligt diagram över data.

```
ggplot(data = tt,aes(x=System,y=Means))+
  geom_point()+
  geom_errorbar(aes(ymin=lower,ymax=upper))+
  ggtitle(label = "Interval plot of the System factor level means")+
  theme_bw()
```



I detta fall sågs ett 95 procentigt konfidensintervall för medelvärdena som ett lämpligt sätt att analysera datan. Det som syns i diagrammet är hur 3 av 4 system har medelvärde som ligger mellan 60 och 50. System 3 är det system som lägst medelvärde. Den generella spridningen är väldigt lik för 3 av 4 system medan system 3 som sagt sticker ut.

b) Ange en modell för data och de förutsättningar som ska vara uppfyllda för test och konfidensintervall

```
Adhesive_c <-aov(Peelstrength~System,data = Adhesive)
Anova(Adhesive_c)
```

```
## Anova Table (Type II tests)
##
## Response: Peelstrength
##             Sum Sq Df  F value     Pr(>F)
## System    5390.458  3 225.3065 1.4034e-15 ***
## Residuals  159.500 20
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

I detta fall har en envägsanova modell ansetts vara den lämpligaste modellen, eftersom att det enbart är en faktornivå.

$$Y_{ij} = \mu_i + \varepsilon_{ij}$$

Där i går från 1 till antalet nivåer och j går från 1 till antalet observationer nivå i. Yij är responsvariabeln som bildas av parametern ui och slumpvariabeln ui.

Förutsättningarna som måste vara uppfyllda

- Variansen inom fördelningarna för observationerna är lika.
- Observationerna inom fördelningarna är normalfördelade.
- Observationerna är slumpmässiga mellan faktornivåerna.

c) Pröva om det finns någon skillnad mellan ” Adhesive system “. Formulera relevant hypotes. Dra slutsatser. Visa beräkningar.

```
anova(Adhesive_c)
```

```
## Analysis of Variance Table
##
## Response: Peelstrength
##           Df   Sum Sq  Mean Sq  F value     Pr(>F)
## System     3 5390.458 1796.819 225.3065 1.4034e-15 ***
## Residuals 20  159.500    7.975
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

H0:Alla ui är lika (i=1,2,3,4)

H1:Alla ui är inte lika

$$F = \frac{1796.82}{7.97} = 225.45$$

Ifall av att värdet från F-testet överstiger det kritiska värdet kan vi med 95 % konfidens förkasta nollhypotesen om att det inte finns några skillnader i medelvärden mellan systemen.

$$F(.95, 3, 20) = 3.098$$

I detta fall då $(225.45 > 3.098)$ samtidigt som att vi får ett p-värde som summeras till $(1.4034e - 15)$ kan vi med 95 % konfidens förkasta H_0 om att det inte finns några skillnader i medelvärden mellan systemen.

d) Om det finns signifikanta skillnader enligt deluppgift c), bilda konfidensintervall för alla parvisa skillnader. Använd 95% simultan konfidensgrad och ange vilken metod du använt. Vilka "behandlingar" kan anses vara skilda från varandra? Visa beräkningar.

```
## (Intercept)     System2     System3     System4
##        57.7        55.5        20.3        49.7
```

Formeln som används för parvisa jämförelser med Tukey-kramers metod

$$\hat{D} \pm \frac{q_{k,n-k_i 1-\alpha}}{\sqrt{2}} \cdot \sqrt{s^2\left\{\hat{D}\right\}}$$

