Understanding Data and Statistical Design (60117)

Chapter 4

Two Factor Experiments

Subject Coordinator: Stephen Woodcock
Lecture notes: Scott Alexander

School of Mathematical and Physical Sciences, UTS

Autumn 2024

Chapter outline

Topics:

- two factor experiment
- two way ANOVA and F-tests
  - assumptions
  - sum of squares
  - F-distribution
  - running factor A test
  - running factor B test
  - running interaction test
- example
  - experimental and blocking/nuisance factor
  - experimental factors with interaction

Two factor experiment

Last chapter we looked at **one factor experiments** where the treatments were assigned randomly so as to create independent samples.

The statistical tool we developed for this situation was **one way ANOVA** and its associated hypothesis test, the ***F*-test**.

This chapter we consider **two factor experiments**, again with treatments randomly assigned so as to create independent samples.

One factor will always be used as an **experimental factor** (controlled by experimenter) while the other will be used as

1. an experimental factor (a second factor under the control of experimenter)
2. a **blocking factor** to control extraneous sources of variability.

Two factor experiment

We introduce the most general design, the $\boldsymbol{a \times b}$ **factorial**, or **two factor crossed** design.

In this setup the first factor A takes levels $i \in \{1, \ldots, a\}$, the second factor B takes levels $j \in \{1, \ldots, b\}$ and there are $a \times b$ unique combinations of factor levels i and j.

For example, for a 3×4 factorial experiment there are $3 \times 4 = 12$ level combinations, as summarised below.

		Factor *B*			
		1	2	3	4
Factor *A*	1	11	12	13	14
	2	21	22	23	24
	3	31	32	33	34

Two factor experiment

If both factors are experimental, then the treatments are the $a \times b$ unique combinations of factor levels.

If factor A is experimental and factor B blocking, then it is customary to refer to the **treatments** as the a levels of factor A and the **blocks** as the b levels of factor B.

But this is just language – mathematically there is no difference between experimental and blocking factors.

Two factor experiment

Examples of two factor experiments include

1. testing dosages of a pharmaceutical (experimental factor 1) in combination with different exercise regimes (experimental factor 2).
2. testing dosages of a pharmaceutical (experimental factor) while controlling for variability due to sex (blocking factor).

The statistical tool we develop is **two way analysis of variance (ANOVA)** with associated ***F*-tests**.

Two factor experiment

The **statistical model** for this type of experiment is defined via $a \times b$ random samples

$$Y_{ijk} = \mu_{ij} + \epsilon_{ijk}$$

for $i \in \{1, \ldots, a\}$, $j \in \{1, \ldots, b\}$ and $k \in \{1, \ldots, n_{ij}\}$, where

- Y_{ijk} is response of the k-th measurement unit to factors A and B at levels i and j respectively
- $\mu_{ij} = \mathbb{E}[Y_{ijk}]$ is population mean response to factors A and B at levels i and j
- ϵ_{ijk} is random effect of the k-th measurement unit to factors A and B at levels i and j (called noise or error terms).

The **sample data** for this experiment are observations of the $a \times b$ random samples

$$y_{ijk} = \mu_{ij} + \epsilon_{ijk}$$

for $i \in \{1, \ldots, a\}$, $j \in \{1, \ldots, b\}$ and $k \in \{1, \ldots, n_{ij}\}$.

Two factor experiment

Becasue we need to distinguish between the effects of the two factors, we decompose the population means as

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \delta_{ij}$$

where

- μ is global mean level
- α_i is effect of factor A at level i
- β_j is effect of factor B at level j
- δ_{ij} is effect of factor A and B interaction at levels i and j respectively.

For technical reasons that we won't go into, this representation also requires

$$\sum_{i=1}^{a} \alpha_i = \sum_{j=1}^{b} \beta_j = \sum_{i=1}^{a} \delta_{i,j} = \sum_{j=1}^{b} \delta_{i,j} = 0.$$

Two way ANOVA and F-tests – assumptions

The F-test is a **parametric test** with assumptions about the random sample.

These assumptions are

- $\epsilon_{ijk} \sim N(0, \sigma)$, i.e. normally distributed with $\mathbb{E}[\epsilon_{ijk}] = 0$ and $\text{var}(\epsilon_{ijk}) = \sigma^2$
- ϵ_{ijk} are all independent from each other.

These assumptions can be re-stated in terms of the response variable as $Y_{ijk} \sim N(\mu_{ij}, \sigma)$ and independent.

So once again we have the three assumptions:

1. normality
2. equal variance
3. independence.

Two way ANOVA and F-tests – sum of squares

We present the case for equal sample sizes $n_{ij} = n$ for $i \in \{1, \ldots, a\}$ and $j \in \{1, \ldots, b\}$, which is called a **balanced design**.

For each random sample we calculate the sample mean

$$\overline{Y}_{ij} = \frac{1}{n} \sum_{k=1}^{n} Y_{ijk}$$

and the factor level means

$$\overline{Y}_{i\bullet} = \frac{1}{bn} \sum_{j=1}^{b} \sum_{k=1}^{n} Y_{ijk} \quad \text{and} \quad \overline{Y}_{\bullet j} = \frac{1}{an} \sum_{i=1}^{a} \sum_{k=1}^{n} Y_{ijk}.$$

We also need the global average across all random samples

$$\overline{Y} = \frac{1}{abn} \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} Y_{ijk}.$$

Two way ANOVA and F-tests – sum of squares

Two way ANOVA is based on the following identity

$$SST = SSTr + SSE$$

where the **sum square total**

$$SST = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(Y_{ijk} - \overline{Y}\right)^2,$$

the **sum square treatment**

$$SSTr = SSA + SSB + SSD,$$

with the **sum square factor A**

$$SSA = bn\sum_{i=1}^{a}(\overline{Y}_{i\bullet} - \overline{Y})^2,$$

Two way ANOVA and F-tests – sum of squares

the **sum square factor B**

$$SSB = an \sum_{j=1}^{b} (\overline{Y}_{\bullet j} - \overline{Y})^2,$$

the **sum square interaction**

$$SSD = n \sum_{i=1}^{a} \sum_{j=1}^{b} (\overline{Y}_{ij} - \overline{Y}_{i\bullet} - \overline{Y}_{\bullet j} + \overline{Y})^2$$

and the **sum square error**

$$SSE = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} (Y_{ijk} - \overline{Y}_{ij})^2.$$

Two way ANOVA and F-tests – sum of squares

The decomposition can be used to define the **coefficient of determination**

$$R^2 = \frac{SSTr}{SST} = \frac{SSA + SSB + SSD}{SST}.$$

As $0 \leq R^2 \leq 1$, we can interpret this quantity as the proportion of the total variation SST captured by the treatments.

