Coding Challenge7 04.03.2025

Katie Clouse

2025-04-03

1. Data | Packages| as.factor variables

```
plantemergence <- read.csv("C:/Users/katie/Downloads/PlantEmergence.csv")
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.4.2
```

```
## Warning: package 'dplyr' was built under R version 4.4.2
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(lme4)
```

```
## Warning: package 'lme4' was built under R version 4.4.2
```

```
## 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'
```

```r
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
```

```r
library(multcompView)
```

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

```r
plantemergence$Treatment <- as.factor(plantemergence$Treatment)
plantemergence$DaysafterPlanting <- as.factor(plantemergence$DaysAfterPlanting)
plantemergence$Rep <- as.factor(plantemergence$Rep)
```

2. Fit the Linear Model

```r
lm.interaction <- lm(Emergence ~ Treatment*DaysafterPlanting, data = plantemergence)
summary(lm.interaction)
```

```
##
## Call:
## lm(formula = Emergence ~ Treatment * DaysafterPlanting, data = plantemergence)
##
## 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(lm.interaction)
```

```
## 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
```

3. Simplified Linear Model

```
simple.lm <- lm(Emergence~Treatment, data = plantemergence)
simple.lm2 <- lm(Emergence~DaysafterPlanting, data = plantemergence)
anova(simple.lm)
```

```
## Analysis of Variance Table
##
## Response: Emergence
##            Df Sum Sq Mean Sq F value    Pr(>F)
## Treatment   8 279366   34921  304.05 < 2.2e-16 ***
## Residuals 135  15505     115
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(simple.lm)
```

```
##
## Call:
## lm(formula = Emergence ~ Treatment, data = plantemergence)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -29.1563  -6.5234  -0.5625   5.9062  23.1250
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  190.156      2.679  70.974  < 2e-16 ***
## Treatment2  -134.531      3.789 -35.506  < 2e-16 ***
## Treatment3     9.750      3.789   2.573  0.01116 *
## Treatment4     2.719      3.789   0.718  0.47428
## Treatment5    10.719      3.789   2.829  0.00538 **
## Treatment6     8.812      3.789   2.326  0.02152 *
## Treatment7    -2.188      3.789  -0.577  0.56468
## Treatment8     7.750      3.789   2.045  0.04276 *
## Treatment9     2.000      3.789   0.528  0.59848
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.72 on 135 degrees of freedom
## Multiple R-squared:  0.9474, Adjusted R-squared:  0.9443
## F-statistic:   304 on 8 and 135 DF,  p-value: < 2.2e-16
```

```
anova(simple.lm2)
```

```
## Analysis of Variance Table
##
```

```
## Response: Emergence
##                     Df Sum Sq Mean Sq F value Pr(>F)
## DaysafterPlanting   3   3116  1038.8  0.4984  0.684
## Residuals         140 291755  2084.0
```

```
summary(simple.lm2)
```

```
##
## Call:
## lm(formula = Emergence ~ DaysafterPlanting, data = plantemergence)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -142.556    7.667   12.625   20.889   38.667
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)          171.611      7.608  22.555   <2e-16 ***
## DaysafterPlanting14    9.722     10.760   0.904    0.368
## DaysafterPlanting21   11.306     10.760   1.051    0.295
## DaysafterPlanting28   10.944     10.760   1.017    0.311
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 45.65 on 140 degrees of freedom
## Multiple R-squared:  0.01057,    Adjusted R-squared:  -0.01063
## F-statistic: 0.4984 on 3 and 140 DF,  p-value: 0.684
```

```
simplified_lm <- lm(Emergence~Treatment + DaysafterPlanting, data = plantemergence)
summary(simplified_lm)
```

```
##
## Call:
## lm(formula = Emergence ~ Treatment + DaysafterPlanting, data = plantemergence)
##
## 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(simplified_lm)
```

```
## 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
```

```
# Intercept and coefficient on treatment2: The intercept for
#Treatment 2 is 182.163 and this is the baseline of for the data.
#The coefficient is the result of adding (-134.531) to 182.163,
#which is 47.632. As Treatment and DaysafterPlanting increase by 1 unit,
#the emergence increases by 47.632.
```

