ST431 Design of Experiments

ANCOVA for wordle

Jason A. Osborne

N. C. State Univ.

```
library(tidyverse)
library(emmeans)
library(multcomp)

#read data
read_delim("wordle-oct23-2023b.csv") -> wordle
wordle %>% mutate(result=as.numeric(result)) -> wordle

# Identify initial guesses with frequency > 6
wordle %>% filter(!is.na(botresult)) %>% group_by(guess1) %>%

summarize(mean(result),mean(botresult),count=n()) %>%
filter(count>6) %>% arrange(-count) %>% select(1) -> topwords

wordle %>% filter(!is.na(botresult),guess1 %in% pull(topwords)) ->
wordle.top


#...which are ...
wordle.top %>% select(guess1) %>% table
wordle.top %>% select(guess1) %>% table %>% prop.table ->
wordle.tab1
```

```
wordle.tab1[order(wordle.tab1,decreasing=TRUE)]

#fit model with two covariates
lm(result ~ guess1 + avg + botresult,data=wordle.top) ->
wordle.fit3

#anova table
wordle.fit3 %>% anova

#summary
wordle.fit3 %>% summary

#means adjusted to avgs of botresult and (nyt)avg
wordle.fit3 %>% emmeans(~guess1) %>% as.data.frame %>%
arrange(emmean)

#pairwise diff between "aisle" and "least"
diffLeast_Aisle <- matrix(c(0,0,-1,0,0,0,1,0,0,0,0,0),1)
glht(wordle.fit3,diffLeast_Aisle) -> diff3.out
diff3.out %>% summary
```

```
#raw means:
wordle.top %>% group_by(guess1) %>%
summarize(mymean=mean(result), std=sd(result),
  mean(avg),mean(botresult),n=n()) %>%
arrange(mymean,decreasing=TRUE)
```

```
R version 4.3.2 (2023-10-31) -- "Eye Holes"
Copyright (C) 2023 The R Foundation for Statistical Computing

> #...which are ...
> wordle.top %>% select(guess1) %>% table
guess1
adieu adios aisle crane lance later least slate stale stare storm
   22     9    19     7     7    14    13    72    16    14     7
> wordle.top %>% select(guess1) %>% table %>% prop.table ->
wordle.tab1
> wordle.tab1[order(wordle.tab1,decreasing=TRUE)]
guess1
slate adieu aisle stale later stare least adios crane lance storm
0.360 0.110 0.095 0.080 0.070 0.070 0.065 0.045 0.035 0.035 0.035
>
> #anova table
> wordle.fit3 %>% anova
Analysis of Variance Table

Response: result
            Df  Sum Sq Mean Sq F value    Pr(>F)
guess1      10   7.850  0.7850  1.2814 0.2435556
avg          1  29.226 29.2262 47.7071 7.608e-11 ***
botresult    1   7.801  7.8008 12.7335 0.0004569 ***
Residuals  186 113.947  0.6126
```

```
> #summary
> wordle.fit3 %>% summary

Call:
lm(formula = result ~ guess1 + avg + botresult, data = wordle.top)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.13721    0.56430  -0.243 0.808151
guess1adios  0.32168    0.31211   1.031 0.304047
guess1aisle -0.23515    0.24525  -0.959 0.338901
guess1crane -0.08019    0.33971  -0.236 0.813644
guess1lance -0.04534    0.34505  -0.131 0.895595
guess1later  0.24788    0.26948   0.920 0.358848
guess1least  0.41413    0.27415   1.511 0.132594
guess1slate  0.10440    0.19144   0.545 0.586189
guess1stale -0.13802    0.26159  -0.528 0.598380
guess1stare -0.04321    0.26832  -0.161 0.872230
guess1storm  0.59563    0.33977   1.753 0.081248 .
avg          0.67540    0.15791   4.277 3.02e-05 ***
botresult    0.33680    0.09439   3.568 0.000457 ***
---
Residual standard error: 0.7827 on 186 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.2826,    Adjusted R-squared:  0.2363
F-statistic: 6.105 on 12 and 186 DF,  p-value: 5.452e-09
```

```
>
> #means adjusted to avgs of botresult and (nyt)avg
> wordle.fit3 %>% emmeans(~guess1) %>% as.data.frame %>%
arrange(emmean)
 guess1    emmean          SE  df lower.CL upper.CL
 aisle   3.462490 0.17978596 186 3.107808 3.817171
 stale   3.559612 0.19868063 186 3.167654 3.951569
 crane   3.617443 0.29603453 186 3.033426 4.201460
 lance   3.652293 0.29965571 186 3.061132 4.243453
 stare   3.654421 0.21165000 186 3.236878 4.071964
 adieu   3.697635 0.16746969 186 3.367251 4.028019
 slate   3.802030 0.09298734 186 3.618585 3.985476
 later   3.945511 0.20995668 186 3.531309 4.359714
 adios   4.019312 0.26234722 186 3.501753 4.536870
 least   4.111765 0.21712832 186 3.683414 4.540115
 storm   4.293262 0.29673150 186 3.707870 4.878654
```

```
Confidence level used: 0.95
>
> #pairwise diff between "aisle" and "least"
> diffLeast_Aisle <- matrix(c(0,0,-1,0,0,0,1,0,0,0,0,0,0),1)
> glht(wordle.fit3,diffLeast_Aisle) -> diff3.out
> diff3.out %>% summary

         Simultaneous Tests for General Linear Hypotheses

Fit: lm(formula = result ~ guess1 + avg + botresult,
        data = wordle.top)

Linear Hypotheses:
       Estimate Std. Error t value Pr(>|t|)
1 == 0   0.6493     0.2818   2.304   0.0223 *
---
(Adjusted p values reported -- single-step method)
```

```
> #raw means:
> wordle.top %>% group_by(guess1) %>%
+ summarize(mymean=mean(result), std=sd(result),mean(avg),
+ mean(botresult),n=n()) %>%
+ arrange(mymean,decreasing=TRUE)
# A tibble: 11 by 6
   guess1 mymean   std `mean(avg)` `mean(botresult)`     n
   <chr>   <dbl> <dbl>       <dbl>                <dbl> <int>
 1 aisle    3.42 0.838        3.91                 3.42    19
 2 stale    3.5  1.10         4.03                 3.12    16
 3 crane    3.57 0.787        3.9                  3.43     7
 4 adieu    3.68 0.839        3.91                 3.5     22
 5 stare    3.71 0.994        3.91                 3.71    14
 6 slate    3.82 0.828       NA                    3.47    72
 7 lance    3.86 0.690        4.27                 3.43     7
 8 adios    3.89 0.782        3.93                 3.11     9
 9 later    3.93 0.917        4.01                 3.29    14
10 least    4.08 1.04         3.94                 3.38    13
11 storm    4.29 1.25         3.89                 3.57     7
>
> proc.time()
   user  system elapsed
  1.033   0.056   1.118
```