LAB5

2025-05-01

Lab 5: Part 1

Paper Airplanes RCBD

Method:

This lab is to determine whether the placement of a paperclip on a paper airplane affects how far the airplane flies. We tested three situations: putting a paperclip on the nose, on the rear, and not using a paperclip at all. Every student threw the same airplane under all three conditions.

We did the experiment in an indoor hallway to avoid wind and outside changes. Each student threw the plane from the same spot, trying to throw it as straight as possible. Other students measured how far the plane flew from the starting point to where it first landed.

To keep the test fair, we used the sample() function in R to randomize the order of the throws for each student. This helped prevent learning or getting tired from affecting the results.

The technical issues happened during the experiment. Sometimes the plane flew way off course or hit someone, so we had to repeat those throws. Also, not every throw was perfect, and small differences in how the paperclip was placed may have caused more variation in the data.

Airplane Data Collection and Constructing the Experimental Data Frame

```r
#in the order of nose, rear, no_p
person_1 <- c(215, 162, 235) #Distances for person 1 in cm
person_2 <- c(180, 145, 242) #Distances for person 2 in cm
person_3 <- c(195, 60, 84) #Distances for person 3 in cm
person_4 <- c(200, 102, 198) #Distances for person 4 in cm
person_5 <- c(152, 91, 170) #Distances for person 5 in cm

# Create data frame
airplane_df <- data.frame(
  score = c(person_1, person_2, person_3, person_4, person_5),
  treatment = rep(c("nose", "rear", "no_p"), 5), #5 people
  block = c(rep("person_1", 3), #3 trials
            rep("person_2",3),
            rep("person_3", 3),
            rep("person_4", 3),
            rep("person_5", 3))
)
sample(airplane_df) #Print out the sample of the data.frame
```

```
##       block treatment score
## 1  person_1      nose   215
## 2  person_1      rear   162
## 3  person_1      no_p   235
## 4  person_2      nose   180
## 5  person_2      rear   145
## 6  person_2      no_p   242
## 7  person_3      nose   195
## 8  person_3      rear    60
## 9  person_3      no_p    84
## 10 person_4      nose   200
## 11 person_4      rear   102
## 12 person_4      no_p   198
## 13 person_5      nose   152
## 14 person_5      rear    91
## 15 person_5      no_p   170
```

Data Visualization

```r
library(ggplot2)
theme_update(text = element_text(size = 20))

ggplot(data = airplane_df,
       aes(y= score, x = treatment, group = block, color = block)) +
  geom_point() +
  geom_line()
```
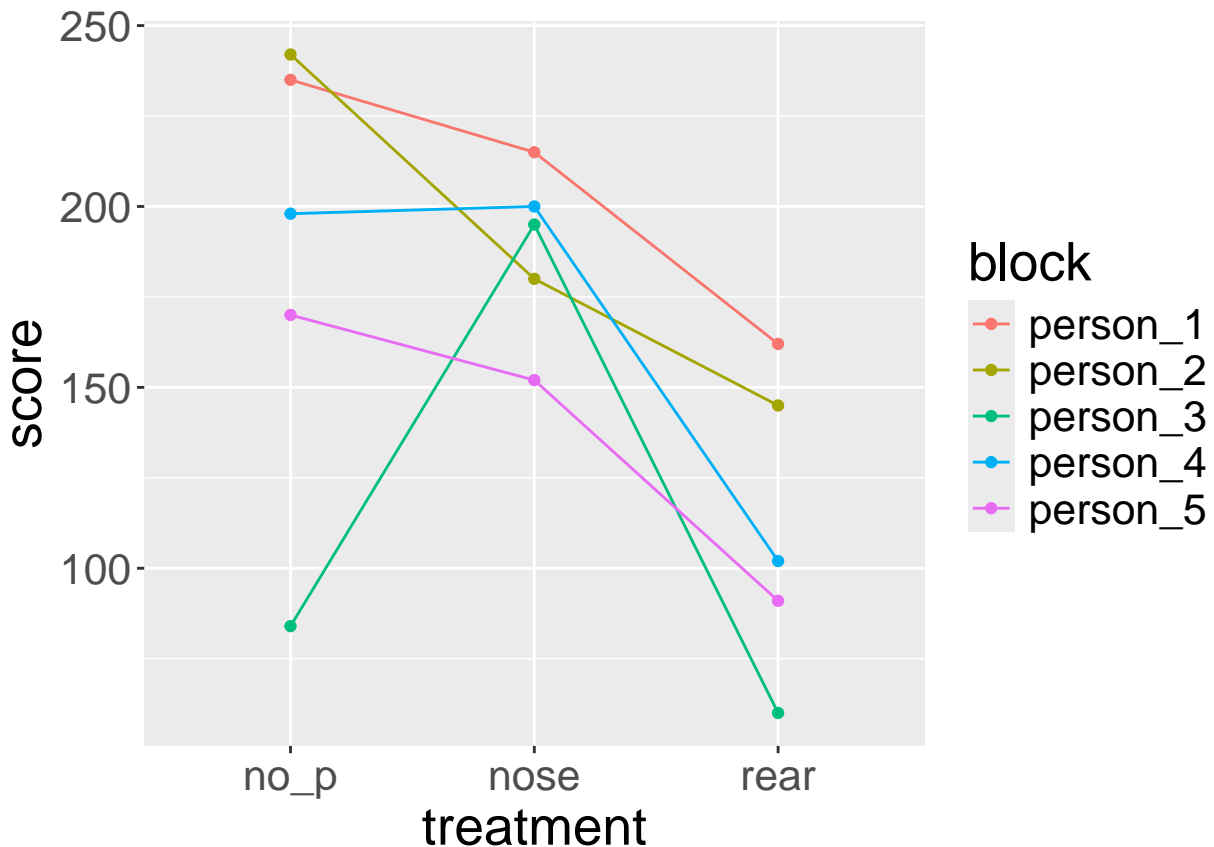
Description for the ggplot:

