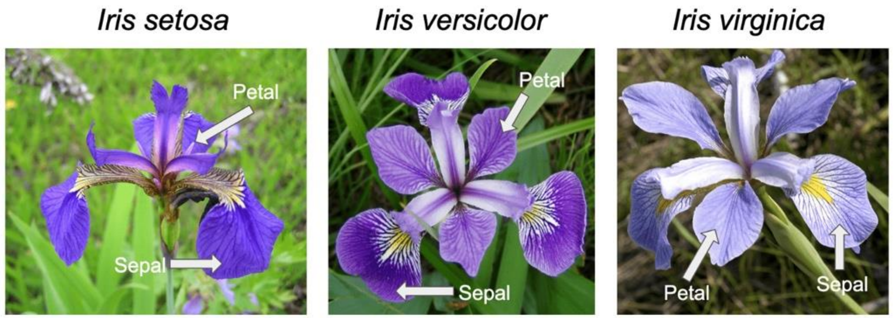
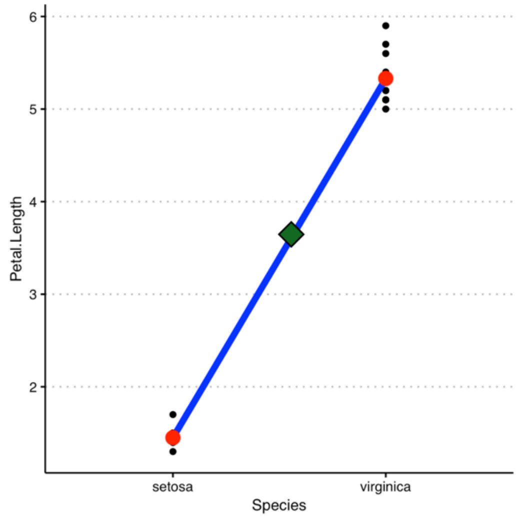
DATA ANALYSIS

Week 11: Special cases (independent t-test)

