LauraTraderscode_testedonPSME—seedlings.R

edeegan

2023-11-09

```r
# Add packages here
library(knitr)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.2     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
## -- 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
```

```r
library(ggplot2)
library(car)
```

```
## Loading required package: carData
##
## Attaching package: 'car'
##
## The following object is masked from 'package:dplyr':
##
##     recode
##
## The following object is masked from 'package:purrr':
##
##     some
```

```r
library(dplyr)
library(car)
library(broom)
library(emmeans)
library(lmerTest)
```

```
## Loading required package: lme4
## Loading required package: Matrix
##
## Attaching package: 'Matrix'
```

```
##
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
##
##
## Attaching package: 'lmerTest'
##
## The following object is masked from 'package:lme4':
##
##     lmer
##
## The following object is masked from 'package:stats':
##
##     step
```

```
library(lme4)
library(pbkrtest)

seedlings <- read_csv("C:/Users/edeegan/OneDrive - DOI/Fire_project/Fire_project/PMSE_data/PSME_Trees -
```

```
## New names:
## Rows: 208 Columns: 25
## -- Column specification
## -------------------------------------------------------- Delimiter: "," chr
## (10): MacroPlot Name, Monitoring Status, Date, Status, AgeCl, UV1, Field... dbl
## (5): Index, SizeClHt, Count, SubFrac, MicroPlotSize lgl (10): AvgDia,
## AvgCrwnRto, Comment, UV2, UV3, UV1Desc, UV2Desc, UV3Desc, ...
## 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.
## * `` -> `...25`
```

```
seedlings<-as_tibble(seedlings)
seedlings=rename(seedlings, plot=`MacroPlot Name`)

seedlings=seedlings %>% separate(Date, c("month", "day", "year"), "/")
seedlings=seedlings %>% separate(year, c("year"), " ")
```

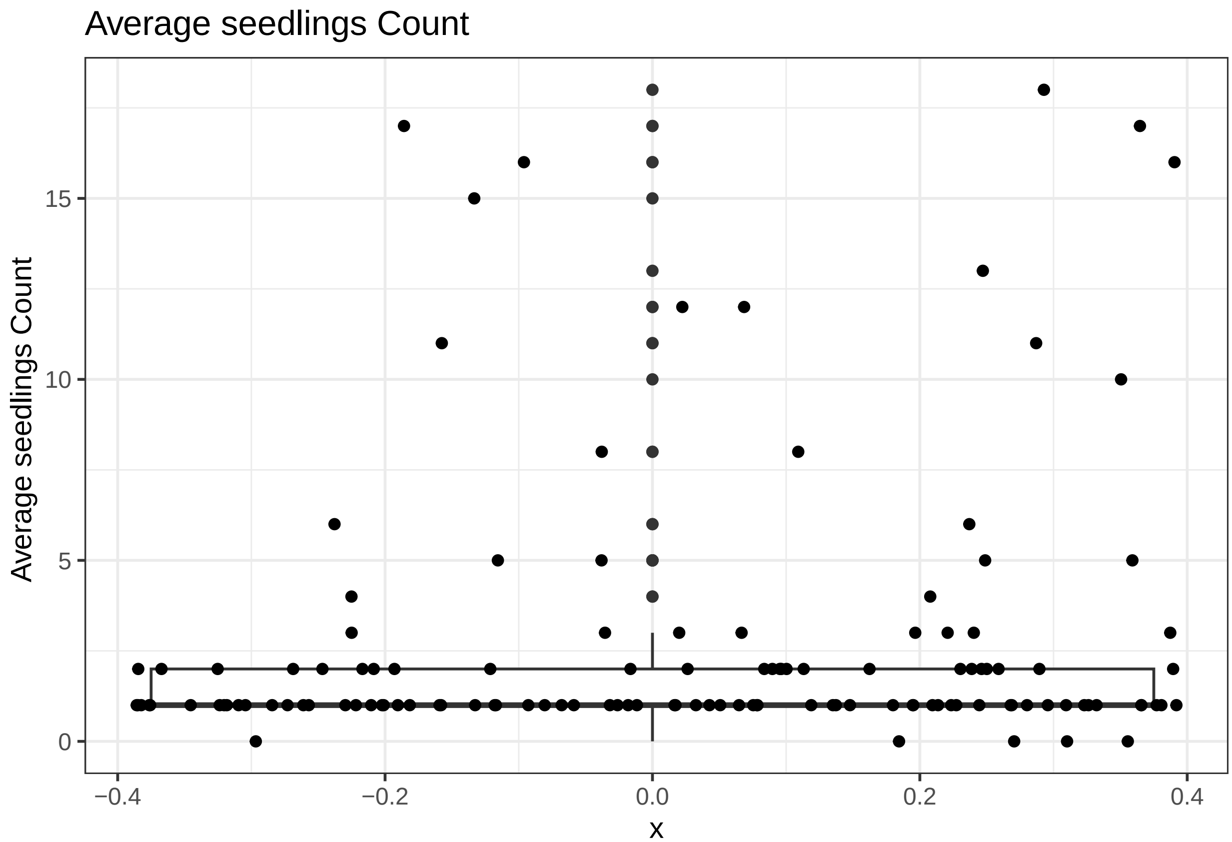
```
## Warning: Expected 1 pieces. Additional pieces discarded in 208 rows [1, 2, 3, 4, 5, 6,
## 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].
```

```
#Fill in the missing plot names, year, and/or species and insert a "zero" for Count

seedlings_fill <- seedlings %>% drop_na(Count)

#Convert "plot" and "year" to factors with mutate. In R, factors are variables that take on a limited n
#of different values; often referred to as categorical variables (e.g. not continuous). Categorical var
#enter into statistical models differently than continuous variables, so storing data as factors ensure
#modeling functions will treat such data correctly.

seedlings_factor1 <- seedlings_fill %>%
  mutate(plot=factor(plot))
```

