

# CodingHW7

Katie Clouse

2025-04-02

Loading Packages

```
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(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(ggplot2)
```

Intro to Regression analysis in R

*#we want to estimate slope, intercept, adn standard deviation. Goal is to minimize distance from line t*

Continuous X and Continuous Y

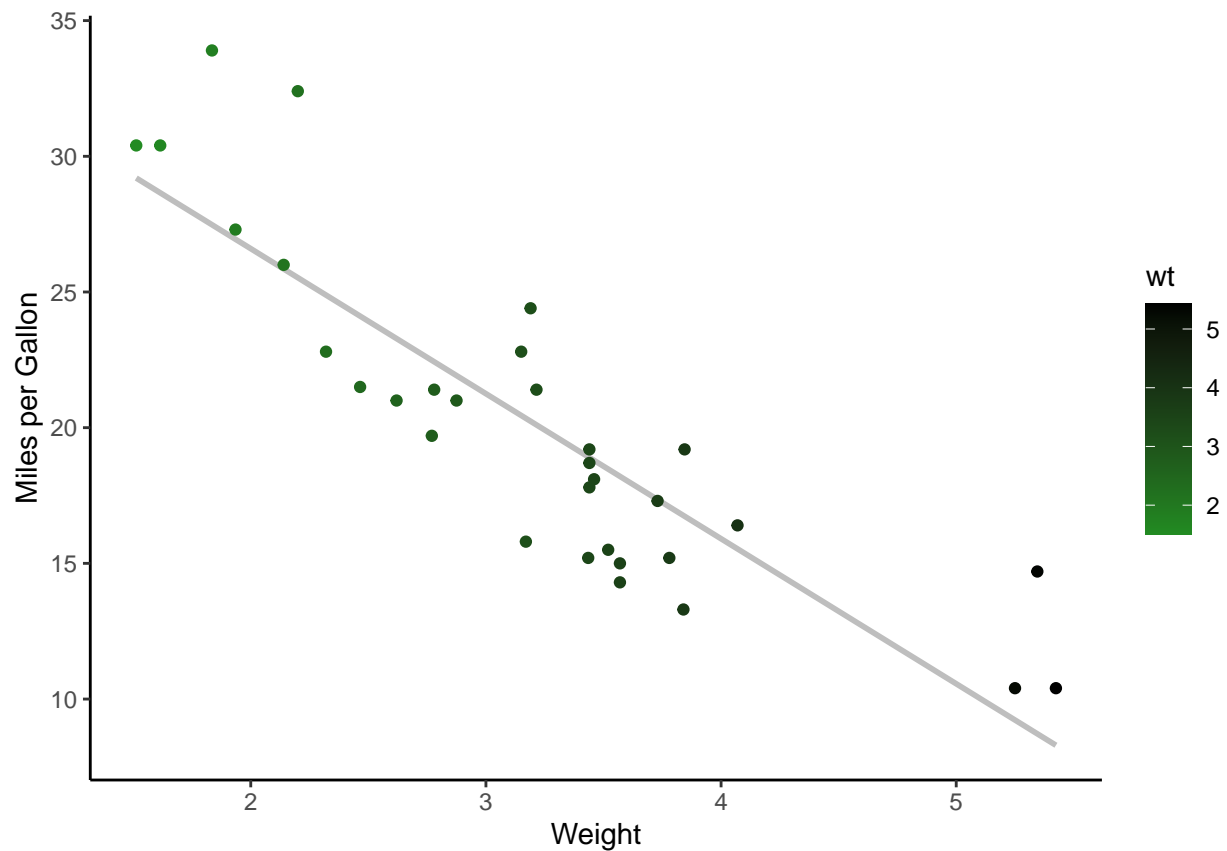
```
data("mtcars")
print(mtcars)
```

```
##           mpg  cyl  disp  hp drat    wt  qsec vs  am gear carb
## Mazda RX4      21.0    6  160.0  110 3.90 2.620 16.46 0  1    4    4
## Mazda RX4 Wag  21.0    6  160.0  110 3.90 2.875 17.02 0  1    4    4
## Datsun 710     22.8    4  108.0   93 3.85 2.320 18.61 1  1    4    1
## Hornet 4 Drive  21.4    6  258.0  110 3.08 3.215 19.44 1  0    3    1
## Hornet Sportabout 18.7    8  360.0  175 3.15 3.440 17.02 0  0    3    2
## Valiant        18.1    6  225.0  105 2.76 3.460 20.22 1  0    3    1
## Duster 360     14.3    8  360.0  245 3.21 3.570 15.84 0  0    3    4
## Merc 240D      24.4    4  146.7   62 3.69 3.190 20.00 1  0    4    2
## Merc 230       22.8    4  140.8   95 3.92 3.150 22.90 1  0    4    2
## Merc 280       19.2    6  167.6  123 3.92 3.440 18.30 1  0    4    4
## Merc 280C      17.8    6  167.6  123 3.92 3.440 18.90 1  0    4    4
## Merc 450SE     16.4    8  275.8  180 3.07 4.070 17.40 0  0    3    3
## Merc 450SL     17.3    8  275.8  180 3.07 3.730 17.60 0  0    3    3
## Merc 450SLC    15.2    8  275.8  180 3.07 3.780 18.00 0  0    3    3
## Cadillac Fleetwood 10.4    8  472.0  205 2.93 5.250 17.98 0  0    3    4
## Lincoln Continental 10.4    8  460.0  215 3.00 5.424 17.82 0  0    3    4
## Chrysler Imperial 14.7    8  440.0  230 3.23 5.345 17.42 0  0    3    4
## Fiat 128       32.4    4   78.7   66 4.08 2.200 19.47 1  1    4    1
## Honda Civic    30.4    4   75.7   52 4.93 1.615 18.52 1  1    4    2
## Toyota Corolla 33.9    4   71.1   65 4.22 1.835 19.90 1  1    4    1
```

