Problem 8

```r
blood = read.csv("blood.csv", header = T)
head(blood)
```

```
##   Type Disease
## 1    O     yes
## 2    O     yes
## 3    O     yes
## 4    O     yes
## 5    O     yes
## 6    O     yes
```

(a) – (b)

```r
# create the observaion table, group by Type and Disease
O = xtabs(~ Type + Disease, data = blood)
test_result = chisq.test(O)
```

```
## Warning in chisq.test(O): Chi-squared approximation may be incorrect
```

```r
test_result
```

```
##
##  Pearson's Chi-squared test
##
## data:  O
## X-squared = 10.654, df = 3, p-value = 0.01375
```

According to the R result, the test-statistic is $T = 10.654$, and p-value is 0.01375.

(c) Since p-value $> \alpha = 0.01$, we fail to reject the null at the 0.01 level of significance. We conclude that blood type and whether to develop a disease are independent.

```r
test_result$observed
```

```
##     Disease
## Type no yes
##   A  12  15
##   AB  7   2
##   B   8  17
##   O   9  30
```

```r
test_result$expected
```

```
##     Disease
## Type    no   yes
##   A   9.72 17.28
##   AB  3.24  5.76
##   B   9.00 16.00
##   O  14.04 24.96
```

(d) The observed frequency for blood type A is 15, while the expected frequency is 17.28. Blood type A is thus less likely to have the disease than what we expected if the null was true.

(e) The observed frequency for blood type A is 30, while the expected frequency is 24.96. Blood type A is thus more likely to have the disease than what we expected if the null was true.

(f)

```r
(test_result$observed - test_result$expected)^2/test_result$expected
```

```
##     Disease
## Type        no       yes
##   A  0.5348148 0.3008333
##   AB 4.3634568 2.4544444
##   B  0.1111111 0.0625000
##   O  1.8092308 1.0176923
```

The group blood type "AB" and no disease contributes most to the test statistic.

Problem 9

```r
IQ = read.csv("IQ.csv")

head(IQ)
```

```
##   group iq
## 1     A 44
## 2     A 40
## 3     A 44
## 4     A 39
## 5     A 25
## 6     A 37
```

(a)

```r
anova = aov(iq ~ group, data = IQ)
summary(anova)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## group        2   1529   764.7   20.02 7.84e-07 ***
## Residuals   42   1604    38.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

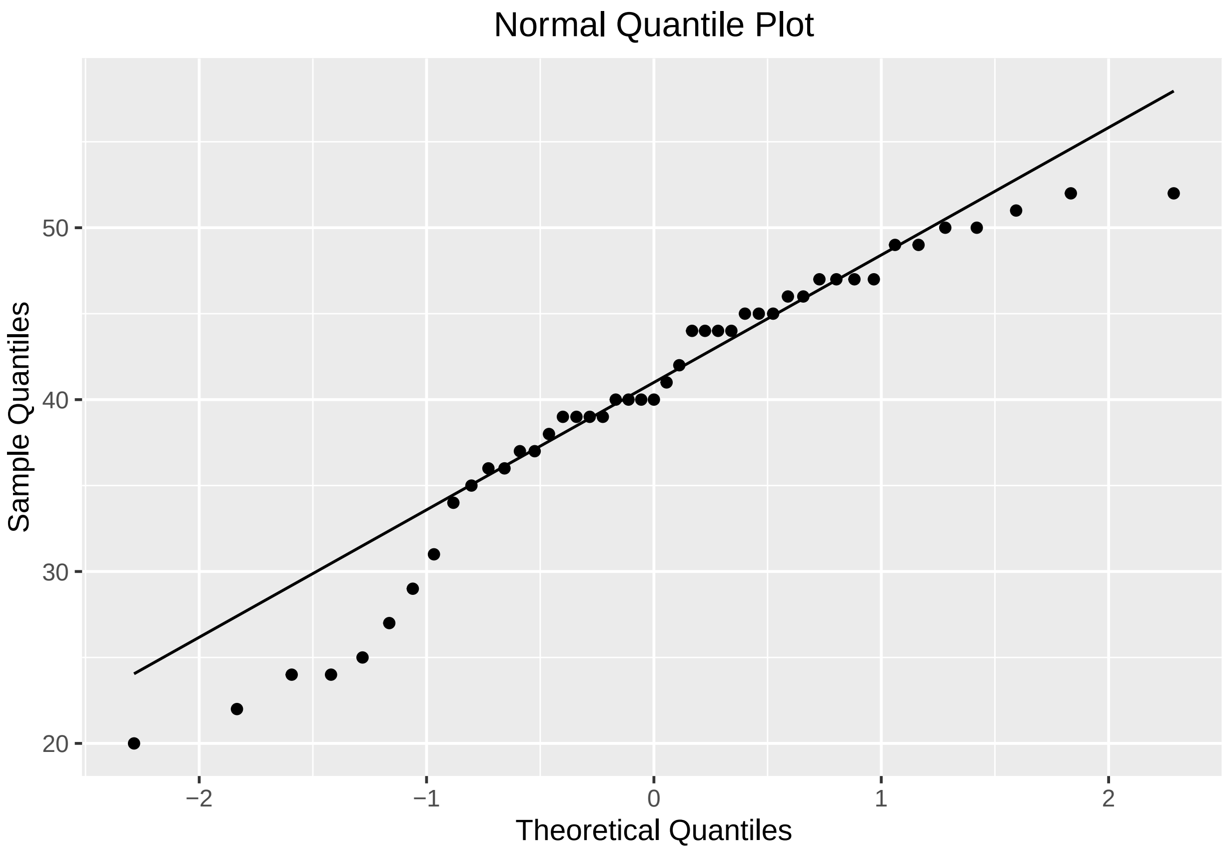
(b) The test statistic is 20.02 and the p-value is 7.84×10^{-7}.

(c) Since p-value $< \alpha = 0.05$, we fail to reject the null at the 0.05 level of significance.

(d) There is significant difference for the mean IQ of students among the three majors.

(e)

