Labbrapport i Statistik

Laboration 16

732G46

Mattias Hällgren, Michael Debebe

Avdelningen för Statistik och maskininlärning
Institutionen för datavetenskap
Linköpings universitet

2022-01-11

Innehåll

Introduktion

I denna laboration kommer tre dataset analyseras.

Det första datasetet kommer från ett företag där de ska internutbilda personalen på deras nya datorutrustning. Man vill därför se om det är skillnad mellan de fyra olika utbildningsprogram soch deras reultat efter testet.

I det andra datasetet vill en tillverkare av frukostflingor göra ett experiment av effekten med en prisreduktion. Det valdes ut två stycken städer, tre affärer och båda städerna testades med prisreduktionen om försäljningsvolymen ökade.

Det sista datasetet vill en biolog undersöka om kackerlackor växer snabbare med hjälp av fuktiga och höga temperaturer. 27 kackerlackor delas upp i nio kombinationer för att se hur de påverkas.

Målet med denna laboration är att se hur mycket man har lärt sig av de tidigare labbarna, att analysera och tolka vilka modeller som ska användas och varför.

Databehandling

```r
#Data Uppgift 1
Resultat <- c(80,72,69,74,90,77,81,70,68,64,77,71,68,69,75,75,84,78,79,92,71,83,90,70,64,72,67,81,70,69,
Program <- c(1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,3,3,3,3,3,3,3,3,4,4,4,4,4,4,4,4)
Utbildning <- data.frame(Resultat, Program)
factosss<-c("Program")
Utbildning[factosss] <- lapply(Utbildning[factosss], factor)

#Data uppgift 2
y <- c(14,14,15,9,7,12,19,17,14,12,13,14)
stader <- c(1,1,1,2,2,2,1,1,1,2,2,2)
prisklass <- c(2,2,2,2,2,2,1,1,1,1,1,1)
Flingor <- cbind.data.frame(y, stader, prisklass)
factoss <-c("stader","prisklass")
Flingor[factoss] <- lapply(Flingor[factoss], factor)
Flingor$prisklass <-ifelse(Flingor$prisklass=="2","Ordinare","Reducerad")

#Data uppgift 3
Storlek <- c(2,2,3,6,6,8,10,10,11,9,10,11,13,14,17,15,18,20,17,18,19,22,23,25,27,28,32)
Luftfuktighet <- c(45,45,45,45,45,45,45,45,45,66,66,66,66,66,66,66,66,66,82,82,82,82,82,82,82,82,82)
Temp <- c(22,22,22,29,29,29,38,38,38,22,22,22,29,29,29,38,38,38,22,22,22,29,29,29,38,38,38)
Kackerlackor <- data.frame(Storlek, Luftfuktighet, Temp)
factos <-c("Luftfuktighet","Temp")
Kackerlackor[factos] <- lapply(Kackerlackor[factos], factor)
```

Uppgift 1

a) Pröva om det finns signifikant skillnad mellan programmen med ett test. Ange också vilken modell du har och vilka antaganden som måste vara uppfyllda för att få utföra testet. Ange hypoteserna. Visa alla uträkningar. Signifikansnivå 5%.

I detta fall har en modell av typen envägs-ANOVA valts. Exempel kommer nedan

$$y_{ij} = \mu_i + \epsilon_{ij}$$

där y_{ij} är reponsvariabeln för observation j och faktornivå i.μ_i är parametrar för faktornivåer. $\epsilon_{ij} \sim N\left(0, \sigma^2\right)$

$$\sum_j e_{ij} = 0 \quad i = 1, \ldots, r$$

Där e_{ji} är de skattade residualerna för observation j och faktor i.r är antal faktornivåer. Detta innebär att residualerna för modellen summerar till noll för varje faktornivå.