## Toyota Corona	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
## Dodge Challenger	15.5	8	318.0	150	2.76	3.520	16.87	0	0	3	2
## AMC Javelin	15.2	8	304.0	150	3.15	3.435	17.30	0	0	3	2
## Camaro Z28	13.3	8	350.0	245	3.73	3.840	15.41	0	0	3	4
## Pontiac Firebird	19.2	8	400.0	175	3.08	3.845	17.05	0	0	3	2
## Fiat X1-9	27.3	4	79.0	66	4.08	1.935	18.90	1	1	4	1
## Porsche 914-2	26.0	4	120.3	91	4.43	2.140	16.70	0	1	5	2
## Lotus Europa	30.4	4	95.1	113	3.77	1.513	16.90	1	1	5	2
## Ford Pantera L	15.8	8	351.0	264	4.22	3.170	14.50	0	1	5	4
## Ferrari Dino	19.7	6	145.0	175	3.62	2.770	15.50	0	1	5	6
## Maserati Bora	15.0	8	301.0	335	3.54	3.570	14.60	0	1	5	8
## Volvo 142E	21.4	4	121.0	109	4.11	2.780	18.60	1	1	4	2

```
ggplot(mtcars, aes(x = wt, y = mpg)) +
  geom_smooth(method = lm, se = FALSE, color = "gray") +
  geom_point(aes(color = wt)) +
  xlab("Weight") +
  ylab("Miles per Gallon") +
  scale_color_gradient(low = "forestgreen", high = "black") +
  theme_classic()
```