$$s^2\left\{\hat{D}\right\} = \frac{2 \cdot MSE}{n}$$

$\hat{D}_{2-1} = 55.50 - 57.67 = -2.17$

$\hat{D}_{3-1} = 20.34 - 57.67 = -37.33$

$\hat{D}_{4-1} = 49.67 - 57.67 = -8$

$\hat{D}_{3-2} = 20.34 - 55.50 = -35.17$

$\hat{D}_{4-2} = 49.67 - 55.50 = -5.84$

$\hat{D}_{4-3} = 49.67 - 20.34 = 29.33$

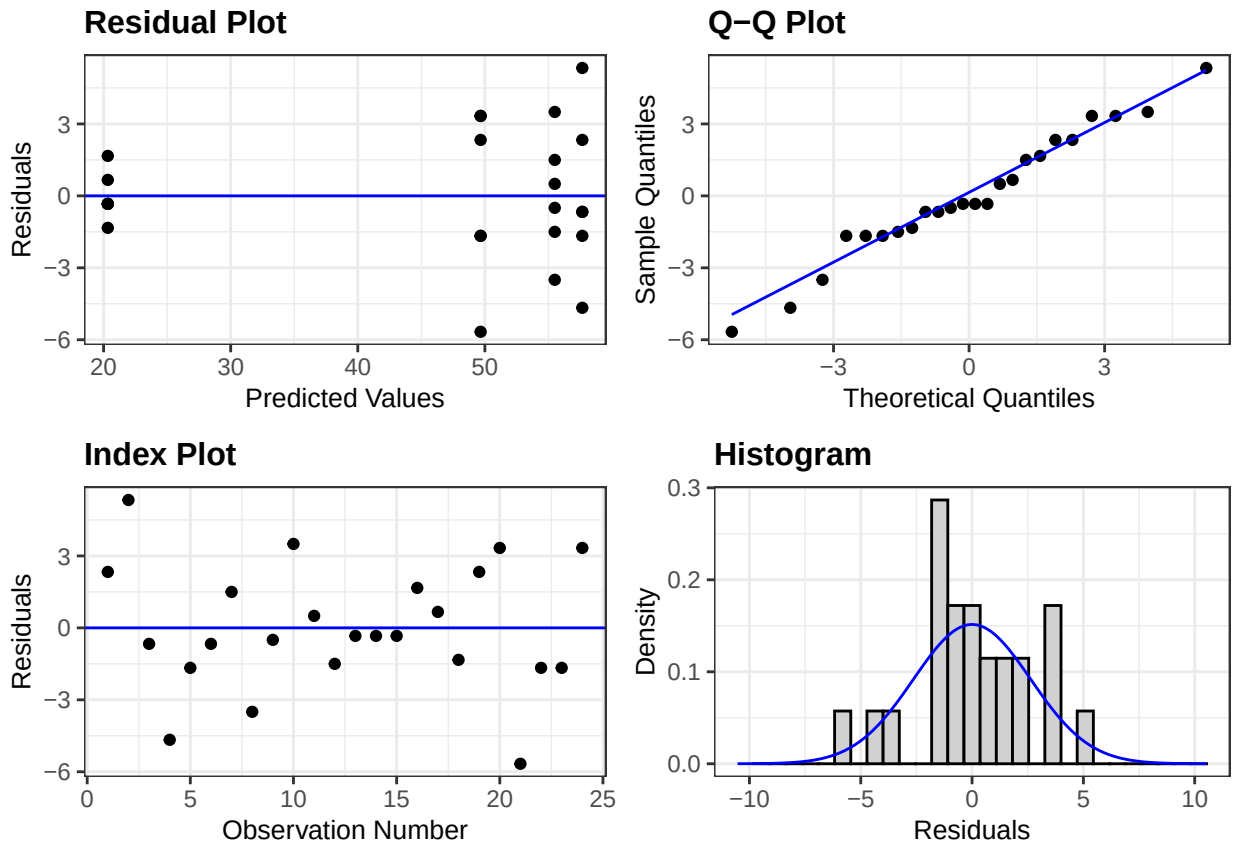
```
TukeyHSD(Adhesive_c,conf.level = 0.95)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Peelstrength ~ System, data = Adhesive)
##
## $System
##       diff    lwr    upr p adj
## 2-1  -2.17  -6.73   2.40 0.556
## 3-1 -37.33 -41.90 -32.77 0.000
## 4-1  -8.00 -12.56  -3.44 0.000
## 3-2 -35.17 -39.73 -30.60 0.000
## 4-2  -5.83 -10.40  -1.27 0.009
## 4-3  29.33  24.77  33.90 0.000
```

I samtliga av fallen förutom jämförelsen mellan grupp 2 och 1 kan vi se att tabellen returernar p-värden som på 5 % signifikansnivå är signifikanta. Detta innebär att det vi kan dra slutsatsen om att det finns signifikanta skillnader mellan de olika limsystemen, för samtliga parvisa jämförelser förutom mellan 1 och 2.

e) Utför en residualanalys och kommentera angående forutsättningarna i deluppgift b)

```
ggResidpanel::resid_auxpanel(residuals = residuals(Adhesive_c),
                             predicted = fitted(Adhesive_c))
```



I jämförelsen mellan residualer och de anpassade värdena är varianserna inom de olika systemen relativt lika och det är mot de sista observationerna som vissa avvikelser kan finnas, samtidigt som att variansen inom observationerna för första systemet är väldigt liten och jämn. I helhet ser däremot spridningen bra ut, med tanke på att enbart 3 av 24 observationer är avvikande.

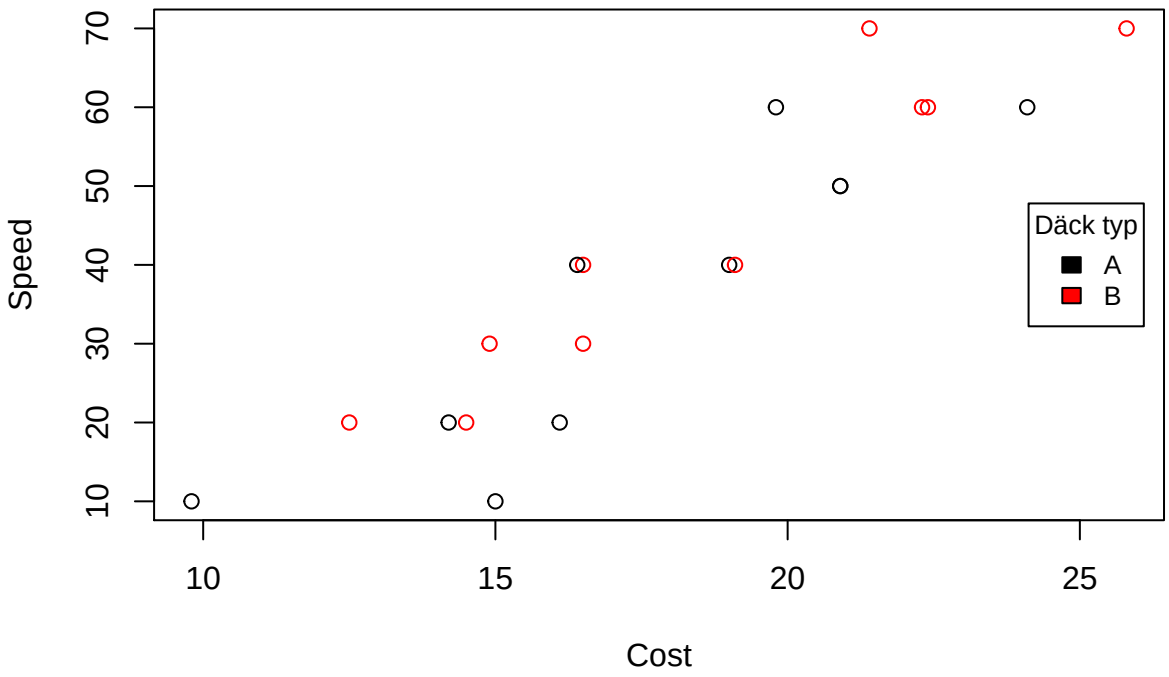
I normalfördelningsplotten syns de hur observationerna följer normalfördelningslinjen moderat bra, tre utomstående observationer syns som dessutom är utmärkta.

