DATA ANALYSIS

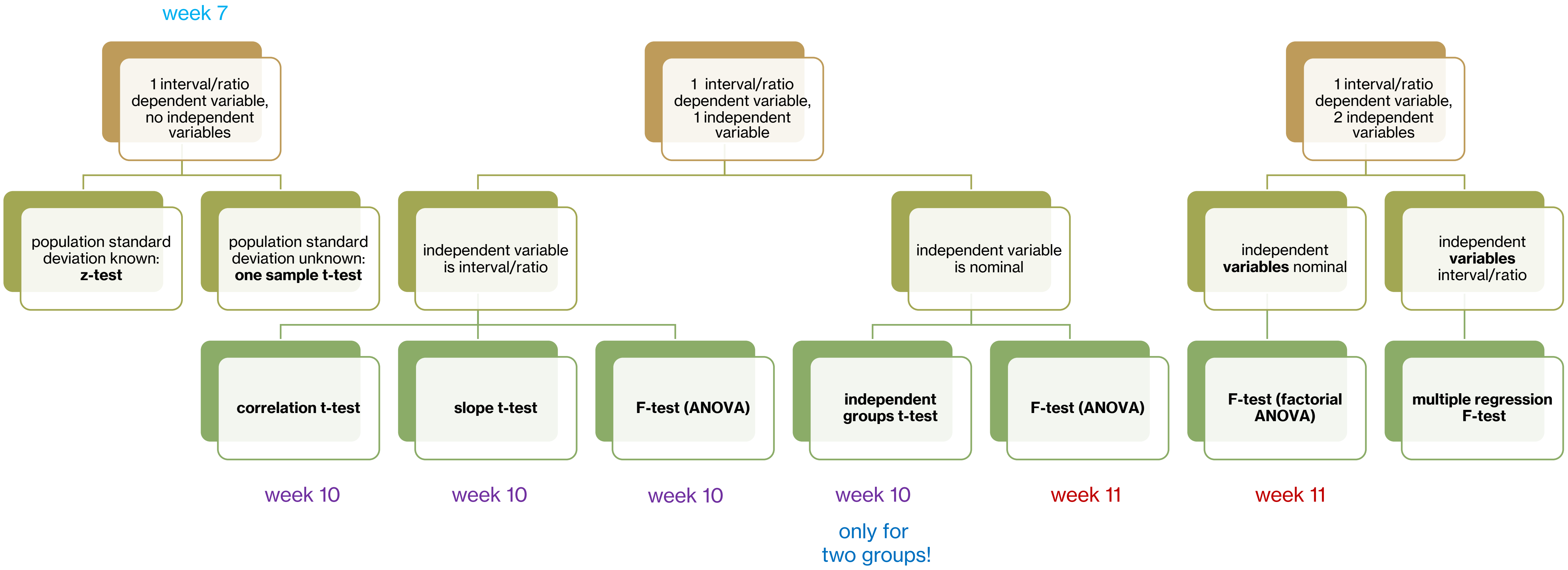
Week 11: F-tests/ANOVAs for nominal data

logistics

- no week 11 or week 12 quiz!
- office hours only 2-4 today
- problem set #4 second attempt due tonight
- problem set #6 (due Apr 9)
- class survey (due Apr 9, Canvas)
- midterm 2 review materials (on Canvas by Friday):
  - midterm 2 practice quiz
  - midterm 2 computational questions
- Apr 10 is a review class for midterm 2

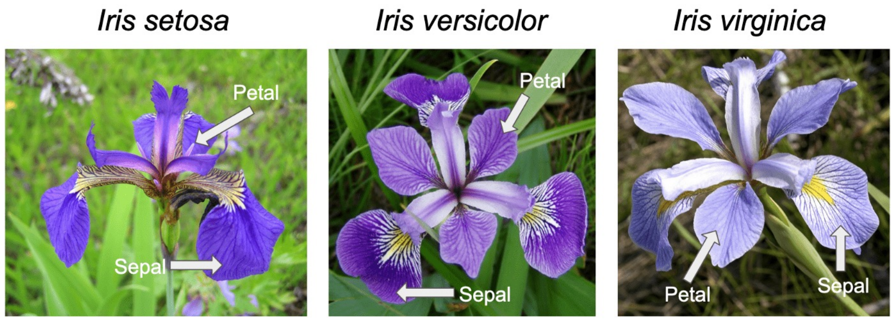
11	W: April 3, 2024	W11: Modeling Relationships II
11	F: April 5, 2024	W11 continued…
12	T: April 9, 2024	**Problem Set 6 due**
12	W: April 10, 2024	W12: Loose Ends / Exam 2 review
12	F: April 12, 2024	**Exam (Midterm) 2**
13	W: April 17, 2024	W13: Non-Independent Data
13	F: April 19, 2024	W13 continued…
14	T: April 23, 2024	**Problem Set Opt-out Deadline 3**
14	W: April 24, 2024	W14: Miscellaneous Data
14	F: April 26, 2024	W14 continued…
15	T: April 30, 2024	**Problem Set 7 due**
15	W: May 1, 2024	W15: Odds and Ends
15	F: May 3, 2024	**Final Exam**
16	W: May 8, 2024	**Wrapping Up!**

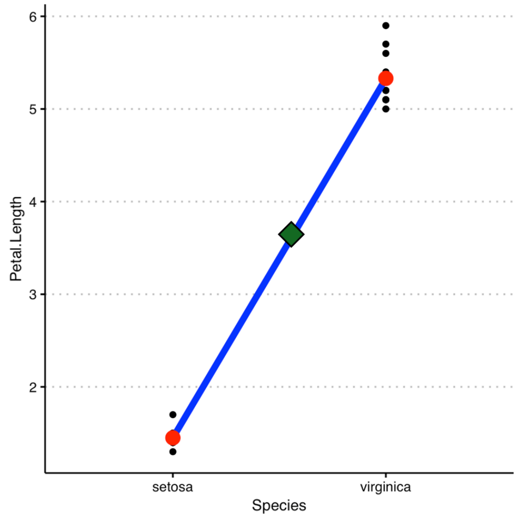
the chart of doom (so far)



review: iris dataset

- our goal is to build the best model for petal lengths
- petal length (Y) = a + b (species:X)
- we conducted an independent groups t-test to evaluate the significance of adding a “species” mean to our model over and above the grand mean
- t-test compared the $M_{setosa}\ to\ M_{virginica}$





review: independent groups t-test

step 1: state the hypotheses

H_0: β = 0 or $\mu_2 - \mu_1 = 0$

H_1: β $\neq 0$ or $\mu_2 - \mu_1 \neq 0$

step 2: set criteria for decision

$\alpha = .05$

find $t_{critical}$ based on one vs. two tailed test and degrees of freedom

$df = n_1 + n_2 - 2$

step 3: collect data

(1) compute $s_{M2-M1} = \sqrt{\frac{s_p{}^2}{n_1} + \frac{s_p{}^2}{n_2}}$ and $b = M_2 - M_1$

(2) compute $t_{observed} = \frac{b - \beta}{s_{M2-M1}}$

(3) find p-value for t-score

step 4: make a decision!

check whether $t_{observed}$ is beyond $t_{critical}$ and p-value < .05. if so, reject null hypothesis!

review: independent groups t-test

- **step 1: state the hypotheses**
  - H_0: $\mu_{virginica} - \mu_{setosa} = 0$: mean petal lengths for both species are equal, i.e.,
  - H_1: $\mu_{virginica} - \mu_{setosa} \neq 0$: mean petal lengths for both species are not equal, i.e.,
- **step 2: set criteria for decision**

$$df = n_1 + n_2 - 2 = 10 + 10 - 2 = 18$$

$$t_{18} = t_{critical} = 2.1009$$

- **step 3: collect data**

$$s_p{}^2 = \frac{SS_1 + SS_2}{df_1 + df_2} = .0525$$

$$s_{M2-M1} = \sqrt{\frac{s_p{}^2}{n_1} + \frac{s_p{}^2}{n_2}} = .1025$$

$$b = M_2 - M_1 = 3.88$$

$$t_{observed} = \frac{b-0}{s_{M2-M1}} = \frac{(3.88)-0}{.1025} = 37.844$$ and p-value <.0001

