101b_hw01_Daren_Sathasivam

Daren Sathasivam

2024-04-14

1. Problem 3.7

- Blank 1(Factor ~ DF):
  $25 + x = 29$
  $x = 4$

- Blank 2(Factor ~ SS):
  $SS_{total} = SS_{Factor} + SS_{Error}$
  $1174.24 = 186.53 + SS_{Factor}$
  $SS_{Factor} = 987.71$

- Blank 3(Error ~ MS):
  $MS_{Error} = \dfrac{SS_{Error}}{DF_{Error}}$
  $MS_{Error} = \dfrac{186.53}{25}$ $MS_{Error} = 7.4612$

- Blank 4(Factor ~ F):
  $F = \dfrac{MS_{Factor}}{MS_{Error}}$
  $F = \dfrac{246.93}{7.4612}$
  $F = 33.0952$

- Blank 5(Factor ~ P):

```
F_stat <- 33.0952
DF_fac <- 4
DF_err <- 25
p_value <- pf(F_stat, DF_fac, DF_err, lower.tail = FALSE)
print(p_value)
```

```
## [1] 1.184659e-09
```

$P = 1.184659e - 09$
The p-value is far smaller than 0.05, indicating that the factor is significant.

2. Problem 3.9-3.10 (skip 3.9b and use boxplot for 3.9f)

3.9

```
mix <- c(rep(1, 4), rep(2, 4), rep(3, 4), rep(4, 4))
tensile <- c(3129, 3000, 2865, 2890, 3200, 3300, 2975, 3150, 2800, 2900, 2985, 3050, 2600, 2700, 2600,
mix <- as.factor(mix)
# mix
# tensile
```

a. Test the hypothesis that mixing technique affect the strenght of the cement. Use $\alpha = 0.05$

```
# a
m1 <- aov(tensile ~ mix)
summary(m1)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## mix           3 489740  163247   12.73 0.000489 ***
## Residuals    12 153908   12826
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- The p-value is 0.000489, meaning we reject the null hypothesis. This indicates that the type of mixing technique does not affect the tensile strength of the cement.

c. Use the Fisher LSD method with $\alpha = 0.05$ to make comparisons between pairs of means

```
# c
# install.packages("DescTools")
library(DescTools)
PostHocTest(m1, method = "lsd")
```

