

# STAT 512 Homework 8

Kathleen Wendt

04/14/2020

## Biomass Study

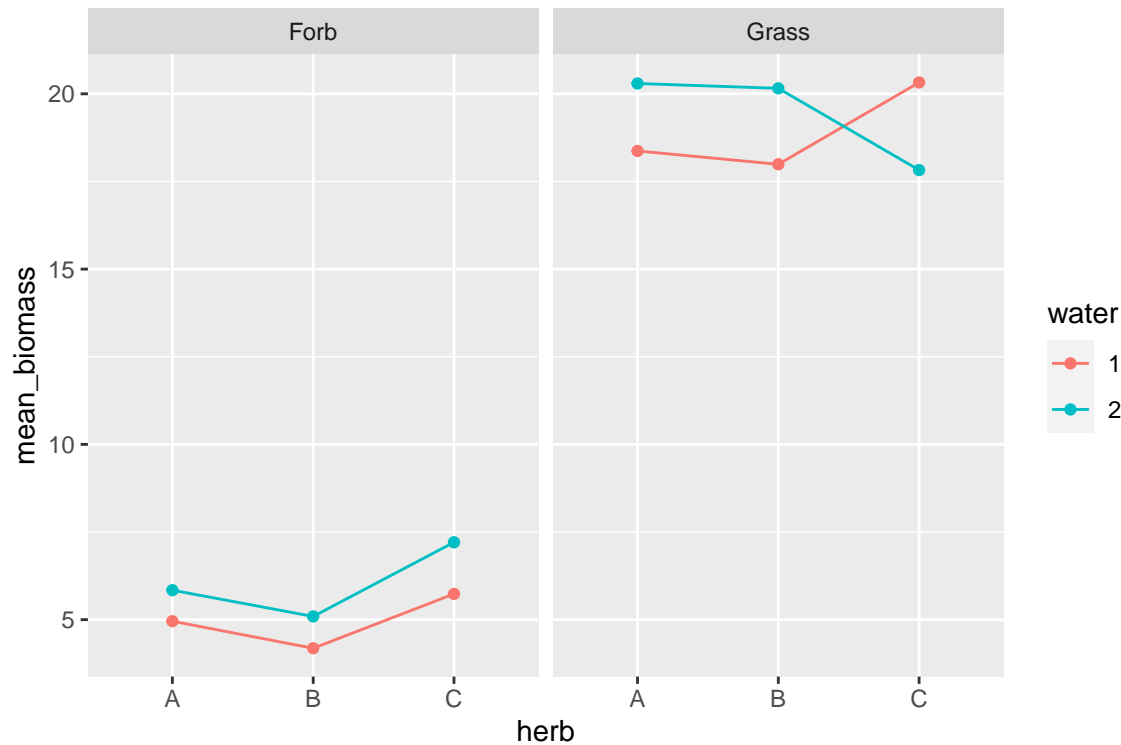
A greenhouse study was done to examine the effect of 3 herbicides (A,B, or C) and 2 water regimes (1 = Low or 2 = High) for 2 plant types (Grass or Forb). The response variable is biomass. There are 3 reps per treatment combination for a total of 36 observations. Each observation was a potted plant. The 36 pots were randomly assigned without restriction to locations in the greenhouse. The data is available from Canvas as “Biomass.csv”.

```
## Observations: 36
## Variables: 4
## $ type      <fct> Grass, Grass, Grass, Grass, Grass, Grass, Grass, Grass, Gra...
## $ herb      <fct> A, A, A, A, A, A, B, B, B, B, B, B, C, C, C, C, C, C, A, A,...
## $ water     <fct> 1, 1, 1, 2, 2, 2, 1, 1, 1, 2, 2, 2, 1, 1, 1, 2, 2, 2, 1, 1,...
## $ biomass   <dbl> 16.395604, 19.859974, 18.853281, 18.213760, 21.211094, 21.4...
```

## Question 1: Plots

Create two-way interaction plots separately for each level of Type. (4 pts) This can be done by first calculating treatment means (using aggregate or dplyr).

type	herb	water	n	mean_biomass	sd_biomass
Forb	A	1	3	4.955550	0.2872469
Forb	A	2	3	5.840386	0.2357440
Forb	B	1	3	4.183644	0.3286999
Forb	B	2	3	5.092050	0.0831322
Forb	C	1	3	5.734037	0.1554219
Forb	C	2	3	7.207097	0.1902947
Grass	A	1	3	18.369620	1.7821087
Grass	A	2	3	20.294048	1.8057830
Grass	B	1	3	17.989272	0.8633054
Grass	B	2	3	20.155989	0.8424093
Grass	C	1	3	20.323887	0.8555735
Grass	C	2	3	17.824716	1.7091929



