Assignment 1

Group 47: Luca Cavellini, Paul Hosek, Robert Satzger

2/20/2023

Preparation. Defining Re-Usable Assumption Checks

To standardize assumption checks throughout the assignment, we use the following functions. `check_normality` plots the histogram, qq-plot, and boxplots for all variables of interest and returns the shapiro-wilk test. `check_vars` extends assumption checks for ANOVAs by showing the qq-plot of the residuals and plotting the fitted values against the residuals.

```
check_normality <- function(column_names, data, groups = NULL){
  shapiros = vector("list",length(column_names))
  par(mfrow=c(length(column_names),3+length(groups)/2))
  for (i in 1:length(column_names)){
    col_name <- column_names[[i]]
    col <- data[[col_name]]
    hist(col, xlab = col_name, main = paste("Histogram of", col_name))
    qqnorm(col)
    qqline(col)
    if (missing(groups)){
      boxplot(col, ylab=col_name, main = paste("Boxplot of", col_name))
    }else{
      for (g in groups){
        boxplot(col ~ data[[g]], ylab=col_name, xlab = g,
                main = paste("Boxplot of", col_name, "by", g))
      }
    }
    shapiros[[i]] <- shapiro.test(col)
  }
  return(shapiros)
}

check_vars <- function(anova_model){
  par(mfrow=c(1,2))
  qqnorm(residuals(anova_model), main = "Normal Q-Q Plot of Residuals")
  plot(fitted(anova_model), residuals(anova_model),
       main = "Residuals vs Fitted Values")
}
```

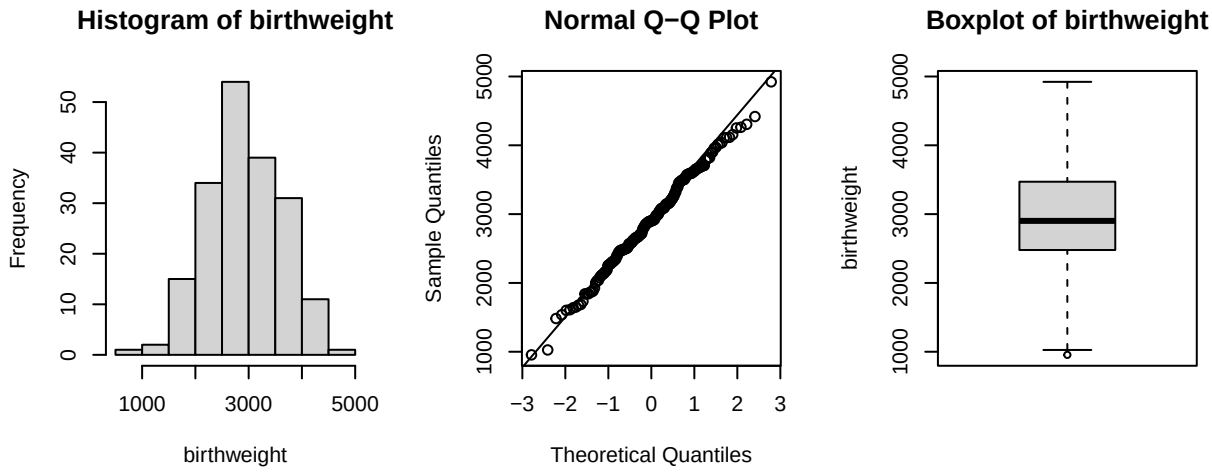
For a consistent report style, we always report two significant decimal digits (at maximum three decimals). Further, all tests use a significance level $\alpha = 0.05$.

Exercise 1. Birthweights

a)

To test for a normal distribution of the sample, the output below show a histogram, a Q-Q plot, and a Shapiro-Wilk test.

```
library(ggplot2)
birthweight <- read.table('data/birthweight.txt', header = TRUE)
shapiro <- check_normality("birthweight", birthweight)
```



The histogram appears to have a bell-shaped curve with no dominant skew to either side, suggesting a normal distribution of the sample data. This is also supported by the roughly linear shape displayed in the Q-Q plot. Moreover, the Shapiro-Wilk test for normality is not significant with $p = 0.90$. Thus, we cannot reject the null hypothesis stating that the sample differs from a normal distribution. Based on these three indications, we therefore assume that the birthweights follow a normal distribution.

```
birthweight <- birthweight$birthweight
alpha = .04
t = qt(1-alpha/2, df = length(birthweight)-1) # t of alpha/2

margin = t * sd(birthweight) / sqrt(length(birthweight))
lower_bound = mean(birthweight) - margin
upper_bound = mean(birthweight) + margin
```

Based on our sample, the 96% confidence interval for μ is $CI = [2808.08, 3018.50]$.

To obtain the minimal sample size for a confidence interval of length 100, we use the z-distribution as an approximation for the t-distribution as the t-distribution would rely on the sample size.

```r
alpha = .04
ci_length = 100
margin = ci_length / 2
q = qnorm(1-alpha/2) # z-scores, t-dist depends on sample size
n = q^2 * sd(birthweight)^2 / margin^2
ceiling(n)
```

```
## [1] 821
```

Ensuring a confidence interval with a maximum length of 100 requires a sample size $n \geq 821$.

```r
alpha = .04
B = 1000
Tstar = numeric(B)
for (i in 1:B){
  Tstar[i] = mean(sample(birthweight, replace = TRUE))
}
TstarLower = quantile(Tstar, alpha/2)
TstarUpper = quantile(Tstar, 1-alpha/2)
upper_bound = 2*mean(birthweight)-TstarUpper
lower_bound = 2*mean(birthweight)-TstarLower
```

The 96% confidence interval resulting from bootstrapping is $CI = [2813.10, 3019.97]$. As expected, the boundaries of the bootstrapped confidence interval are very similar to the calculated confidence interval reported prior because the sample follows an approximately normal distribution.

b)

Based on the assumption of normality established in part **a**, we perform a right-sided one-sample t-test.

```r
mu = 2800
ttest <- t.test(birthweight, mu = mu, alternative = 'g')
print(ttest)
```

```
##
##  One Sample t-test
##
## data:  birthweight
## t = 2, df = 187, p-value = 0.01
## alternative hypothesis: true mean is greater than 2800
## 95 percent confidence interval:
##  2829  Inf
## sample estimates:
## mean of x
##      2913
```

The t-test returns $p = 0.014$. This means we reject the null hypothesis $H0 : \mu <= 2800$. The expert's claim is therefore justified. Moreover, the given confidence interval suggests that the population mean lies anywhere higher than 2829.20. In 95% of replications, it will contain the population mean. The undefined upper boundary (displayed as "Inf") is a consequence of the right-sided test - we are only interested in the lower boundary of the CI.

