

HW4: Random Effects + Intro to Factorial Designs

MATH/STAT 571B

DUE: 3/22/2024 11:59pm

Homework Guidelines

Please submit your answers on Gradescope as a PDF with pages matched to question answers.

One way to prepare your solutions to this homework is with R Markdown, which provides a way to include mathematical notation, text, code, and figures in a single document. A template .Rmd file is available through D2L.

Make sure all solutions are clearly labeled, and please utilize the question pairing tool on Gradescope. You are encouraged to work together, but your solutions, code, plots, and wording should always be your own. Come and see me during office hours or schedule an appointment when you get stuck and can't get unstuck.

I. Mathematical Foundations [13 pts]

- (1) [7 pts] Suppose we are running a single-factor experiment with a levels modeled by $y_{ij} = \mu_i + \epsilon_{ij}$, $\epsilon_{ij} \sim N(0, \sigma^2)$ and are interested in testing $H_0: \mu_i = 0$, $i = 1, \dots, a$ using the usual F -statistic. We know that the power of the test depends directly on the "between variance", $V_b = \frac{1}{a} \sum_{i=1}^a (\mu_i - \bar{\mu})^2$, and that power increases monotonically with increasing V_b (i.e., the more spread out the group means are, the more power we will have to reject H_0). We would like to calculate the minimum power of our test under the assumption that at least one pair of means differs by an amount D (i.e., $|\mu_i - \mu_j| = D$ for some i, j). Thus, we need to know the minimum possible value of V_b consistent with our assumption.

Without loss of generality, assume that the index i on the means corresponds to a sorted, non-decreasing order such that $\mu_1 \leq \mu_2 \leq \dots \leq \mu_a$.

- (a) [2 pts] Provide an argument for why V_b is minimized when the absolute difference $|\mu_i - \mu_j| = D$ occurs between the largest and smallest means (i.e., $\mu_a - \mu_1 = D$).
- (b) [3 pts] Show that V_b is minimized when $\mu_i = \bar{\mu}$ for all $1 < i < a$. That is, the smallest value of V_b occurs when all of the means besides μ_1 and μ_a are equal to $\bar{\mu} = \frac{1}{a} \sum_{i=1}^a \mu_i$.
- (c) [2 pts] Show that the minimum value for V_b is $\frac{D^2}{2a}$.

answers on last page

- (2) [6 pts] Recall Henry's example about coffee quality from the first day of class. In the story Henry told, there were three factors that were thought to impact a person's rating of their coffee:
- (i) Cafe (4 levels: Luce, Slot Canyon, Starbucks, and Snakes and Lattes),
 - (ii) Time of day (2 levels: AM and PM), and
 - (iii) Faculty member (4 levels: HS, XT, JW, and HZ).

Assume that the results of the coffee experiment are expected to be shared with all faculty in the Statistics GIDP to help them optimize their coffee intake. Determine whether each factor should be treated as a fixed or random effect and explain your reasoning. **(Cafe) and (Time of Day) should be considered**

fixed effects because the levels can be set. The choice of cafe and the time of day can be set for repeated experiments in the future, and we do care about the possible cafe effect on the coffee. (Faculty member) should be considered a random effect because future experiments may not be able to choose the exact same faculty members, and they can be assumed to come from some normal population with some standard deviation. The levels are not as important since the coffee testers are not a factor for which you can set levels.

II. Applications [27 pts]

(3) [7 pts] (DAE 5.1)

(a) [3 pts] Give exact *p*-values.

DF values - A: 1 df, B = AB interaction: 2 df (since $ab_df = b_df * a_df$ and $a_df = 1$)

SS interaction - 45.348

MS values - A: 0.322, AB interaction: 22.674

F values - A: 0.03668554, AB interaction: 2.58325453

p values - A: 0.8513093, B: 0.03307743, AB interaction: 0.1166798

```
# DF
leftover_df <- 17 - 12 - 1
leftover_df / 2

## [1] 2

# SS (orthogonal since factorial design)
ss_interaction <- 231.551 - 105.327 - 80.554 - 0.322
ss_interaction

## [1] 45.348

# MS
ss_interaction / 2

## [1] 22.674

# F values
f_A <- 0.322 / 8.7773
f_AB <- 22.674 / 8.7773
c(f_A, f_AB)

## [1] 0.03668554 2.58325453

# p values
pf(f_A, 1, 12, lower.tail=F)

## [1] 0.8513093
```

```
f_B <- 4.59
pf(f_B, 2, 12, lower.tail=F)
```

```
## [1] 0.03307743
```

```
pf(f_AB, 2, 12, lower.tail=F)
```

```
## [1] 0.1166798
```

(b) [1 pts]

There are 3 levels for factor B (since B has 2 df)

(c) [1 pts]

18 / 6 = 3 replicates

(d) [2 pts]

Based on the p-values, it appears that factor B has a significant effect on the response at a 0.05 alpha level. It has a p-value of 0.033 which would be lower than an $\alpha = 0.05$. At least one level varies significantly from the others in factor B.