4. Finding Least Square Means

```
lsmeans <- emmeans(simple.lm, ~Treatment)
Results_lsmeans <- cld(lsmeans, alpha = 0.05, details = TRUE)
print(Results_lsmeans)
```

```
## $emmeans
##  Treatment emmean   SE  df lower.CL upper.CL .group
##  2           55.6 2.68 135     50.3     60.9  1
##  7          188.0 2.68 135    182.7    193.3   2
##  1          190.2 2.68 135    184.9    195.5   23
##  9          192.2 2.68 135    186.9    197.5   23
##  4          192.9 2.68 135    187.6    198.2   23
##  8          197.9 2.68 135    192.6    203.2   23
##  6          199.0 2.68 135    193.7    204.3   23
##  3          199.9 2.68 135    194.6    205.2   23
##  5          200.9 2.68 135    195.6    206.2    3
##
## 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.
##
## $comparisons
##  contrast               estimate   SE  df t.ratio p.value
```

```
##  Treatment7 - Treatment2  132.344 3.79 135  34.928  <.0001
##  Treatment1 - Treatment2  134.531 3.79 135  35.506  <.0001
##  Treatment1 - Treatment7    2.188 3.79 135   0.577  0.9997
##  Treatment9 - Treatment2  136.531 3.79 135  36.033  <.0001
##  Treatment9 - Treatment7    4.188 3.79 135   1.105  0.9726
##  Treatment9 - Treatment1    2.000 3.79 135   0.528  0.9998
##  Treatment4 - Treatment2  137.250 3.79 135  36.223  <.0001
##  Treatment4 - Treatment7    4.906 3.79 135   1.295  0.9313
##  Treatment4 - Treatment1    2.719 3.79 135   0.718  0.9985
##  Treatment4 - Treatment9    0.719 3.79 135   0.190  1.0000
##  Treatment8 - Treatment2  142.281 3.79 135  37.551  <.0001
##  Treatment8 - Treatment7    9.938 3.79 135   2.623  0.1871
##  Treatment8 - Treatment1    7.750 3.79 135   2.045  0.5149
##  Treatment8 - Treatment9    5.750 3.79 135   1.518  0.8455
##  Treatment8 - Treatment4    5.031 3.79 135   1.328  0.9212
##  Treatment6 - Treatment2  143.344 3.79 135  37.831  <.0001
##  Treatment6 - Treatment7   11.000 3.79 135   2.903  0.0971
##  Treatment6 - Treatment1    8.812 3.79 135   2.326  0.3344
##  Treatment6 - Treatment9    6.812 3.79 135   1.798  0.6835
##  Treatment6 - Treatment4    6.094 3.79 135   1.608  0.7988
##  Treatment6 - Treatment8    1.062 3.79 135   0.280  1.0000
##  Treatment3 - Treatment2  144.281 3.79 135  38.079  <.0001
##  Treatment3 - Treatment7   11.938 3.79 135   3.151  0.0503
##  Treatment3 - Treatment1    9.750 3.79 135   2.573  0.2079
##  Treatment3 - Treatment9    7.750 3.79 135   2.045  0.5149
##  Treatment3 - Treatment4    7.031 3.79 135   1.856  0.6450
##  Treatment3 - Treatment8    2.000 3.79 135   0.528  0.9998
##  Treatment3 - Treatment6    0.938 3.79 135   0.247  1.0000
##  Treatment5 - Treatment2  145.250 3.79 135  38.335  <.0001
##  Treatment5 - Treatment7   12.906 3.79 135   3.406  0.0237
##  Treatment5 - Treatment1   10.719 3.79 135   2.829  0.1167
##  Treatment5 - Treatment9    8.719 3.79 135   2.301  0.3490
##  Treatment5 - Treatment4    8.000 3.79 135   2.111  0.4701
##  Treatment5 - Treatment8    2.969 3.79 135   0.784  0.9972
##  Treatment5 - Treatment6    1.906 3.79 135   0.503  0.9999
##  Treatment5 - Treatment3    0.969 3.79 135   0.256  1.0000
##
## P value adjustment: tukey method for comparing a family of 9 estimates
```

