

# MA50259: Statistical Design of Investigations

## Coursework 1 (2024)

### Disclaimer:

AI software (RStudio built-in Co-Pilot and ChatGPT) are used in this coursework for code, RMarkdown formatting, explanation suggestions, grammar check, and debugging purposes.

## Part 1: Baking Powder Experiment

In an experiment to study the effect of the amount of baking powder in a biscuit dough upon the rise heights (measured in centimetre) of the biscuits, four levels of baking powder were tested and four replicate biscuits were made with each level in a random order. The results are shown in the table below (tsp stands for teaspoon)

.25 tsp	0.5 tsp	.75 tsp	1 tsp
11.4	27.8	47.6	61.6
11.0	29.2	47.0	62.4
11.3	26.8	47.3	63.0
9.5	26.0	45.5	63.9

1. What is the experimental unit and which type of experimental design is this?

---

### Answer:

**Definition** of an *experimental unit*: The item under study upon which something is changed. This could be things, human subjects, or just a point in time.

The *experimental unit* in this experiment is each individual biscuit.

**Definition** of a *experimental design*: A collection of experiments or runs that is planned in advance of the actual execution. The particular runs selected in an experimental design will depend upon the purpose of the design.

This type of experimental design is known as completely randomised design (CRD).

---

2. Read-in the data in the table above in R as a `data.frame` in such a way that it can be used by the `lm` function without any modification, that is, your data frame should have 2 columns and 16 rows. The variables in the dataframe should be called `riseht` and `type`. The data frame should be called `biscuits`

```
options(width = 50)
biscuits <- data.frame(riseht=c(11.4,11.0,11.3,9.5,27.8,29.2,26.8,26.0,47.6,47.0
                              ,47.3,45.5,61.6,62.4,63.0,63.9),
                      type=rep(c(0.25,0.5,0.75,1),each=4))
biscuits$type <- factor(biscuits$type)
```

3. Construct in R, the design matrix  $\mathbf{X}$  corresponding to the model:

$$y_{ij} = \mu + \tau_j + \epsilon_{ij}$$

where  $\tau_{0.25}, \tau_{0.50}, \tau_{0.75}$  and  $\tau_{1.0}$  are the effects of the levels of the baking powder. All  $\epsilon_{ij} \sim N(0, \sigma^2)$  and are mutually independent.

```
y <- biscuits$riseht
t <- length(unique(factor(biscuits$type))) # number of treatments
r <- nrow(biscuits)/t # number of replicates per treatment
n <- nrow(biscuits) # total number of experimental units
levels <- unique(factor(biscuits$type)) # levels of the baking powder
fact <- gl(t,r,labels=levels) # factor variable
Z <- model.matrix(~fact-1) # matrix Z as the means model of a CRD
X <- cbind(1,Z) # design matrix X as the treatment effects model of a CRD

# naming the columns
colnames(X) <- c("reference", "0.25 tsp", "0.50 tsp", "0.75 tsp", "1.00 tsp")

# Display the design matrix X
kable(X, align = "c", caption = "Design Matrix X")
```

Table 1: Design Matrix X

reference	0.25 tsp	0.50 tsp	0.75 tsp	1.00 tsp
1	1	0	0	0
1	1	0	0	0
1	1	0	0	0
1	1	0	0	0
1	0	1	0	0
1	0	1	0	0
1	0	1	0	0
1	0	1	0	0
1	0	0	1	0
1	0	0	1	0
1	0	0	1	0
1	0	0	0	1
1	0	0	0	1
1	0	0	0	1
1	0	0	0	1

4. How many unknown parameters are in the model?

---

Answer:

5 unknown parameters:

- i. One reference rise height of the biscuit when there is no treatment  $\mu$
- ii. Four treatment effects  $\tau_i$  for  $i = 1, 2, 3, 4$

---

5. Find the rank of the design matrix  $\mathbf{X}$ . You should justify your answer.

```
cat("Rank of design matrix X: ", qr(X)$rank)
```

```
## Rank of design matrix X: 4
```

---

Answer:

The rank of the design matrix  $\mathbf{X}$  is 4. This is because the design matrix  $\mathbf{X}$  has 5 columns and the rank of a matrix is the maximum number of linearly independent column vectors in the matrix. The first column of  $\mathbf{X}$  is a column of ones, and thus the remaining 4 columns are dependent on the first column but not to one another. Therefore, the rank of  $\mathbf{X}$  is 4 as shown in the QR decomposition command above.

---

6. Perform the analysis of variance to test the hypothesis of no treatment effect. You should do the computations using the theory given in the lectures and not simply use the command `aov`.

```
# The ANOVA table is computed using the following steps:
# Calculate grand mean
grand_mean <- mean(y)

# Calculate the total sum of squares (SSTotal)
SSTotal <- sum((y - grand_mean)^2)

# Calculate the treatment sum of squares (SSTreatment)
means <- tapply(y, biscuits$type, mean)
SSTreatment <- sum(r*means^2) - n*grand_mean^2

# Calculate the error sum of squares (SSError)
SSError <- SSTotal - SSTreatment

# Alternative code for SSError
# t(matrix(y,ncol=1))%*(diag(n)-(1/n)*rep(1,n)%*t(rep(1,n)))*%matrix(y,ncol= 1)

# Calculate the degrees of freedom for the treatment (dfTreatment)
dfTreatment <- t - 1

# Calculate the degrees of freedom for the error (dfError)
dfError <- n - t

# Calculate the mean square for the treatment (MSTreatment)
MSTreatment <- SSTreatment/dfTreatment

# Calculate the mean square for the error (MSError)
MSError <- SSError/dfError

# Calculate the F-statistic
F_stat <- MSTreatment/MSError
```

```
# Calculate the p-value
p_value <- 1 - pf(F_stat,dfTreatment,dfError)

kable(
  data.frame(
    `Source` = c("Treatment", "Error", "Total"),
    `Sum of Squares` = c(SSTreatment, SSError, SSTotal),
    `Degrees of Freedom` = c(dfTreatment, dfError, n - 1),
    `Mean Square` = c(round(MSTreatment, 4), round(MSError, 4), ""),
    `F-statistic` = c(round(F_stat, 4), "", ""),
    `p-value` = c(sprintf("%.2e", p_value), "", "")
  ),
  align = "c"
)
```

Source	Sum.of.Squares	Degrees.of.Freedom	Mean.Square	F.statistic	p.value
Treatment	6145.7319	3	2048.5773	1822.6452	3.33e-16
Error	13.4875	12	1.124		
Total	6159.2194	15			

Answer:

The hypothesis test is as follows:

The null hypothesis,  $H_0$ :  $\tau_1 = \tau_2 = \tau_3 = \tau_4$ .

The alternative hypothesis,  $H_a$ : at least two of the  $\tau$ s differs.