## 'geom\_smooth()' using formula = 'y ~ x'



```
lm1 <- lm(mpg~wt, data =mtcars)
#y~x (dependent~independent) (criterion~predictor)
# use this for the linear model equation and this is a correlation. You have to use a continuous y and
summary(lm1)
```

```
##
## Call:
## lm(formula = mpg ~ wt, data = mtcars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.5432 -2.3647 -0.1252  1.4096  6.8727
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  37.2851     1.8776   19.858 < 2e-16 ***
## wt          -5.3445     0.5591   -9.559 1.29e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.046 on 30 degrees of freedom
## Multiple R-squared:  0.7528, Adjusted R-squared:  0.7446
## F-statistic: 91.38 on 1 and 30 DF,  p-value: 1.294e-10
```

```
summary(lm(mpg~wt, data = mtcars))
```

```
##
## Call:
## lm(formula = mpg ~ wt, data = mtcars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.5432 -2.3647 -0.1252  1.4096  6.8727
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  37.2851     1.8776   19.858 < 2e-16 ***
## wt          -5.3445     0.5591   -9.559 1.29e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.046 on 30 degrees of freedom
## Multiple R-squared:  0.7528, Adjusted R-squared:  0.7446
## F-statistic: 91.38 on 1 and 30 DF,  p-value: 1.294e-10
```

```
anova(lm(mpg~wt, data = mtcars))
```

```
## Analysis of Variance Table
##
## Response: mpg
##           Df Sum Sq Mean Sq F value    Pr(>F)
```

```
## wt          1 847.73  847.73  91.375 1.294e-10 ***
## Residuals 30 278.32    9.28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

*# essentially the linear model, regression, and ANOVA are going to give you all the same values*

```
cor.test(mtcars$wt, mtcars$mpg) #you have to spell out the variables that you want
```

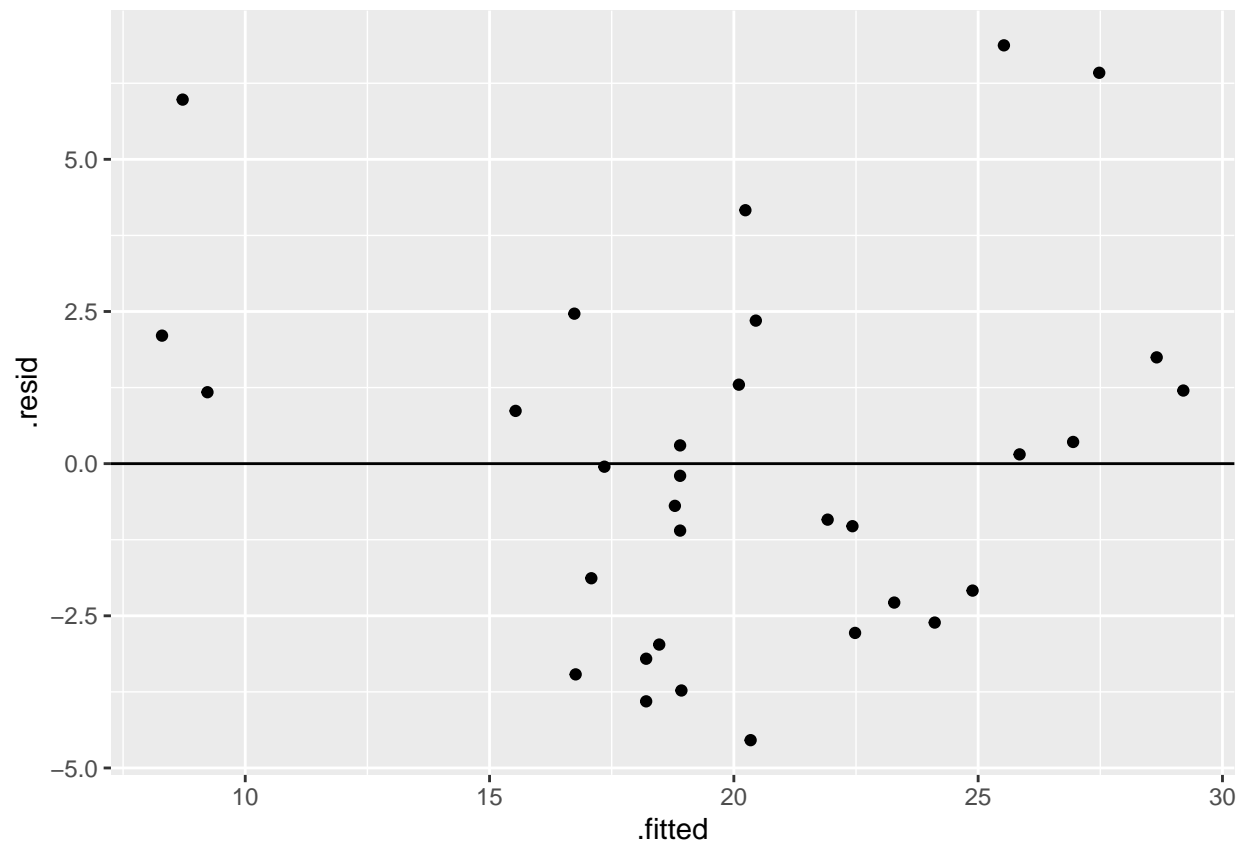
```
##
## Pearson's product-moment correlation
##
## data:  mtcars$wt and mtcars$mpg
## t = -9.559, df = 30, p-value = 1.294e-10
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.9338264 -0.7440872
## sample estimates:
##          cor
## -0.8676594
```

Assumptions