Two way ANOVA and F-tests – F-distribution

Now define the **mean squares** for the factors and interaction as

$$MSA = \frac{SSA}{a-1}, \quad MSB = \frac{SSB}{b-1}, \quad MSD = \frac{SSD}{(a-1)(b-1)}$$

and **mean square error** as

$$MSE = \frac{SSE}{abn - ab}.$$

Under the assumptions, it can be shown that if

$$\alpha_1 = \alpha_2 = \cdots = \alpha_a = 0$$

then the RV

$$F_A^* = \frac{MSA}{MSE} \sim F(a-1, abn-ab). \tag{1}$$

Two way ANOVA and F-tests – F-distribution

Similarly, if

$$\beta_1 = \beta_2 = \cdots = \beta_b = 0$$

then

$$F_B^* = \frac{MSB}{MSE} \sim F(b-1, abn-ab). \tag{2}$$

Also, if

$$\delta_{ij} = 0 \quad \text{for all } i, j$$

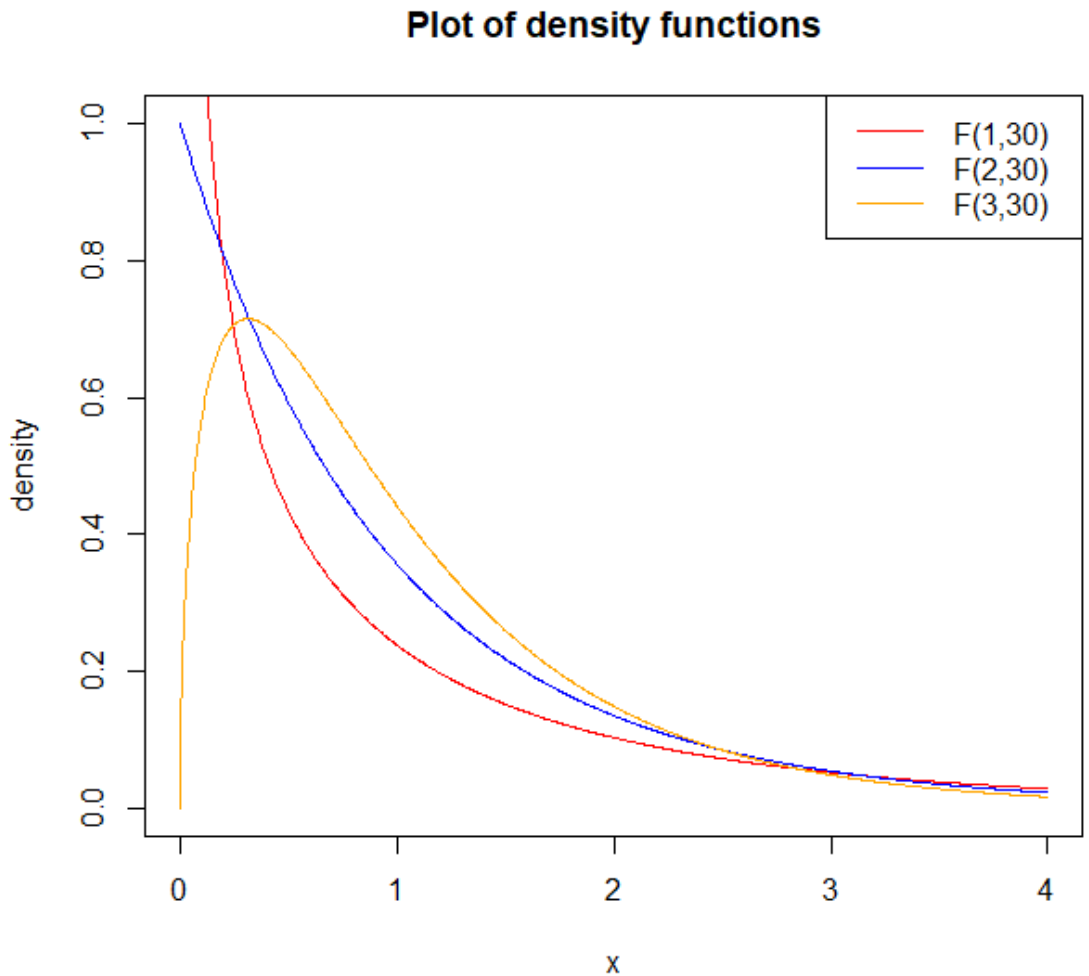
then

$$F_D^* = \frac{MSD}{MSE} \sim F((a-1)(b-1), abn-ab). \tag{3}$$

These RVs can be used as test statistics in F-tests.

Two way ANOVA and F-tests – F-distribution

The following plots show F-densities for a variety of parameter values.



Two way ANOVA and F-tests – running factor A test

Hypotheses

The **null and alternative hypotheses** for this test are

$$H_0: \ \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0$$

$$H_A: \ \text{at least two } \alpha_i \neq 0.$$

Test statistic

The test statistic is calculated from the sample data as

$$f_A^* = \frac{msa}{mse}$$

where msa and mse are observations of the RVs MSA and MSE respectively.

Under H_0, f_A^* is an observation of the $F(a-1, abn-ab)$ RV defined in (1).

Two way ANOVA and F-tests – running factor A test

Test decision

H_0 is rejected in favour of H_A at significance level α if

$$f_A^* > f_{1-\alpha},$$

where the quantile $f_{1-\alpha}$ is from $F(a-1, abn-ab)$ distribution.