Alternatively, the claim can be tested with a sign test by comparing the proportion of birth weights above 2800 to $H0 : p \leq 0.5$.

```
binomtest <- binom.test(sum(birthweight>mu),length(birthweight),p=.5,alt='g')
print(binomtest)
```

```
##
##  Exact binomial test
##
## data:  sum(birthweight > mu) and length(birthweight)
## number of successes = 107, number of trials = 188, p-value = 0.03
## alternative hypothesis: true probability of success is greater than 0.5
## 95 percent confidence interval:
##  0.507 1.000
## sample estimates:
## probability of success
##                  0.569
```

Based on the p-value of $p = 0.034$, we can reject the null hypothesis. In accordance with the t-test, the binomial test supports the expert's claim.

c)

The power of the two tests can be compared by simulation. To do so, one needs to repeatedly sample a normal distribution lying within the alternative hypothesis (with $\mu > 2800$). Each of the samples X_b^* then produce a sample mean $\overline{X}_b^*$ which is tested against the true mean with both the t-test and the sign test. Finally, the power of each test is found by calculating the proportions of times each test correctly rejected the null hypothesis.

It can be expected that the t-test will have a higher power than the sign test because it correctly assumes a normal distribution of the sample. In contrast, the sign test discards a lot of information. This makes it more robust than the t-test but reduces power in cases like these where assumptions aren't violated.

d)

Deriving the point estimate $\hat{p}$ from our sample, we can use the margin of error to compute the right side of the confidence interval.

```
p_hat = sum(birthweight<2600)/length(birthweight)
p_left = .25
me = p_hat - p_left
p_right = p_hat + me
z = (p_left - p_hat) / sqrt(p_hat*(1-p_hat)/length(birthweight))
conf_level = 1-pnorm(z)*2
```

This results in the confidence interval $CI = [0.25, 0.41]$. The interval has a confidence level of 98%.

e)

Since we are not given the distribution of birth weights by gender, we cannot perform a test for difference in means. Thus, we perform a test for difference in proportions.

```
k_male = 34
k_female = 28
n_male = k_male + 61
n_female = k_female + 65
p_female = k_female/n_female

proptest = prop.test(c(k_male,k_female), c(n_male, n_female))
print(proptest)
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  c(k_male, k_female) out of c(n_male, n_female)
## X-squared = 0.5, df = 1, p-value = 0.5
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.0879  0.2016
## sample estimates:
## prop 1 prop 2
##  0.358  0.301
```

The proportion test returns $p = 0.50$. Since it is larger than the standard significance level, we reject the null hypothesis stating that the two proportions are significantly different. Thus, the expert's claim is not supported by the data.

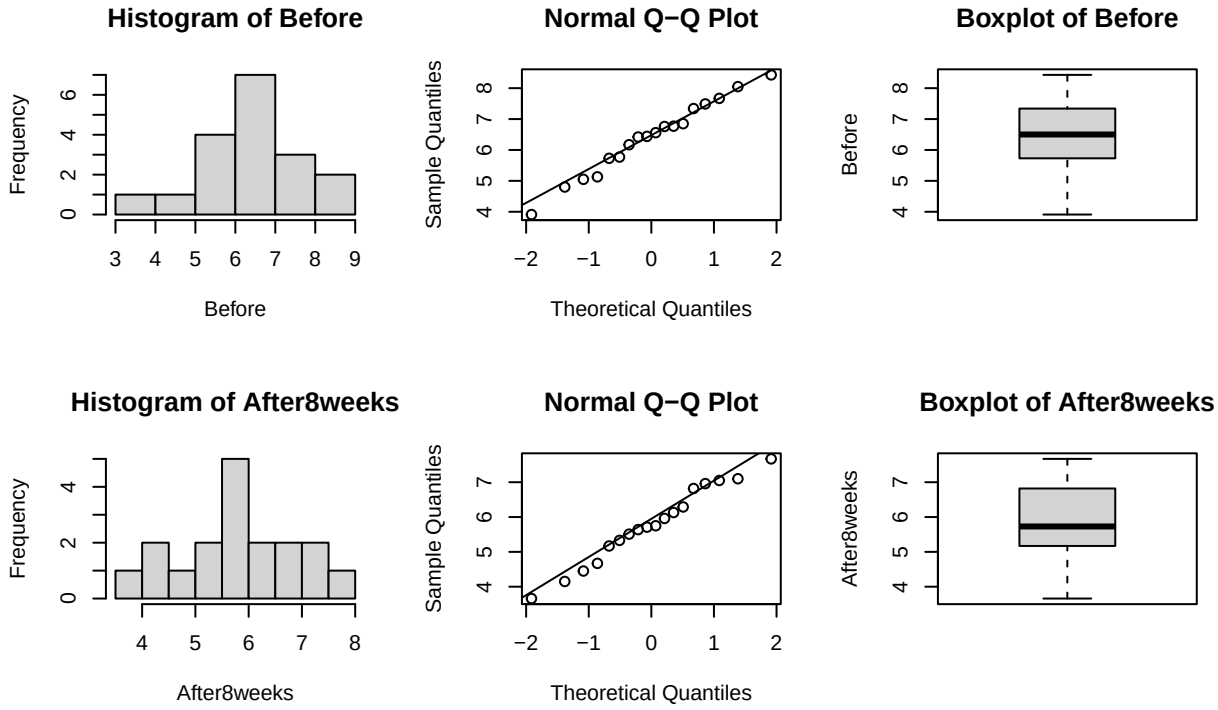
Exercise 2. Cholesterol

```
library(ggplot2)
library(dplyr)
library(ggpubr)
data <- read.table("data/cholesterol.txt", header = TRUE, sep = " ")
```

a)

First plotting the discrete distribution of values, we see that values appear approximately normally distributed. Further, QQ-plots indicate that samples from both measurements are similarly distributed by inspection. However, we need to confirm this with thorough statistical testing.