```
##
##   Posthoc multiple comparisons of means : Fisher LSD
##     95% family-wise confidence level
##
## $mix
##          diff     lwr.ci     upr.ci    pval
## 2-1    185.25    10.77016  359.72984  0.0392 *
## 3-1    -37.25 -211.72984  137.22984  0.6501
## 4-1  -304.75 -479.22984 -130.27016  0.0025 **
## 3-2  -222.50 -396.97984  -48.02016  0.0167 *
## 4-2  -490.00 -664.47984 -315.52016 5.2e-05 ***
## 4-3  -267.50 -441.97984  -93.02016  0.0059 **
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Technique 2 shows the greatest tensile strength and is significantly better than both technique 1 and 3
- Technique 4 is the lowest strength and significantly underperforms relative to the other techniques.
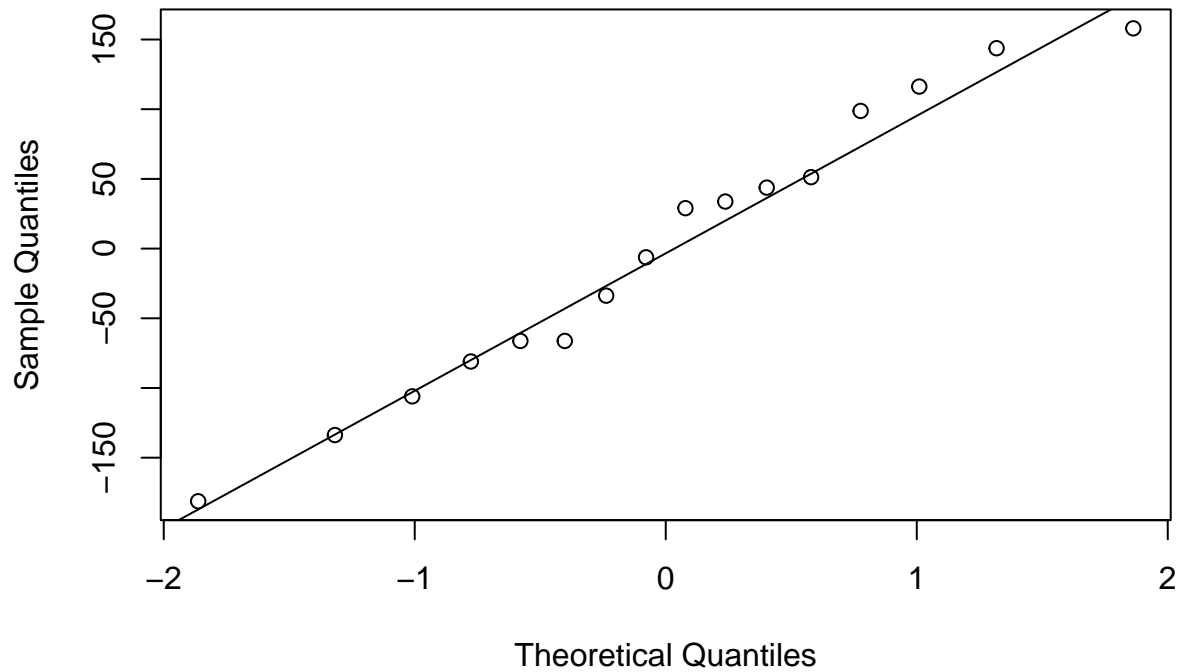
d. Construct a normal probability plot of the residuals. What conclusion would you draw about the validity of the normality assumption?

```
# d
res1 <- m1$residuals
qqnorm(res1); qqline(res1)
```
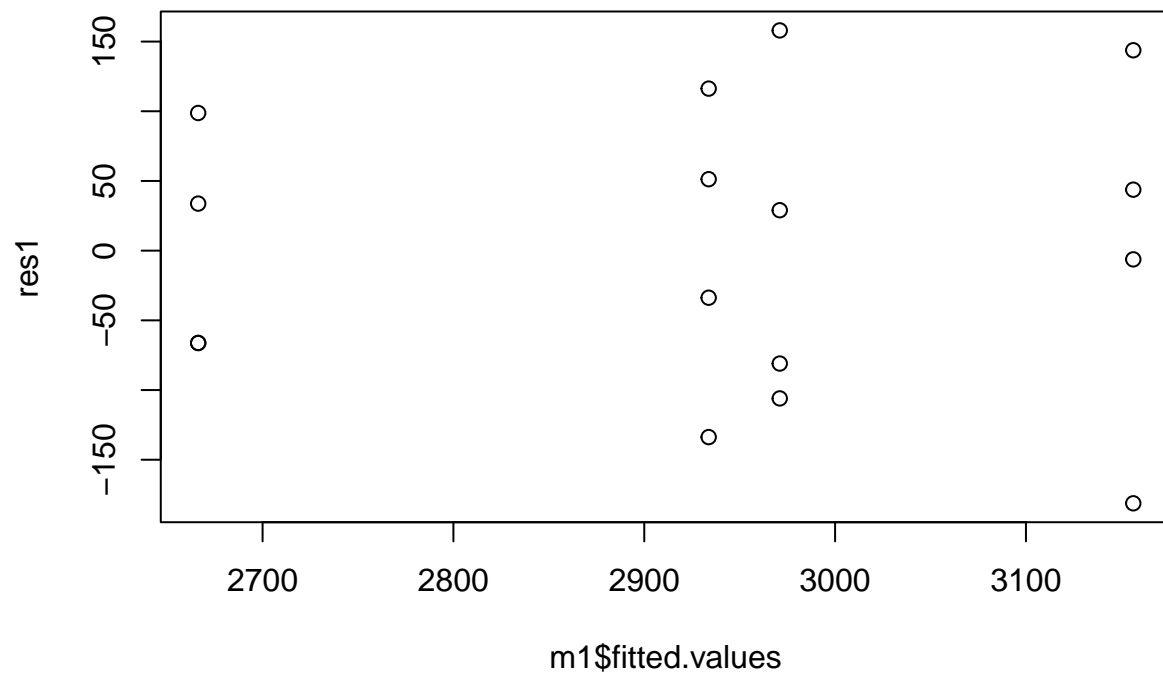
2

Normal Q–Q Plot

- Data has satisfied the normality assumption due to the observed points proximity to the qq line.