Let the significance level be  $\alpha = 0.05$ . The null hypothesis is that there is no treatment effect, that is,  $\tau_1 = \tau_2 = \tau_3 = \tau_4$ . The alternative hypothesis is that at least one of the  $\tau$ s are not equal. Given that we obtained a p-value of  $3.33 \times 10^{-16}$ , which is less than the significance level of 0.05 (5% is commonly used in a typical hypothesis test), we reject the null hypothesis and conclude that there is a treatment effect.

7. Formulate a contrast to test the hypothesis that increase in rise height is a linear function of the increase in baking powder in the dough, and test this hypothesis.

```
# Contrast Reference: https://online.stat.psu.edu/stat555/node/73/
# Other references:
# https://en.wikipedia.org/wiki/Contrast_(statistics)
# Statistical Inference by Casella and Berger
# Design and Analysis of Experiments with R by John Lawson
# ChatGPT used only to understand the concept of contrast in simple terms ONLY

# Formulate the contrast
# Since we are testing whether the increase in rise height is a linear function of the
# increase in baking powder in the dough, we can use the contrast:
linear_contrast_c <- c(0, -3, -1, 1, 3)
ginverse <- rbind(0, cbind(0, solve(t(Z) %*% Z)))
beta <- ginverse %*% t(X) %*% y
```

```

# Calculate the contrast estimate
contrast_estimate <- linear_contrast_c %*% beta

# Calculate the standard error of the contrast estimate
# The standard error of the contrast estimate is calculated using the formula:
# SE = sqrt(MSError * sum(c^2)/t)
# Refer: https://websites.umich.edu/~gonzo/coursenotes/file3.pdf page 3-3
SE_contrast_estimate <- sqrt(MSError * sum(linear_contrast_c^2)/t)

# Calculate the t-statistic
t_stat <- contrast_estimate / SE_contrast_estimate

# Calculate the p-value in a two-tailed test
# Refer: https://www.geeksforgeeks.org/how-to-calculate-the-p-value-of-a-t-score-in-r/
p_value_contrast <- 2 * pt(-abs(t_stat), dfError)

kable(
  data.frame(
    `Contrast` = c("Contrast Test"),
    `Contrast Estimate` = round(contrast_estimate, 2),
    `Contrast Standard Error` = round(SE_contrast_estimate, 2),
    `t-Statistic` = round(t_stat, 2),
    `p-Value` = sprintf("%.2e", p_value_contrast)
  ),
  align = "c"
)

```

Contrast	Contrast.Estimate	Contrast.Standard.Error	t.Statistic	p.Value
Contrast Test	175.18	2.37	73.89	2.51e-17

### Answer:

A contrast is a linear combination  $\mathbf{c}^T \beta = c_0 \mu + \sum_{i=1}^t c_i \tau_i$  of the treatment means, where  $\sum_{i=0}^t c_i = 0$ .

A contrast allows us to test the hypothesis that increase in rise height is a linear function of the increase in baking powder in the dough.

Let  $c_0 = 0$  as we are only interested to check for presence of a linear trend due to varying the amount of baking powder used.

Since we are testing whether the increase in rise height is a linear function of the increase in baking powder in the dough, we can use the contrast:

$$\mathbf{c} = (0, -3, -1, 1, 3)$$

which is a classical orthogonal contrast coefficients for testing the linear trend in the treatment means. This also fulfils the requirement that the sum of the contrast weights is equal to zero  $\sum \mathbf{c} = 0 - 3 - 1 + 1 + 3 = 0$ .

The contrast weights are chosen such that the first treatment level is weighted by -3, the second treatment level is weighted by -1, the third treatment level is weighted by 1, and the fourth treatment level is weighted by 3. Doing so allows the  $\tau_2$  and  $\tau_3$  to serve as a pivot point for the linear trend, while multiplying  $\tau_1$  by -3 allows us to “pull” the contrast value down, and multiplying  $\tau_4$  by 3 allows us to “push” the contrast value up. Thus allows symmetrical distribution of the contrast weights around the pivot point, making it suitable to check a simple linear trend.

Setting up our hypothesis test:

Let:

The null hypothesis,  $H_0$ :  $\tau_1 = \tau_2 = \tau_3 = \tau_4$ .

The alternative hypothesis,  $H_a$ : at least two of the  $\tau$ s differs.

Calculating beta:

From the equation:

$$\mathbf{X}^T \mathbf{X} \boldsymbol{\beta} = \mathbf{X}^T \mathbf{y}$$

we can rewrite the equation to find  $\tilde{\boldsymbol{\beta}}$  as:

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^- \mathbf{X}^T \mathbf{y}$$

where  $(\mathbf{X}^T \mathbf{X})^-$  is the generalized inverse of  $\mathbf{X}^T \mathbf{X}$ .

A generalized inverse of a square matrix  $\mathbf{A}$  is another square matrix  $\mathbf{A}^-$  satisfying

$$\mathbf{A} \mathbf{A}^- \mathbf{A} = \mathbf{A}$$

The generalized inverse of a non-full rank matrix is used to estimate the variance of the experimental error using the technique

$$(\mathbf{X}^T \mathbf{X})^- = \begin{pmatrix} 0 & 0 \\ 0 & (\mathbf{Z}^T \mathbf{Z})^{-1} \end{pmatrix}$$

where  $\mathbf{X}$  is the design matrix and  $\mathbf{Z}$  is the matrix of means.

```
kable(means,
      col.names = c(
        "Treatment Level",
        "Means Within Individual Treatment Levels"
      ),
      align = "c"
    )
```

Treatment Level	Means Within Individual Treatment Levels
0.25	10.800
0.5	27.450
0.75	46.850
1	62.725

The contrast estimates to 175.18 using the formula:

$$c^T \hat{\boldsymbol{\beta}} = 0 + c^T \hat{\tau}_i = \text{contrast weight} \times \text{mean within individual treatment levels} = (-3 \times 10.800) + (-1 \times 27.450) + (1 \times 46.850) + (3 \times 62.725) = 175.18$$

Next, we calculate the standard error of the contrast using the formula:

$$\text{standard error (contrast)} = \sqrt{MSE \times \frac{\sum_{i=1}^t c_i^2}{\text{no. of treatment levels checked in this contrast}}} = \sqrt{1.124 \times \frac{(-3)^2 + (-1)^2 + (1)^2 + (3)^2}{4}} = 2.37$$

where

$MSE$  = mean square error

$c_i$  = contrast weight

$t$  = number of treatments

The t-statistic is calculated using the formula:

$$t = \frac{\text{contrast estimate}}{\text{standard error (contrast)}} = \frac{175.18}{2.37} = 73.89$$

The p-value is calculated using a t-score of 73.89 and 12 degrees of freedom ( $n-t = 16-4 = 12$ ). We obtain a p-value of  $2.51 \times 10^{-17}$ , which is less than the significance level of 0.05 (*5% is commonly used in a typical hypothesis test*). Therefore, we reject the null hypothesis and conclude that there is a linear trend in the treatment means given that the p-value is less than 0.05 ( $2.51 \times 10^{-17} < 0.05$ ).