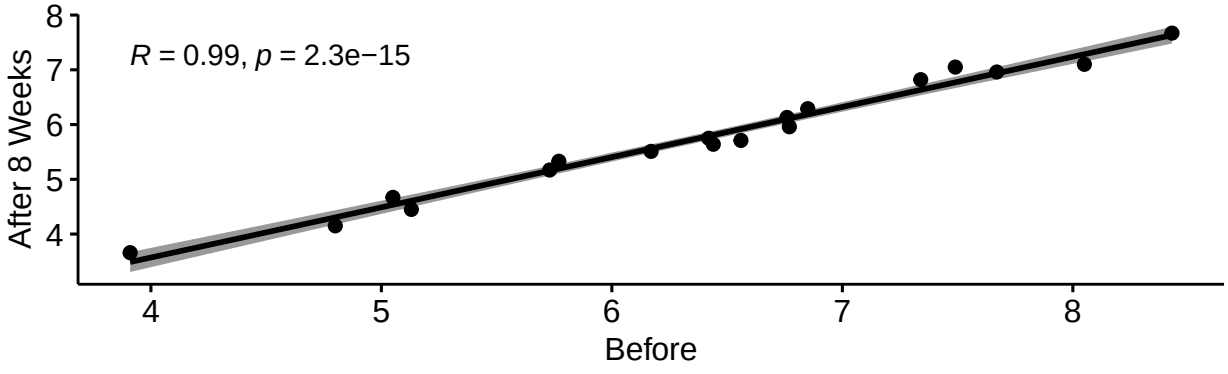
```
shapiros <- check_normality(list("Before", "After8weeks"), data)
```



Shapiro-Wilk tests for normality indicate that the assumption of normality is met for both pre- and post-measurement. P-values are $p = 0.97$ for the before measurement and $p = 0.92$ for the post measurement respectively.

Further, we observe a very high correlation of the data. The Pearson correlation coefficient is $r = 0.99$ with significant correlation of pre and post measurement $p = 0.99$.

```
ggscatter(data, x = colnames(data)[1], y = colnames(data)[2],
          add = "reg.line",cor.coef = TRUE, conf.int = TRUE,
          cor.method = "pearson",
          xlab = "Before", ylab = "After 8 Weeks");
```



b) Since both measurements are taken from the same experimental units, we perform two paired samples tests. Based on the assumption of normality in **a**, we perform a dependent, two-samples t-test. For comparison, we also perform a non-parametric alternative: The Mann-Whitney rank-based test.

```
ttest <- t.test(data$Before, data$After8weeks, paired = TRUE, alternative = "two.sided");
print(ttest)
```

```
##
##  Paired t-test
##
## data:  data$Before and data$After8weeks
## t = 15, df = 17, p-value = 3e-11
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  0.540 0.718
## sample estimates:
## mean difference
##           0.629
```

```
wilcoxtest <- wilcox.test(data$Before, data$After8weeks,
                          paired = TRUE, alternative = "two.sided");
print(wilcoxtest)
```

```
##
##  Wilcoxon signed rank exact test
##
## data:  data$Before and data$After8weeks
## V = 171, p-value = 8e-06
## alternative hypothesis: true location shift is not equal to 0
```

Both the t-test and the Mann-Whitney test concur in that there is a significant effect of the diet on weight, with $p = 0.00$ and $p = 0.00$, respectively. However, the statistical effect does not imply a large effect size. Therefore, the practical utility of the diet is not necessarily given.

The permutation test is applicable, since we have a test statistic that expresses a difference between dependent measures. While the assumption of normality is met in this case, the permutation test would also apply in case of a violation of this assumption.

c)

Based on the Central Limit Theorem, we know that the test statistic of any distribution will be normally distributed as we resample.

```
nsamples <- 1000;

theta <- replicate(nsamples, max(sample(data$After8weeks,size=18, replace = TRUE)));
Tstar25=quantile(theta,0.025)
```

```
Tstar975=quantile(theta,0.975)
T1 = max(data$After8weeks)
CI = c(2*T1-Tstar975,2*T1-Tstar25)
print(CI)
```

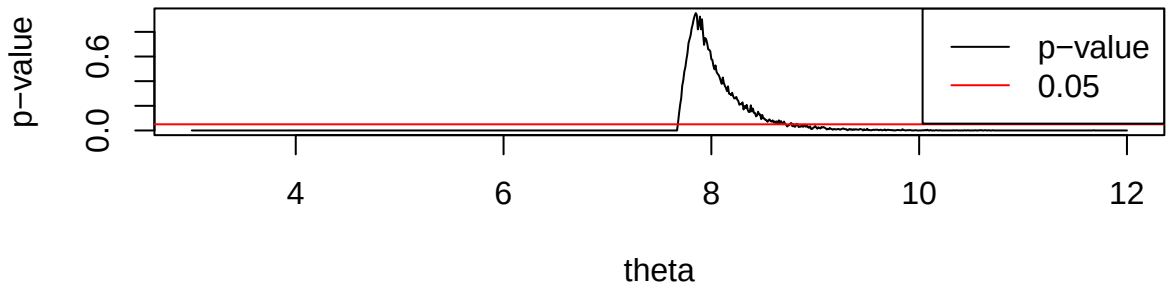
```
## 97.5%  2.5%
##  7.67  8.38
```

The confidence interval is $CI = [7.67, 8.38]$.

This confidence interval could be improved by resampling more often. However, since the sample size is only 18, conclusions about the true population mean are limited. As such, we recommend to collect more data.

d)

```
nsamples <- 1000;
thetas <- seq(3, 12, by = .01);
t <- max(data$After8weeks)
p_vals <- numeric(length(thetas));
for (i in 1:length(thetas)) {
  res <- replicate(nsamples,max(runif(18,3, thetas[i])))
  pl=sum(res<t)/nsamples
  pr=sum(res>t)/nsamples
  p_vals[i]=2*min(pl,pr)
}

plot(thetas, p_vals, type = "l", xlab = "theta", ylab = "p-value");
abline(h = 0.05, col = "red");
legend("topright", legend = c("p-value", "0.05"),
       col = c("black", "red"), lty = c(1, 1));
```



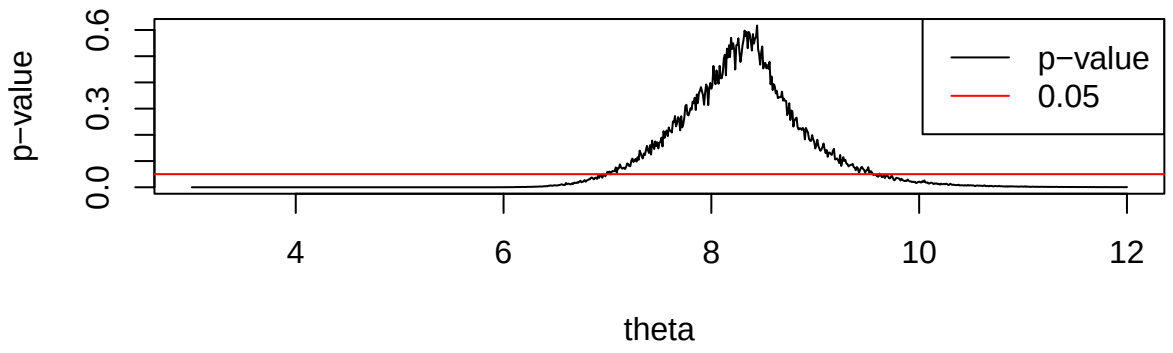
```
t <- thetas[p_vals >.05]
I <- c(min(t), max(t))
print(I)
```

```
## [1] 7.68 8.82
```

The thetas for which the hypothesis is not rejected lie in the interval $I = [7.68, 8.82]$.

