

Labbrapport i Statistik

Laboration 14

732G46

Mattias Hällgren, Michael Debebe

Avdelningen för Statistik och maskininlärning
Institutionen för datavetenskap
Linköpings universitet

2021-12-13

Innehåll

Introduktion	1
Databehandling	2
Uppgifter	3
Uppgift 1 Analys av ett obalanserat tvåvägsförsk	3
a) Ta fram beskrivande statistik (minst medelvärde och standardavvikelse) för Sales price grupperat på Pool och Quality (alltså på båda variablerna samtidigt). Kommentera resultatet. Har vi en obalanserad datamängd?	3
Ta fram korstabell med frekvenser för Pool och Quality.	4
b) Skapa en interaktions-plot. Kommentera	5
c) Skriv upp den fulla tvåvägsmodellen (förklara dess delar), skatta och analysera den sedan. Kan modellen reduceras till endast huvudeffekter på 10% nivå? Stämmer ditt resultat överens med det du kom fram till i b)?	6
d) Utför analysen av den reducerade modellen.	8
e) Är den slutliga reducerade modellen bra?	9
f) Fullfölj analysen i den valda modellen (fulla eller reducerade) med lämpliga multipla jämförelser. Tolka resultatet. en slutliga reducerade modellen bra?	10
Uppgift 2 Analys av ett obalanserat tvåvägsförsk	11
a) Ta fram beskrivande statistik (minst medelvärde och standardavvikelse) grupperat på alla de förklarande variablerna. Kommentera. Ser du några problem?	11
b) Skapa en interaktions-plot. Kommentera	12
c) Skatta den fullständiga trefaktormodellen med hjälp av aov(). Kan modellen reduceras? Titta på treinteraktionen och därefter tvåinteraktionerna. Stämmer ditt resultat överens med det du kom fram till i b)?	13
d) Presentera din valda reducerade modell.	15
e) Analysera den reducerade modellen. Är den bra?	16
f) 1. Använd den reducerade modellen för att lösa övning 24.14, se figur 3	17

Introduktion

I denna laboration kommer två dataset analyseras.

Det första datasetet kommer från en stadsskatttjänsteman och syftar till att förutspå huspriser i en medelvästern stad tack vare en funktion av varierande karaktärsdrag för hemmen. Datat består av 522 observationer

I det andra datasetet har det sammanställts data om produktivitetsförbättringar det senaste året för ett urval av företag, där de tillverkar elektronisk datautrustning. Resultat av bevisen mäts på en skala 0 till 100.

Målet med denna laboration är att lära sig utvidga tidigare modeller till tre eller fler faktorer och analysera resultatet med rätt program. Även analysera obalanserade försök med hjälp av R

Databehandling

```
Real_eS <-read.table("https://raw.githubusercontent.com/MatttiasH/Labbar/main/Real_estate_sales")
cols1 <-c("ID", "Price", "x1", "x2", "x3", "x4", "x5", "Pool", "x7", "Quality", "x9", "x10", "x11")
colnames(Real_eS) <-cols1
factores <-c("Pool", "Quality")
Real_eS[factores] <- lapply(Real_eS[factores], factor)

Electronic_assembly <-read.table("https://raw.githubusercontent.com/MatttiasH/Labbar/main/Labb%2014%20up")
cols2 <-c("y", "Gender", "Sequency", "Experience", "observation")
colnames(Electronic_assembly) <-cols2
factos <-c("Gender", "Sequency", "Experience", "observation")
Electronic_assembly[factos] <- lapply(Electronic_assembly[factos], factor)
```

Uppgifter

Uppgift 1 Analys av ett obalanserat tvåvägsförsk

a) Ta fram beskrivande statistik (minst medelvärde och standardavvikelse) för Sales price grupperat på Pool och Quality (alltså på båda variablerna samtidigt). Kommentera resultatet. Har vi en obalanserad datamängd?

```
##      Pool Quality Mean_Sales_price Sd_Sales_price
## [1,] 1      1      544782.9      130489.3
## [2,] 2      1      535925.6      72907.93
## [3,] 1      2      270825.5      85078.41
## [4,] 2      2      309590.9      81902.3
## [5,] 1      3      173968.6      37826.14
## [6,] 2      3      208400      25131.75
```

Ovan visas beskrivande statistik för priset på fastigheter beroende på kombination av kvalitet på husets konstruktion och huruvida huset har en pool eller inte. För variabeln Pool kännetecknar värde (2) att fastigheten har pool och värde (1) att fastigheten inte har pool. Kvalitet beskrivs i ett intervall från 1 till 3 där 1 kännetecknar hög kvalite, 2 medel och 3 låg kvalite på husets konstruktion.

Från tabellen ovan syns det hur fastigheter med hög kvalite på konstruktion utan pool har ett väldigt högt medelvärde vad gäller pris, däremot är standardavvikelsen betydligt mycket högre medan hus med hög kvalite med pool däremot inte har så stor skillnad i medelpris men med en nästan hälften så stor standardavvikelse. Detta kan innebära att en viss interaktion kan komma att synas i interaktionsplotten. Generellt oberoende av kombination visas en tydlig minskning i standardavvikelse för fastigheter med pool, samtidigt som att de har ett högre genomsnittligt pris.

Ta fram korstabell med frekvenser för Pool och Quality.