I histogrammet ser residualerna fint fördelade ut och följer normalfördelningskurvan bra, däremot skulle man kunna säga att datat är snäppet negativt skev. I sin helhet uppfyller de kritierierna i detta fall.

Uppgift 2

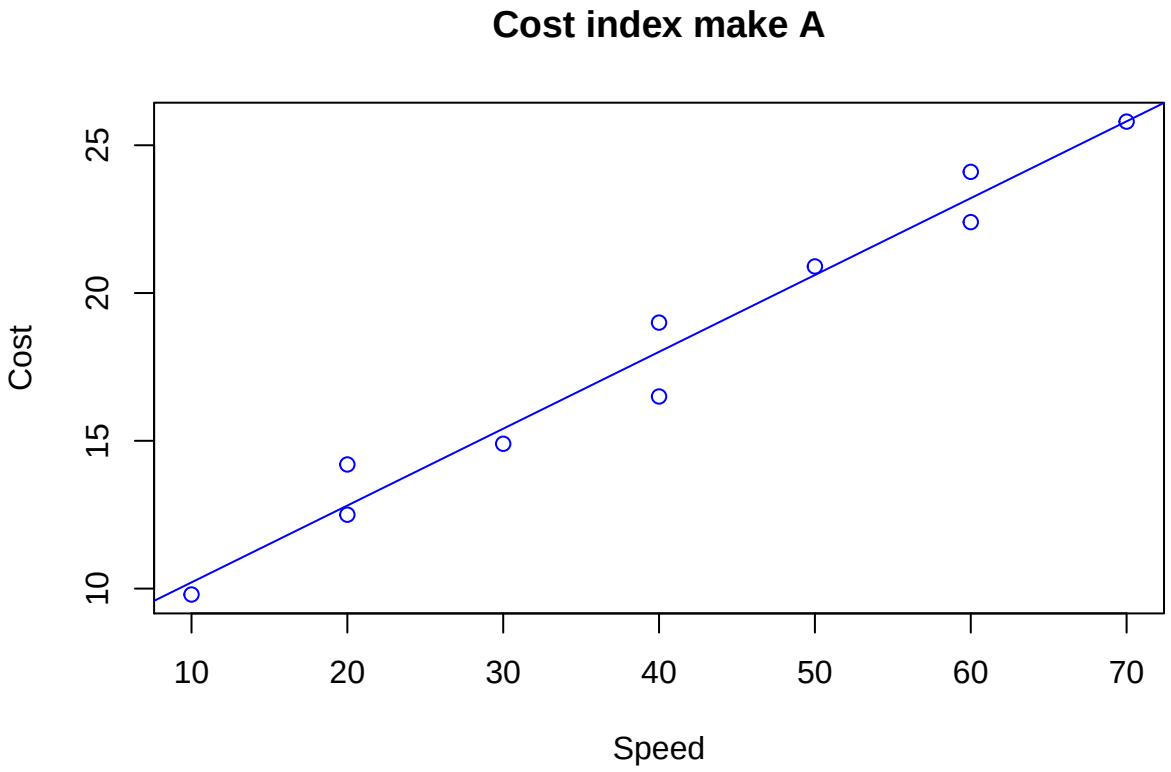
a) Gör ett lämpligt diagram över data.

