HW9

Wenxiao Yang

11/15/2021

2

(a)

```
library(faraway)
data(butterfat)
head(butterfat)
```

```
##   Butterfat    Breed    Age
## 1      3.74 Ayrshire Mature
## 2      4.01 Ayrshire  2year
## 3      3.77 Ayrshire Mature
## 4      3.78 Ayrshire  2year
## 5      4.10 Ayrshire Mature
## 6      4.06 Ayrshire  2year
```

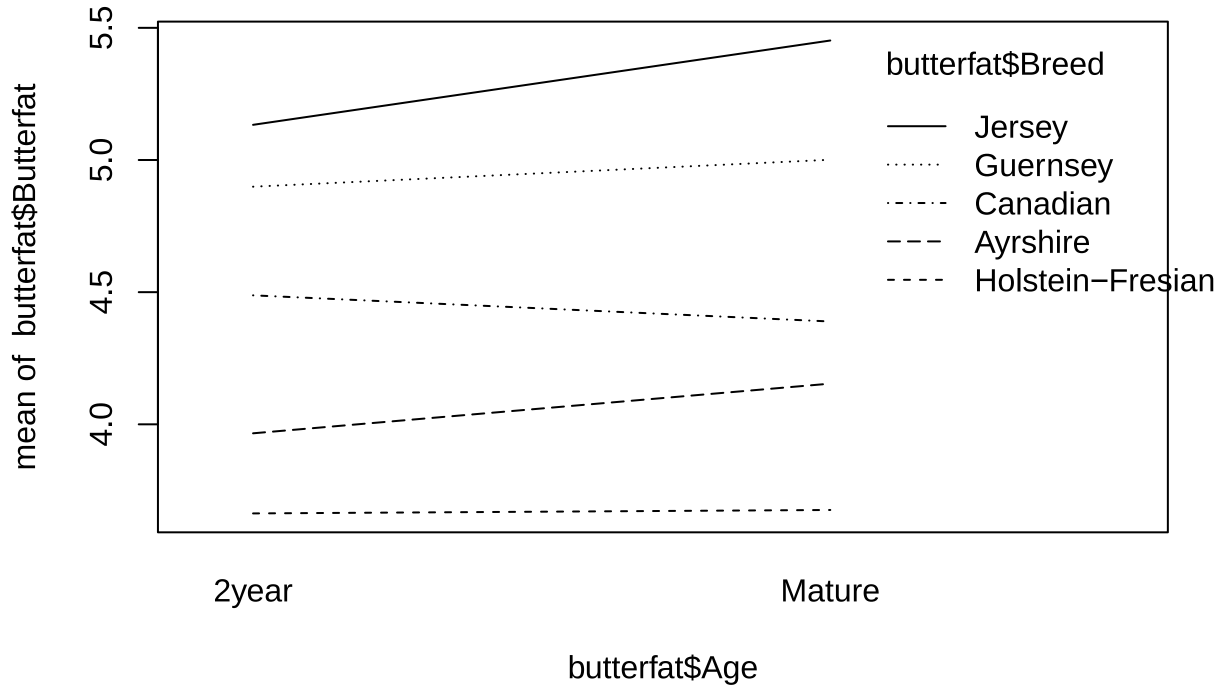
The factor effects model is as follows:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

