STAT 516 hw 6

Solutions

Chp 6 Ex 9

Here we read in the data. We can use `read.csv` with the `colClasses` argument as specified below to read in the data such that the columns are assigned the appropriate classes.

```
tensile <- read.csv("Data Tables 4th edition/Chapter 6/datatab_6_26.prn",
                    sep = " ",
                    colClasses = c("factor","numeric"))
```

We assume the model $Y_{ij} = \mu + S_i + \varepsilon_{ij}$ for $i = 1, \ldots, 4$ and $j = 1, \ldots, 4$, where μ is the overall mean,d $S_1, \ldots, S_4$ are independent Normal$(0, \sigma_S^2)$ random effects for the suppliers, and the ε_{ij} are Normal$(0, \sigma_\varepsilon^2)$ are error terms.

```
boxplot(tensile ~ supplier, data = tensile)
```
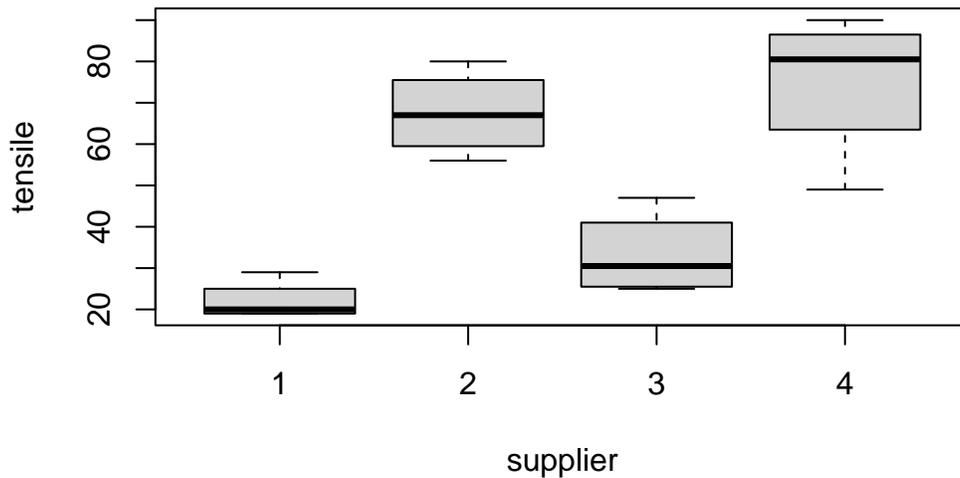

We can obtain the test statistic for testing H_0: $\sigma_S^2 = 0$ with the `anova()` function on the `lm()` output.

```r
lm_out <- lm(tensile ~ supplier, data = tensile)
anova(lm_out)
```

```
Analysis of Variance Table

Response: tensile
          Df Sum Sq Mean Sq F value    Pr(>F)
supplier   3 7978.2 2659.40  19.044 7.401e-05 ***
Residuals 12 1675.7  139.65
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The very small p value indicates strong evidence of significant variability in the tensile strength of the sheet metal from different suppliers.

The `lmer()` function provides estimates of the variance components.

```r
library(lmerTest)
```

```
Loading required package: lme4
```

```
Loading required package: Matrix
```

```
Attaching package: 'lmerTest'
```

```
The following object is masked from 'package:lme4':

    lmer
```

```
The following object is masked from 'package:stats':

    step
```

```r
lmer_out <- lmer(tensile ~ 1 + (1|supplier), data = tensile)
lmer_out
```

```
Linear mixed model fit by REML ['lmerModLmerTest']
Formula: tensile ~ 1 + (1 | supplier)
   Data: tensile
REML criterion at convergence: 128.2676
Random effects:
 Groups    Name         Std.Dev.
 supplier (Intercept) 25.10
 Residual             11.82
Number of obs: 16, groups:  supplier, 4
Fixed Effects:
(Intercept)
      49.44
```

We obtain $\hat{\sigma}_S = 25.10$ and $\hat{\sigma}_\varepsilon = 11.82$.

It also makes sense to check the residuals versus fitted values plot.

