4 Statistical Frameworks Study Guide (and other notes?)

EA30 2023 Students and Marc

2023-03-27

# Using R to Create R Study Guide

## Generic Template to Create Word Documents

I created this document to give you a “head-start” in creating the study guide with a format that everyone can use. By using Rstudio, you will have the capacity to create an integrated approach that merges R commands and outputs and text.

I have defined the output as a word document to start with – but we can create a pdf or html output if we decide that is better. But by putting into Word, we can use Word to spot typos easier!

There are ways to share these documents via github.com – but frankly, it works great if the class is smaller, e.g. 10-12. So, we’ll use an assortment of sakai upload tools – not ideal, but I’ll make it work!

## Developing Content

Each week, I intend to dedicate 40-60 minutes to develop content.

We’ll start with creating an example. But as we do, we want to be answering these following questions:

* What questions can be answered with the method?
* Are their limitations to the method?
* What are the assumptions?
* How can the assumptions be tested?

The example should include data to be analyzed – either imported or created in R, then the R codes to analyze the results and then a discussion about how to interpret the data.

### Defining Variable Types

#### Dependent vs. Independent Data (add text here…)

#### Continuous vs. Count vs. Categorical

NOTE: Spatial Data and Time Series data have special concerns… “auto-correlation”

## Class Assignments

The study guide development will go for five weeks, so everyone can have a change to look at each of the for frameworks and then one session to integrate everything into one document. See sakai to see your group assignment.

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

summary(cars) # extract summary statistics from built-in cars dataset.

## speed dist   
## Min. : 4.0 Min. : 2.00   
## 1st Qu.:12.0 1st Qu.: 26.00   
## Median :15.0 Median : 36.00   
## Mean :15.4 Mean : 42.98   
## 3rd Qu.:19.0 3rd Qu.: 56.00   
## Max. :25.0 Max. :120.00

I added the comment line to help the reader remember what’s happening. These comments are useful to help everyone see the intentions of the author.

### Including Plots

You can also embed plots, for example:



Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot. But as a guide, we might want to show the code – which can be done with echo = TRUE, which is the default.

# Four Frameworks

## Linear Regression

Example #2

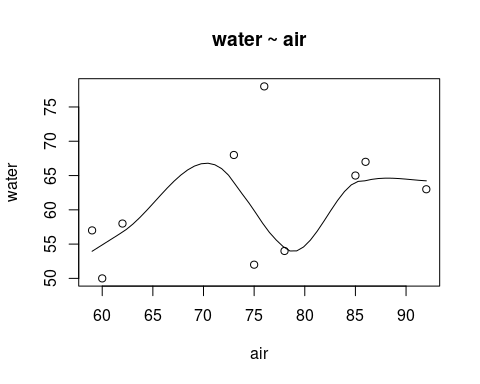
Is water and air temperature corrlated?

water = c(50, 52, 65, 67, 68, 78, 54, 63, 57, 58)  
air = c(60, 75, 85, 86, 73, 76, 78, 92, 59,62)  
lm(formula = water ~ air)

##   
## Call:  
## lm(formula = water ~ air)  
##   
## Coefficients:  
## (Intercept) air   
## 36.3333 0.3333

making a graph…

scatter.smooth(x=air, y=water, main="water ~ air") # scatterplot



## Logistic Regression

## ANOVA

### Example #1

##Research question: Which fertilizer produces the most crop yield? ##Independent variable/predictor: Fertilizer A, Fertilizer B, Fertilizer C ##Dependent variable/response: amount of crop yield

# install.packages(c(“ggplot2”, “ggpubr”, “tidyverse”, “broom”, “AICcmodavg”))   
  
# Read Data  
# crop.data.csv = file.choose()  
  
crop.data.csv = "/home/mwl04747/beginnersluck/Stats/EA30\_Study\_Guide/ANOVA\_cropdata.csv"  
crop.data <- read.csv(crop.data.csv)  
  
names(crop.data)

## [1] "Density" "Block" "Fertilizer" "Yield"

crop.data.aov <- aov(Yield ~ Fertilizer, data=crop.data)  
  
summary(crop.data.aov)

## Df Sum Sq Mean Sq F value Pr(>F)   
## Fertilizer 1 5.74 5.743 14.91 0.000207 \*\*\*  
## Residuals 94 36.21 0.385   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Tests for Association

# Other Types the Might be Important (Marc might discuss if we have time…)

## Multiple Regression

## Time Series Data

## Count Data

## Mixed Effects Models