This is the ggplot shows that most individuals achieved their longest flight distances with no paperclip or when the paperclip was on the nose, while placing the paperclip on the rear generally resulted in shorter flights for all students.

Report the null and alternative hypotheses Treatment Groups and Block

To find out whether the paperclip position (treatment) or the person throwing the airplane (block) has a statistically significant effect on flight distance, I performed a two-way ANOVA. For both factors, I clearly stated the null and alternative hypotheses, and tested them using a significance level of $\alpha = 0.05$.

```
model_treatment_block <- aov(score ~ treatment + block, data = airplane_df)
#model1 <- aov(score ~ treatment, data = airplane_df) need the block
summary(model_treatment_block)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## treatment     2  18817    9408   8.426 0.0107 *
## block         4  16524    4131   3.700 0.0545 .
## Residuals     8   8932    1117
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Report the null and alternative hypotheses for Treatment Groups (Paperclip Positions)

$$H_0 : \mu_{\text{no p}} = \mu_{\text{nose}} = \mu_{\text{rear}}$$

$$H_0 : \text{The mean flight distance is the same for all paperclip positions.}$$

$$H_a : \text{The mean flight distance is different for at least one paperclip position.}$$

$$H_a : \text{At least one position has a different mean.}$$

F value: 8.426 p-value: 0.0107

Report the null and alternative hypotheses for Block (Student Throwing)

$$H_0 : \mu_{\text{person 1}} = \mu_{\text{person 2}} = \mu_{\text{person 3}} = \mu_{\text{person 4}} = \mu_{\text{person 5}}$$

$$H_0 : \text{The mean flight distance is the same for all individual students.}$$

$$H_a : \text{The mean flight distance is different for at least one student.}$$

$$H_a : \text{At least one student has a different mean.}$$

F value: 3.700 p-value: 0.0545

```
library(knitr)
anova_table <- as.data.frame(summary(model_treatment_block)[[1]])
anova_table$Factor <- rownames(anova_table)
anova_table <- anova_table[, c( "Df", "Sum Sq", "Mean Sq", "F value", "Pr(>F)")]
kable(anova_table, caption = "ANOVA Table for Paper Airplane Data")
```

Table 1: ANOVA Table for Paper Airplane Data

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
treatment	2	18816.93	9408.467	8.426373	0.0107365
block	4	16523.60	4130.900	3.699700	0.0545233
Residuals	8	8932.40	1116.550	NA	NA

Result:

I used a Randomized Complete Block Design (RCBD) to analyze the data. In this design, each student is a block, and each paperclip position is a treatment.

For the paperclip position, the ANOVA table showed an F value of 8.4264 and a p-value of 0.0107. Since the p-value is smaller than 0.05, we reject the null hypothesis. This means that the paperclip position has a statistically significant effect on how far the airplane flies.

For the student (block) effect, the F value was 3.6997 with a p-value of 0.0545. Since this p-value is greater than 0.05, we do not reject the null hypothesis. This means that the difference between students is not statistically significant.

Discussion:

This experiment tested if the paperclip's position changes how far a paper airplane flies. We found that the position matters, but who throws the plane doesn't make a big difference.

Flights were usually shortest with the rear paperclip, while the nose and no paperclip conditions flew farther. Randomizing the order helped reduce bias, but small errors in throwing or setup might have affected results. Overall, rear paperclip placement seems to reduce flight distance, and future studies with more trials could give stronger results.

Lab 5: Part 2

RCBD from scratch

2a

```
# read the xlsx
library(readxl)
vasc <- read_excel("VascGraft.xlsx")
```

Report the null and alternative hypotheses for PSI and Block

To determine whether either the PSI factor or the block factor has statistically significant on the outcome variable yield by using a two-way ANOVA. Below the null and alternative hypotheses for each factor, using a significance level of $\alpha = 0.05$.

