Labbrapport i Statistik

Laboration 15

732G46

Mattias Hällgren, Michael Debebe

Avdelningen för Statistik och maskininlärning
Institutionen för datavetenskap
Linköpings universitet

2022-01-09

Innehåll

Introduktion 1

Databehandling 2

Uppgifter 3

Uppgift 1 Envägs ANOVA med slumpmässiga effekter 3

25.9 Coil winding Machines 3

a) Test whether or not the mean coil characteristic is the same for all machines in the plant use a=0.10, State the alternatives, decision rule and conclusion. What is the P-value of the test? 3

b) Estimate the mean coil characteristic for all coil winding machines in the plant; use a 90 percent confidence interval 4

25.10 Refer to Coil winding machines problem 25.9 5

a) Estimate with a 90 percent confindence interval. Interpret your interval estimate 5

b) Test whether or not are equal use a=.10. State the alternatives, deicision rule and conclusion, 6

c) Estimate σ^2 with a 90 percent confidence interval. Interpret your interval estimate. 6

d) Obtain a point estimate of σ. Interpret your interval estimate 7

e) Obtain separate approximate 90 percent confidence intervals for using the satterhwaite process. Comment 7

Uppgift 2 Två-vägs ANOVA med alla typer av effektkombinationer 8

a) Test for interaction effects; use a = .05. State the alternatives,decision rule, and conclusion. What is the p-value for the test? 9

Test for interaction effect 9

b) Test for factor A and factor B main effects. For each tests, use a =.05 and state the alternatives, decision rule and conclusion. WHat is the p-value for all tests? 10

Test for Fixed factor effect (Factor A, Coats) 10

Test for random effect (Factor B, Batch) 11

c) Estimte D1 = u2-u1 and D2 = u3-u2 by means of the bonferroni procedure with a 90 percent family confidence coefficent. State your finings 12

d) Use the Satterhwaite procedure to obtain an approximate 95 percent confidence interval for u2, Interpret your confidence interval 13

e) Use the Satterhwaite procedure to obtain an approximate 90 percent confidence interval for o2b Does 02b appear to be larged compared to o2? 14

Introduktion

I denna laboration kommer två dataset analyseras.

Det första datasetet kommer från en studie gneomförd av en produktionsanalytiker som studerar en särskild typ av karaktär hos spolar som blir lindade av en spollindningsmaskin, detta genom att slumpmässigt välja ut fyra maskiner och studera 10 slumpmässigt utvalda spolar.

I det andra datasetet är det data från preliminär forskning om framställniong av imiterade pärlor som syftar till studera effekten av antalet lager av en speciell lack (Factor Coats), applicerad på en opaliserad plastpärla som används för bas för pärlans marknadsvärde. Fyra satser om 12 pärlor (Faktor Batch) användes i studien. Detta data är fixed.

Målet med denna laboration är att förstå skilnaden mellan fixa och slumpmässiga effekter, formulera modeller med fixa, slumpmässiga och blandade modeller. Utföra relevant inferens i de olika situtionerna.

Databehandling

```
Coilwinding <-read.table("https://raw.githubusercontent.com/MichaelDebebe/6555/main/Coilwinding")
cols1 <-c("Y","Machine","Coil")
colnames(Coilwinding) <-cols1
factores <-c("Machine","Coil")
Coilwinding[factores] <- lapply(Coilwinding[factores], factor)

Ipearls <-read.table("https://raw.githubusercontent.com/MichaelDebebe/6555/main/Ipearls")
cols2 <-c("Y","Coats","Batch","Vet ej")
colnames(Ipearls) <-cols2
factos <-c("Coats","Batch")
Ipearls[factos] <- lapply(Ipearls[factos], factor)
```

Uppgifter

Uppgift 1 Envägs ANOVA med slumpmässiga effekter

Envägs ANOVA med slumpmässiga effekter

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

där y_{ij} är reponsvariabeln för observation j och faktornivå $i.\mu_i$ är parametrar för faktornivåer. $\epsilon_{ij} \sim N\left(0, \sigma^2\right)$ Ekvation (16.21):

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

Där e_{ji} är de skattade residualerna för observation j och faktor $i.r$ är antal faktornivåer. Detta innebär att residualerna för modell (16.2) summerar till noll för varje faktornivå.