Equivalently, H_0 is rejected if the p-value

$$p = \text{Prob}(F > f_A^*) < \alpha$$

where $F \sim F(a-1, abn-ab)$.

The null hypothesis H_0 is retained in any other case.

Two way ANOVA and F-tests – running factor B test

Hypotheses

The **null and alternative hypotheses** for this test are

$$H_0: \ \beta_1 = \beta_2 = \cdots = \beta_b = 0$$

$$H_A: \ \text{at least two } \beta_j \neq 0.$$

Test statistic

The test statistic is calculated from the sample data as

$$f_B^* = \frac{msb}{mse}$$

where msb and mse are observations of the RVs MSB and MSE respectively.

Under H_0, f_B^* is an observation of the $F(b-1, abn-ab)$ RV defined in (2).

Two way ANOVA and F-tests – running factor B test

Test decision

H_0 is rejected in favour of H_A at significance level α if

$$f_B^* > f_{1-\alpha},$$

where the quantile $f_{1-\alpha}$ is from $F(b-1, abn-ab)$ distribution.

Equivalently, H_0 is rejected if the p-value

$$p = \text{Prob}(F > f_B^*) < \alpha$$

where $F \sim F(b-1, abn-ab)$.

The null hypothesis H_0 is retained in any other case.

Two way ANOVA and F-tests – running interaction test

Hypotheses

The **null and alternative hypotheses** for this test are

$$H_0\text{: } \delta_{ij} = 0 \text{ for all } i \in \{1, \ldots, a\} \text{ and } j \in \{1, \ldots, b\}$$

$$H_A\text{: at least two } \delta_{ij} \neq 0.$$

Test statistic

The test statistic is calculated from the sample data as

$$f_D^* = \frac{msd}{mse}$$

where msd and mse are observations of the RVs MSD and MSE respectively.

Under H_0, f_D^* is an observation of the $F((a-1)(b-1), abn-ab)$ RV defined in (3).

Two way ANOVA and F-tests – running interaction test

Test decision

H_0 is rejected in favour of H_A at significance level α if

$$f_D^* > f_{1-\alpha},$$

where the quantile $f_{1-\alpha}$ is from $F((a-1)(b-1), abn-ab)$ distribution.

Equivalently, H_0 is rejected if the p-value

$$p = \text{Prob}(F > f_D^*) < \alpha$$

where $F \sim F((a-1)(b-1), abn-ab)$.

The null hypothesis H_0 is retained in any other case.

Example – experimental and blocking/nuisance factor

In this example we assess the effect of 3 different concentrations of fertilizer on crop yield.

There are 2 farms available for the experiment, with each farm having 3 plots to give a total of 6 plots.

The variables we consider are summarised in the table below (data in chapter4a.csv on Canvas).

Name	**Type**	**Description**
yield	numerical	crop yield
concentration	factor	fertilizer concentration: 1 (low), 2 (medium), 3 (high)
farm	factor	1 (farm 1), 2 (farm 2)

Example – experimental and blocking/nuisance factor

We would expect the 3 plots of farm 1 to provide homogenous growing conditions, as we would the 3 plots of farm 2.

However, it is possible that the 2 farms themselves have different growing conditions (e.g. if the 2 farms are in different geographical regions).

This means there are two ways that this experiment may be run.

Example – experimental and blocking/nuisance factor

ONE WAY CRD

If each farm has similar soil, growing conditions etc., this experiment could be conducted as a one factor, three level experiment with each fertilizer concentration tested twice.

To run the experiment, the 3 fertilizer concentrations are randomly allocated twice each across the 6 plots.

This is an example of a **completely randomised design (CRD)** where the data is analysed using one way ANOVA, as considered in the previous chapter.

The components of this design are

- *yield* is **response**
- the 3 fertilizer concentrations are the **treatments**
- the 3 groups of 2 plots each allocated a treatment are the **experimental units**
- the 6 plots are the **measurement units**.

Example – experimental and blocking/nuisance factor

TWO WAY CRBD

If the growing conditions on the farms are not homogenous, then we can expect different crop yields from the farms.

This extraneous source of variability is a **nuisance factor** that should be recognised and incorporated into the design as a **blocking factor**.

To run the experiment, the 3 fertilizer concentrations are allocated once each across the 3 plots of farm 1, and then again across the 3 plots of farm 2.

The components of this design are

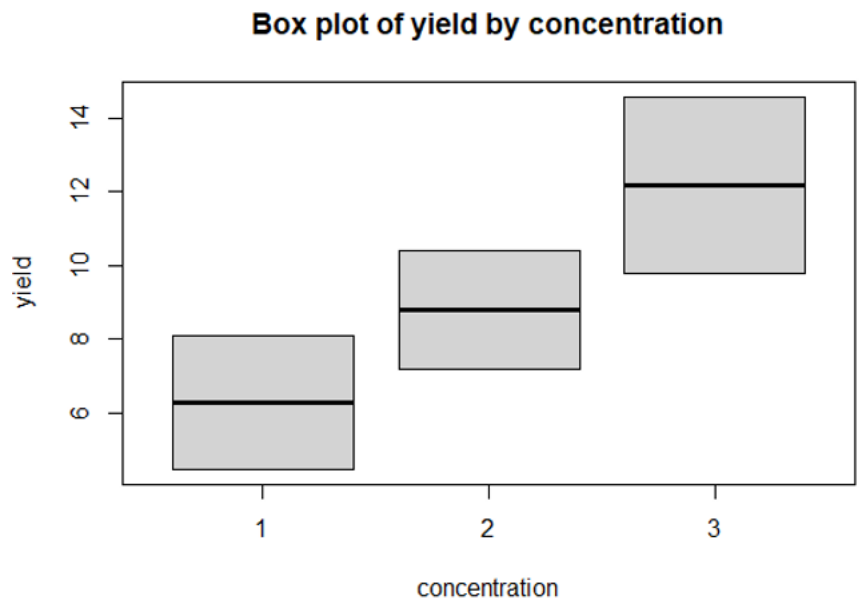
- *yield* is **response**
- the 3 fertilizer concentrations are the **treatments**
- the 2 farms are the **blocks**
- the 6 plots are the **experimental units**.
- the 6 plots are the **measurement units**.