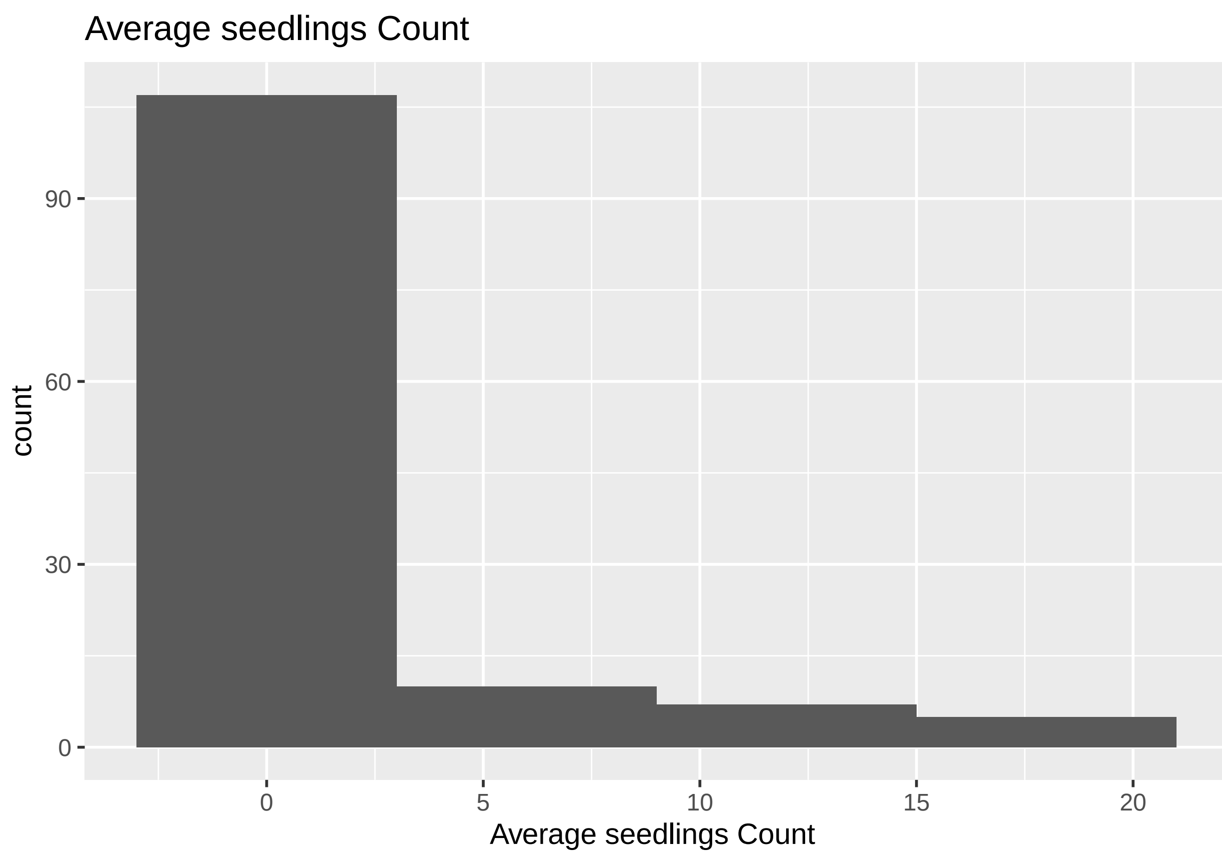
```r
seedlings_factor2 <- seedlings_factor1 %>%
  mutate(year=factor(year))


#Calculate summary statistics. Create a table with sample size, minimum value, maximum value, mean,
#standard deviation, and standard error. Use kable for an improved table.

seedlings_factor2_summ <- seedlings_factor2 %>%
  group_by(`Year`=year) %>%
  summarise(n=n(),
            `Min.`=min(Count),
            `Max.`=max(Count),
            `Mean`=mean(Count),
            `SD`=sd(Count),
            `SE`=sd(Count)/sqrt(n))

kable(seedlings_factor2_summ, booktabs=T, digits=3)
```