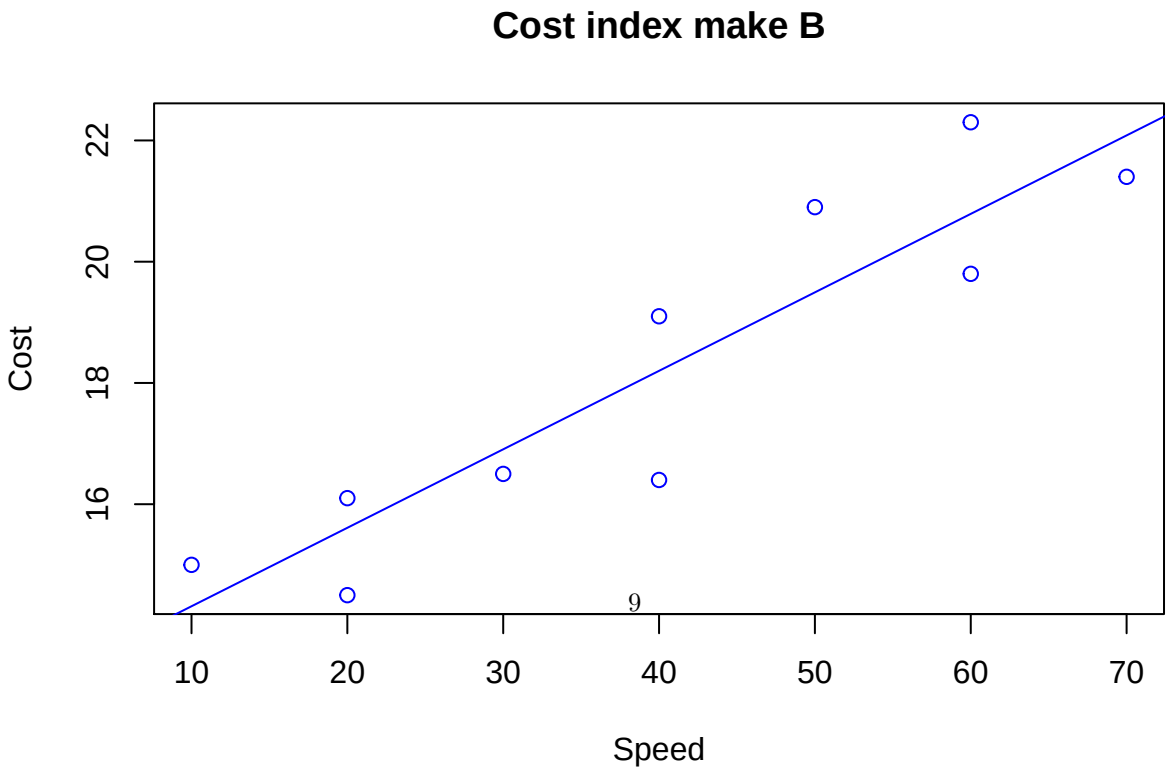
```
df$Type <-as.factor(df$Type)
plot(df$Speed~df$Cost, col=c("black", "red"),ylab="Speed",xlab="Cost")
legend("right", inset=.02, title="Däck typ",
   c("A","B"), fill=c("black", "red"), horiz=FALSE, cex=0.8)
```



I diagrammet ovan syns spridningen för de två däcktyperna, det syns tydligare hur typ B har en bättre och jämnare spridning till skillnad från typ A.

b) Anpassa en rät linje till data separat för var och en av tillverkarna.Avgör från dessa anpassningar om residualvariansen kan anses vara ungefär lika stor för de båda tillverkarna. Motivera varför vi skulle vilja veta detta.





I figurerna ovan visas residualerna samt en anpassad regressionslinje för de båda däcktyperna Det syns tydligt hur regressionslinjen anpassat sig bättre till residualerna för däck typ A, till skillnad från B. Trots att residualerna skiljer sig mellan däcktyperna, dras slutsatsen om att de har liknande residualspridning. Varför vi skulle vilja veta detta beror på att de avgör huruvida modellantaganden uppfylls eller ej, i form av att vi kan se hur de oberoende mätningarna har en lik varians eller ej. I detta fall är inte variansen mellan Däck typ A och Däck typ B helt lik men går att anta vara snarlik.

Visa skattade regressionsekvationer för de två modellerna.

Regressions ekvation för däck A.

$$\hat{y} = b_0 + b_1 x_1 = 7.61 + 0.26 x_1$$

Regressions ekvation för däck B.

$$\hat{y} = b_0 + b_1 x_1 = 13.022 + 0.129 x_1$$

Genom att skapa detta diagram kan jämförelser av hur modellerna ter sig ses och analyseras samtidigt som att man kan se hur väl linjerna är anpassade sina till residualer.

c) Ange en (1) regressionsmodell, med indikatorvariabel, som resulterar i en rät linje för var och en av tillverkarna. Skriv upp ekvationen för modellen, och förklara de olika delarna.

```
##
## Call:
## lm(formula = Cost ~ Speed * Type, data = df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -1.800 -0.715 -0.160  0.892  1.511
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   7.6100     0.8063    9.44  6.1e-08 ***
## Speed         0.2600     0.0182   14.28  1.6e-10 ***
## TypeB         5.4122     1.1403    4.75  0.00022 ***
## Speed:TypeB  -0.1306     0.0258   -5.07  0.00011 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.09 on 16 degrees of freedom
## Multiple R-squared:  0.941,  Adjusted R-squared:  0.93
## F-statistic: 84.8 on 3 and 16 DF,  p-value: 4.88e-10
```

$$\hat{y} = 7.61 + 0.26x_1 + 5.4122x_2 - 0.1306x_3$$

Ovan visas ekvationen för regressionsmodellen som skapats, ekvationen består av fyra parametrar. De två första parametrarna är precis samma parametrar som skulle finnas i en linjär modell med enbart däck A, skillnaden i detta fall är dock att de två sista parametrarna används för att justera värdet ifall man vill mäta kostnad för däck B. I fall av mätning för däck typ B, multipliceras x_2 med 1 och både x_1 samt x_3 med hastigheten. Detta för att ändra startvärdet samtidigt som att kostnaden inte ökar lika drastiskt för däck typ B.

d) Skatta parametrarna i modellen från deluppgift c) Överensstämmer parameterskattningarna med de från deluppgift b)?

```
##
## Call:
## lm(formula = Cost ~ Speed * Type, data = df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -1.800 -0.715 -0.160  0.892  1.511
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   7.6100     0.8063    9.44  6.1e-08 ***
## Speed         0.2600     0.0182   14.28  1.6e-10 ***
## TypeB         5.4122     1.1403    4.75  0.00022 ***
## Speed:TypeB  -0.1306     0.0258   -5.07  0.00011 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.09 on 16 degrees of freedom
## Multiple R-squared:  0.941,  Adjusted R-squared:  0.93
## F-statistic: 84.8 on 3 and 16 DF,  p-value: 4.88e-10
```