25.9 Coil winding Machines

a) Test whether or not the mean coil characteristic is the same for all machines in the plant use a=0.10, State the alternatives, decision rule and conclusion. What is the P-value of the test?

```
##             Df Sum Sq Mean Sq F value     Pr(>F)
## Machine      3  602.5 200.833 28.0886 1.5397e-09 ***
## Residuals   36  257.4   7.150
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$H0 : \sigma\mu^2 = 0$$

$$H1 : \sigma\mu^2 > 0$$

Uträkning för F-värdet

$$F = \frac{MSR}{MSE} = \frac{602.5/3}{257.4/36} = \frac{200.83}{7.15} = 28.089$$

kritiska värdet

$$F(0.90, 3, 36) = 2.2426$$

I fall av att teststatistikan överstiger det kritiska värdet ($F > 2.2426$) förkastas H_0 om att medelkaraktären är den samma för alla maskiner.

I detta fall då $(28.089 > 2.2426)$ samtidigt som att vi får ett p-värde som summeras till $(1.54e - 09)$ kan vi med 90 % konfidens förkasta H_0 om att medelkaraktären är den samma för alla maskiner.

b) Estimate the mean coil characteristic for all coil winding machines in the plant; use a 90 percent confidence interval

```
mean(Coilwinding$Y)
```

```
## [1] 205.05
```

Formeln som används för konfidensintervallet

$$\bar{Y} \pm t(1-\alpha/2; r-1) \cdot s\left\{\bar{Y}\right\}$$

Formel för variansen

$$s^2\left\{\bar{Y}_{ij}\cdot\right\} = \frac{MSE}{n} \Rightarrow \frac{200.83}{40} = 5.02075$$

Uträkning för kritiska värdet

$$t(.95, 4-1) = 2.353$$

Uträkning för konfidensintervall i antal lindande spolar

$$205.05 \pm 2.353 \cdot 2.240 = 199.779 : 210.32$$

Medelvärdet för spollindningsmaskinen är 205.05 maskiner. I ett konfidensintervall kommer medelvärdet för antalet lindade spolar att ligga mellan 199.78 och 210.32 med 90 % konfidens.

25.10 Refer to Coil winding machines problem 25.9

a) Estimate with a 90 percent confindence interval. Interpret your interval estimate

I denna uppgift kommer en skattning av kvoten $\sigma_\mu^2/\left(\sigma_\mu^2+\sigma^2\right)$ F-värdet som används för den lägre gränsen

$$F(.95,3,36)=2.8663$$

F-värdet som används för den övre gränsen

$$F(.05,3,36)=0.11625$$

Formel för kvoten av den lägre konfidensgränsen

$$L=\frac{1}{n}\left[\frac{\text{MSTR}}{\text{MSE}}\left(\frac{1}{F(1-\alpha/2;r-1,r(n-1))}\right)-1\right]$$

Uträkning för kvoten av den lägre konfidensgränsen

$$L=\frac{1}{10}\left[\frac{200.83}{7.15}\left(\frac{1}{2.8663}\right)-1\right]=0.879$$

Konfidensintervall för kvoten av den lägre gränsen

$$L^*=\frac{L}{1+L}=>\frac{0.879}{1+0.879}=0.4678$$

Formel för kvoten av den övre konfidensgränsen

$$U=\frac{1}{n}\left[\frac{\text{MSTR}}{\text{MSE}}\left(\frac{1}{F(\alpha/2;r-1,r(n-1))}\right)-1\right]$$

Uträkning för kvoten av den övre konfidensgränsen

$$U=\frac{1}{10}\left[\frac{200.83}{7.15}\left(\frac{1}{0.11625}\right)-1\right]=24.061$$

Konfidensintervall för kvoten av den övre gränsen