```
plot(lmer_out)
```
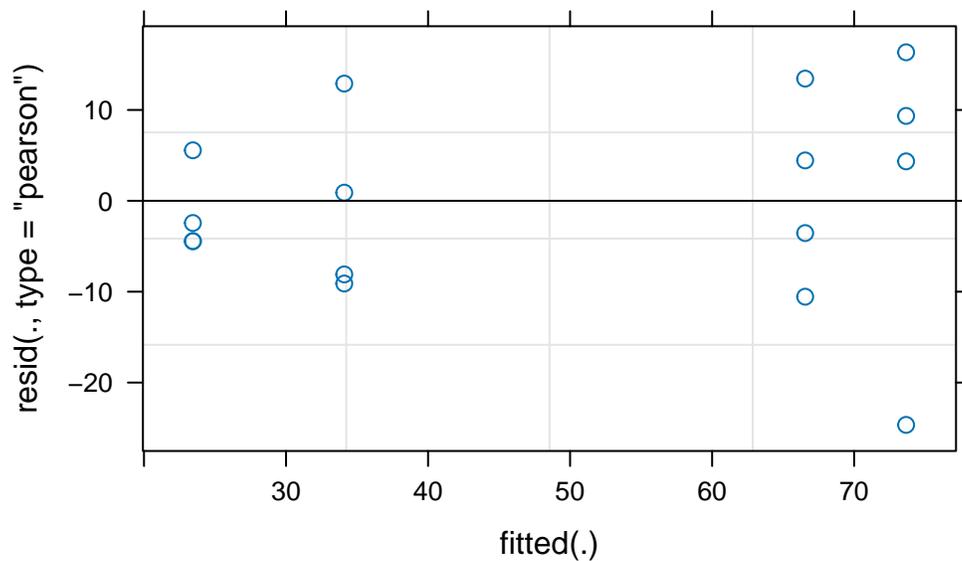

We see that there is some fanning in the residuals, with a wider residual variance at larger fitted values.

```
yhat <- predict(lmer_out)
ehat <- tensile$tensile - yhat
qqnorm(scale(ehat))
```

```
abline(0,1)
```
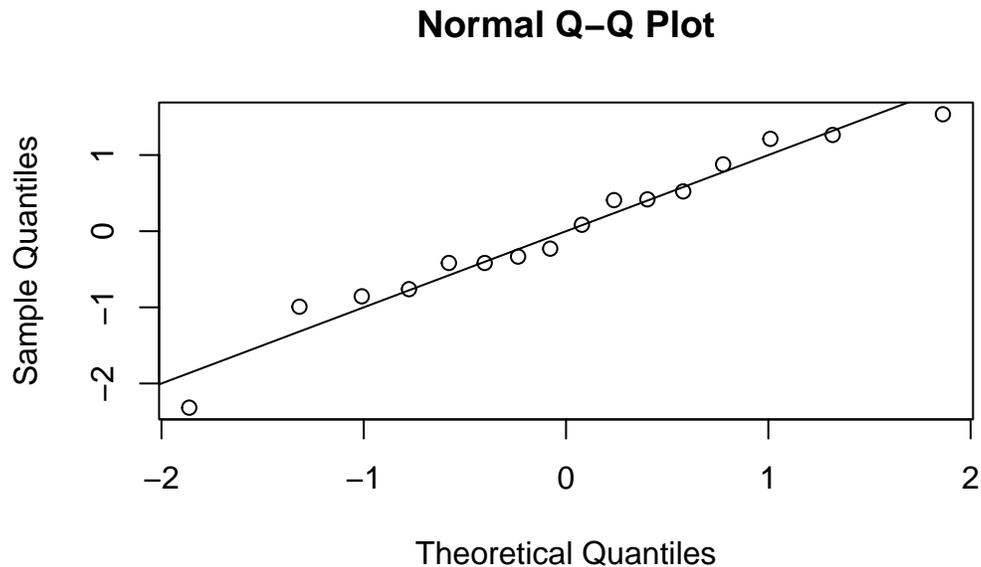
Normal Q–Q Plot

The Normal Q-Q plot of the residuals, however, does not show any great departure from Normality, so it is perhaps best not to transform the data.

Chp 10 Ex 4

Here we read in the data.