Alternatively, we can use the Kolmogorov-Smirnov test to test if the distributions are the same. But we must be careful to not accept distributions with a maximum value smaller than the maximum of the sample, because else the maximum lies outside the given range.

```
thetas <- seq(3, 12, by = .01);

# Loop over theta values and perform bootstrap test
ks_res <- numeric();
p_vals <- numeric(length(thetas));
for (i in 1:length(thetas)) {
  theta <- thetas[i];
  cur_unif <- runif(10000,3,thetas[i]);
  p_vals[i] <- ks.test(data$After8weeks, cur_unif)['p.value'];
}
```



e)

```
binom_e <- binom.test(sum(data$After8weeks < 6), nrow(data), alternative = "less")
```

The binomial test returns an insignificant p-value of $p = 0.88$. Therefore, we cannot reject the null hypothesis stating that the median post measurement is less than 6.

First, we find true fraction of cholesterol levels less than 4.5.

```r
binom_e1 <- binom.test(sum(data$After8weeks < 4.5), nrow(data),
                       p = 0.25, alternative = "l")
print(binom_e1)
```

```
##
##  Exact binomial test
##
## data:  sum(data$After8weeks < 4.5) and nrow(data)
## number of successes = 3, number of trials = 18, p-value = 0.3
## alternative hypothesis: true probability of success is less than 0.25
## 95 percent confidence interval:
##  0.000 0.377
## sample estimates:
## probability of success
##                  0.167
```

```r
binom_e2 <- binom.test(sum(data$After8weeks < 4.5), nrow(data),
                       p = 0.2500001, alternative = "l")
print(binom_e2)
```

```
##
##  Exact binomial test
##
## data:  sum(data$After8weeks < 4.5) and nrow(data)
## number of successes = 3, number of trials = 18, p-value = 0.3
## alternative hypothesis: true probability of success is less than 0.25
## 95 percent confidence interval:
##  0.000 0.377
## sample estimates:
## probability of success
##                  0.167
```

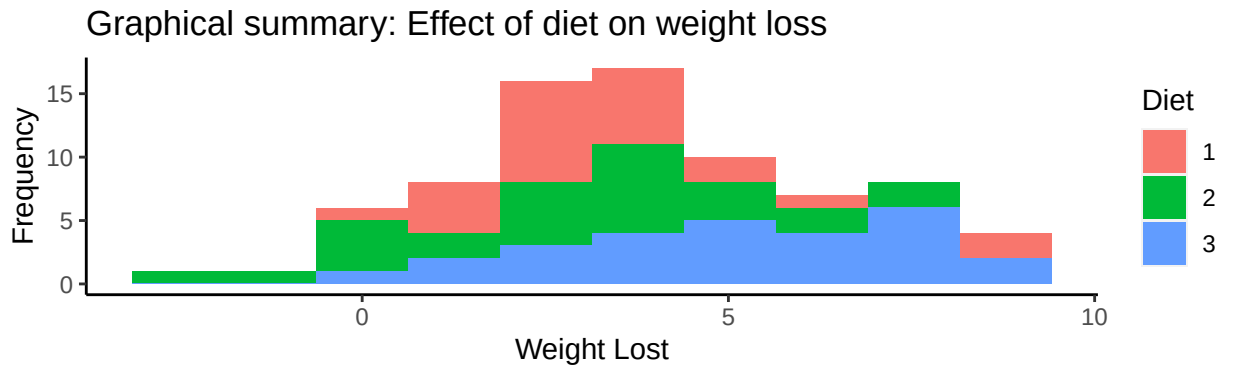
With the null hypothesis H0: p >= 0.25, the binomial test returns an insignificant p-value of $p = 0.31$. However, the alternative hypothesis should include the value 0.25. Therefore, we corroborated our findings with a second test, including the value 0.25. Therefore, we can not confirm the claim that the fraction of cholesterol levels less than 4.5 is at most 25%.

Exercise 3. Diet

a)

```r
data <- read.table("data/diet.txt", header=TRUE)
data$diet <- factor(data$diet)
data$gender <- factor(data$gender)
```

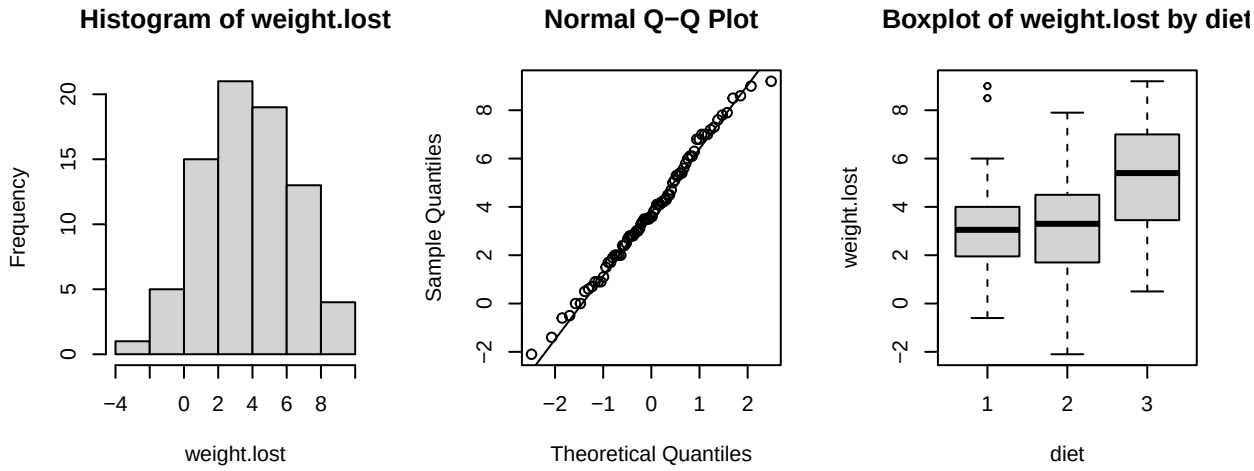
```
data$weight.lost <- data$preweight - data$weight6weeks

ggplot(data, aes(x = weight.lost, fill = diet)) +
  geom_histogram(alpha = 1, position = "stack", bins = 10) +
  labs(title = "Graphical summary: Effect of diet on weight loss"
       , x = "Weight Lost", y = "Frequency", fill = "Diet")
```



To test if the diet has a general effect on weight loss, we can do a paired t-test. Here, we test if there is a significant weight loss between the measurements.