Där vi gör antagandena:

- Homogena varianser: Variansen på den aktuella variabeln är densamma i alla populationer.
- Variabeln är normalfördelad i alla populationer
- Observationer är oberoende.

$$H_0: \text{ Alla } \mu_i \text{ är lika } (i = 1, \ldots, 5)$$

$$H_1: \text{Alla } \mu_i \text{ är inte lika } (i = 1, \ldots, 5)$$

```
anova(aov(Resultat~Program, data = Utbildning))
```

```
## Analysis of Variance Table
##
## Response: Resultat
##           Df Sum Sq Mean Sq F value  Pr(>F)
## Program    3  524.5 174.833  4.2219 0.01392 *
## Residuals 28 1159.5  41.411
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$F^* = \frac{MSTR}{MSE} = \frac{174.833}{41.411} = 4.2219$$
$$F(.95, 3, 28) = 2.947$$

I fall av att teststatistikan överstiger det kritiska värdet ($F > 2.947$) förkastas H_0 om att variabeln program inte har en statistiskt signifikant påverkan på resultatet.

I detta fall då ($4.2219 < 2.947$) samtidigt som att vi får ett p-värde som summeras till (0.0139) kan vi med 95 % konfidens inte förkasta H_0 om att det inte råder en statistiskt signifikant skillnad mellan programmen. Alltså är skillnaden mellan programmen statistiskt signifikant på 5 % signifikansnivå.

b) Pröva om det är skillnad i populationsmedelvärde mellan program 2 och 3. Signifikansnivå 5%.

$$H0 : \mu_2 - \mu_3 = 0$$

$$H1 : \mu_2 - \mu_3 \neq 0$$

```
tretva <-TukeyHSD(aov(Resultat~Program, data = Utbildning),ordered = TRUE,conf.level = 0.95)
tretva$Program[3,1:4]
```

```
##        diff         lwr         upr       p adj
## 10.00000000  1.21506453 18.78493547  0.02106626
```

I ett 95 procentigt konfidensintervall kommer skillnaden mellan program 2 och 3 att ligga mellan 1.215 och 18.785, från utskriften returneras ett statistiskt signifikant p-värde (0.021) på 5 % signifikansnivå. Alltså är skillnaden mellan dem vad gäller populationsmedelvärde signifikant på 5 % signifikansnivå.

c) Låt vara populationsmedelvärde för program nr i. Pröva nollhypotesen med signifikansnivå 5%

$$\text{H0} : \frac{\mu_1 + \mu_2}{2} = \frac{\mu_3 + \mu_4}{2}$$

$$\text{H1} : \frac{\mu_1 + \mu_2}{2} \neq \frac{\mu_3 + \mu_4}{2}$$

```
c <-c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2)
Utbildning1 <-cbind(Utbildning,c)
anova(aov(Resultat~as.factor(c), data = Utbildning1))
```