(4) [6 pts] (DAE 5.6)

```
tv <- read.csv("./Datasets/Q5-6.csv")
tv$Phosphorous <- as.factor(tv$Phosphorous)
tv$Glass <- as.factor(tv$Glass)
tv_aov <- aov(Current ~ Phosphorous*Glass, data=tv)
summary(tv_aov)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Phosphorous    2    933     467   8.842 0.00436 **
## Glass          1 14450  14450 273.789 1.26e-09 ***
## Phosphorous:Glass 2    133      67   1.263 0.31780
## Residuals     12    633      53
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(a) [2 pts]

The glass factor appears to have a significant effect on tv brightness with a p-value of 1.64e-08 which is less than $\alpha = 0.05$.

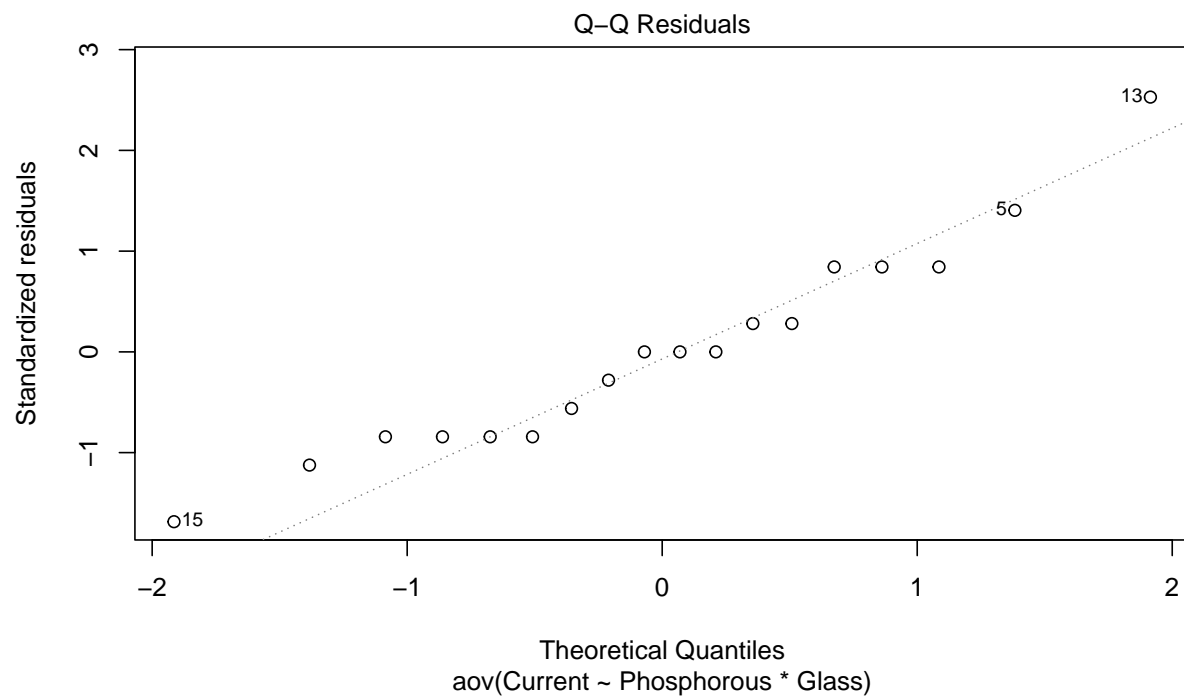
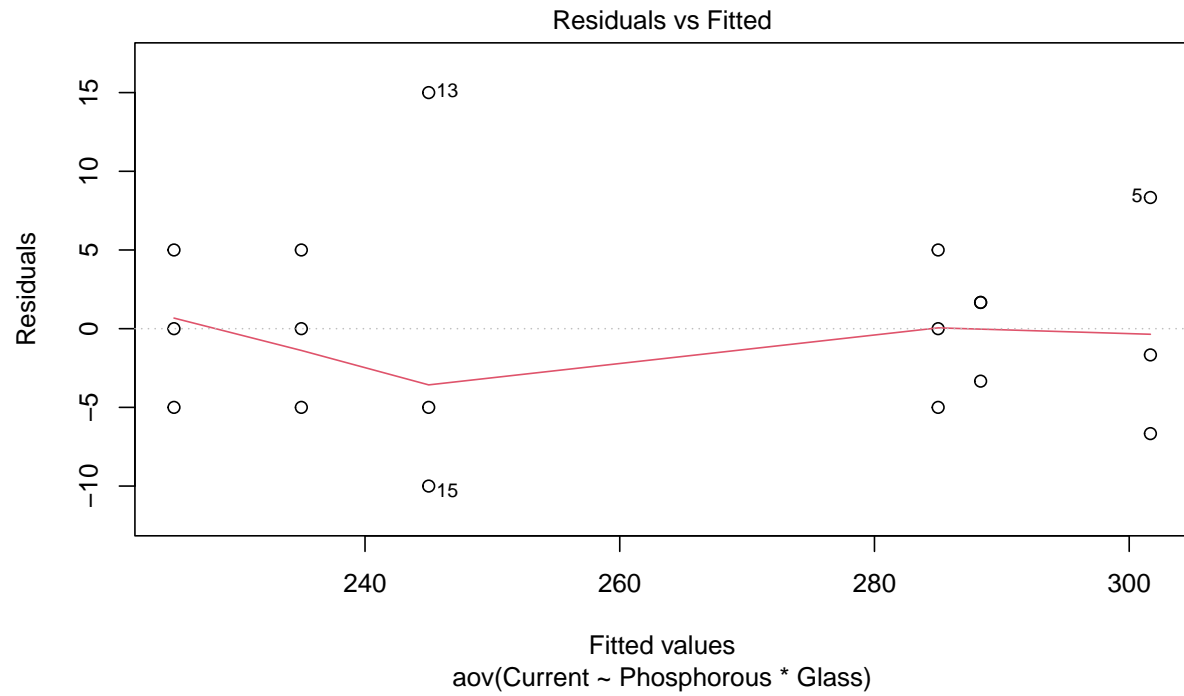
(b) [2 pts]

With p-value = 0.288, there does not appear to be an interaction between phosphorous and glass factors at an alpha level of 0.05.

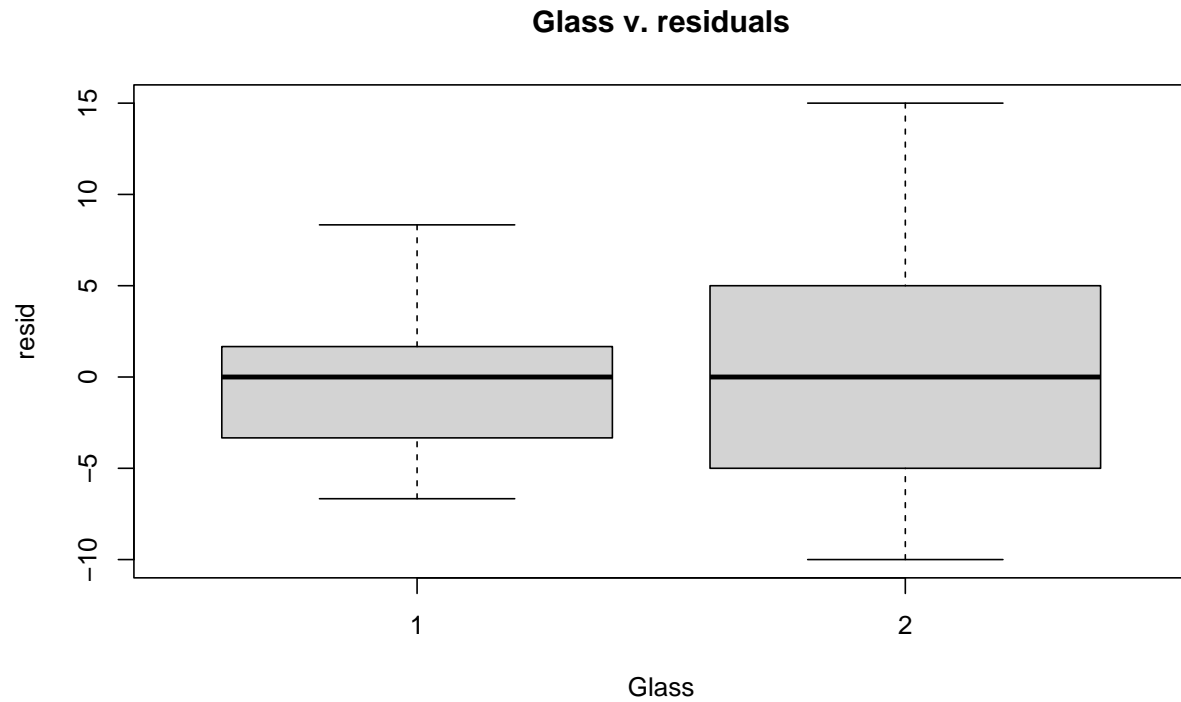
(c) [2 pts]

The plots below show that the data is pretty normally distributed (QQ plot), the residuals seem to have constant variance with only slightly larger residuals for phosphorous type 3.

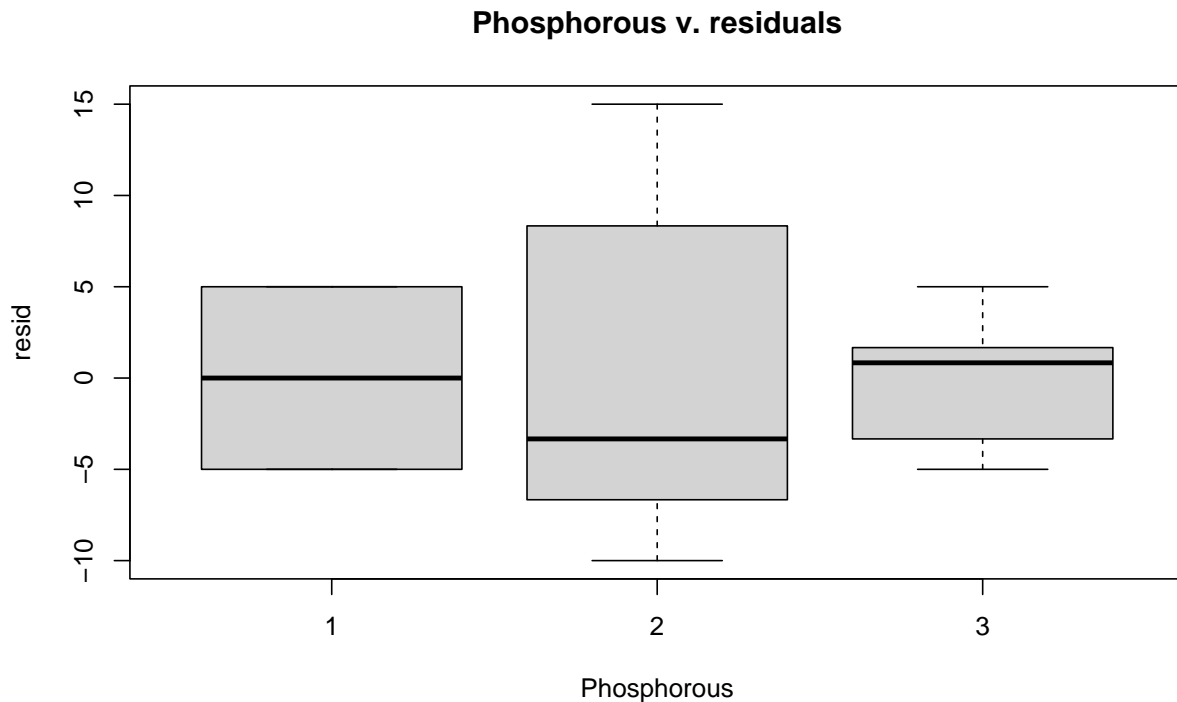
```
plot(tv_aov, c(1,2))
```