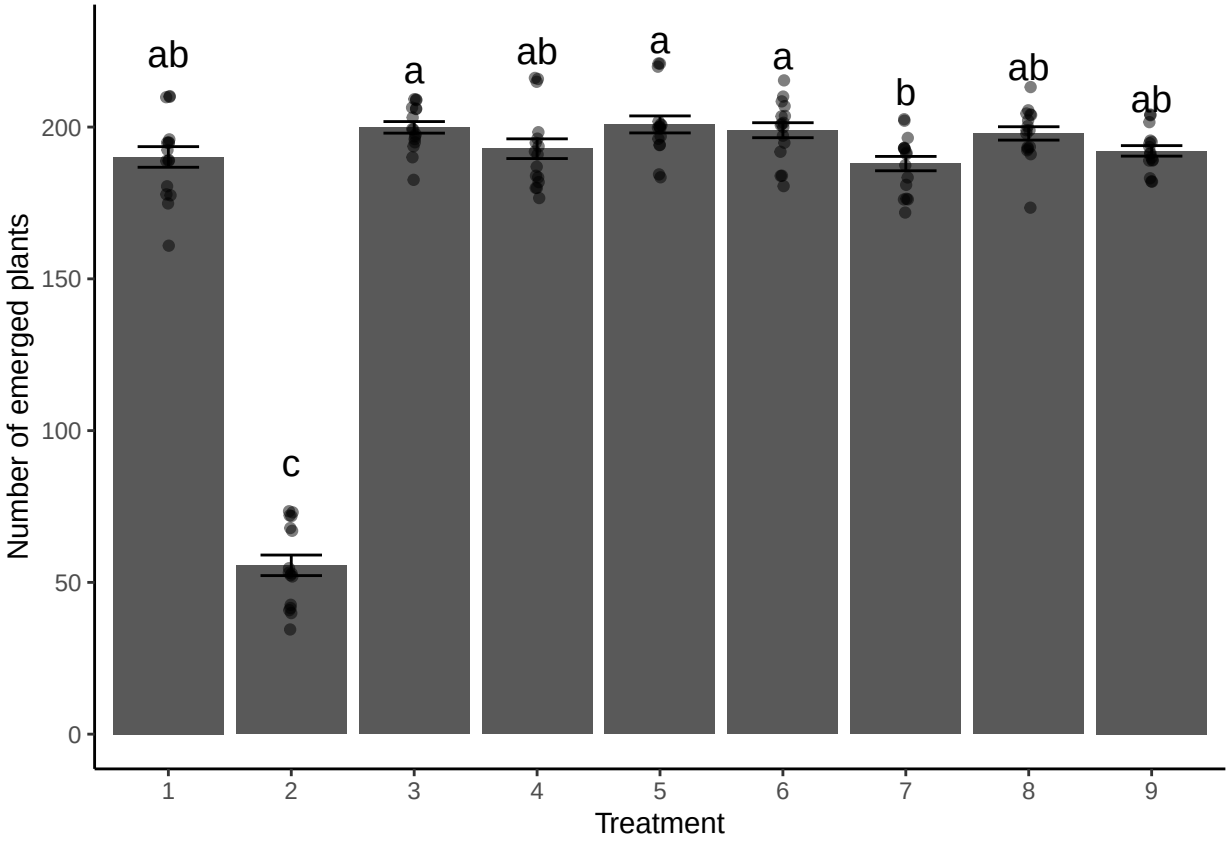
```
# It provides estimated means for each of the groups. The pairwise allows
#you to see the statistical comparisons between groups to see if there was
#any significance. Based on these results, there are 8 comparisons that have
#a pvalue of less than 0.0001.
```

6. Generating Plots

```
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
```

```r
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(simplified_lm, "Treatment")
```



Link to my github