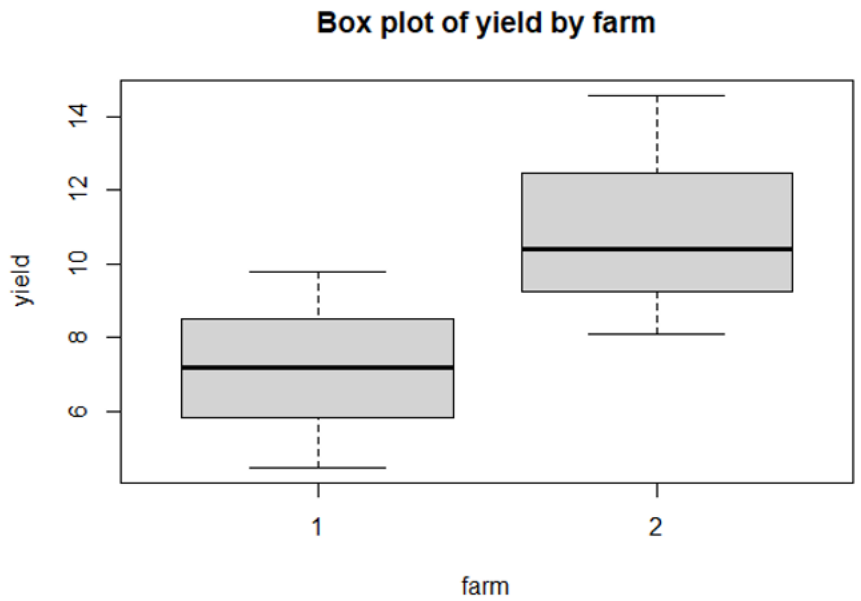
We see that a **CRBD** consists of a **CRD in each block**.

Example – experimental and blocking/nuisance factor

RUNNING THE TWO WAY CRBD

R produced the following boxplots.





The first plots suggests a significant effect of fertilizer concentration on crop yield.

The second plots suggests a significant effect of farm on crop yield – this supports including the factor *farm* as a blocking factor.

Example – experimental and blocking/nuisance factor

The sample data collected for the CRBD can be described as

$$yield_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$$

for $i \in \{1, 2, 3\}$ and $j \in \{1, 2\}$ where

- $yield_{ij}$ is crop yield of j-th farm using *fertilizer* i
- μ is global mean crop yield
- α_i treatment effect of *fertilizer* i
- β_j is block effect of *farm* j
- ϵ_{ij} is random effect of *farm* j using *fertilizer* i.

Note that the subscript k is not required and has been has been dropped as sample size $n = 1$.

The small sample also means that an interaction term cannot be included.

Example – experimental and blocking/nuisance factor

***F*-tests**

The hypotheses for the test on experimental factor *fertilizer* are

H_0: $\alpha_1 = \alpha_2 = \alpha_3 = 0$
H_A: at least two $\alpha_i \neq 0$

and for the test on blocking factor *farm* are

H_0: $\beta_1 = \beta_2 = 0$
H_A: at least two $\beta_j \neq 0$.

R produced the following output.

```
              Df Sum Sq Mean Sq F value Pr(>F)
concentration  2  34.74  17.371   50.63 0.0194 *
farm           1  22.25  22.250   64.85 0.0151 *
Residuals      2   0.69   0.343
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Both null hypotheses can be rejected (p-values both below our usual significance level 0.05) and we conclude that both fertilizer concentration and farm have significant effects on mean yield.

Example – experimental and blocking/nuisance factor

Post-hoc analysis

From the first F-test we know that mean yield from at least one fertilizer concentration is different from the others.

Tukey analysis can determine which one(s), with R producing the following output.

```
 contrast estimate    SE df lower.CL upper.CL t.ratio p.value
 1 - 2       -2.50 0.586  2    -5.95   0.9489  -4.271 0.0910
 1 - 3       -5.87 0.586  2    -9.32  -2.4223 -10.026 0.0178
 2 - 3       -3.37 0.586  2    -6.82   0.0792  -5.756 0.0523

