InClassChallenge7

Alex Berry

2025-04-02

Table of Contents

# QUESTION 1

## Load in the appropriate libraries

library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.1 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.4 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(lme4)

## Loading required package: Matrix  
##   
## Attaching package: 'Matrix'  
##   
## The following objects are masked from 'package:tidyr':  
##   
## expand, pack, unpack

library(emmeans)

## Welcome to emmeans.  
## Caution: You lose important information if you filter this package's results.  
## See '? untidy'

library(multcomp)

## Loading required package: mvtnorm  
## Loading required package: survival  
## Loading required package: TH.data  
## Loading required package: MASS  
##   
## Attaching package: 'MASS'  
##   
## The following object is masked from 'package:dplyr':  
##   
## select  
##   
##   
## Attaching package: 'TH.data'  
##   
## The following object is masked from 'package:MASS':  
##   
## geyser

library(multcompView)

## Read in the dataset

plant.emergence <- read.csv("PlantEmergence.csv")

## Turn Treatment, DaysAfterPlanting and Rep into Factors

plant.emergence$Treatment <- as.factor(plant.emergence$Treatment)  
plant.emergence$DaysAfterPlanting <- as.factor(plant.emergence$DaysAfterPlanting)  
plant.emergence$Rep <- as.factor(plant.emergence$Rep)

# QUESTION 2

## Fit a linear model to predict Emergence using Treatment and DaysAfterPlanting along with the interaciton

lm1 <- lm(Emergence ~ Treatment \* DaysAfterPlanting, data = plant.emergence)

## Summary of the linear model and ANOVA results

summary(lm1)

##   
## Call:  
## lm(formula = Emergence ~ Treatment \* DaysAfterPlanting, data = plant.emergence)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -21.250 -6.062 -0.875 6.750 21.875   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.823e+02 5.324e+00 34.229 <2e-16 \*\*\*  
## Treatment2 -1.365e+02 7.530e+00 -18.128 <2e-16 \*\*\*  
## Treatment3 1.112e+01 7.530e+00 1.477 0.142   
## Treatment4 2.500e+00 7.530e+00 0.332 0.741   
## Treatment5 8.750e+00 7.530e+00 1.162 0.248   
## Treatment6 7.000e+00 7.530e+00 0.930 0.355   
## Treatment7 -1.250e-01 7.530e+00 -0.017 0.987   
## Treatment8 9.125e+00 7.530e+00 1.212 0.228   
## Treatment9 2.375e+00 7.530e+00 0.315 0.753   
## DaysAfterPlanting14 1.000e+01 7.530e+00 1.328 0.187   
## DaysAfterPlanting21 1.062e+01 7.530e+00 1.411 0.161   
## DaysAfterPlanting28 1.100e+01 7.530e+00 1.461 0.147   
## Treatment2:DaysAfterPlanting14 1.625e+00 1.065e+01 0.153 0.879   
## Treatment3:DaysAfterPlanting14 -2.625e+00 1.065e+01 -0.247 0.806   
## Treatment4:DaysAfterPlanting14 -6.250e-01 1.065e+01 -0.059 0.953   
## Treatment5:DaysAfterPlanting14 2.500e+00 1.065e+01 0.235 0.815   
## Treatment6:DaysAfterPlanting14 1.000e+00 1.065e+01 0.094 0.925   
## Treatment7:DaysAfterPlanting14 -2.500e+00 1.065e+01 -0.235 0.815   
## Treatment8:DaysAfterPlanting14 -2.500e+00 1.065e+01 -0.235 0.815   
## Treatment9:DaysAfterPlanting14 6.250e-01 1.065e+01 0.059 0.953   
## Treatment2:DaysAfterPlanting21 3.500e+00 1.065e+01 0.329 0.743   
## Treatment3:DaysAfterPlanting21 -1.000e+00 1.065e+01 -0.094 0.925   
## Treatment4:DaysAfterPlanting21 1.500e+00 1.065e+01 0.141 0.888   
## Treatment5:DaysAfterPlanting21 2.875e+00 1.065e+01 0.270 0.788   
## Treatment6:DaysAfterPlanting21 4.125e+00 1.065e+01 0.387 0.699   
## Treatment7:DaysAfterPlanting21 -2.125e+00 1.065e+01 -0.200 0.842   
## Treatment8:DaysAfterPlanting21 -1.500e+00 1.065e+01 -0.141 0.888   
## Treatment9:DaysAfterPlanting21 -1.250e+00 1.065e+01 -0.117 0.907   
## Treatment2:DaysAfterPlanting28 2.750e+00 1.065e+01 0.258 0.797   
## Treatment3:DaysAfterPlanting28 -1.875e+00 1.065e+01 -0.176 0.861   
## Treatment4:DaysAfterPlanting28 3.264e-13 1.065e+01 0.000 1.000   
## Treatment5:DaysAfterPlanting28 2.500e+00 1.065e+01 0.235 0.815   
## Treatment6:DaysAfterPlanting28 2.125e+00 1.065e+01 0.200 0.842   
## Treatment7:DaysAfterPlanting28 -3.625e+00 1.065e+01 -0.340 0.734   
## Treatment8:DaysAfterPlanting28 -1.500e+00 1.065e+01 -0.141 0.888   
## Treatment9:DaysAfterPlanting28 -8.750e-01 1.065e+01 -0.082 0.935   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 10.65 on 108 degrees of freedom  
## Multiple R-squared: 0.9585, Adjusted R-squared: 0.945   
## F-statistic: 71.21 on 35 and 108 DF, p-value: < 2.2e-16