**e. Plot the residuals versus tensile strength. Comment on the plot**

```
plot(m1$fitted.values, res1)
```


- The plot displays groups of points at roughly 2600, 2925, 2975, and 3200. Due to the points clustering around these points, this may suggest a difference in variances across different techniques.

f. Prepare a scatterplot of the residuals to aid the interpretation of the results of this experiment.

```
boxplot(tensile ~ mix)
```
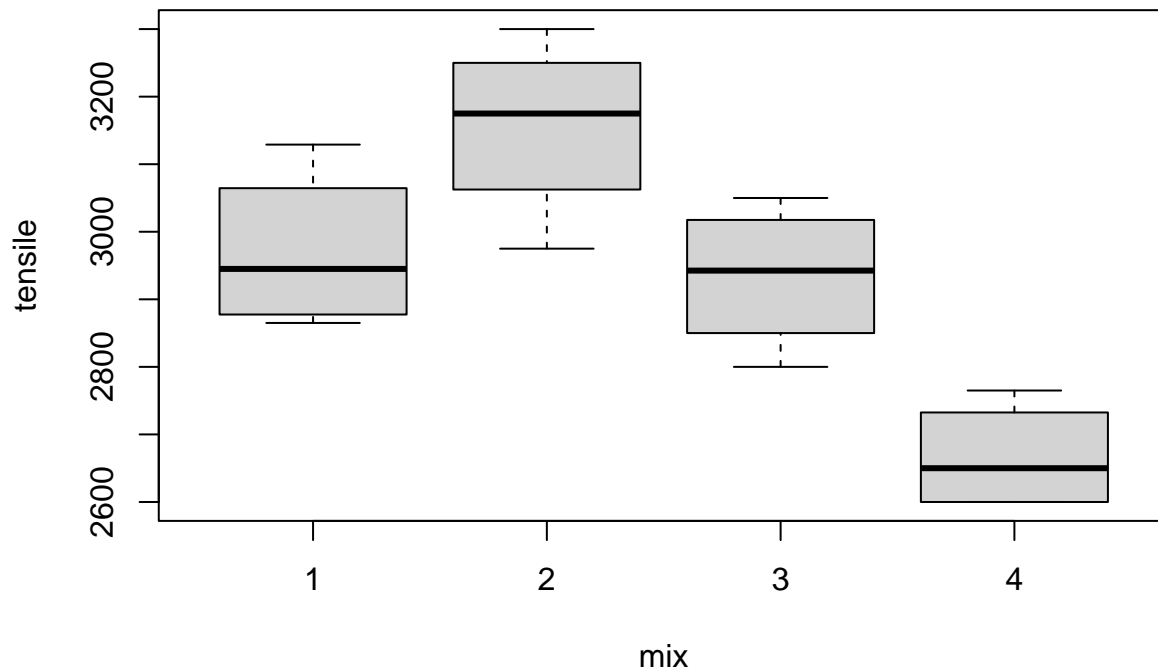

- After observing the boxplot, we can confirm our earlier interpretation where technique 2 is the strongest and technique 4 as the weakest technique.

3.10

a.Rework part(c) or fproblem 3.9 using Tukey's test with $\alpha = 0.05$. Do you get the same conclusions from Tukey's test that you did from the graphical procedure and/or the Fisher LSD method?

```
TukeyHSD(m1)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = tensile ~ mix)
##
## $mix
##         diff        lwr        upr     p adj
## 2-1  185.25  -52.50029  423.00029 0.1493561
## 3-1  -37.25 -275.00029  200.50029 0.9652776
## 4-1 -304.75 -542.50029  -66.99971 0.0115923
## 3-2 -222.50 -460.25029   15.25029 0.0693027
## 4-2 -490.00 -727.75029 -252.24971 0.0002622
## 4-3 -267.50 -505.25029  -29.74971 0.0261838
```

- When using Tukey's test, it can be observed that the greatest difference between technique 1 and the others is technique 2. Additionally, the weakest in comparison to technique 1 is technique 4.

b. Explain the difference between the Tukey and Fisher procedures

- Tukey's method provides a more precise estimate than the Bonferroni's and tends to be more conservative in comparison to Fisher's LSD. Fisher's LSD is a one-at-a-time method that does not control the experiment-wise or family error rate.

