Linear Model coding challenge

Karamjit Baryah and Prativa Chhetri

2025-04-03

Plant.Emer=read.csv("PlantEmergence.csv")

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 1.0.4   
## ── 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

library(lme4)

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

library(emmeans)

## Warning: package 'emmeans' was built under R version 4.4.3

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

library(multcomp)

## Warning: package 'multcomp' was built under R version 4.4.3

## Loading required package: mvtnorm

## Warning: package 'mvtnorm' was built under R version 4.4.3

## Loading required package: survival  
## Loading required package: TH.data

## Warning: package 'TH.data' was built under R version 4.4.3

## 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)

## Warning: package 'multcompView' was built under R version 4.4.3

Plant.Emer$Treatment <- as.factor(Plant.Emer$Treatment)  
Plant.Emer$DaysAfterPlanting <- as.factor(Plant.Emer$DaysAfterPlanting)  
Plant.Emer$Rep <- as.factor(Plant.Emer$Rep)

str(Plant.Emer)

## 'data.frame': 144 obs. of 7 variables:  
## $ Plot : int 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 180.5 54.5 195 198.5 202 ...  
## $ DatePlanted : chr "9-May-22" "9-May-22" "9-May-22" "9-May-22" ...  
## $ DateCounted : chr "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 ...

lm\_model <- lm(Emergence ~ Treatment \* DaysAfterPlanting, data = Plant.Emer)

summary(lm\_model)

##   
## Call:  
## lm(formula = Emergence ~ Treatment \* DaysAfterPlanting, data = Plant.Emer)  
##   
## 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\_results <- anova(lm\_model)  
print(anova\_results)

## 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

# Fit the simplified linear model (without interaction)  
lm\_model\_simplified <- lm(Emergence ~ Treatment + DaysAfterPlanting, data = Plant.Emer)

# Summary of the simplified model  
summary(lm\_model\_simplified)

##   
## Call:  
## lm(formula = Emergence ~ Treatment + DaysAfterPlanting, data = Plant.Emer)  
##   
## 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 results  
anova\_results\_simplified <- anova(lm\_model\_simplified)  
print(anova\_results\_simplified)

## 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

# Compute least squares means for Treatment  
lsmeans\_treatment <- emmeans(lm\_model\_simplified, ~ Treatment)  
print(lsmeans\_treatment)

## Treatment emmean SE df lower.CL upper.CL  
## 1 190.2 2.42 132 185.4 194.9  
## 2 55.6 2.42 132 50.8 60.4  
## 3 199.9 2.42 132 195.1 204.7  
## 4 192.9 2.42 132 188.1 197.7  
## 5 200.9 2.42 132 196.1 205.7  
## 6 199.0 2.42 132 194.2 203.8  
## 7 188.0 2.42 132 183.2 192.8  
## 8 197.9 2.42 132 193.1 202.7  
## 9 192.2 2.42 132 187.4 196.9  
##   
## Results are averaged over the levels of: DaysAfterPlanting   
## Confidence level used: 0.95

# Perform Tukey's multiple comparisons  
tukey\_results <- pairs(lsmeans\_treatment, adjust = "tukey")  
print(tukey\_results)

## contrast estimate SE df t.ratio p.value  
## Treatment1 - Treatment2 134.531 3.43 132 39.277 <.0001  
## Treatment1 - Treatment3 -9.750 3.43 132 -2.847 0.1120  
## Treatment1 - Treatment4 -2.719 3.43 132 -0.794 0.9969  
## Treatment1 - Treatment5 -10.719 3.43 132 -3.129 0.0535  
## Treatment1 - Treatment6 -8.812 3.43 132 -2.573 0.2083  
## Treatment1 - Treatment7 2.188 3.43 132 0.639 0.9993  
## Treatment1 - Treatment8 -7.750 3.43 132 -2.263 0.3724  
## Treatment1 - Treatment9 -2.000 3.43 132 -0.584 0.9997  
## Treatment2 - Treatment3 -144.281 3.43 132 -42.124 <.0001  
## Treatment2 - Treatment4 -137.250 3.43 132 -40.071 <.0001  
## Treatment2 - Treatment5 -145.250 3.43 132 -42.406 <.0001  
## Treatment2 - Treatment6 -143.344 3.43 132 -41.850 <.0001  
## Treatment2 - Treatment7 -132.344 3.43 132 -38.638 <.0001  
## Treatment2 - Treatment8 -142.281 3.43 132 -41.540 <.0001  
## Treatment2 - Treatment9 -136.531 3.43 132 -39.861 <.0001  
## Treatment3 - Treatment4 7.031 3.43 132 2.053 0.5099  
## Treatment3 - Treatment5 -0.969 3.43 132 -0.283 1.0000  
## Treatment3 - Treatment6 0.938 3.43 132 0.274 1.0000  
## Treatment3 - Treatment7 11.938 3.43 132 3.485 0.0187  
## Treatment3 - Treatment8 2.000 3.43 132 0.584 0.9997  
## Treatment3 - Treatment9 7.750 3.43 132 2.263 0.3724  
## Treatment4 - Treatment5 -8.000 3.43 132 -2.336 0.3288  
## Treatment4 - Treatment6 -6.094 3.43 132 -1.779 0.6957  
## Treatment4 - Treatment7 4.906 3.43 132 1.432 0.8832  
## Treatment4 - Treatment8 -5.031 3.43 132 -1.469 0.8678  
## Treatment4 - Treatment9 0.719 3.43 132 0.210 1.0000  
## Treatment5 - Treatment6 1.906 3.43 132 0.557 0.9998  
## Treatment5 - Treatment7 12.906 3.43 132 3.768 0.0074  
## Treatment5 - Treatment8 2.969 3.43 132 0.867 0.9943  
## Treatment5 - Treatment9 8.719 3.43 132 2.545 0.2204  
## Treatment6 - Treatment7 11.000 3.43 132 3.212 0.0425  
## Treatment6 - Treatment8 1.062 3.43 132 0.310 1.0000  
## Treatment6 - Treatment9 6.812 3.43 132 1.989 0.5538  
## Treatment7 - Treatment8 -9.938 3.43 132 -2.901 0.0978  
## Treatment7 - Treatment9 -4.188 3.43 132 -1.223 0.9502  
## Treatment8 - Treatment9 5.750 3.43 132 1.679 0.7583  
##   
## Results are averaged over the levels of: DaysAfterPlanting   
## P value adjustment: tukey method for comparing a family of 9 estimates

# Generate compact letter display for Treatment  
cld\_results <- cld(lsmeans\_treatment, Letters = letters)  
print(cld\_results)

## 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.

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)  
}

# Generate the plot for Treatment  
plot\_cldbars\_onefactor(lm\_model\_simplified, "Treatment")