Year	n	Min.	Max.	Mean	SD	SE
1990	6	1	15	4.833	5.307	2.167
1991	4	1	11	4.000	4.761	2.380
1992	21	1	5	1.619	1.203	0.263
1997	20	1	4	1.400	0.821	0.184
2001	27	1	4	1.481	0.802	0.154
2004	8	0	2	1.000	0.756	0.267
2023	43	0	18	4.930	5.625	0.858

```r
#Create a boxplot. A boxplot displays the distribution of a dataset based on its five number summary of
#points: minimum, 1st quartile (25th percentile), median, third quartile (75th percentile), and maximum
#boxplot can be used to show the symmetry, skew, variance, and outliers of a dataset.

ggplot() +
  theme_bw() +
  geom_boxplot(aes(y=Count),
               data=seedlings_factor2) +
  geom_jitter(aes(y=Count, x=0),
              height=0,
              data=seedlings_factor2) +
  ylab("Average seedlings Count ") +
  ggtitle("Average seedlings Count ")
```




```
#Create a histogram. A histogram is an approximate representation of the distribution of continuous dat
#Histograms can be used to identify patterns in data, such as the shape of the distribution (e.g. norma
# distributed, skewed), the spread of the data, and outliers. The height of each bar represents the fre
#(count) of data points within the corresponding bin (x-axis).


ggplot() +
  geom_histogram(aes(x=Count),
                 bins=4,
                 data=seedlings_factor2) +
  xlab("Average seedlings Count ") +
  ggtitle("Average seedlings Count ")
```




```
#Fit a mixed model for a repeated measures design using lmer and indicate "plot" as a random effect. Ca
#summary on the model. We have repeated measurements on individual plots (experimental units) and those
#measurements will be correlated (not independent). A mixed model accounts for the correlated responses
#We indicate "plot" as a random effect to account for the correlation between measurements that arise f
#the same plot. "Plot" has random variation and is not of primary interest in this analysis.

seedlings_lmer <- lmer(Count~year + (1|plot),
                    data=seedlings_factor2)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
summary(seedlings_lmer)
```

