

PSYC 7710 Lab

Lab 11 Activity

Andrew Graves, Department of Psychology, University of Virginia

Directions:

- Answer the following questions and save the code you used in an R script.
- For each question, use the **iris** dataset in base R to name the appropriate statistical test, apply the model as both a traditional test and a linear model, ensure equivalence of the two statistical models, and plot the results as a linear model.
- You have until the end of lab to complete.

```
library(tidyverse)
theme_set(theme_bw())
theme_update(text = element_text(family = "serif"),
axis.title.y = element_text(margin = margin(r = 20)))

data("iris")
```

Questions:

- Are there differences between the means of **Petal.Width** for any of the three **Species**?

```
# One-way analysis of variance (ANOVA)
group_means <- iris %>%
  group_by(Species) %>%
  summarize(pet_width = mean(Petal.Width))

set_mean <- group_means %>%
  filter(Species == "setosa") %>%
  select(pet_width) %>%
  pull()

vers_mean <- group_means %>%
  filter(Species == "versicolor") %>%
  select(pet_width) %>%
  pull()

virg_mean <- group_means %>%
  filter(Species == "virginica") %>%
  select(pet_width) %>%
  pull()

aov(Petal.Width ~ Species, data = iris) %>%
  summary()
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## Species      2   80.41    40.21    960 <2e-16 ***
## Residuals   147    6.16     0.04
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

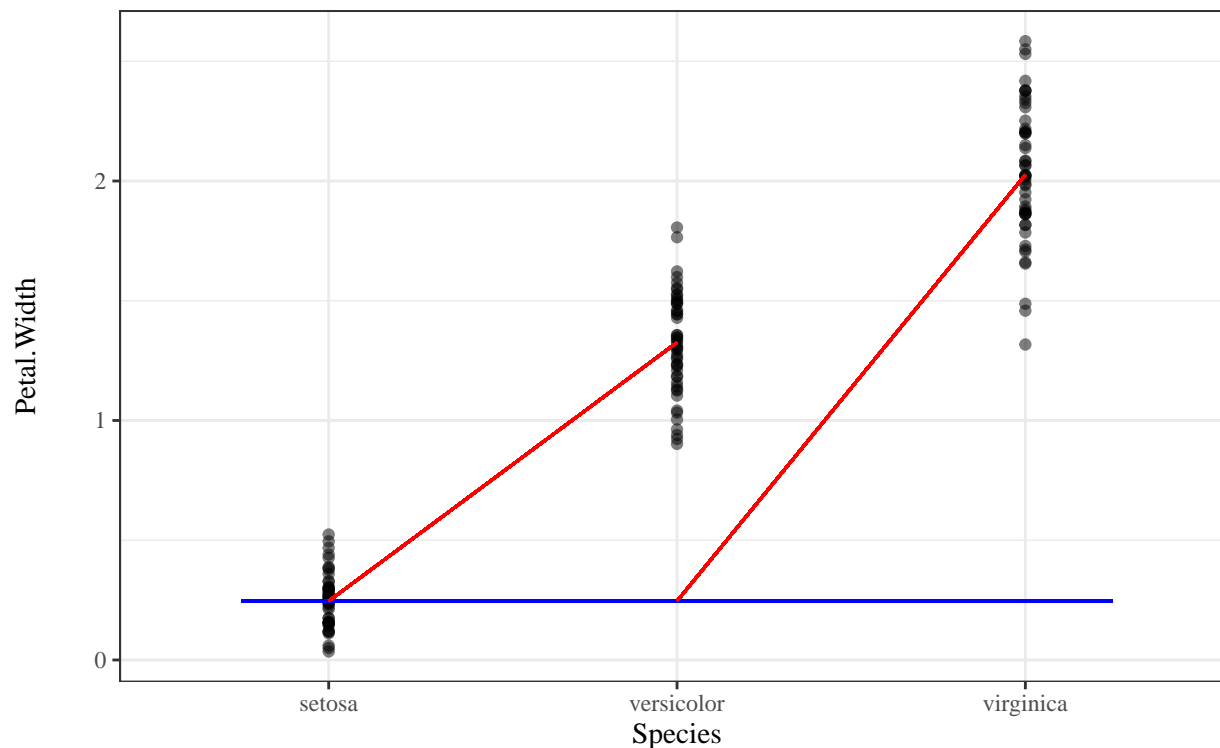
```

lm(Petal.Width ~ Species, data = iris) %>%
  summary()