$$U^*=\frac{U}{1+U}\Rightarrow\frac{24.061}{1+24.061}=0.96$$

Konfidensintervallet

$$0.4678\leq\frac{\sigma_\mu^2}{\sigma_\mu^2+\sigma^2}\leq 0.96$$

I ett 95% konfidensintervall så är den undre konfidensgränsen på 0.4678 och den övre 0.96. Eftersom att den ligger mellan 0 och 1 så kan kvoten ses som oändlig.

b) Test whether or not are equal use a=.10. State the alternatives, deicision rule and conclusion,

$$\mathrm{H}_0 : \sigma_\mu^2 = \sigma^2$$

$$\mathrm{H}_1 : \sigma_\mu^2 \neq \sigma^2$$

$$F = \frac{MSR}{MSE} \cdot \frac{1}{n+1} \Rightarrow \frac{200.83}{7.15} \cdot \frac{1}{10+1} = 2.553$$

$$F(.95, 3, 36) = 2.8663$$

$$F(.05, 3, 36) = 0.11625$$

I fall av att teststatistikan över eller understiger de kritiska värdena ($0.11625 < F < 2.8663$) förkastas H_0 om att varianserna är lika.

I detta fall då ($0.11625 < 2.553 < 2.8663$) kan vi med 90 % konfidens inte förkasta H_0 om att varianserna är lika.

c) Estimate σ^2 with a 90 percent confidence interval. Interpret your interval estimate.

$$\frac{r(n-1)MSE}{\chi^2(1-\alpha/2; r(n-1))} \leq \sigma^2 \leq \frac{r(n-1)MSE}{\chi^2(\alpha/2; r(n-1))}$$

$$\chi^2(.95, 36) = 50.998$$

$$\chi^2(.05, 36) = 23.269$$

$$\frac{4(10-1) \cdot 7.15}{50.998} \leq \sigma^2 \leq \frac{4(10-1) \cdot 7.15}{23.269}$$

$$5.0472 \leq \sigma^2 \leq 11.062$$

I ett 90 procentigt konfidensintervall kommer σ^2 att ligga mellan 5.0472 och 11.062.

d) Obtain a point estimate of σ. Interpret your interval estimate

$$\hat{\sigma}_\mu^2 = s_\mu^2 = \frac{MSTR - MSE}{n} \Rightarrow \frac{200.83 - 7.15}{10} = 19.37$$

Med hjälpa av punktskattning av urvalsvariansen kommer att summera till 19.37.

e) Obtain separate approximate 90 percent confidence intervals for using the satterhwaite process. Comment

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

$$df = \frac{(10 * 19.37)^2}{\frac{(200.83)^2}{4-1} + \frac{(7.15)^2}{4(10-1)}} = 2.790$$

Övre kritiska chisqvärdet

$$\chi^2(.95, 2.79) = 7.4474$$

Lägre kritiska chisqvärdet

$$\chi^2(.05, 2.79) = 0.289$$

$$\frac{2.79 \cdot \ 19.37}{7.4474} \leq \sigma_\mu^2 \leq \frac{2.79 \cdot \ 19.37}{0.289}$$
$$7.257 \leq \sigma_\mu^2 \leq \ 186.99$$

Med 10 % signifikans kan vi säga att konfidensintervall för σ_μ^2 kommer att ligga mellan 7.257 och 186.99 vilket är ett väldigt stort intervall!

Uppgift 2 Två-vägs ANOVA med alla typer av effektkombinationer

Två-vägs ANOVA modell med en fix och en slumpmässig faktor

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

Där vi gör antagandena

- $\mu_{..}$ är en konstant
- α_i är konstanter med begränsningen $\sum_i \alpha_i = 0$
- β_j är oberoende och $N\left(0, \sigma_\beta^2\right)$
- $(\alpha\beta)_{ij}$ 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

a) Test for interaction effects; use a = .05. State the alternatives,decision rule, and conclusion. What is the p-value for the test?