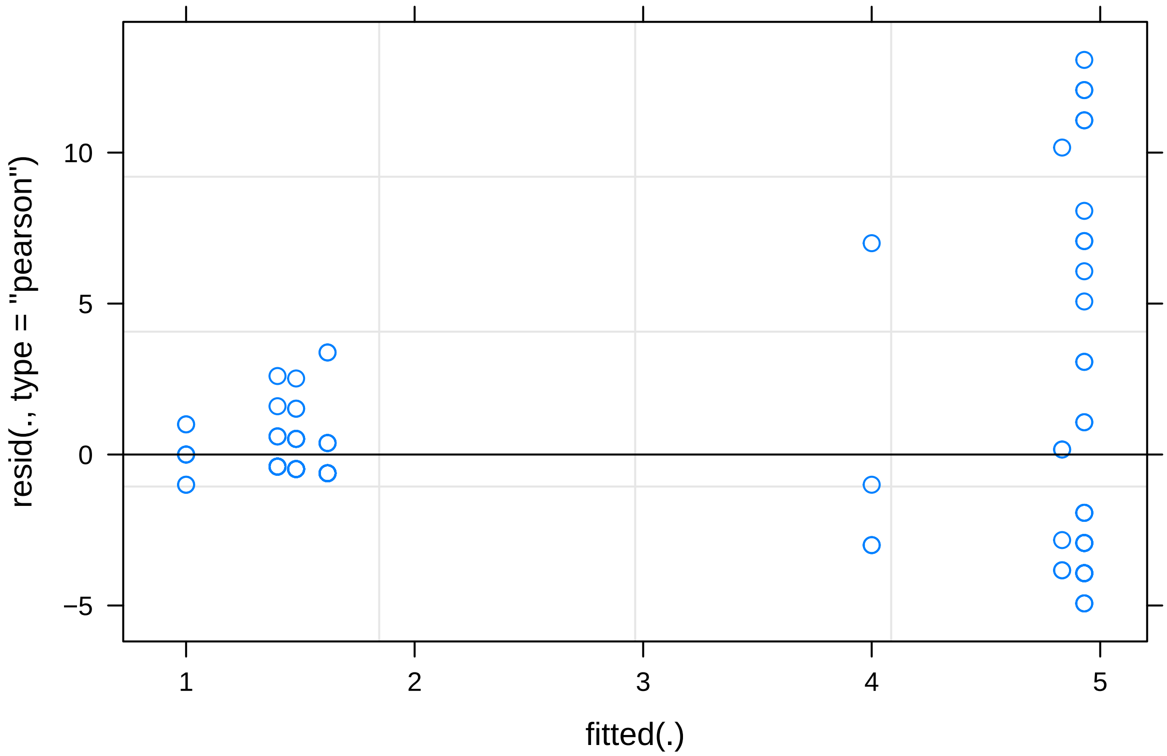
```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: Count ~ year + (1 | plot)
##    Data: seedlings_factor2
##
## REML criterion at convergence: 678.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.3614 -0.5330 -0.1329  0.1432  3.6089
##
## Random effects:
```

```
##  Groups   Name        Variance Std.Dev.
##  plot     (Intercept)  0.00    0.000
##  Residual             13.12    3.622
## Number of obs: 129, groups:  plot, 10
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)   4.8333     1.4785 122.0000   3.269   0.0014 **
## year1991     -0.8333     2.3377 122.0000  -0.356   0.7221
## year1992     -3.2143     1.6765 122.0000  -1.917   0.0575 .
## year1997     -3.4333     1.6858 122.0000  -2.037   0.0438 *
## year2001     -3.3519     1.6345 122.0000  -2.051   0.0424 *
## year2004     -3.8333     1.9559 122.0000  -1.960   0.0523 .
## year2023      0.0969     1.5783 122.0000   0.061   0.9511
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) yr1991 yr1992 yr1997 yr2001 yr2004