I korstabellen nedan syns fördelningen mellan frekvenser för Pool och quality, Pool (x6), där kodningen är skedd på så sätt att fastigheter utan pool kodas till 0 och fastigheter med pool kodas till 1. För kvalite på fastigheternas konstruktion, kodas dom i ett intervall med 1 och 3, där 1 känntecknar hög kvalite, 2 medelhög kvalite och 3 låg kvalité. en fastighet.

```
crosstab(Real_eS, row.vars = "Pool", col.vars = "Quality", type = "frequency", style = "long")
```

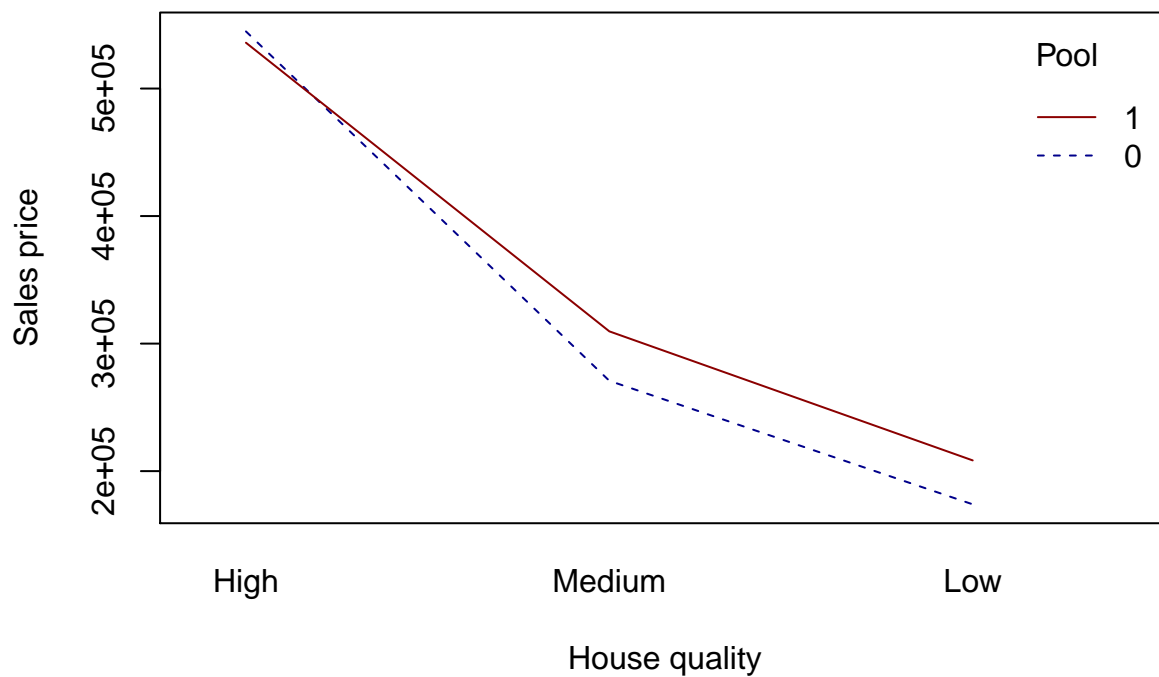
```
##      Quality    1    2    3 Sum
## Pool
## 0           59 268 159 486
## 1             9  22   5  36
## Sum           68 290 164 522
```

Främst syns den lilla proportionen 6.9 % fastigheter som har pool, bland dessa fastigheter består 25 % av fastigheter med hög kvalite av konstruktion, 61 % av fastigheter med medelhög konstruktion och drygt 14 % av låg kvalitetskonstruktion. Sett till andel av hus med pool kopplat till kvalite är pool som vanligast bland fastigheter med hög kvalite, följt av medel och därefter låg.

Fördelningen kan ses som allt annat än balanserad då antalet av hus med högkvalitativ konstruktion har dubbelt så hög proportion av hus med pool till skillnad från hus med låg konstruktion, trots att husen med låg konstruktion är trippelt så många.

b) Skapa en interaktions-plot. Kommentera

```
levels(Real_eS$Quality) =c("High", "Medium","Low")
interaction.plot(x.factor = Real_eS$Quality,
  trace.factor = Real_eS$Pool, response = Real_eS$Price,
  fun = mean, type = "l",
  ylab = "Sales price", col = c("blue4", "red4"),
  xlab = "House quality", trace.label = "Pool", xpd = TRUE)
```



I diagrammet ovan syns tydligt hur fastigheter med pool vanligtvis har ett högre pris gentemot fastigheter utan pool oberoende av kvalite på fastighetens konstruktion. I samband med att kvaliteten på fastigheten sjunker, sjunker också priset för fastigheten. Det syns en liten interaktion mellan fastigheter utan samt med i höjd med fastighetspris 500 000, detta är däremot den enda interaktionen som syns

c) Skriv upp den fulla tvåvägsmodellen (förklara dess delar), skatta och analysera den sedan. Kan modellen reduceras till endast huvudeffekter på 10% nivå? Stämmer ditt resultat överens med det du kom fram till i b)?

