KG help

Gina Nichols

4/13/2020

Install the R package with the data

```
devtools::install_github("vanichols/PFIweeds2020")
```

```
## Skipping install of 'PFIweeds2020' from a github remote, the SHA1 (384d6f2e) has not changed since la
##   Use `force = TRUE` to force installation
```

```
library(PFIweeds2020)
```

Other libraries

```
library(tidyverse)
```

```
## -- Attaching packages ------------------------------------------------------------ tidyver
```

```
## v ggplot2 3.3.0      v purrr   0.3.3
## v tibble  3.0.0      v dplyr   0.8.5
## v tidyr   1.0.2      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.4.0
```

```
## Warning: package 'ggplot2' was built under R version 3.6.3
```

```
## Warning: package 'tibble' was built under R version 3.6.3
```

```
## Warning: package 'dplyr' was built under R version 3.6.3
```

```
## -- Conflicts --------------------------------------------------------------- tidyverse_con
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Tranformed Data Question

We wanted to know if cover-cropping (*cc_trt*) affects the number of seeds in the weed seedbank (*totseeds_m2*).
We measured this at 4 sites (*site_sys*).

```
dat <- pfifun_sum_byeu(pfi_ghobsraw) %>%
  unite(site_name, sys_trt, col = "site_sys", remove = F)

dat
```

```
## # A tibble: 46 x 9
## # Groups:   site_name, field, sys_trt, cc_trt, rep [46]
##    site_sys   site_name field sys_trt cc_trt  rep blockID totseeds totseeds_m2
##    <chr>      <chr>     <chr> <chr>   <chr>  <dbl> <chr>      <dbl>       <dbl>
## 1 Boyd_grain Boyd      B42   grain   no         1 B42_1         34        663.
## 2 Boyd_grain Boyd      B42   grain   no         2 B42_2         12        234.
## 3 Boyd_grain Boyd      B42   grain   no         3 B42_3          9        175.
## 4 Boyd_grain Boyd      B42   grain   no         4 B42_4          7        136.
```

```
##  5 Boyd_grain Boyd      B42   grain  no       5 B42_5      17      331.
##  6 Boyd_grain Boyd      B42   grain  rye      1 B42_1     116     2261.
##  7 Boyd_grain Boyd      B42   grain  rye      2 B42_2      79     1540.
##  8 Boyd_grain Boyd      B42   grain  rye      3 B42_3       6      117.
##  9 Boyd_grain Boyd      B42   grain  rye      4 B42_4      29      565.
## 10 Boyd_grain Boyd      B42   grain  rye      5 B42_5      12      234.
## # ... with 36 more rows
```

Fitting the models

```r
library(lme4) #--for mixed models
```

```
## Loading required package: Matrix

##
## Attaching package: 'Matrix'

## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
```

```r
library(lmerTest) #--to get significances
```

```
##
## Attaching package: 'lmerTest'

## The following object is masked from 'package:lme4':
##
##     lmer

## The following object is masked from 'package:stats':
##
##     step
```

```r
library(broom)
library(emmeans)
```

```
## Welcome to emmeans.
## NOTE -- Important change from versions <= 1.41:
##     Indicator predictors are now treated as 2-level factors by default.
##     To revert to old behavior, use emm_options(cov.keep = character(0))
```

I want the *rye* treatment to appear first, so I renamed some things.