```
## Analysis of Variance Table
##
## Response: Resultat
##               Df Sum Sq Mean Sq F value Pr(>F)
## as.factor(c)  1     50  50.000   0.918 0.3457
## Residuals    30   1634  54.467
```

$$F^* = \frac{MSTR}{MSE} = \frac{50}{54.467} = 0.918$$

$F(0.975, 2-1, 2(16-1)) = 5.568$

I fall av att teststatistikan överstiger det kritiska värdet ($F > 5.568$) förkastas H_0 de grupperade medelvärdena är lika.

I detta fall då ($0.918 < 2.947$) samtidigt som att vi får ett p-värde som summeras till (0.346) kan vi med 95 % konfidens inte förkasta H_0 om att det inte råder en statistiskt signifikant skillnad mellan de grupperade medelvärdena. Alltså är skillnaden mellan de grupperade medelvärdena inte statistiskt signifikant på 5 % signifikansnivå.

Genom att först räkna på medelvärdet för program 1 och program 2 för att därefter jämföra dem med medelvärdet för de 16 första observationerna drogs slutsatsen att lägga till en faktor som kännetecknar de 16 första observationerna och en faktor som kännetecknar de 16 sista faktorerna och sedan genomföra ett f-test på dessa.

d) Om antalet program som fanns tillgängliga var mycket stort och man hade erhållit de fyra programmen med ett slumpmässigt urval, hur ser då modellen ut? Antaganden? Beräkna ett 95% konfidensintervall för i detta scenario

$$\hat{\sigma}_\mu^2 = s_\mu^2 = \frac{MSTR - MSE}{n} \Rightarrow \frac{174.833 - 41.411}{8} = 16.68$$

Generella formeln för konfidensintervallet

$$\frac{(df)\sigma_\mu^2}{\chi^2(1-\alpha/2; df)} \leq \sigma_\mu^2 \leq \frac{(df)\sigma_\mu^2}{\chi^2(\alpha/2; df)}$$

Formel för frihetsgraden

$$df = \frac{(8 * 16.68)^2}{\frac{(174.833)^2}{4-1} + \frac{(41.411)^2}{4(8-1)}} = 1.7372$$

Övre kritiska chisqvärdet

$$\chi^2(.975, 1.7372) = 6.812$$

Lägre kritiska chisqvärdet

$$\chi^2(.025, 1.7372) = 0.02722$$

$$\frac{1.7372 \cdot \ 16.68}{6.812213} \leq \sigma_\mu^2 \leq \frac{1.7372 \cdot \ 16.68}{0.02722}$$

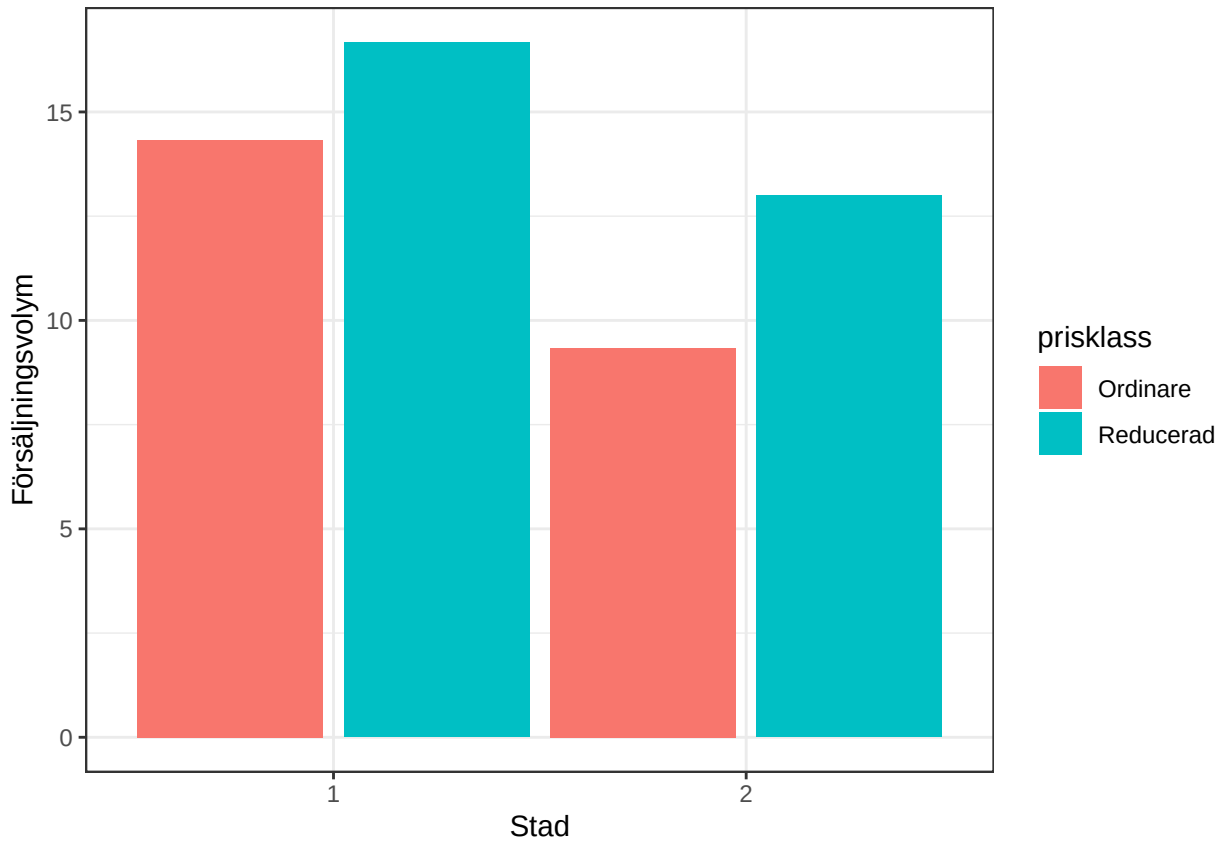
$$4.2536 \leq \sigma_\mu^2 \leq \ 1064.53$$

Med 5 % signifikans kan vi säga att konfidensintervall för σ_μ^2 kommer att ligga mellan 4.2536 och 1064.53.

Uppgift 2

a) Rita ett så kallat Treatment means plot och tolka diagrammet.