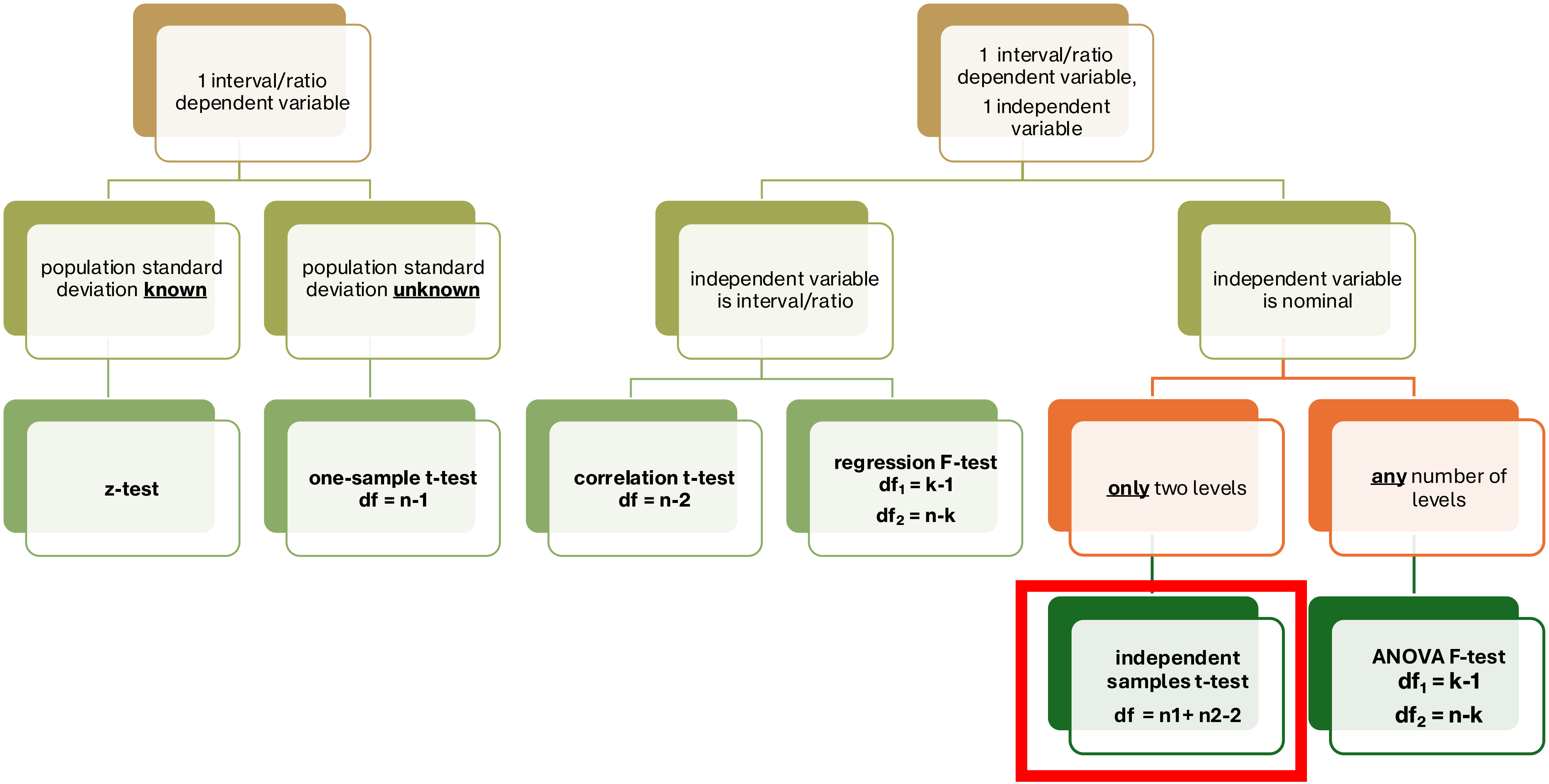
review: iris dataset

- our goal is to build the best model for petal lengths
- petal length (Y) = a + b (species:X)
- we conducted an ANOVA to evaluate the significance of adding “species” means to our model over and above the grand mean
- if we have **only two groups/levels being compared**, the general ANOVA approach can be reduced to an independent samples t-test