```r
dstat <-
  dat %>%
  mutate(cc_trt2 = recode(cc_trt,          ##--I want rye to appear first alphabetically?
                          no = "none",
                          rye = "aryecc"))
```
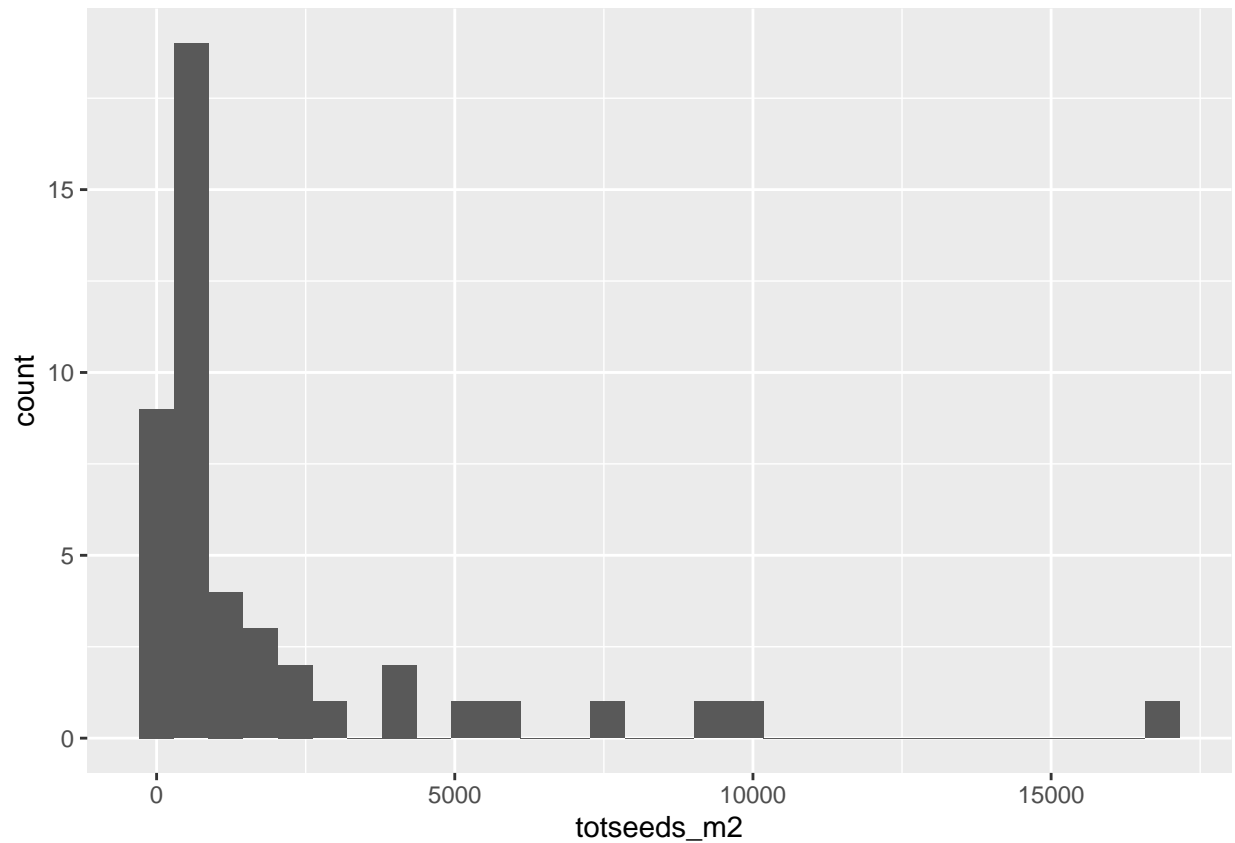
There is a one value that is very large. I need to look into how to assess 'outlier-ness', so just know I filter that value out.

The response variable of interest, totseeds_m2, is skewed:

```r
dat %>%
  ggplot(aes(totseeds_m2)) +
  geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Because of this (I think it was because it was skewed), we fit our statistical models to a log-transformed response varible. Is there a better way to justify this? Looking at the residual plots, I am not sure there is a clear answer.

```r
library(ggResidpanel)

m1 <- lmer(log(totseeds_m2) ~ site_sys * cc_trt2 + (1|blockID), data = filter(dstat, totseeds_m2 < 15000
m2 <- lmer((totseeds_m2) ~ site_sys * cc_trt2 + (1|blockID), data = filter(dstat, totseeds_m2 < 15000))

ggResidpanel::resid_compare(list(m1, m2))
```
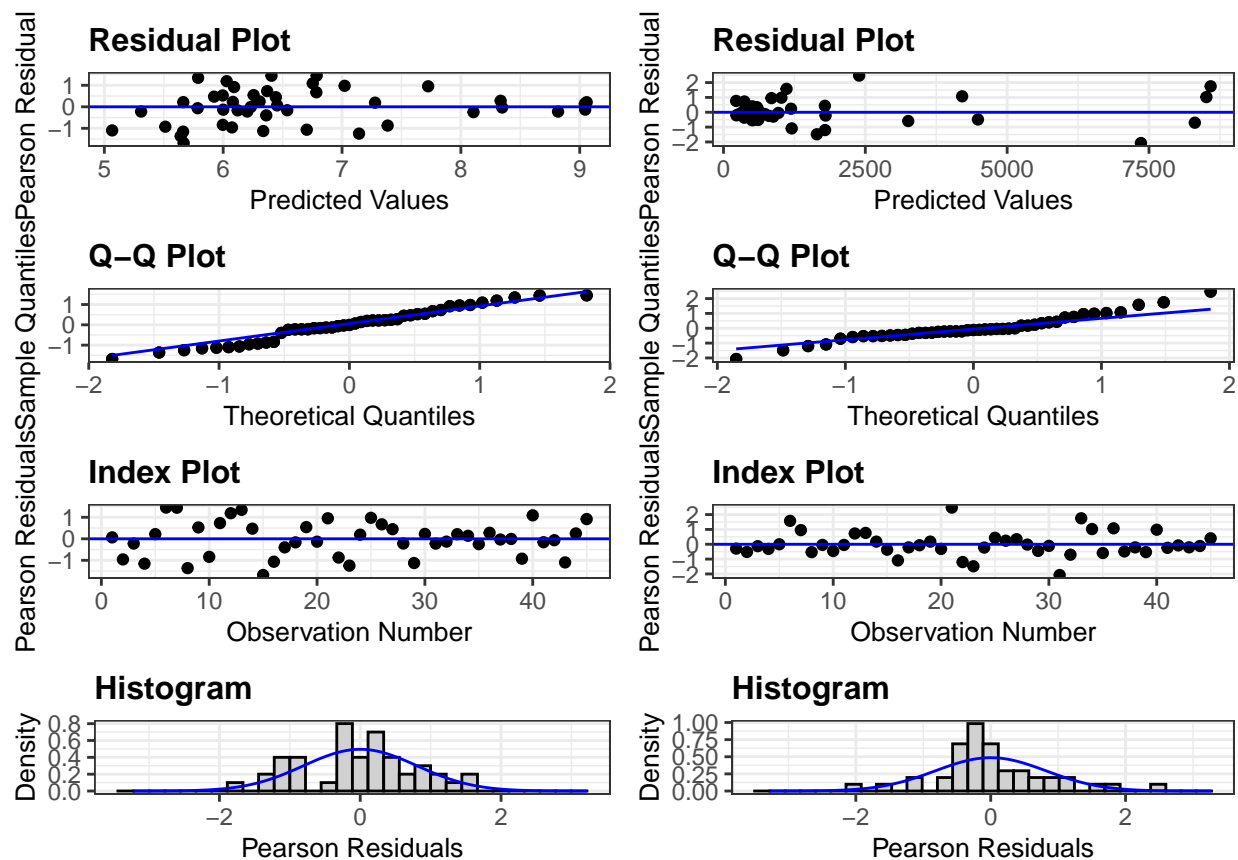
Pretending the log-transformed response variable is the best. . .

