Stat631 HW 1

Brandon Keck

1.

(a)

2 factors: Fertilizer type and Watering Irrigation

(b)

Levels: Fertilizer type (4 levels). Watering treatments (2 levels)

(c)

Treatments: 4 fertilizer types x 2 watering treatments = 8 treatments total

1. Fertilizer A + Standard Watering
2. Fertilizer A + Enhanced Watering
3. Fertilizer B + Standard Watering
4. Fertilizer B + Enhanced Watering
5. Fertilizer C + Standard Watering
6. Fertilizer C + Enhanced Watering
7. Fertilizer D + Standard Watering
8. Fertilizer D + Enhanced Watering

(d)

Total number of experimental Treatments: 8 experimental treatments.

(e)

Experimental Units: The sub-fields of cornfield.

(f)

Total number of experimental units: 8 treatments x 5 sub-fields = 40 experimental units

(g)

Measurement Units: The 10 corn plants selected from each sub-field.

(h)

Total number of measurement units: 5 sub-fields x 10 plants x 8 treatments

Total measurement units = 400

(i)

Number of replications: 5

(2)

(a)

```
library(readr)
wafers <- read_csv("~/Documents/EastBay/Spring2025/Stat631Anova/wafers.csv", show_col_type
```

```
head(wafers)
```

```
# A tibble: 6 x 2
     low  high
   <dbl> <dbl>
1 11.2    5.26
2  7.09   6.75
3  8.10   7.46
4 11.7    7.02
5 11.3    8.13
6 10.8    7.42
```
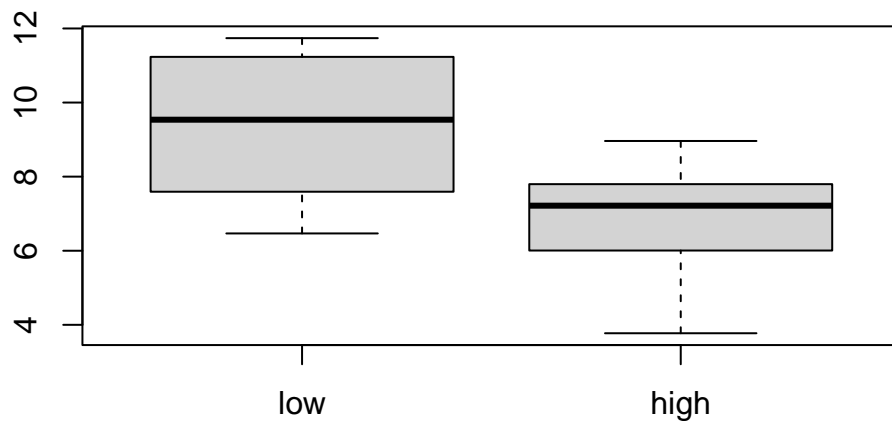
Step 1: Hypothesis

$H_0 : \mu_1 = \mu_2$

$H_A : \mu_1 \neq \mu_2$

Where group 1 is 95 degrees and group 2 is 100 degrees.

Step 2: Check conditions (assumptions):

```
# Generate boxplot
boxplot(wafers)
```


There are no noticeable outliers shown in the boxplot. Both the high and low temperature groups appear to be normally distributed and symmetric around the center. We can assume normality and proceed.

Step 3: Test Statistic:

```
t.test(wafers$low, wafers$high, var.equal = TRUE)
```

```
    Two Sample t-test

data:  wafers$low and wafers$high
t = 2.6751, df = 14, p-value = 0.01812
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
```

```
  0.4995743 4.5404257
sample estimates:
mean of x mean of y
 9.366625  6.846625
```

Our test statistic is 2.6751.

Step 4: p-value We have a p-value of 0.01812

Step 5: Decision With a p-value of 0.01812 which is less than our significance value of $\alpha = 0.05$ we reject the H_0

Step 6 Conclusion: The t-statistic is 2.675 and the p-value is 0.01812. Since the p-value is less than 0.05 we have enough evidence to conclude that there is a significant different in the mean photoresist thickness between wafers baked at 95 degrees Celsius and 100 degrees Celsius.

(b)

We are 95% confident that the true difference in mean of photoresist thickness for wafers baked at 95 degrees Celsius and wafers baked at 100 degrees Celsius is between 0.4996 kA and 4.540 kA.