- **step 4: make a decision!**

Sheets solution
video tutorial

$$s_p{}^2 = \frac{SS_1 + SS_2}{df_1 + df_2}$$

$$s_{M2-M1} = \sqrt{\frac{s_p{}^2}{n_1} + \frac{s_p{}^2}{n_2}}$$

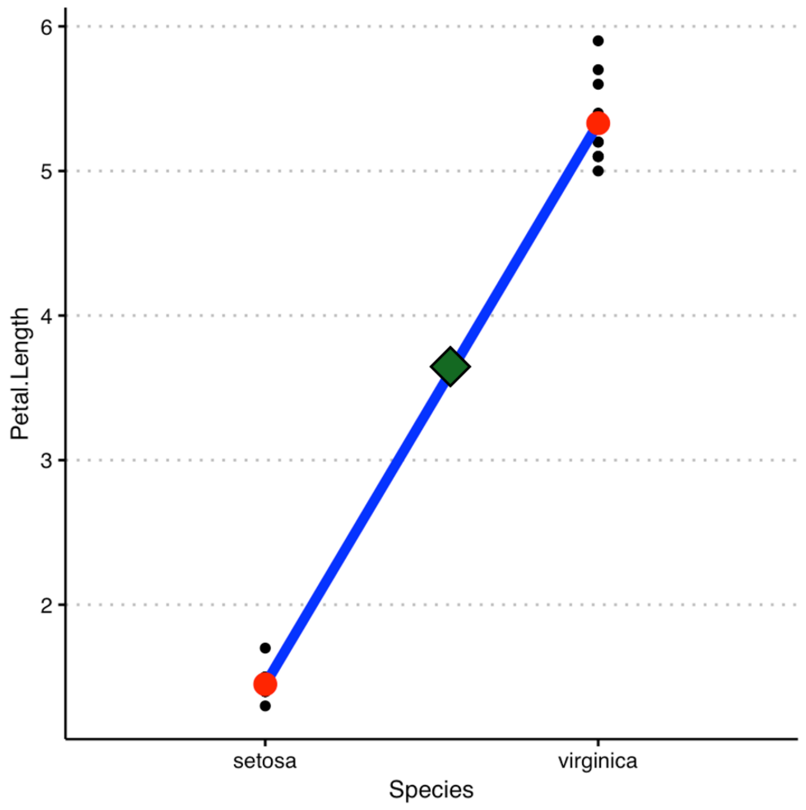
assumptions: t-test

- interval/ratio dependent variable
- independent observations (between-subjects design)
- normality
  - when data are not normal, the t-test is not appropriate
  - BUT: t-tests are fairly robust to minor violations for large n
- homogeneity of variances
  - we assume that the populations from which samples are drawn have equal variances to compute a “pooled” estimate of variance for the independent groups t-test
  - Welch’s test is done for unequal variances

questions?

F-test for two groups

- just as we did an "overall" test for linear regression, we can do the same here for the iris dataset, where we compare the grand mean model with the species mean model
- recall that $F = \frac{MS_{model}}{MS_{error}} = \frac{SS_{model}/df_{model}}{SS_{error}/df_{error}}$
- and SS_{total} = SS_{model} + SS_{error}
- how do we obtain SS_{total}, SS_{error}, and SS_{model}?



F-test for two groups

- SS_{total} represents score deviations from grand mean (M_Y)

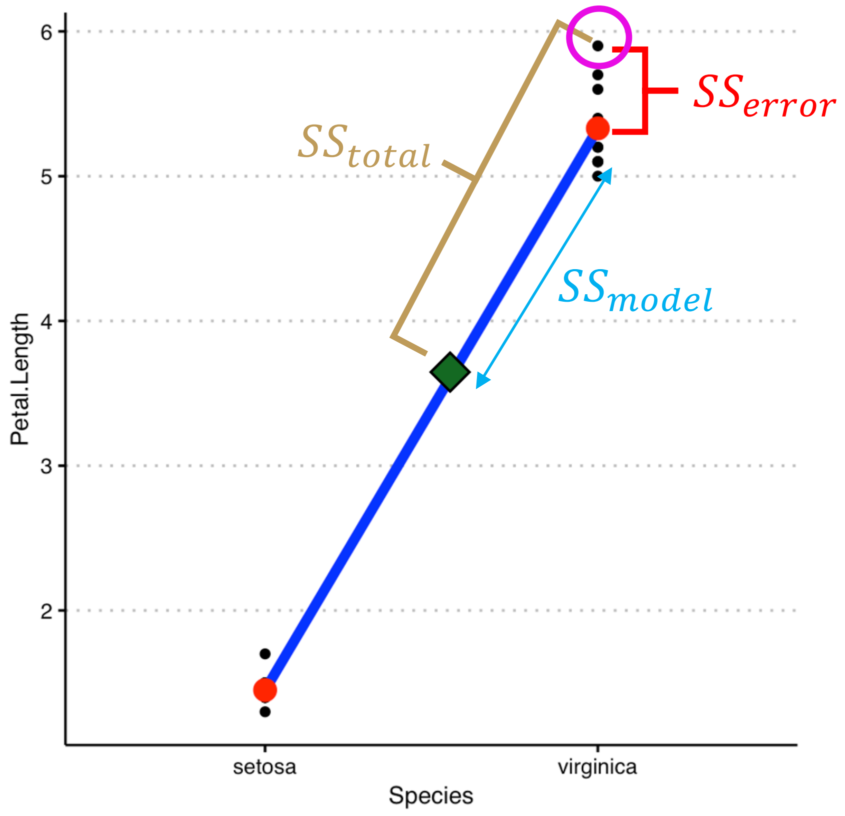
$$SS_{total} = \sum (Y - M_Y)^2$$

- SS_{error} represents the deviations of each score from its group mean

$$SS_{error} = \sum (Y - \hat{Y})^2 = \sum (Y - M_{group})^2$$

- SS_{model} represents the ***gains*** we get if we substitute each score with the group mean instead of the grand mean

$$SS_{model} = \sum \sum n_i \, (M_{group} - M_Y)^2 = SS_{total} - SS_{error}$$



NHST for two independent groups (F-test)

step 1: state the hypotheses

step 2: set criteria for decision

step 3: collect data

step 4: make a decision!