```
#Assumptions for regression, linear model, ANOVA:
# y is continuous
# normal distribution
#linear relationship
# homoskedasticity
# sigma is consistent
# independent samples
```

```
#how we can get residuals:
model <- lm(mpg~wt, data = mtcars)
```

```
ggplot(model, aes(y= .resid, x = .fitted)) +
  geom_point()+
  geom_hline(yintercept = 0)
```



Messin around w bullrichness data

```
bull.rich <- read.csv("C:/Users/katie/Downloads/Bull_richness.csv")
library(tidyverse)

bull.rich.subset <- bull.rich %>%
  filter(GrowthStage == "V8" & Treatment == "Conv.")
# this will show that there is a control and a fungicide treated group

t.test(richness ~ Fungicide, data = bull.rich.subset)

##
## Welch Two Sample t-test
##
## data: richness by Fungicide
## t = 4.8759, df = 17.166, p-value = 0.0001384
## alternative hypothesis: true difference in means between group C and group F is not equal to 0
## 95 percent confidence interval:
##  4.067909 10.265425
## sample estimates:
## mean in group C mean in group F
##      11.750000      4.583333

#null is that means would be equal to zero adn w this data, we can see that the means are not zero

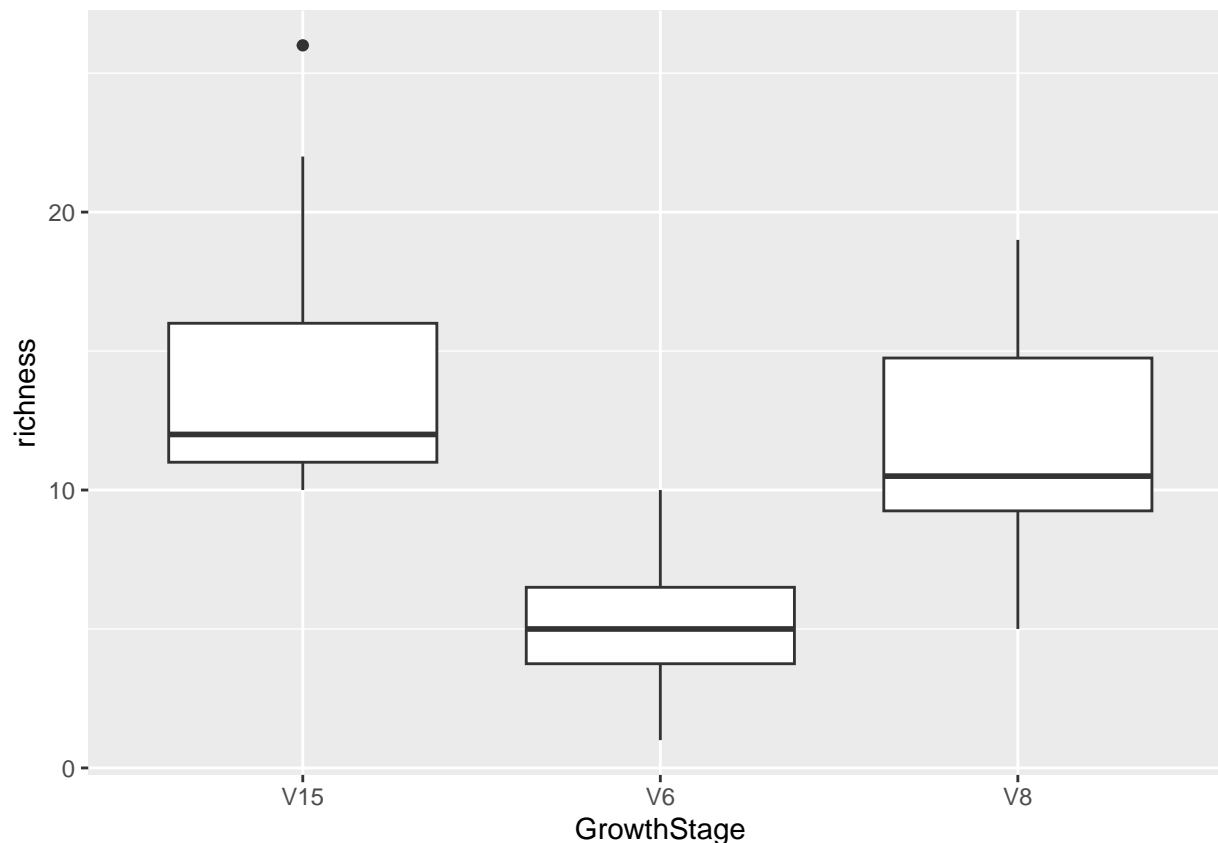
summary(lm(richness~Fungicide, data = bull.rich.subset))
```

```
##
## Call:
## lm(formula = richness ~ Fungicide, data = bull.rich.subset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.7500 -1.7500 -0.6667  2.2500  7.2500
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    11.750      1.039   11.306 1.24e-10 ***
## FungicideF      -7.167      1.470   -4.876 7.12e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.6 on 22 degrees of freedom
## Multiple R-squared:  0.5194, Adjusted R-squared:  0.4975
## F-statistic: 23.77 on 1 and 22 DF,  p-value: 7.118e-05
```