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

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

μ gemensamma medelvärde (grand mean)

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

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

e_{ij} slumpfel

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

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

```
## Anova Table (Type III tests)
##
## Response: Price
##           Sum Sq Df  F value Pr(>F)
## (Intercept) 1.7511e+13 1 2711.5370 <2e-16 ***
## Quality      5.9185e+12 2  458.2479 <2e-16 ***
## Pool         6.1263e+08 1    0.0949 0.7582
## Quality:Pool 1.3122e+10 2    1.0160 0.3628
## Residuals    3.3322e+12 516
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

I den fulla modellen som visas ovan är varken huvudeffekten Pool samt interaktionstermen Pool:Quality statistiskt signifikanta variabler på 10 % signifikansnivå sett till p-värden. Däremot är det inte lämpligt att reducera modellen med en variabel och därav reduceras enbart modellen med interaktionstermen.

$$H_0 : (\alpha\beta)_{ij} = 0$$

$$H_1 : (\alpha\beta)_{ij} \neq 0$$

$$F_{Test}(RC) = \frac{MSRC}{MSE} = 1.0160$$

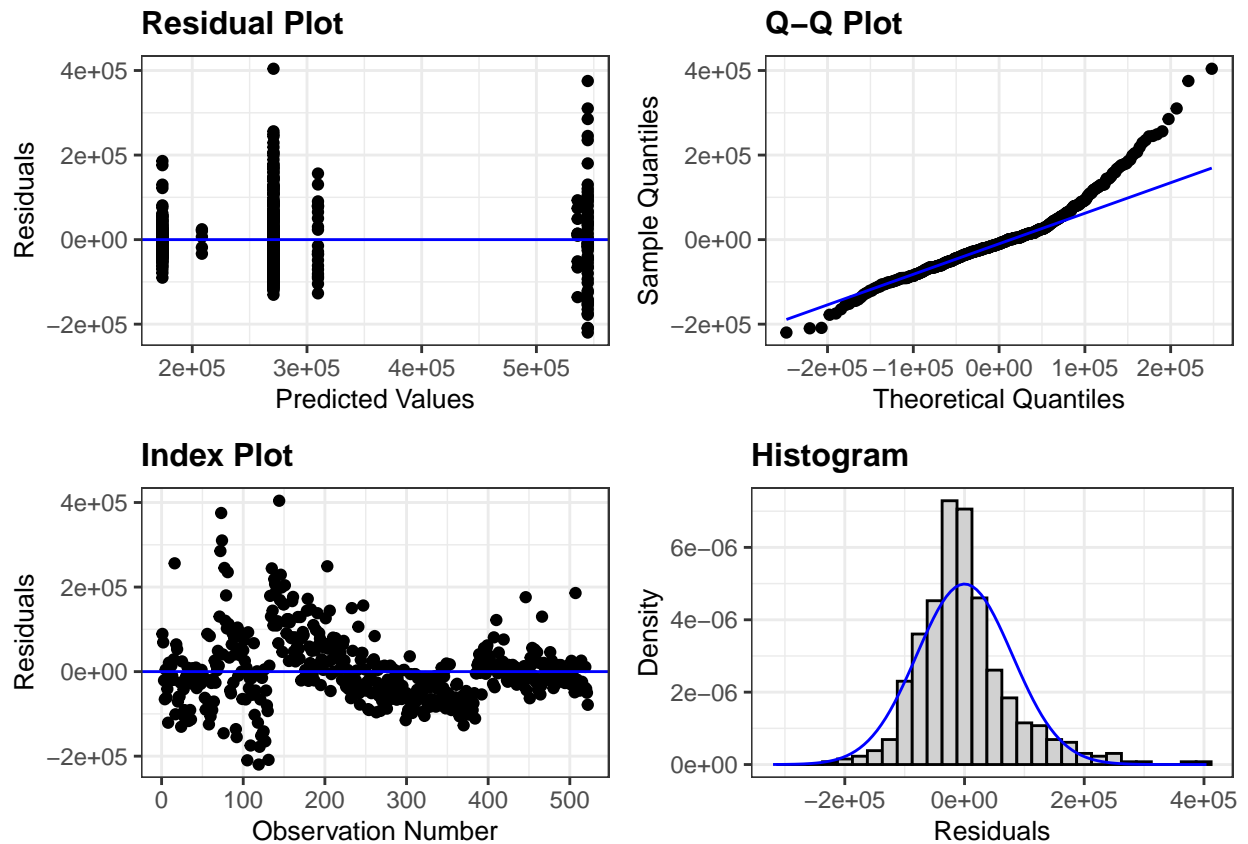
$$F = (0.95, (2 - 1)(3 - 1), 516) = 2.313$$

Ifall av att värdet från F-testet $F > 2.313$ överstiger det kritiska värdet kan vi med 95 % konfidens förkasta H_0 om att det inte råder någon interaktionseffekt mellan effekterna.

I detta fall där teststatistikan inte överstiger det kritiska värdet ($1.0160 < 2.313$) antas H_0 om att det inte råder någon interaktionseffekt mellan effekterna. P-värdet från testet summerar till 0.3628 vilket inte heller är signifikant på 5 % signifikansnivå.

Slutsatsen blir att interaktionstermen inte följer med till den reducerade modellen.