$H_0: \mu_2 - \mu_1 = 0$
$H_1: \mu_2 - \mu_1 \neq 0$

$\alpha = .05$
find $F_{critical}$ based on **right** tailed test and degrees of freedom
$df_1 = k - 1$
$df_2 = n - k$

(1) compute SS_{model} and SS_{error}

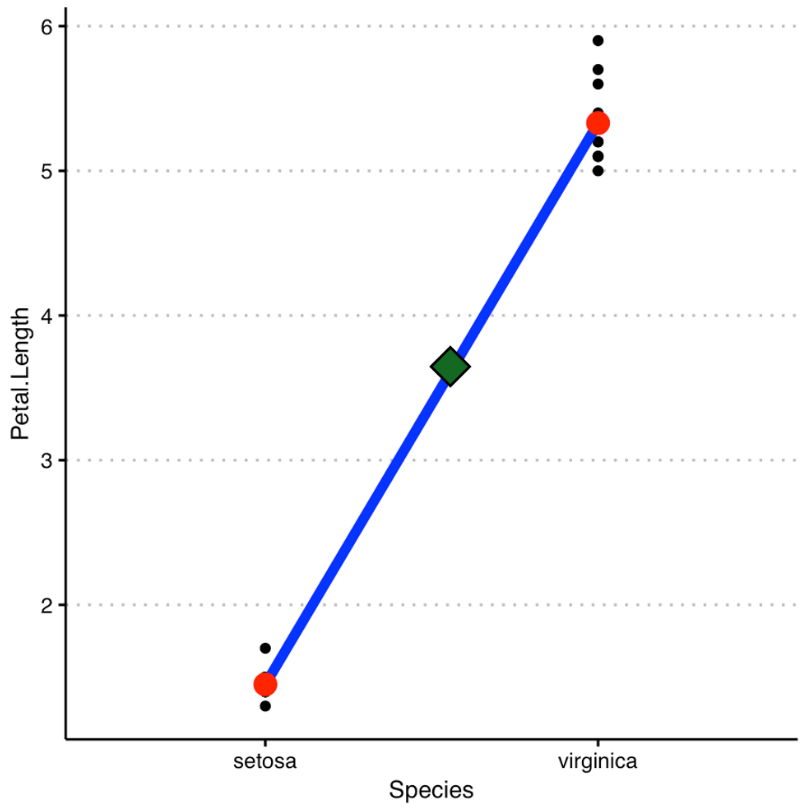
(2) compute $F_{observed} = \frac{MS_{model}}{MS_{error}}$

(3) find p-value for F-score

check whether F is beyond $F_{critical}$ and p-value < .05. if so, reject null hypothesis!

activity: F-test for iris dataset

- conduct the F test for the iris dataset



F-test for iris dataset

- **step 1: state the hypotheses**
  - H$_0$: $\mu_{virginica} - \mu_{setosa} = 0$: petal lengths for both species are equal
  - H$_1$: $\mu_{virginica} - \mu_{setosa} \neq 0$: petal lengths for species are different
- **step 2: set criteria for decision**

$k = 2$: number of levels of independent variable OR estimated parameters

$$df_1 = k - 1 = 2 - 1 = 1$$

$$df_2 = n - k = 20 - 2 = 18$$

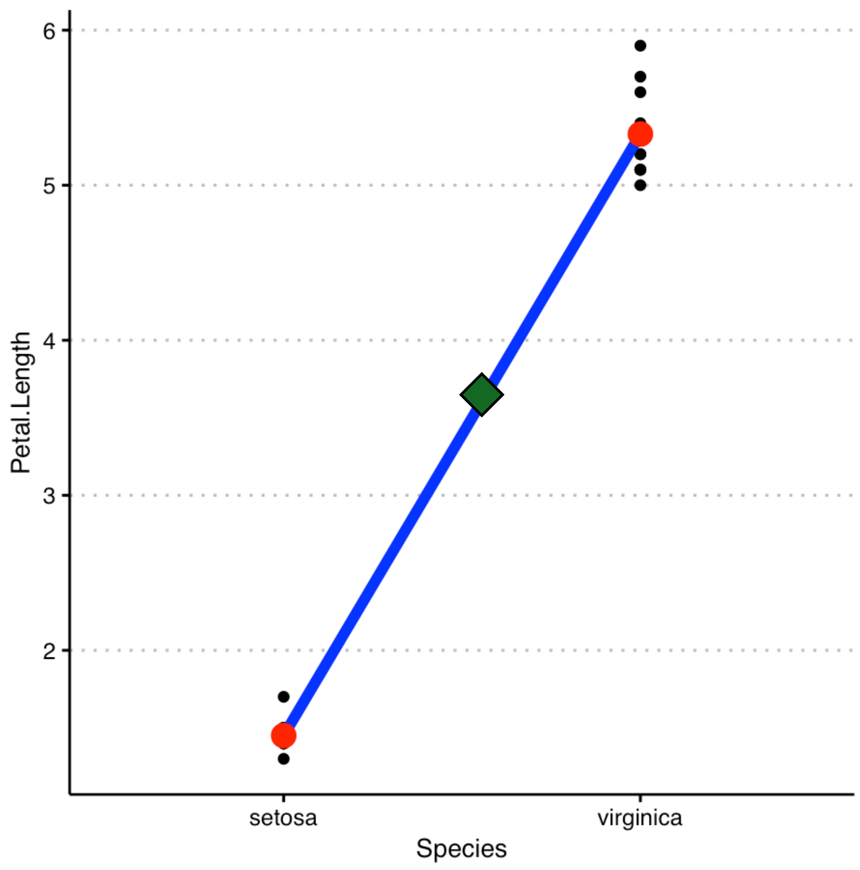
$$F(df_1, df_2) = F(1,18) = F_{critical} = 4.414$$

step 3a: obtaining SS_{total}

- what is SS_{total}? SS_{total} is the error left over after the grand mean has been fit to the data

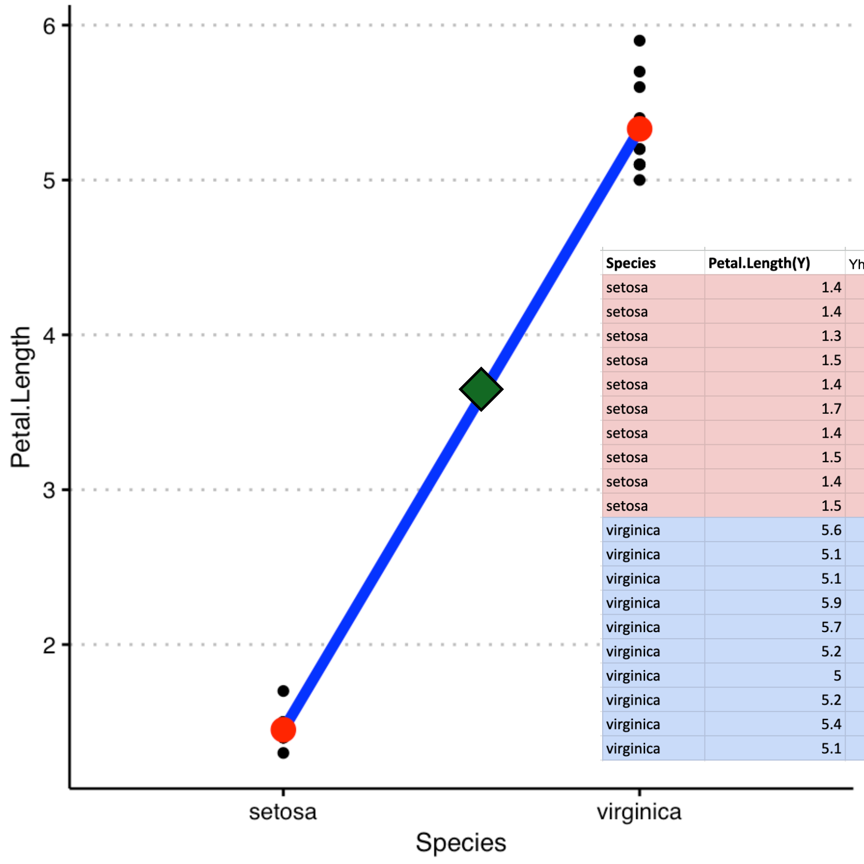
$$SS_{total} = \sum (Y - M_Y)^2$$

- for iris, SS_{total}= 76.218



step 3b: obtaining SS_{error}

- SS_{error} is the error that is left over after our species model has been fit
- our species model substitutes each raw score with the mean of the specific species
- $SS_{error} = \sum(Y - \hat{Y})^2 = \sum(Y - M_{group})^2$
- for iris, SS_{error}= .946



