

Question 1 (Show work to get full points)

Suppose there are 4 levels of a single treatment factor to be compared in a balanced incomplete block design. There are only 6 blocks available each with only 2 experimental units.

(a). Compute the number of times each treatment level appears in the whole design.

4 levels of treatment means $a = 4$; 6 blocks means $b = 6$; 2 experimental units means $k = 2$;

$$N = kb = ra \implies r = kb/a = 2 \times 6 \div 4 = 3$$

(b). Compute the number of times each pair of treatment levels appears in the whole design.

$$\lambda = \frac{r(k-1)}{a-1} = \frac{3(2-1)}{4-1} = 1$$

(c). Provide a random assignment of treatment levels so that columns represent blocks and rows represent treatment levels.

| | block1 | block2 | block3 | block4 | block5 | block6 |
|-------|--------|--------|--------|--------|--------|--------|
| unit1 | 1 | 2 | 3 | 4 | 1 | 2 |
| unit2 | 2 | 3 | 4 | 1 | 3 | 4 |

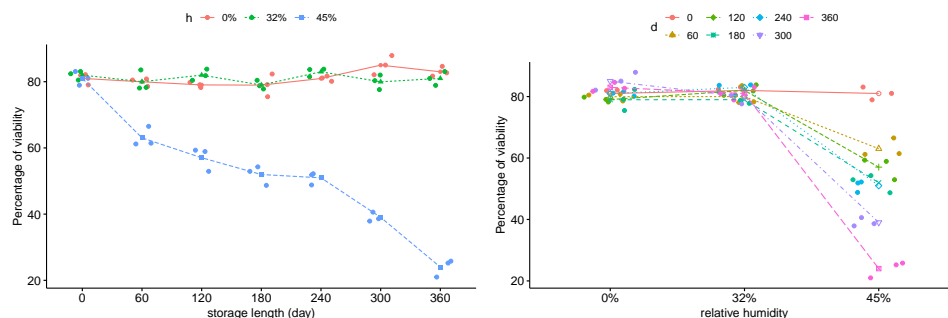
Question 2 (Use software to analyze the given data)

Big sagebrush is often planted in range restoration projects. An experiment is performed to determine the effects of storage length and relative humidity on the viability of seeds. Sixty-three batches of 300 seeds each are randomly divided into 21 groups of three. These 21 groups each receive a different treatment, namely the combinations of storage length (0, 60, 120, 180, 240, 300, or 360 days) and storage relative humidity (0, 32, or 45). After the storage time, the seeds are planted, and the response is the percentage of seeds that sprout. The data are shown above and given in Sagebrush Excel file.

Part I: Analyze these data for the effects of the factors on viability. Do not report code here. Your analysis should include the following:

(a). Plot of the data (a plot with data and/or means of treatment combinations). Do not report code here.

