

STAT 512 Homework 6

Kathleen Wendt

03/31/2020

Part 1: Irrigation data

A study was done to investigate the effectiveness of five methods for the irrigation of blueberry shrubs. Ten farms were included in the study. Each of the five treatments was evaluated at each of the ten farms (with irrigation treatments randomly assigned to plots). The response variable is weight of the harvested fruit. The data is available from Canvas as “Irrigation.csv”. Note: Be sure to define Farm **as.factor**!

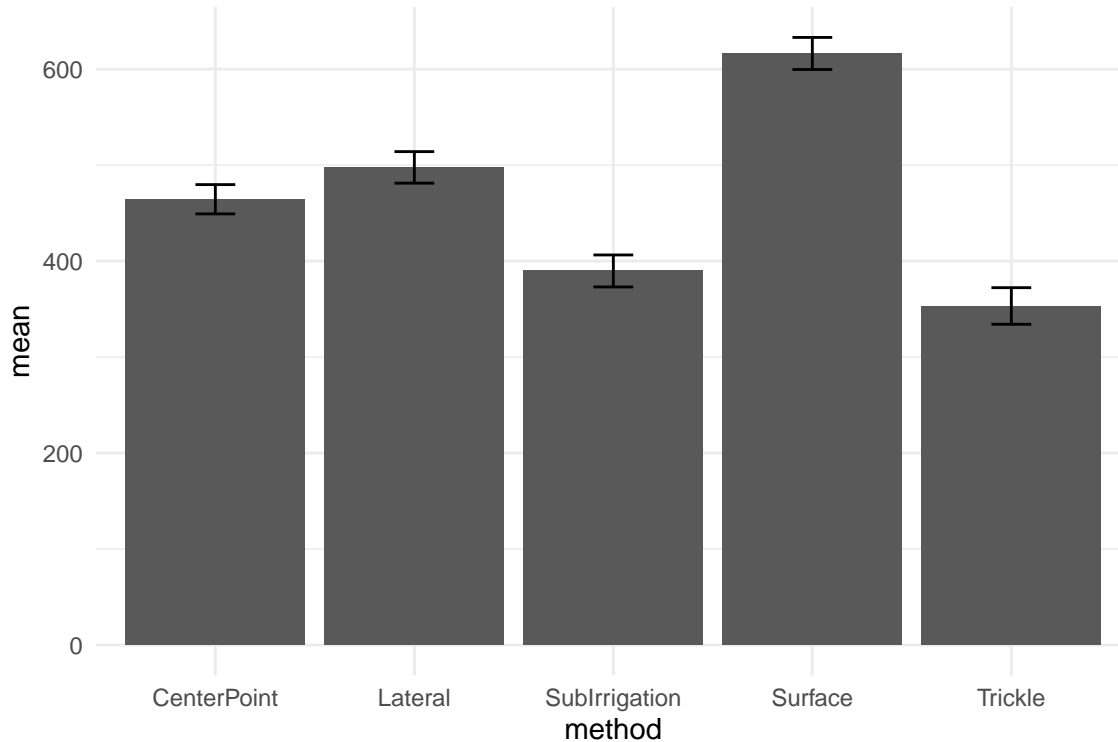
Question A: Summary table

Calculate the sample size, simple mean and SE for each method (averaging over farms). Include the resulting summary table in your assignment.

method	n	mean	sd	se
CenterPoint	10	464.4	48.23369	15.25283
Lateral	10	497.6	52.00684	16.44601
SubIrrigation	10	389.7	52.73635	16.67670
Surface	10	616.3	52.82266	16.70399
Trickle	10	353.2	60.32836	19.07750

Question B: Bar chart

Create a bar chart (with SE bars) to summarize the data. Include the resulting graph in your assignment.



Question C: RCB model and assumptions

Fit the RCB model. Inspect the diagnostic plots (Resids vs Fitted and Normal QQplot of Resids), and comment on what you see. Do the assumptions appear to be satisfied? Note: You do not have to include the diagnostic plot in your assignment, just comment on each graph. (4 pts)

Yes, assumptions appear to be satisfied.

- Residuals vs Fitted: Evenly distributed residuals. Supports assumption of linearity and constant variance.
- Normal Q-Q: Residuals tightly follow line. Supports assumption of normality.

Question D: Type 3 ANOVA for RCB

Continuing with the RCB model from the previous question, include the Type3 ANOVA table in your assignment.

term	sumsq	df	statistic	p.value
(Intercept)	577042.29	1	335.776676	0.0000000
farm	66311.52	9	4.287354	0.0007685
method	421212.52	4	61.275119	0.0000000
Residuals	61867.08	36	NA	NA

Question E: Differences by method

Can we conclude that there is a difference between the irrigation methods? Justify your response with a test statistic and p-value.