First, we test whether the normality assumption is met.

```
shapiro <- check_normality("weight.lost", data, "diet")
```



Upon inspection of the histogram and the Q-Q plot, the assumption of a normal distribution appears to be met. The Shapiro-Wilk test does not find a violation of the normality assumption, either, with $p = 0.80$.

```
t.test(data$weight6weeks, data$preweight,paired=TRUE)
```

```
##
```

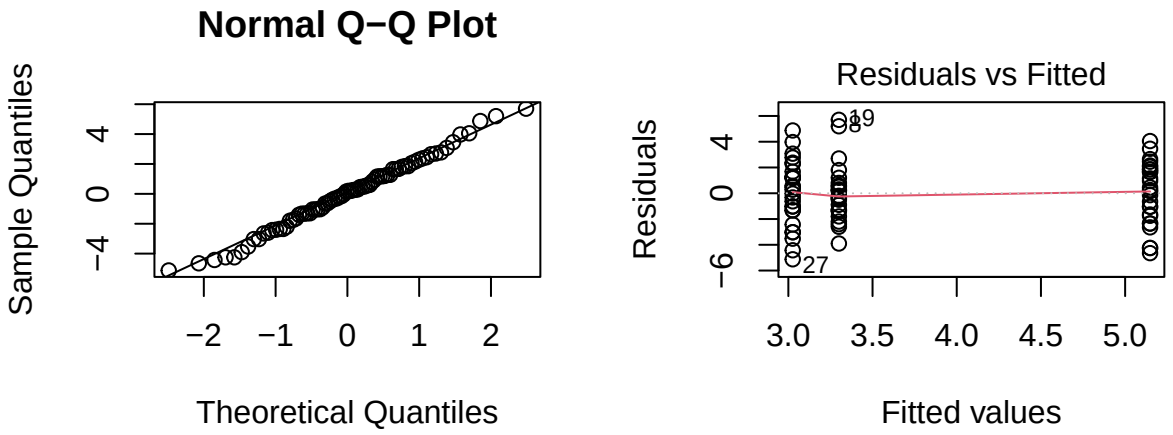
```
##  Paired t-test
##
## data:  data$weight6weeks and data$preweight
## t = -13, df = 77, p-value <2e-16
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -4.42 -3.27
## sample estimates:
## mean difference
##           -3.84
```

The results of the repeated measures t-test show us that there is a significant difference between pre- and post weight. This suggests an effect of diet on weight. However, since there are three different types of diets, it is possible that, when combined, they have a significant effect on weight loss, but there are differences between the diets. It is possible, that one of the diets leads to no weight loss, but the other two do. As such, we need to conduct a one-way ANOVA to investigate the effect of the type of diet.

b)

After establishing normality in **a**, we additionally check for homoscedasticity here.

```
one_w_anova <- lm(weight.lost ~ diet, data)
par(mfrow=c(1,2))
qqnorm(one_w_anova$residuals); qqline(one_w_anova$residuals)
plot(one_w_anova, which = 1)
```



The plots show no pattern in the distribution of residuals in relation to fitted values.

We now can do a one-way ANOVA to test the effect of type of diet.

```
res_ow <-anova(one_w_anova)
res_ow
```

```
## Analysis of Variance Table
##
## Response: weight.lost
##           Df Sum Sq Mean Sq F value Pr(>F)
## diet       2     71    35.5     6.2 0.0032 **
## Residuals 75    430     5.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The result indicate, that there is a significant difference in the type of diet on weight loss. Based on the result of the ANOVA alone, we cannot conclude which type of diet is different.

To identify which type of diet is best for losing weight, we will conduct post-hoc t-tests between all pairs of the ANOVA. Here, we must account for the multiple comparisons and adjust our significance level accordingly. As such, we will use Bonferroni correction.

```
post_hoc <- pairwise.t.test(data$weight.lost, data$diet, p.adjust.method = "bonferroni")
post_hoc
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  data$weight.lost and data$diet
##
##   1     2
## 2 1.000 -
## 3 0.022 0.005
##
## P value adjustment method: bonferroni
```

We can see that there is a significant difference between diets 1 and 3 and diets 2 and 3. However, no significant difference between diet 1 and 2 was found.

```
one_sample_t_test <- function(X){
  return(t.test(X)$p.value)
}
means <- aggregate(data$weight.lost, list(data$diet), mean)
means$p.value <- aggregate(data$weight.lost, list(data$diet), one_sample_t_test)$x
means$significant <- means$p.value < 0.05
names(means)[2] <- "Average weight loss"
names(means)[1] <- "Diet Type"
means
```