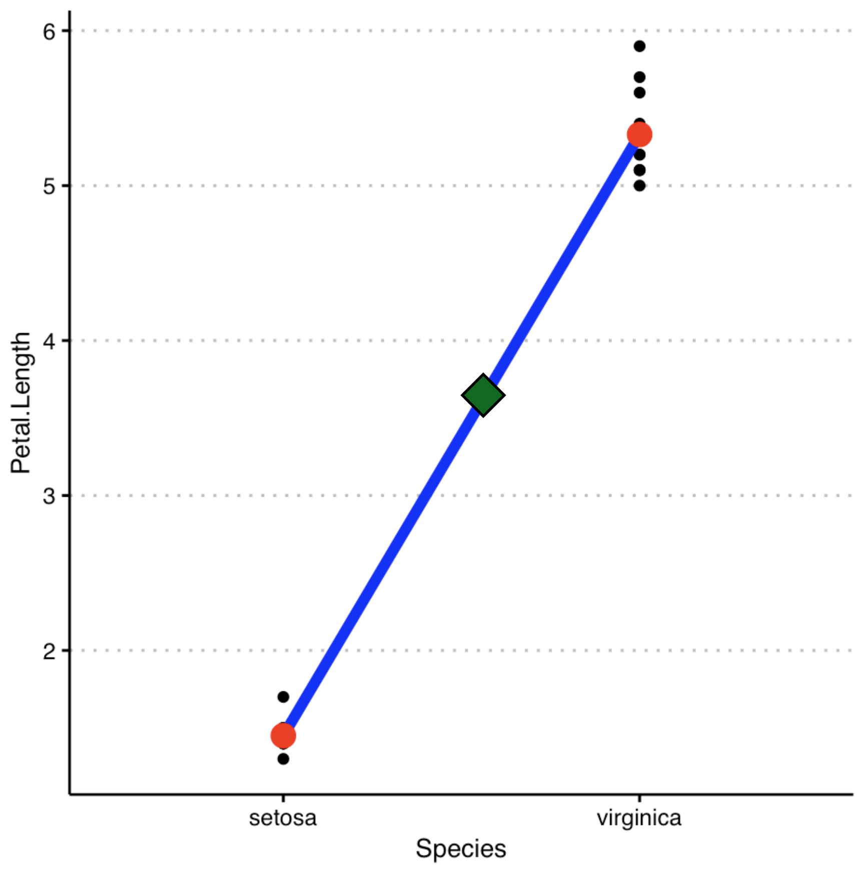


hypothesis testing flowchart



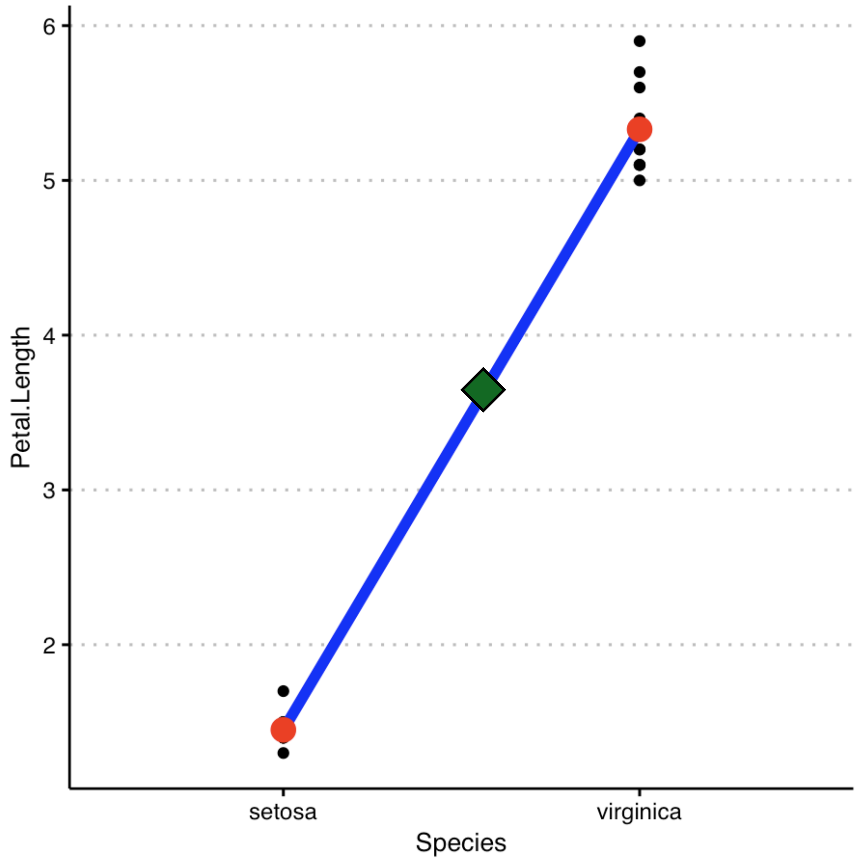
comparing two groups

- Y (petal lengths) = a + b X (species) + error
- X is now an indicator of *levels* of the independent variable
  - when X= 0 (first group), Y = a
  - when X = 1 (second group), Y = a + b
- the intercept is the mean for the first group (X = 0)
  - a = M_{group1}
- the slope is the change in means from first to second group (X = 1)
  - b= $M_{group2} - M_{group1}$
- essentially, we are fitting a line to the data but slope and intercept are the group means!