```
plot(tv
     $Glass, resid(tv_aov), main = 'Glass v. residuals', xlab='Glass', ylab='resid')
```



```
plot(
  tv$Phosphorous,
  resid(tv_aov),
  main = 'Phosphorous v. residuals',
  xlab = 'Phosphorous',
  ylab = 'resid'
)
```



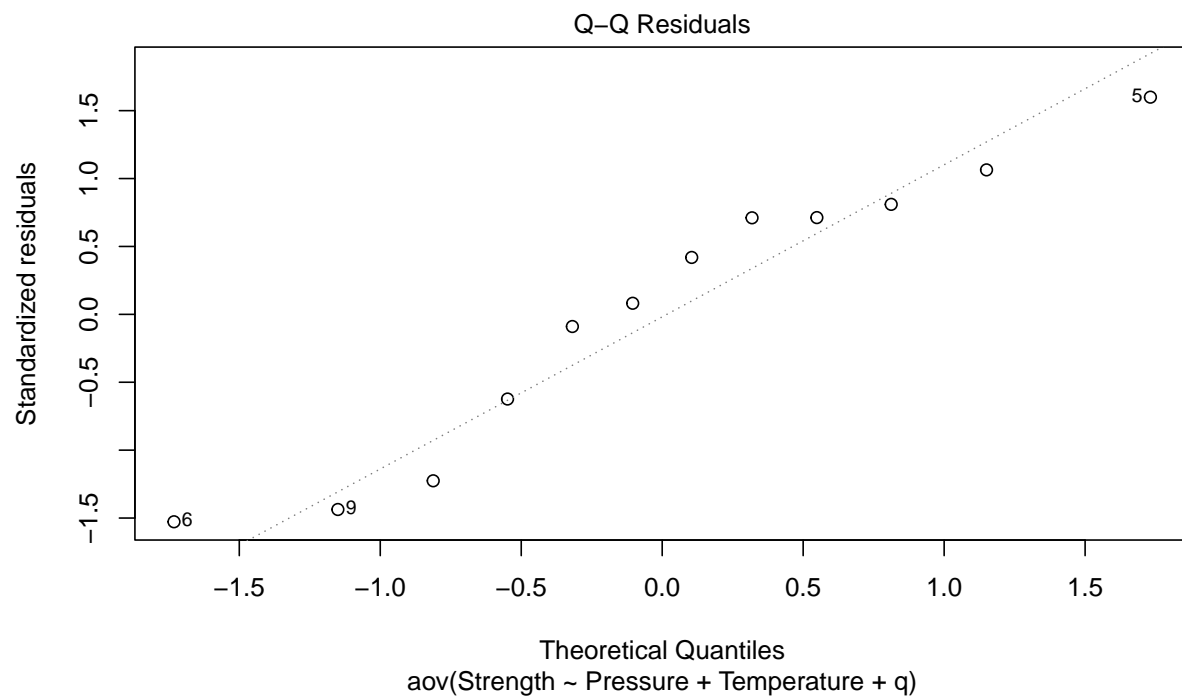
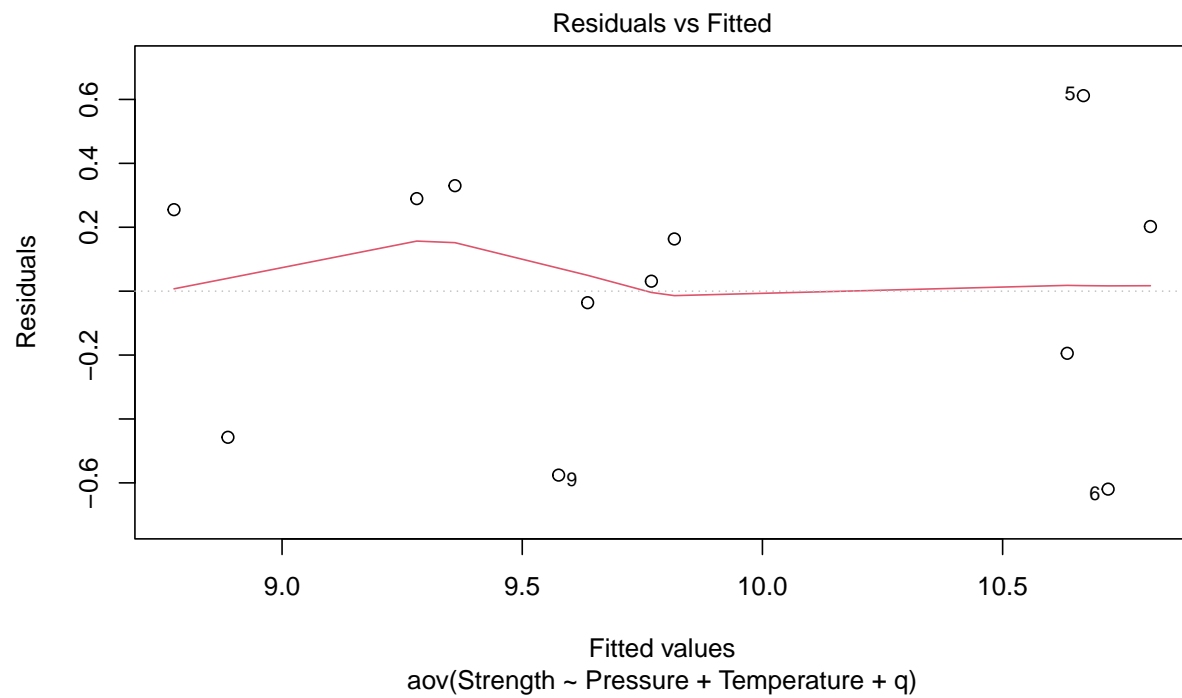
- (5) [7 pts] (DAE 5.16) Use Tukey's test for additivity. Your analysis should include: relevant hypothesis tests and a check for violations of basic assumptions. There appears to be no additivity in the factors (Pressure and Temperature) at an $\alpha = 0.05$. The p-value for the q term is 0.2974 which is not a significant interaction. The Temperature factor appears to be the only significant factor with a p-value of 0.0356. The qq plot shows normal residuals, and the resid v. fitted, resid v. Temperature, and resi v. Pressure all show constant residuals.