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    63 obs. of  5 variables:
## $ Humidity: num  0 0 0 0 0 0 0 0 0 0 0 ...
## $ Days    : num  0 0 0 60 60 60 120 120 120 180 ...
## $ Percent : num  82.1 79 81.9 78.6 80.8 80.5 79.8 79.1 78.2 82.3 ...
## $ h       : Factor w/ 3 levels "0%","32%","45%": 1 1 1 1 1 1 1 1 1 1 ...
## $ d       : Factor w/ 7 levels "0","60","120",...: 1 1 1 2 2 2 3 3 3 4 ...
```



(b). Description of the observed relationship between two factors based on the graph.

The line plot shows that not all lines are parallel. Difference in average percentage between humidity is not same for different storage length. There could be an interaction effect. The graph also show that,

When relative humidity is 0 or 32%, the average percentage is around 80% and have not obvious difference between different storage length.

When humidity is 45%, the average percentage of viability of seeds have obvious difference between different storage length. The longer storage length have lower average percentage of viability of seeds.

When storage length is 0 day, the average percentage of viability of seeds has not obvious difference between different relative humidity.

When any storage length is applied, the average percentage have not obvious difference between 0 and 32% humidity.

When storage length is 60 to 360 days, the average percentage of viability have significant difference between 45% humidity and other levels of humidity. The higher humidity has lower average percentage of viability of seeds.

(c). Tabular form of the numerical summaries for each treatment combination and factor levels separately.

| h | min | Q1 | median | Q3 | max | mean | sd | n | missing |
|-----|------|------|--------|------|------|----------|-----------|----|---------|
| 0% | 75.5 | 79.1 | 81.1 | 82.1 | 87.9 | 81.13333 | 2.670268 | 21 | 0 |
| 32% | 77.6 | 78.9 | 81.0 | 83.1 | 83.8 | 80.99524 | 2.128491 | 21 | 0 |
| 45% | 21.0 | 40.6 | 52.9 | 61.2 | 83.1 | 52.43333 | 17.196114 | 21 | 0 |

* h=relative humidity

| d | min | Q1 | median | Q3 | max | mean | sd | n | missing |
|-----|------|------|--------|------|------|----------|-----------|---|---------|
| 0 | 78.9 | 80.5 | 81.9 | 82.4 | 83.1 | 81.33333 | 1.599219 | 9 | 0 |
| 60 | 61.2 | 66.5 | 78.3 | 80.5 | 83.6 | 74.33333 | 8.767554 | 9 | 0 |
| 120 | 52.9 | 59.3 | 79.1 | 80.4 | 83.8 | 72.68889 | 11.984203 | 9 | 0 |
| 180 | 48.7 | 54.3 | 77.8 | 79.1 | 82.3 | 69.97778 | 13.708462 | 9 | 0 |
| 240 | 48.8 | 52.2 | 81.1 | 81.7 | 83.8 | 71.64444 | 15.580606 | 9 | 0 |
| 300 | 37.9 | 40.6 | 80.3 | 82.1 | 87.9 | 68.00000 | 21.921451 | 9 | 0 |
| 360 | 21.0 | 25.8 | 81.0 | 82.7 | 84.6 | 62.66667 | 29.070690 | 9 | 0 |

* d=storage length (days)

| h.d | min | Q1 | median | Q3 | max | mean | sd | n | missing |
|---------|------|-------|--------|-------|------|----------|-----------|---|---------|
| 0%.0 | 79.0 | 80.45 | 81.9 | 82.00 | 82.1 | 81.00000 | 1.7349352 | 3 | 0 |
| 32%.0 | 80.5 | 81.45 | 82.4 | 82.75 | 83.1 | 82.00000 | 1.3453624 | 3 | 0 |
| 45%.0 | 78.9 | 79.95 | 81.0 | 82.05 | 83.1 | 81.00000 | 2.1000000 | 3 | 0 |
| 0%.60 | 78.6 | 79.55 | 80.5 | 80.65 | 80.8 | 79.96667 | 1.1930353 | 3 | 0 |
| 32%.60 | 78.1 | 78.20 | 78.3 | 80.95 | 83.6 | 80.00000 | 3.1192948 | 3 | 0 |
| 45%.60 | 61.2 | 61.30 | 61.4 | 63.95 | 66.5 | 63.03333 | 3.0038864 | 3 | 0 |
| 0%.120 | 78.2 | 78.65 | 79.1 | 79.45 | 79.8 | 79.03333 | 0.8020806 | 3 | 0 |
| 32%.120 | 80.4 | 81.10 | 81.8 | 82.80 | 83.8 | 82.00000 | 1.7088007 | 3 | 0 |
| 45%.120 | 52.9 | 55.90 | 58.9 | 59.10 | 59.3 | 57.03333 | 3.5851546 | 3 | 0 |
| 0%.180 | 75.5 | 77.30 | 79.1 | 80.70 | 82.3 | 78.96667 | 3.4019602 | 3 | 0 |
| 32%.180 | 77.8 | 78.30 | 78.8 | 79.60 | 80.4 | 79.00000 | 1.3114877 | 3 | 0 |
| 45%.180 | 48.7 | 50.80 | 52.9 | 53.60 | 54.3 | 51.96667 | 2.9143324 | 3 | 0 |
| 0%.240 | 80.1 | 80.60 | 81.1 | 81.40 | 81.7 | 80.96667 | 0.8082904 | 3 | 0 |
| 32%.240 | 81.5 | 82.60 | 83.7 | 83.75 | 83.8 | 83.00000 | 1.3000000 | 3 | 0 |
| 45%.240 | 48.8 | 50.35 | 51.9 | 52.05 | 52.2 | 50.96667 | 1.8823744 | 3 | 0 |
| 0%.300 | 82.1 | 83.55 | 85.0 | 86.45 | 87.9 | 85.00000 | 2.9000000 | 3 | 0 |
| 32%.300 | 77.6 | 78.95 | 80.3 | 81.15 | 82.0 | 79.96667 | 2.2188586 | 3 | 0 |
| 45%.300 | 37.9 | 38.25 | 38.6 | 39.60 | 40.6 | 39.03333 | 1.4011900 | 3 | 0 |
| 0%.360 | 81.7 | 82.20 | 82.7 | 83.65 | 84.6 | 83.00000 | 1.4730920 | 3 | 0 |
| 32%.360 | 78.9 | 79.95 | 81.0 | 82.05 | 83.1 | 81.00000 | 2.1000000 | 3 | 0 |
| 45%.360 | 21.0 | 23.10 | 25.2 | 25.50 | 25.8 | 24.00000 | 2.6153394 | 3 | 0 |

* h.d=interaction items

(d). The complete (theoretical, not the estimated) linear model and explain the terms for this experiment.

This is a two-factor factorial model with fixed effect.

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{jik}$$

for $i = 1, 2, \dots, a$ represents 0, 32%, and 45% humidity; $j = 1, 2, \dots, b$ represents 7 levels of storage length; $k = 1, 2, \dots, n$ represents 3 experimental units.

y_{ijk} is the the average percentage of viability of seeds for the k^{th} Exerimental Unit when i^{th} level of relative humidity and j^{th} level of storage length are applied.

a is the number of levels of relative humidity being compared;

b is the number of levels of storage length being compared;

τ_i is fixed main effect of i^{th} level of relative humidity (Treatment effect of relative humidity);

β_j is fixed main effect of j^{th} level of storage length (Treatment effect of storage length);

$(\tau\beta)_{ij}$ is fixed interaction effect of i^{th} level of relative humidity and j^{th} level of storage length (Interaction effect of relative humidity and storage length)