anova(lm1)

## Analysis of Variance Table  
##   
## Response: Emergence  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Treatment 8 279366 34921 307.9516 < 2.2e-16 \*\*\*  
## DaysAfterPlanting 3 3116 1039 9.1603 1.877e-05 \*\*\*  
## Treatment:DaysAfterPlanting 24 142 6 0.0522 1   
## Residuals 108 12247 113   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# QUESTION 3

## Simplified linear model without interaction but still testing both main effects

lm2 <- lm(Emergence ~ Treatment + DaysAfterPlanting, data = plant.emergence)  
summary(lm2)

##   
## Call:  
## lm(formula = Emergence ~ Treatment + DaysAfterPlanting, data = plant.emergence)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -21.1632 -6.1536 -0.8542 6.1823 21.3958   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 182.163 2.797 65.136 < 2e-16 \*\*\*  
## Treatment2 -134.531 3.425 -39.277 < 2e-16 \*\*\*  
## Treatment3 9.750 3.425 2.847 0.00513 \*\*   
## Treatment4 2.719 3.425 0.794 0.42876   
## Treatment5 10.719 3.425 3.129 0.00216 \*\*   
## Treatment6 8.812 3.425 2.573 0.01119 \*   
## Treatment7 -2.188 3.425 -0.639 0.52416   
## Treatment8 7.750 3.425 2.263 0.02529 \*   
## Treatment9 2.000 3.425 0.584 0.56028   
## DaysAfterPlanting14 9.722 2.283 4.258 3.89e-05 \*\*\*  
## DaysAfterPlanting21 11.306 2.283 4.951 2.21e-06 \*\*\*  
## DaysAfterPlanting28 10.944 2.283 4.793 4.36e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 9.688 on 132 degrees of freedom  
## Multiple R-squared: 0.958, Adjusted R-squared: 0.9545   
## F-statistic: 273.6 on 11 and 132 DF, p-value: < 2.2e-16

anova(lm2)

## Analysis of Variance Table  
##   
## Response: Emergence  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Treatment 8 279366 34921 372.070 < 2.2e-16 \*\*\*  
## DaysAfterPlanting 3 3116 1039 11.068 1.575e-06 \*\*\*  
## Residuals 132 12389 94   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

### Based on the results of the ANOVA in question 2, the interaction term was not significant between Treatment and DaysAfterPlanting (p = 1), meaning there is no eveidence that suggests taht the effect of treatmetn varies by days after planting. Thus, the interaction term is not necessary in the model. However, in the simplified model (lm2), which includes Treatment and DaysAfterPlanting as main effects but removes the interaction, the ANOVA results reveal Treatment and DaysAfterPlanting are highly significant, meaning both main effects are statistically significant in predicting Emergence (p < 2.2-16, p = 1.575e-06, respectively). For Treatment 1, the predicted emergence on Day 0 is 182.163. For Treatment 2, emergence is predicted to be 134.531 units lower than for Treatment 1, signifying Treatment 2’s strong negative impact on emergence. Basically, it simplifies the model because you don’t need to include the interaction since it is not statistically significant.