##
## Call:
## lm(formula = Petal.Width ~ Species, data = iris)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.626 -0.126 -0.026  0.154  0.474
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.24600    0.02894   8.50 1.96e-14 ***
## Speciesversicolor  1.08000    0.04093  26.39 < 2e-16 ***
## Speciesvirginica   1.78000    0.04093  43.49 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2047 on 147 degrees of freedom
## Multiple R-squared:  0.9289, Adjusted R-squared:  0.9279
## F-statistic: 960 on 2 and 147 DF, p-value: < 2.2e-16

# Plot w/ treatment contrasts
iris %>%
  ggplot(aes(x = Species, y = Petal.Width)) +
  geom_jitter(height = .1, width = 0, alpha = 0.5) +
  geom_segment(aes(x = .75, xend = 3.25, y = set_mean, yend = set_mean), color = "blue") +
  geom_segment(aes(x = 1, xend = 2, y = set_mean, yend = vers_mean), color = "red") +
  geom_segment(aes(x = 2, xend = 3, y = set_mean, yend = virg_mean), color = "red")

```



There is a difference between the means of petal width for at least one of the species.

2. Is **Sepal.Width** for the Species *virginica* less than 3 cm?

```
# One sample t-test
one_samp_data <- iris %>%
  filter(Species == "virginica") %>%
  select(Species, Sepal.Width)

t.test(one_samp_data$Sepal.Width, mu = 3)

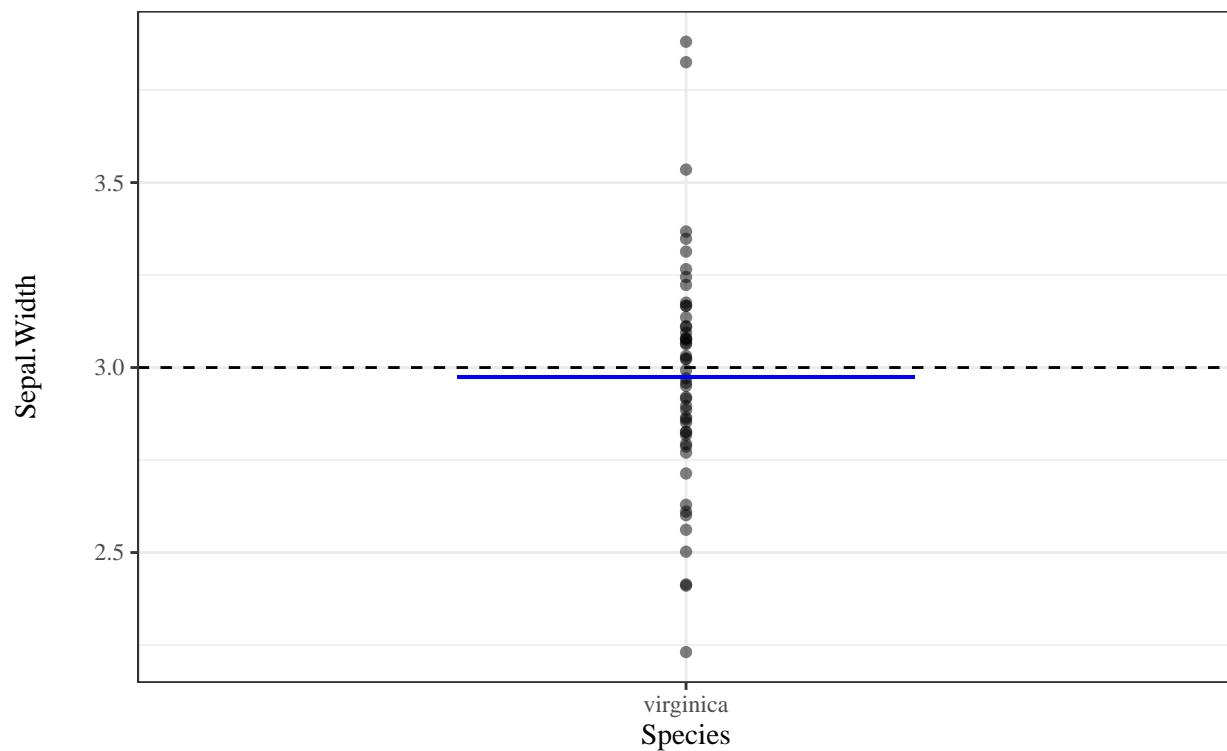
##
## One Sample t-test
##
## data: one_samp_data$Sepal.Width
## t = -0.57008, df = 49, p-value = 0.5712
## alternative hypothesis: true mean is not equal to 3
## 95 percent confidence interval:
##  2.882347 3.065653
## sample estimates:
## mean of x
##      2.974