8. Estimate the variance  $\sigma^2$  of the experimental error.

```
# Estimate the variance of the experimental error
# The variance of the experimental error is estimated using the formula:
# \hat{\sigma}^2 = SSE / (n - t)
# Reference: Design and Analysis of Experiments with R by John Lawson

SSE_matrix_method <- t(y) %*% (diag(n) - X %*% ginverse %*% t(X)) %*% y
variance_of_experimental_error <- round(SSE_matrix_method / (n - t), 4)
cat("Variance of the experimental error: ", variance_of_experimental_error)

## Variance of the experimental error: 1.124
```

Answer:

The variance of the experimental error is estimated to be 1.124 using the formula  $\hat{\sigma}^2 = \frac{SSE}{n-t}$ .

Sum of Square Error (SSE) is calculated using the formula

$$\mathbf{y}^T (\mathbf{I} - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T) \mathbf{y}$$

where:

- $\mathbf{y}$  is the response vector
- $\mathbf{I}$  is the identity matrix
- $n$  is the total number of experimental units
- $t$  is the number of treatments.

---

9. Discuss whether the assumptions for the considered linear model are justified.

---

Answer:

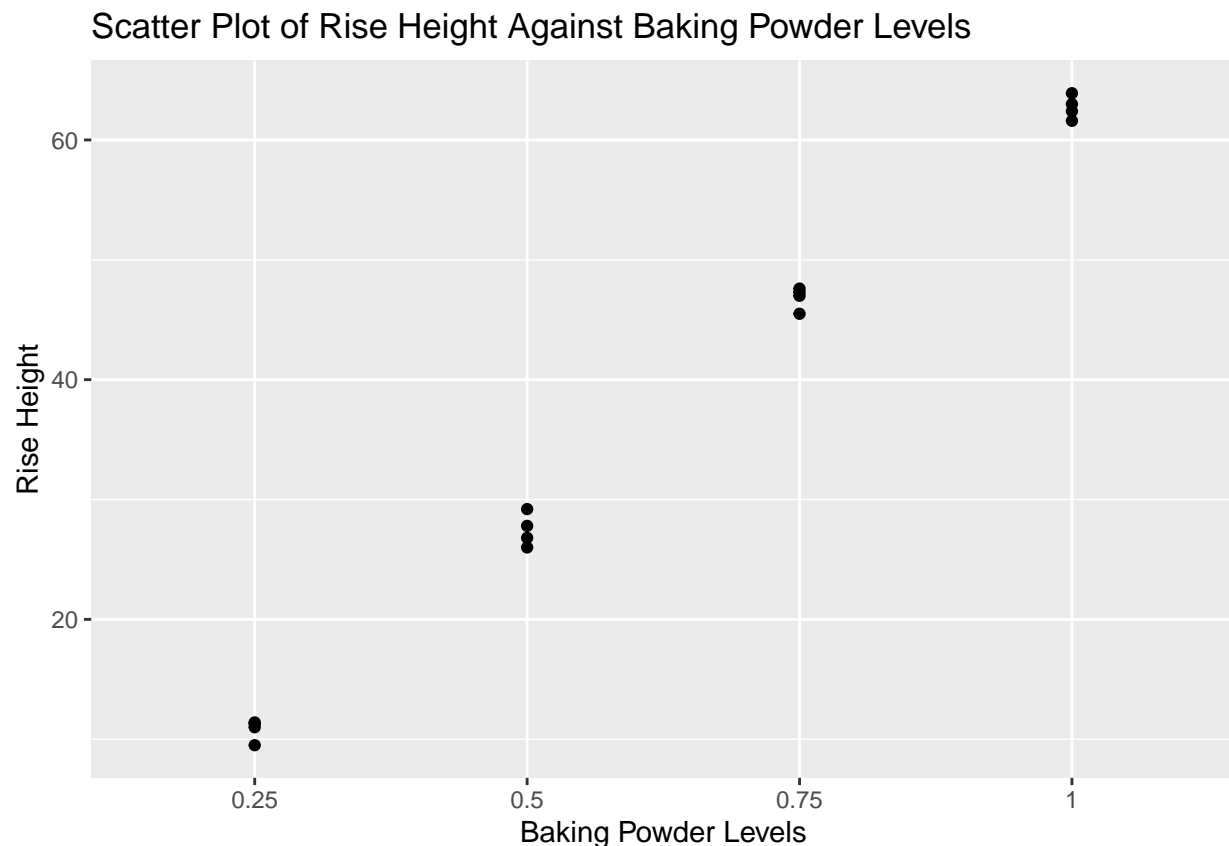
The assumptions for the considered linear model are:

Assumption 1: There is an additive effect of the treatment levels on the means, e.g.  $\mu_i = \mu + \tau_i$ .

The assumption is justified as based on our contrast test, we have found that there is a linear trend in the treatment means. This means that the increase in rise height is a linear function of the increase in baking powder in the dough.

Not to mention a simple scatter plot also shows a linear trend in the treatment means as we increase the amount of baking powder in the dough.

```
library(ggplot2)
ggplot(biscuits, aes(x=type, y=riseht)) +
  geom_point() +
  labs(
    title="Scatter Plot of Rise Height Against Baking Powder Levels",
    x="Baking Powder Levels",
    y="Rise Height"
  )
```

