Coding Challenge 7

Muhtarin Khayer Brohee

2025-04-08

Table of Contents

[Load libraries 1](#_Toc195021066)

[Load data 2](#_Toc195021067)

[Linear Model with Interaction 3](#_Toc195021068)

[Simplified Linear Model without Interaction 5](#_Toc195021069)

[Least Square Means and Tukey’s HSD with CLD 6](#_Toc195021070)

[Plotting Function and Interpretation 7](#_Toc195021071)

[Click the Github Link](https://github.com/mzb0226/PLPA-6820/tree/main/Coding%20Challenge%207)

### Load libraries

## Loading required package: 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 1.0.4   
## ── 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  
## Loading required package: lme4  
##   
## Loading required package: Matrix  
##   
##   
## Attaching package: 'Matrix'  
##   
##   
## The following objects are masked from 'package:tidyr':  
##   
## expand, pack, unpack  
##   
##   
## Loading required package: emmeans  
##   
## Welcome to emmeans.  
## Caution: You lose important information if you filter this package's results.  
## See '? untidy'  
##   
## Loading required package: 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  
##   
##   
## Loading required package: multcompView

### Load data

STAND <- readr::read\_csv("PlantEmergence.csv")

## Rows: 144 Columns: 7  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (2): DatePlanted, DateCounted  
## dbl (5): Plot, Treatment, Rep, Emergence, DaysAfterPlanting  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

# Convert to factors  
STAND$Treatment <- as.factor(STAND$Treatment)  
STAND$DaysAfterPlanting <- as.factor(STAND$DaysAfterPlanting)  
STAND$Rep <- as.factor(STAND$Rep)  
  
# View structure  
str(STAND)

## spc\_tbl\_ [144 × 7] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ Plot : num [1:144] 101 102 103 104 105 106 107 108 109 201 ...  
## $ Treatment : Factor w/ 9 levels "1","2","3","4",..: 1 2 3 4 5 6 7 8 9 6 ...  
## $ Rep : Factor w/ 4 levels "1","2","3","4": 1 1 1 1 1 1 1 1 1 2 ...  
## $ Emergence : num [1:144] 180.5 54.5 195 198.5 202 ...  
## $ DatePlanted : chr [1:144] "9-May-22" "9-May-22" "9-May-22" "9-May-22" ...  
## $ DateCounted : chr [1:144] "16-May-22" "16-May-22" "16-May-22" "16-May-22" ...  
## $ DaysAfterPlanting: Factor w/ 4 levels "7","14","21",..: 1 1 1 1 1 1 1 1 1 1 ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. Plot = col\_double(),  
## .. Treatment = col\_double(),  
## .. Rep = col\_double(),  
## .. Emergence = col\_double(),  
## .. DatePlanted = col\_character(),  
## .. DateCounted = col\_character(),  
## .. DaysAfterPlanting = col\_double()  
## .. )  
## - attr(\*, "problems")=<externalptr>

### Linear Model with Interaction

model1 <- lm(Emergence ~ Treatment \* DaysAfterPlanting, data = STAND)  
summary(model1)

##   
## Call:  
## lm(formula = Emergence ~ Treatment \* DaysAfterPlanting, data = STAND)  
##   
## 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(model1)

## 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

#### Interpretation of model1

* Intercept: 182.3 represents mean emergence in Treatment 1 at 7 days after planting (reference levels).
* Treatment2 has a large negative coefficient (-136.5) and is highly significant (p < 2e-16), meaning it had substantially lower emergence.
* Most interaction terms (e.g., Treatment2:DaysAfterPlanting14) are not significant (p > 0.05).
* ANOVA:Treatment and DaysAfterPlanting are significant.
* The interaction term (Treatment:DaysAfterPlanting) is not significant (F = 0.0522, p = 1), suggesting it can be removed.