```
anova(lm(richness ~ Fungicide, data = bull.rich.subset))
```

```
## Analysis of Variance Table
##
## Response: richness
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Fungicide  1 308.17  308.167   23.774 7.118e-05 ***
## Residuals 22 285.17   12.962
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
bull.rich.subset2<- bull.rich%>%
  filter(Fungicide == "C" & Treatment == "Conv." & Crop == "Corn")
ggplot(bull.rich.subset2, aes(x = GrowthStage, y = richness)) +
  geom_boxplot()
```



```
summary(lm(richness~GrowthStage, data = bull.rich.subset2)) #make sure you don't have spaces between th
```

```
##
## Call:
## lm(formula = richness ~ GrowthStage, data = bull.rich.subset2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.750 -2.625 -1.000  2.250 11.583
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    14.417     1.208   11.939 1.60e-13 ***
## GrowthStageV6    -9.167     1.708   -5.368 6.23e-06 ***
## GrowthStageV8    -2.667     1.708   -1.562  0.128
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.183 on 33 degrees of freedom
## Multiple R-squared:  0.4803, Adjusted R-squared:  0.4488
## F-statistic: 15.25 on 2 and 33 DF,  p-value: 2.044e-05
```

```
anova((lm(richness~GrowthStage, data = bull.rich.subset2)))
```

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



```
##
## Response: richness
##           Df Sum Sq Mean Sq F value    Pr(>F)
## GrowthStage  2 533.56 266.778  15.247 2.044e-05 ***
## Residuals   33  577.42  17.497
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# this says the model is a good fit
# after the ANOVA you would run a pairwise comparison or a post hoc
```

```
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) #multiple comparisons
sessionInfo()
```