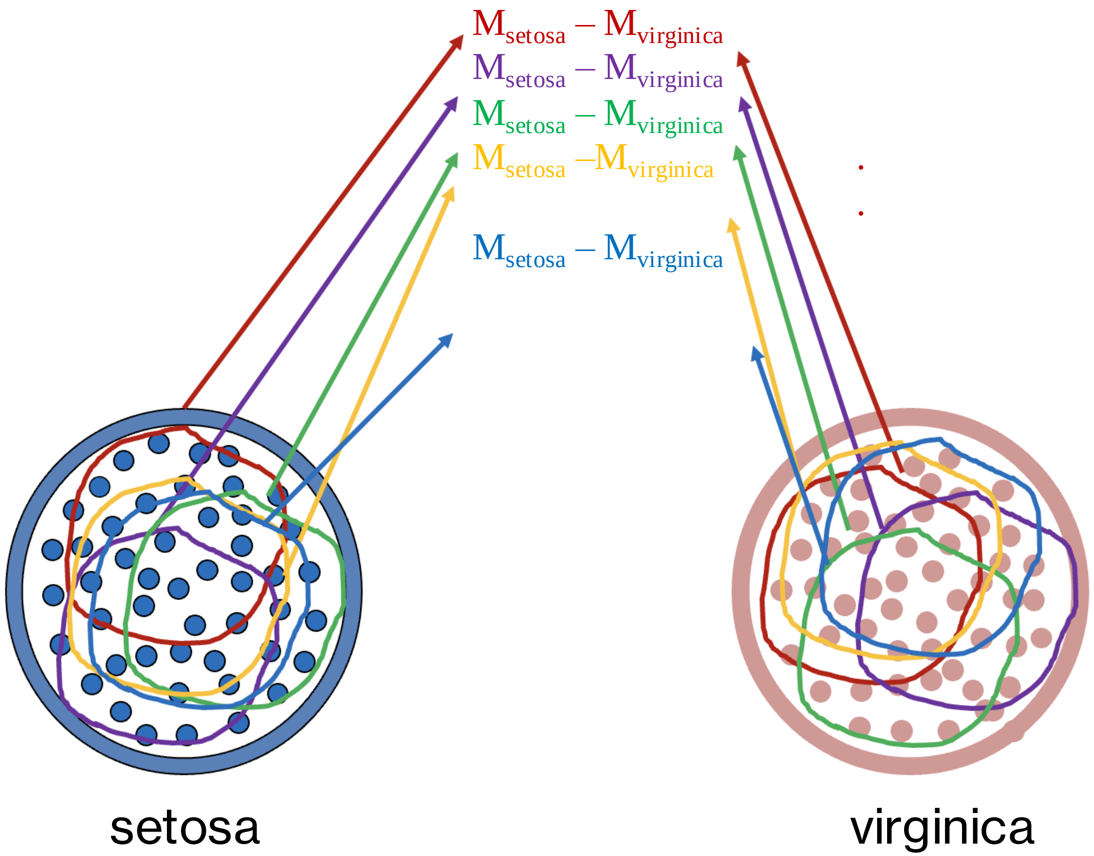
comparing two groups

- we have the slope and intercept formulation
  - a = M_{group1}
  - b= $M_{group2} - M_{group1}$
- we compute the slope and intercept for iris data
  - a = M_{setosa} = 1.45
  - b = $M_{virginica} - M_{setosa}$ = 5.33 – 1.45 = 3.88
- line's equation is:
  - Y = 1.45 + 3.88 (X) where X = {0,1}
- as with regression testing, we will test whether the "slope" is significant, i.e., is this difference $M_{virginica} - M_{setosa}$ significant?