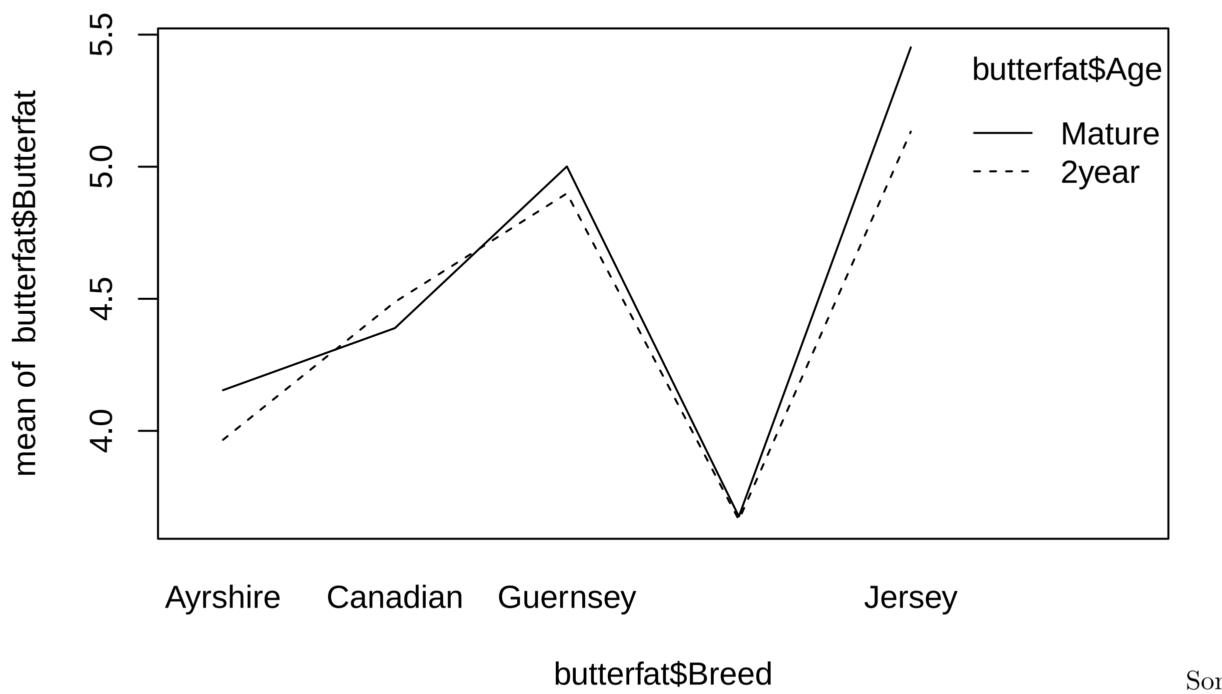
Y_{ijk}: butterfat for age i, breed j, cow k μ : mean butterfat for all samples α_i : effect of age i on butterfat β_j : effect of breed j on butterfat $(\alpha\beta)_{ij}$: interaction term The error terms satisfy the usual assumption $\varepsilon_{ijk} \sim \mathcal{N}\left(0, \sigma^2\right)$. Sum Constraints: $\sum_i \alpha_i = 0, \sum_j \beta_j 0, \sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} = 0$

(b)

```
interaction.plot(butterfat$Age, butterfat$Breed, butterfat$Butterfat)
```




```
interaction.plot(butterfat$Breed, butterfat$Age, butterfat$Butterfat)
```



Some of the lines are intersect, so interactions are present.

(c)

```
model1=lm(log(Butterfat)~Age*Breed,data=butterfat)
anova(model1)
```

```
## Analysis of Variance Table
##
```

```
## Response: log(Butterfat)
##           Df  Sum Sq Mean Sq F value Pr(>F)
## Age        1 0.01367 0.01367  1.8141 0.1814
## Breed      4 1.70334 0.42584 56.5179 <2e-16 ***
## Age:Breed  4 0.02232 0.00558  0.7406 0.5668
## Residuals 90 0.67811 0.00753
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The hypothesis we test is

$$\begin{cases} H_0 : (\alpha\beta)_{ij} = 0, \forall i, j \\ H_\alpha : \text{ not all } (\alpha\beta)_{ij} = 0 \end{cases}$$

Since the p-value is large, we conclude that the interaction term is not statistically significant. So, we can remove it from the model.

(d)

```
model2=lm(log(Butterfat)~Age+Breed,data=butterfat)
```

$$\begin{cases} H_0 : Y_{ijk} = \mu + \beta_j + \varepsilon_{ijk} \\ H_\alpha : Y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk} \end{cases}$$