Regressionsekvationer för de olika däcktyperna

Däck A

$$\hat{y} = 7.61 + 0.26 + 5.4122 * 0 - 0.1306 * 0 = 7.61 + 0.26x_1$$

Regressions ekvation för däck A från b)

$$\hat{y} = b_0 + b_1 x_1 = 7.61 + 0.26x_1$$

Däck B

$$\hat{y} = 7.61 + 0.26 + 5.4122 * 1 - 0.1306 * 1 = 13.022 + 0.1294x_1$$

Regressions ekvation för däck B från b)

$$\hat{y} = b_0 + b_1 x_1 = 13.022 + 0.129x_1$$

Ja, parameterskattningar i detta fall stämmer överens med de från deluppgift b.

e) Avgör om sambanden mellan kostnad och hastighet är lika för de båda tillverkarna. Ange hypoteser och testvariabel. Visa beräkningar.

H0: B2=B3=0

H1: B2=B3 $\neq$ 0

```
lm_a <-lm(Cost~Speed,data=df)
model_c <-lm(Cost~Speed*Type,data=df)
anova(lm_a, model_c)
```

```
## Analysis of Variance Table
##
## Model 1: Cost ~ Speed
## Model 2: Cost ~ Speed * Type
##   Res.Df  RSS Df Sum of Sq    F  Pr(>F)
## 1     18 50.0
## 2     16 19.1  2      30.9 12.9 0.00046 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$F = \left(\frac{\text{SSE}(R) - \text{SSE}\,(F')}{df_R - df_F}\right) \div \left(\frac{\text{SSE}(F)}{df_F}\right) = \frac{49.969 - 19.108}{18 - 16} \div \frac{19.108}{16} = 12.92$$

I fall av teststatistiskan överstiger det kritiska värdet förkastar vi H_0 om att det inte råder någon statistiskt signifikant skillnad i samband mellan kostnad och hastighet för tillverkarna.

$$F = (0.95, 2, 16) = 3.634$$

I detta fall då $(12.92 > 3.634)$ samtidigt som ett signifikant p-värde på (0.0004) returneras, kan vi med 95 % konfidens förkasta H_0 om att det inte finns några skillnader i medelvärden mellan systemen. Alltså råder det statistiskt signifikanta skillnader i samband mellan kostnad och hastighet för tillverkarna.

f) Om slutsatsen i deluppgift e) är att de inte är lika, avgör vad det är som skiljer sambanden åt (lutning och/eller intercept). Ange hypoteser och testvariabler. Visa beräkningar.

I denna uppgift avser jag att undersöka vad som skiljer sambanden däcktyperna emellan för att göra det genomförs två tester för att se ifall de skiljer sig däcktyperna emellan. Signifikansnivån som detta genomförs på är 5 %.

```
##
## Call:
## lm(formula = Cost ~ Speed * Type, data = df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -1.800 -0.715 -0.160  0.892  1.511
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   7.6100     0.8063    9.44  6.1e-08 ***
## Speed         0.2600     0.0182   14.28  1.6e-10 ***
## TypeB         5.4122     1.1403    4.75  0.00022 ***
## Speed:TypeB  -0.1306     0.0258   -5.07  0.00011 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.09 on 16 degrees of freedom
## Multiple R-squared:  0.941,  Adjusted R-squared:  0.93
## F-statistic: 84.8 on 3 and 16 DF,  p-value: 4.88e-10
```

$$H0 : B2 = 0$$

$$H1 : B2 \neq 0$$

$$T_1 = \frac{\beta_2}{s\beta_2} \Rightarrow \frac{5.41222}{1.1404} = 4.7458$$

Ifall av att värdet från T-testet överstiger det kritiska värdet kan vi förkasta H0 om att lutningen för sambanden är lika.

$$T_{krit} = qt(1 - (0.05/2), 16) = 2.12$$

I detta fall då ($4.7458 > 2.12$) samtidigt som att vi får ett p-värde som summeras till (0.00022) kan vi med 95 % konfidens förkasta H_0 om att det inte finns några skillnader i sambanden. Alltså finns de skillnader i lutning för sambanden.

$$\text{H0: B3} = 0$$

$$H1 : B3 \neq 0$$

$$T_2 = \frac{\beta_3}{s\beta_3} \Rightarrow \frac{-0.1306}{0.02576} = -5.0686$$

I fall av att värdet från T-testet understiger det nedre kritiska värdet samtidigt som vi har ett p-värde som summerar till(0.0001) kan vi med 95 % konfidens förkasta H0 om att sambanden är lika. Alltså finns de statistiskt signifikanta skillnader i lutning för sambanden på 5 % signifikansnivå.

$$T_{krit} = qt((0.05/2), 16) = -2.12$$

I detta fall då ($-5.0686 < -2.12$) samtidigt som att vi får ett p-värde som summeras till (0.0001) kan vi med 95 % konfidens förkasta H_0 om att det inte finns några skillnader i lutning mellan däcktyperna.

I samband med att H0 förkastades i båda prövningarna dras slutsatsen om att lutningen är de som skiljer de båda sambanden åt.

g) Bilda ett 95%-igt konfidensintervall för parametern som gör att det blir olika lutning för de båda märkena. Visa beräkningar.

95 % konfidensintervall för paramtern Speed:TypeB