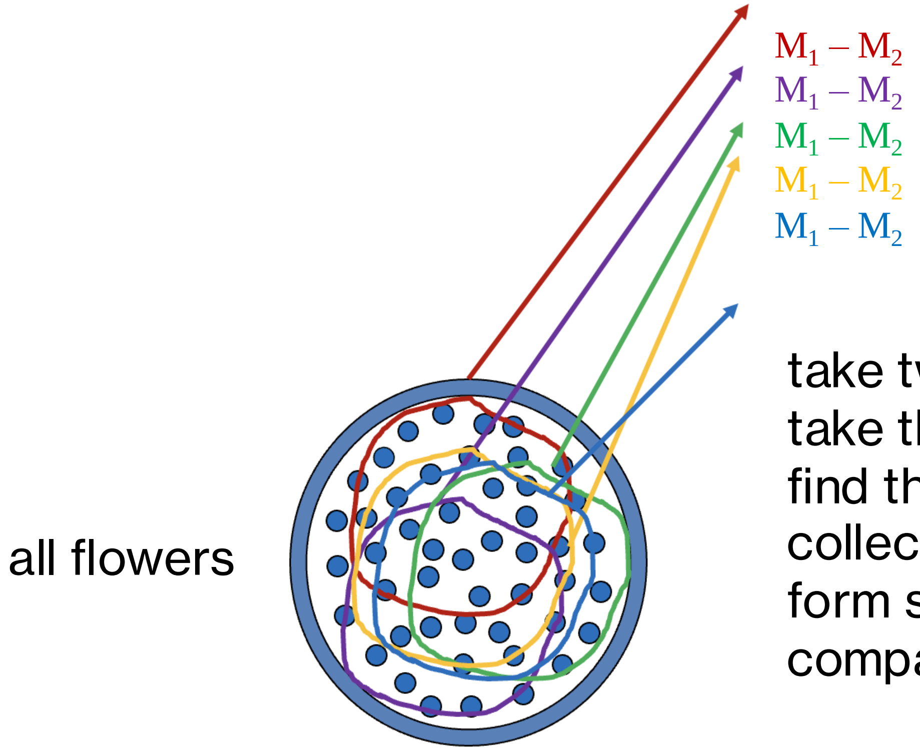
t-test for two independent groups

alternative hypothesis: flowers come from different populations,
i.e., the observed difference between means is **larger** than what would be expected under the null



t-test for two independent groups

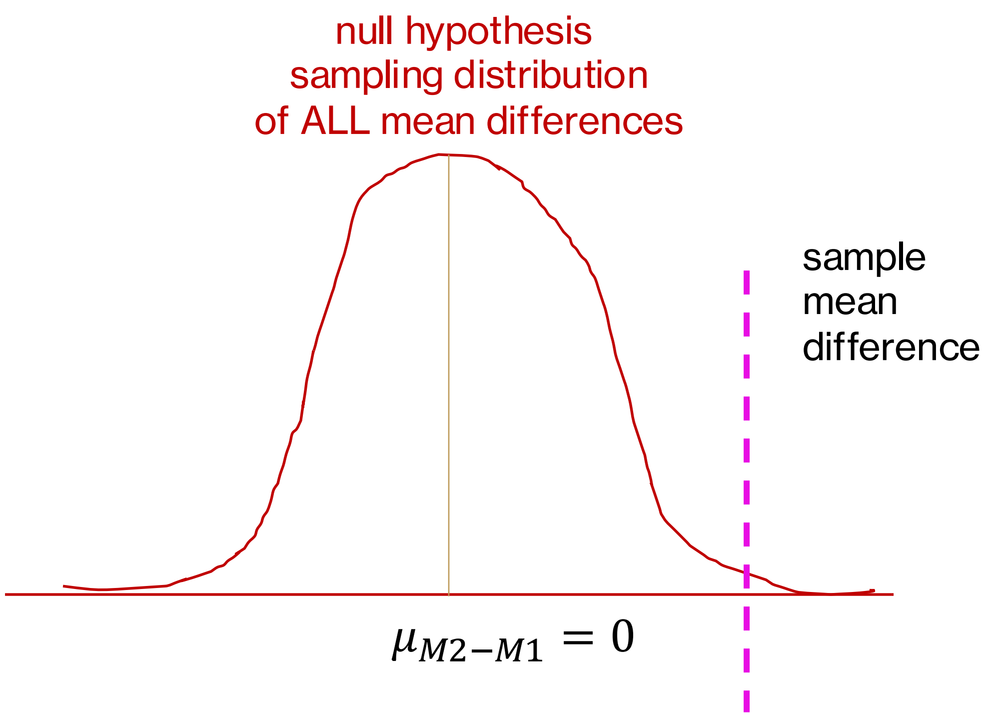
null hypothesis: flowers don't come from different populations, i.e., obtained difference between means is not too different from what would be expected if two samples were drawn from the same population



take two random samples (1 and 2)
take the mean of each sample (M_1 and M_2)
find the difference in means ($M_1 - M_2$)
collect ALL such samples of size n
form sampling distribution of **mean differences**
compare to observed mean difference from actual data

t-test for two independent groups

- under the null hypothesis, this sampling distribution of mean differences for all pairs of possible samples of size n follows a t-distribution with a mean μ_{M2-M1} = 0
- next, we need to figure out degrees of freedom for this t-distribution and standard error (or standard deviation of this sampling distribution)



independent measures t-test

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

$$t_{df} = \frac{sample\ statistic\ (M_2 - M_1) - population\ parameter\ (\beta)}{standard\ error} = t_{critical}$$