## year1991 -0.632
## year1992 -0.882  0.558
## year1997 -0.877  0.555  0.773
## year2001 -0.905  0.572  0.798  0.793
## year2004 -0.756  0.478  0.667  0.663  0.684
## year2023 -0.937  0.592  0.826  0.822  0.847  0.708
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

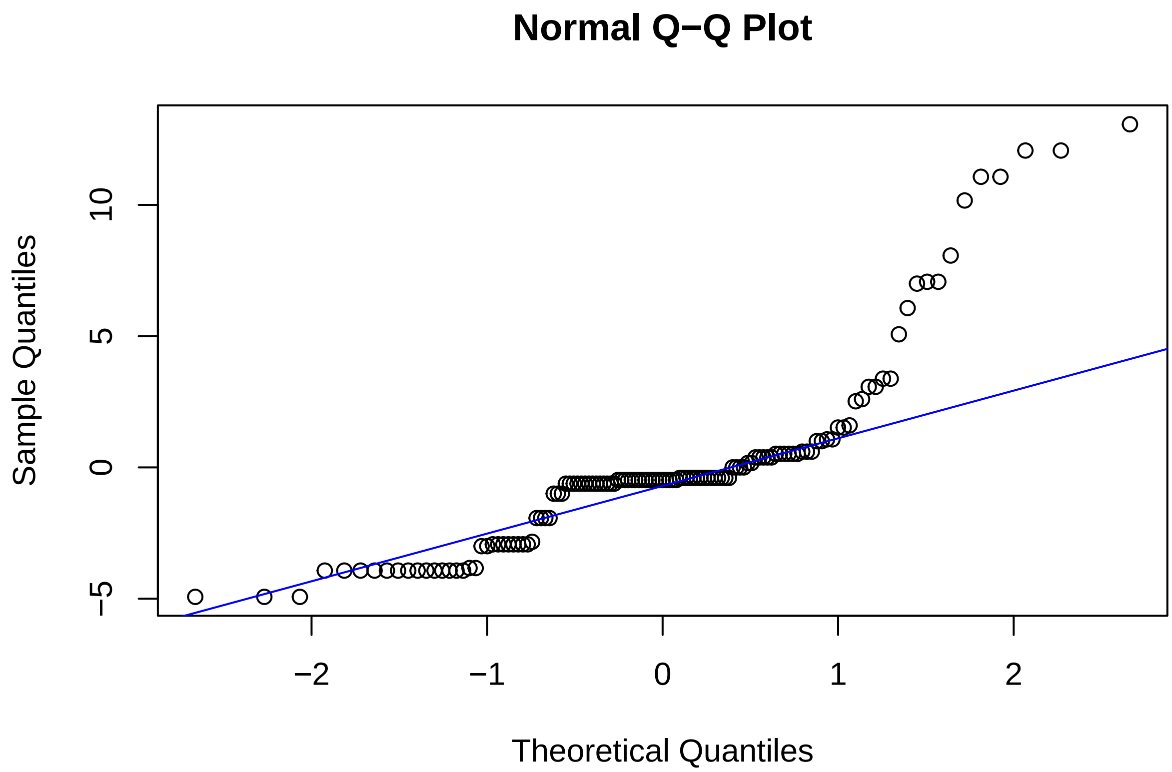
```
#Call anova (Analysis of Variance) on the model. Question: "Is there a difference in mean seedlings Cou
#between years?"
#Null hypothesis (HO): There is no difference in mean seedlings Count between years.
#Alternative hypothesis (HA): There is a difference in mean seedlings Count between years.
#A small p-value (less than alpha 0.05) will reject the null hypothesis that there is no difference in 
#seedlings Count between years. You can then conclude that there is evidence of a difference in mean se
#Count between years.