```
Sum <-Summarize(y ~ stader*prisklass,
data=Flingor)
pd = position_dodge(1)
ggplot(Sum,aes(x = stader,y = mean,fill=prisklass)) +
geom_bar(stat = "identity",position = pd)+
theme_bw()+
  ylab("Försäljningsvolym")+
  xlab("Stad")
```



I diagrammet ovan syns det hur försäljningsnivån är betydligt mycket högre för städer med en prisreduktion oberoende av vilken stad de är. Det som blir intressant är att volymökningen ser proportionell ut sett till volymen på försäljningen.

b) Ange också vilken modell du har och vilka antaganden som måste vara uppfyllda för att få utföra testet.

Begränsad ANOVA-modell med en fix och en slumpmässig faktor.

Faktor A är en fix och faktor B är slumpmässig

$$Y_{ijk} = \mu_{..} + \alpha_i + \beta_i + (a\beta)_{ij} + \epsilon_{ijk}$$
$$i = 1, \ldots, a \quad j = 1, \ldots, b \quad k = 1, \ldots, n$$

där vi gör antagandena

- $\mu\ldots$ är en konstant
- α_i är konstanter med begränsningen $\sum_i \alpha_i = 0$
- β_j är oberoende och $N\left(0, \sigma_B^2\right)$
- $(\alpha\beta)_{ij} är N\left(0, \frac{a-1}{a}\sigma_{\alpha\beta}^2\right)$, med begränsningarna:

$$-\sum_i (\alpha\beta)_{ij} = 0 \text{ för alla } j$$
$$-\sigma\left[(\alpha\beta)_{ij}, (\alpha\beta)_{i'j}\right] = -\frac{1}{a}\sigma_{\alpha\beta}^2 \quad i \neq i'$$

- ϵ_{ijk} är oberoende och $N\left(0, \sigma^2\right)$
- $\beta_j, (a\beta)_{ij}$ och ϵ_{ijk} är parvis oberoende Egenskaper

$$E\left[Y_{ijk}\right] = \mu_{..} + \alpha_i$$
$$\sigma^2\left[Y_{ijk}\right] = \sigma_Y^2 = \sigma_\beta^2 + \frac{a-1}{a}\sigma_{\alpha\beta}^2 + \sigma^2$$

b) Pröva om försäljningsvolymen har förändrats vid prisreduktionen. Visa hypoteser. Signifikansnivå 5%.

$$H_0 : \sigma_\alpha^2 = 0$$
$$H_1 : \sigma_\alpha^2 > 0$$

```
anova(aov(y~prisklass*stader,data = Flingor))
```

```
## Analysis of Variance Table
##
## Response: y
##                   Df Sum Sq Mean Sq F value   Pr(>F)
## prisklass          1 27.000  27.000  7.7143 0.024019 *
## stader             1 56.333  56.333 16.0952 0.003885 **
## prisklass:stader   1  1.333   1.333  0.3810 0.554250
## Residuals          8 28.000   3.500
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Prisreduktion

$$F^* = \frac{MSA}{MSAB} \Rightarrow \frac{27}{1.333} = 20.26$$

$$F(.95, 1, 1) = 161.45$$

I fall av att teststatistikan överstiger det kritiska värdet ($F > 161.45$) förkastas H_0 om att försäljningsvolymen inte förändrats av prisreduktionen.

I detta fall då ($20.26 < 161.45$) kan vi med 95 % konfidens anta H_0 om försäljningsvolymen inte förändrats av prisreduktionen. Alltså kan vi med 95 % konfidens säga att prisreduktion inte förändrat försäljningsvolymen.

c) Pröva om det är skillnad i försäljningsvolym mellan alla städer. Visa hypoteser. Signifikansnivå 5%.

$$H_0 : \sigma_\beta^2 = 0$$
$$H_1 : \sigma_\beta^2 > 0$$

Stad

$$F^* = \frac{MSB}{MSE} \Rightarrow \frac{56.33}{3.5} = 16.09$$

$$F(.95, 1, 1) = 161.45$$

I fall av att teststatistikan överstiger det kritiska värdet ($F > 161.45$) förkastas H_0 om att stad inte har en statistiskt signifikant påverkan på försäljningsvolym.