```
model3=lm(log(Butterfat)~Breed,data=butterfat)
anova(model3,model2)
```

```
## Analysis of Variance Table
##
## Model 1: log(Butterfat) ~ Breed
## Model 2: log(Butterfat) ~ Age + Breed
##   Res.Df     RSS Df Sum of Sq      F Pr(>F)
## 1     95 0.71410
## 2     94 0.70043  1  0.013668 1.8343 0.1789
```

The p-value is larger than 0.05, then we can't reject null hypothesis. So, we conclude that Age is statistically insignificant.

$$\begin{cases} H_0 : Y_{ijk} = \mu + \alpha_i + \varepsilon_{ijk} \\ H_\alpha : Y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk} \end{cases}$$

```
model4=lm(log(Butterfat)~Age,data=butterfat)
anova(model4,model2)
```

```
## Analysis of Variance Table
##
## Model 1: log(Butterfat) ~ Age
## Model 2: log(Butterfat) ~ Age + Breed
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1     98 2.40377
## 2     94 0.70043  4    1.7033 57.149 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value is smaller than 0.05, then we reject null hypothesis. So, we conclude that Breed is statistically significant.

(e)

```r
anova(model2)
```

```
## Analysis of Variance Table
##
## Response: log(Butterfat)
##           Df  Sum Sq Mean Sq F value Pr(>F)
## Age        1 0.01367 0.01367  1.8343 0.1789
## Breed      4 1.70334 0.42584 57.1486 <2e-16 ***
## Residuals 94 0.70043 0.00745
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$D = \mu_{1.} - \mu_{2.}$$

```r
butterfat['logbutfat']=log(butterfat$Butterfat)
mean.mature=mean(butterfat$logbutfat[butterfat$Age=="Mature"])
mean.2year=mean(butterfat$logbutfat[butterfat$Age=="2year"])
mean.mature-mean.2year
```

```
## [1] 0.02338223
```

$$\hat{D} = \hat{Y}_{1..} - \hat{Y}_{2..} = 0.02338223$$

```r
0.00745/25
```

```
## [1] 0.000298
```

$$MSE = 0.00745, s^2_{\hat{D}} = \frac{2MSE}{10 * 5} = 0.000298$$

$$D \in 0.02338223 \pm T_{94}(0.05/2)\sqrt{0.000298}$$

(f)

```r
anova(model3)
```

```
## Analysis of Variance Table
##
## Response: log(Butterfat)
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Breed      4 1.7033 0.42584  56.651 < 2.2e-16 ***
## Residuals 95 0.7141 0.00752
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(butterfat$Breed)
```

```
##          Ayrshire          Canadian          Guernsey Holstein-Fresian
##                20                20                20               20
##            Jersey
##                20
```

$$L = \frac{\mu_{1..} + \mu_{2..}}{2} - \frac{\mu_{3..} + \mu_{5..}}{2}$$

```
mean.1=mean(butterfat$logbutfat[butterfat$Breed=="Ayrshire"])
mean.2=mean(butterfat$logbutfat[butterfat$Breed=="Canadian"])
mean.3=mean(butterfat$logbutfat[butterfat$Breed=="Guernsey"])
mean.5=mean(butterfat$logbutfat[butterfat$Breed=="Jersey"])
(mean.1+mean.2-mean.3-mean.5)/2
```

```
## [1] -0.1843879
```

$$\hat{L} = \frac{1}{2}\bar{Y}_{1..} + \frac{1}{2}\bar{Y}_{2..} - \frac{1}{2}\bar{Y}_{3..} - \frac{1}{2}\bar{Y}_{5...} = -0.1843879$$

$$s^2_{\hat{L}} = \frac{0.00752}{10*2}\left((1/2)^2 + (1/2)^2 + (1/2)^2 + (1/2)^2\right) = 0.000376$$

So, the interval is

$$L \in (-0.1843879 \pm T_{95}(0.05/2)\sqrt{0.000376})$$