```
ggResidpanel::resid_auxpanel(residuals = residuals(mod.full),
                             predicted = fitted(mod.full))
```



Vad gäller residualanalysen som visas ovan är den enda plotten som visar normalfördelade residualer histogrammet, vad gäller indexplotten är residualerna långt ifrån jämnt fördelade och koncentrerade i den lägre halvan av diagrammet, q-qplotten visar ett krökt samband där det ett stort antal residualer inte ligger på linjen. Vad gäller residualplotten är variansen inom de olika grupperna inte heller jämn. Slutsatsen blir att modellen måste reduceras och att residualerna tydligt visar att modellen inte är tillförlig.

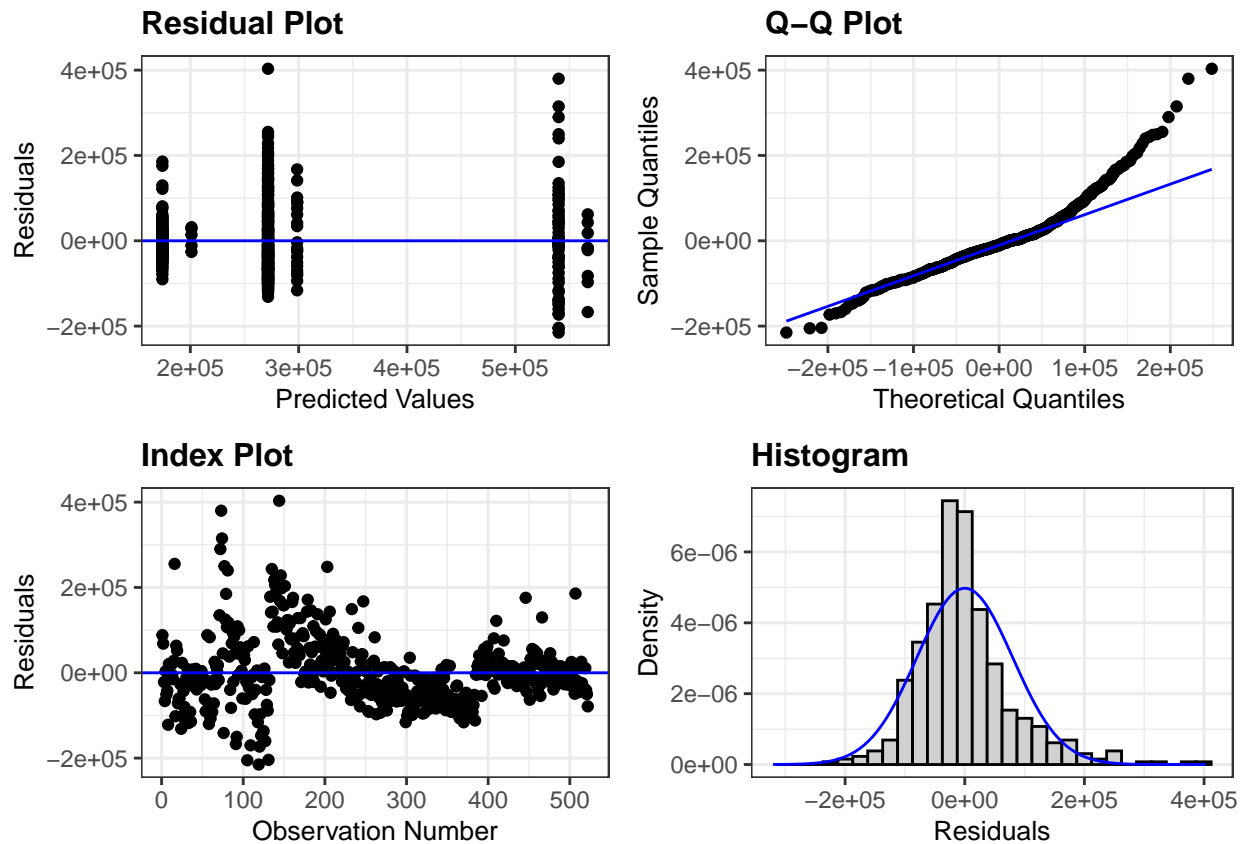
d) Utför analysen av den reducerade modellen.

```
mod.red <- aov(Price ~Pool+Quality,data = Real_eS)
anova(mod.red)

## Analysis of Variance Table
##
## Response: Price
##          Df      Sum Sq   Mean Sq F value    Pr(>F)
## Pool       1 2.1303e+11 2.1303e+11   32.987 1.583e-08 ***
## Quality    2 6.3525e+12 3.1763e+12 491.821 < 2.2e-16 ***
## Residuals 518 3.3453e+12 6.4582e+09
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

I den reducerade modellen som visas ovan är samtliga variabler statistiskt signifikanta på 10 % signifikansnivå, vad gäller huvudeffekterna är kvaliteten på husets konstruktion en betydligt mer signifikant variabel i modellen till skillnad från huruvida huset har en pool eller ej.

```
ggResidpanel::resid_auxpanel(residuals = residuals(mod.red),
                             predicted = fitted(mod.red))
```



Residualerna är snarlika residualerna från den fulla modellen vilket talar negativt för den reducerade modellen, trots att interaktionstermen som hade ett så högt p-värde som (0.57) är borta kan fortfarande inte huvudeffekterna förklara variationen i responsvariabeln särskilt bra, slutsatsen blir att trots de signifikanta värdena i tidigare utskrift är fortfarande inte modellen tillförlig.

e) Är den slutliga reducerade modellen bra?