```
adhesive <- read.csv("../Datasets/Q5-16.csv")
adhesive$Pressure <- as.factor(adhesive$Pressure)
adhesive$Temperature <- as.factor(adhesive$Temperature)

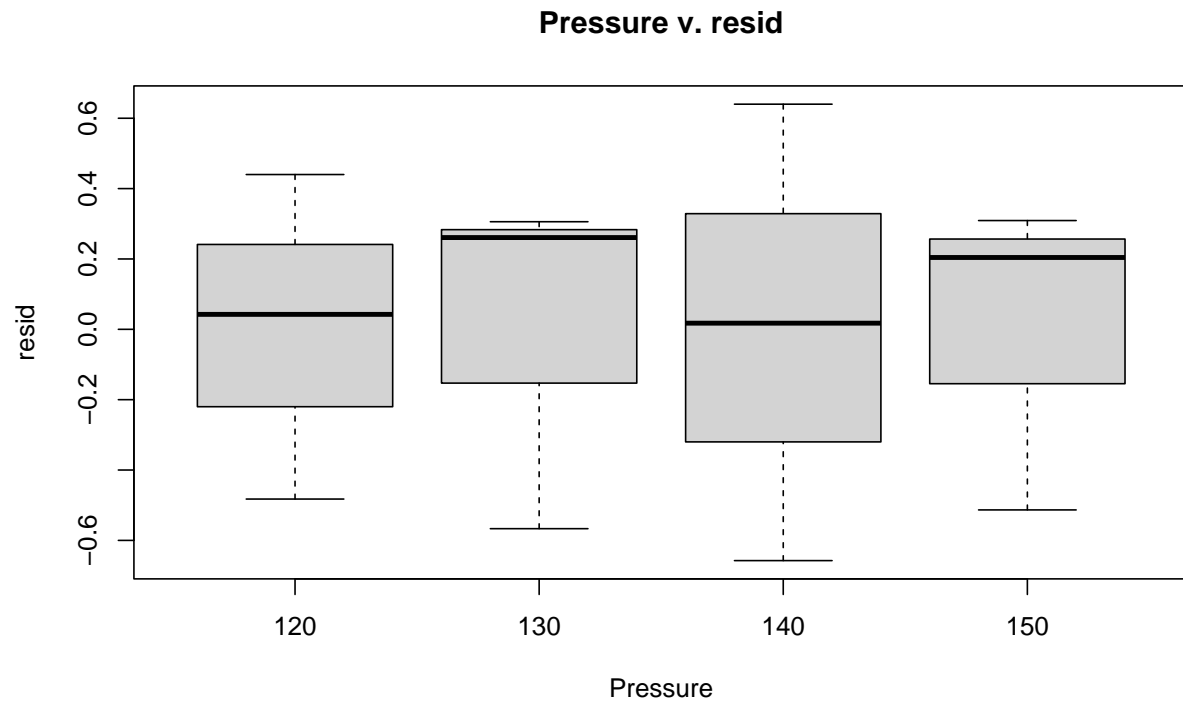
adhesive_aov <- aov(Strength ~ Pressure + Temperature, data = adhesive)
adhesive$q <- adhesive_aov$fitted.values^2
adhesive_tukey_add <- aov(Strength ~ Pressure + Temperature + q, data = adhesive)
summary(adhesive_tukey_add)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## Pressure    3  0.581   0.1936    0.581  0.6524
## Temperature  2  4.658   2.3288    6.996  0.0356 *
## q           1  0.489   0.4895    1.470  0.2794
## Residuals    5  1.664   0.3329
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