### Simplified Linear Model without Interaction

model2 <- lm(Emergence ~ Treatment + DaysAfterPlanting, data = STAND)  
summary(model2)

##   
## Call:  
## lm(formula = Emergence ~ Treatment + DaysAfterPlanting, data = STAND)  
##   
## 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(model2)

## 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

#### Interpretation:

The linear model showed a significant overall fit, with an F-statistic of 273.6 (p < 2.2e-16), and an adjusted R-squared of 0.9545, indicating that about 95.45% of the variation in plant emergence is explained by the model. The intercept is 182.163, meaning the average emergence for Treatment 1 at 7 days after planting is approximately 182 plants. Treatment 2 had a significantly lower emergence, with a coefficient of -134.531 (p < 2e-16), indicating an average decrease of about 135 plants compared to Treatment 1. According to the ANOVA results, both Treatment (F = 372.07, p < 2.2e-16) and DaysAfterPlanting (F = 11.07, p = 1.575e-06) have significant effects on plant emergence.

### Least Square Means and Tukey’s HSD with CLD

lsm <- emmeans(model2, ~ Treatment)  
cld(lsm, Letters = letters)

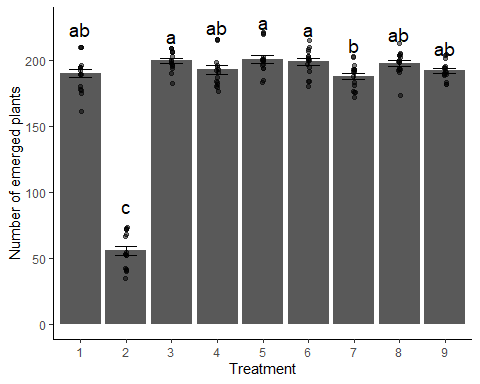
## Treatment emmean SE df lower.CL upper.CL .group  
## 2 55.6 2.42 132 50.8 60.4 a   
## 7 188.0 2.42 132 183.2 192.8 b   
## 1 190.2 2.42 132 185.4 194.9 bc   
## 9 192.2 2.42 132 187.4 196.9 bc   
## 4 192.9 2.42 132 188.1 197.7 bc   
## 8 197.9 2.42 132 193.1 202.7 bc   
## 6 199.0 2.42 132 194.2 203.8 c   
## 3 199.9 2.42 132 195.1 204.7 c   
## 5 200.9 2.42 132 196.1 205.7 c   
##   
## Results are averaged over the levels of: DaysAfterPlanting   
## Confidence level used: 0.95   
## P value adjustment: tukey method for comparing a family of 9 estimates   
## significance level used: alpha = 0.05   
## NOTE: If two or more means share the same grouping symbol,  
## then we cannot show them to be different.  
## But we also did not show them to be the same.

#### Interpretation of model2

* Treatment 2 had the lowest plant emergence (55.6) and was significantly different from all other treatments (group “a”).
* Treatments 3, 5, and 6 had the highest emergence (~199–201) and were grouped as “c”, meaning they were not significantly different from each other.
* Treatments 1, 4, 8, and 9 had moderately high emergence and were in group “bc”.
* Treatment 7 was in group “b”, slightly lower than the top-performing treatments but still much better than Treatment 2.
* Treatments that share the same letter are not significantly different; those with different letters are significantly different in plant emergence.

### Plotting Function and Interpretation

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)))  
 Results\_lsmeans <- cld(lsmeans, alpha = 0.05, reversed = TRUE, details = TRUE, Letters = letters)  
   
 sig.diff.letters <- data.frame(Results\_lsmeans$emmeans[,1],   
 str\_trim(Results\_lsmeans$emmeans[,7]))  
 colnames(sig.diff.letters) <- c(factor, "Letters")  
   
 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(model2, "Treatment")



#### Significance of letters in Plot

* The letters above the bars in the plot represent statistical groupings based on Tukey’s HSD test from the cld() function.
* Treatments that share the same letter (e.g., “a”, “ab”) are not significantly different from each other in terms of plant emergence. Treatments with different letters (e.g., “a” vs. “c”) are significantly different.