```
## R version 4.4.1 (2024-06-14 ucrt)
## Platform: x86_64-w64-mingw32/x64
## Running under: Windows 11 x64 (build 26100)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United States.utf8
## [2] LC_CTYPE=English_United States.utf8
## [3] LC_MONETARY=English_United States.utf8
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.utf8
##
## time zone: America/Chicago
## tzcode source: internal
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
## other attached packages:
## [1] emmeans_1.11.0  multcomp_1.4-28 TH.data_1.1-3  MASS_7.3-61
## [5] survival_3.6-4  mvtnorm_1.3-3   lme4_1.1-36    Matrix_1.7-0
## [9] lubridate_1.9.3 forcats_1.0.0   stringr_1.5.1  dplyr_1.1.4
## [13] purrr_1.0.2     readr_2.1.5     tidyr_1.3.1    tibble_3.2.1
## [17] ggplot2_3.5.1   tidyverse_2.0.0
##
## loaded via a namespace (and not attached):
## [1] sandwich_3.1-1  utf8_1.2.4      generics_0.1.3  stringi_1.8.4
## [5] lattice_0.22-6  hms_1.1.3       digest_0.6.37   magrittr_2.0.3
```

```
## [9] estimability_1.5.1 evaluate_0.24.0 grid_4.4.1 timechange_0.3.0
## [13] fastmap_1.2.0 mgcv_1.9-1 fansi_1.0.6 scales_1.3.0
## [17] codetools_0.2-20 Rdpack_2.6.2 reformulas_0.4.0 cli_3.6.3
## [21] rlang_1.1.4 rbibutils_2.3 munsell_0.5.1 splines_4.4.1
## [25] withr_3.0.1 yaml_2.3.10 tools_4.4.1 tzdb_0.4.0
## [29] nloptr_2.1.1 minqa_1.2.8 colorspace_2.1-1 boot_1.3-30
## [33] vctr_0.6.5 R6_2.5.1 zoo_1.8-13 lifecycle_1.0.4
## [37] pkgconfig_2.0.3 pillar_1.9.0 gtable_0.3.5 glue_1.7.0
## [41] Rcpp_1.0.14 xfun_0.51 tidyselect_1.2.1 rstudioapi_0.16.0
## [45] knitr_1.50 xtable_1.8-4 farver_2.1.2 htmltools_0.5.8.1
## [49] nlme_3.1-164 labeling_0.4.3 rmarkdown_2.29 compiler_4.4.1
```

```
lm3 <- lm(richness~GrowthStage, data = bull.rich.subset2)
emmeans(lm3, ~GrowthStage)
```

```
## GrowthStage emmean SE df lower.CL upper.CL
## V15 14.42 1.21 33 11.96 16.87
## V6 5.25 1.21 33 2.79 7.71
## V8 11.75 1.21 33 9.29 14.21
##
## Confidence level used: 0.95
```

```
lsmeans <- emmeans(lm3, ~GrowthStage) #lsmeans = least squared means
#cld = compact letter display

results_lsmeans <- cld(lsmeans, alpha = 0.05, details = TRUE)
# outputs which group are different from the other
```

Looking at Interactions

```
bull.rich.subset3 <- bull.rich%>%
  filter(Treatment == "Conv." & Crop == "Corn")

bull.rich.subset3$GrowthStage <- factor(bull.rich.subset3$GrowthStage, levels = c("V6", "V8", "V15"))

# lm.interaction <- lm(richness ~ GrowthStage + Fungicide + GrowthStage:Fungicide))
# this allows you to add the variables that you want to see the interaction in between but you can do t

lm.interaction <- lm(richness ~ GrowthStage*Fungicide, data = bull.rich.subset3)
summary(lm.interaction)