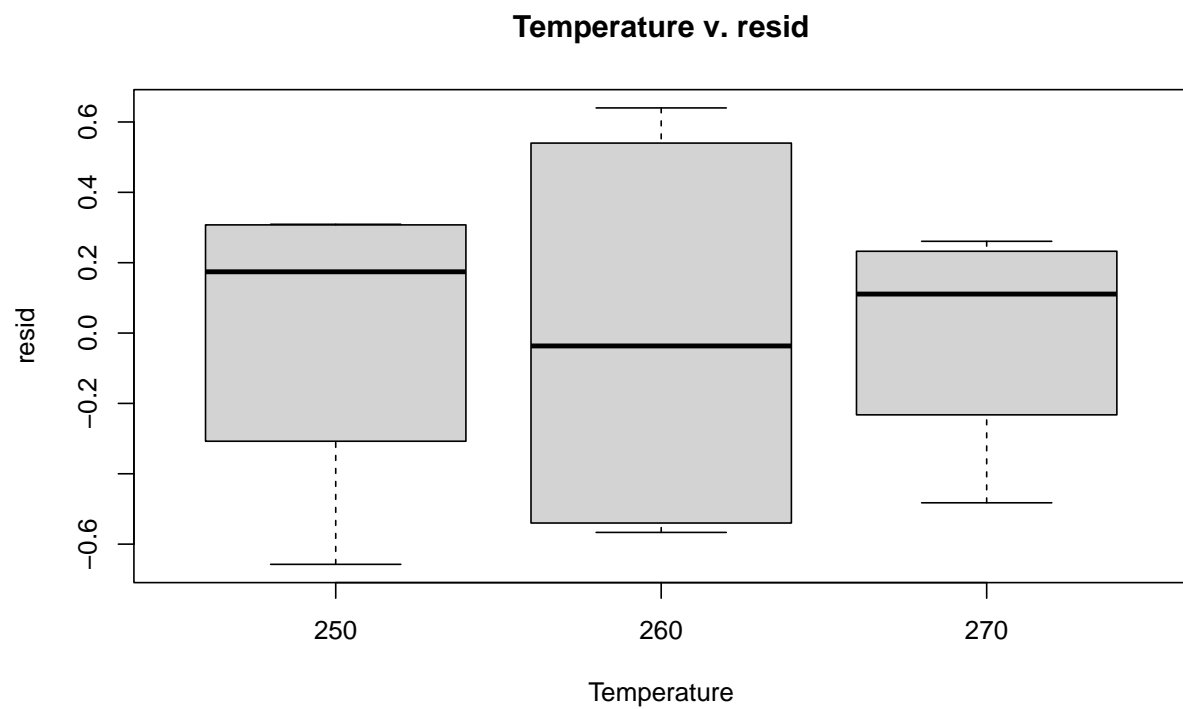
```
# Model assumptions
plot(adhesive_tukey_add, c(1,2))
```



```
plot(adhesive$Pressure, resid(adhesive_aov), xlab="Pressure", ylab="resid", main="Pressure v. resid")
```



```
plot(adhesive$Temperature, resid(adhesive_aov), xlab="Temperature", ylab="resid", main="Temperature v. .")
```



(6) [7 pts] (DAE 5.21) *Perform your analysis assuming Day is a random effect. Your analysis should include:*

relevant tests for each main effect and their interaction, a test for the null hypothesis of zero random effect, and a check for violations of basic assumptions.

At $\alpha = 0.05$, both Temperature and Pressure appear to be significant factors with p-values 2.778e-06 and 0.03599, respectively. However, their interaction is not significant (p-value = 0.17331). The Day random effect also appears to be significant at a p-value of 0.003851. The diagnostics reveal no violations to the basic assumptions (qq plot shows normal residuals and residual plots (against fitted, temperature, pressure, day) indicate constant variance).

```
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
```

```
chemicals <- read.csv("./Datasets/Q5-21.csv")
chemicals$Pressure <- as.factor(chemicals$Pressure)
chemicals$Day <- as.factor(chemicals$Day)
chemicals$Temperature <- as.factor(chemicals$Temperature)
```

```
chemicals_RE <- lmer(Yield ~ Pressure * Temperature + (1 | Day), data = chemicals)
anova(chemicals_RE)
```

```
## Type III Analysis of Variance Table with Satterthwaite's method
```

```
##               Sum Sq Mean Sq NumDF DenDF F value    Pr(>F)
## Pressure           5.508   2.754      2      8  5.1838  0.03599 *
## Temperature       99.854  49.927      2      8 93.9806 2.778e-06 ***
## Pressure:Temperature  4.452   1.113      4      8  2.0952  0.17331
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
ranova(chemicals_RE)
```

```
## ANOVA-like table for random-effects: Single term deletions
```

```
##
```

```
## Model:
```

```
## Yield ~ Pressure + Temperature + (1 | Day) + Pressure:Temperature
```

```
##               npar logLik   AIC    LRT Df Pr(>Chisq)
```