I fit my model:

```r
m1 <- lmer(log(totseeds_m2) ~ site_sys * cc_trt2 + (1|blockID), data = filter(dstat, totseeds_m2 < 15000
```

Now I use the **emmeans** package to compare the rye/none treatments in each site_sys, and the **broom** package to make it tidy.

```r
m1em <- (emmeans(m1, pairwise ~ cc_trt2|site_sys, type = "response"))
m1res <- broom::tidy(m1em$contrasts)
m1res
```

```
## # A tibble: 4 x 7
##   contrast      site_sys     ratio std.error    df statistic p.value
##   <fct>         <fct>        <dbl>     <dbl> <dbl>     <dbl>   <dbl>
## 1 aryecc / none Boyd_grain   1.40      0.403  22.4     1.15    0.261
## 2 aryecc / none Boyd_silage  0.390     0.160  22.4    -2.30    0.0310
## 3 aryecc / none Funcke_grain 0.491     0.250  24.9    -1.39    0.176
## 4 aryecc / none Stout_grain  0.639     0.292  22.4    -0.979   0.338
```

Question

The ratio this produces - how should I interpret/report this? For example, in the *Boyd_silage* site_sys:

```r
m1res %>%
  filter(site_sys == "Boyd_silage") %>%
  select(contrast, site_sys, ratio, p.value)
```

```
## # A tibble: 1 x 4
##   contrast       site_sys    ratio p.value
##   <fct>          <fct>       <dbl>   <dbl>
## 1 aryecc / none Boyd_silage 0.390  0.0310
```

This means there were more weeds in the *none* treatment compared to the *aryecc* treatment. How much did cover-cropping reduce the weed seedbank? Should I interpret this as a percentage? So rye reduces weed seedbanks by:

```
(1-0.39)*100
```

```
## [1] 61
```

61%?

Alternatively, emmeans offers something that seems to deal with responses that were transformed:

```
m1em_alt1 <- emmeans(m1, pairwise ~ cc_trt2|site_sys, type = "response")  #--ratio, must be default
broom::tidy(m1em_alt1$contrasts)
```

```
## # A tibble: 4 x 7
##   contrast       site_sys     ratio std.error    df statistic p.value
##   <fct>          <fct>        <dbl>     <dbl> <dbl>     <dbl>   <dbl>
## 1 aryecc / none Boyd_grain    1.40     0.403  22.4     1.15    0.261
## 2 aryecc / none Boyd_silage  0.390     0.160  22.4    -2.30    0.0310
## 3 aryecc / none Funcke_grain 0.491     0.250  24.9    -1.39    0.176
## 4 aryecc / none Stout_grain  0.639     0.292  22.4    -0.979   0.338
```

So, this seems to be the exponential of the estimates from above. But... what does this value actually mean?

```
m1em_alt2 <- emmeans(m1, pairwise ~ cc_trt2|site_sys, type = "lp") #--log scale? back-transformed? conf
broom::tidy(m1em_alt2$contrasts)
```

```
## # A tibble: 4 x 8
##   level1 level2 site_sys     estimate std.error    df statistic p.value
##   <chr>  <chr>  <fct>           <dbl>     <dbl> <dbl>     <dbl>   <dbl>
## 1 aryecc none   Boyd_grain      0.333     0.289  22.4     1.15    0.261
## 2 aryecc none   Boyd_silage    -0.941     0.409  22.4    -2.30    0.0310
## 3 aryecc none   Funcke_grain   -0.711     0.510  24.9    -1.39    0.176
## 4 aryecc none   Stout_grain    -0.447     0.457  22.4    -0.979   0.338
```