Sett till p-värden och teststatistikor är modellen bra och signifikant, vad gäller residualer och normalfördelning uppfyller däremot inte modellen antagandena som krävs, vilket innebär att modellen inte kan ses som bra.

f) Fullfölj analysen i den valda modellen (fulla eller reducerade) med lämpliga multipla jämförelser. Tolka resultatet, en slutliga reducerade modellen bra?

```
options(digits=7,scipen=7)
kk <-TukeyHSD(mod.red)
kk$Pool
```

```
##          diff      lwr      upr      p adj
## 1-0 79724.36 52454.37 106994.4 0.00000001606264
```

```
kk$Quality
```

```
##          diff      lwr      upr      p adj
## Medium-High -265340.67 -290791.2 -239890.10 0.0000000002299340
## Low-High     -360471.22 -387715.6 -333226.83 0.0000000002299340
## Low-Medium   -95130.55 -113585.7  -76675.45 0.0000000002299343
```

Formeln som används för parvisa jämförelser med Tukey-kramers metod

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

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

Det syns tydligt i hur det finns stora samt signifikanta skillnader i faktorermedelvärden för båda variablerna, främst i form av de extremt små p-värdena tillsammans med att inget intervall täcker 0. Den största skillnaden återfinns mellan fastigheter med låg konstruktionskvalite och fastigheter med hög konstruktionskvalite vilket är logiskt.

Uppgift 2 Analys av ett obalanserat tvåvägsförsk

a) Ta fram beskrivande statistik (minst medelvärde och standardavvikelse) grupperat på alla de förklarande variablerna. Kommentera. Ser du några problem?

```
f2 <-function(x){data.frame(Mean_Time_Assembly = mean(x),  
                             Sd_Time_Assembly=sd(x))  
}  
DS <-do.call("cbind", aggregate(y~Gender*Experience*Sequency, Electronic_assembly, FUN =f2))  
DS
```

```
##      Gender Experience Sequency Mean_Time_Assembly Sd_Time_Assemblye  
## [1,] 1      1      1      1218.6      32.63893  
## [2,] 2      1      1      1036.4      34.21695  
## [3,] 1      2      1      1051      41.57523  
## [4,] 2      2      1      870.6      16.7869  
## [5,] 1      1      2      1274.2      32.26763  
## [6,] 2      1      2      1077.4      37.09852  
## [7,] 1      2      2      1122.4      24.79516  
## [8,] 2      2      2      931.6      22.46775  
## [9,] 1      1      3      1218.2      24.28374  
## [10,] 2      1      3      1020.4      12.8569  
## [11,] 1      2      3      1051.2      23.09112  
## [12,] 2      2      3      860.4      34.99714
```

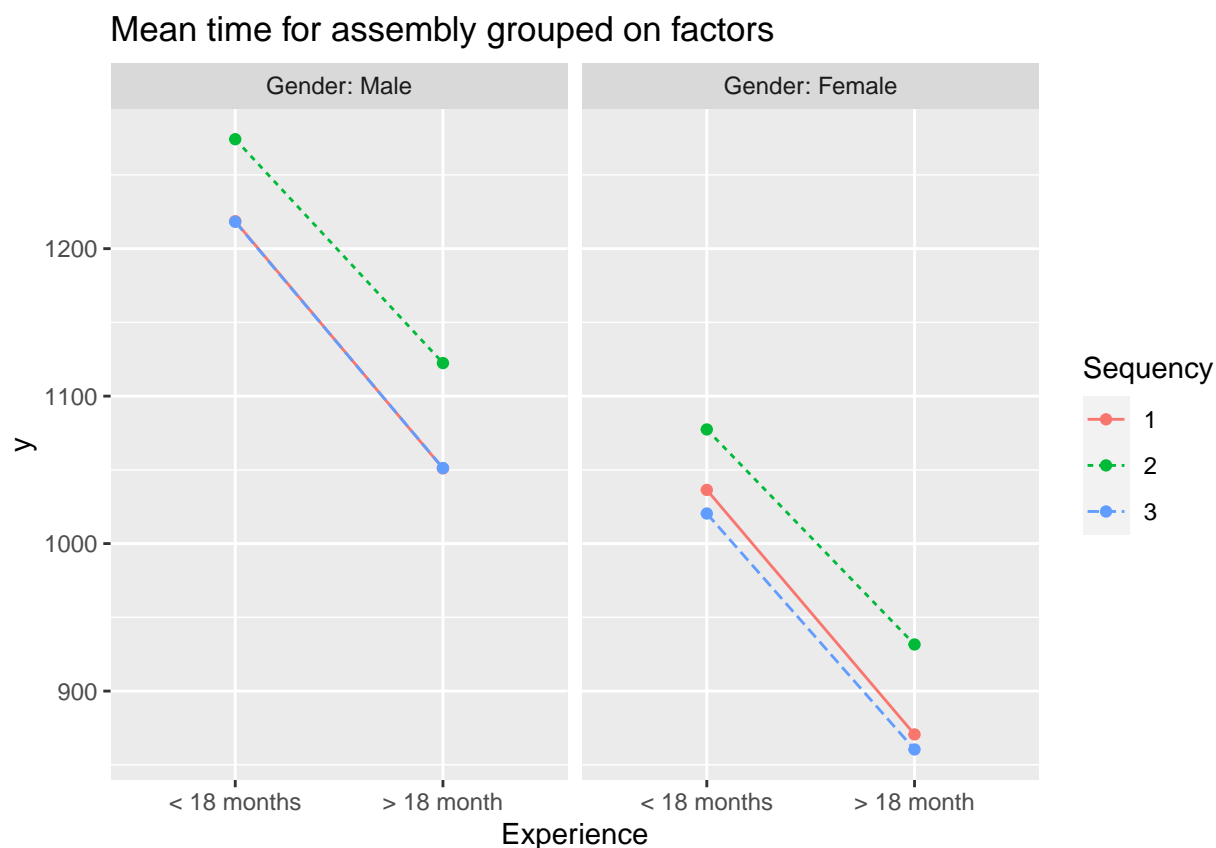
Tiden(y) för att göra 50 stycken brädor har ett medelvärde på 1058 minuter, standardavvikelsen är 131 minuter vilket säger att det är den varierande tiden för operatörerna att göra 50 brädor.