Yes, there is a statistically significant difference in blueberry weights by irrigation method, based on the test statistic (61.275), with a corresponding p-value ($1.4338943 \times 10^{-15}$) $< \alpha = 0.05$.

Question F: Effectiveness of blocking

Make a conclusion about the effectiveness of the blocking in this example. Justify your response with a test statistic and p-value.

Yes, blocking by farm was effective, based on the test statistic (4.287), with a corresponding p-value (0.001) $< \alpha = 0.05$.

Question G: Multiple comparisons

The investigators are interested in which irrigation methods are significantly different from each other. Use `emmeans()` function from the `emmeans` package to get Tukey-adjusted p-values for comparing treatments. Then use this information to create a “cld” display, where methods that are NOT significantly different from each other are given the same number grouping. Hint: Use code something like this:

Question H: Simple means and SEs

Are the simple means (part A) and `emmeans` (part G) the same for this analysis? What about the simple SE's (part A) versus SE's returned by `emmeans` (part G)?

Question I: One-way ANOVA

Run the analysis as a one-way ANOVA using just Method in the model. (In practice I would not do this, but try it here for illustration.) Include the ANOVA table in your assignment. How does `dfResid` compare to the RCB model? How does `MSResid` compare to the RCB model? (4 pts) Hint: Recall that `MSResid = SSResid/dfResid`.

Part 2: Fertilizer data

A fertilizer trial on a range grass (blue grama) was conducted in a randomized complete block design. Five fertilizer treatments were randomly assigned to the plots in each of five blocks, but two observations have missing values. The response variable (Y) represents phosphorous. The data is available from Canvas as “GrassMiss.csv”. Note: Be sure to define Block `as.factor`!

Question A: Summary table

Calculate the simple mean for each `trt` (averaging over blocks). Include the resulting summary table in your assignment. Hint: Because of the NA values, it is easiest to use `aggregate()` here.

Question B: RCB model with Type 3 ANOVA

Fit the RCB model and include the Type3 ANOVA table in your assignment.

Question C: EM Means with CIs

Calculate the emmeans and corresponding confidence intervals for each `trt` and include them in your assignment. Note that the SE is larger (and CIs are wider) for treatments that have missing values.

Question D: Means

Are the simple means (part A) and emmeans (part C) the same for this analysis?

Question E: Predict NA values

Use the coefficient estimates (from the `summary()` output) to compute predicted values for the two missing observations. Show your work for full credit. (Note that you can verify these using the `predict()` function.) (4 pts)

Question F: N50wP average

Verify that the emmean for N50wP is the average of the five predicted values (one from each block) for N50wP. Show your work for full credit.

Appendix

```
# load packages
library(tidyverse)
library(janitor)
# library(broom)
# library(kableExtra)
# library(MASS)
# library(emmeans)
# library(MuMIn)
# library(car)
# library(ResourceSelection)
# set global options
knitr::opts_chunk$set(fig.width = 6,
                        fig.height = 4,
                        fig.path = "figs/",
                        echo = FALSE,
                        warning = FALSE,
                        message = FALSE)

# 1. load blueberry irrigation data
berry_data <- readr::read_csv("data/Irrigation.csv") %>%
  janitor::clean_names() %>%
  dplyr::mutate(method = as.factor(method),
                farm = as.factor(farm))

# 1a. create summary statistics table for blueberry data
berry_table <- berry_data %>%
  dplyr::group_by(method) %>%
  dplyr::summarize(
    n = n(),
    mean = mean(weight),
    sd = sd(weight),
    se = sd/sqrt(n)
  )

kableExtra::kable(berry_table)

# 1b. create summary bar chart
berry_plot <- berry_table %>%
  ggplot2::ggplot(aes(x = method, y = mean)) +
  geom_bar(stat = "identity") +
  geom_errorbar(aes(ymin = mean - se,
                    ymax = mean + se),
                width = 0.2) +
  theme_minimal()

berry_plot

# 1c. fit rcb model for blueberry data
berry_rcb_lm <- lm(weight ~ farm + method, data = berry_data)
# 1c. rcb diagnostic plots
par(mfrow = c(2, 2))
plot(berry_rcb_lm)

# 1d. call and tidy anova type 3 on rcb berry
berry_type3 <- broom::tidy(car::Anova(berry_rcb_lm, type = 3))
# 1d. kable type 3 table
```

```
kableExtra::kable(berry_type3)
# library(emmeans)
# library(multcomp)
# emout <- emmeans(Model1, pairwise ~ Method)
# emout
# cld(emout$emmeans)
```