

# 565\_HW\_4

Matthew Stoebe

2025-04-15

## Question 1

```
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

site <- c(rep("A", 10), rep("B", 10), rep("C", 10), rep("D", 10))

defective <- c(57, 11, 57, 30, 119, 50, 31, 21, 67, 31,      # Site A
              39, 42, 133, 86, 3, 76, 41, 12, 114, 39,      # Site B
              28, 97, 34, 117, 6, 109, 33, 115, 65, 74,      # Site C
              72, 128, 73, 8, 74, 62, 78, 32, 8, 74)         # Site D
batch_size <- c(1000, 100, 1000, 500, 2000, 1000, 500, 500, 1000, 500,
              500, 500, 2000, 1000, 100, 1000, 500, 100, 2000, 500,
              500, 2000, 500, 2000, 100, 2000, 500, 2000, 1000, 1000,
              1000, 2000, 1000, 100, 1000, 1000, 1000, 500, 100, 1000)

chips_df <- data.frame(site, defective, batch_size)

agg_data <- chips_df %>%
  group_by(site) %>%
  summarize(total_def = sum(defective),
            total_chips = sum(batch_size),
            prop = total_def / total_chips)

x <- agg_data$total_def
n <- agg_data$total_chips
```

```

# Use the chi-square test for homogeneity (without continuity correction)
prop_test_result <- prop.test(x, n, correct = FALSE)
print(prop_test_result)

##
## 4-sample test for equality of proportions without continuity correction
##
## data:  x out of n
## X-squared = 22.599, df = 3, p-value = 4.895e-05
## alternative hypothesis: two.sided
## sample estimates:
##      prop 1      prop 2      prop 3      prop 4
## 0.05851852 0.07134146 0.05844828 0.07000000

```

We see that there is a significant differences between the factories. the rate at factory 4 and factor 2 is higher than the rate at factory 1 and 3. this is found at a  $\alpha = .01$  level

## Question 2

Need to get the dataset set up

## Question 3

```

beta0 <- 0.5
beta1 <- 2
beta2 <- 1.5
beta3 <- -1
sigma <- 0.5

x_val <- 1.2
d_clay <- 1
expected_height_clay <- beta0 + beta1 * x_val + beta2 * d_clay + beta3 * (x_val * d_clay)
cat("Expected height for clay soil with 1.2 gallons/week:", expected_height_clay, "inches\n")

## Expected height for clay soil with 1.2 gallons/week: 3.2 inches

x_val <- 1
expected_height_sandy <- beta0 + beta1 * x_val # since d=0
expected_height_clay <- beta0 + beta1 * x_val + beta2 * 1 + beta3 * (x_val * 1)
cat("At 1 gallon/week:\n")

## At 1 gallon/week:

cat(" Sandy soil:", expected_height_sandy, "inches\n")

## Sandy soil: 2.5 inches

```

```

cat(" Clay soil:", expected_height_clay, "inches\n")

## Clay soil: 3 inches

x_val <- 2
expected_height_sandy <- beta0 + beta1 * x_val # d=0
expected_height_clay <- beta0 + beta1 * x_val + beta2 * 1 + beta3 * (x_val * 1)
cat("At 2 gallons/week:\n")

## At 2 gallons/week:

cat(" Sandy soil:", expected_height_sandy, "inches\n")

## Sandy soil: 4.5 inches

cat(" Clay soil:", expected_height_clay, "inches\n")

## Clay soil: 4 inches

mu <- beta0 + beta1 * 1 + beta2 * 1 + beta3 * (1 * 1) # mean = 3.0 inches
threshold <- 4

p_above4 <- 1 - pnorm(threshold, mean = mu, sd = sigma)
cat("Probability that grass in clay soil with 1 gallon/week is taller than 4 inches:", p_above4, "\n")

## Probability that grass in clay soil with 1 gallon/week is taller than 4 inches: 0.02275013

```

## Question 4

(a) **Notation & Model:** \ Let  $Y_i$  be the dry weight of the  $i$ th seedling, and let  $x_i$  be the dose (in ounces) applied. Define

$$D_i = \begin{cases} 0, & \text{if seedling } i \text{ receives booster A,} \\ 1, & \text{if seedling } i \text{ receives booster B.} \end{cases}$$

For the 12 treated seedlings (ignoring controls), a suitable linear model is

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 D_i + \beta_3 (x_i \cdot D_i) + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2).$$

(b) **Parameter Interpretation:** \  $\beta_0$ : Baseline dry weight for booster A (when  $x_i = 0$ ). \ [4mm]  $\beta_1$ : Increase in dry weight per ounce of dose for booster A. \ [4mm]  $\beta_2$ : Difference in baseline weight (at  $x_i = 0$ ) between booster B and A. \ [4mm]  $\beta_3$ : Difference in the dose effect between boosters; so for booster B the dose-response slope is  $\beta_1 + \beta_3$ .

(c) **Hypotheses:** \ To test if the two boosters are equally effective (i.e., have the same baseline and dose effect), we set

$$H_0 : \beta_2 = 0 \text{ and } \beta_3 = 0, \\ H_a : \text{at least one of } \beta_2 \text{ or } \beta_3 \text{ differs from 0.}$$