##
## Call:
## lm(formula = richness ~ GrowthStage * Fungicide, data = bull.rich.subset3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.5000 -2.4167 -0.4167  2.0625 11.5833
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      5.2500     1.1029   4.760 1.10e-05 ***
```

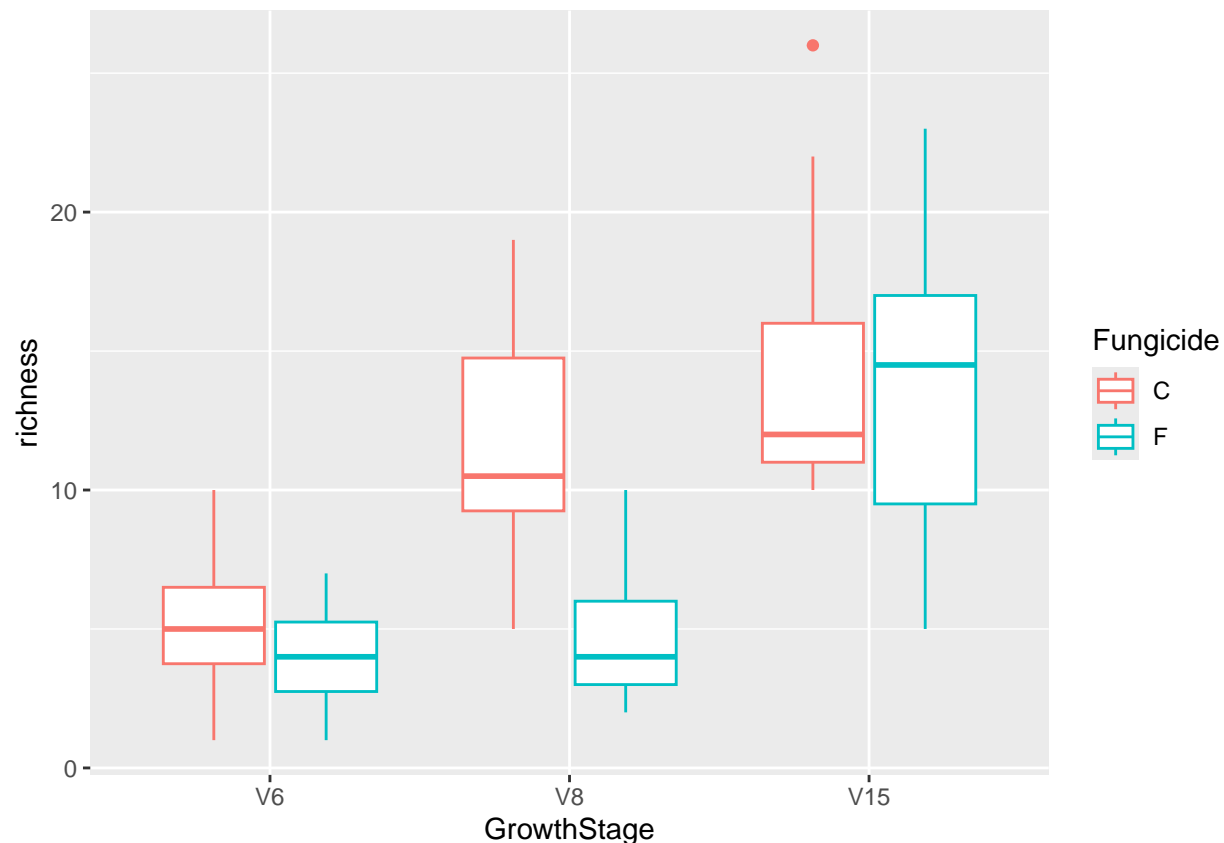
```
## GrowthStageV8          6.5000      1.5597      4.167 9.15e-05 ***
## GrowthStageV15         9.1667      1.5597      5.877 1.51e-07 ***
## FungicideF             -1.2500      1.5597     -0.801 0.42575
## GrowthStageV8:FungicideF -5.9167      2.2057     -2.682 0.00923 **
## GrowthStageV15:FungicideF 0.3333      2.2057      0.151 0.88034
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.82 on 66 degrees of freedom
## Multiple R-squared:  0.5903, Adjusted R-squared:  0.5593
## F-statistic: 19.02 on 5 and 66 DF,  p-value: 1.144e-11
```