anova(seedlings_lmer, ddf="Kenward-Roger")
```

```
## Type III Analysis of Variance Table with Kenward-Roger's method
##      Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)
## year 345.67  57.611     6 103.68  4.3739 0.0005604 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Create a plot of standardized residuals vs. fitted values for the model to assess the assumption of co
#variance. Fitted values are the values predicted by the model. Residuals are the differences between t
#observed values (data) and the corresponding fitted values. This plot displays the fitted values of th
#along the x-axis and the residuals of the fitted values along the y-axis. If the spread of the residua
#equal at each level of the fitted values, the constant variance assumption is met. The residuals shoul
#scattered randomly about zero, with no obvious pattern emerging

plot(seedlings_lmer)
```




```
#Create a Quantile-Quantile (QQ) plot to assess normality of the residuals (normal distribution). A QQ
#compares two probability distributions by plotting their quantiles against each other: the quantiles o
#sample data versus the theoretical quantile values from a normal distribution (or what we would expect
# a normal distribution). Data points should fall on a fairly straight line to indicate linearity. If d
#deviate largely from a straight line, it suggests that the two data sets do not have the same distribu

qqnorm(residuals(seedlings_lmer))
qqline(residuals(seedlings_lmer), col = "blue")
```




```
#Create an emmeans object and conduct Tukey-adjusted pairwise comparisons ("contrasts") between years.
#These contrasts provide estimates of the pairwise differences in average live seedlings Count between y
#Include confidence intervals and p-values that have been adjusted for multiple comparisons.
#Multiple comparison problem: each time you run a hypothesis test, there is a small chance you will obta
#a "false" significant result (you will reject the null hypothesis when it is actually true, also calle
#Error Rate). If you run multiple tests, the number of "false positives" increases with each test, so t
#this Type I Error Rate, the p-values can be adjusted (Tukey is one method and used here) to be more
#conservative (less false positives).
#Call tidy to create an emmeans table. Copy tidy code and include kable for an improved table. Use the
#table column headers in the first tidy table to create the headers in the second (and final) tidy tabl

seedlings_lmer_emm <- emmeans(seedlings_lmer, ~year)
contrast(seedlings_lmer_emm, "pairwise" , infer=TRUE, conf.int=TRUE)
```

```
##  contrast              estimate    SE    df lower.CL upper.CL t.ratio p.value
##  year1990 - year1991     0.8333 2.489  73.2    -6.71    8.379   0.335  0.9999
##  year1990 - year1992     3.2143 1.913  42.1    -2.71    9.134   1.681  0.6322
##  year1990 - year1997     3.4333 1.823  83.9    -2.08    8.942   1.883  0.4972
##  year1990 - year2001     3.3519 1.797  69.4    -2.10    8.808   1.865  0.5097
##  year1990 - year2004     3.8333 2.039 115.9    -2.28    9.951   1.880  0.4975
##  year1990 - year2023    -0.0969 1.743  66.7    -5.40    5.202  -0.056  1.0000
##  year1991 - year1992     2.3810 2.002 103.1    -3.64    8.403   1.189  0.8968
##  year1991 - year1997     2.6000 2.011 103.8    -3.45    8.646   1.293  0.8537
##  year1991 - year2001     2.5185 1.962 111.9    -3.37    8.409   1.284  0.8580
##  year1991 - year2004     3.0000 2.247 116.3    -3.74    9.744   1.335  0.8341
```

```
##  year1991 - year2023  -0.9302 1.920 103.7    -6.70    4.843  -0.484  0.9990
##  year1992 - year1997   0.2190 1.156 122.0    -3.25    3.686   0.189  1.0000
##  year1992 - year2001   0.1376 1.075 120.8    -3.09    3.362   0.128  1.0000
##  year1992 - year2004   0.6190 1.555 116.8    -4.05    5.285   0.398  0.9997
##  year1992 - year2023  -3.3112 0.994 118.4    -6.29   -0.329  -3.330  0.0193
##  year1997 - year2001  -0.0815 1.077 121.0    -3.31    3.148  -0.076  1.0000
##  year1997 - year2004   0.4000 1.530 121.5    -4.19    4.989   0.261  1.0000
##  year1997 - year2023  -3.5302 0.999 121.9    -6.53   -0.535  -3.534  0.0101
##  year2001 - year2004   0.4815 1.472 120.0    -3.93    4.896   0.327  0.9999
##  year2001 - year2023  -3.4488 0.907 121.4    -6.17   -0.729  -3.803  0.0041
##  year2004 - year2023  -3.9302 1.413 121.7    -8.17    0.306  -2.782  0.0878
##
## Degrees-of-freedom method: kenward-roger
## Confidence level used: 0.95
## Conf-level adjustment: tukey method for comparing a family of 7 estimates
## P value adjustment: tukey method for comparing a family of 7 estimates
```