lm(one_samp_data$Sepal.Width ~ 3 - 1) %>%
  summary()

##
## Call:
## lm(formula = one_samp_data$Sepal.Width ~ 3 - 1)
##
## Residuals:
```

```
##      Min      1Q Median      3Q      Max
## -0.774 -0.174  0.026  0.201  0.826
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.02600    0.04561   -0.57   0.571
##
## Residual standard error: 0.3225 on 49 degrees of freedom
```

```
one_samp_data %>%
  ggplot(aes(x = Species, y = Sepal.Width)) +
  geom_jitter(height = .1, width = 0, alpha = 0.5) +
  geom_hline(yintercept = 3, lty = 2) +
  geom_segment(aes(x = .75, xend = 1.25,
    y = mean(one_samp_data$Sepal.Width),
    yend = mean(one_samp_data$Sepal.Width)), color = "blue")
```



There is little to no evidence that virginica sepal width is less than 3 cm.

3. Is there a difference between the means of **Petal.Length** for the **Species** *versicolor* and *virginica*?

```
# Independent samples t-test
ind_samp_data <- iris %>%
  filter(Species %in% c("versicolor", "virginica")) %>%
  select(Species, Petal.Length)

ind_samp_data_vers <- ind_samp_data %>%
  filter(Species == "versicolor") %>%
  select(Petal.Length) %>%
  pull()
```

```

ind_samp_data_virg <- ind_samp_data %>%
  filter(Species == "virginica") %>%
  select(Petal.Length) %>%
  pull()

t.test(ind_samp_data_virg, ind_samp_data_vers, var.equal = TRUE)