```
cotton <- read.csv("Data Tables 4th edition/Chapter 10/datatab_10_28.prn", sep = " ",
                   colClasses = c("factor","factor","factor","numeric","numeric"))
head(cotton)
```

```
  rep var trt brate mrate
1   1   1   1  2.98    42
2   1   1   2  3.08    42
3   1   2   1  2.75    42
4   1   2   2  2.75    42
5   1   3   1  2.83    42
6   1   3   2  3.11    42
```

The experiment is a randomized complete block split-plot design; the whole plot factor is the variety, the split-plot factor is the treatment, and the blocks are the replications.

Analysis with `mrate` as the response

First print the ANOVA table using `anova()` on the `lm()` output.

```
lm_out <- lm(mrate ~ var + trt + var:trt + rep + rep:var, data = cotton)
anova(lm_out)
```

```
Analysis of Variance Table

Response: mrate
          Df  Sum Sq Mean Sq F value Pr(>F)
var         5  160.94  32.188  0.6099 0.6935
trt         1    4.69   4.687  0.0888 0.7691
rep         3  127.23  42.410  0.8037 0.5081
var:trt     5  481.94  96.387  1.8265 0.1583
var:rep    15 1235.65  82.376  1.5610 0.1828
Residuals  18  949.87  52.771
```

We must correct the F statistic and p value for `var`.

```
F_var <- 32.188 / 82.376
p_var <- 1 - pf(F_var,5,15)
```

The correct F statistic is 0.3907449 and the correct p value is 0.8475051.

If we run `anova()` on the `lmer()` output using the R package `lmerTest`, we see that the p value in the output does not match our adjusted p value:

```
library(lmerTest)
lmer_out <- lmer(mrate ~ var + trt + var:trt + (1|rep) + (1|rep:var), data = cotton)
```

```
boundary (singular) fit: see help('isSingular')
```

```
anova(lmer_out)
```

```
Type III Analysis of Variance Table with Satterthwaite's method
         Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
var      112.17  22.434     5    18  0.4251 0.8252
trt        4.69   4.687     1    18  0.0888 0.7691
var:trt  481.94  96.387     5    18  1.8265 0.1583
```

It turns out this is because the REML estimate of the variance component associated with `rep` was equal to zero:

```
lmer_out
```

```
Linear mixed model fit by REML ['lmerModLmerTest']
Formula: mrate ~ var + trt + var:trt + (1 | rep) + (1 | rep:var)
   Data: cotton
REML criterion at convergence: 268.072
Random effects:
 Groups   Name        Std.Dev.
 rep:var  (Intercept) 3.387
 rep      (Intercept) 0.000
 Residual             7.264
Number of obs: 48, groups:  rep:var, 24; rep, 4
Fixed Effects:
(Intercept)          var2          var3          var4          var5          var6
      42.25         -3.25          1.25         -3.25          9.00         -5.25
       trt2     var2:trt2     var3:trt2     var4:trt2     var5:trt2     var6:trt2
      -2.75          5.50          2.50          4.00        -10.00         10.75
optimizer (nloptwrap) convergence code: 0 (OK) ; 0 optimizer warnings; 1 lme4 warnings
```

In this case, the `lmer()` package has a way of computing the p value for the whole plot factor which is different from ours.

We see from our adjusted p-value for the effect of `var` and the p values for `trt` and the `var-trt` interaction, that none of these have a significant effect on the response.

The interaction plot below may make it appear that the mean response of variety 5 under treatment 1 is higher than the response mean at other variety and treatment combinations, but, since the p values are large for all effects, we cannot conclude that the apparent difference in the interaction plot indicates any real difference in treatment means.

