MA50259: Statistical Design of Investigations

Coursework 1 (2024)

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 coursework 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)
```

```
# Delete this later
library(ggplot2)
ggplot(biscuits,aes(x=type,y=riseht))+geom_point(alpha=0.5)+xlab("Baking Powder (tsp)")+ylab("Rise Heig
```
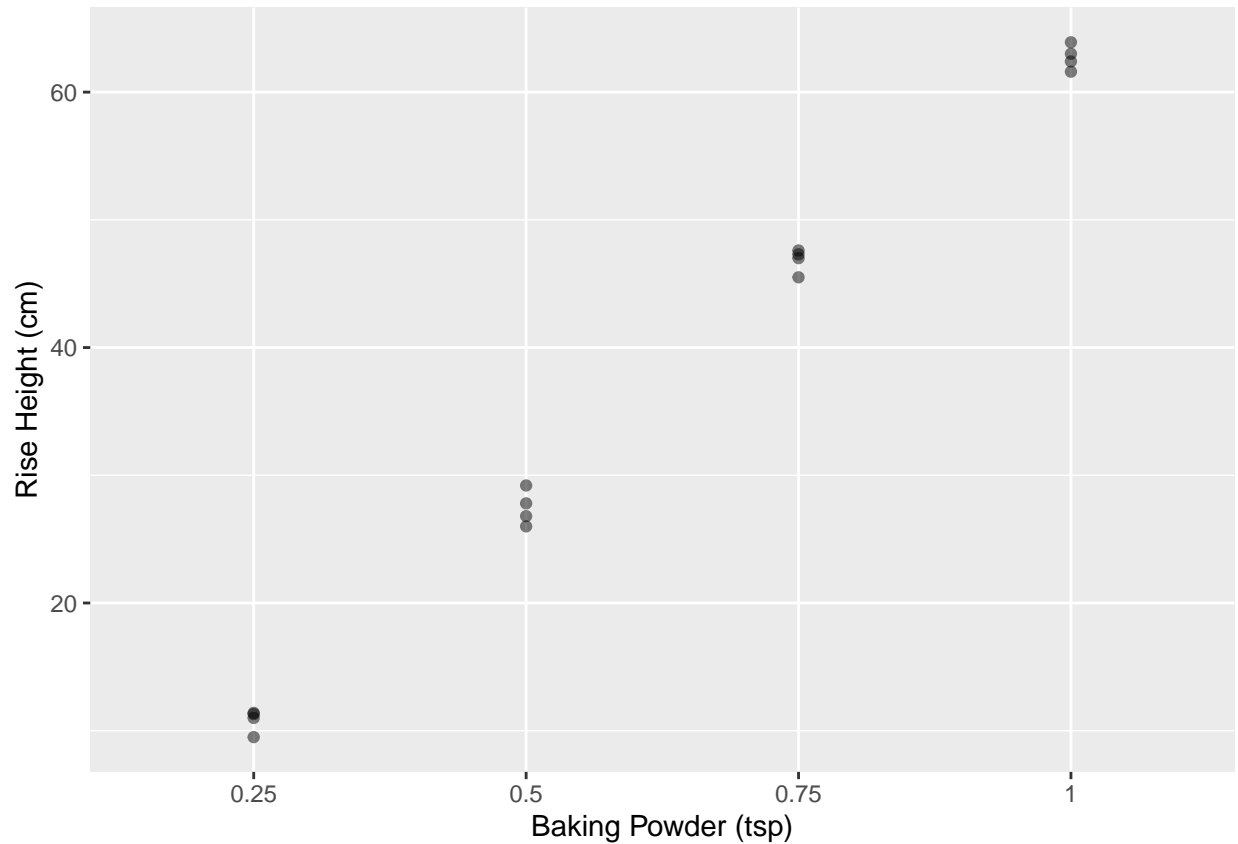

3. Construct in R, the design matrix X corresponding to the model:

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

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.