## Question 2: Three-way ANOVA model

Fit the three-way model with all interactions. Include the Type 3 ANOVA table in your assignment. You should find evidence of a 3 way interaction.

term	sumsq	df	statistic	p.value
(Intercept)	5473.8021497	1	5539.1323374	0.0000000
type	1678.7362340	1	1698.7720610	0.0000000
herb	5.0683466	2	2.5644188	0.0978717
water	5.9007118	1	5.9711372	0.0222710
type:herb	5.3729130	2	2.7185195	0.0862592
type:water	0.7008443	1	0.7092089	0.4080205
herb:water	7.9002258	2	3.9972578	0.0317416
type:herb:water	13.1311975	2	6.6439596	0.0050550
Residuals	23.7169368	24	NA	NA

There is evidence supporting a three-way interaction between plant type, herbicide, and water pattern.

## Question 3: Assumptions

Consider the diagnostics plots and discuss whether model assumptions are satisfied.

Residuals vs Fitted: Residuals in megaphone pattern and not evenly distributed. Assumptions of linearity and constant variance likely not supported.

Normal Q-Q: Residuals do not tightly follow line. Heavy tails. Assumption of normality likely not supported, although this may be due in part to the small sample size.

water	herb	type	emmean	SE	df	lower.CL	upper.CL
1	A	Forb	4.955550	0.5739354	24	3.771005	6.140094
2	A	Forb	5.840386	0.5739354	24	4.655842	7.024931
1	B	Forb	4.183644	0.5739354	24	2.999099	5.368188
2	B	Forb	5.092050	0.5739354	24	3.907505	6.276594
1	C	Forb	5.734037	0.5739354	24	4.549493	6.918582
2	C	Forb	7.207097	0.5739354	24	6.022553	8.391642
1	A	Grass	18.369620	0.5739354	24	17.185075	19.554164
2	A	Grass	20.294048	0.5739354	24	19.109504	21.478593
1	B	Grass	17.989272	0.5739354	24	16.804728	19.173817
2	B	Grass	20.155989	0.5739354	24	18.971444	21.340533
1	C	Grass	20.323887	0.5739354	24	19.139342	21.508432
2	C	Grass	17.824716	0.5739354	24	16.640171	19.009260
contrast	herb	type	estimate	SE	df	t.ratio	p.value
1 - 2	A	Forb	-0.8848361	0.8116673	24	-1.090146	0.2864735
1 - 2	B	Forb	-0.9084060	0.8116673	24	-1.119185	0.2741377
1 - 2	C	Forb	-1.4730603	0.8116673	24	-1.814857	0.0820648
1 - 2	A	Grass	-1.9244287	0.8116673	24	-2.370958	0.0261147
1 - 2	B	Grass	-2.1667163	0.8116673	24	-2.669464	0.0134117
1 - 2	C	Grass	2.4991714	0.8116673	24	3.079059	0.0051391

#### Question 4: Pairwise comparisons

Use `emmeans` to calculate pairwise comparisons of Water (1 vs 2) for each level of Herb and Type. You can use code something like this: `emmeans(Model1, pairwise ~ Water|Herb*Type)`

#### Question 5: Two-way model

Now fit a two-way model (including interaction) for Forb only. Include the Type 3 ANOVA table in your assignment.

term	sumsq	df	statistic	p.value
(Intercept)	544.9213058	1	10432.15625	0.0000000
herb	10.1742346	2	97.38948	0.0000000
water	5.3343660	1	102.12289	0.0000003
herb:water	0.3326989	2	3.18465	0.0777198
Residuals	0.6268173	12	NA	NA

No interaction indicated at  $\alpha = 0.05$ .

#### Question 6: Assumptions

Consider the diagnostics plots and discuss whether model assumptions are satisfied.

Residuals vs Fitted: Residuals evenly distributed across level but still some cluster. Assumptions of linearity and constant variance seem reasonable.

Normal Q-Q: Residuals close to line. Assumption of normality is supported.