- step 3: collect data

$$t_{df} = \frac{(M_2 - M_1) - 0}{SE}$$

- step 4: make a decision!

step 2a: setting criteria for t-test

- to set criteria for decision, we need to define which t-distribution we will use
- for each group, since we will estimate a mean, we take away 1 degree of freedom

$$df = df_1 + df_2 = (n_1 - 1) + (n_2 - 1) = n_1 + n_2 - 2$$

$$t_{n_1 + n_2 - 2} = t_{critical}$$

step 2b: standard error for t-test

- how do we calculate standard error? typically, when we had data from a single sample of scores, we looked at the “average” deviation from the mean

$$one\ sample\ t = \frac{M - \mu}{s_M} \text{ and } s_M = \frac{s}{\sqrt{n}}$$

- we now have two different samples, so the standard error needs to take into account the variation from both samples (it is the average error to be expected between any two sample means under the null hypothesis)
- $s_{M_2 - M_1}$: standard error for independent samples t-test

step 2b: pooled variance

- when $n_1 = n_2$, $s_{M2-M1} = \sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}$ (equally weight both samples)
- when $n_1 \neq n_2$, bigger samples should yield more accurate estimates of population variance than smaller samples and this formula may not be appropriate
- an estimate of pooled variance can be computed in such cases

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

- then we can compute standard error as follows:

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

step 3 and 4: collect data + decide

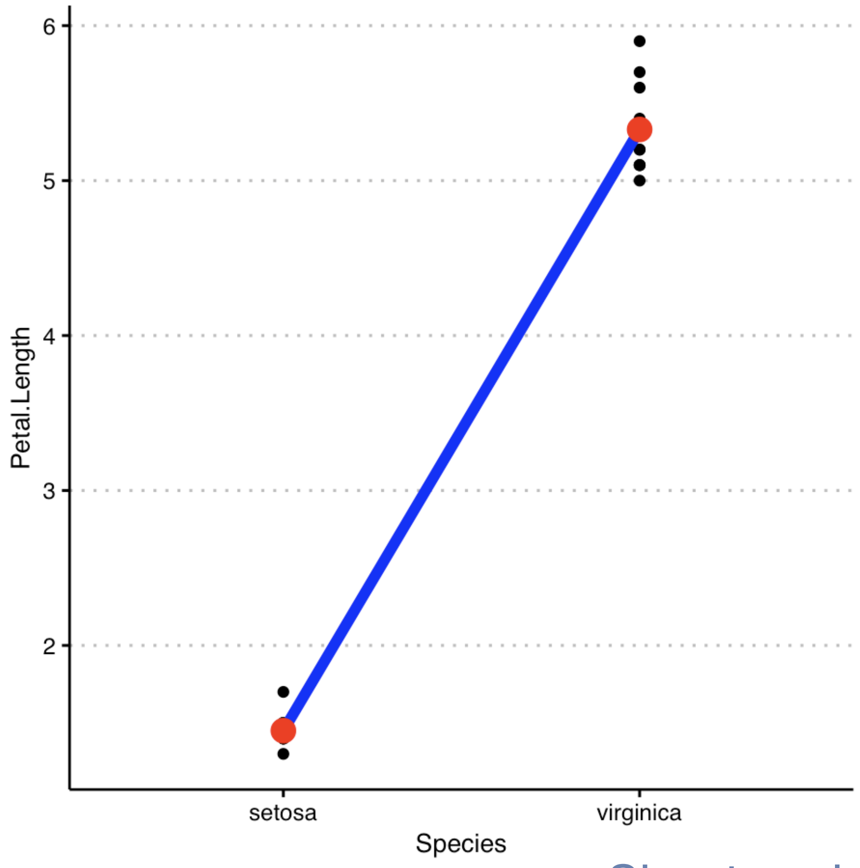
- we compute the standard error for our sample

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

- we compute $t_{observed} = \frac{M_2 - M_1}{s_{M_2 - M_1}}$

$$t_{18} = \frac{b - \beta}{standard\ error} = \frac{b - 0}{s_{M2-M1}} = \frac{(M_2 - M_1) - 0}{s_{M_2 - M_1}} = \frac{3.88}{.1025} = 37.845$$

- $t_{18} = 37.845, p < .0001$
- therefore, we reject the null hypothesis that the petal lengths do not differ by species, i.e., species have different petal lengths!