```
##
## Call:
## lm(formula = Cost ~ Speed * Type, data = df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -1.800 -0.715 -0.160  0.892  1.511
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   7.6100     0.8063    9.44  6.1e-08 ***
## Speed         0.2600     0.0182   14.28  1.6e-10 ***
## TypeB         5.4122     1.1403    4.75  0.00022 ***
## Speed:TypeB  -0.1306     0.0258   -5.07  0.00011 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.09 on 16 degrees of freedom
## Multiple R-squared:  0.941,  Adjusted R-squared:  0.93
## F-statistic: 84.8 on 3 and 16 DF,  p-value: 4.88e-10
```

$$T_{krit} = qt(1 - (0.05/2), 16) = 2.12$$

$$-0.1305 \pm 2.12 \cdot 0.02576 = -0.1852 : -0.076$$

Alltså kommer indikatorvariabeln för lutningen (Speed:TypeB) att i ett 95 % konfidensintervall ligga mellan -0.1852 och -0.076.

h) Ge en uppskattning av skillnaden i kostnadsindex mellan de båda tillverkarna vid hastigheten 20.

Regressionsekvation med indikatorvariabel för däck typ B.

$$\hat{y} = 7.61 + 0.26 \cdot 20 + 5.41222 \cdot 1 - 0.13056 \cdot 20 = 15.611$$

Regressionsekvation med indikatorvariabel för däck typ A.

$$\hat{y} = 7.61 + 0.26 \cdot 20 + 5.41222 \cdot 0 - 0.13056 \cdot 0 = 12.81$$

Uppskattning av skillnad i kostnadsindex vid hastigheten 20 mellan däcktyperna.

$$D_{f-r} = 15.611 - 12.81 = 2.801$$

Alltså kommer det att skilja 2.801 kostnadsenheter mellan tillverkarna i 20 hastighetsenheter.

Uppgift 3

Ange en lämplig modell och förutsättningarna som behövs för en statistiskanalys. Analysera data och kontrollera förutsättningarna. Finns det signifikantaskillnader mellan behandlingarna

I detta fall har jag tänkt använda mig av Flerfaktorsförsök (Randomiserade block)

Modell: $y_{ij} = \mu + \alpha_i + b_j + e_{ij} \quad i = 1, 2, \ldots, a$

y_{ij} vara en observation för den i:te behandlingen och det j:e blocket

μ gemensamma medelvärdet(grand mean)

α_i effekt av behandling $i \quad i = 1, 2, \ldots, a$

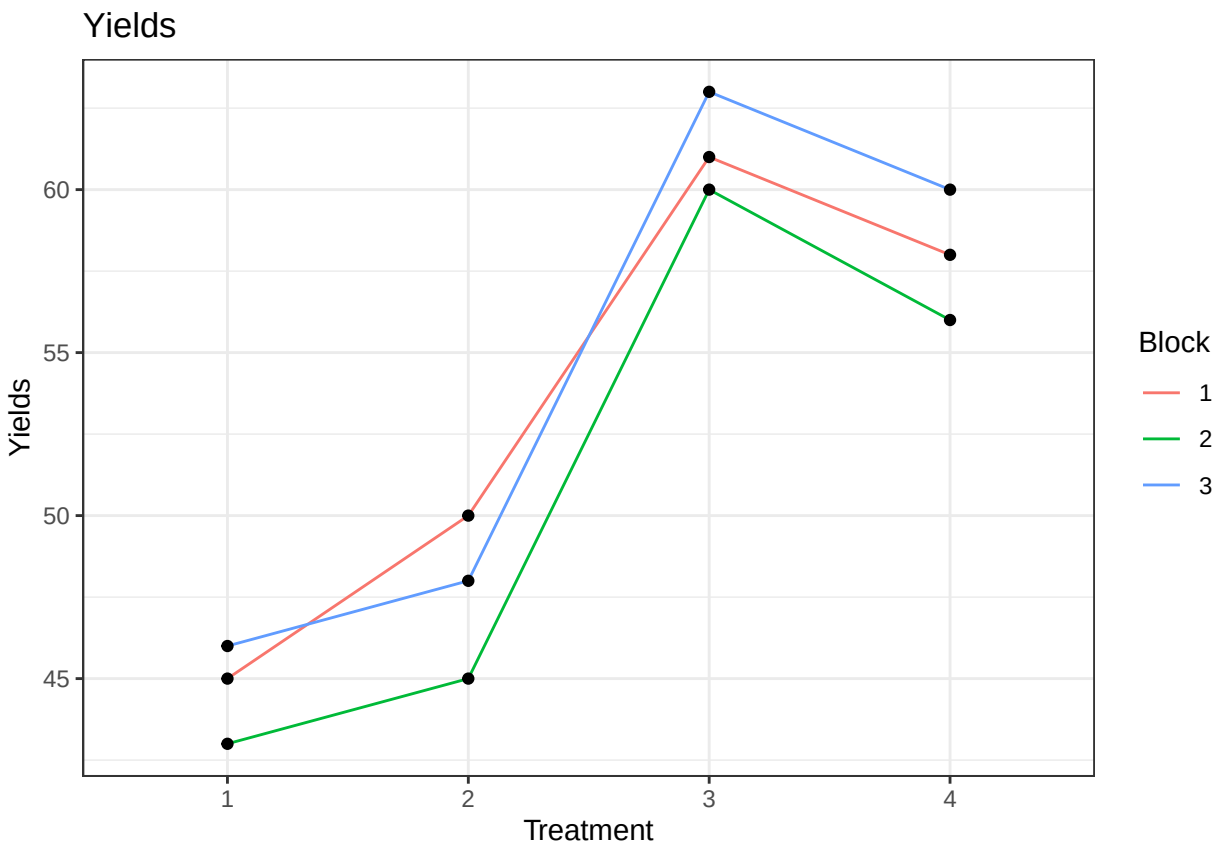
b_j effekt av block$j \quad j = 1, 2, \ldots, b$

e_{ij} slumpfel

Förutsättningar:$e_{ij} = NID\left(0, \sigma_e^2\right)$

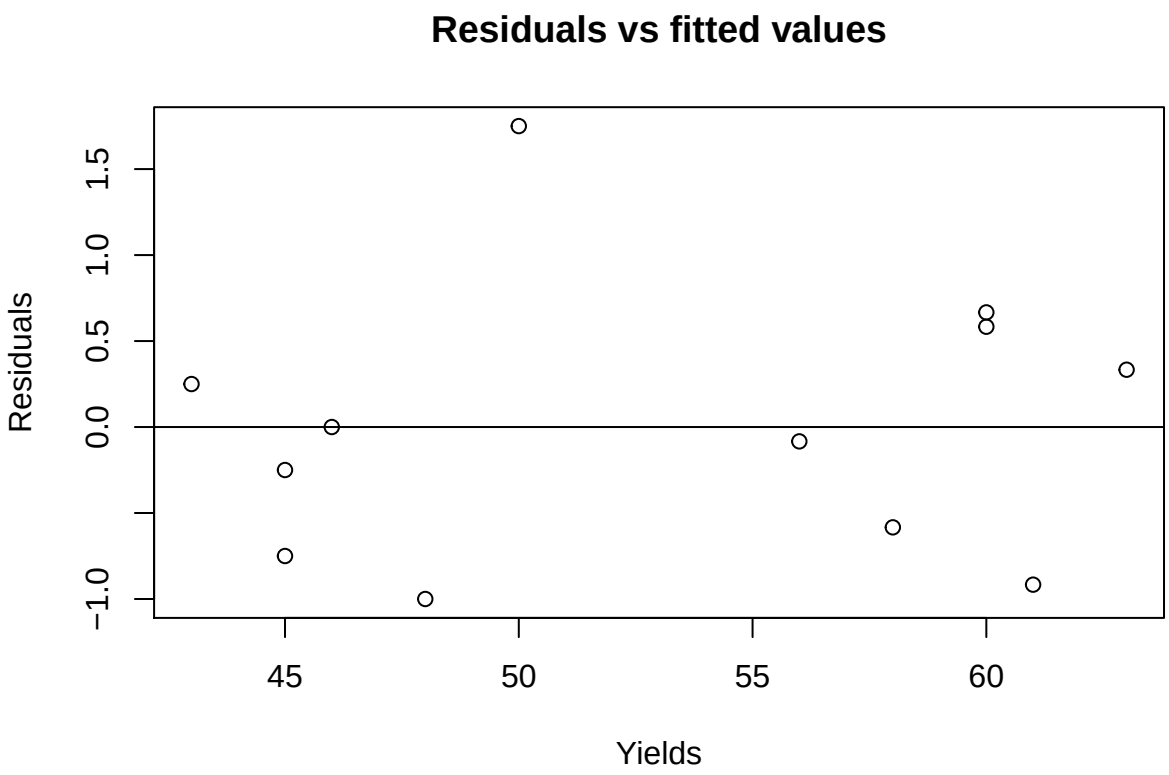
Förutsättningar:Alla behandlingar ska genomföras under lika förutsättningar och blocken ska vara randomiserade.