ε_{jik} is random error for the k^{th} Exerimental Unit when i^{th} level of relative humidity and j^{th} level of storage length are applied.

The model includes below assumptions:

$$\varepsilon_{ijk} \sim iidN(0, \sigma^2); \sum_i^a \tau_i = 0; \sum_j^b \beta_j = 0; \sum_i^a (\tau\beta)_{ij} = 0; \sum_j^b (\tau\beta)_{ij} = 0$$

(e). The complete ANOVA table for the fitted model you have in part (d). Do not report code here.

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|------------------|----|----------|---------|---------|-----------|
| humidity | 2 | 11476 | 5738 | 1179 | 1.271e-37 |
| days | 6 | 1789 | 298.1 | 61.25 | 3.504e-19 |
| h:d | 12 | 4154 | 346.2 | 71.12 | 6.511e-24 |
| Residuals | 42 | 204.4 | 4.867 | | |
| Total | 62 | 17623.78 | | | |

(f). Conclusion from the test of significant interaction effect along with the p value. What do you recommend as the next step in your analysis? (That is, should you test for main effects or simple effects? Why? Explain clearly and provide reasons).

According to ANOVA table, there is a significant interaction effect from relative humidity and storage length on the average percentage at 5% significance level (P-value= $6.511 \times e^{-24}$). That means, effect of humidity and effect of storage length on average percentage of viability is not independent. Therefore, simple effects examinations are recommended.

(g). According to your answer in part (f), conclusion(s) from the appropriate tests about the effect of factors along with the p-value(s).

The pairwise test adjusted by Tukey method show the simple effects (full table and P-values in Part III):

- In comparisons of storage length Least Squares Means by humidity,

When 0 humidity is applied, **NONE** of comparisons of storage length has significant difference on the

average percentage around 5% significance level (all P-value>0.05).

When 32% humidity is applied, **NONE** of comparisons of storage length has significant difference on the average percentage around 5% significance level (all P-value>0.05).

When 45% humidity applied, **MOST** of comparisons of storage length have significant difference on the average percentage around 5% significance level (P-value<0.05) **EXCEPT** 60 versus 120 days, 120 versus 180 days, 120 versus 240 days, and 180 versus 240 days (P-value=0.0815, 0.2452, 0.0747, 1.0000, respectively)

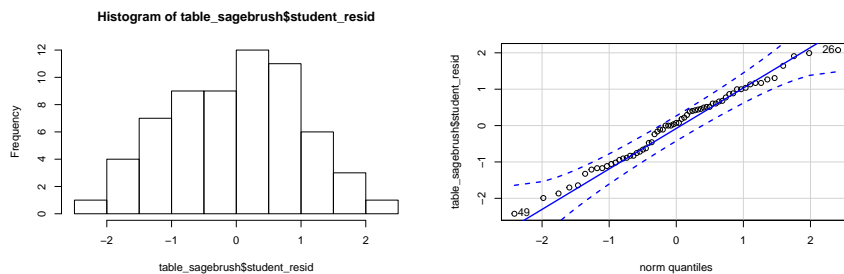
- In comparisons of humidity Least Squares Means by storage length,

When 0 day applied, **NONE** of comparisons of humidity have significant difference on the average percentage around 5% significance level (all P-value>0.05).

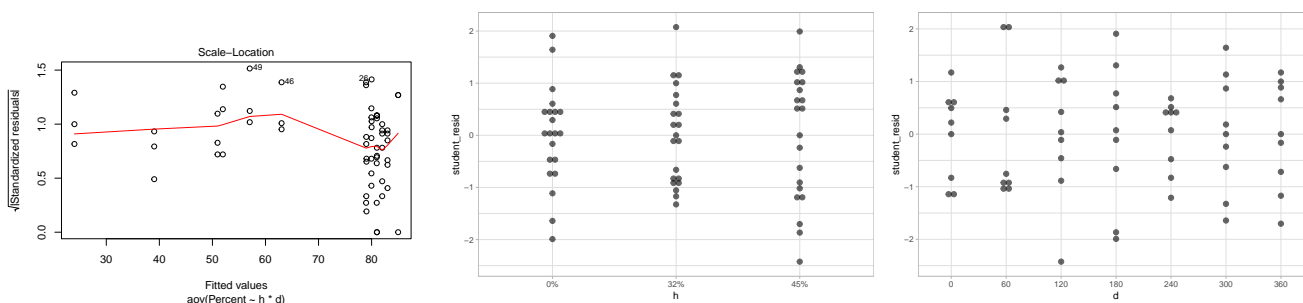
When 60 to 360 day applied, **NONE** of 0 versus 32% have significant difference on the average percentage around 5% significance level (all P-value>0.05). **ALL** of 0 versus 45%, 32% versus 45% have significant difference on the average percentage around 5% significance level (all P-value<0.05).

The overall conclusion is that humidity lower than 45% can keep the average percentage of viability of seeds being around 80%. If we cannot control the high humidity, the shorter storage length can give higher average percentage of viability of seeds.

(h). Check the assumptions of the fitted model and if you find any violation of assumptions, recommend solutions to fix them. Clearly explain and provide all the graphs and/or tables used to answer this question here. (Do not provide any code here).



[1] 49 26



The histogram of residuals shows an acceptable normal distribution. The QQ plot didn't show obvious violation of normality except a few data points are not on the line and flattening at the extremes, which is acceptable.

The plot of studentized residual versus predicted (fitted) value shows that except few outliers, the residuals are evenly distributed about zero at each predicted value (zero mean) and vertical deviations of residuals from zero are about same for each predicted value (constant variance).

The plots of studentized residual versus humanity and storage length levels didn't show obvious violation of zero mean and constant variance.

Part II: The researchers are interested in testing whether the mean percentage of seeds that sprout is significantly different at humidity of 45% compared to average of percentage of seeds that sprout at humidity of 0 and 32% when the storage length is 60 days.

(a). To test this, write a contrast statement in terms of the treatment means.

$$\Gamma : \mu_{32} - \frac{1}{2}(\mu_{12} + \mu_{22}) = 0 \implies \tau_{32} + (\tau\beta)_{32} - \frac{1}{2}(\tau_{12} + (\tau\beta)_{12} + \tau_{22} + (\tau\beta)_{22}) = 0$$

where subscript "32" represents 45% humidity and 60-days storage; "12" represents 0 humidity and 60-days storage; "22" represents 32% humidity and 60-days storage.

The H_0 for contrast is $\mu_{32} - \frac{1}{2}(\mu_{12} + \mu_{22}) = 0$ or $\tau_{32} + (\tau\beta)_{32} - \frac{1}{2}(\tau_{12} + (\tau\beta)_{12} + \tau_{22} + (\tau\beta)_{22}) = 0$.

The H_1 is they don't equal zero.

(b). test your contrast using the software and provide the conclusion along with p value.

The estimated value of the contrast is -16.9 and the adjusted Tukey's p value for testing the above hypotheses for contrast is less than 0.0001, which is small enough. We can reject the H_0 and conclude that the average percentage of viability of seeds for 60-days storage at 45% humidity versus that at 0 and 32% humidity are different at 5% significance level.

Part III: Provide your full code and/or output (only the ones used to answer above questions) here.