3. Problem 3.25

```
noise <- c(19, 20, 19, 30, 8, 80, 61, 73, 56, 80, 47, 26, 25, 35, 50, 95, 46, 83, 78, 97)
noise
```

```
##  [1] 19 20 19 30  8 80 61 73 56 80 47 26 25 35 50 95 46 83 78 97
```

```
design <- c(rep(1, 5), rep(2, 5), rep(3, 5), rep(4, 5))
design <- as.factor(design)
```

a. Is the same amount of noise present for all four designs? Use $\alpha = 0.05$

```
m2 <- aov(noise ~ design)
summary(m2)
```

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## design       3  12042    4014   21.78 6.8e-06 ***
## Residuals   16   2949     184
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Since the p-value is 6.8e-06, we reject the null hypothesis. This suggests that there are different noise levels between the different designs.

b. Analyze the residuals from this experiment. Are the analysis of variance assumptions satisfied?

```
res2 <- m2$residuals
par(mfrow = c(1, 2))
qqnorm(res2); qqline(res2)
plot(m2$fitted.values, res2)
```

Normal Q–Q Plot
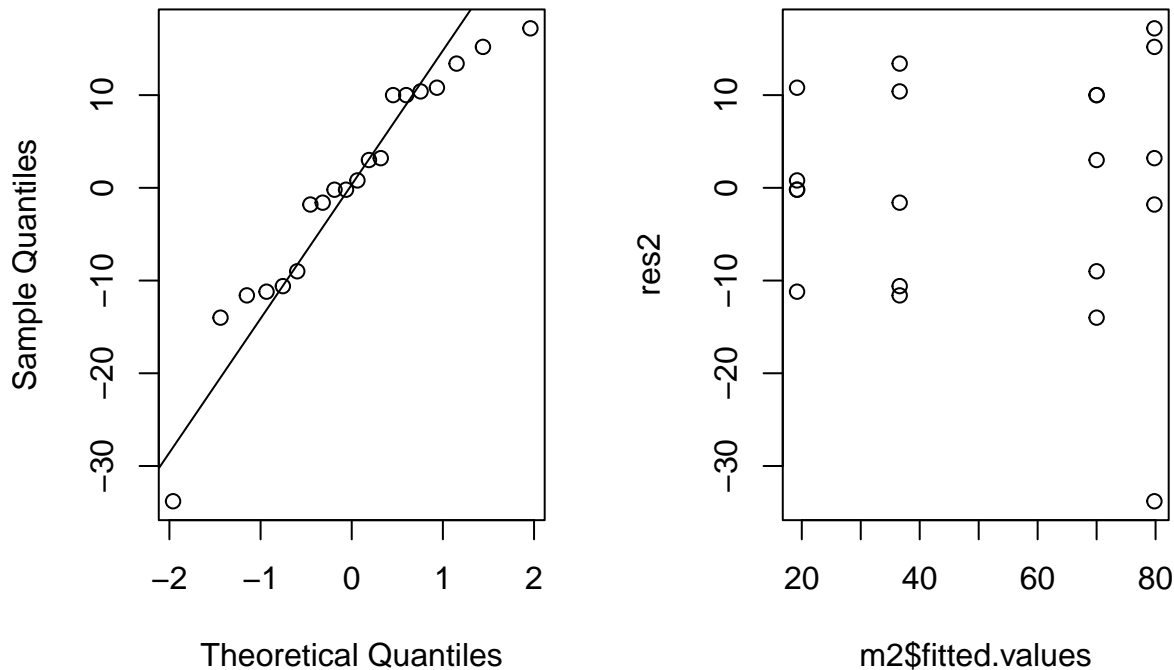
- The points on the residual plot can be observed to be relatively near the line, indicating that the model does not violate the normality assumption. Additionally, the residuals vs fitted values displays a grouping of values, indicating that it does not violate the variance assumption.

c. Which circuit design would you select for use? Low noise is best

```
PostHocTest(m2, method = "lsd")
```

```
##
##   Posthoc multiple comparisons of means : Fisher LSD
##     95% family-wise confidence level
##
## $design
##       diff      lwr.ci     upr.ci     pval
## 2-1   50.8  32.5984185   69.00158 2.2e-05 ***
## 3-1   17.4  -0.8015815   35.60158 0.05971 .
## 4-1   60.6  42.3984185   78.80158 2.7e-06 ***
## 3-2 -33.4 -51.6015815  -15.19842 0.00130 **
## 4-2    9.8  -8.4015815   28.00158 0.27050
## 4-3   43.2  24.9984185   61.40158 0.00012 ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Circuit design 1 is the best as the difference between design 1 and all other designs are positive, indicating that design 1 has the lowest noise.

4. Problem 3.29

- Hint for (c): Consider square root transformation of the data to satisfy the model assumption.