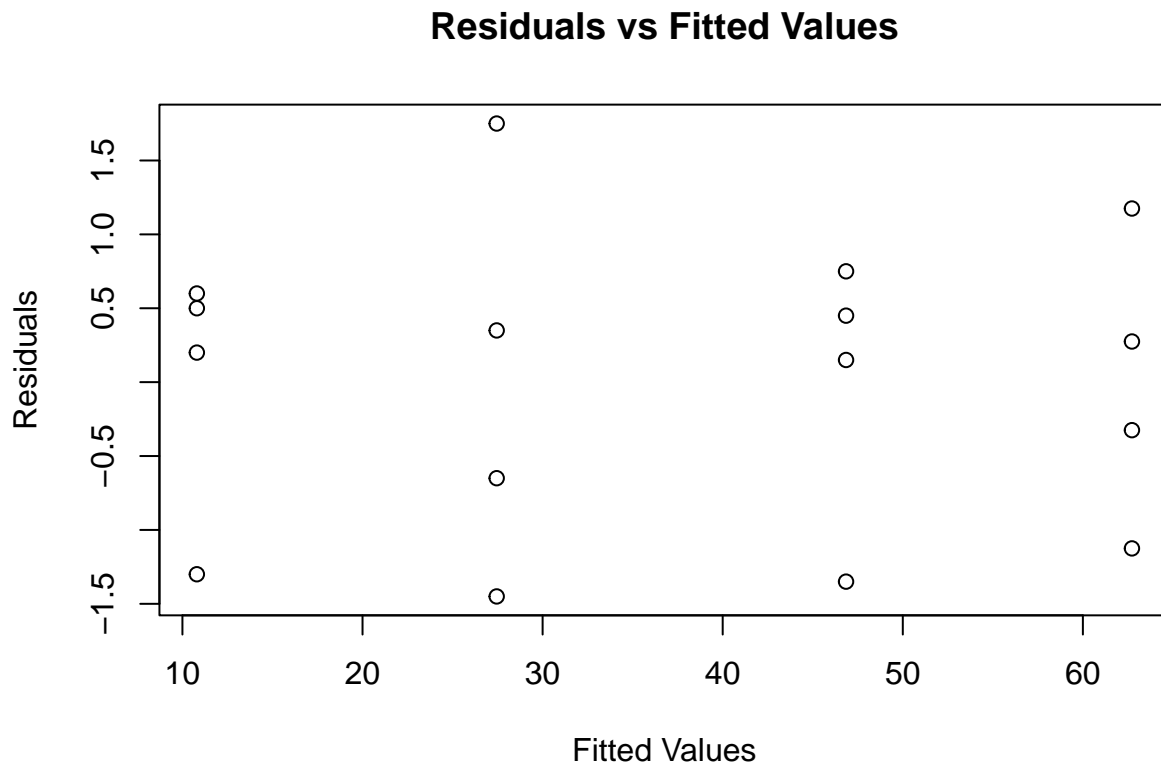




Assumption 2: If the units are homogeneous, the probability distribution of the response  $y_{ij}$  or the experimental error  $\epsilon_{ij}$ , under the same treatment level, is the same.

To check the assumption of homogeneity of variance, we can use the residuals vs fitted values plot. The residuals vs fitted values plot will help us to assess whether the variance of the residual is consistent across all levels of the independent variable. If the variance of the residuals is consistent across all levels of the independent variable, then the assumption of homogeneity of variance is met.

```
# Reference: Design and Analysis of Experiments with R by John Lawson
residuals <- residuals(aov(riseht ~ type, data = biscuits))
plot(
  fitted(aov(riseht ~ type, data = biscuits)),
  residuals, xlab = "Fitted Values",
  ylab = "Residuals",
  main = "Residuals vs Fitted Values"
)
```

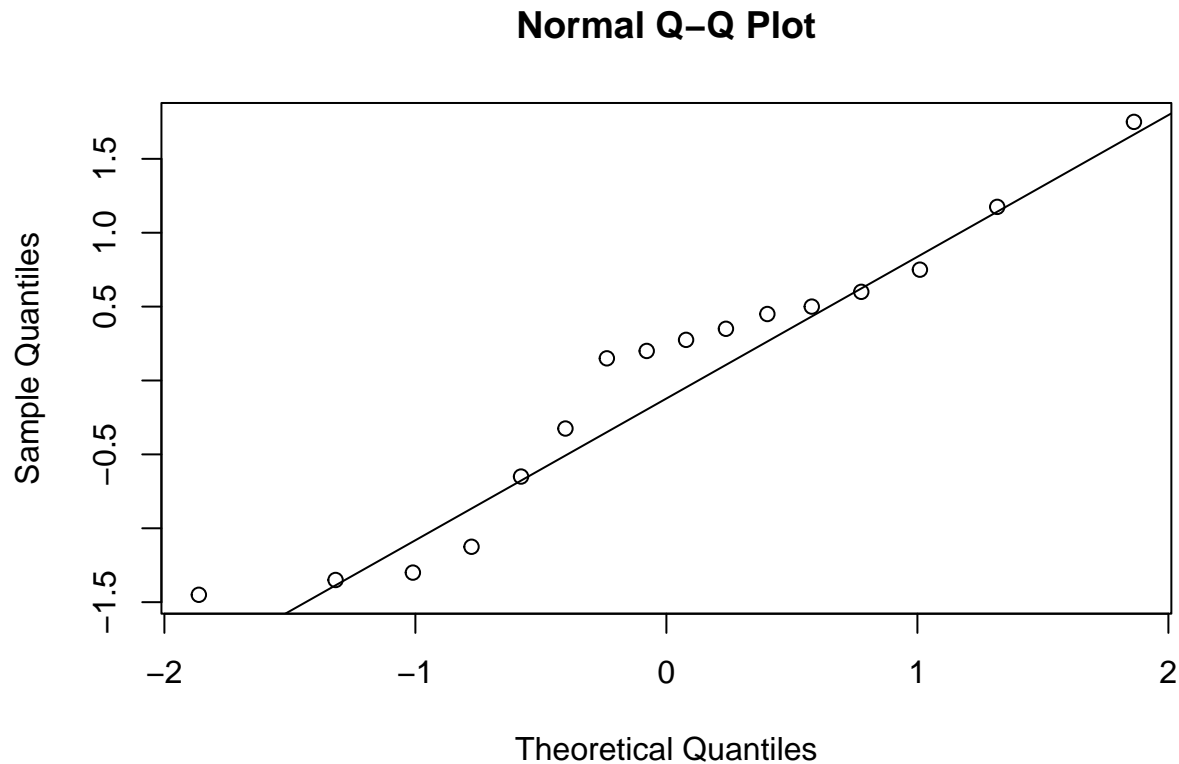


As shown in the residuals vs fitted values plot, the variance of the residuals is fairly consistent (*no significant deviation on the datapoints observed*) with minor deviations. However, with only 4 replicates per treatment, it is difficult to truly assess the homogeneity of variance. Ideally, we would like to increase the number of replicates, to obtain a more accurate assessment of the homogeneity of variance to see if the spread is consistent across treatment levels. Therefore, for assumption #2, if we **ONLY** based on only 4 replicate per treatment, the assumption of homogeneity of variance is justified.

Assumption 3: The probability distribution of the response  $y_{ij}$  or the experimental error  $\epsilon_{ij}$ , is Gaussian.

A (Quantile-Quantile) QQ plot can be used to check the whether the residuals are normally distributed (Gaussian). If the residuals are normally distributed, the tail endpoints in the QQ plot will fall along a straight line without significant deviation.

```
qqnorm(residuals)
qqline(residuals)
```



As shown in the QQ plot, the points at both tail ends falls approximately along the straight line without significant deviations (*Observe the 3 points when theoretical quantiles are less than -1 and the 3 points when theoretical quantiles more than 1*). This suggests that the residuals are normally distributed but problem is we do not have sufficient datapoint to really tell. Again, like assumption #2, the more datapoints, the more accurate our assessment of the normality of the residuals will be as 3 datapoints at each tail end is insufficient. Therefore, for assumption #3, if we **ONLY** based on observing the 3 datapoints per tail endpoints to check for significant deviation, the assumption of normality of the residuals is justified.