```
Perla <-aov(Y~Coats*Batch,data = Ipearls)
Anova(Perla)
```

```
## Anova Table (Type II tests)
##
## Response: Y
##                Sum Sq Df  F value     Pr(>F)
## Coats       150.38792  2 15.59097 1.3271e-05 ***
## Batch       152.85167  3 10.56426 3.9841e-05 ***
## Coats:Batch   1.85208  6  0.06400    0.99883
## Residuals   173.62500 36
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Test for interaction effect

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

$$F^* = \frac{MSAB}{MSE} = \frac{1.852/6}{173.625/36} = \frac{0.309}{4.823} = 0.064$$

$$F(.95, 6, 36) = 2.36$$

I fall av att teststatistikan överstiger det kritiska värdet ($F > 2.36$) förkastas H_0 om att det råder interaktionseffekt mellan variablerna.

I detta fall då ($0.064 < 2.36$) samtidigt som att vi får ett p-värde som summeras till (0.99883) kan vi med 95 % konfidens inte förkasta 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.

Anledningen till varför ett test för interaktionstermen blir viktigt i just denna modell är för att uppfylla modellkraven vad gäller att ($\beta_j, (a\beta)_{ij}$ och ϵ_{ijk} är parvis oberoende)

b) Test for factor A and factor B main effects. For each tests, use a =.05 and state the alternatives, decision rule and conclusion. WHat is the p-value for all tests?

Test for Fixed factor effect (Factor A, Coats)

$$H0 : \alpha_1 = \alpha 2 = \alpha 3 = 0$$

$$H_1 : \alpha_1 \neq \alpha_2 \neq \alpha_3 \neq \ 0$$

$$F^* = \frac{MSA}{MSAB} = \frac{152.85167/2}{1.85208/6} = \frac{75.194}{0.30868} = 243.346$$

$$F(0.95, 2, 6) = 5.14326$$

I fall av att teststatistikan överstiger det kritiska värdet ($F > 5.143$) förkastas H_0 om att Faktor A inte har någon statistiskt signifikant påverkan på marknadspriset.

I detta fall då ($243.346 > 5.143$) samtidigt som att vi får ett p-värde som summeras till ($1.3271e - 05$) kan vi med 95 % konfidens förkasta H_0 om att faktor A (Coats) inte har en statistiskt signifikant påverkan på marknadspriset. Alltså har faktor A en statistiskt signifikant påverkan på marknadspriset på 5 % signifikansgrad.

Test for random effect (Factor B, Batch)

$$H0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$$

$$H0 : \beta_1 \neq \beta_2 \neq \beta_3 \neq \beta_4 \neq 0$$

$$F^* = \frac{MSB}{MSE} = \frac{152.852/3}{173.625/36} = \frac{50.9506}{4.823} = 10.56$$

$$F(0.95, 3, 6) = 4.7570$$

I fall av att teststatistikan överstiger det kritiska värdet ($F > 4.7570$) förkastas H_0 om att Faktor B (Batch) inte har en statistiskt signifikant påverkan på marknadspriset.

I detta fall då ($10.56 > 5.14326$) samtidigt som att vi får ett p-värde som summeras till ($3.9841e - 05$) kan vi med 95 % konfidens förkasta H_0 om att faktor B (Batch) inte har en statistiskt signifikant påverkan på marknadspriset.Alltså har faktor B en statistiskt signifikant påverkan på marknadspriset på 5 % signifikansgrad.

c) Estimte D1 = u2-u1 and D2 = u3-u2 by means of the bonferroni procedure with a 90 percent family confidence coefficent. State your finings

```
by(Ipearls$Y, Ipearls$Coats, mean)
```

```
## Ipearls$Coats: 1
## [1] 73.10625
## ------------------------------------------------------------
## Ipearls$Coats: 2
## [1] 76.79375
## ------------------------------------------------------------
## Ipearls$Coats: 3
## [1] 76.925
```

Formel för konfidensintervallet

$$\hat{D} \pm t(1-\alpha/2; (a-1)(b-1)) \cdot s[\hat{D}]$$

$$D1_{(2-1)} = 76.79375 - 73.10625 = 3.6875 \quad D2_{(3-2)} = 76.925 - 76.79375 = 0.13125$$