Species	Petal.Length(Y)	Yhat
setosa	1.4	1.45
setosa	1.4	1.45
setosa	1.3	1.45
setosa	1.5	1.45
setosa	1.4	1.45
setosa	1.7	1.45
setosa	1.4	1.45
setosa	1.5	1.45
setosa	1.4	1.45
setosa	1.5	1.45
virginica	5.6	5.33
virginica	5.1	5.33
virginica	5.1	5.33
virginica	5.9	5.33
virginica	5.7	5.33
virginica	5.2	5.33
virginica	5	5.33
virginica	5.2	5.33
virginica	5.4	5.33
virginica	5.1	5.33

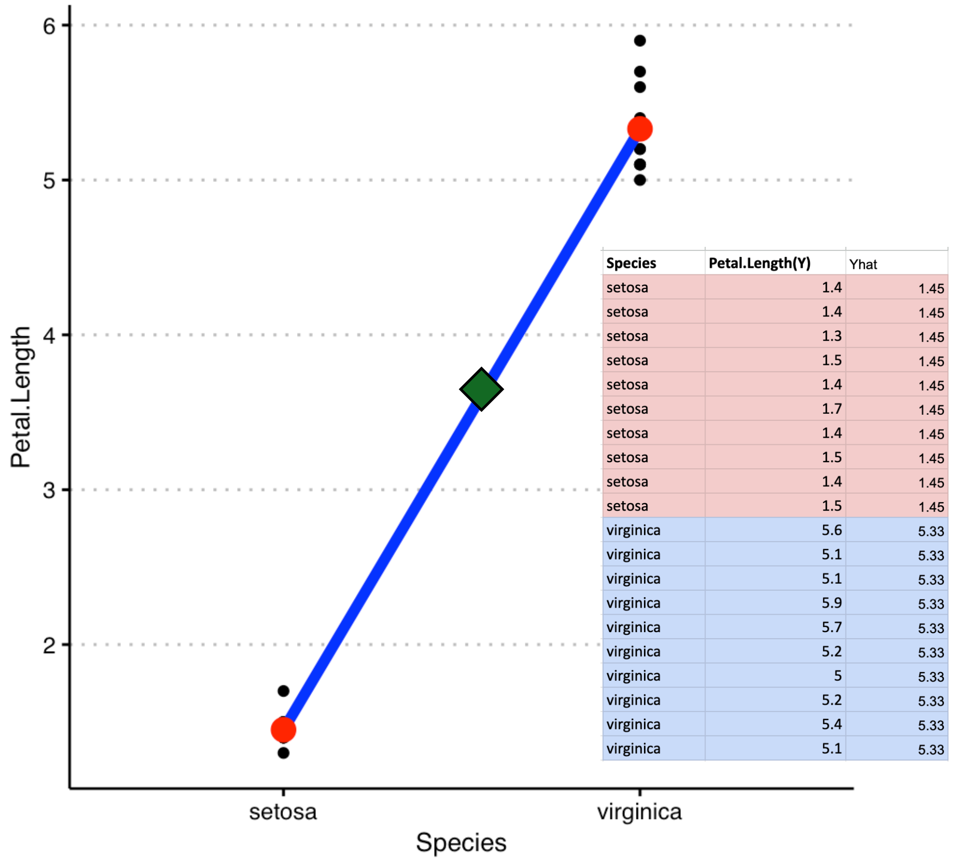
step 3c: obtaining SS_{model}

- how can we obtain SS_{model}?

$$SS_{total} = SS_{model} + SS_{error}$$

$$\text{thus, } SS_{model} = SS_{total} - SS_{error}$$

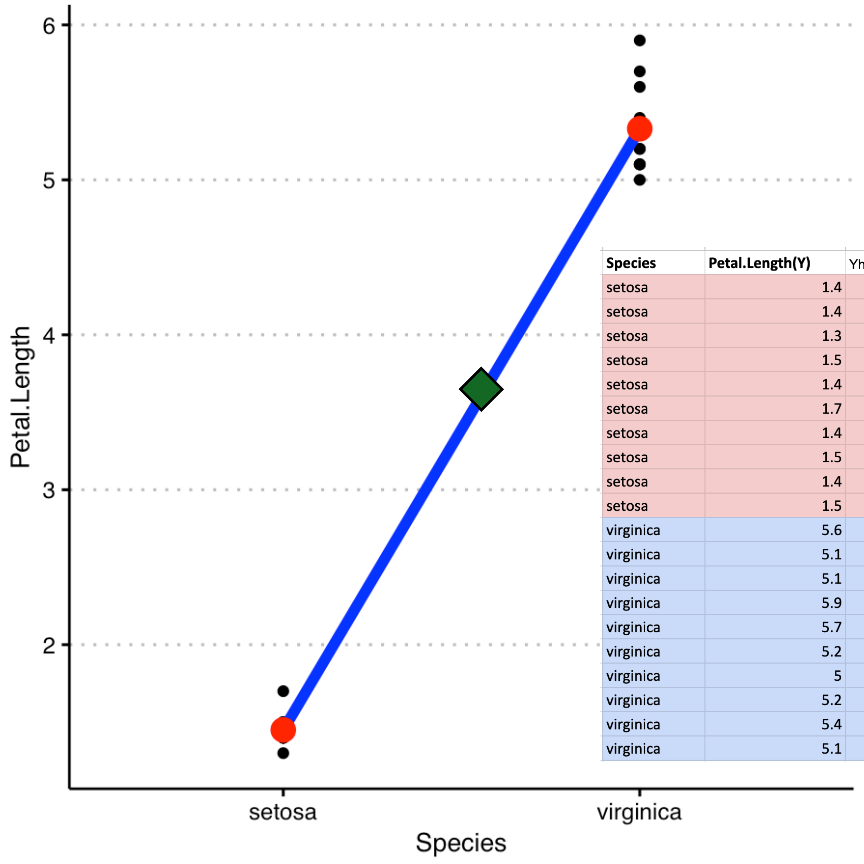
- for iris, SS_{total}= 76.218 and SS_{error}= .946
- SS_{model}= 75.272



Species	Petal.Length(Y)	Yhat
setosa	1.4	1.45
setosa	1.4	1.45
setosa	1.3	1.45
setosa	1.5	1.45
setosa	1.4	1.45
setosa	1.7	1.45
setosa	1.4	1.45
setosa	1.5	1.45
setosa	1.4	1.45
setosa	1.5	1.45
virginica	5.6	5.33
virginica	5.1	5.33
virginica	5.1	5.33
virginica	5.9	5.33
virginica	5.7	5.33
virginica	5.2	5.33
virginica	5	5.33
virginica	5.2	5.33
virginica	5.4	5.33
virginica	5.1	5.33

step 3d: obtaining $F_{observed}$

- $F_{observed} = \frac{MS_{model}}{MS_{error}} = \frac{SS_{model}/df_{model}}{SS_{error}/df_{error}}$ = 1432.24
- p-value = <.0001
- $F_{critical} = 4.414$
- thus, $F(1{,}18) =$ 1432.24, p < .0001
  - we can reject the null hypothesis
  - petal lengths of setosa and virginica are significantly different
  - also, $t^2 = F$!!