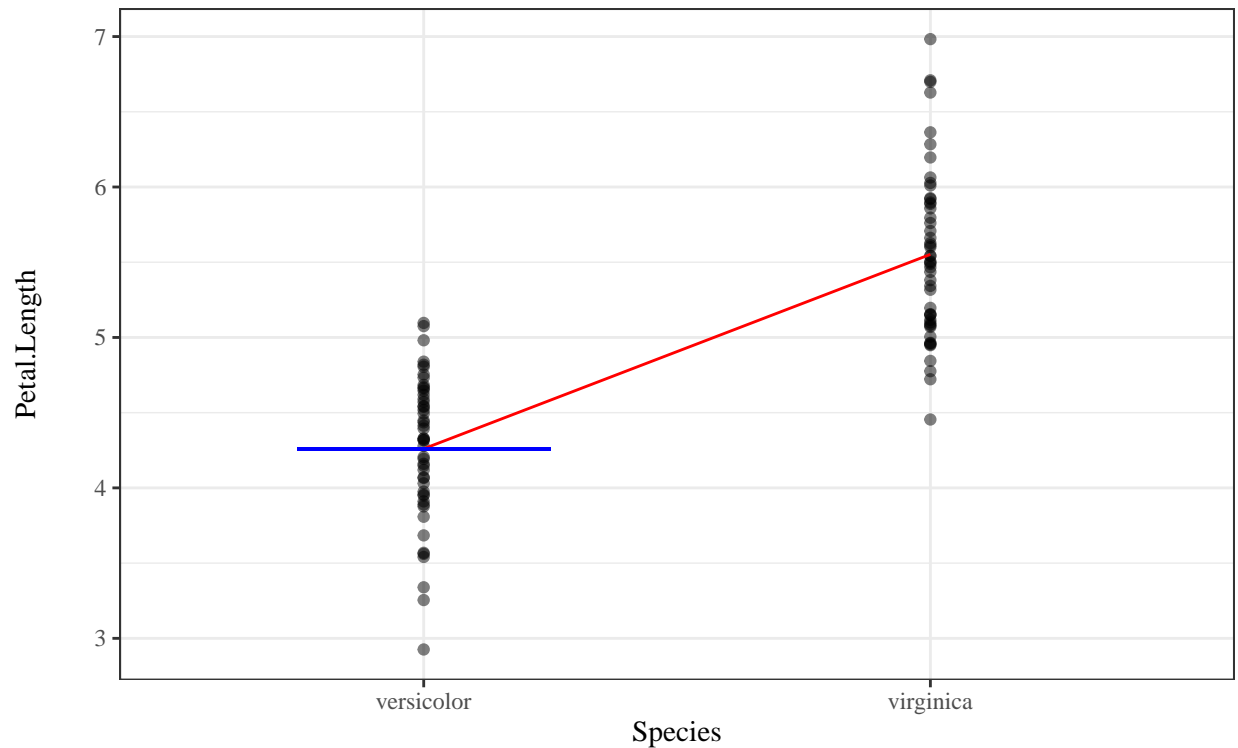
##
## Two Sample t-test
##
## data: ind_samp_data_virg and ind_samp_data_vers
## t = 12.604, df = 98, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.088574 1.495426
## sample estimates:
## mean of x mean of y
##    5.552    4.260

lm(Petal.Length ~ Species, data = ind_samp_data) %>%
  summary()

##
## Call:
## lm(formula = Petal.Length ~ Species, data = ind_samp_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.260 -0.360  0.044  0.340  1.348
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.26000    0.07248   58.77  <2e-16 ***
## Speciesvirginica 1.29200    0.10251   12.60  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5125 on 98 degrees of freedom
## Multiple R-squared:  0.6185, Adjusted R-squared:  0.6146
## F-statistic: 158.9 on 1 and 98 DF,  p-value: < 2.2e-16

ind_samp_data %>%
  ggplot(aes(x = Species, y = Petal.Length, group = 1)) +
  geom_jitter(height = .1, width = 0, alpha = 0.5) +
  stat_summary(fun.y = mean, geom = "line", color = "red") +
  geom_segment(aes(x = .75, xend = 1.25, y = mean(ind_samp_data_vers),
    yend = mean(ind_samp_data_vers)), color = "blue")

```



The virginica petal length is larger than the versicolor petal length.

4. Across all the flowers, are the petals shorter than the sepals?

Paired samples t-test