Results are averaged over the levels of: farm
Confidence level used: 0.95
Conf-level adjustment: tukey method for comparing a family of 3 estimates
P value adjustment: tukey method for comparing a family of 3 estimates
```

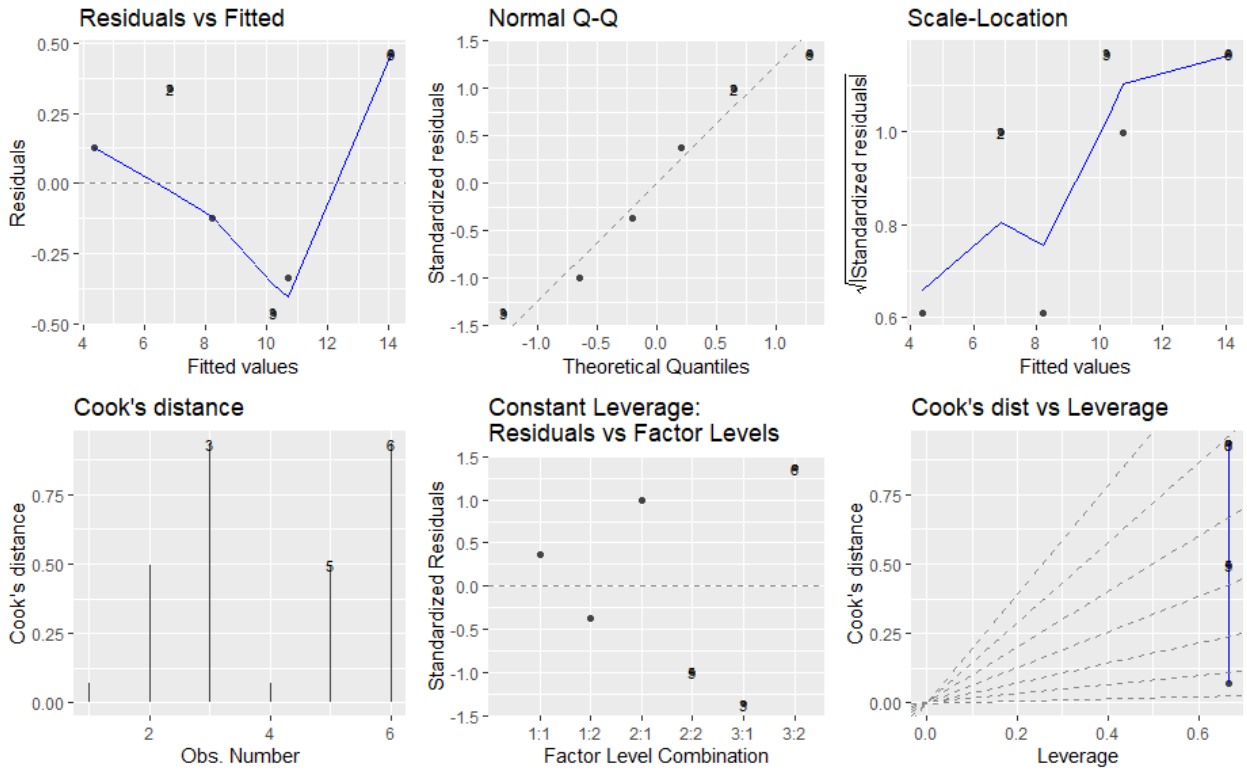
At the 0.05 significance level we see that the only statistically-significant difference in mean crop yield is between fertilizer concentrations 1 and 3 (p-value $p = 0.0178 < 0.05$).

As for the farm blocking factor, Tukey analysis is not required as this factor has only two levels.

Example – experimental and blocking/nuisance factor

Assumptions

We now check the assumptions via the plots of the residuals produced by R below.



With only six data points there is not much to go on (see Chapter 3 for a short summary of what to look for).

Example – experimental and blocking/nuisance factor

Normality test

Finally we test the normality of the residuals with hypotheses

H_0: the residuals $\hat{\epsilon}_{ij}$ are normally distributed

H_A: the residuals $\hat{\epsilon}_{ij}$ are not normally distributed.

We use significance level $\alpha = 0.05$ and from R obtain the following output.