Species	Petal.Length(Y)	Yhat
setosa	1.4	1.45
setosa	1.4	1.45
setosa	1.3	1.45
setosa	1.5	1.45
setosa	1.4	1.45
setosa	1.7	1.45
setosa	1.4	1.45
setosa	1.5	1.45
setosa	1.4	1.45
setosa	1.5	1.45
virginica	5.6	5.33
virginica	5.1	5.33
virginica	5.1	5.33
virginica	5.9	5.33
virginica	5.7	5.33
virginica	5.2	5.33
virginica	5	5.33
virginica	5.2	5.33
virginica	5.4	5.33
virginica	5.1	5.33

F-table

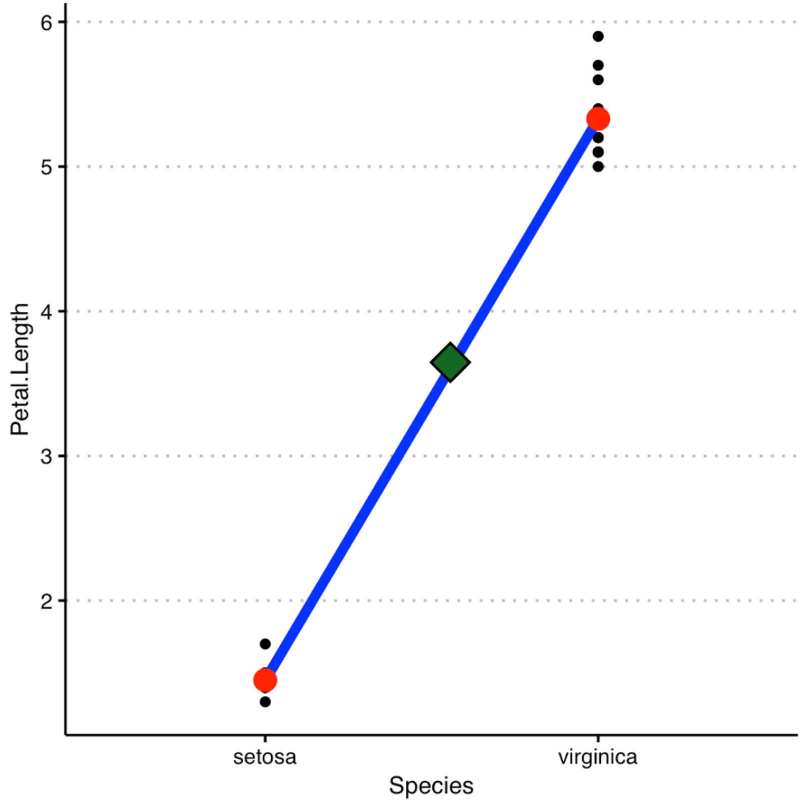
		SS	df	MS	F	p-value
SS_{model}	species	75.272	1	75.272	1432.24	<.0001
SS_{error}	residual	0.946	18	0.0526		

```
Response: Petal.Length
          Sum Sq Df F value    Pr(>F)
Species   75.272  1  1432.2 < 2.2e-16 ***
Residuals  0.946 18
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

[Sheets solution video tutorial]

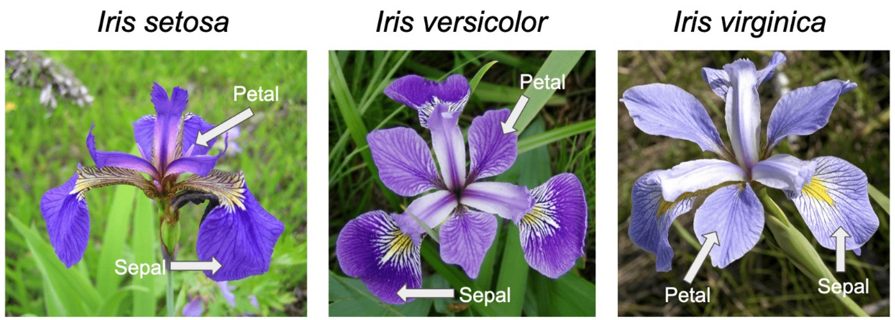
testing more than two groups

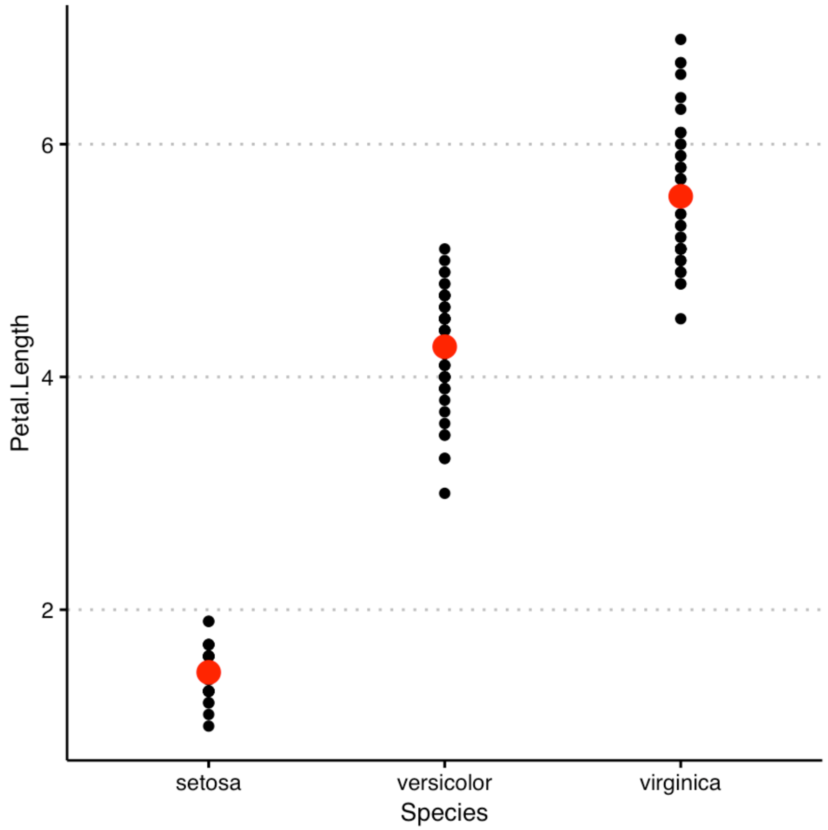
- an independent groups t-test is a special case of linear models ($Y = a + bX$) when X is nominal
- it is *also* a special case of only comparing two groups
- example of comparing more than two groups?



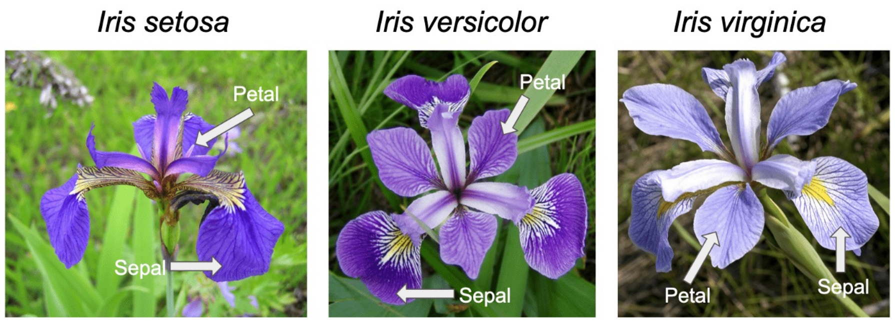
revisiting iris

- recall that the iris dataset actually contains information about three species (setosa, virginica, and versicolor)

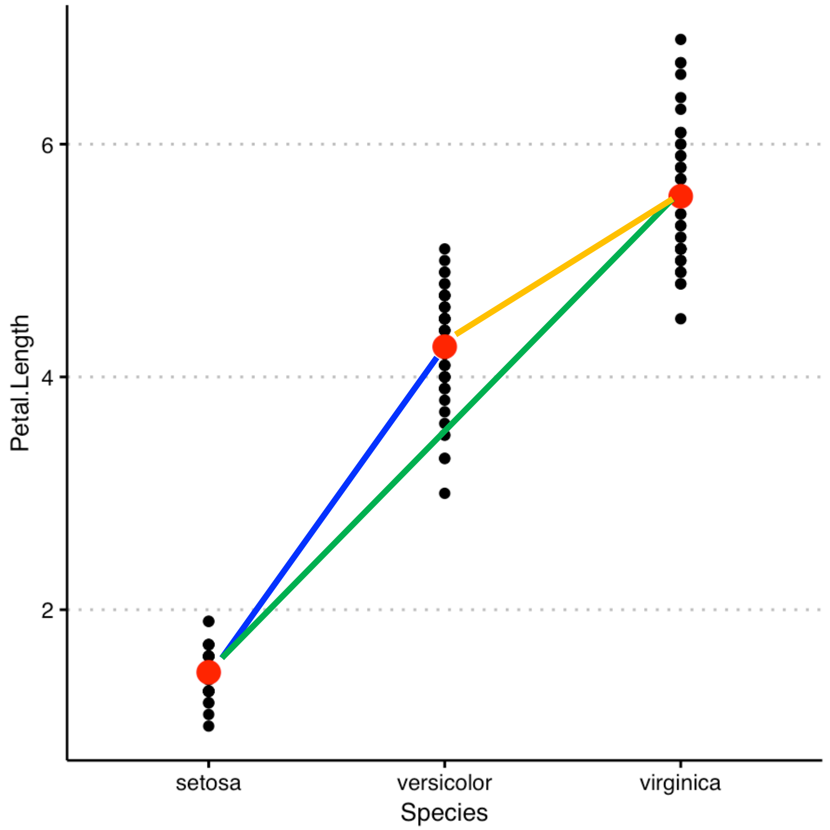




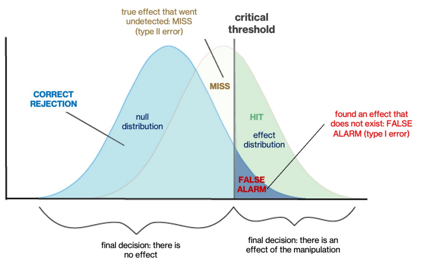
revisiting iris



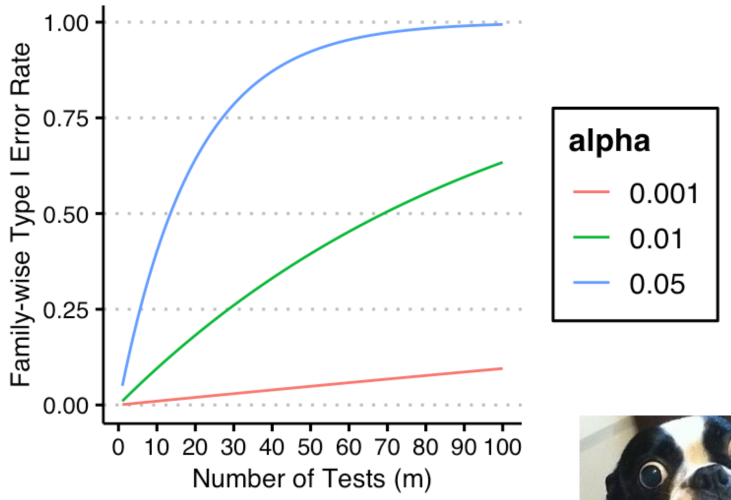
- what if we wanted to look at all three species?
- how many possible comparisons are involved?
  - $M_{virginica} - M_{setosa}$
  - $M_{versicolor} - M_{setosa}$
  - $M_{virginica} - M_{versicolor}$
- we could fit individual linear models for each comparison and conduct the t-test/F-test for each comparison



multiple tests and type I errors

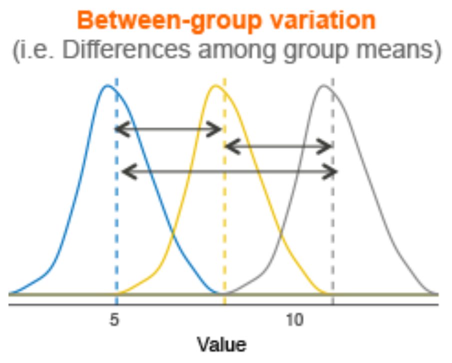


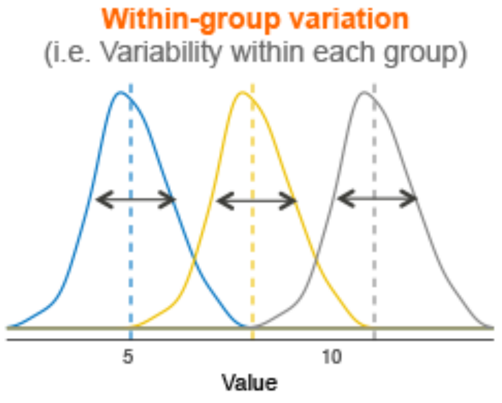
- each time a hypothesis test is conducted with some α-level, there is α % probability of making a type I error
- as more tests are conducted, this probability increases
  - P(type I error in one test) = α
  - P (no type I error in one test) = $1 - \alpha$
  - P (no type I error in *m* tests) = $(1 - \alpha)^m$
  - P (at least one type I error in *m* tests) = $1 - (1 - \alpha)^m$