```
#yield is the response (dependent) variable
#PSI is the factor (grouping variable, like treatment)
# This code is a two-way ANOVA model to the data in vasc which is testing whether PSI and/or block have
model_yield_PSI <- aov(yield ~ PSI + block, data = vasc)
summary(model_yield_PSI)
```

```
##              Df Sum Sq Mean Sq F value  Pr(>F)
## PSI           1 171.60  171.60  12.609 0.00189 **
## block         1  22.91   22.91   1.684 0.20850
## Residuals    21 285.79   13.61
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Report the null and alternative hypotheses for PSI.

$$H_0 : \text{The mean yield is the same for all PSI levels.}$$

$$H_0 : \mu_{\text{PSI level 1}} = \mu_{\text{PSI level 2}} = \ldots = \mu_{\text{PSI level k}}$$

$$H_a : \text{At least one PSI level has a different mean yield.}$$

$$H_a : \text{At least one PSI group mean is different}$$

5

Report both statistical conclusion (reject / fail to reject H0)

F value: 12.609 p-value: 0.00189

Report the null and alternative hypotheses for Block

$$H_0 : \text{The mean yield is the same for all blocks.}$$

$$H_0 : \mu_{\text{block 1}} = \mu_{\text{block 2}} = \ldots = \mu_{\text{block k}}$$

$$H_a : \text{At least one block has a different mean yield.}$$

$$H_a : \text{At least one block group mean is different}$$

Report both statistical conclusion (reject / fail to reject H0)

F value: 1.684 p-value: 0.20850

ANOVA Table for VascGraft Data (From Scratch Calculation)

Explainsion of MS_T from Chatgpt

1. Sum of Squares for Treatment (SS_T)

$$SS_T = b \sum_{i=1}^{a} (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$$

2. Mean Square for Treatment (MS_T)

$$MS_T = \frac{SS_T}{a - 1}$$

3. Sum of Squares for Blocks (SS_B)

$$SS_B = a \sum_{j=1}^{b} (\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot})^2$$

4. Sum of Squares for Error (SS_E)

$$SS_E = \text{SSEA}_{\text{(total within-group)}} - SS_B$$

5. Mean Square for Error (MS_E)

$$MS_E = \frac{SS_E}{(a - 1)(b - 1)}$$

6. F-Statistic for Treatment

$$F = \frac{MS_T}{MS_E}$$

Why We Leave Block F and p-values as NA

In RCBD, blocks are used to reduce variability, not to test for significance. Since our goal is to determine whether PSI levels treatments affect yield, we do not compute F-statistics or p-values for blocks. These values are left as `NA`, following standard practice.

```r
library(dplyr)
```

```
## 
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
## 
##     filter, lag

## The following objects are masked from 'package:base':
## 
##     intersect, setdiff, setequal, union
```

```r
# Initialize the Output Matrix
output <- matrix(NA, nrow = 3, ncol = 5)

#Calculate Treatment Sum of Squares (SST)
group_means <- vasc %>%
  group_by(PSI) %>%
  summarize(mean = mean(yield))

#Degrees of Freedom and Mean Square for Treatment
output[1,1] <- 3
output[1,2] <- sum(6*(group_means$mean - mean(vasc$yield))^2)
output[1,3] <- output[1,2] / output[1,1]

#Calculate Block Sum of Squares (SSB)
block_means <- vasc %>%
  group_by(block) %>%
  summarize(mean = mean(yield))

#Degrees of Freedom and Mean Square for Block
output[2,1] <- 5
output[2,2] <- sum(4*(block_means$mean - mean(vasc$yield))^2)
output[2,3] <- output[2,2] / output[2,1]

#Calculate Residual (Error) Sum of Squares (SSE)
group_var <- vasc %>%
  group_by(PSI) %>%
  summarize(n = n(), var = var(yield))
SSEA <- sum((group_var$n-1)*group_var$var)
output[3,2] <- SSEA - output[2,2] #SSE_block = SSE_all - SS_teat

#Degrees of Freedom and Mean Square for Residual
output[3,1] <- 15
output[3,3] <- output[3,2] / output[3,1]
```

```
#F-statistic and p-value for Treatment
output[1,4] <- output[1,3] / output[3,3]
output[1,5] <- pf(output[1,4], df1 = 3, df2 = 15, lower.tail = FALSE)

#Leave Block F and p-value as NA
output[2,4] <- NA
output[2,5] <- NA
output[3,4] <- NA
output[3,5] <- NA
colnames(output) <- c("DF", "Sum Sq", "Mean Sq", "F value", "p-value")
rownames(output) <- c("treatment", "block", "residual")
kable(output, caption = "ANOVA Table for VascGraft Data (Calculated from Scratch)")
```