(c)

```
# pooling variance
var_pooled <- (var(wafers$low) + var(wafers$high)) / 2

sd_pooled <- sqrt(var_pooled)

# creating power test
power.t.test(n = 8, delta = 2.5, sd = sd_pooled)
```

```
    Two-sample t test power calculation

              n = 8
          delta = 2.5
             sd = 1.884034
      sig.level = 0.05
          power = 0.6945829
    alternative = two.sided

NOTE: n is number in *each* group
```

The power of this test for detecting an actual difference in means of 2.5 kA is 0.695

(d)

```
# another power test
power.t.test(delta = 1.5, sd = sd_pooled, power = 0.9)
```

```
     Two-sample t test power calculation

              n = 34.14244
          delta = 1.5
             sd = 1.884034
      sig.level = 0.05
          power = 0.9
    alternative = two.sided
```

NOTE: n is number in *each* group

The sample size that would be necessary to detect an actual difference in means of 1.5 kA with a power of at least 0.9 is approximately 35.

3

(a)

Step 1: Hypothesis Testing

H_0 : The new textbook and old textbook lead to the same rating distributions

H_A : The new textbook results in higher ratings.

$H_0 : p_1 = p_2$

$H_A : p_1 < p_2$

where group 1 is the old textbook and group 2 is the new textbook ratings.

```
# Create the vectors of the new and old textbook scores
control <- c(37,47,62) # old book ratings
treatment <- c(46,45,32,56) # new book ratings

# obs is observed test statistic
```

```
obs <- (mean(control) - mean(treatment))

# pool data
pd <- c(control, treatment)

nsim <- 10000
results <- c() # the empty space to store for loop value


set.seed(631) # set the seed for reproducibility

# create the for loop
for(i in 1:nsim){
id <- sample(1:7, 3, replace = FALSE)
results[i] <- (mean(pd[id])-mean(pd[-id]))
}

mean(results <= obs)
```
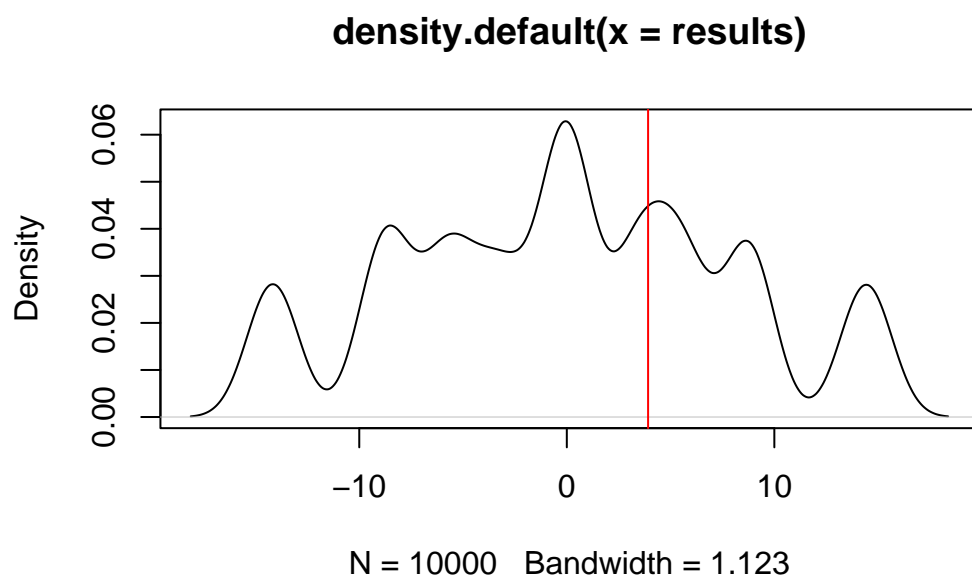
[1] 0.6884

Since the p-value is large at 0.6884, we fail to reject H_0. There is insufficient evidence that the new testbook leads to higher ratings.

```
# plot the distribution
plot(density(results))
abline(v = c(obs), col = "red")
```

density.default(x = results)

N = 10000   Bandwidth = 1.123

The red vertical line represents the observed difference in means.

```
t.test(treatment, control, alternative = "greater")$p.value
```

[1] 0.6599919