```
# Z <- model.matrix(~type-1,data=biscuits)
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
colnames(X) <- c("reference", "0.25 tsp", "0.50 tsp", "0.75 tsp", "1.00 tsp") # naming the columns
X
```

```
##   reference 0.25 tsp 0.50 tsp 0.75 tsp 1.00 tsp
## 1         1        1        0        0        0
## 2         1        1        0        0        0
```

```
## 3              1        1        0        0        0
## 4              1        1        0        0        0
## 5              1        0        1        0        0
## 6              1        0        1        0        0
## 7              1        0        1        0        0
## 8              1        0        1        0        0
## 9              1        0        0        1        0
## 10             1        0        0        1        0
## 11             1        0        0        1        0
## 12             1        0        0        1        0
## 13             1        0        0        0        1
## 14             1        0        0        0        1
## 15             1        0        0        0        1
## 16             1        0        0        0        1
```

4. How many unknown parameters are in the model?

<u>Answer:</u>

5 unknown parameters:

   i. Reference rise height of the biscuit when there is no treatment μ

  ii. Treatment effects τ_i for $i = 1, 2, 3, 4$

---

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

```r
# Figure out how does QR decomposition allows us to find the rank of a matrix
qr(X)$rank
```

```
## [1] 4
```

<u>Answer:</u>

The rank of the design matrix $\boldsymbol{X}$ is 4. This is because the design matrix $\boldsymbol{X}$ 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 $\boldsymbol{X}$ 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 $\boldsymbol{X}$ is 4 as shown in the QR decomposition of $\boldsymbol{X}$ above.

```r
# Delete this later
anova(lm(riseht~type,data=biscuits))
```

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

```
## type         3 6145.7 2048.58  1822.6 3.231e-16 ***
## Residuals 12   13.5     1.12
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Alternative
aov(riseht ~ type, data = biscuits)
```

```
## Call:
##    aov(formula = riseht ~ type, data = biscuits)
##
## Terms:
##                    type Residuals
## Sum of Squares  6145.732    13.487
## Deg. of Freedom        3        12
##
## Residual standard error: 1.060169
## Estimated effects may be unbalanced
```

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`.

```r
# Compute the ANOVA table
# Am I suppose to use the reduced model here? SSE_0 = y^T * A * y (Refer to week 3 lecture 1)
# Another possibility is that since the non full rank model is the treatment effect model. No treatment

# ChatGPT
# The ANOVA table is computed using the following steps:
# Calculate grand mean
grand_mean <- mean(biscuits$riseht)

# Calculate the total sum of squares (SSTotal)
SSTotal <- sum((biscuits$riseht - grand_mean)^2)

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

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

# 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)

# Create the ANOVA table
ANOVA_table <- 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(MSTreatment,MSError,""),
                          `F-statistic`=c(F_stat,"",""),
                          `p-value`=c(p_value,"",""))

ANOVA_table
```

```
##        Source Sum.of.Squares Degrees.of.Freedom
## 1 Treatment      6145.7319                    3
## 2     Error        13.4875                   12
## 3     Total      6159.2194                   15
##          Mean.Square      F.statistic
## 1 2048.57729166667 1822.64522706281
## 2 1.12395833333289
## 3
##                  p.value
## 1 3.33066907387547e-16
## 2
## 3
```

Answer:

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 τs are not equal. Given that we obtained a p-value of 3.33e-16, we reject the null hypothesis and conclude that there is a treatment effect.

```
# Using SSE under the null model (and I suspect the full rank model)
y <- biscuits$riseht
# SSE_0 <- t(y) %*% (diag(n) - (1/n) * rep(1,n) %*% t(rep(1,n))) %*% y

# alternative formula from pg 26 of the book
# (t(matrix(y, ncol = 1)) %*% y) - (t(rep(1,n)) %*% y ) ^ 2 / (t(rep(1,n)) %*% rep(1,n))

# More rigour to avoid the auto-convert from %*%
SSE_0 <- t(matrix(y, ncol = 1)) %*% (diag(n) - (1/n) * rep(1,n) %*% t(rep(1,n))) %*% matrix(y, ncol = 1)
SSE_0
```

```
##          [,1]
## [1,] 6159.219
```

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/



# 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(-3, -1, 1, 3)

# Calculate the contrast estimate
# Reusing the 'means' variable obtained when calculating the SSTreatment for ANOVA table
contrast_estimate <- linear_contrast_c %*% means

# 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)

# List the contrast estimate, standard error, t-statistic, and p-value
contrast_results <- 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))
kable(contrast_results, 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.

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:

$$c = (-3, -1, 1, 3)$$

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.

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 τ_2 and τ_3 to serve as a pivot point for the linear trend, while multiplying τ_1 by -3 allows us to "pull" the contrast value down, and multiplying τ_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 τs differs.

The contrast estimates to 175.18 using the formula:

$c^T \hat{\tau}_i$ = contrast weight × 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:

standard error (contrast) $= \sqrt{MSE \times \frac{\sum_{i=1}^{t} c_i^2}{t}} = \sqrt{1.124 \times \frac{(-3)^2 + (-1)^2 + (1)^2 + (3)^2}{4}} = 2.37$

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 2.51×10^{-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 σ^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)

# Formulate the ginverse for non-full rank treatment model
ginverse <- rbind(0, cbind(0, solve(t(Z) %*% Z)))
```

```
SSE_test2 <- t(y) %*% (diag(n) - X %*% ginverse %*% t(X)) %*% y
variance_of_experimental_error <- round(SSE_test2 / (n - t), 4)
variance_of_experimental_error
```

```
##       [,1]
## [1,] 1.124
```