I detta fall då ($16.09 < 161.45$), kan vi med 95 % konfidens anta H_0 om att försäljningsvolymen inte påverkats av stad. Alltså kan vi med 95 % konfidens säga att stad inte har en statistiskt signifikant påverkan på försäljningsvolymen.

d) Skatta varianskomponenten för städer med ett 95% konfidensintervall. Använd Satterthwaites metod

$$\frac{(df)\sigma_\beta^2}{\chi^2(1-\alpha/2;df)} \leq \sigma_\beta^2 \leq \frac{(df)\sigma_\beta^2}{\chi^2(\alpha/2;df)}$$

Formel för variansen

$$s_\beta^2 = \frac{MSB - MSE}{nb} \Rightarrow \frac{56.33 - 3.5}{6} = 8.805$$

Formel för frihetsgraderna

$$df = \frac{\left(ns_\beta^2\right)^2}{\frac{(MSTR)^2}{r-1} + \frac{(MSE)^2}{r(n-1)}} \Rightarrow \frac{(6 \cdot 8.805)^2}{\frac{(56.33)^2}{2-1} + \frac{(3.5)^2}{2(6-1)}} = 0.8873$$

$$\frac{0.8873 \cdot \ 8.805}{\chi^2(0.95; 0.8873)} \leq \sigma_\beta^2 \leq \frac{0.8873 \cdot \ 8.805}{\chi^2(\ 0.05; 0.8873)}$$

$$1.66 \leq \sigma_\beta^2 \leq \ 20968.82$$

I ett 95 % konfidensintervall kommer varianskomponenten för städer att ligga mellan 1.66 och 20968.82

e) Skatta differensen i priseffekt med ett 95% konfidensintervall.

```
TukeyHSD(aov(y~stader*prisklass,data = Flingor),"prisklass",ordered = TRUE)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##     factor levels have been ordered
##
## Fit: aov(formula = y ~ stader * prisklass, data = Flingor)
##
## $prisklass
##                     diff       lwr      upr     p adj
## Reducerad-Ordinare    3 0.5092309 5.490769 0.0240185
```

I ett 95 procentigt konfidensintervall kommer differensen i priseffekt att ligga mellan 0.5092 och 5.490769 osm inte täcker noll samtidigt som att vi har ett p-värde som är statistiskt signifikant på 95 % konfidensnivå, alltså råder det statistiskt signifikant skillnad i priseffekt på 5 % signifikansnivå.

f) testa om varianskomponenten, o2ab, för interaktionen är noll. Anta därefter att den är noll och testa återigen om försäljningsvolymen har förändrats vid prisreduktionen. Signifikansnivå 5%.

```
Anova(aov(y ~ prisklass * stader, data = Flingor))
```

```
## Anova Table (Type II tests)
##
## Response: y
##                   Sum Sq Df F value   Pr(>F)
## prisklass         27.000  1  7.7143 0.024019 *
## stader            56.333  1 16.0952 0.003885 **
## prisklass:stader   1.333  1  0.3810 0.554250
## Residuals         28.000  8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$H_0 : \sigma^2_{\alpha\beta} = 0$$
$$H_1 : \sigma^2_{\alpha\beta} > 0$$

$$F^* = \frac{MSAB}{MSE} \Rightarrow \frac{1.3333}{3.5} = 0.381$$
$$F(.95, 1, 8) = 5.3177$$

I fall av att teststatistikan överstiger det kritiska värdet ($F > 5.3177$) förkastas H_0 om att det råder interaktionseffekt mellan variablerna.

I detta fall då ($0.1285 < 5.3177$) samtidigt som att vi får ett p-värde som summeras till (0.942) kan vi med 95 % konfidens anta H_0 om att det inte råder en interaktionseffekt mellan variablerna. Alltså kan vi med 95 % konfidens säga att det inte råder någon interaktionseffekt mellan variablerna.

Pröva därefter om försäljningssvolym har förändrats vid prisreduktionen. Signifikansnivå 5 %.

$$H_0 : \sigma^2_{\alpha} = 0$$
$$H_1 : \sigma^2_{\alpha} > 0$$

```
anova(aov(y~prisklass+stader,data = Flingor))
```

```
## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value   Pr(>F)
## prisklass  1 27.000  27.000  8.2841 0.018231 *
## stader     1 56.333  56.333 17.2841 0.002457 **
## Residuals  9 29.333   3.259
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$F^* = \frac{MSA}{MSE} \Rightarrow \frac{27}{3.259} = 8.285$$

$$F(.95, 1, 8) = 5.3177$$

I fall av att teststatistikan överstiger det kritiska värdet ($F > 5.3177$) förkastas H_0 om att försäljningsvolymen inte förändrats av prisreduktionen.