Som uppgiften visar så får man reda på att det är 15 män och 15 kvinnor, medelvärdet för kön ligger på 1.5 vilket är detsamma. Det här gäller även för Sequency, experience.

Problemet här för alla värden förutom tid är att det inte visar någon intressant fakta var för sig, utan denna data ska integreras med varandra för att få intressanta påverkningar och resultat.

b) Skapa en interaktions-plot. Kommentera

```
levels(Electronic_assembly$Gender) = c("Male", "Female")
levels(Electronic_assembly$Experience) = c("< 18 months", "> 18 month")
interaction.ABC.plot(y, x.factor=Experience,
                     groups.factor=Sequence, trace.factor=Gender,
                     data=Electronic_assembly,
                     title="Mean time for assembly grouped on factors")
```



Ovan visas en interaktionsplot av samtliga variabler i datasetet, plottarna är grupperade på kön, där den vänstra visar medeltiden kvinnor och den högra visar medeltiden för män. På x-axeln återfinns variabeln Experience (Experience) och är grupperade på mindre eller mer än 18 månaders erfarenhet. Slutligen visar linjerna de olika sekvenserna som behandlas.

I diagrammet syns två interaktioner, ena interaktionen syns mellan Sequence 1 och 3 för kvinnor i form av att oberoende av erfarenhet så är tid väldigt lik. Den andra är för män med mindre än 18 månaders erfarenhet där medeltiden för sekvens 1 och 3 återigen är lik.

Bortsett från interaktionerna tar de generellt sett mer tid för män till skillnad från kvinnor oberoende av sekvens att montera de elektroniska komponenterna. Det syns även hur erfarenhet påverkar den genomsnittliga tiden av montering i form av att de tar mindre tid för anställda med mer erfarenhet till skillnad från de med mindre.