```
## # I(a)
## table_sagebrush <- read_xlsx("Sagebrush.xlsx")
## table_sagebrush$h <- factor(table_sagebrush$Humidity, levels = c(0, 32, 45),
##   labels = c("0%", "32%", "45%"))
## table_sagebrush$d <- factor(table_sagebrush$Days, levels = c(0, 60, 120, 180,
##   240, 300, 360), labels = c("0", "60", "120", "180", "240", "300", "360"))
## str(table_sagebrush)
## ggline(table_sagebrush, "d", "Percent", add = c("mean", "jitter"), shape = "h",
##   color = "h", linetype = "h", ylab = "viability of seeds", xlab = "storage length (day)")
## ggline(table_sagebrush, "h", "Percent", add = c("mean", "jitter"), shape = "d",
##   color = "d", linetype = "d", ylab = "viability of seeds", xlab = "relative humidity")
## # I(c)
## favstats(Percent ~ h, data = table_sagebrush)
## favstats(Percent ~ d, data = table_sagebrush)
## favstats(Percent ~ h + d, data = table_sagebrush)
## # I(e)
## model_sagebrush <- aov(Percent ~ h * d, data = table_sagebrush)
## summary(model_sagebrush)
## # I(g)
h_d <- pairs(lsmmeans(model_sagebrush, ~h | d))
d_h <- pairs(lsmmeans(model_sagebrush, ~d | h))
kable(test(rbind(d_h, h_d), adjust = "tukey"), format = "latex") %>% kable_styling("condensed",
  full_width = F, font_size = 6) %>% row_spec(c(43:63, 68, 69, 71, 72, 74,
  75, 77, 78, 80, 81, 83, 84), bold = T) %>% row_spec(c(49, 54, 55, 58), background = "#EAF8E8")
footnote(symbol = "Adjusted by Tukey's Method")
```

| h | d | contrast | estimate | SE | df | t.ratio | p.value |
|-----|-----|-----------|------------|----------|----|------------|-----------|
| 0% | . | 0 - 60 | 1.0333333 | 1.801352 | 42 | 0.5736433 | 0.9999965 |
| 0% | . | 0 - 120 | 1.9666667 | 1.801352 | 42 | 1.0917728 | 0.9967578 |
| 0% | . | 0 - 180 | 2.0333333 | 1.801352 | 42 | 1.1287820 | 0.9955898 |
| 0% | . | 0 - 240 | 0.0333333 | 1.801352 | 42 | 0.0185046 | 1.0000000 |
| 0% | . | 0 - 300 | -4.0000000 | 1.801352 | 42 | -2.2205548 | 0.5995036 |
| 0% | . | 0 - 360 | -2.0000000 | 1.801352 | 42 | -1.1102774 | 0.9962109 |
| 0% | . | 60 - 120 | 0.9333333 | 1.801352 | 42 | 0.5181294 | 0.9999989 |
| 0% | . | 60 - 180 | 1.0000000 | 1.801352 | 42 | 0.5551387 | 0.9999976 |
| 0% | . | 60 - 240 | -1.0000000 | 1.801352 | 42 | -0.5551387 | 0.9999976 |
| 0% | . | 60 - 300 | -5.0333333 | 1.801352 | 42 | -2.7941981 | 0.2537612 |
| 0% | . | 60 - 360 | -3.0333333 | 1.801352 | 42 | -1.6839207 | 0.9000001 |
| 0% | . | 120 - 180 | 0.0666667 | 1.801352 | 42 | 0.0370092 | 1.0000000 |
| 0% | . | 120 - 240 | -1.9333333 | 1.801352 | 42 | -1.0732681 | 0.9972374 |
| 0% | . | 120 - 300 | -5.9666667 | 1.801352 | 42 | -3.3123276 | 0.0851541 |
| 0% | . | 120 - 360 | -3.9666667 | 1.801352 | 42 | -2.2020502 | 0.6119457 |
| 0% | . | 180 - 240 | -2.0000000 | 1.801352 | 42 | -1.1102774 | 0.9962109 |
| 0% | . | 180 - 300 | -6.0333333 | 1.801352 | 42 | -3.3493368 | 0.0780433 |
| 0% | . | 180 - 360 | -4.0333333 | 1.801352 | 42 | -2.2390594 | 0.5870368 |
| 0% | . | 240 - 300 | -4.0333333 | 1.801352 | 42 | -2.2390594 | 0.5870368 |
| 0% | . | 240 - 360 | -2.0333333 | 1.801352 | 42 | -1.1287820 | 0.9955898 |
| 0% | . | 300 - 360 | 2.0000000 | 1.801352 | 42 | 1.1102774 | 0.9962109 |
| 32% | . | 0 - 60 | 2.0000000 | 1.801352 | 42 | 1.1102774 | 0.9962109 |
| 32% | . | 0 - 120 | 0.0000000 | 1.801352 | 42 | 0.0000000 | 1.0000000 |
| 32% | . | 0 - 180 | 3.0000000 | 1.801352 | 42 | 1.6654161 | 0.9067389 |
| 32% | . | 0 - 240 | -1.0000000 | 1.801352 | 42 | -0.5551387 | 0.9999976 |
| 32% | . | 0 - 300 | 2.0333333 | 1.801352 | 42 | 1.1287820 | 0.9955898 |
| 32% | . | 0 - 360 | 1.0000000 | 1.801352 | 42 | 0.5551387 | 0.9999976 |
| 32% | . | 60 - 120 | -2.0000000 | 1.801352 | 42 | -1.1102774 | 0.9962109 |
| 32% | . | 60 - 180 | 1.0000000 | 1.801352 | 42 | 0.5551387 | 0.9999976 |
| 32% | . | 60 - 240 | -3.0000000 | 1.801352 | 42 | -1.6654161 | 0.9067389 |
| 32% | . | 60 - 300 | 0.0333333 | 1.801352 | 42 | 0.0185046 | 1.0000000 |
| 32% | . | 60 - 360 | -1.0000000 | 1.801352 | 42 | -0.5551387 | 0.9999976 |