```
        Shapiro-Wilk normality test

data:  res
W = 0.94651, p-value = 0.7119
```

We retain the null (p-value $p = 0.7119 > 0.05$) and conclude that there is little evidence suggesting the residuals are not normally distributed.

Example – experimental and blocking/nuisance factor

Actually, we should be careful using two way ANOVA here because of the very small sample.

Although no gross departure from normality has been detected, with such a small sample detection will always be difficult.

Likewise for the assumptions of constant variance and independence – departures from these assumptions will be difficult to detect with such a small sample.

For one way ANOVA, we presented a non-parametric alternative, the Kruskal-Wallis test, but we will not explore this for two way analysis.

Example – experimental and blocking/nuisance factor

RUNNING THE ONE WAY CRD

Now let us re-run our analysis, but this time without the farm blocking factor.

If we do not use the blocking factor we are left with a **nuisance factor**, which is a known source of variation not controlled for in our model.

The sample data collected for the CRD can be described as

$$yield_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

for $i \in \{1, 2, 3\}$ and $j \in \{1, 2\}$ where

- $yield_{ij}$ is crop yield of the j-th plot using *fertilizer* i
- μ is global mean crop yield
- α_i is treatment effect of *fertilizer* i
- ϵ_{ij} is random effect of the j-th plot using *fertilizer* i.

Example – experimental and blocking/nuisance factor

***F*-test**

The hypotheses for the test on experimental factor *fertilizer* are

$$H_0: \ \alpha_1 = \alpha_2 = \alpha_3 = 0$$

$$H_A: \ \text{at least two } \alpha_i \neq 0.$$

R provides the following output.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
concentration	2	34.74	17.371	2.272	0.251
Residuals	3	22.94	7.646		

In this case we retain the null hypothesis (p-value $p = 0.251 > 0.05$) and conclude that fertilizer concentration does not have a significant effect on mean crop yield.

Example – experimental and blocking/nuisance factor

DIFFERENT CONCLUSIONS FROM CRBD AND CRD

So what explains the different conclusions drawn from the two way and one way analyses?

The mathematical explanation can be seen by comparing the ANOVA tables.

Without the blocking factor in place, the sum squares from the *farm* nuisance factor are included in *sse*, which is 22.94 compared to 0.69 with the blocking factor in place.

This results in the lower test statistic f_A^* (2.272 down from 50.63) with associated higher p-value (0.251 up from 0.0194).

It also results in the lower R^2.

Lesson. When designing an experiment, carefully consider all possible sources of variation and, where likely to be large, control for these in your experiment design.

Example – experimental factors with interaction

In the next example we assess the effect of 4 different vitamin and 3 different diet regimes on the weight gain of farmed pigs.

A total of 240 pigs were used in the study, with the variables we consider summarised in the table below (data in chapter4b.csv on Canvas).

Name	**Type**	**Description**
gain	numerical	weight gain
vitamin	factor	vitamin regime: 1, 2, 3, 4
diet	factor	diet regime: 1, 2, 3

Example – experimental factors with interaction

To run the experiment, the $4 \times 3 = 12$ combinations of vitamin and diet regimes were each randomly allocated to 20 pigs.

The components of this design are

- *gain* is **response**
- the $4 \times 3 = 12$ combinations of vitamin and diet regimes are the **treatments**
- the 12 groups of 20 pigs each allocated a treatment are the **experimental units**.
- the 240 pigs are the **measurement units**.

This is another example of a **CRD**.

Example – experimental factors with interaction

The sample data collected for the CRD can be described as

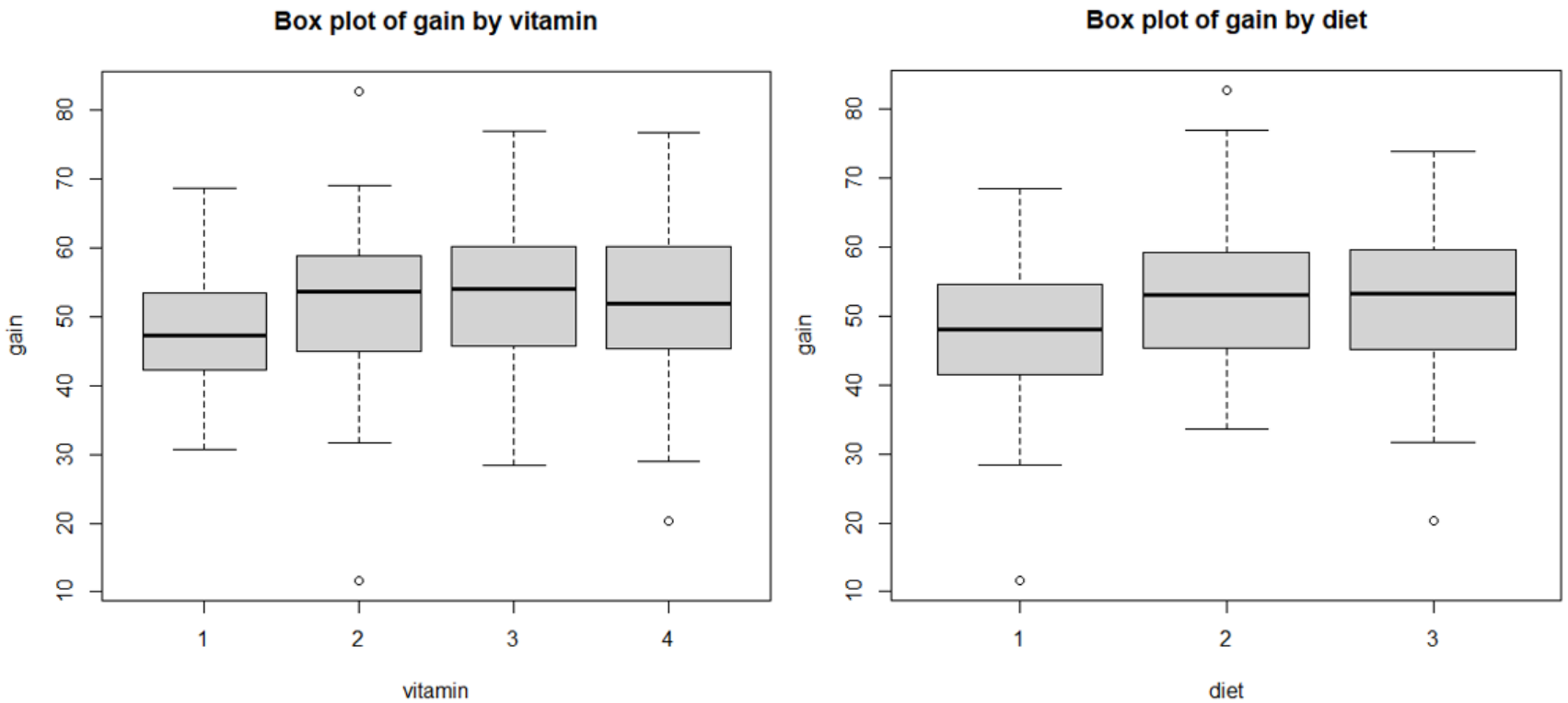
$$gain_{ijk} = \mu + \alpha_i + \beta_j + \delta_{ij} + \epsilon_{ijk}$$

for $i \in \{1, 2, 3, 4\}$, $j \in \{1, 2, 3\}$ and $k \in \{1, 2, \ldots, 20\}$ where

- $gain_{ijk}$ is weight gain of the k-th pig receiving *vitamin* i and *diet* j
- μ as global mean weight gain
- α_i is the treatment effect of *vitamin* i
- β_j is the treatment effect of *diet* j
- δ_{ij} treatment effect of *vitamin* and *diet* interaction at levels i and j
- ϵ_{ijk} is the random effect of the k-th pig receiving *vitamin* i and *diet* j.

Example – experimental factors with interaction

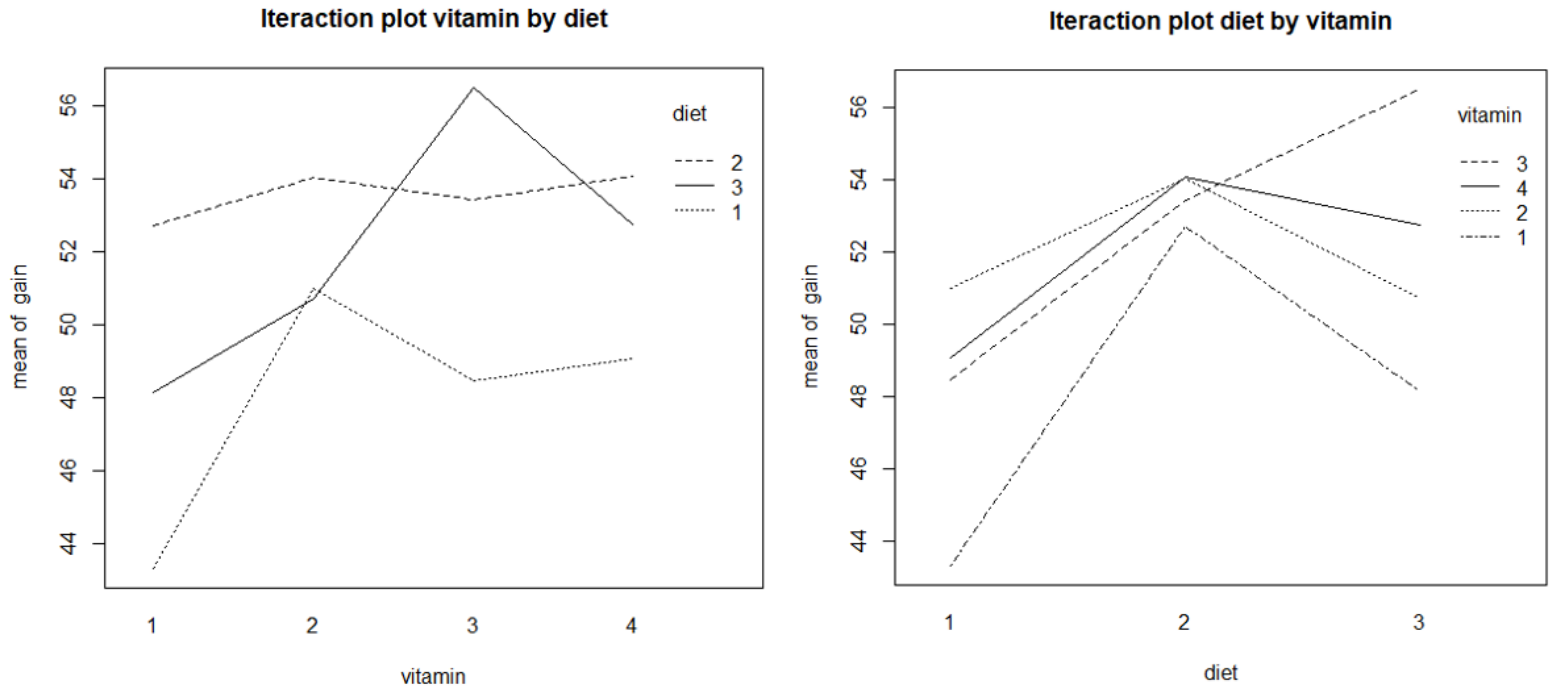
R produced the following boxplots.



From the plots we see differences in weight gain due to different vitamin and different diet regimes, although the comparatively large within treatment variation makes it difficult to assess whether the differences will turn out to be statistically-significant.

Example – experimental factors with interaction

We can also have R produce **interaction plots**.



To detect interaction we look for non-parallel **trace lines**.

In this case most of the trace lines seem parallel, indicating that factor interaction may not be statistically-significant.

Example – experimental factors with interaction

***F*-tests**

The hypotheses for the test on factor *vitamin* are

$$H_0: \ \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$$
$$H_A: \ \text{at least two } \alpha_i \neq 0,$$

for the test on factor *diet* are

$$H_0: \ \beta_1 = \beta_2 = \beta_3 = 0$$
$$H_A: \ \text{at least two } \beta_j \neq 0$$

and for factor interaction are

$$H_0: \ \delta_{i,j} = 0 \text{ for all } i \in \{1, \ldots, 4\} \text{ and } j \in \{1, \ldots, 3\}$$
$$H_A: \ \text{at least two } \delta_{i,j} \neq 0.$$

Example – experimental factors with interaction

R produced the following output.