- two solutions
  - correct for multiple comparisons
  - do an “overall” test before jumping in



F test aka ANOVA!

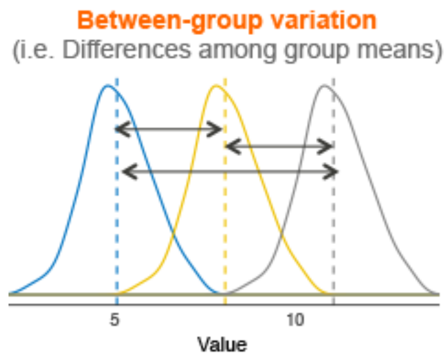
- we already saw how an analysis of variance / F-test can help us assess "overall" fit of the model
- formally, ANOVA is a *generalized* t-test for more than two means/groups!
- we first evaluate whether the overall model explains variance over and above random chance
  - $F = \frac{MS_{model}}{MS_{error}} = \frac{SS_{model}/df_{model}}{SS_{error}/df_{error}}$
  - If F > 1, the group differences are greater than what would be expected as random variation within groups
- if this test is significant, we then go in to look for pairwise differences between groups

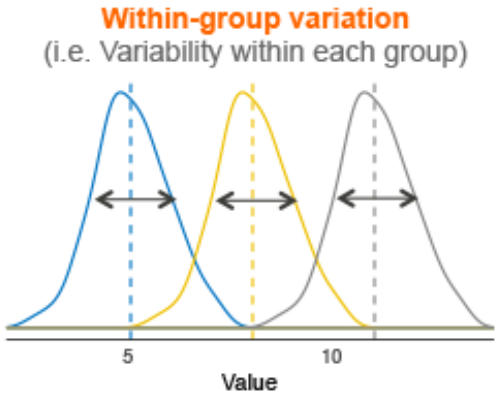




types of ANOVAs = complex linear models

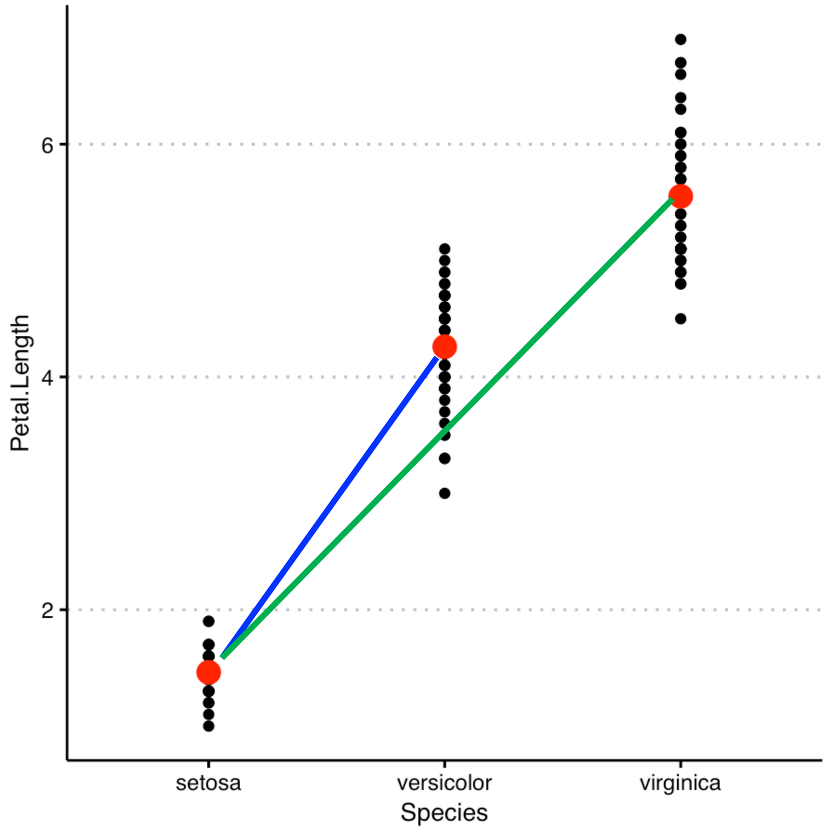
- n (independent variables)
    - one-way: one independent variable
    - two-way / three-way
    - (n>3)-way: crazy land
- within or between subjects
    - between subjects: regular ANOVA
        - independent observations: each raw score comes from *different* individuals!
    - within-subjects: repeated measures ANOVA
        - non-independent observations: multiple raw scores from from the same individuals





revisiting iris

- when more than two groups are involved, we need to expand our model to include multiple groups
- petal length = a + b (virginica) + c (versicolor)
  - a = M_{setosa}
  - b = $M_{virginica} - M_{setosa}$
  - c = $M_{versicolor} - M_{setosa}$



NHST for one-way ANOVA

- **step 1: state the hypotheses**
  - H_0: no change in mean petal lengths due to species, i.e., $\mu_{setosa} = \mu_{virginica} = \mu_{versicolor}$
  - H_1: there is at least one mean difference (no claims about where!)
- **step 2: set criteria for decision**

$$F(df_1, df_2) = F_{critical}$$

- **step 3: collect data**
- **step 4: make a decision!**

NHST for one-way ANOVA

step 1: state the hypotheses	step 2: set criteria for decision	step 3: collect data	step 4: make a decision!