#### Question 7: Pairwise comparisons

Use `emmeans` to calculate pairwise comparisons of Water (1 vs 2) for each level of Herb.

water	herb	emmean	SE	df	lower.CL	upper.CL
1	A	4.955550	0.131953	12	4.668049	5.243051
2	A	5.840386	0.131953	12	5.552885	6.127887
1	B	4.183644	0.131953	12	3.896143	4.471145
2	B	5.092050	0.131953	12	4.804549	5.379551
1	C	5.734037	0.131953	12	5.446536	6.021538
2	C	7.207097	0.131953	12	6.919597	7.494598

contrast	herb	estimate	SE	df	t.ratio	p.value
1 - 2	A	-0.8848361	0.1866097	12	-4.741640	0.0004789
1 - 2	B	-0.9084060	0.1866097	12	-4.867946	0.0003863
1 - 2	C	-1.4730603	0.1866097	12	-7.893804	0.0000043

water	emmean	SE	df	lower.CL	upper.CL
1	4.957744	0.0761831	12	4.791755	5.123732
2	6.046511	0.0761831	12	5.880522	6.212500

contrast	estimate	SE	df	t.ratio	p.value
1 - 2	-1.088768	0.1077392	12	-10.10559	3e-07

### Question 8: LSD - REVISIT

Calculate the LSD value (ME for difference between means) corresponding to the comparisons from the previous question. Show your work for full credit. *Hint: See ExpDesign2 slide 81. Recall that you can check your calculation using the SE value given in the output from the previous question.*

ME = TV x SE

## [1] 36.46639

## [1] 0.70499

### Question 9: Pairwise comparisons

Use `emmeans` to calculate the comparison of Water (1 vs 2) averaging over the levels of Herb.

### Question 10: LSD - REVISIT

Calculate the LSD value (ME for difference between means) corresponding to the comparisons from the previous question. Show your work for full credit. *See hints from #8.*

### Question 11: Power

Comparing your LSD values from questions #8 (interaction comparison) and #10 (main effect comparison), which comparison has higher power? Briefly discuss.

As sample size increases and margin of error decreases, power increases.

### Question 12: Two-way model

Now fit a two-way model (including interaction) for Grass only. Include the Type 3 ANOVA table in your assignment.

water	herb	emmean	SE	df	lower.CL	upper.CL
1	A	18.36962	0.8008697	12	16.62467	20.11456
2	A	20.29405	0.8008697	12	18.54910	22.03899
1	B	17.98927	0.8008697	12	16.24433	19.73422
2	B	20.15599	0.8008697	12	18.41104	21.90093
1	C	20.32389	0.8008697	12	18.57894	22.06883
2	C	17.82472	0.8008697	12	16.07977	19.56966
contrast	herb	estimate	SE	df	t.ratio	p.value
1 - 2	A	-1.924429	1.132601	12	-1.699124	0.1150484
1 - 2	B	-2.166716	1.132601	12	-1.913045	0.0798986
1 - 2	C	2.499171	1.132601	12	2.206578	0.0475719

term	sumsq	df	statistic	p.value
(Intercept)	6607.617078	1	3433.9971610	0.0000000
herb	0.267025	2	0.0693868	0.9333373
water	1.267190	1	0.6585622	0.4328704
herb:water	20.698724	2	5.3785926	0.0214969
Residuals	23.090120	12	NA	NA

Evidence of interaction.

### Question 13: Assumptions

Consider the diagnostics plots and discuss whether model assumptions are satisfied.

Residuals vs Fitted: Residuals evenly distributed across residual values but not fitted values. Assumptions of linearity and constant variance seem reasonable, but there there may be evidence of a systemic difference between the “two sides” of the data (i.e., interaction effect).

Normal Q-Q: Residuals close to line but tails look heavy. Assumption of normality is okay, given the small sample size.

### Question 14: Pairwise comparisons

Use `emmeans` to calculate pairwise comparisons of Water (1 vs 2) for each level of Herb.

### Question 15: Reflection

Would it be appropriate to calculate the comparison of Water (1 vs 2) averaging over the levels of Herb? Briefly discuss.

No, it would not be appropriate to average over herbicide because there is evidence of an interaction between water patter and herbicide among plant type **grass**.

### Question 16: Model comparison

Now we compare the three-way model to the separate two-way models. The comparisons from #7 and #14 were also considered in #4. Considering the corresponding results from the combined analysis (#4) versus the split analysis (#7, #14), briefly discuss what is the same and what is different. Specifically, consider the estimated difference, corresponding SE and df. (4 pts)

The combined model (#4) had 24 degrees of freedom, whereas the split-type models (#7 and #14) each had 12 degrees of freedom. Split model 1 for **forb** (#7) had the smallest standard errors, but split model 2 for **grass** (#14) had the largest standard errors. The estimated marginal means and estimated differences across the models (i.e., combined vs. split 1; combined vs. split 2) were the same.