| 32% | . | 120 - 180 | 3.0000000 | 1.801352 | 42 | 1.6654161 | 0.9067389 |
| 32% | . | 120 - 240 | -1.0000000 | 1.801352 | 42 | -0.5551387 | 0.9999976 |
| 32% | . | 120 - 300 | 2.0333333 | 1.801352 | 42 | 1.1287820 | 0.9955898 |
| 32% | . | 120 - 360 | 1.0000000 | 1.801352 | 42 | 0.5551387 | 0.9999976 |
| 32% | . | 180 - 240 | -4.0000000 | 1.801352 | 42 | -2.2205548 | 0.5995036 |
| 32% | . | 180 - 300 | -0.9666667 | 1.801352 | 42 | -0.5366341 | 0.9999984 |
| 32% | . | 180 - 360 | -2.0000000 | 1.801352 | 42 | -1.1102774 | 0.9962109 |
| 32% | . | 240 - 300 | 3.0333333 | 1.801352 | 42 | 1.6839207 | 0.9000001 |
| 32% | . | 240 - 360 | 2.0000000 | 1.801352 | 42 | 1.1102774 | 0.9962109 |
| 32% | . | 300 - 360 | -1.0333333 | 1.801352 | 42 | -0.5736433 | 0.9999965 |
| 45% | . | 0 - 60 | 17.9666667 | 1.801352 | 42 | 9.9739919 | 0.0000000 |
| 45% | . | 0 - 120 | 23.9666667 | 1.801352 | 42 | 13.3048241 | 0.0000000 |
| 45% | . | 0 - 180 | 29.0333333 | 1.801352 | 42 | 16.1175268 | 0.0000000 |
| 45% | . | 0 - 240 | 30.0333333 | 1.801352 | 42 | 16.6726655 | 0.0000000 |
| 45% | . | 0 - 300 | 41.9666667 | 1.801352 | 42 | 23.2973206 | 0.0000000 |
| 45% | . | 0 - 360 | 57.0000000 | 1.801352 | 42 | 31.6429057 | 0.0000000 |
| 45% | . | 60 - 120 | 6.0000000 | 1.801352 | 42 | 3.3308322 | 0.0815323 |
| 45% | . | 60 - 180 | 11.0666667 | 1.801352 | 42 | 6.1435349 | 0.0000187 |
| 45% | . | 60 - 240 | 12.0666667 | 1.801352 | 42 | 6.6986736 | 0.0000031 |
| 45% | . | 60 - 300 | 24.0000000 | 1.801352 | 42 | 13.3233287 | 0.0000000 |
| 45% | . | 60 - 360 | 39.0333333 | 1.801352 | 42 | 21.6689138 | 0.0000000 |
| 45% | . | 120 - 180 | 5.0666667 | 1.801352 | 42 | 2.8127027 | 0.2451864 |
| 45% | . | 120 - 240 | 6.0666667 | 1.801352 | 42 | 3.3678414 | 0.0746835 |
| 45% | . | 120 - 300 | 18.0000000 | 1.801352 | 42 | 9.9924965 | 0.0000000 |
| 45% | . | 120 - 360 | 33.0333333 | 1.801352 | 42 | 18.3380816 | 0.0000000 |
| 45% | . | 180 - 240 | 1.0000000 | 1.801352 | 42 | 0.5551387 | 0.9999976 |
| 45% | . | 180 - 300 | 12.9333333 | 1.801352 | 42 | 7.1797938 | 0.0000006 |
| 45% | . | 180 - 360 | 27.9666667 | 1.801352 | 42 | 15.5253789 | 0.0000000 |
| 45% | . | 240 - 300 | 11.9333333 | 1.801352 | 42 | 6.6246551 | 0.0000039 |
| 45% | . | 240 - 360 | 26.9666667 | 1.801352 | 42 | 14.9702402 | 0.0000000 |
| 45% | . | 300 - 360 | 15.0333333 | 1.801352 | 42 | 8.3455851 | 0.0000000 |
| . | 0 | 0% - 32% | -1.0000000 | 1.801352 | 42 | -0.5551387 | 0.9999976 |
| . | 0 | 0% - 45% | 0.0000000 | 1.801352 | 42 | 0.0000000 | 1.0000000 |
| . | 0 | 32% - 45% | 1.0000000 | 1.801352 | 42 | 0.5551387 | 0.9999976 |
| . | 60 | 0% - 32% | -0.0333333 | 1.801352 | 42 | -0.0185046 | 1.0000000 |
| . | 60 | 0% - 45% | 16.9333333 | 1.801352 | 42 | 9.4003486 | 0.0000000 |
| . | 60 | 32% - 45% | 16.9666667 | 1.801352 | 42 | 9.4188532 | 0.0000000 |
| . | 120 | 0% - 32% | -2.9666667 | 1.801352 | 42 | -1.6469115 | 0.9131836 |
| . | 120 | 0% - 45% | 22.0000000 | 1.801352 | 42 | 12.2130513 | 0.0000000 |
| . | 120 | 32% - 45% | 24.9666667 | 1.801352 | 42 | 13.8599628 | 0.0000000 |
| . | 180 | 0% - 32% | -0.0333333 | 1.801352 | 42 | -0.0185046 | 1.0000000 |
| . | 180 | 0% - 45% | 27.0000000 | 1.801352 | 42 | 14.9887448 | 0.0000000 |
| . | 180 | 32% - 45% | 27.0333333 | 1.801352 | 42 | 15.0072494 | 0.0000000 |
| . | 240 | 0% - 32% | -2.0333333 | 1.801352 | 42 | -1.1287820 | 0.9955898 |
| . | 240 | 0% - 45% | 30.0000000 | 1.801352 | 42 | 16.6541609 | 0.0000000 |
| . | 240 | 32% - 45% | 32.0333333 | 1.801352 | 42 | 17.7829429 | 0.0000000 |
| . | 300 | 0% - 32% | 5.0333333 | 1.801352 | 42 | 2.7941981 | 0.2537612 |
| . | 300 | 0% - 45% | 45.9666667 | 1.801352 | 42 | 25.5178754 | 0.0000000 |
| . | 300 | 32% - 45% | 40.9333333 | 1.801352 | 42 | 22.7236773 | 0.0000000 |
| . | 360 | 0% - 32% | 2.0000000 | 1.801352 | 42 | 1.1102774 | 0.9962109 |
| . | 360 | 0% - 45% | 59.0000000 | 1.801352 | 42 | 32.7531831 | 0.0000000 |
| . | 360 | 32% - 45% | 57.0000000 | 1.801352 | 42 | 31.6429057 | 0.0000000 |

* Adjusted by Tukey's Method