I detta fall då ($8.285 >$) kan vi med 95 % konfidens förkasta H_0 om försäljningsvolymen inte förändrats av prisreduktionen. Alltså kan vi med 95 % konfidens säga att prisreduktion förändrat försäljningsvolymen.

Uppgift 3

a) Vilken ANOVA-modell är lämplig i detta fall? Skriv upp modellen och dess antaganden

I detta fall är det en Två-vägs Anova med alla typer av effektkombinationer som är utav intresse, alltså en Mixed effects model (två slumpmässiga faktorer), det som i figur 3, tabell 25.5 kallas Random ANOVA model. Detta utfirån att vi har balanserade celler med två slumpmässiga faktorer.

$$Y_{ijk} = \mu_{..} + \alpha_i + \beta_j + (a\beta)_{ij} + \epsilon_{ijk}$$
$$i = 1, \ldots, a \quad j = 1, \ldots, b \quad k = 1, \ldots, n$$

där vi gör antagandena:

- $\mu_{..}$ är en konstant
- α_i, β_j och $(a\beta)_{ij}$ är oberoende normalfördelade slumpvariabler med väntevärde 0 och varians $\sigma^2_\alpha, \sigma^2_\beta$ och $\sigma^2_{\alpha\beta}$ respektive
- ϵ_{ijk} är oberoende och $N\left(0, \sigma^2\right)$
- $\alpha_i, \beta_j, (a\beta)_{ij}$ och ϵ_{ijk} är parvís oberoende Egenskaper

$$E\left[Y_{ijk}\right] = \mu_{...}$$
$$\sigma^2\left[Y_{ijk}\right] = \sigma^2_Y = \sigma^2_\alpha + \sigma^2_\beta + \sigma^2_{\alpha\beta} + \sigma^2$$
$$Y_{ijk} \sim N\left(\mu_{...}, \sigma^2_Y\right)$$

b) Testa om det finns interaktionseffekt på 5% signifikansnivå. Formulera hypoteser, utför testet och dra slutsats.

$$H_0 : \sigma^2_{\alpha\beta} = 0$$
$$H_1 : \sigma^2_{\alpha\beta} > 0$$

```
model7 <- aov(Storlek ~ Luftfuktighet * Temp, data = Kackerlackor)
anova(model7)
```

```
## Analysis of Variance Table
##
## Response: Storlek
##                     Df  Sum Sq Mean Sq  F value    Pr(>F)
## Luftfuktighet        2 1304.67  652.33 244.6250 8.922e-14 ***
## Temp                 2  356.22  178.11  66.7917 4.695e-09 ***
## Luftfuktighet:Temp   4   11.11    2.78   1.0417    0.4132
## Residuals           18   48.00    2.67
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$F^* = \frac{MSAB}{MSE} \Rightarrow \frac{2.78}{2.67} = 1.042$$
$$F(.95, 4, 18) = 2.928$$

I fall av att teststatistikan överstiger det kritiska värdet ($F > 2.928$) förkastas H_0 om att det inte råder interaktionseffekt mellan variablerna.

I detta fall då ($0.1285 < 2.928$) samtidigt som att vi får ett p-värde som summeras till (0.4132) kan vi med 95 % konfidens anta H_0 om att det inte råder en interaktionseffekt mellan variablerna. Alltså kan vi med 95 % konfidens säga att det inte råder någon interaktionseffekt mellan variablerna.

c) Pröva om temperatur och luftfuktighet har effekt på kackerlackornas växthastighet. Använd 5 % signifikansnivå, formulera hypoteser, utför testen och dra slutsatser

```
model5 <- aov(Storlek ~ Luftfuktighet * Temp, data = Kackerlackor)
anova(model5)
```

```
## Analysis of Variance Table
##
## Response: Storlek
##                     Df  Sum Sq Mean Sq  F value    Pr(>F)
## Luftfuktighet        2 1304.67  652.33 244.6250 8.922e-14 ***
## Temp                 2  356.22  178.11  66.7917 4.695e-09 ***
## Luftfuktighet:Temp   4   11.11    2.78   1.0417    0.4132
## Residuals           18   48.00    2.67
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Luftfuktighet

$$H_0 : \sigma_\alpha^2 = 0$$
$$H_1 : \sigma_\alpha^2 > 0$$
$$F^* = \frac{MSA}{MSAB} = \frac{652.33}{2.78} = 234.651$$

$$F(.95, 2, 4) = 6.944$$

I fall av att teststatistikan överstiger det kritiska värdet ($F > 6.944$) förkastas H_0 om att luftfuktighet inte har effekt på kackerlackornas växthastighet.