```
              Df Sum Sq Mean Sq F value  Pr(>F)
vitamin        3    810   269.9   2.721 0.04523 *
diet           2   1341   670.4   6.758 0.00141 **
vitamin:diet   6    605   100.9   1.017 0.41480
Residuals    228  22617    99.2
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The null hypotheses for both factor F-tests can be rejected (p-values both below our usual significance level 0.05) and we conclude that both vitamin and diet have significant effects on mean weight gain.

The null hypothesis for the interaction test is retained (p-value $p = 0.4148 > 0.05$) and we conclude the vitamin and diet interaction has insignificant effect on mean weight gain.

Example – experimental factors with interaction

Post-hoc analysis
Because of the presence of the interaction term, post-hoc analysis is more complicated.

As the interaction term is insignificant we could re-run the two way analysis without interaction and then proceed with post-hoc analysis individually on each of the two factors.

However, for the purpose of illustration we will retain the interaction term.

In this case, the correct way to proceed is to perform the post-hoc analysis for one factor at each level of the other factor.

Example – experimental factors with interaction

R produced the following output for factor *vitamin* for each level of *diet*.

```
diet = 1:
 contrast estimate   SE  df lower.CL upper.CL t.ratio p.value
 1 - 2     -7.6766 3.15 228   -15.83    0.475 -2.437  0.0730
 1 - 3     -5.1355 3.15 228   -13.29    3.016 -1.631  0.3635
 1 - 4     -5.7448 3.15 228   -13.90    2.407 -1.824  0.2647
 2 - 3      2.5411 3.15 228    -5.61   10.692  0.807  0.8512
 2 - 4      1.9319 3.15 228    -6.22   10.083  0.613  0.9278
 3 - 4     -0.6092 3.15 228    -8.76    7.542 -0.193  0.9974

diet = 2:
 contrast estimate   SE  df lower.CL upper.CL t.ratio p.value
 1 - 2     -1.3073 3.15 228    -9.46    6.844 -0.415  0.9759
 1 - 3     -0.7267 3.15 228    -8.88    7.425 -0.231  0.9957
 1 - 4     -1.3385 3.15 228    -9.49    6.813 -0.425  0.9742
 2 - 3      0.5806 3.15 228    -7.57    8.732  0.184  0.9978
 2 - 4     -0.0312 3.15 228    -8.18    8.120 -0.010  1.0000
 3 - 4     -0.6118 3.15 228    -8.76    7.540 -0.194  0.9974

diet = 3:
 contrast estimate   SE  df lower.CL upper.CL t.ratio p.value
 1 - 2     -2.5634 3.15 228   -10.71    5.588 -0.814  0.8479
 1 - 3     -8.3550 3.15 228   -16.51   -0.204 -2.653  0.0422
 1 - 4     -4.6176 3.15 228   -12.77    3.534 -1.466  0.4597
 2 - 3     -5.7916 3.15 228   -13.94    2.360 -1.839  0.2579
 2 - 4     -2.0542 3.15 228   -10.21    6.097 -0.652  0.9147
 3 - 4      3.7374 3.15 228    -4.41   11.889  1.187  0.6359

Confidence level used: 0.95
Conf-level adjustment: tukey method for comparing a family of 4 estimates
P value adjustment: tukey method for comparing a family of 4 estimates
```

The only significant difference in mean weight gain (at 0.05 significance level) is between *vitamin* levels 1 and 3 when $diet = 3$.