- 
10. If the dough were made in batches and the four replicate biscuit rise heights in each column (shown in the table above) were all from the same batch, would your answer to (a) be different? How could the data be analysed if this were the case?

---

Answer:

Changing the experiment to use the same dough batch for all replicates within each baking powder level would significantly alter the experimental design. This approach would eliminates batch-to-batch variation but it

introduces new complications as randomisation plays a pivotal role in experimental design by ensuring that treatment groups are comparable and that observed effects are attributable to the treatments themselves, rather than underlying differences between groups. By using the same batch for all replicates of a given treatment, we control for batch variation, but at the risk of other forms of variability. If the batches have inherent differences not related to baking powder levels—such as variations in ingredient quality or environmental conditions during preparation—these could confound the results. This makes it difficult to isolate the true effects of the baking powder levels. Therefore, while using the same batch for each treatment level controls for one type of variability, it may inadvertently introduce another, potentially impacting the experiment’s validity.

The assumption of independence of the errors  $\epsilon_{ij}$  would be violated. This is because the errors  $\epsilon_{ij}$  would not be independent of one another. In other words, the error associated from one experimental unit (biscuit) should not be influenced by error of any other biscuit for it to be considered independent. However, if the biscuits are made all from the same batches, that meant that all replicate within treatments come from the same batch and are thus manufactured under the same conditions (e.g. same temperature, same humidity, same manufacturing equipment used etc.). This would introduce the possibility of correlated error, as conditions within a batch could affect all the biscuits in a similar way.

In this case, the data should be analysed using a randomised complete block design where the blocks would be the different batches of dough and the treatments would be the levels of the baking powder. This would allow us to account for the batch-to-batch variation by ensuring that all treatments are represented within each batch.

---

## Part 2: Free fall of Papers of Different Dimension

In this part of the coursework you will design a factorial experiment, collect the data and then analyse it. You will carry out this experiment to identify how different design factors influence the free fall time of a piece of paper.

### Experimental Units

In this experiment each experimental unit will correspond to a single piece of paper. Replicates will mean you will have to use as many papers of different dimension as needed.

### Instructions

In this experiment you will need to drop many pieces of paper from a certain height. Please follow the following instructions:

- Use white printer paper (if needed you can use the university printer papers).
- Label your papers according to the factor variables (described below) and the replicate number.
- Perform the experiment in a large open space, mostly isolated from wind.
- Always allow the papers to fall approximately from the same height, straight towards the floor.
- You can do some practice before starting the main experiment, but I recommend to use different papers for this. The practice papers can be used more than once.
- Do not do all the experiment in one go. Take a few breaks to make sure your arm does not get tired and always use the same arm for the experiment.
- Recycle all the paper after finishing the experiment.

## Response

Time in seconds (time to be recorded to the nearest whole second) of the free fall of the paper. To measure the free fall time, follow the following instructions:

- Locate a position from where you would drop the piece of paper. You can mark that with your bag/other item. Stand there and drop the piece of paper from a certain height. Use your mobile/ a stopwatch to record the free fall time of the paper.
- Record the times in seconds (time to be recorded to the nearest whole second) in a spreadsheet making reference to each level of the factor.
- If anything goes wrong, for example if the paper touches your body before falling down, repeat that particular run with a different (pristine) paper.
- Repeat the experiment by dropping the pieces of paper from (approximately) the same height every time.

## Factor variables

- Factor 1 (length of the paper): (2 levels) 5 cm, 6 cm. You can measure the length using a ruler.
- Factor 2 (Width of the paper): (2 levels) 5.5 cm, 6.5 cm. You can measure the width using a ruler.

## Randomization

Remember to assign the factor level combinations to the experimental units (the piece of paper) completely at random. You can do this in R using the `sample` function. For example, for a balanced design with  $r = 3$  replicates, you can assign the level combinations to a total of 12 papers using the following command

```
f<-rep(c("5,5.5", "6,5.5", "5,6.5", "6,6.5"), each=3)
sample(f, 12)
```

```
## [1] "6,5.5" "5,5.5" "6,6.5" "6,5.5" "6,6.5"
## [6] "5,6.5" "6,5.5" "5,5.5" "5,5.5" "6,6.5"
## [11] "5,6.5" "5,6.5"
```

The experiment will be balanced in the sense that each factor level combination will have the same number of replicates  $r$  but you will have to determine the number of replicates. For this, consider the model with interactions:

$$y_{ijk} = \mu + \tau_i + \alpha_j + \gamma_{ij} + \epsilon_{ijk}$$

where  $\tau_{3.5}, \tau_{4.5}$  are the effects of the two different lengths of the papers,  $\alpha_3, \alpha_4$  are the effects of the two different widths of the papers,  $\gamma_{ij}$  are the effects of the interactions. All  $\epsilon_{ijk} \sim N(0, \sigma^2)$  and are mutually independent.

Before running the main experiment, you will need to perform a simple pilot experiment in order to estimate the variance parameter  $\sigma^2$ . You should perform  $n_{pilot} = 10$  runs with 10 different papers of length 5 cm and width 5.5 cm.

Specifically of length 5 cm and width 5.5 cm respectively. Denote by  $t_1, \dots, t_{10}$  the observed times in seconds. You should estimate the variance in the usual way, that is,

$$\hat{\sigma}^2 = \frac{1}{n_{pilot} - 1} \sum_{i=1}^{10} (t_i - \bar{t})^2$$

a) Report your 10 observations from the pilot experiment as well as your estimated value of  $\sigma^2$ .

```
# Pilot experiment
pilot_raw <- c(3.28, 3.52, 3.56, 3.54, 3.66, 3.62, 2.98, 3.10, 3.28, 2.98)

# Round to nearest whole second
pilot <- round(pilot_raw)

kable(
  data.frame(
    `Number of Observations` = 1:10,
    `Free Fall Time` = pilot_raw,
    `Free Fall Time Rounded to Nearest Second` = pilot
  ),
  align = "c"
)
```