c) Skatta den fullständiga trefaktormodellen med hjälp av `aov()`. Kan modellen reduceras? Titta på treinteraktionen och därefter tvåinteraktionerna. Stämmer ditt resultat överens med det du kom fram till i b)?

```
mod.f <-aov(y~(Gender+Experience+Sequency)^3,data = Electronic_assembly)
Anova(mod.f)
```

```
## Anova Table (Type II tests)
##
## Response: y
##
##              Sum Sq Df  F value    Pr(>F)
## Gender          540361  1 629.7603    < 2.2e-16 ***
## Experience       382402  1 445.6679    < 2.2e-16 ***
## Sequency         49320  2  28.7396 0.00000000622 ***
## Gender:Experience    91  1   0.1064    0.7457
## Gender:Sequency     543  2   0.3161    0.7305
## Experience:Sequency  911  2   0.5310    0.5914
## Gender:Experience:Sequency  19  2   0.0111    0.9890
## Residuals         41186 48
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

En modell där samtliga tre variabler och dess interaktioner har tagits med visas ovan, samtliga tre förklaringsvariabler är signifikanta sett till p och f-värde på 5 % signifikansnivå. Ingen av deras interaktionsvariabler (Varken två eller treinteraktionerna) är statistiskt signifikanta och har en påverkan på resultatet, därav bedöms inte vara med. Sett till plotten i b, är det tydligt hur det Sequency är den variabeln som är minst statistiskt signifikant.

Interaktion mellan Gender, Experience och Sequency

$$H_0 : (\alpha\beta\gamma)_{ijk} = 0$$

$$H_1 : (\alpha\beta\gamma)_{ijk} \neq 0$$

$$F_{Test} = \frac{MS_{ABC}}{MSE} \Rightarrow \frac{9.5}{858.042} = 0.0111$$

$$F = (0.95, (2-1)(2-1)(3-1), 48) = 3.190727$$

Ifall av att värdet från F-testet $F > 3.190727$ överstiger det kritiska värdet kan vi med 95 % konfidens förkasta H_0 om att det inte råder någon interaktionseffekt mellan effekterna Gender, Experience och Sequency.

I detta fall där teststatistikan inte överstiger det kritiska värdet ($0.0111 < 3.190727$) antas H_0 om att det inte råder någon interaktionseffekt mellan effekterna Gender, Experience och Sequency. P-värdet från testet summerar till 0.9890 vilket inte heller är signifikant på 5 % signifikansnivå.

Interaktion mellan Gender och Experience

$$H_0 : (\alpha\beta)_{ij} = 0$$

$$H_1 : (\alpha\beta)_{ij} \neq 0$$

$$F_{Test} = \frac{MSAB}{MSE} \Rightarrow \frac{91}{858.042} = 0.1064$$

$$F = (0.95, (2-1)(2-1)(3-1), 48) = 3.190727$$

Ifall av att värdet från F-testet $F > 3.190727$ överstiger det kritiska värdet kan vi med 95 % konfidens förkasta H_0 om att det inte råder någon interaktionseffekt mellan effekterna Gender och Experience.

I detta fall där teststatistikan inte överstiger det kritiska värdet ($0.1064 < 3.190727$) antas H_0 om att det inte råder någon interaktionseffekt mellan effekterna Gender och Experience. P-värdet från testet summerar till 0.7457 vilket inte heller är signifikant på 5 % signifikansnivå.

Interaktion mellan Gender och Sequency

$$H_0 : (\alpha\gamma)_{ik} = 0$$

$$H_1 : (\alpha\gamma)_{ik} \neq 0$$

$$F_{Test} = \frac{MSAC}{MSE} \Rightarrow \frac{271.250}{858.042} = 0.3161$$

$$F = (0.95, (2-1)(2-1)(3-1), 48) = 3.190727$$

Ifall av att värdet från F-testet $F > 3.190727$ överstiger det kritiska värdet kan vi med 95 % konfidens förkasta H_0 om att det inte råder någon interaktionseffekt mellan effekterna Gender och Sequency.

I detta fall där teststatistikan inte överstiger det kritiska värdet ($0.3161 < 3.190727$) antas H_0 om att det inte råder någon interaktionseffekt mellan effekterna Gender och Sequency. P-värdet från testet summerar till 0.7305 vilket inte heller är signifikant på 5 % signifikansnivå.

Interaktion mellan Experience och Sequency

$$H_0 : (\beta\gamma)_{jk} = 0$$

$$H_1 : (\beta\gamma)_{jk} \neq 0$$

$$F_{Test} = \frac{MSBC}{MSE} \Rightarrow \frac{455.617}{858.042} = 0.5310$$

$$F = (0.95, (2-1)(2-1)(3-1), 48) = 3.190727$$

Ifall av att värdet från F-testet $F > 3.190727$ överstiger det kritiska värdet kan vi med 95 % konfidens förkasta H_0 om att det inte råder någon interaktionseffekt mellan effekterna Experience och Sequency.

I detta fall där teststatistikan inte överstiger det kritiska värdet ($0.5310 < 3.190727$) antas H_0 om att det inte råder någon interaktionseffekt mellan effekterna Experience och Sequency. P-värdet från testet summerar till 0.5914 vilket inte heller är signifikant på 5 % signifikansnivå.

Slutsatsen som blir från samtliga partiella f-test ovan är att ingen av interaktionstermerna varken de parvisa eller den trippla är signifikant på 5 % signifikansnivå och kan därmed redcueras bort.

d) Presentera din valda reducerade modell.

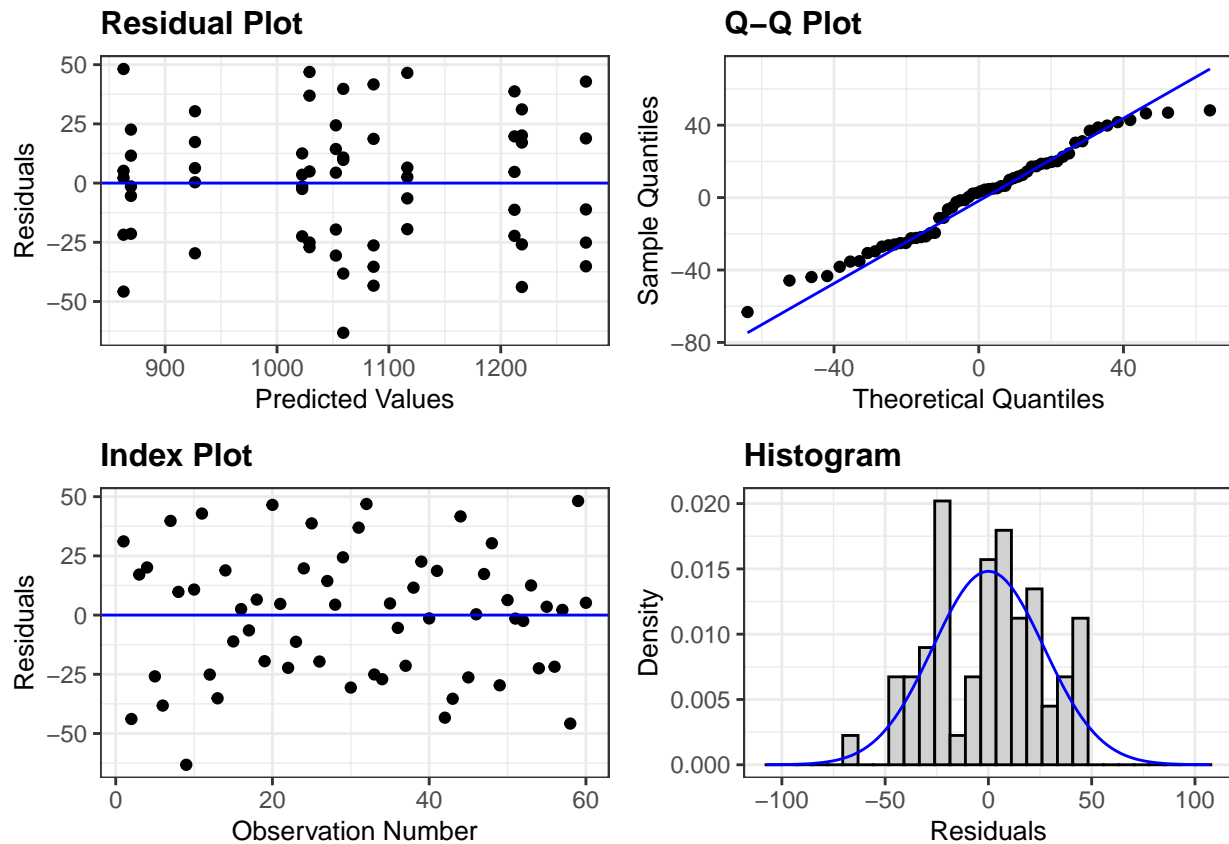
```
mod.r <-aov(y~Experience+Gender+Sequency,data = Electronic_assembly)
Anova(mod.r)
```