```
##   Diet Type Average weight loss  p.value significant
## 1         1                3.30 2.40e-07        TRUE
## 2         2                3.03 1.36e-06        TRUE
## 3         3                5.15 2.03e-11        TRUE
```

Comparing the means, we can see the direction of the relationships. Based on the significant differences found before, we can conclude that diet 3 is more effective than diets 1 and 2. It is the best diet for losing weight. But, all three diets lead to weight loss, as the one-sample t-tests compared to $\mu = 0$ show. They are all significantly different from 0.

It is also possible to alternatively conduct a Kruskal-Wallis test instead of the one-way ANOVA. Again, a significant difference at the same significance level between groups was found.

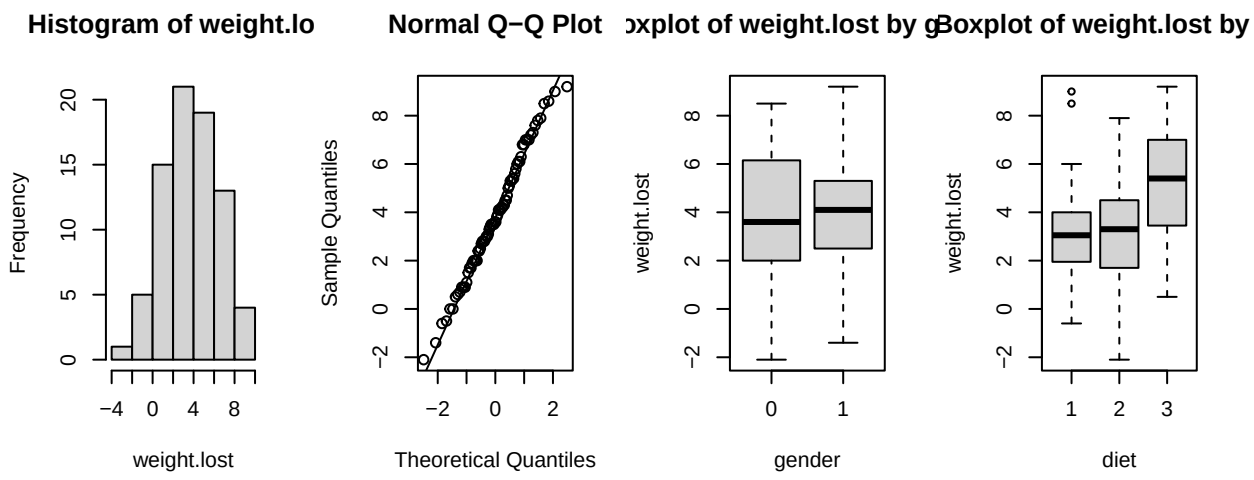
```
kruskal.test(weight.lost ~ diet, data)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  weight.lost by diet
## Kruskal-Wallis chi-squared = 10, df = 2, p-value = 0.005
```

c)

Before performing a two-way ANOVA, we run normality and variance checks on the lost weight.

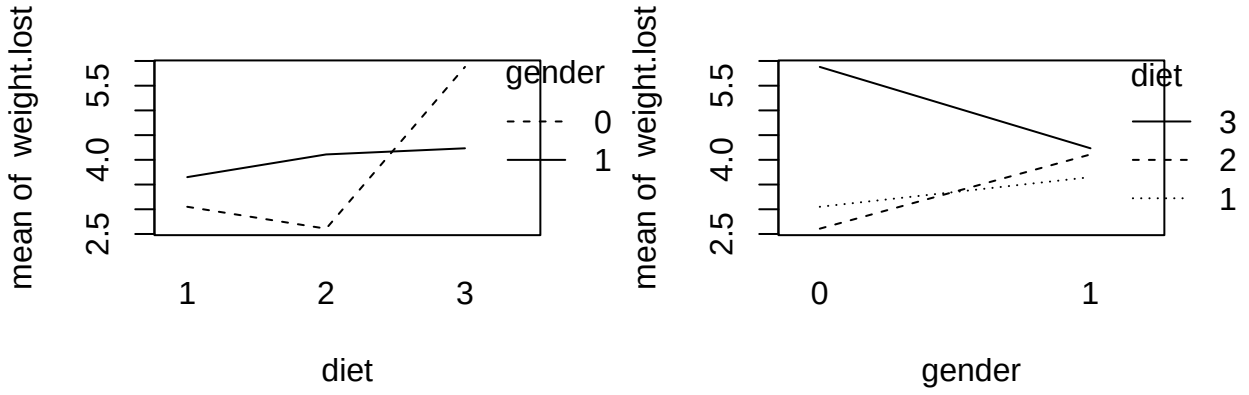
```
shapiro <- check_normality("weight.lost", data, list("gender", "diet"))
```



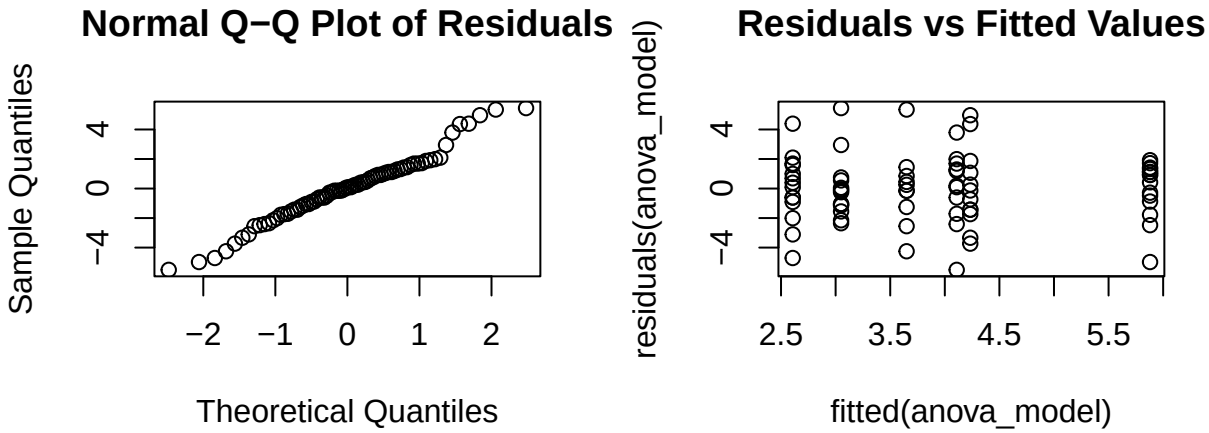
Based on the plots, the data appear to be normally distributed. The Shapiro-Wilk test also does not find a violation of the normality assumption, with $p = 0.80$. Further, we check the assumption of a homogeneity of variances.

To decide whether the model should contain an interaction term, we plot the means of lost weight by gender and diet.

```
attach(data)
par(mfrow=c(1,2))
interaction.plot(diet, gender, weight.lost)
interaction.plot(gender, diet, weight.lost)
```




```
two_w_anova <- lm(weight.lost ~ diet * gender, data)
check_vars(two_w_anova)
```



The Q-Q plot shows deviations of the residuals' sample quantiles from the theoretical quantiles. Thus, the interpretations of the ANOVA may be impacted. Nonetheless, we will proceed with the ANOVA for the purposes of the exercise.

As the non-parallel lines in the plots suggest, we can assume an interaction of diet and gender on lost weight.

```
anova_res <- anova(two_w_anova)
print(anova_res)
```

```
## Analysis of Variance Table
##
## Response: weight.lost
##             Df Sum Sq Mean Sq F value Pr(>F)
```

```
## diet         2     61   30.26    5.63 0.0054 **
## gender       1      0    0.17    0.03 0.8599
## diet:gender  2     34   16.95    3.15 0.0488 *
## Residuals   70    376    5.38
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The full two-way ANOVA results in an interaction effect of diet and gender on lost weight ($p = 0.049$). Additionally, the analysis finds a main effect of diet on lost weight ($p = 0.005$) while a main effect of gender was not present ($p = 0.86$).

e)

As found in **c**, the effect of diet on weight loss depends on gender. As such, we prefer the two-way ANOVA as the model in **b** does not take this information into account. Specifically, the first model will result in more accurate diet recommendations to maximize weight loss.

```
diet1 <- two_w_anova$coefficients[[1]]
diet2 <- diet1 + two_w_anova$coefficients[[2]]
diet3 <- diet2 + two_w_anova$coefficients[[3]]

dietgroups <- c(diet1,diet2,diet3)
print(dietgroups)
```

```
## [1] 3.05 2.61 5.44
```

The model from the two-way ANOVA predicts weight losses of $3.05, 2.61, 5.44$ for diet group 1, 2, and 3, respectively.