```
ggplot(data=Cowpeas, aes(x=Treatment, y=Yields,group=Block)) +
geom_line(aes(color=Block))+
geom_point() +
labs(title = 'Yields', x = 'Treatment', y = 'Yields')+
theme_bw()
```



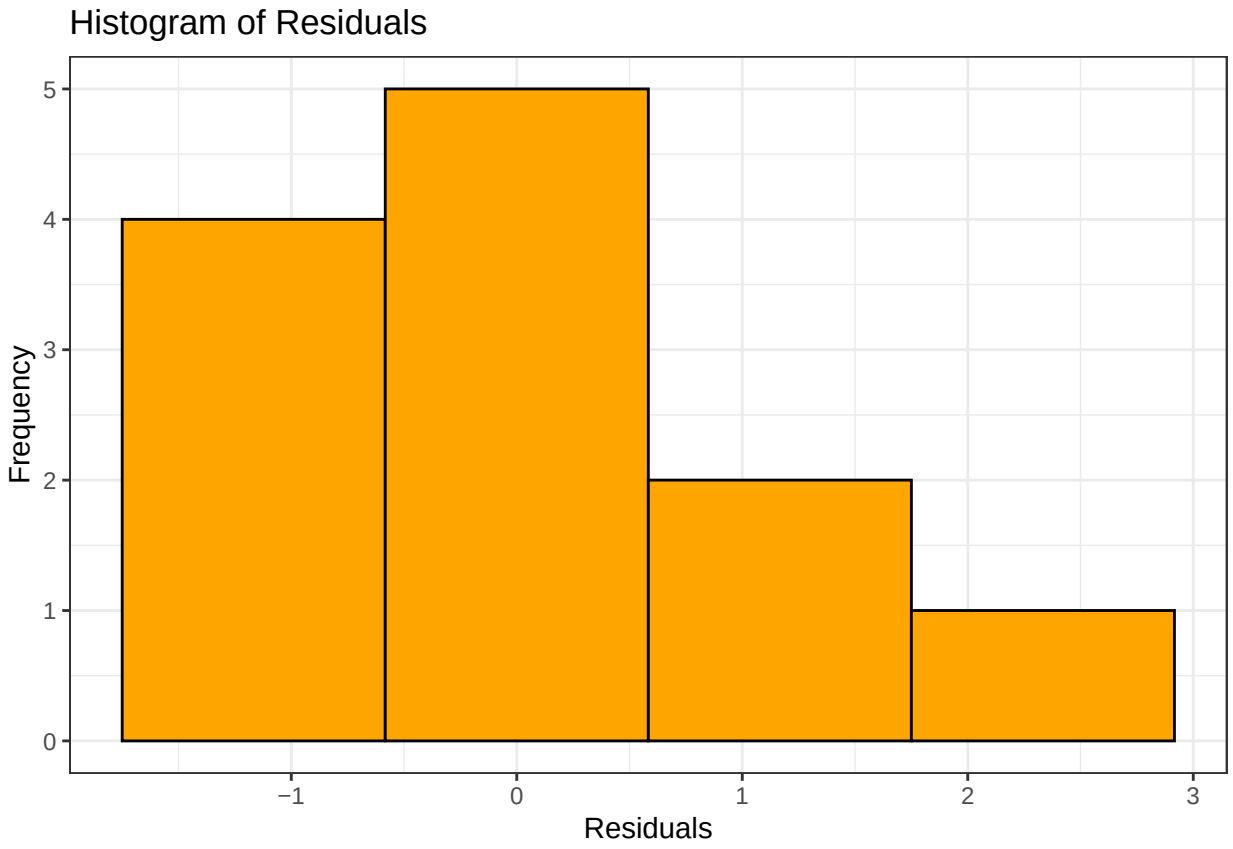
I diagrammet visas de olika behandlingar gentemot blocken, utifrån grafen syns det hur Block 3 som kännetecknar land typ 3 är den typen av land som ger bäst avkastning i 3 av 4 fall. Land 2 är landtypen som ger sämst avkastning oberoende av behandling. Vad gäller behandlingar är behandling 3 den behandling som får bäst avkastning oberoende av block, den behandling som får sämst avkastning är behandling 1.