```
method <- c(rep(1, 5), rep(2, 5), rep(3, 5))
method <- as.factor(method)
count <- c(31, 10, 21, 4, 1, 62, 40, 24, 30, 35, 53, 27, 120, 97, 68)
```

a. Do all methods have the same effect on mean particle count?

```
m3 <- aov(count ~ method)
summary(m3)
```

```
##              Df Sum Sq Mean Sq F value  Pr(>F)
## method        2   8964    4482   7.914 0.00643 **
## Residuals    12   6796     566
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Since the p-value is 0.00643, we reject the null hypothesis. This indicates that not all of the methods are equally effective in reducing particle count.

b. Plot the residuals versus the predicted response. Construct a normal probability plot of the residuals. Are there potential concerns about the validity of the assumptions?

```
res3 <- m3$residuals
par(mfrow = c(1, 2))
qqnorm(res3); qqline(res3)
plot(m3$fitted.values, res3)
```
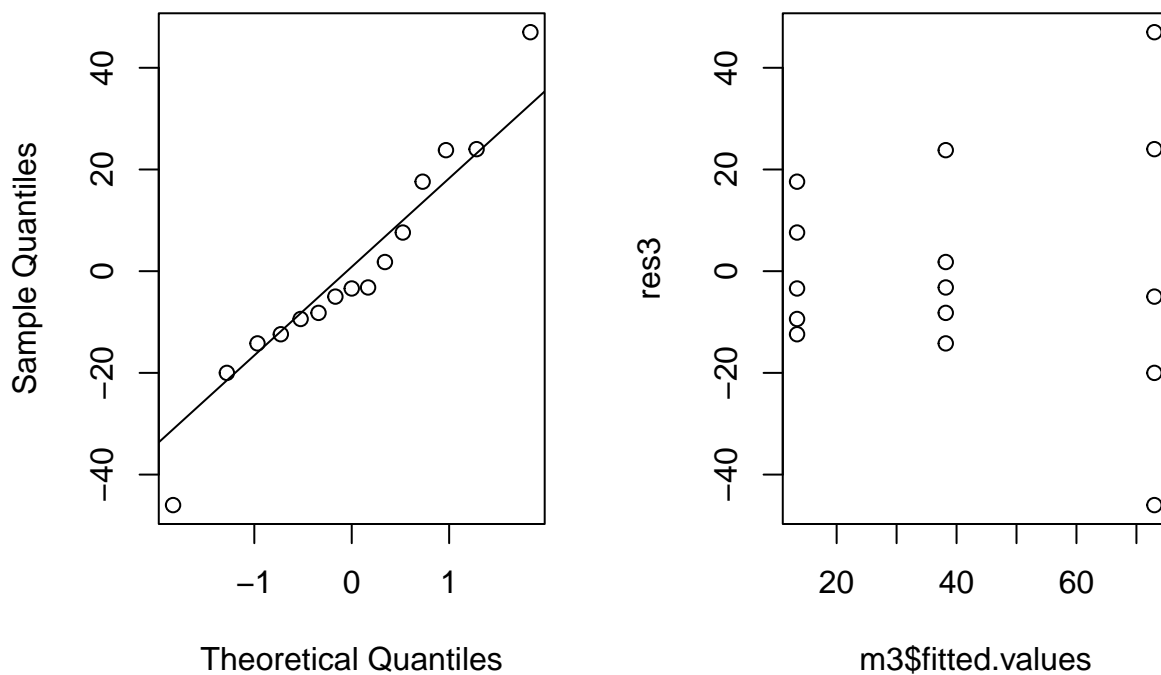