```
## # I(h)
## table_sagebrush$student_resid <- rstudent(model_sagebrush)
## hist(table_sagebrush$student_resid)
## qqPlot(table_sagebrush$student_resid)
## plot(model_sagebrush, 3)
## ggplot(table_sagebrush, aes(h, student_resid)) + geom_dotplot(binaxis = "y",
##   stackdir = "center", binwidth = 0.1, alpha = 0.6) + theme_light()
## ggplot(table_sagebrush, aes(d, student_resid)) + geom_dotplot(binaxis = "y",
##   stackdir = "center", binwidth = 0.1, alpha = 0.6) + theme_light()
## # II(b)
table_sagebrush$hd <- interaction(table_sagebrush$h, table_sagebrush$d)
kable(summary(contrast(lsmmeans(lm(Percent ~ hd, table_sagebrush), "hd"), list(c32_12.22 = c(re
  3), -0.5, -0.5, 1, rep(0, 15))))), adjust = "tukey")) %>% kable_styling("condensed",
  full_width = F)
```

| contrast | estimate | SE | df | t.ratio | p.value |
|-----------|----------|----------|----|-----------|---------|
| c32_12.22 | -16.95 | 1.560016 | 42 | -10.86527 | 0 |

Question 3 (Show main steps of your work to get full points)

Let $y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \delta_k + \varepsilon_{ijk}$; $\varepsilon_{ijk} \sim iidN(0, \sigma^2)$; for $i = 1, \dots, a$; $j = 1, \dots, b$; $k = 1, \dots, n$;
 Derive the normal equations from the least squares parameter estimation method and estimate the model
 parameters using the following constraints. $\sum_i^a \hat{\tau}_i = 0$; $\sum_j^b \hat{\beta}_j = 0$; $\sum_i^a (\widehat{\tau\beta})_{ij} = 0$; $\sum_j^b (\widehat{\tau\beta})_{ij} = 0$; $\sum_k^n \hat{\delta}_k = 0$

This is a two-factor factorial in RCBD model with fixed effect.

$$SSE = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \mu - \tau_i - \beta_j - (\tau\beta)_{ij} - \delta_k)^2$$

Derive

$$\left. \frac{\partial SSE}{\partial \mu} \right|_{\hat{\mu}, \hat{\tau}_i, \hat{\beta}_j, (\widehat{\tau\beta})_{ij}, \hat{\delta}_k} = 2 \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \hat{\mu} - \hat{\tau}_i - \hat{\beta}_j - (\widehat{\tau\beta})_{ij} - \hat{\delta}_k)(-1) = 0$$

For $i = 1, \dots, a$,

$$\left. \frac{\partial SSE}{\partial \tau_i} \right|_{\hat{\mu}, \hat{\tau}_i, \hat{\beta}_j, (\widehat{\tau\beta})_{ij}, \hat{\delta}_k} = 2 \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \hat{\mu} - \hat{\tau}_i - \hat{\beta}_j - (\widehat{\tau\beta})_{ij} - \hat{\delta}_k)(-1) = 0$$

For $j = 1, \dots, b$,

$$\left. \frac{\partial SSE}{\partial \beta_j} \right|_{\hat{\mu}, \hat{\tau}_i, \hat{\beta}_j, (\widehat{\tau\beta})_{ij}, \hat{\delta}_k} = 2 \sum_{i=1}^a \sum_{k=1}^n (y_{ijk} - \hat{\mu} - \hat{\tau}_i - \hat{\beta}_j - (\widehat{\tau\beta})_{ij} - \hat{\delta}_k)(-1) = 0$$

For $i = 1, \dots, a$; $j = 1, \dots, b$,

$$\left. \frac{\partial SSE}{\partial (\tau\beta)_{ij}} \right|_{\hat{\mu}, \hat{\tau}_i, \hat{\beta}_j, (\widehat{\tau\beta})_{ij}, \hat{\delta}_k} = 2 \sum_{k=1}^n (y_{ijk} - \hat{\mu} - \hat{\tau}_i - \hat{\beta}_j - (\widehat{\tau\beta})_{ij} - \hat{\delta}_k)(-1) = 0$$

For $k = 1, \dots, n$,

$$\left. \frac{\partial SSE}{\partial \delta_k} \right|_{\hat{\mu}, \hat{\tau}_i, \hat{\beta}_j, (\widehat{\tau\beta})_{ij}, \hat{\delta}_k} = 2 \sum_{i=1}^a \sum_{j=1}^b (y_{ijk} - \hat{\mu} - \hat{\tau}_i - \hat{\beta}_j - (\widehat{\tau\beta})_{ij} - \hat{\delta}_k)(-1) = 0$$