```
## Anova Table (Type II tests)
##
## Response: y
##           Sum Sq Df F value    Pr(>F)
## Experience 382402  1 491.978    < 2.2e-16 ***
## Gender      540361  1 695.200    < 2.2e-16 ***
## Sequency    49320  2  31.726 0.0000000006879 ***
## Residuals   42750 55
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

I den reducerade modellen är samtliga huvudeffekter från den fulla modellen kvar, samtliga har p-värden som på 5 % signifikans nivå är signifikanta, även teststatistikorna från F-testen som genomförs i modellen är signifikanta. Slutsatsen blir att modellen är en bra och tillförlig modell utifrån utskriften. Däremot måste även residualerna kontrolleras.

e) Analysera den reducerade modellen. Är den bra?

```
ggResidpanel::resid_auxpanel(residuals = residuals(mod.r),  
                             predicted = fitted(mod.r))
```



I residual plotten visas en relativt jämn varians, vad gäller q-qplotten följer ett stort antal residualer linjen men däremot en handfull observationer som inte gör det. Index plotten visar en jämn fördelning utspridd över hela diagrammet, histogrammet är nog den plott som visar sämst fördelningen, i sin helhet, går däremot residualerna för modellen att klassas som normalfördelade för att uppfylla modellantagandena.

f) 1. Använd den reducerade modellen för att lösa övning 24.14, se figur 3

```
PostHocTest(mod.r, method = "bonferroni", conf.level = 0.90)

##
## Posthoc multiple comparisons of means : Bonferroni
## 90% family-wise confidence level
##
## $Experience
##               diff    lwr.ci    upr.ci    pval
## > 18 month-< 18 months -159.6667 -171.71 -147.6234 <2e-16 ***
##
## $Gender
##               diff    lwr.ci    upr.ci    pval
## Female-Male -189.8 -201.8433 -177.7567 <2e-16 ***
##
## $Sequency
##               diff    lwr.ci    upr.ci    pval
## 2-1   57.25   38.0053   76.4947 0.0000000762 ***
## 3-1   -6.60  -25.8447   12.6447    1.0000
## 3-2  -63.85  -83.0947  -44.6053 0.0000000045 ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Formeln som används för parvisa jämförelser i faktormedelvärde med bonferronimetoden

$$\hat{D} \pm t(1 - \alpha/(2g); (n-1) \cdot abc) \cdot \sqrt{s^2 \left\{ \hat{D} \right\}}$$

$$B = t(1 - \alpha/(2g); (n-1)abc) = (1 - .10/(2 * 5); (4)12) = 2.4066$$

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

I detta fall där den reducerade modellen ska användas, har det blivit 5 jämförelser med 5 kategorier, detta då faktornivåerna som används enbart haft två nivåer.

Det som kan utläsas från tabellen är att experience är en statistiskt signifikant variabel på 10 % signifikansnivå och har en betydelse för hur lång tid de kommer ta att sätta ihop den elektroniska komponenten, alltså går det snabbare för de med mer än 18 månaders erfarenhet att sätta ihop komponenten.

Även variabeln kön har en statistiskt signifikant påverkan på tiden det kommer ta för montören att sätta ihop komponenten, med 90 % konfidens kan vi säga att medeltiden för kvinnor vad gäller att sätta ihop komponenten är lägre än medeltiden för män.

```
model.tables(mod.r,type="means")
```

f) 2. Använd den reducerade modellen för att lösa övning 24.14, se figur 3

```
## Tables of means
## Grand mean
##
## 1061.033
##
## Experience
## Experience
## < 18 months > 18 month
##      1140.9      981.2
##
## Gender
## Gender
##   Male Female
## 1155.9  966.1
##
## Sequency
## Sequency
##      1      2      3
## 1044.1 1101.4 1037.5
```

$$\bar{Y}_{ijk} \pm t(1 - \alpha/2; (n - 1)abc) \cdot s \{ \bar{Y}_{ijk} \}$$

$$t = t(1 - \alpha/2; (n - 1)abc) = (1 - .05/2; (5 - 1)12) = 2.011$$

$$s^2 \{ \bar{Y}_{ijk} \} = \frac{MSE}{n} \Rightarrow \frac{777.27}{5} = 155.454$$

$$\hat{\mu}_{ijk} = \bar{Y}_{ijk} = 1020.4$$

$$1020.4 \pm 2.011 \cdot 12.47 = 995.32 \leq \mu_{231} \leq 1045.47$$

Alltså kommer medeltiden att sätta ihop komponenten för en kvinna med mindre erfarenhet än 18 månader med 95 % konfidens att ligga mellan 995.32 och 1045 minuter, intervallet är relativt litet sett till konfidensgraden, speciellt sett till att standardavvikelsen för de enskilda faktormedelvärdena ligger mellan 12 och 41 enheter.