### **Question 17: Reflection**

Give (at least) one reason why we might prefer to split the analysis by Type (running separate 2way ANOVAs for Grass and Forb). Your answer should be based on the output.

Forb and Grass had very different patterns of mean biomass by water pattern and herbicide. Splitting the data would simplify the analysis and specify comparisons of interest consistent with the research questions (e.g., maybe huge differences by plant type are already known). Also, splitting the data might help with the concern of unequal variance.

### **Question 18: Reflection**

Give (at least) one weakness of splitting the analysis “by type” as compared to the full 3-way ANOVA model.

Splitting the data reduces dfResid, therefore reducing power. In this case, the three-way model had 24 dfResid, but each two-way model only had 12 dfResid.

## Appendix

```
# load packages
library(tidyverse)
library(janitor)
library(kableExtra)
library(car)
library(broom)
library(emmeans)
# set global options
knitr::opts_chunk$set(fig.width = 6,
                        fig.height = 4,
                        fig.path = "figs/",
                        echo = FALSE,
                        warning = FALSE,
                        message = FALSE)
# read and prepare biomass data
bio_data <- readr::read_csv("data/Biomass.csv") %>%
  janitor::clean_names() %>%
  dplyr::mutate(type = as.factor(type),
                herb = as.factor(herb),
                water = as.factor(water))
tibble::glimpse(bio_data)
# 1. build summary stats by type
bio_table <- bio_data %>%
  dplyr::group_by(type, herb, water) %>%
  dplyr::summarize(
    n = n(),
    mean_biomass = mean(biomass),
    sd_biomass = sd(biomass)
  )
# 1. kable table
kableExtra::kable(bio_table)
# 1. build two-way interaction plots
ggplot2::qplot(x = herb,
               y = mean_biomass,
               color = water,
               group = water,
               data = bio_table) +
  geom_line() +
  facet_grid(. ~ type)
# 2. adjust contrast default
options(contrasts = c("contr.sum", "contr.poly"))
# 2. fit three-way anova model
bio_anova_3 <- lm(biomass ~ type*herb*water,
                  data = bio_data)
# 2. call tidy anova type 3 table
kableExtra::kable(broom::tidy(car::Anova(bio_anova_3, type = 3)))
# 3. diagnostic plots for three-way anova model
plot(bio_anova_3)
# 4. pairwise comparisons of water for each herb and type level/combo
bio_water_pairs <- emmeans::emmeans(bio_anova_3, pairwise ~ water|herb*type)
```

```

# 4. kable emmeans and contrasts
kableExtra::kable(bio_water_pairs)
# 5. subset data to forb type only
forb_data <- bio_data %>% dplyr::filter(type == "Forb")
# 5. adjust contrast default
options(contrasts = c("contr.sum", "contr.poly"))
# 5. fit two-way anova model for forb data
forb_anova <- lm(biomass ~ herb*water,
                 data = forb_data)
# 5. call tidy anova type 3 table
kableExtra::kable(broom::tidy(car::Anova(forb_anova, type = 3)))
# 6. diagnostic plots for two-way anova model of type forb
plot(forb_anova)
# 7. pairwise comparisons of water for each herb level
forb_water_pairs <- emmeans::emmeans(forb_anova, pairwise ~ water|herb)
# 7. kable emmeans and contrasts
kableExtra::kable(forb_water_pairs)
qt(0.975, 3.77, 12)

3.77*0.187
# 9. pairwise comparisons of water averaging over herb
forb_water_pairs_avg <- emmeans::emmeans(forb_anova, pairwise ~ water)
# 9. kable emmeans and contrasts
kableExtra::kable(forb_water_pairs_avg)
# 5. subset data to grass type only
grass_data <- bio_data %>% dplyr::filter(type == "Grass")
# 5. adjust contrast default
options(contrasts = c("contr.sum", "contr.poly"))
# 5. fit two-way anova model for grass data
grass_anova <- lm(biomass ~ herb*water,
                  data = grass_data)
# 5. call tidy anova type 3 table
kableExtra::kable(broom::tidy(car::Anova(grass_anova, type = 3)))
# 6. diagnostic plots for two-way anova model of type grass
plot(grass_anova)
# 7. pairwise comparisons of water for each herb level
grass_water_pairs <- emmeans::emmeans(grass_anova, pairwise ~ water|herb)
# 7. kable emmeans and contrasts
kableExtra::kable(grass_water_pairs)

```