$H_0: \mu_1 = \mu_2 = \cdots = \mu_n$ $H_1: at\ least\ one\ mean\ difference$	$\alpha = .05$ find $F_{critical}$ based on **right** tailed test and degrees of freedom $df_1 = k - 1$ $df_2 = n - k$	(1) compute SS_{model} and SS_{error} (2) compute $F_{observed} = \frac{MS_{model}}{MS_{error}}$ (3) find p-value for F-score	check whether F is beyond $F_{critical}$ and p-value < .05. if so, reject null hypothesis!

step 2: set criteria for decision

- we need to find the values of df_1 and df_2 and the corresponding critical F value

$$F(df_1, df_2) = F_{critical}$$

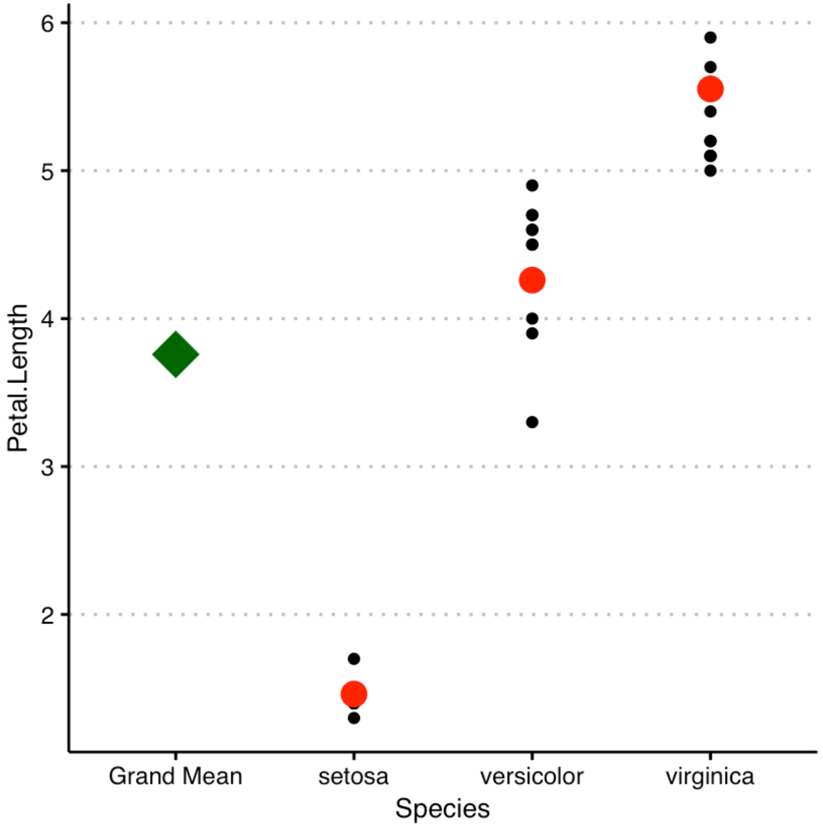
- k denotes the number of levels of the independent variable OR estimated parameters
  - $df_{model} = k - 1 = 3 - 1 = 2$
  - $df_{error} = n - k = 30 \; - 3 = 27$
- $F_{critical} = F(df_1, df_2)$ at α = 0.05
- $F(2{,}27)$ at α = 0.05 = 3.59

step 3: obtaining $F_{observed}$

- we need to obtain estimates of

$$F_{observed} = \frac{MS_{model}}{MS_{error}} = \frac{SS_{model}/df_{model}}{SS_{error}/df_{error}}$$

- $SS_{total} = SS_{model} + SS_{error}$
    - SS_{total} represents error left over after grand mean
    - SS_{error} represents error left over after species mean model
    - $SS_{model} = SS_{total} - SS_{error}$



step 3: obtaining $F_{observed}$

- SS_{total} represents deviations from grand mean (M_Y)

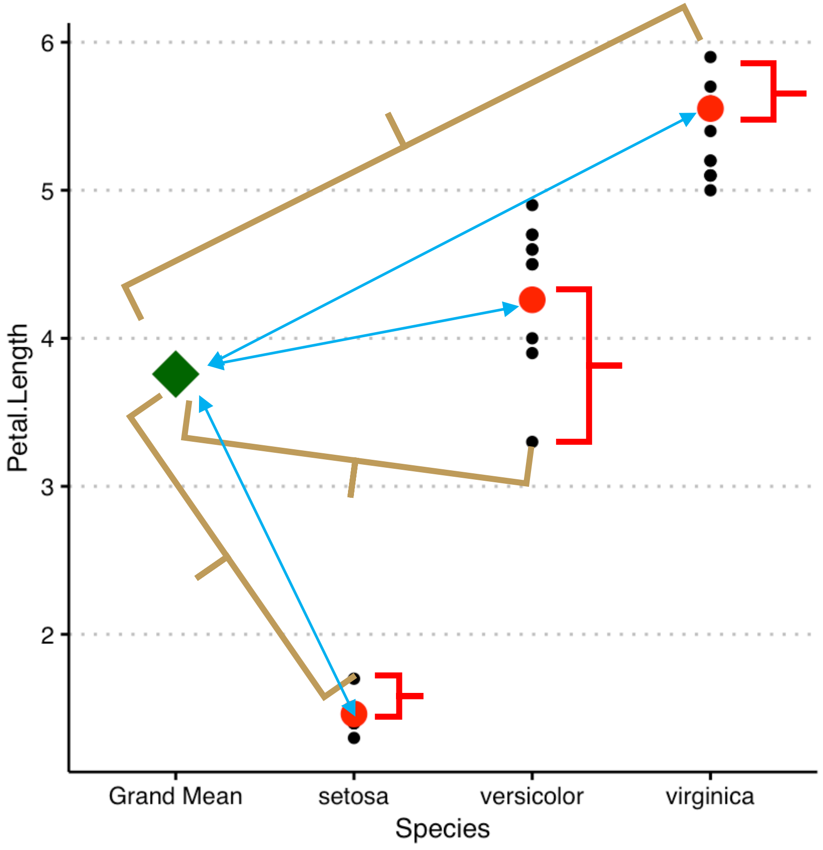
$$SS_{total} = \sum (Y - M_Y)^2$$

- SS_{error} represents the deviations of each score from its group mean

$$SS_{error} = \sum (Y - \hat{Y})^2 = \sum (Y - M_{group})^2$$

- SS_{model} represents the ***gains*** we get if we substitute each score with the group mean instead of the grand mean

$$SS_{model} = \sum \sum n_i \, (M_{group} - M_Y)^2 = SS_{total} - SS_{error}$$



step 3 and 4: obtaining F & decisions

- calculate each of these for iris
  - SS_{total}= 84.76
  - SS_{error} = 3.087
  - SS_{model}= 81.67
- calculate $F_{observed} = \frac{MS_{model}}{MS_{error}} = \frac{SS_{model}/df_{model}}{SS_{error}/df_{error}} = 357.1778$, $p < .0001$
- $F_{critical}$ = 3.59 at α = 0.05
- therefore, the null hypothesis can be rejected! i.e., species differ in their petal lengths
- $F(2,27) = 357.18, p < .0001$

F-table

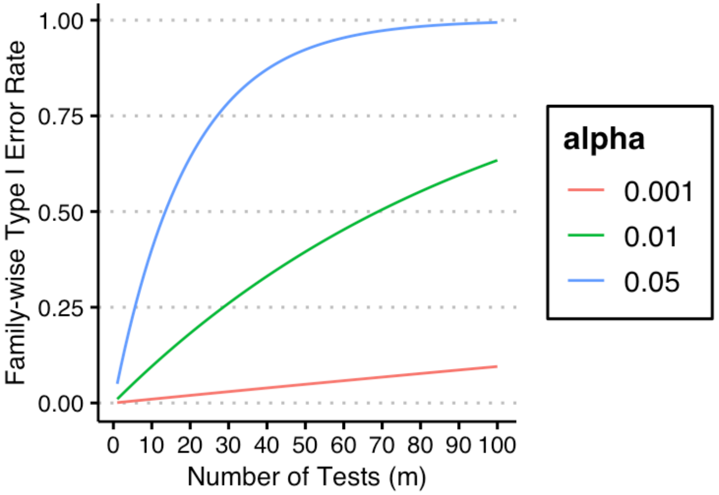
		SS	df	MS	F	p-value
SS_{model}	species	81.67	2	75.272	357.1778	<.0001
SS_{error}	residual	3.087	27	0.0526		