```
interaction.plot(cotton$var, cotton$trt, cotton$mrate)
```
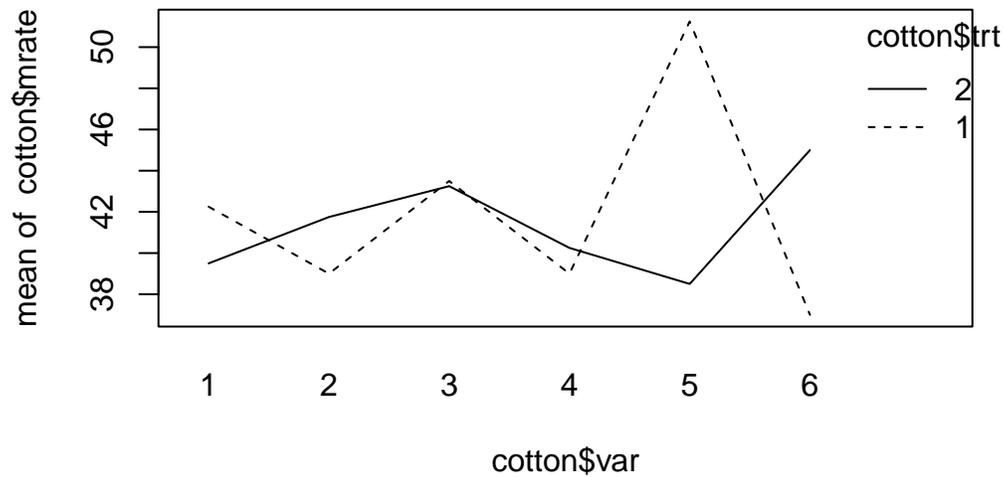
Analysis with `brate` as the response

We obtain the ANOVA table using `anova()` on the `lm()` output.

```
lm_out <- lm(brate ~ var + trt + var:trt + rep + rep:var, data = cotton)
anova(lm_out)
```

```
Analysis of Variance Table

Response: brate
          Df Sum Sq Mean Sq F value      Pr(>F)
var        5 4.1428 0.82856 16.7205 3.269e-06 ***
trt        1 1.0710 1.07102 21.6134 0.0001996 ***
rep        3 0.8161 0.27204  5.4898 0.0074012 **
var:trt    5 0.5896 0.11791  2.3795 0.0799096 .
var:rep   15 2.1228 0.14152  2.8560 0.0182033 *
Residuals 18 0.8920 0.04955
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We must adjust the F statistic and p value for the `var` effect.

```
F_var <- 0.82856 / 0.14152
p_var <- 1 - pf(F_var,5,15)
```

The correct F statistic is 5.8547202 and the correct p value is 0.0034086.

The `anova()` function on the `lmer()` output within the package `lmerTest` gives the same p value.

```
lmer_out <- lmer(brate ~ var + trt + var:trt + (1|rep) + (1|rep:var), data = cotton)
anova(lmer_out)
```

```
Type III Analysis of Variance Table with Satterthwaite's method
        Sum Sq Mean Sq NumDF DenDF F value     Pr(>F)
var     1.45058 0.29012     5    15  5.8546 0.0034088 **
trt     1.07102 1.07102     1    18 21.6133 0.0001996 ***
var:trt 0.58957 0.11791     5    18  2.3795 0.0799102 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

It appears that the whole-plot and split-plot factors both have significant main effects; the evidence for a interaction effect is fairly weak (p value 0.0799).

We make interaction plots:

```
interaction.plot(cotton$var,cotton$trt,cotton$brate)
```


From these plots, and due to the small p values for testing for main effects, we can claim that treatment 2 generally produces a higher response mean. Moreover, the apparent difference in the response mean across varieties reflects true differences in means. We will not perform any specific mean comparisons.

Chp 10 Ex 6

This is a two-way factorial design with randomized blocks; the panel is the block, and time and temperature are the fixed factors.

We first read in the data:

```
meat <- read.csv("Data Tables 4th edition/Chapter 10/datatab_10_30.prn", sep = " ",
                  colClasses = c("factor","factor","factor","numeric","factor"))
head(meat)
```

```
  OBS TIME PANEL QUAL TEMP
1   1    0     1 2.38    2
2   5    0     2 2.19    2
3   9    2     1 2.74    2
4  13    2     2 2.50    2
5  17    4     1 2.75    2
6  21    4     2 2.74    2
```

Now we look at the ANOVA table.