Number.of.Observations	Free.Fall.Time	Free.Fall.Time.Rounded.to.Nearest.Second
1	3.28	3
2	3.52	4
3	3.56	4
4	3.54	4
5	3.66	4
6	3.62	4
7	2.98	3
8	3.10	3
9	3.28	3
10	2.98	3

```
# Estimating the variance of the experimental error using the given formula
variance_of_pilot <- sum((pilot - mean(pilot))^2) / (length(pilot) - 1)
cat("Variance of The Experimental Error: ", round(variance_of_pilot, 4))
```

```
## Variance of The Experimental Error: 0.2778
```

b) Denote  $\Delta_5 := \mu_{5,5.5} - \mu_{5,6.5}$  and  $\Delta_6 := \mu_{6,5.5} - \mu_{6,6.5}$  where  $\mu_{ij}$  is the mean response for length  $i \in \{5, 6\}$  and width  $j \in \{5.5, 6.5\}$ . State the null and the alternative hypotheses for a two-sided test of  $\Delta_5 = \Delta_6$ . Let  $\Delta = |\Delta_5 - \Delta_6|$ . We would like to detect an absolute difference  $\Delta = 1$  with high power. Determine the number of replicates  $r$  necessary to achieve at least 90% power using a significance level  $\alpha = 0.05$ .

```
# Reference: Design and Analysis of Experiments with R by John Lawson

library(daewr)
# Power calculation
# Minimum number of replicates needs to be at least 2 given that there
# are 2 levels present here
rmin <- 2
# We set an arbitrary value of 10 here for maximum number of replicate
# given that the pilot experiment was done with 10 papers
rmax <- 10
alpha <- 0.05
```

```

delta <- 1
nlev <- c(2,2)
nreps <- c(rmin:rmax)
power <- Fpower2(alpha, nlev, nreps, delta, sqrt(variance_of_pilot))
kable(power, align = "c")

```

alpha	a	b	nreps	Delta	sigma	powera	powerb
0.05	2	2	2	1	0.5270463	0.5302213	0.5302213
0.05	2	2	3	1	0.5270463	0.8199630	0.8199630
0.05	2	2	4	1	0.5270463	0.9353854	0.9353854
0.05	2	2	5	1	0.5270463	0.9781279	0.9781279
0.05	2	2	6	1	0.5270463	0.9929507	0.9929507
0.05	2	2	7	1	0.5270463	0.9978190	0.9978190
0.05	2	2	8	1	0.5270463	0.9993481	0.9993481
0.05	2	2	9	1	0.5270463	0.9998108	0.9998108
0.05	2	2	10	1	0.5270463	0.9999465	0.9999465

Answer:

The null hypothesis,  $H_0: \Delta_5 = \Delta_6$ .

The alternative hypothesis,  $H_a: \Delta_5 \neq \Delta_6$ .

*powera* and *powerb* exceed 0.9 or 90% when the number of replicates *nrep* is 4. We would like to detect an absolute difference  $\Delta = 1$  with high power, thus, the number of replicates *r* necessary to achieve at least 90% power using a significance level  $\alpha = 0.05$  is *nrep* = 4. This is the minimum number of replicates required to achieve at least 90% power.

What is happening here is that given the following parameters:

- variance of the experimental error  $\sigma^2 = 0.2778$
- standard deviation of the experimental error  $\sigma = \sqrt{0.2778}$
- delta,  $\Delta = 1$
- significance level,  $\alpha = 0.05$
- number of levels of the factor, *nlev* = 2 for each factor
- desired power, *Power*  $\geq 0.9$

We then iteratively calculate power for increasing number of replicates from 2 to 10 using the non-centrality parameter  $\lambda$ .

The non-centrality parameters for a two-factor ANOVA (factors: length and width) can be calculated as follows:

- For length:

$$\lambda_{\text{length}} = b \times nrep \times \sum_{i=1}^a \frac{(\mu_{\text{length},i} - \mu_{\text{overall}})^2}{\sigma^2}$$

- For width:

$$\lambda_{\text{width}} = a \times nrep \times \sum_{i=1}^b \frac{(\mu_{\text{width},i} - \mu_{\text{overall}})^2}{\sigma^2}$$

where *b* is the number of levels in width, *a* is the number of levels in length,  $\mu_{\text{width},i}$  is the mean for each level of width, and  $\mu_{\text{overall}}$  is the overall mean.

- Next, we use the non-centrality parameters to calculate the power using the non-central F-distribution using the *daepr* library.
- We then pick the minimum number of replicates required to achieve at least 90% power which is 4

---

```
# Generating the data
set.seed(37) # Set seed for reproducibility
f<-rep(c("5,5.5","6,5.5","5,6.5","6,6.5"),each=4)
seed_37_randomised_design <- sample(f,16)
seed_37_randomised_design
```

```
## [1] "6,5.5" "6,6.5" "6,5.5" "5,5.5" "5,5.5"
## [6] "6,6.5" "5,5.5" "5,6.5" "5,6.5" "5,6.5"
## [11] "6,5.5" "6,6.5" "5,6.5" "6,5.5" "5,5.5"
## [16] "6,6.5"
```

```
kable(
  data.frame(
    `Run Order` = 1:16,
    `Length` = c(6, 6, 6, 5, 5, 6, 5, 5, 5, 5, 6, 6, 5, 6, 5, 6),
    `Width` = c(5.5, 6.5, 5.5, 5.5, 5.5, 6.5, 5.5, 6.5, 6.5, 6.5, 5.5, 6.5, 6.5,
                5.5, 5.5, 6.5)
  ), align = "c"
)
```

Run.Order	Length	Width
1	6	5.5
2	6	6.5
3	6	5.5
4	5	5.5
5	5	5.5
6	6	6.5
7	5	5.5
8	5	6.5
9	5	6.5
10	5	6.5
11	6	5.5
12	6	6.5
13	5	6.5
14	6	5.5
15	5	5.5
16	6	6.5

```
# Inputting the collect data into a dataframe (Exclude the non-rounded observed values)
df <- data.frame(
  `Length` = c(6, 6, 6, 5, 5, 6, 5, 5, 5, 5, 6, 6, 5, 6, 5, 6),
  `Width` = c(5.5, 6.5, 5.5, 5.5, 5.5, 6.5, 5.5, 6.5, 6.5, 6.5, 5.5, 6.5, 6.5,
              5.5, 5.5, 6.5),
  `Fall_Time` = c(3, 3, 4, 4, 3, 3, 3, 3, 3, 3, 4, 3, 4, 4, 4, 4)
```

```

)

# Displaying the collected data in a table
kable(
  data.frame(
    `Number of Observations` = 1:16,
    `Length` = c(6, 6, 6, 5, 5, 6, 5, 5, 5, 5, 6, 6, 5, 6, 5, 6),
    `Width` = c(5.5, 6.5, 5.5, 5.5, 5.5, 6.5, 5.5, 6.5, 6.5, 6.5, 5.5, 6.5, 6.5,
                5.5, 5.5, 6.5),
    `Observed_Value` = c(3.35, 3.45, 3.57, 3.58, 3.45, 3.3, 3.2, 3.16, 2.85, 3.07,
                        3.4, 3.72, 3.2, 3.7, 3.68, 3.98),
    `Observed_Value_Rounded` = c(3, 3, 4, 4, 3, 3, 3, 3, 3, 3, 4, 3, 4, 4, 4, 4)
  ),
  align = "c"
)

```

Number.of.Observations	Length	Width	Observed_Value	Observed_Value_Rounded
1	6	5.5	3.35	3
2	6	6.5	3.45	3
3	6	5.5	3.57	4
4	5	5.5	3.58	4
5	5	5.5	3.45	3
6	6	6.5	3.30	3
7	5	5.5	3.20	3
8	5	6.5	3.16	3
9	5	6.5	2.85	3
10	5	6.5	3.07	3
11	6	5.5	3.40	3
12	6	6.5	3.72	4
13	5	6.5	3.20	3
14	6	5.5	3.70	4
15	5	5.5	3.68	4
16	6	6.5	3.98	4

c) Use ANOVA and your collected data to test if length of paper has a significant effect. Clearly state the hypotheses you are testing. Repeat the test but now consider width instead of length of paper.