Kritiska t-värdet

$$t(0.975, (3-1) \cdot (4-1)) = 2.447$$

$$s^2[\hat{D}] = \frac{MSAB}{bn} \sum c_i^2 \Rightarrow \frac{0.3087}{3} \cdot 0.3748 = 0.1964$$

$$D1 \Rightarrow 3.68750 \pm 2.447 \cdot 0.1964 = 3.207 : 4.1681$$

Med 95 % konfidens kan vi säga att konfidensintervall för D1 kommer att ligga mellan 3.207 och 4.1681. Detta intervall är bra i form av att det inte täcker 0 samt inte är särskilt stort.

$$D2 \Rightarrow 0.13125 \pm 2.447 \cdot 0.02576 = -0.3493 : 0.612$$

Med 95 % konfidens kan vi säga att konfidensintervall för D2 kommer att ligga mellan -0.3493 och 0.612. Detta intervall täcker 0 vilket gör att skillnaden mellan medelvärdena inte går att bekräfta.

d) Use the Satterhwaite procedure to obtain an approximate 95 percent confidence interval for u2, Interpret your confidence interval

Den generella formlen för konfidensintervallet

$$\hat{\mu}_i \pm t(1-\alpha/2; df) \cdot s\left[\hat{\mu}_i\right]$$

Formel för frihetsgraderna

$$df = \frac{\left(\frac{a-1}{nab}MSAB + \frac{1}{nab}MSB\right)^2}{\frac{\left(\frac{a-1}{nab}MSAB\right)^2}{(a-1)(b-1)} + \frac{\left(\frac{1}{nab}MSB\right)^2}{(b-1)}} \Rightarrow \frac{\left(\frac{3-1}{48} \cdot .309 + \frac{-1}{48} \cdot 50.951\right)^2}{\frac{\left(\frac{3-1}{48} \cdot .309\right)^2}{(3-1)(4-1)} + \frac{\left(\frac{1}{48} \cdot 50.951\right)^2}{(4-1)}} = 3.073$$

Formel för standardavvikelsen

$$s^2\left[\hat{\mu}_i\right] = c_1 MSAB + c_2 MSB \Rightarrow \frac{3-1}{48} \cdot .309 + \frac{1}{48} \cdot 50.951 = 1.0743$$

Det kritiska t-värdet

$$t(.975, 3.073) = 3.140$$

Konfidensintervallet

$$\hat{\mu}_2 \pm t(.975, 3.073) \cdot s\left[\hat{\mu}_i\right] \Rightarrow \; 76.794 \pm 3.140 \cdot 1.0365 = 73.54 : 80.05$$

Alltså kommer medelmarknadspriset för en pärla med 8 lager lack att ligga mellan 73.54 och 80.05 prisenheter i ett konfidens intervall på 5 % signifikansnivå.

e) Use the Satterhwaite procedure to obtain an approximate 90 percent confidence interval for o2b Does 02b appear to be larged compared to o2?

Generella formeln för satterwhaite intervall för varians

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

Formel för variansen $s_\beta^2 = \frac{MSB-MSE}{na} \Rightarrow \frac{50.951-4.823}{16} = 2.883$

Formel för frihetsgraderna

$$df = \frac{(c_1MS_1 + \ldots + c_hMS_h)^2}{\frac{(c_1MS_1)^2}{df_1} + \ldots + \frac{(c_hMS_h)^2}{df_h}} \Rightarrow \frac{\left(\frac{1}{12}50.951 - \frac{1}{12}4.823\right)^2}{\frac{\left(\frac{1}{12}50.951\right)^2}{3} - \frac{(\frac{1}{12}4.823)^2}{36}} = 2.4811$$

$$\frac{2.4811 \cdot 2.883}{\chi^2(0.95; 2.4811)} \leq \sigma_\beta^2 \leq \frac{2.4811 \cdot \ 3.844}{\chi^2(\ 0.05; 2.4811)}$$

$$1.0376 \leq \sigma_\beta^2 \leq \ 34.7180$$

I ett 90 procentigt konfidensintervall kommer variansen för Faktor B (Batch) att ligga mellan 1.0376 och 34.7180, vilket är ett intervall som inte täcker 0. σ_β^2 skiljer inte så mycket från σ^2