```
lm_out <- lm(QUAL ~ TIME + TEMP + TIME:TEMP + PANEL, data = meat)
anova(lm_out)
```

```
Analysis of Variance Table

Response: QUAL
          Df  Sum Sq Mean Sq F value    Pr(>F)
TIME       4 15.2717  3.8179 31.2138 4.109e-08 ***
TEMP       3 20.5439  6.8480 55.9863 1.269e-09 ***
PANEL      1  0.7784  0.7784  6.3640  0.020724 *
TIME:TEMP 12  4.8861  0.4072  3.3289  0.009518 **
Residuals 19  2.3240  0.1223
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The small p-values indicate significant interaction and main effects for the fixed factors as well as a significant variability between the two panels.

```
interaction.plot(meat$TIME,meat$TEMP,meat$QUAL)
```
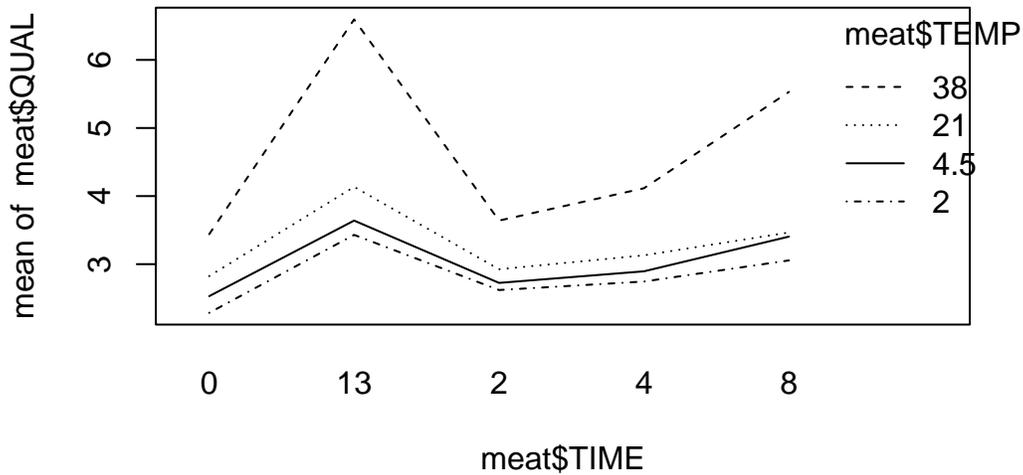
```
interaction.plot(meat$TEMP,meat$TIME,meat$QUAL)
```
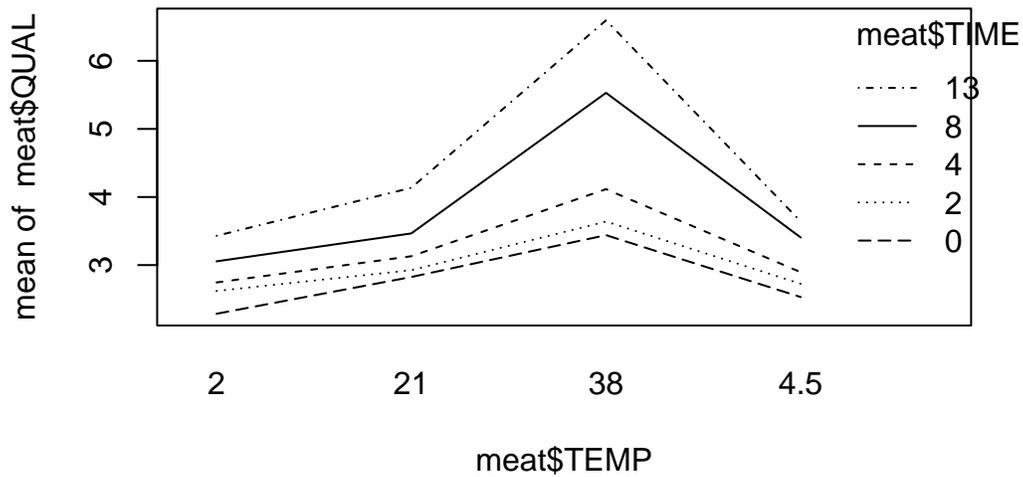

We will omit specific comparisons of means (though in order to be fully rigorous in our analysis, we should make specific comparisons) and simply remark, in light of the fact that there *are* significant main and interaction effects, that it appears higher temps and longer times generally improve the ratings, and that at higher temperatures, the longer time seemed to have a greater impact on the rating.