- Based on the QQ-plot, we can see that the data does not violate the assumption of normality. However, there may be indications of heavier tails as there are points that are much lower and higher than the line at the start and end of plot, creating the "S"-shape. Based off of the residuals vs. fitted values plot,

we can see that it does not violate the variance assumption due to the groupings of values at certain fitted values.
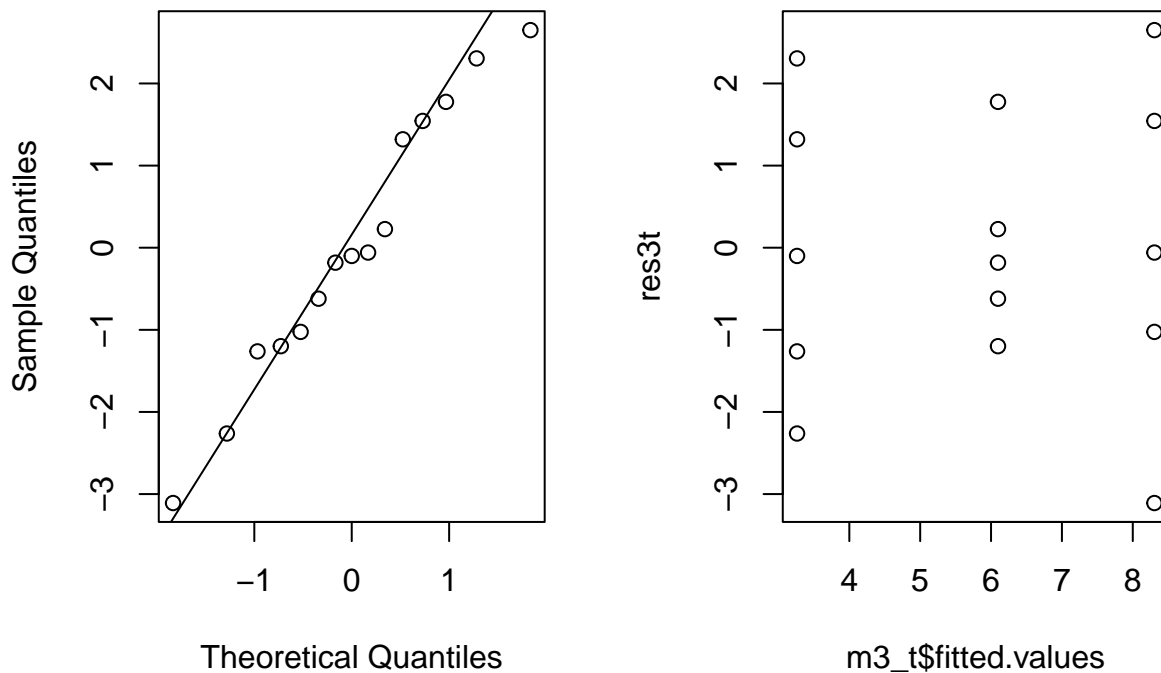
c. Based on your answers to part (b), conduct another analysis of the particle count data and draw appropriate conclusions.

```r
ty <- sqrt(count)
m3_t <- aov(ty ~ method)
summary(m3_t)
```

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## method       2  63.90   31.95    9.84 0.00295 **
## Residuals   12  38.96    3.25
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
res3t <- m3_t$residuals
par(mfrow = c(1, 2))
qqnorm(res3t); qqline(res3t)
plot(m3_t$fitted.values, res3t)
```


- Based on part (b), we perform a root transformation on the model. This assists the "S"-shape in the QQ-plot, indicating a better fit to the normality assumption. The residuals vs fitted values stays relatively similar.

5. Problem 3.42

Refer to Problem 3.23. If we wish to detect a maximum difference in mean response times of 10 milliseconds with a probability of at least 0.90, what sample size should be used? How would you obtain a preliminary estimate of σ^2?

```r
# From 3.23
circuit <- c(rep("1", 5), rep("2", 5), rep("3", 5))
time <- c(9, 12, 10, 8, 15, 20, 21, 23, 17, 30, 6, 5, 8, 16, 7)

m4 <- aov(time ~ circuit)
summary(m4)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## circuit      2  543.6   271.8   16.08 0.000402 ***
## Residuals   12  202.8    16.9
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
d <- 10; MSE <- 16.9
library(pwr)
pwr.anova.test(k = 3, f = d/sqrt(MSE), power = 0.9)
```

```
##
##      Balanced one-way analysis of variance power calculation
##
##              k = 3
##              n = 2.107858
##              f = 2.432521
##      sig.level = 0.05
##          power = 0.9
##
## NOTE: n is number in each group
```

- The sample size that should be used is 3 since the power ANOVA test resulted in n = 2.107858. To obtain a prelminary estimate of σ^2, we can observe the MSE value from the summary of the ANOVA test being 16.9.