---

Answer:

For a two-factor factorial:

The means model:

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

where:

- $i$  represents the level of the first factor
- $j$  represents the level of the second factor
- $k$  represents the replicate number



- $\mu_{ij}$  represents the expected response in the  $(i, j)$ th cell

and its corresponding effects model:

$$y_{ijk} = \mu + \tau_i + \alpha_j + \gamma_{ij} + \epsilon_{ijk}$$

where:

- $\tau_i$  (length of paper) and  $\alpha_j$  (width of paper) are the main treatment effects.
- $\gamma_{ij}$  is the interaction effect between the cell mean,  $\mu_{ij}$  and  $\mu + \tau_i + \alpha_j$ .
- $\epsilon_{ijk}$  is the experimental error such that  $E(\epsilon_{ijk}) = 0$  and  $\epsilon \sim MVN(0, \sigma^2 I)$ .

---

```
# Creating the design matrix X
# Define factor variables
length_ <- factor(df$Length)
width_ <- factor(df$Width)

# Creating the interaction model
inter_ <- interaction(df$Length, df$Width, sep = ",")

# Creating the design matrix
X1 <- model.matrix(~ length_ - 1, data = df)
X2 <- model.matrix(~ width_ - 1, data = df)
X3 <- model.matrix(~ inter_ - 1, data = df)

X <- cbind(1, X1, X2, X3)
colnames(X)[1] <- "Intercept"
kable(X)
```

Intercept	length_5	length_6	width_5.5	width_6.5	inter_5,5.5	inter_6,5.5	inter_5,6.5	inter_6,6.5
1	0	1	1	0	0	1	0	0
1	0	1	0	1	0	0	0	1
1	0	1	1	0	0	1	0	0
1	1	0	1	0	1	0	0	0
1	1	0	1	0	1	0	0	0
1	0	1	0	1	0	0	0	1
1	1	0	1	0	1	0	0	0
1	1	0	0	1	0	0	1	0
1	1	0	0	1	0	0	1	0
1	1	0	0	1	0	0	1	0
1	0	1	1	0	0	1	0	0
1	0	1	0	1	0	0	0	1
1	1	0	0	1	0	0	1	0
1	0	1	1	0	0	1	0	0
1	1	0	1	0	1	0	0	0
1	1	0	1	0	1	0	0	0
1	0	1	0	1	0	0	0	1

```
X0 <- cbind(1, X1, X2)
colnames(X0)[1] <- "Intercept"
kable(X0)
```

Intercept	length_5	length_6	width_5.5	width_6.5
1	0	1	1	0
1	0	1	0	1
1	0	1	1	0
1	1	0	1	0
1	1	0	1	0
1	0	1	0	1
1	1	0	1	0
1	1	0	0	1
1	1	0	0	1
1	1	0	0	1
1	0	1	1	0
1	0	1	0	1
1	1	0	0	1
1	0	1	1	0
1	1	0	1	0
1	0	1	0	1

To test for:

- null hypothesis,  $H_0: \gamma_{ij} = 0$  (no interaction effect)
- alternative hypothesis,  $H_a: \gamma_{ij} \neq 0$  (interaction effect)

```
library(MASS)
```

```
##
## Attaching package: 'MASS'

## The following objects are masked from 'package:daewr':
##
##      cement, chem
```

```
X0 <- cbind(1, X1, X2)
G <- ginv(t(X) %*% X)
G0 <- ginv(t(X0) %*% X0)

HX <- X %*% G %*% t(X)
HX0 <- X0 %*% G0 %*% t(X0)

y <- df$Fall_Time
n <- length(y)

r <- (diag(n) - HX) %*% y
r0 <- (diag(n) - HX0) %*% y
```

```

SSE <- t(r) %*% r
SSE0 <- t(r0) %*% r0
SSEInt <- SSE0 - SSE

kable(
  data.frame(
    `SSE` = SSE,
    `SSE0` = SSE0,
    `SSEInt` = SSEInt
  ),
  align = "c",
  caption = "Sum of Squares Error"
)

```

Table 11: Sum of Squares Error

SSE	SSE0	SSEInt
3	3.25	0.25

After finding the *SSE* errors, we can compute the degree of freedom associated with *SSInt* and *SSE* and then compute the F-test statistics as follows:

$$F = \frac{SSInt/df_{SSInt}}{SSE/df_{SSE}}$$

where  $df_{INT}$  and  $df_{SSE}$  are the degrees of freedom associated with *SSInt* and *SSE* respectively.

```

# Degrees of freedom
df_SSE <- qr(diag(n)-HX)$rank
df_SSInt <- qr(HX-HX0)$rank
kable(
  data.frame(
    `df_SSE` = df_SSE,
    `df_SSInt` = df_SSInt
  ),
  align = "c",
  caption = "Degrees of Freedom"
)

```

Table 12: Degrees of Freedom

df_SSE	df_SSInt
12	1

```

# F-test statistics
MSInt <- SSEInt / df_SSInt
MSE <- SSE / df_SSE
Fval <- MSInt / MSE
cat("F-test statistics: ", Fval)

```

```
## F-test statistics: 1
```

---

Next, we compute the p-value associated using the fact that, under the null hypothesis:

$$F = \frac{SSInt/df_{SSInt}}{SSE/df_{SSE}} \sim F_{(df_{SSInt}, df_{SSE}, 0)}$$

a central F distribution with  $df_{SSInt}$  in the numerator and  $df_{SSE}$  in the denominator.

```
# P-value
cat("P-value: ", round(pf(Fval, df_SSInt, df_SSE, lower.tail = FALSE), 3))
```

```
## P-value: 0.337
```

---

The above method the method of computing the p-value is tedious. Alternatively, we can use the built-in R function to compute the p-value as follows:

---

```
# Alternative method, a faster method suggested by ChatGPT
df_alt <- data.frame(y, length_, width_)
model <- aov(y ~ length_ * width_, data = df_alt)
test_results <- anova(model)

# Convert ANOVA table to a dataframe
df_anova <- as.data.frame(test_results)

rownames(df_anova) <- c("Length", "Width", "Interaction between Length & Width", "Residuals")
colnames(df_anova) <- c("Degrees of Freedom", "Sum of Squares", "Mean Square", "F-Value", "Pr(>F)")
df_anova["Pr(>F)"] <- round(df_anova["Pr(>F)"], 3) # Round the p-value to 3 decimal places
kable(df_anova, align = "c")
```

---

	Degrees of Freedom	Sum of Squares	Mean Square	F-Value	Pr(>F)
Length	1	0.25	0.25	1	0.337
Width	1	0.25	0.25	1	0.337
Interaction between Length & Width	1	0.25	0.25	1	0.337
Residuals	12	3.00	0.25	NA	NA

---

Answer:

The hypothesis test for the length of paper is as follows:

The null hypothesis,  $H_0$ :  $\tau_i = 0$ . The alternative hypothesis,  $H_a$ :  $\tau_i \neq 0$ .