Sheets solution video tutorial

NHST for two independent groups (t-test)

step 1: state the hypotheses

H_0: β = 0 or $\mu_2 - \mu_1 = 0$

H_1: β ≠ 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 $M_2 - M_1$

(2) compute $t_{observed} = \frac{(M_2 - M_1) - \beta}{s_{M2-M1}}$

(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**

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

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

$$M_2 - M_1 = 3.88$$

$$t_{observed} = \frac{(M_2 - M_1) - 0}{s_{M2-M1}} = \frac{(3.88) - 0}{.1025} = 37.844$$ 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

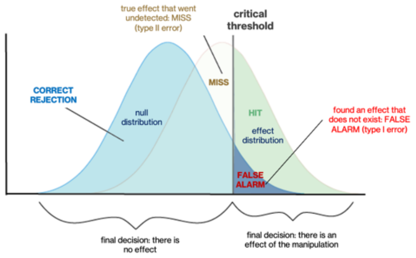
- interval/ratio dependent variable
- independent observations (between-subjects design)
  - when observations are not independent, then we conduct a paired/dependent-samples t-test (later!)
- normality
  - when data are not normal, the t-test is not appropriate
  - BUT: t-tests are fairly robust to 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?

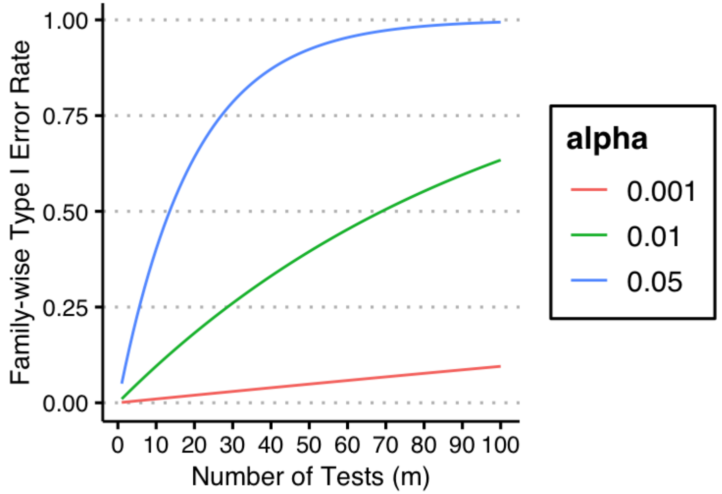
W11 Activity 3

- complete on Canvas

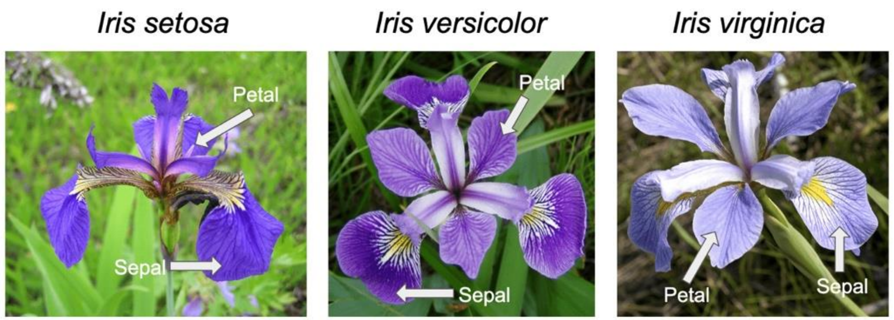
multiple tests and type I errors



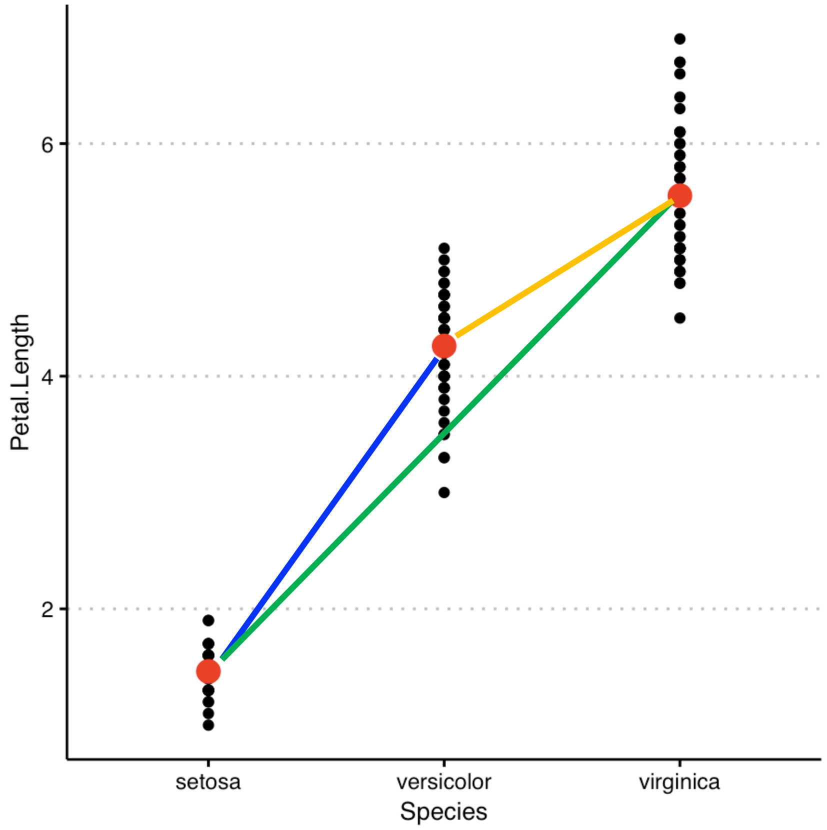
- each time a hypothesis test is conducted with some α-level, there is α % probability of making a type I error
- as more tests are conducted, this probability increases
  - P(type I error in one test) = α
  - P (no type I error in one test) = $1 - \alpha$
  - P (no type I error in *m* tests) = $(1 - \alpha)^m$
  - 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