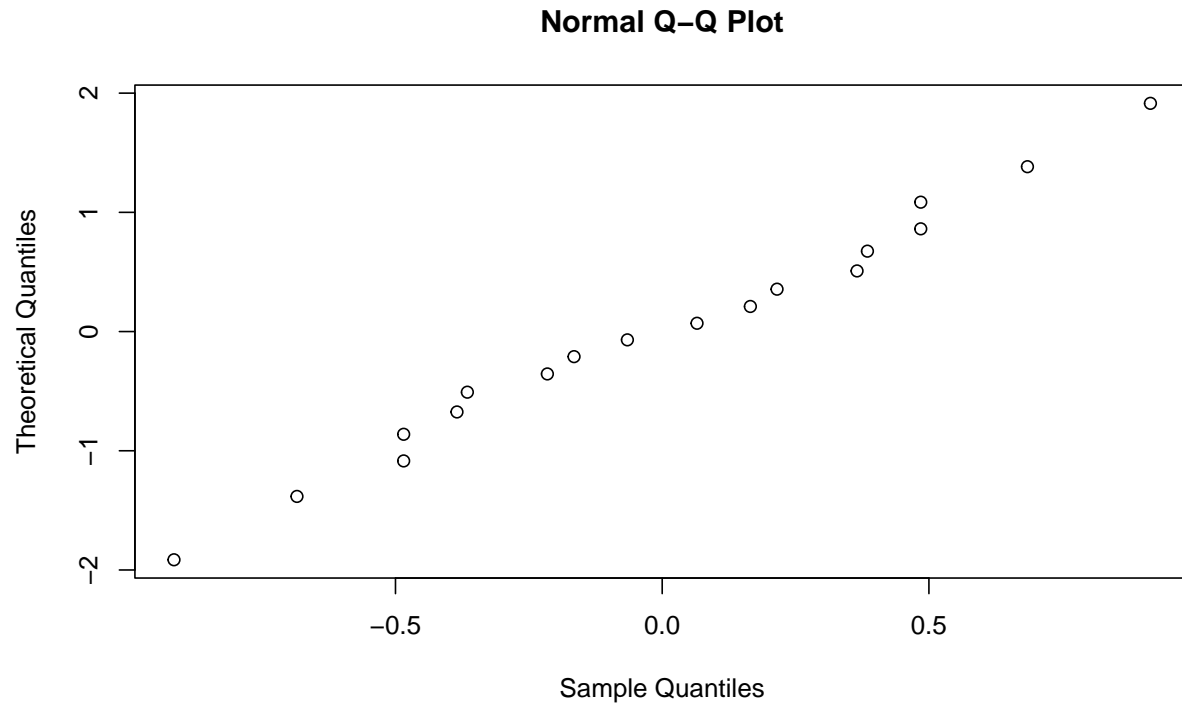
```
## <none>          11 -14.642 51.284
```

```
## (1 | Day)       10 -18.819 57.637 8.3527  1  0.003851 **
```