$$\begin{cases} y_{...} = abn\hat{\mu} + bn \sum_{i=1}^a \hat{\tau}_i + an \sum_{j=1}^b \hat{\beta}_j + n \sum_{i=1}^a \sum_{j=1}^b \widehat{(\tau\beta)}_{ij} + ab \sum_{k=1}^n \hat{\delta}_k \\ y_{i..} = bn\hat{\mu} + bn\hat{\tau}_i + n \sum_{j=1}^b \hat{\beta}_j + n \sum_{j=1}^b \widehat{(\tau\beta)}_{ij} + b \sum_{k=1}^n \hat{\delta}_k \\ y_{.j.} = an\hat{\mu} + n \sum_{i=1}^a \hat{\tau}_i + an\hat{\beta}_j + n \sum_{i=1}^a \widehat{(\tau\beta)}_{ij} + a \sum_{k=1}^n \hat{\delta}_k \\ y_{ij.} = n\hat{\mu} + n\hat{\tau}_i + n\hat{\beta}_j + n\widehat{(\tau\beta)}_{ij} + \sum_{k=1}^n \hat{\delta}_k \\ y_{..k} = ab\hat{\mu} + b \sum_{i=1}^a \hat{\tau}_i + a \sum_{j=1}^b \hat{\beta}_j + \sum_{i=1}^a \sum_{j=1}^b \widehat{(\tau\beta)}_{ij} + ab\hat{\delta}_k \end{cases}$$

For $\sum_{i=1}^a \hat{\tau}_i = 0$; $\sum_{j=1}^b \hat{\beta}_j = 0$; $\sum_{i=1}^a \widehat{(\tau\beta)}_{ij} = 0$; $\sum_{j=1}^b \widehat{(\tau\beta)}_{ij} = 0$; $\sum_{k=1}^n \hat{\delta}_k = 0$,

and for $\hat{\mu}$ is constant, $\hat{\tau}_i, \hat{\beta}_j, \widehat{(\tau\beta)}_{ij}, \hat{\delta}_k$ are constants for summations on other parameters,

and for $y_{...} = abn\bar{y}_{...}$, $y_{i..} = bn\bar{y}_{i..}$, $y_{.j.} = an\bar{y}_{.j.}$, $y_{ij.} = n\bar{y}_{ij.}$, $y_{..k} = ab\bar{y}_{..k}$, then

$$\begin{cases} abn\bar{y}_{...} = abn\hat{\mu} + 0 + 0 + 0 + 0 \\ bn\bar{y}_{i..} = bn\hat{\mu} + bn\hat{\tau}_i + 0 + 0 + 0 \\ an\bar{y}_{.j.} = an\hat{\mu} + 0 + an\hat{\beta}_j + 0 + 0 \\ n\bar{y}_{ij.} = n\hat{\mu} + n\hat{\tau}_i + n\hat{\beta}_j + n\widehat{(\tau\beta)}_{ij} + 0 \\ ab\bar{y}_{..k} = ab\hat{\mu} + 0 + 0 + 0 + ab\hat{\delta}_k \end{cases} \implies \begin{cases} \hat{\mu} = \bar{y}_{...} \\ \hat{\tau}_i = \bar{y}_{i..} - \bar{y}_{...} \\ \hat{\beta}_j = \bar{y}_{.j.} - \bar{y}_{...} \\ \widehat{(\tau\beta)}_{ij} = \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...} \\ \hat{\delta}_k = \bar{y}_{..k} - \bar{y}_{...} \end{cases}$$

Question 4 (Show main steps of your work to get full points)

Consider the two-factor factorial model with fixed effects

$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}$, for $i = 1, 2, \dots, a$; $j = 1, 2, \dots, b$; $k = 1, 2, \dots, n$. $\varepsilon_{ijk} \sim iidN(0, \sigma^2)$; $\sum_{i=1}^a \tau_i = 0$ and $\sum_{j=1}^b \beta_j = 0$; $\sum_{i=1}^a (\tau\beta)_{ij} = 0$, $\sum_{j=1}^b (\tau\beta)_{ij} = 0$. $\bar{y}_{i..} = \frac{1}{bn} \sum_{j=1}^b \sum_{k=1}^n y_{ijk}$; $\bar{y}_{.j.} = \frac{1}{an} \sum_{i=1}^a \sum_{k=1}^n y_{ijk}$; $y_{...} = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n y_{ijk}$; $\bar{y}_{...} = \frac{y_{...}}{abn}$. $SS_A = bn \sum_{i=1}^a (\bar{y}_{i..} - \bar{y}_{...})^2$; $SS_B = an \sum_{j=1}^b (\bar{y}_{.j.} - \bar{y}_{...})^2$; $SS_{AB} = n \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2$

- Step 1: Expand the sum-of-squares terms

$$\begin{aligned} n \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{ij.} - \bar{y}_{...})^2 &= n \sum_{i=1}^a \sum_{j=1}^b [(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}) + (\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{.j.} - \bar{y}_{...})]^2 \\ &= n \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 + bn \sum_{i=1}^a (\bar{y}_{i..} - \bar{y}_{...})^2 + an \sum_{j=1}^b (\bar{y}_{.j.} - \bar{y}_{...})^2 + n \sum_{i=1}^a \sum_{j=1}^b [\text{three cross product terms}] \end{aligned}$$