The hypothesis test for the width of paper is as follows:

The null hypothesis,  $H_0: \alpha_j = 0$ . The alternative hypothesis,  $H_a: \alpha_j \neq 0$ .

where  $i$  and  $j$  are the levels of the length and width of the paper respectively.

The p-value for the standalone effect of the length of the paper  $\tau_i$  is 0.337 which is greater than the significance level of 0.05. Therefore, we fail to reject the null hypothesis and conclude that the length of the paper does not have a significant effect on the free fall time of the paper.

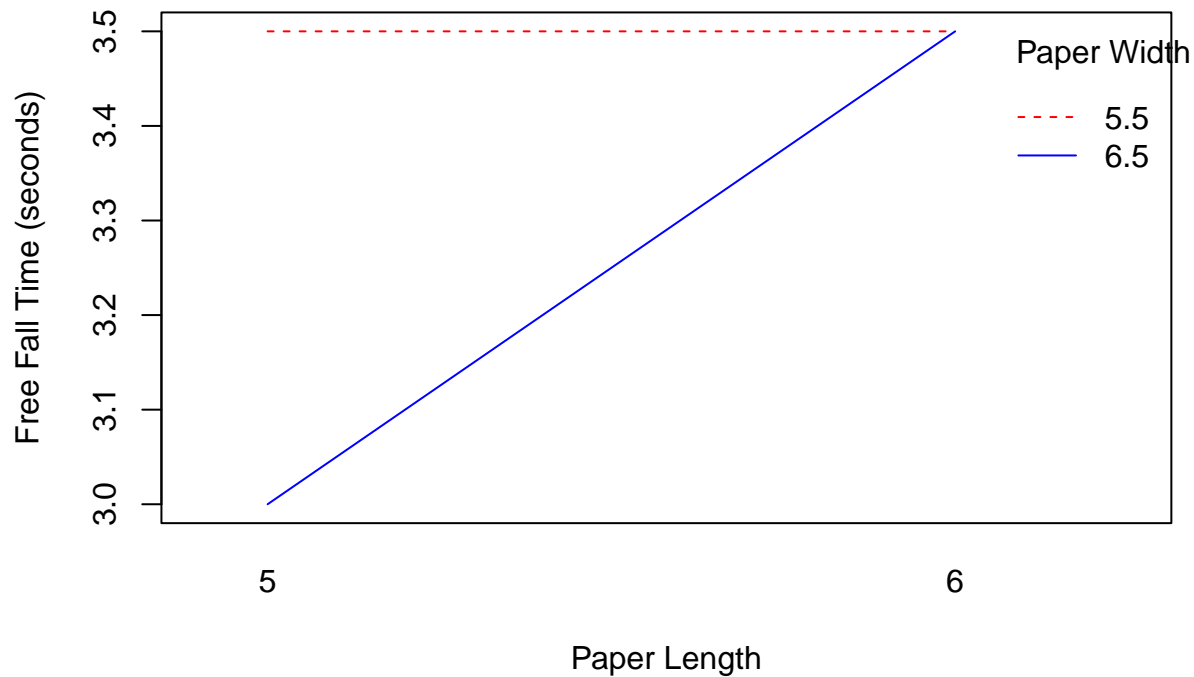
The p-value for the standalone effect of the width of the paper  $\alpha_j$  is 0.337 which is greater than the significance level of 0.05. Therefore, we fail to reject the null hypothesis and conclude that the width of the paper does not have a significant effect on the free fall time of the paper.

---

- d) Use the previous ANOVA table to comment on if the interaction effect is significant or not. Clearly state the hypotheses you are testing to check if the interaction effect is significant or not. Also, comment on the same using an appropriate plot. Mention any lurking variables one needs to be aware of.

```
# Interaction plot
interaction.plot(
  df$Length,
  df$Width,
  df$Fall_Time,
  xlab = "Paper Length",
  ylab = "Free Fall Time (seconds)",
  trace.label = "Paper Width",
  legend = TRUE,
  col = c("red", "blue"),
  lwd = 1,
  main = "Interaction Plot of Length and Width of Paper"
)
```

### Interaction Plot of Length and Width of Paper



---

Answer:

The hypothesis test for the interaction effect is as follows:

The null hypothesis,  $H_0: \gamma_{ij} = 0$ .

The alternative hypothesis,  $H_a: \gamma_{ij} \neq 0$

where  $i$  and  $j$  are the levels of the length and width of the paper respectively.

The p-value for the interaction effect  $\gamma_{ij}$  is 0.337 which is greater than the significance level of 0.05. Therefore, we fail to reject the null hypothesis and conclude that the interaction effect is not significant.

Based on the interaction plot, the lines do not cross and the slope change is evident in one of the width levels but based on our ANOVA test, we fail to reject the null hypothesis and conclude that the interaction effect is not significant. Not to mention, given that the increase in free fall time from a paper length of 5cm to 6cm is very minimal, it might be possible that the slight gradient can be considered as near parallel to the other line (Generally, a pair of perfect parallel lines means that there is no interaction effect between the two).

However, it might be possible that by increasing the number of levels within each factor, we may improve our ability to detect a significant interaction effect. This could be particularly important if there is a non-linear relationship between the factors and the response variable that isn't captured with fewer levels. This is something to be aware of.

In terms of lurking variables, we consider factors such as:

- Reaction time in starting and stopping the stopwatch that makes this an imperfect measurement

- The height from which the paper is dropped (too low will not give enough time to measure, too high increases the risk of wind interference)
- The environmental conditions such as wind, temperature, and humidity (despite being in a empty room with little airflow, it is hard to achieve a room with perfect measurement throughout the experiment)
- Condition of the paper such as its weight, thickness, and texture (variability in manufacturing)

These are all factors that could potentially influence the free fall time of the paper.

Also, we note that the increment from 5cm to 6cm in length and 5.5cm to 6.5cm in width is quite small. It might be possible that the effect of the length and width of the paper on the free fall time is not significant due to the small increment in the length and width of the paper does not sufficiently allows us to conclude that the length and width of the paper has a significant effect on the free fall time. For future study, it would be better to use a larger increment in the length and width of the paper to see if the effect is significant as well as increasing the amount of factor levels to see if the interaction effect becomes significant over a healthy span of length and width selections.

---