Example – experimental factors with interaction

R produced the following output for factor *diet* for each level of *vitamin*.

```
vitamin = 1:
 contrast estimate   SE  df lower.CL upper.CL t.ratio p.value
 1 - 2      -9.397 3.15 228   -16.83   -1.967 -2.984  0.0088
 1 - 3      -4.825 3.15 228   -12.26    2.605 -1.532  0.2779
 2 - 3       4.572 3.15 228    -2.86   12.002  1.452  0.3164

vitamin = 2:
 contrast estimate   SE  df lower.CL upper.CL t.ratio p.value
 1 - 2      -3.028 3.15 228   -10.46    4.402 -0.961  0.6021
 1 - 3       0.288 3.15 228    -7.14    7.718  0.091  0.9954
 2 - 3       3.316 3.15 228    -4.11   10.746  1.053  0.5445

vitamin = 3:
 contrast estimate   SE  df lower.CL upper.CL t.ratio p.value
 1 - 2      -4.988 3.15 228   -12.42    2.442 -1.584  0.2547
 1 - 3      -8.045 3.15 228   -15.48   -0.615 -2.554  0.0303
 2 - 3      -3.057 3.15 228   -10.49    4.374 -0.970  0.5963

vitamin = 4:
 contrast estimate   SE  df lower.CL upper.CL t.ratio p.value
 1 - 2      -4.991 3.15 228   -12.42    2.439 -1.585  0.2544
 1 - 3      -3.698 3.15 228   -11.13    3.732 -1.174  0.4698
 2 - 3       1.293 3.15 228    -6.14    8.723  0.410  0.9114

Confidence level used: 0.95
Conf-level adjustment: tukey method for comparing a family of 3 estimates
P value adjustment: tukey method for comparing a family of 3 estimates
```

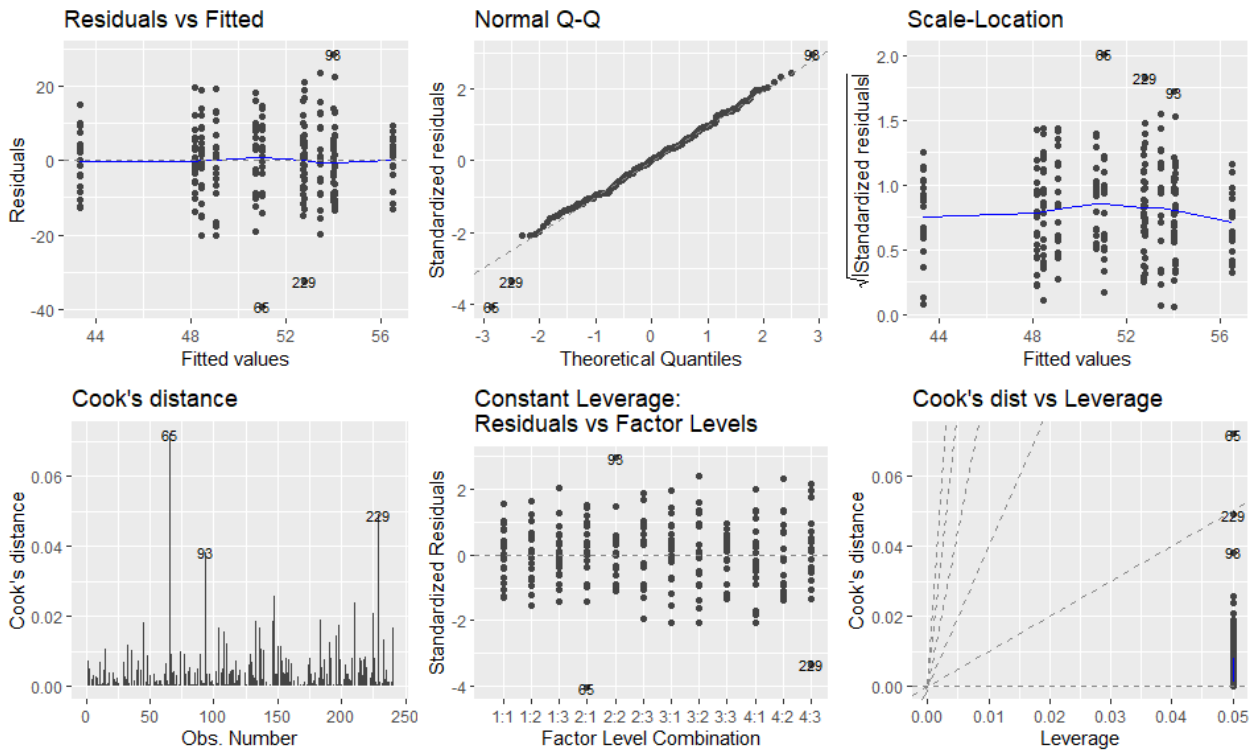
The significant differences in mean weight gain (at 0.05 significance level) are between

- *diet* levels 1 and 2 when $vitamin = 1$
- *diet* levels 1 and 3 when $vitamin = 3$.

Example – experimental factors with interaction

Assumptions

We now check the assumptions via the plots of the residuals produced by R below.



Nothing in these plots suggest any obvious issues with the modelling assumptions (see Chapter 3 for a short summary of what to look for).

Example – experimental factors with interaction

Normality test

Finally we test the normality of the residuals with hypotheses

H_0: the residuals $\hat{\epsilon}_{ij}$ are normally distributed

H_A: the residuals $\hat{\epsilon}_{ij}$ are not normally distributed.

We use significance level $\alpha = 0.05$ and from R obtain the following output.

```
        Shapiro-Wilk normality test

data:  res
W = 0.98947, p-value = 0.07792
```

We retain the null (p-value $p = 0.07792 > 0.05$) and conclude that there is no strong evidence suggesting the residuals are not normally distributed.

References I