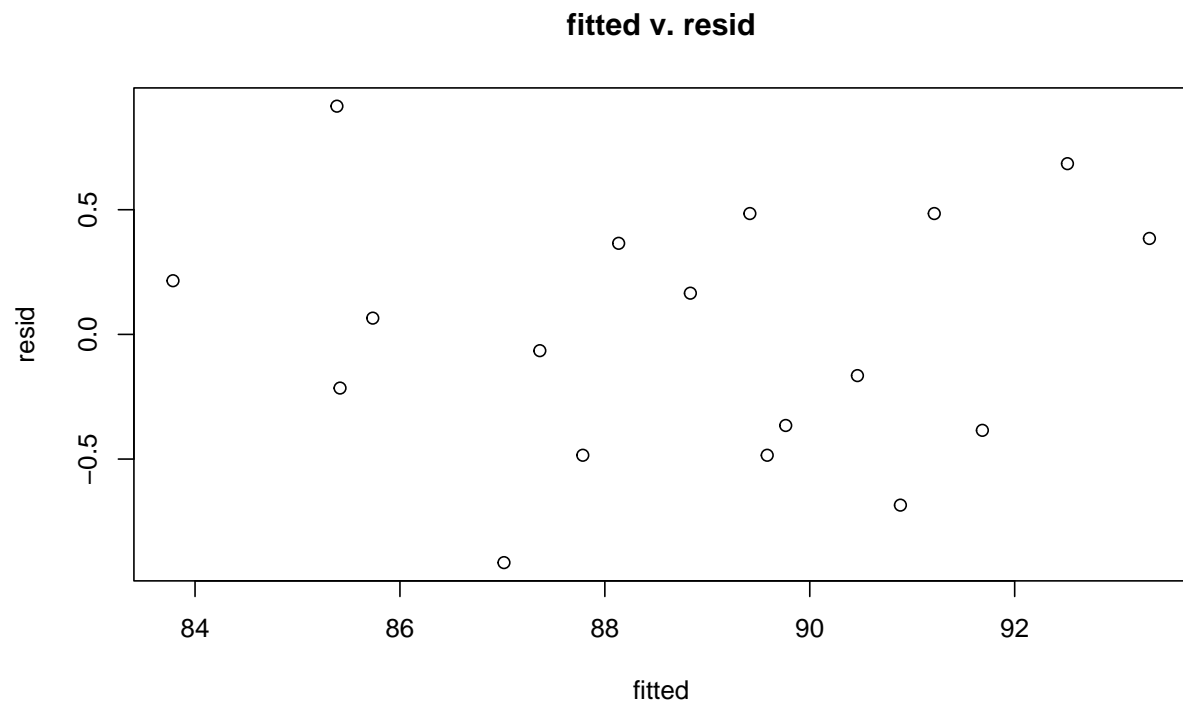
```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

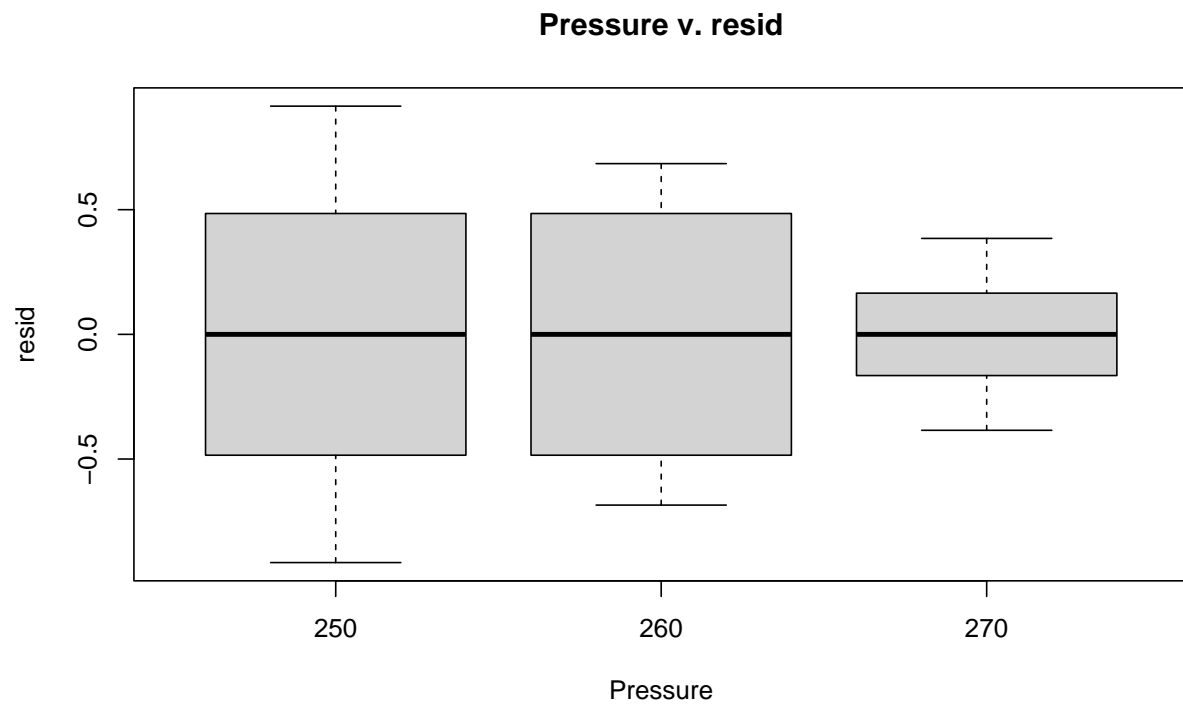
```
# Check for model assumptions  
resid <- resid(chemicals_RE)  
qqnorm(resid, datax = T)
```



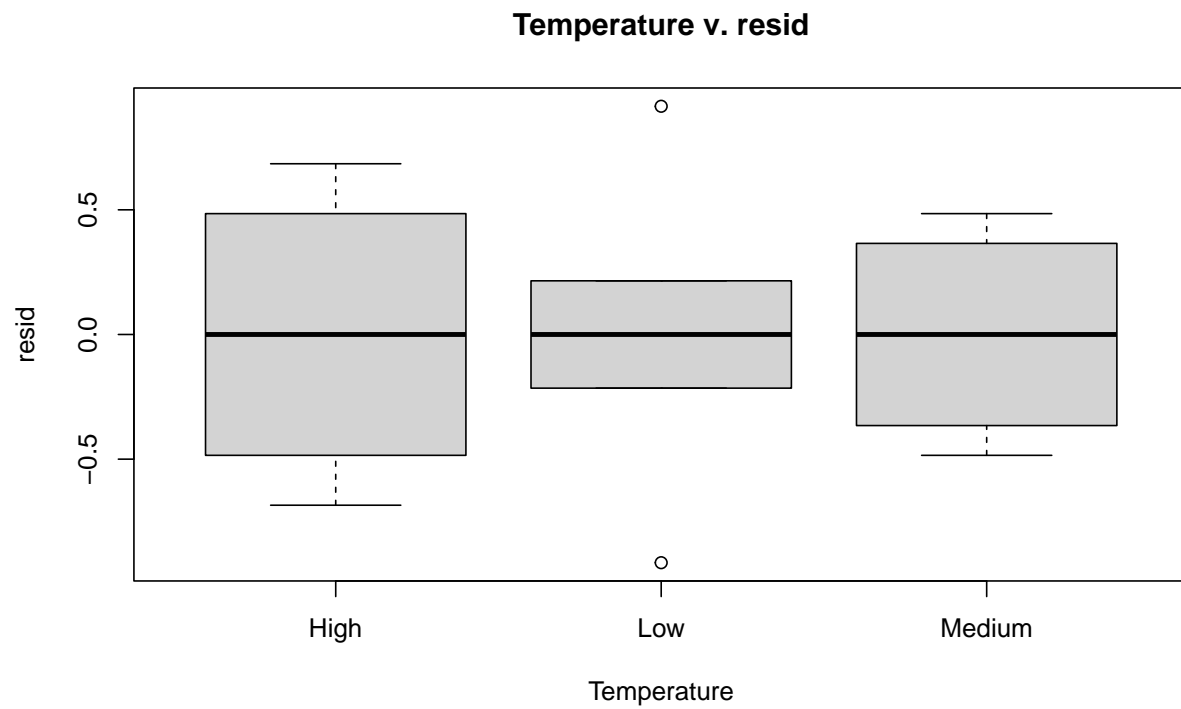
```
plot(fitted(chemicals_RE), resid, main="fitted v. resid", xlab="fitted", ylab="resid")
```



```
plot(chemicals$Pressure, resid, main="Pressure v. resid", xlab="Pressure", ylab="resid")
```



```
plot(chemicals$Temperature, resid, main="Temperature v. resid", xlab="Temperature", ylab="resid")
```



```
plot(chemicals$Day, resid, main="Day v. resid", xlab="Day", ylab="resid")
```