```r
Cowpeas_c <-aov(Yields~Block+Treatment,data = Cowpeas)
 plot(Cowpeas$Yields, Cowpeas_c$residuals,
    ylab="Residuals", xlab="Yields",
    main="Residuals vs fitted values")
 abline(0, 0)
```



Spridningen av residualernas anses se bra ut, det förekommer en utomliggande observation i det mittersta hörnet, bortsett från de anses fördelning vara någorlunda fin och jämn.

```
ggplot(data = Cowpeas_c, aes(x = Cowpeas_c$residuals)) +
    geom_histogram(fill = 'orange', color = 'black',binwidth = 1.1666) +
    labs(title = 'Histogram of Residuals', x = 'Residuals', y = 'Frequency')+
  theme_bw()
```



I histogrammet syns residualerna för det randomiserade blocket, fördelningen anses vara aningen positvt skev men datasetet än användbart.

Pröva om det finns någon skillnad mellan Behandlingarna

```
Cowpeas_c <-aov(Yields~Treatment+Block,data = Cowpeas)
anova(Cowpeas_c)
```

```
## Analysis of Variance Table
##
## Response: Yields
##           Df Sum Sq Mean Sq F value  Pr(>F)
## Treatment  3    577   192.3   168.8 3.5e-06 ***
## Block      2     23    11.6    10.2   0.012 *
## Residuals  6      7     1.1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$H0 : \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$$

$$H1 : \tau_1 \neq \tau_2 \neq \tau_3 \neq \tau_4 = 0$$

F-test för variabeln Treatment

$$F = (192.305556/1.1380 = 168.854)$$

I fall av att teststatistikan överstiger det kritiska värdet ($F > 4.757$) förkastas $ H_{0} $ om att alla medelvärden för behandlingarna är lika.

$$F(.95, 3, 6) = 4.757$$

I detta fall då ($168.854 > 4.757$) samtidigt som att vi får ett p-värde som summeras till ($3.4934e - 065$) kan vi med 95 % konfidens förkasta H_0 om att det inte finns några skillnader i medelvärden mellan behandlingar.

Finns det signifikanta skillnader mellan behandlingarna?

```
coef <-coef(aov(Yields~Treatment,data=Cowpeas))
Means <-coef+c(0, 44.67, 44.67,44.67)
Means
```

```
## (Intercept)  Treatment2  Treatment3  Treatment4
##        44.7        47.7        61.3        58.0
```

```
TukeyHSD(Cowpeas_r,conf.level = 0.95)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Yields ~ Treatment, data = Cowpeas)
##
## $Treatment
##      diff   lwr   upr p adj
## 2-1  3.00 -2.06  8.06 0.301
## 3-1 16.67 11.60 21.73 0.000
## 4-1 13.33  8.27 18.40 0.000
## 3-2 13.67  8.60 18.73 0.000
## 4-2 10.33  5.27 15.40 0.001
## 4-3 -3.33 -8.40  1.73 0.229
```

I detta fall där enbart behandlingar och inte Block studeras, kan vi tillsammans med p-värdena se att de är alla parvisa jämförelser mellan behandlingarna förutom den första och den sista som är statistiskt signifikant skillda sett till p-värde på 5 % signifikansnivå.

Lärdomar, problem, övriga kommentarer

Laborationen har varit extremt givande, att få ta egna beslut samtidigt som man återigen brottas med latex och att hitta rätt formler. Användningen av funktioner och metoder börjar komma väldigt naturligt.