I detta fall då ($234.651 > 6.944$) samtidigt som att vi får ett p-värde som summeras till ($8.92e-14$) kan vi med 95 % konfidens förkasta H_0 om att luftfuktighet inte har effekt på kackerlackornas växthastighet. Alltså kan vi med 95 % konfidens säga att luftfuktighet inte har effekt på kackerlackornas växthastighet.

Temperatur

$$H_0 : \sigma_\beta^2 = 0$$
$$H_1 : \sigma_\beta^2 > 0$$
$$F^* = \frac{MSB}{MSAB} \Rightarrow \frac{178.11}{2.78} = 64.07$$
$$F(.95, 2, 4) = 6.944$$

I fall av att teststatistikan överstiger det kritiska värdet ($F > 6.944$) förkastas H_0 om att temperatur inte har effekt på kackerlackornas växthastighet.

I detta fall då ($64.07 > 6.944$) samtidigt som att vi får ett p-värde som summeras till ($6.453e-10$) kan vi med 95 % konfidens förkasta H_0 om att temperatur inte har effekt på kackerlackornas växthastighet. Alltså kan vi med 95 % konfidens säga att temperatur har effekt på kackerlackornas växthastighet.

d) Skattta varianskomponenten för Temperatur med ett 95 % konfidensintervall, använd Satterhwaites metod.

$$\frac{(df)\sigma_\beta^2}{\chi^2(1-\alpha/2;df)} \leq \sigma_\beta^2 \leq \frac{(df)\sigma_\beta^2}{\chi^2(\alpha/2;df)}$$

Formel för variansen

$$s_\beta^2 = \frac{MSB - MSE}{nb} \Rightarrow \frac{178.111 - 2.687}{9} = 19.492$$

Formel för frihetsgraderna

$$df = \frac{\left(ns_\beta^2\right)^2}{\frac{(MSTR)^2}{r-1} + \frac{(MSE)^2}{r(n-1)}} \Rightarrow \frac{(9 \cdot 19.492)^2}{\frac{(178.11)^2}{3-1} + \frac{(2.687)^2}{3(9-1)}} = 1.940$$

$$\frac{1.940 \cdot 19.492}{\chi^2(0.975; 1.940)} \leq \sigma_\beta^2 \leq \frac{1.940 \cdot \ 19.492}{\chi^2(\ 0.05; 1.940)}$$

$$5.2150 \leq \sigma_\beta^2 \leq \ 848.954$$

I ett 95 % konfidensintervall kommer varianskomponenten för temperatur att ligga mellan 5.215 och 848.954.

e) Är det av intresse att göra multipla jämförelser i detta fall? Om ja, utför dessa. Om nej, motivera varför.

I fall av att de förklarande variablerna tillsammans hade skapat en interaktionseffekt hade de varit nödvändigt att göra multipla jämförelser för att separera dem, när de däremot inte gör det, är det inte heller nödvändigt med multipla jämförelser. Den andra förklaringen är att trots att de inte finns någon interaktionseffekt, är fortfarande båda variabler viktiga för att kunna skapa en så stark och förklarande modell som möjligt; att enbart titta på luftfuktighet eller temperatur kan absolut bidra till en bra "skattning" då båda är statistiskt signifikanta. Men inte ett tillförtligt resultat att luta sig tillbaka mot.

f) Oavsett resultat på tidigare uppgifter: Resonera hurvida det är rimligt att anta att ”längdökning på kackerlackorna i millimeter” är en normalfördelad slumpvariabel eller ej. Motivera ditt svar.

Till att börja med så är en av grundreglerna för att anta att något är “normalfördelat” att urvalsstorlekn är större än >30, vilket det inte är i detta fall. Däremot har kackerlackorna varit identiska från början och sedan placerats i olika slumpmässiga kombinationer av grupper och därav dras slutsatsen att kackerlackorna i sig är en slumpvariabel men att de är för litet urval för att kallas normalfördelat, slutsatsen blir att det inte är en normalfördelad slumpvariabel.

Lärdomar

Många.