Exercise 4. Yield of peas

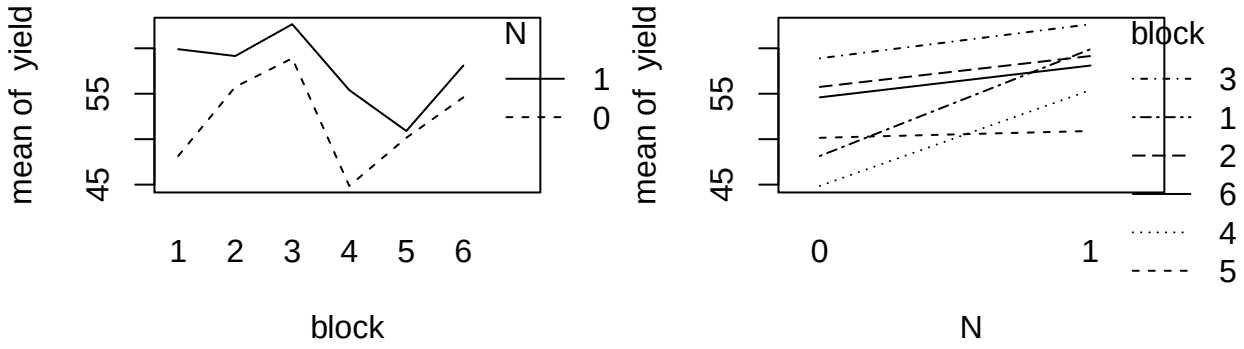
a)

```
random_plots <- cbind(
  rep(1:24),
  rep(1:6, each = 4),
  replicate(3, c(replicate(6, sample(c(1,1,0,0)))))
)
plots_df <- data.frame(random_plots)
header <- c("plot", "block", "N", "P", "K")
colnames(plots_df) <- header
plots_df[1:4,]
```

```
##   plot block N P K
## 1    1     1 1 1 0
## 2    2     1 1 0 1
## 3    3     1 0 0 1
## 4    4     1 0 1 0
```

b)

The following plots show the average yield per block for plots treated with or without Nitrogen (N).



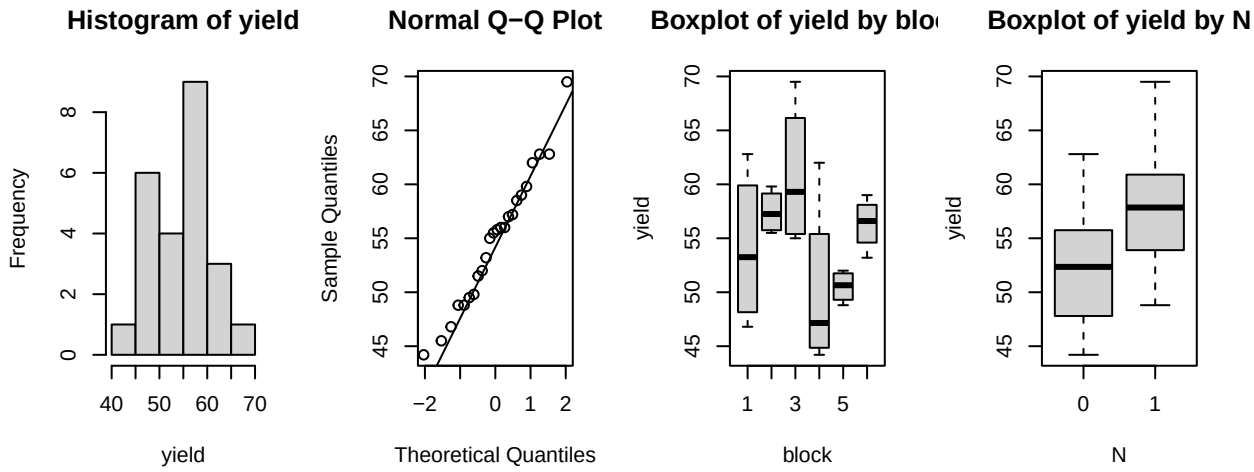
It is noteworthy that the study generally follows an incomplete block design. This would suggest interactions cannot be evaluated. However, taking only the factors block and N into account, there are two observations per condition. This means, the interactions may be evaluated.

Even though the yield is consistently higher for plots treated with Nitrogen, the size of this effect may differ between blocks. Taking this potential interaction of block and Nitrogen into account can help explain some of the variance in the treatment effect. Therefore, it can help improve the model fit. By performing a full two-way ANOVA, we can test whether the interaction is, in fact, present.

c)

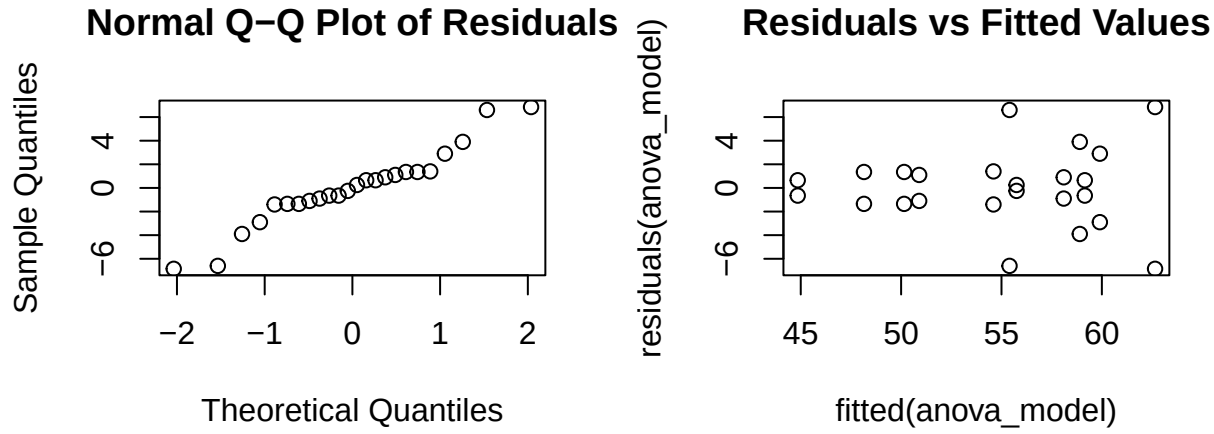
To perform an ANOVA, we first check the normality assumption with the following plots.

```
npk$block <- as.factor(npk$block)
npk$N <- as.factor(npk$N)
shapiro <- check_normality("yield", npk, c("block","N"))
```