```
tidy(contrast(seedlings_lmer_emm, "pairwise", infer=TRUE), conf.int=TRUE)
```

```
## # A tibble: 21 x 10
##    term  contrast          null.value estimate std.error    df conf.low conf.high
##    <chr> <chr>                  <dbl>    <dbl>     <dbl> <dbl>    <dbl>     <dbl>
##  1 year  year1990 - year~           0   0.833       2.49  73.2    -6.71      8.38
##  2 year  year1990 - year~           0   3.21        1.91  42.1    -2.71      9.13
##  3 year  year1990 - year~           0   3.43        1.82  83.9    -2.08      8.94
##  4 year  year1990 - year~           0   3.35        1.80  69.4    -2.10      8.81
##  5 year  year1990 - year~           0   3.83        2.04 116.     -2.28      9.95
##  6 year  year1990 - year~           0  -0.0969      1.74  66.7    -5.40      5.20
##  7 year  year1991 - year~           0   2.38        2.00 103.     -3.64      8.40
##  8 year  year1991 - year~           0   2.6         2.01 104.     -3.45      8.65
##  9 year  year1991 - year~           0   2.52        1.96 112.     -3.37      8.41
## 10 year  year1991 - year~           0   3           2.25 116.     -3.74      9.74
## # i 11 more rows
## # i 2 more variables: statistic <dbl>, adj.p.value <dbl>
```

```
tidy(contrast(seedlings_lmer_emm, "pairwise", infer=TRUE), conf.int=TRUE) %>%
  select(Contrast=contrast, Estimate=estimate,
         SE=std.error, df=df, `CI-low`=conf.low, `CI-high`=conf.high,
         `P-value`='adj.p.value') %>%
  kable(digits=c(2, 2, 2, 2, 2, 2, 6), booktabs=T)
```

Contrast	Estimate	SE	df	CI-low	CI-high	P-value
year1990 - year1991	0.83	2.49	73.18	-6.71	8.38	0.999879
year1990 - year1992	3.21	1.91	42.10	-2.71	9.13	0.632194
year1990 - year1997	3.43	1.82	83.90	-2.08	8.94	0.497236
year1990 - year2001	3.35	1.80	69.36	-2.10	8.81	0.509732
year1990 - year2004	3.83	2.04	115.91	-2.28	9.95	0.497462
year1990 - year2023	-0.10	1.74	66.70	-5.40	5.20	1.000000
year1991 - year1992	2.38	2.00	103.10	-3.64	8.40	0.896780
year1991 - year1997	2.60	2.01	103.80	-3.45	8.65	0.853686
year1991 - year2001	2.52	1.96	111.93	-3.37	8.41	0.857959

Contrast	Estimate	SE	df	CI-low	CI-high	P-value
year1991 - year2004	3.00	2.25	116.26	-3.74	9.74	0.834140
year1991 - year2023	-0.93	1.92	103.68	-6.70	4.84	0.999003
year1992 - year1997	0.22	1.16	122.00	-3.25	3.69	0.999996
year1992 - year2001	0.14	1.08	120.85	-3.09	3.36	1.000000
year1992 - year2004	0.62	1.56	116.78	-4.05	5.28	0.999679
year1992 - year2023	-3.31	0.99	118.43	-6.29	-0.33	0.019312
year1997 - year2001	-0.08	1.08	121.04	-3.31	3.15	1.000000
year1997 - year2004	0.40	1.53	121.48	-4.19	4.99	0.999973
year1997 - year2023	-3.53	1.00	121.90	-6.53	-0.53	0.010142
year2001 - year2004	0.48	1.47	120.02	-3.93	4.90	0.999898
year2001 - year2023	-3.45	0.91	121.38	-6.17	-0.73	0.004133
year2004 - year2023	-3.93	1.41	121.67	-8.17	0.31	0.087760

```r
#Eva's timeline graph
seedlings_factor2_summ %>%
  ggplot(aes(x=Year, y=Mean, size=Mean))+geom_point()+theme_classic()+ylab("Seedlings Count")
```