# QUESTION 4

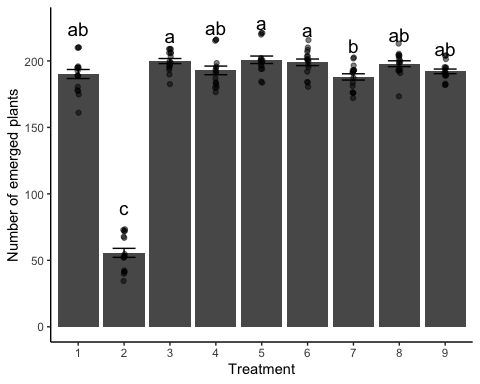
## Calculate the least squares mean for Treatment and perform Tukey separation then interpret the results

lsmeans <- emmeans(lm2, ~ Treatment)  
results\_lsmeans <- cld(lsmeans, alpha = 0.05, reversed = TRUE, details = TRUE)

### Based off the results of the least squares mean for Treatment and Tukey separation, Treatment 2 mean (55.6) is significantly lower than all others, with p < 0.0001. Treatments 5, 3, and 6 have similar means (199.0-200.9) as indicated by same group number (1). Treatments 8, 4, 9, and 1 also have similar means (190.2-197.9) as indicated by the same group number (12). Other significant differences revealed through Tukey separation include Treatment6-Treatment7 (p = 0.0425), Treatment3-Treatment7 (p = 0.0187), Treatment5-Treatment7 (p = 0.0074), along with each Treatment compared to Treatment 2 (p < 0.0001). Treatment 7 would be significantly lower than 8, 4, 9, and 1, then Treatment 2 would be less than Treatment 7.

# QUESTION 5

plot\_cldbars\_onefactor <- function(lm\_model, factor) {  
 data <- lm\_model$model  
variables <- colnames(lm\_model$model)  
dependent\_var <- variables[1]  
independent\_var <- variables[2:length(variables)]  
lsmeans <- emmeans(lm\_model, as.formula(paste("~", factor))) # estimate  
lsmeans  
Results\_lsmeans <- cld(lsmeans, alpha = 0.05, reversed = TRUE, details =  
TRUE, Letters = letters) # contrast with Tukey adjustment by default.  
# Extracting the letters for the bars  
sig.diff.letters <- data.frame(Results\_lsmeans$emmeans[,1],  
str\_trim(Results\_lsmeans$emmeans[,7]))  
colnames(sig.diff.letters) <- c(factor, "Letters")  
# for plotting with letters from significance test  
ave\_stand2 <- lm\_model$model %>%  
group\_by(!!sym(factor)) %>%  
dplyr::summarize(  
ave.emerge = mean(.data[[dependent\_var]], na.rm = TRUE),  
se = sd(.data[[dependent\_var]]) / sqrt(n())  
) %>%  
left\_join(sig.diff.letters, by = factor) %>%  
mutate(letter\_position = ave.emerge + 10 \* se)  
plot <- ggplot(data, aes(x = !! sym(factor), y = !! sym(dependent\_var))) +  
stat\_summary(fun = mean, geom = "bar") +  
stat\_summary(fun.data = mean\_se, geom = "errorbar", width = 0.5) +  
ylab("Number of emerged plants") +  
geom\_jitter(width = 0.02, alpha = 0.5) +  
geom\_text(data = ave\_stand2, aes(label = Letters, y = letter\_position),  
size = 5) +  
xlab(as.character(factor)) +  
theme\_classic()  
return(plot)  
}  
  
plot\_cldbars\_onefactor(lm2, "Treatment")

 ### Based off the letters in the plot, if a bar has the same letter, then it will not be significantly different than the other bars with the same letter. If a bar has a different letter than another, then the they are significantly different. Bars labeled “ab” mean the treatment is not significantly different from both groups a and b. Therefore, bars labeled just a or b will not be significantly different than bars labeled ab.

# QUESTION 6

[Link to GitHub](https://github.com/alexberry8/In-Class-Coding-Challenges)