```
t.test(iris$Petal.Length, iris$Sepal.Length, paired = TRUE)
```

```
##
## Paired t-test
##
## data: iris$Petal.Length and iris$Sepal.Length
## t = -22.813, df = 149, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -2.265959 -1.904708
## sample estimates:
## mean of the differences
## -2.085333
```

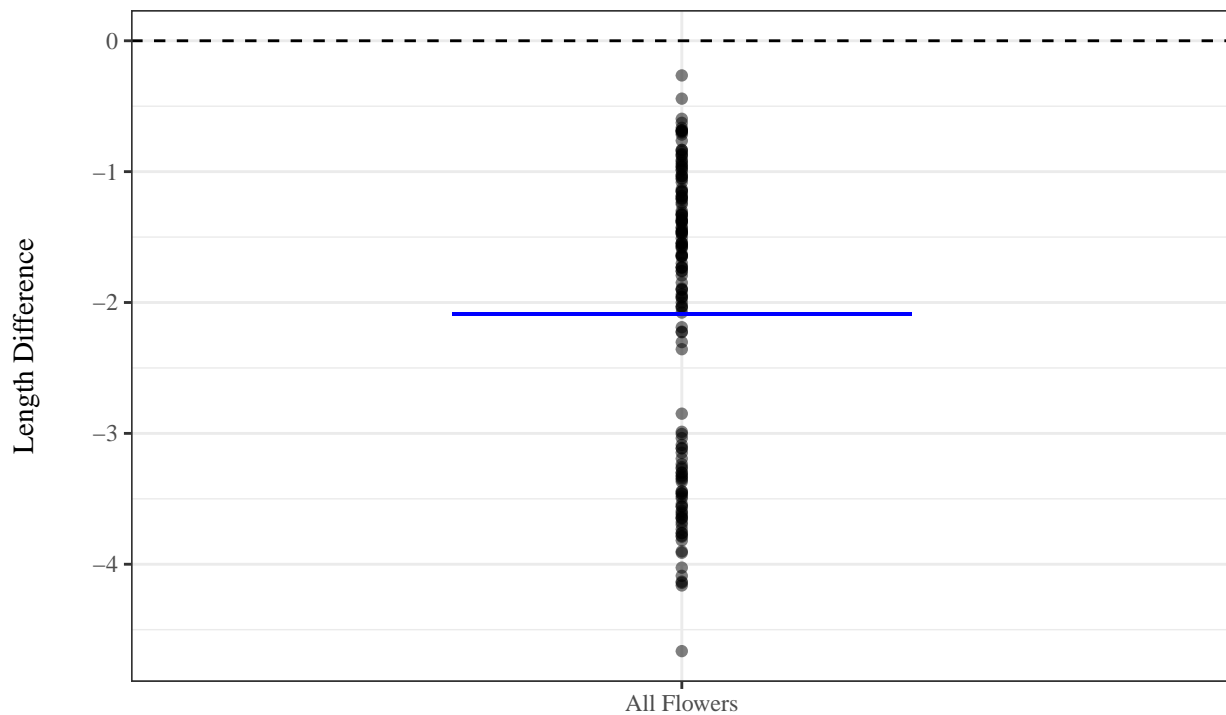
```
lm(Petal.Length ~ Sepal.Length - 1, data = iris) %>%
summary()
```

```
##
## Call:
## lm(formula = Petal.Length ~ Sepal.Length - 1, data = iris)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.5147 -1.1147  0.3853  0.8853  1.7853
##
## Coefficients:
```

```
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.08533    0.09141  -22.81  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.12 on 149 degrees of freedom
```

```
add_length_diff <- iris %>%
  mutate(`Length Difference` = Petal.Length - Sepal.Length,
         `All Flowers` = "All Flowers")

add_length_diff %>%
  ggplot(aes(x = `All Flowers`, y = `Length Difference`)) +
  geom_jitter(height = .1, width = 0, alpha = 0.5) +
  geom_hline(yintercept = 0, lty = 2) +
  geom_segment(aes(x = .75, xend = 1.25, y = mean(add_length_diff$`Length Difference`),
               yend = mean(add_length_diff$`Length Difference`)), color = "blue") +
  labs(x = "")
```



```
# Petals are shorter than the sepals.
```