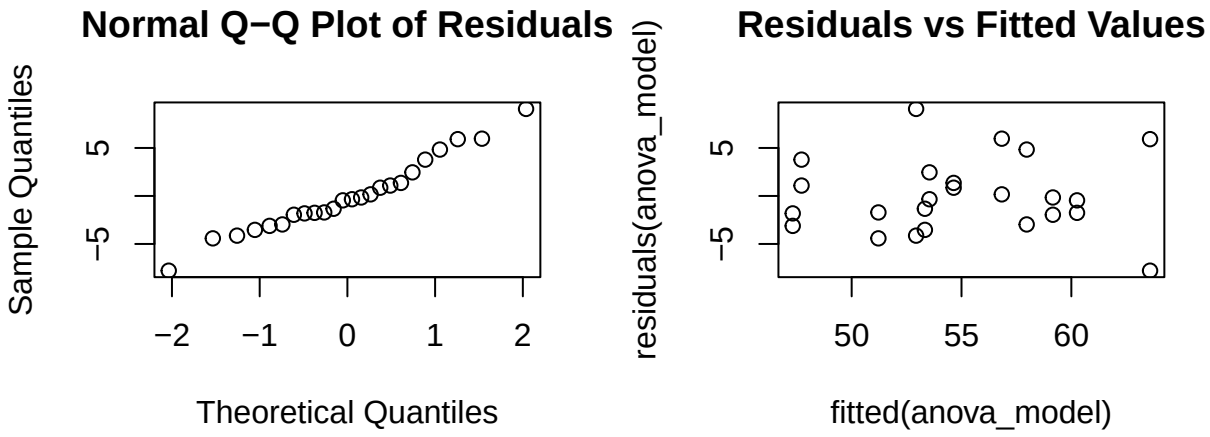
The histogram and the QQ-plot indicate a normal distribution. Moreover, the Shapiro-Wilk test returns $p = 0.87$. Therefore, we do not find a violation of the normality assumption. Second, we check for homogeneity of variances.

```
lm_aov1 = lm(yield ~ block * N, data = npk)
check_vars(lm_aov1)
```



The Q-Q plot of the residuals indicates a violation of the assumption of homogeneity of variances. Therefore, we drop the interaction term from the model.

```
lm_aov12 = lm(yield ~ block + N, data = npk)
check_vars(lm_aov12)
```



With the removal of the interaction term, equal variances may be assumed. For completeness, we will proceed with both models.

```
# Anova without interaction term.
aov1 = anova(lm_aov1)
print(aov1)
```

```
## Analysis of Variance Table
##
## Response: yield
##           Df Sum Sq Mean Sq F value Pr(>F)
## block      5    343    68.7    3.36   0.04 *
## N          1    189   189.3    9.26   0.01 *
## block:N    5     99    19.7    0.96   0.48
## Residuals 12    245    20.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Anova with the interaction term.
aov2 = anova(lm_aov12)
print(aov2)
```

```
## Analysis of Variance Table
##
## Response: yield
##           Df Sum Sq Mean Sq F value Pr(>F)
## block      5    343    68.7    3.40 0.0262 *
## N          1    189   189.3    9.36 0.0071 **
## Residuals 17    344    20.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The model without the interaction term and with the interaction term both paint a similar picture. There does not appear to be an interaction effect of block x Nitrogen on yield. Therefore, block does not affect our research question about the effect of Nitrogen on yield. A more sensible way to take block into account in the analysis would be in an additive model.

The Friedman test is not applicable in this case because there are two observations of the outcome variable (yield) per block x Nitrogen combination. In contrast, the Friedman test only applies to situations with 1 observation per factor combination.

d)

```
# Define the models
models <- list(
  lm(yield ~ block * N + P + K, npk),
  lm(yield ~ block * N + K, npk),
  lm(yield ~ block * N + P, npk),
  lm(yield ~ block * N, npk),
  lm(yield ~ block + N, npk),

  lm(yield ~ block * P + K + N, npk),
  lm(yield ~ block * K + N + P, npk)
)

result_matrix <- matrix("-",nrow = 8, ncol = length(models),
                        dimnames = list(c(
                          "Residual sum of Squares","block", "N", "P",
                          "K","block:N","block:P","block:K"),
                          paste0("model", 1:length(models))))

# loop through each model
for (i in 1:length(models)) {
  anova_table <- anova(models[[i]])
  terms_in_model <- attr(terms(models[[i]]), "term.labels")
  count_terms <- 1
  for (term in terms_in_model){

    result_matrix[term, i] <-
      ifelse(anova_table[count_terms, "Pr(>F)"] < 0.05, paste0(term, "*"), term)
    count_terms <- count_terms+1
  }
  result_matrix["Residual sum of Squares", i] <-
    round(anova(models[[i]])[nrow(anova(models[[i]])), "Sum Sq"],2)

}
kable(result_matrix)
```

	model1	model2	model3	model4	model5	model6	model7
Residual sum of Squares	141.67	150.07	236.87	245.27	343.79	168.79	169.92
block	block*	block*	block	block*	block*	block*	block*
N	N*	N*	N*	N*	N*	N*	N*
P	P	-	P	-	-	P	P
K	K*	K*	-	-	-	K*	K*
block:N	block:N	block:N	block:N	block:N	-	-	-
block:P	-	-	-	-	-	block:P	-
block:K	-	-	-	-	-	-	block:K

Our result table compare different models. For each possible term in the model, we indicated whether the term is part of the model by writing the term in the cell (e.g., N) or not ("-"). Further, if the term was significant it is written with an asterisk (e.g., K*). As such, we can reconstruct how each model looks from this table. We also provide the Sum of Square of the model as an indicator of model fit. We can see that dropping P from the model between models 1 and 2 reduces model fit much less than dropping K (model3). Further, model 1 and 2 have the best fit. We also observer that if we drop the interaction term from the model (model5), model fit severely suffers. Based on our previous analysis and the current observations (and Occam's razor), we prefer model 2. Model 2 does not include P (phosphate), so it is much more parsimonious. If predictions must be as precise as possible and measuring phosphate is not a lot of work, model 1 should be preferred instead.

e)

Below, we base our normality assumption on the checks done in **c**.

```
library(lme4)
mixed_model <- lmer(yield ~ N + (1|block), REML=FALSE, data = npk)
mixed_model1 <- lmer(yield ~ (1|block), REML=FALSE, data = npk)
mm_aov <- anova(mixed_model1, mixed_model)
print(mm_aov)
```

```
## Data: npk
## Models:
## mixed_model1: yield ~ (1 | block)
## mixed_model: yield ~ N + (1 | block)
##              npar AIC BIC logLik deviance Chisq Df Pr(>Chisq)
## mixed_model1    3 159 163  -76.7      153
## mixed_model     4 154 158  -72.7      146   7.9  1      0.005 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The mixed effects model returns a significant effect of Nitrogen on yield, with $p = 0.005$. Therefore, it is in agreement with the additive, fixed-effects model in **c**. In contrast to the fixed effects model, this analysis does not allow for interpretations on the effect of block on yield because block was assumed to be a random factor.