```
Response: Petal.Length
          Sum Sq Df F value    Pr(>F)
Species   81.675  2  357.18 < 2.2e-16 ***
Residuals  3.087 27
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Sheets solution video tutorial

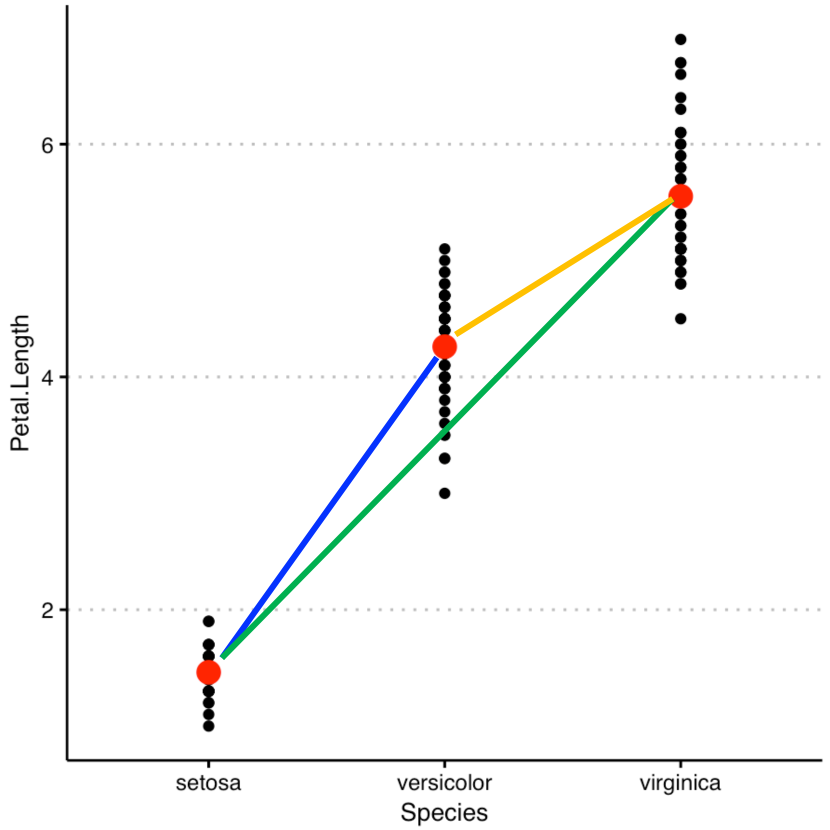
what now?

- once we know that the "overall" test has detected something meaningful, we can look for specific differences by conducting pairwise t-tests
    - $M_{virginica} - M_{setosa}$
    - $M_{versicolor} - M_{setosa}$
    - $M_{virginica} - M_{versicolor}$
- BUT...what about the type I error??
- we correct for multiple comparisons



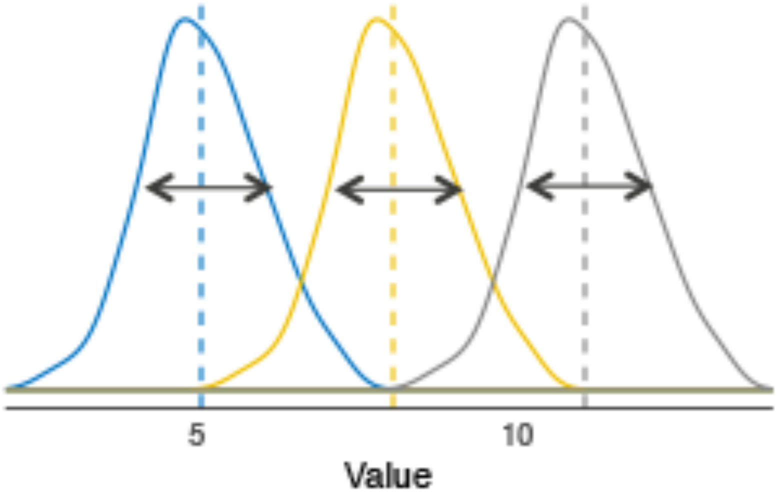
post-hoc tests

- when pairwise t-tests are conducted after an “overall” / omnibus test (ANOVA), they are called post-hoc tests
- several corrections exist in the literature
  - Tukey’s Honest Significant Difference Test: moderately conservative
  - Scheffe’s test: very conservative
  - Fisher LSD: very liberal
- most statistical software will allow you to apply a correction, so we will not cover the specifics
- visual inspection is useful in these situations



one-way ANOVA assumptions

- independent observations within each sample
- normality
- homogeneity of variances



next time

- **before** class
  - *watch:* Hypothesis Testing (two-groups: F-test) [12 min]
  - *watch:* Hypothesis Testing (one-way ANOVA: F-test) [18 min]
  - *read*: [optional] Read Chapter 14 from the Gravetter & Wallnau (2017) textbook.
  - *start*: Problem Set #6 (Chapter 10 and Chapter 12 problems)
- **during** class
  - two independent variables / two-way ANOVA