Table 2: ANOVA Table for VascGraft Data (Calculated from Scratch)

	DF	Sum Sq	Mean Sq	F value	p-value
treatment	3	178.1712	59.39042	8.107077	0.0019163
block	5	192.2521	38.45042	NA	NA
residual	15	109.8863	7.32575	NA	NA

Interpretation:

The ANOVA table for the VascGraft data shows that the treatment (PSI level) has a statistically significant effect on yield, with an F-value of 8.11 and a p-value of 0.00192. Since the p-value is less than 0.05, we reject the null hypothesis and conclude that the yield does vary between different PSI levels.

The block effect was included to account for differences between groups, but we're not actually trying to test whether the blocks themselves are different. Our main interest is in the effect of PSI. That's why the F-value and p-value for the block were not calculated (marked as NA) which is normal in this kind of analysis when blocks are just used for control, not for testing.

2b

ANOVA Table for Paper Airplane RCBD Data (From Scratch Caluculation)

```
library(dplyr)
library(knitr)
output <- matrix(NA, nrow = 4, ncol = 5)
a <- 3
b <- 5
#Treatment (paperclip position) effect
group_means <- airplane_df %>%
  group_by(treatment) %>%
  summarize(mean = mean(score))
output[1,1] <- a-1 # treatment: 3 treatments - 1
output[1,2] <- sum(5 * (group_means$mean - mean(airplane_df$score))^2) # SS_treatment: 5 blocks
output[1,3] <- output[1,2] / output[1,1]
# MS_treatment
```

```r
# Block (student) effect
block_means <- airplane_df %>%
  group_by(block) %>%
  summarize(mean = mean(score))

output[2,1] <- b-1 # block: 5 blocks - 1
output[2,2] <- sum(3 * (block_means$mean - mean(airplane_df$score))^2) # SS_block: 3 treatments
output[2,3] <- output[2,2] / output[2,1]
# MS_block

#Error (residual)
group_var <- airplane_df %>%
  group_by(treatment) %>%
  summarize(n = n(), var = var(score))

SSEA <- sum((group_var$n-1)*group_var$var) #SSE_all

output[3,1] <- (a-1) * (b-1) # Error: (3-1) * (5-1) = 8
output[3,2] <- SSEA - output[2,2] #SSR
output[3,3] <- output[3,2] / output[3,1] # MS_residual

#F and p value for treatment
output[1,4] <- output[1,3] / output[3,3]
output[1,5] <- pf(output[1,4], df1 = output[1,1], df2 = output[3,1], lower.tail = FALSE)

# Block F and p left blank (optional: can calculate if you want)
output[2,4] <- NA
output[2,5] <- NA
output[3,4] <- NA
output[3,5] <- NA

#Total row
output[4,1] <- a*b - 1
output[4,2] <- sum((airplane_df$score - mean(airplane_df$score^2))) #Total Sum of Squares(SST)
output[4,3] <- NA
output[4,4] <- NA
output[4,5] <- NA
colnames(output) <- c("DF", "Sum Sq", "Mean Sq", "F value", "p-value")
rownames(output) <- c("treatment", "block", "residual", "total")
kable(output, caption = "ANOVA Table for Paper Airplane Data (Calculated from Scratch)")
```

Table 3: ANOVA Table for Paper Airplane Data (Calculated from Scratch)

	DF	Sum Sq	Mean Sq	F value	p-value
treatment	2	18816.93	9408.467	8.426373	0.0107365
block	4	16523.60	4130.900	NA	NA
residual	8	8932.40	1116.550	NA	NA
total	14	-435826.00	NA	NA	NA

Interpretation:

The from-scratch ANOVA table for the paper airplane experiment shows that the paperclip position treatment has a statistically significant effect, with an F-value of 8.43 and a p-value of 0.0107. Since the p-value is below the 0.05 significance level, we reject the null hypothesis and conclude that the position of the paperclip does affect how far the airplane flies.

The student block variable was included to help control for individual differences in throwing ability, but we're not focused on testing whether the students themselves differ significantly. Because of this, the F-value and p-value for the block row are marked as NA, which is normal when blocks are used only for control and not for direct analysis.