- Step 2: Proof the first cross-product terms equal 0

$$\sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{i..} - \bar{y}_{...})(\bar{y}_{.j.} - \bar{y}_{...}) = \sum_{i=1}^a (\bar{y}_{i..} - \bar{y}_{...}) \sum_{j=1}^b (\bar{y}_{.j.} - \bar{y}_{...}) = (a\bar{y}_{...} - a\bar{y}_{...})(b\bar{y}_{...} - b\bar{y}_{...}) = 0$$

- Step 3: Proof the second cross-product terms equal 0

$$\begin{aligned}
\sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{ij} - \bar{y}_{i..} - \bar{y}_{.j} + \bar{y}_{...})(\bar{y}_{i..} - \bar{y}_{...}) &= \sum_{i=1}^a \sum_{j=1}^b \left[\bar{y}_{ij} \bar{y}_{i..} - \bar{y}_{i..}^2 - \bar{y}_{i..} \bar{y}_{.j} + \bar{y}_{i..} \bar{y}_{...} - \bar{y}_{ij} \bar{y}_{...} + \bar{y}_{i..} \bar{y}_{...} + \bar{y}_{.j} \bar{y}_{...} - \bar{y}_{...}^2 \right] \\
&= \sum_{i=1}^a \bar{y}_{i..} \sum_{j=1}^b \bar{y}_{ij} - \sum_{i=1}^a \sum_{j=1}^b \bar{y}_{i..}^2 - \sum_{i=1}^a \bar{y}_{i..} \sum_{j=1}^b \bar{y}_{.j} + 2\bar{y}_{...} \sum_{i=1}^a \sum_{j=1}^b \bar{y}_{i..} - \bar{y}_{...} \sum_{i=1}^a \sum_{j=1}^b \bar{y}_{ij} + \bar{y}_{...} \sum_{i=1}^a \sum_{j=1}^b \bar{y}_{.j} - \sum_{i=1}^a \sum_{j=1}^b \bar{y}_{...}^2 \\
&= \sum_{i=1}^a \bar{y}_{i..} b \bar{y}_{i..} - \sum_{i=1}^a b \bar{y}_{i..}^2 - a \bar{y}_{...} b \bar{y}_{...} + 2\bar{y}_{...} a b \bar{y}_{...} - \bar{y}_{...} a b \bar{y}_{...} + \bar{y}_{...} a b \bar{y}_{...} - a b \bar{y}_{...}^2 = 0
\end{aligned}$$

- Step 4: Proof the third cross-product terms equal 0

$$\begin{aligned}
\sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{ij} - \bar{y}_{i..} - \bar{y}_{.j} + \bar{y}_{...})(\bar{y}_{.j} - \bar{y}_{...}) &= \sum_{i=1}^a \sum_{j=1}^b \left[\bar{y}_{ij} \bar{y}_{.j} - \bar{y}_{i..} \bar{y}_{.j} - \bar{y}_{.j}^2 + \bar{y}_{.j} \bar{y}_{...} - \bar{y}_{ij} \bar{y}_{...} + \bar{y}_{i..} \bar{y}_{...} + \bar{y}_{.j} \bar{y}_{...} - \bar{y}_{...}^2 \right] \\
&= \sum_{j=1}^b \bar{y}_{.j} \sum_{i=1}^a \bar{y}_{ij} - \sum_{i=1}^a \bar{y}_{i..} \sum_{j=1}^b \bar{y}_{.j} - \sum_{i=1}^a \sum_{j=1}^b \bar{y}_{.j}^2 + 2\bar{y}_{...} \sum_{i=1}^a \sum_{j=1}^b \bar{y}_{.j} - \bar{y}_{...} \sum_{i=1}^a \sum_{j=1}^b \bar{y}_{ij} + \bar{y}_{...} \sum_{i=1}^a \sum_{j=1}^b \bar{y}_{.j} - \sum_{i=1}^a \sum_{j=1}^b \bar{y}_{...}^2 \\
&= \sum_{j=1}^b \bar{y}_{.j} a \bar{y}_{.j} - a \bar{y}_{...} b \bar{y}_{...} - \sum_{i=1}^a b \bar{y}_{.j}^2 + 2\bar{y}_{...} a b \bar{y}_{...} - \bar{y}_{...} a b \bar{y}_{...} + \bar{y}_{...} a b \bar{y}_{...} - a b \bar{y}_{...}^2 = 0
\end{aligned}$$

- Step 5: Complete the proof

For $n \sum_{i=1}^a \sum_{j=1}^b (\text{three cross product terms}) = 0$, then

$$n \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{ij} - \bar{y}_{...})^2 = n \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{ij} - \bar{y}_{i..} - \bar{y}_{.j} + \bar{y}_{...})^2 + b n \sum_{i=1}^a (\bar{y}_{i..} - \bar{y}_{...})^2 + a n \sum_{j=1}^b (\bar{y}_{.j} - \bar{y}_{...})^2 = SS_{AB} + SS_A + SS_B$$

Therefore,

$$SS_{AB} = n \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{ij} - \bar{y}_{...})^2 - SS_A - SS_B$$