```
anova(lm.interaction)
```

```
## Analysis of Variance Table
##
## Response: richness
##
##           Df Sum Sq Mean Sq F value    Pr(>F)
## GrowthStage      2 1065.58   532.79  36.5027 2.113e-11 ***
## Fungicide         1  174.22   174.22  11.9363 0.0009668 ***
## GrowthStage:Fungicide 2  148.36    74.18   5.0823 0.0088534 **
## Residuals        66  963.33    14.60
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
lsmeans <- emmeans(lm.interaction, ~Fungicide|GrowthStage)
#we are seeing fungicide WITHIN each growthstage
results_lsmeans <- cld(lsmeans, alpha = 0.05, details = TRUE)
#fungicide didnt't have an effect until v8

ggplot(bull.rich.subset3, aes(x = GrowthStage, y = richness, color = Fungicide))+
  geom_boxplot()
```



Mixed effects models

*# these have a fixed effect and a random effect. This means random effects impacts the variation in y a*

```
library(lme4)
```

```
# lm.interaction2 <- lmer(richness ~ GrowthStage*Fungicide, data = bull.rich.subset3)
```

*# if you just ran the code above, it would give you an error saying that you need to specify a random e*

```
lm.interaction2 <- lmer(richness ~ GrowthStage*Fungicide + (1|Rep), data = bull.rich.subset3)
summary(lm.interaction2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: richness ~ GrowthStage * Fungicide + (1 | Rep)
## Data: bull.rich.subset3
##
## REML criterion at convergence: 378.3
##
## Scaled residuals:
##    Min      1Q  Median      3Q     Max
## -2.4664 -0.5966 -0.1788  0.6257  2.9101
##
## Random effects:
## Groups Name Variance Std.Dev.
## Rep (Intercept) 0.7855 0.8863
## Residual 13.9533 3.7354
## Number of obs: 72, groups: Rep, 4
```

```
##
## Fixed effects:
##               Estimate Std. Error t value
## (Intercept)      5.2500     1.1658   4.503
## GrowthStageV8      6.5000     1.5250   4.262
## GrowthStageV15     9.1667     1.5250   6.011
## FungicideF       -1.2500     1.5250  -0.820
## GrowthStageV8:FungicideF -5.9167     2.1566  -2.743
## GrowthStageV15:FungicideF  0.3333     2.1566   0.155
##
## Correlation of Fixed Effects:
##      (Intr) GrwSV8 GrSV15 FngcdF GSV8:F
## GrowthStgV8 -0.654
## GrwthStgV15 -0.654  0.500
## FungicideF  -0.654  0.500  0.500
## GrwthSV8:FF  0.462 -0.707 -0.354 -0.707
## GrwtSV15:FF  0.462 -0.354 -0.707 -0.707  0.500
```

```
summary(lm.interaction)
```

```
##
## Call:
## lm(formula = richness ~ GrowthStage * Fungicide, data = bull.rich.subset3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.5000 -2.4167 -0.4167  2.0625 11.5833
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      5.2500     1.1029   4.760 1.10e-05 ***
## GrowthStageV8      6.5000     1.5597   4.167 9.15e-05 ***
## GrowthStageV15     9.1667     1.5597   5.877 1.51e-07 ***
## FungicideF       -1.2500     1.5597  -0.801  0.42575
## GrowthStageV8:FungicideF -5.9167     2.2057  -2.682  0.00923 **
## GrowthStageV15:FungicideF  0.3333     2.2057   0.151  0.88034
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.82 on 66 degrees of freedom
## Multiple R-squared:  0.5903, Adjusted R-squared:  0.5593
## F-statistic: 19.02 on 5 and 66 DF,  p-value: 1.144e-11
```

*# when the std. error goes down, that means you are better able to predict the means of the betas in the*

```
lsmeans <- emmeans(lm.interaction2, ~Fungicide|GrowthStage)
results_lsmeans <- cld(lsmeans, alpha = 0.05, details = TRUE)
```

*# we are better able to detect differences in the linear models*