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

<u>Answer:</u>

The assumptions for the considered linear model are:

    i. The errors ϵ_{ij} are independent and identically distributed $N(0, \sigma^2)$.

    ii. The errors ϵ_{ij} are independent of the treatment effects τ_i.

    iii. The errors ϵ_{ij} are independent of one another.

    iv. The errors ϵ_{ij} are normally distributed.

    v. The errors ϵ_{ij} have constant variance σ^2.

The first four assumptions are justified because the experimental design is a completely randomised design (CRD) and the errors ϵ_{ij} are assumed to be independent and identically distributed $N(0, \sigma^2)$. The fifth assumption is justified because the variance of the experimental error σ^2 is constant as shown in the ANOVA table above.

---

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 analyzed if this were the case?

<u>Answer:</u>

If the dough were made in batches and the four replicate biscuit rise heights in each column were all from the same batch, then the assumption of independence of the errors ϵ_{ij} would be violated. This is because the errors ϵ_{ij} would not be independent of one another. In this case, the data could be analysed using a randomised complete block design (RCBD) where the blocks would be the batches of dough and the treatments would be the levels of the baking powder. The ANOVA table would be computed using the same method as in question 6 but the design matrix $\boldsymbol{X}$ would be different.

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" "6,5.5" "5,5.5" "6,6.5" "5,5.5"
##  [6] "6,6.5" "6,5.5" "6,6.5" "5,6.5" "5,5.5"
## [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, α_3, α_4 are the effects of the two different widths of the papers, γ_{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 σ^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 $t_1, \ldots, t_{10}$ the observed times in seconds. You should estimate the variance in the usual way, that is,

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

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

```
# Pilot experiment # SEED 1234
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)

# Estimate the variance of the experimental error
variance_of_experimental_error_pilot <- sum((pilot - mean(pilot))^2) / (length(pilot) - 1)
variance_of_experimental_error_pilot
```

```
## [1] 0.2777778
```

b) Denote $\Delta_5 := \mu_{5,5.5} - \mu_{5,6.5}$ and $\Delta_6 := \mu_{6,5.5} - \mu_{6,6.5}$ where μ_{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 $\alpha = 0.05$.

```
# Power calculation
# The power of a test is the probability of rejecting the null hypothesis when the alternative hypothes
# The power of a test is calculated using the formula:
# power = 1 - pnorm(qnorm(1 - alpha/2) - abs(delta), 0, sqrt(sigma^2/n))
# where alpha is the significance level, delta is the absolute difference, sigma^2 is the variance of t

# Calculate the number of replicates necessary to achieve at least 90% power
```

10

```
alpha <- 0.05
power <- 0.9
delta <- 1
sigma <- variance_of_experimental_error_pilot
n <- 10
r <- ceiling((qnorm(1 - alpha/2) - qnorm(power)) ^ 2 * sigma / delta ^ 2)
r
```

[1] 1

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.

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.

```
# Assuming you have the knitr package installed

# Creating the dataframe
data <- 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),
  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
  Observed_Value_Rounded = c(3, 3, 4, 4, 3, 3, 3, 3, 3, 3, 3, 4, 3, 4, 4, 4)
)

# Converting the dataframe to a Markdown table
kable(data)
```

Length	Width	Observed_Value	Observed_Value_Rounded
6	5.5	3.35	3
6	6.5	3.45	3
6	5.5	3.57	4
5	5.5	3.58	4
5	5.5	3.45	3
6	6.5	3.30	3
5	5.5	3.20	3
5	6.5	3.16	3
5	6.5	2.85	3
5	6.5	3.07	3
6	5.5	3.40	3
6	6.5	3.72	4
5	6.5	3.20	3
6	5.5	3.70	4
5	5.5	3.68	4
6	6.5	3.98	4