```r
library(ggplot2)
ggplot(IQ,
       aes(sample = iq)) +
  stat_qq() +
  stat_qq_line() +
  labs(x = "Theoretical Quantiles",
       y = "Sample Quantiles",
       title = "Normal Quantile Plot") +
  theme(plot.title = element_text(hjust = 0.5))
```



This data do not appear to be approximately normally distributed.

(f)

```
library(asbio)
```

```
## Loading required package: tcltk
```

```
bonfCI(y = IQ$iq, x = factor(IQ$group), conf.level = 0.95)
```

```
##
## 95% Bonferroni confidence intervals
##
##                Diff     Lower    Upper  Decision Adj. p-value
## muA-muB  -0.06667  -5.69471  5.56138    FTR H0            1
## muA-muC     -12.4 -18.02805 -6.77195 Reject H0        6e-06
## muB-muC -12.33333 -17.96138 -6.70529 Reject H0        7e-06
```

(g) The confidence intervals for $\mu_A - \mu_C$ and $\mu_B - \mu_C$ suggest a significant difference in the means.

Problem 10

```
fitness = read.csv("fitness.csv")
head(fitness)
```

```
##   Tread  Run
## 1   7.5 43.5
## 2   7.8 45.2
## 3   7.9 44.9
```

```
## 4   8.1 41.1
## 5   8.3 43.8
## 6   8.7 44.4
```

(a)

```
reg = lm(Run ~ Tread, data = fitness)
summary(reg)
```

```
##
## Call:
## lm(formula = Run ~ Tread, data = fitness)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.9440 -1.5788  0.1860  0.7863  4.5603
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  59.9211     3.1166  19.226 1.90e-13 ***
## Tread        -1.9601     0.3164  -6.194 7.59e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.921 on 18 degrees of freedom
## Multiple R-squared:  0.6807, Adjusted R-squared:  0.6629
## F-statistic: 38.37 on 1 and 18 DF,  p-value: 7.589e-06
```

The slope and intercept of the fitted regression line are -1.9601 and 59.9211.

(b)

```
confint(reg, 'Tread', level = 0.95)
```

```
##           2.5 %    97.5 %
## Tread -2.624957 -1.295313
```

(c) From the summary table, we find that $s_e = 1.921$.

(d) From the summary table, we find that $r^2 = 0.6807$.

(e) Yes, the interval suggests a significant linear relationship.