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



instead: do overall 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

iris: overall ANOVA results

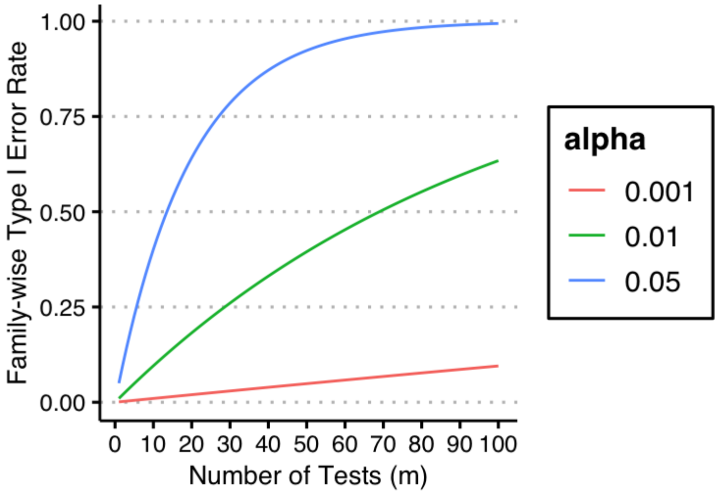
		SS	df	MS	F	p-value
SS_{model}	species	81.67	2	75.272	357.1778	<.0001
SS_{error}	residual	3.087	27	0.0526		

```
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

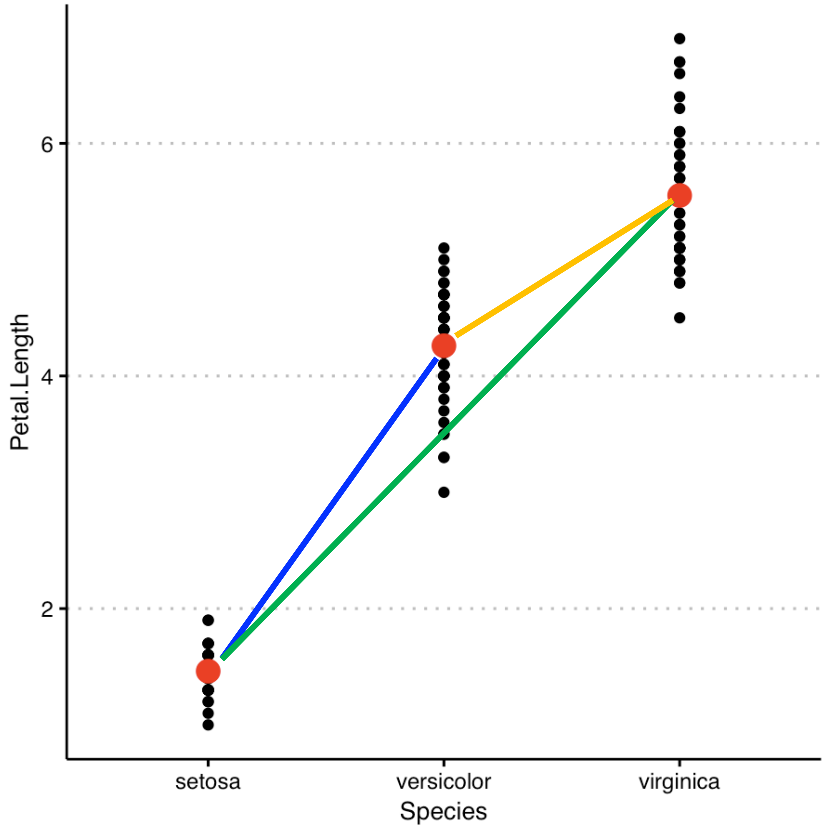
what now?

- 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



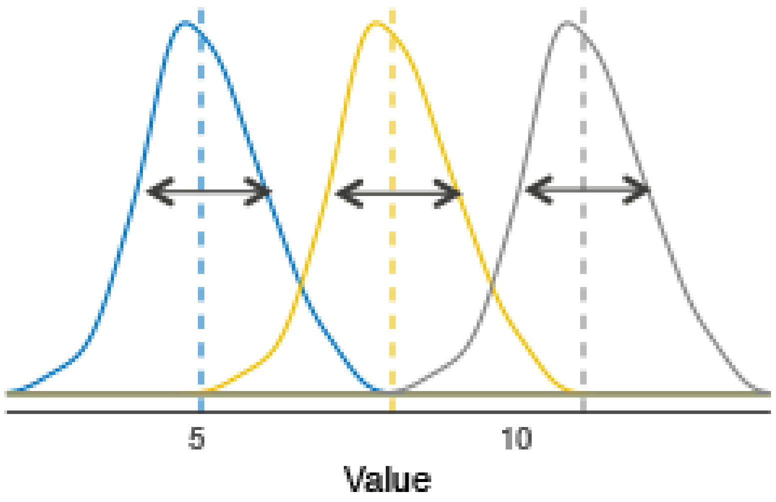
post-hoc tests

- 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

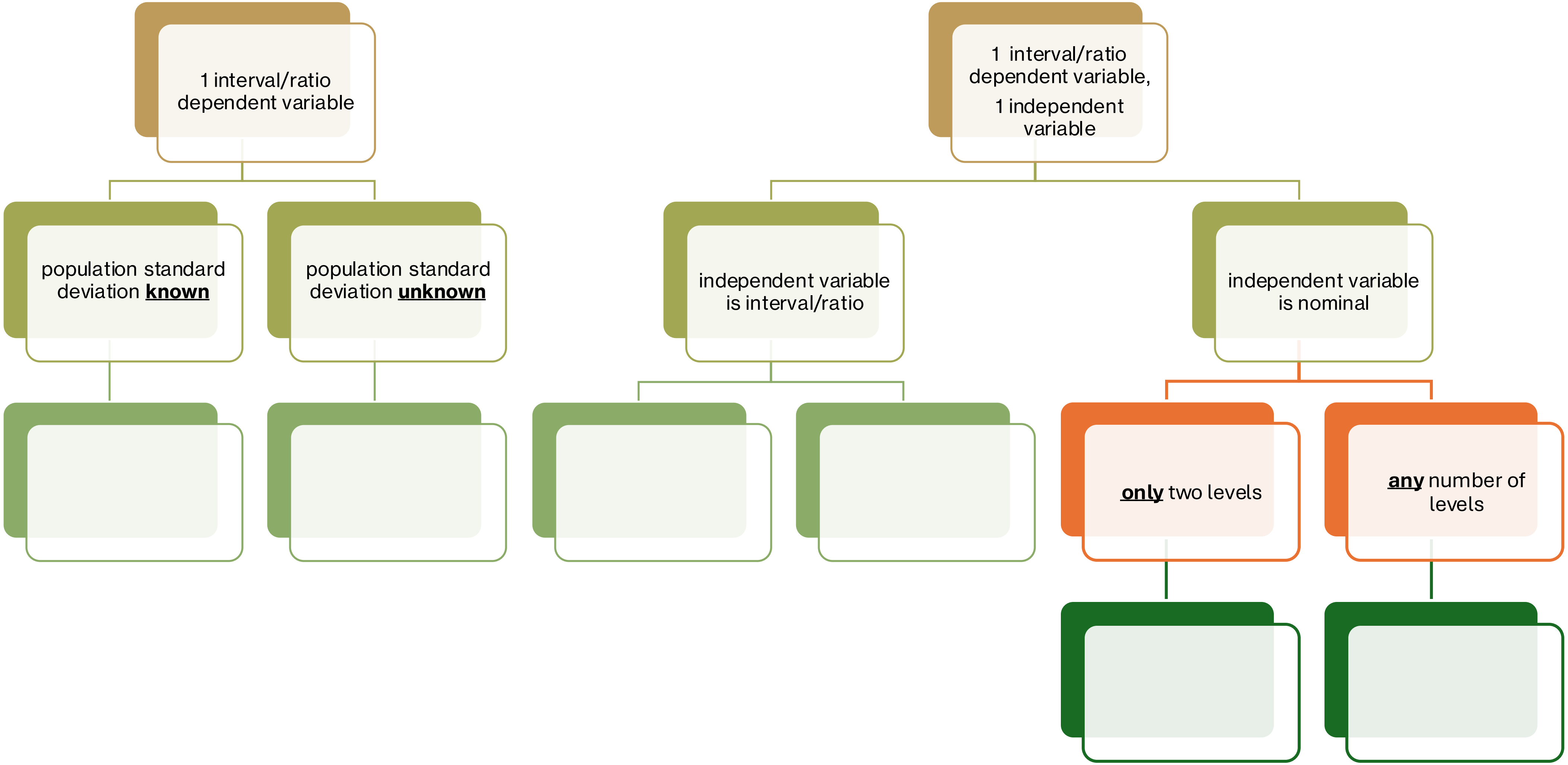


one-way ANOVA assumptions

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



hypothesis testing flowchart

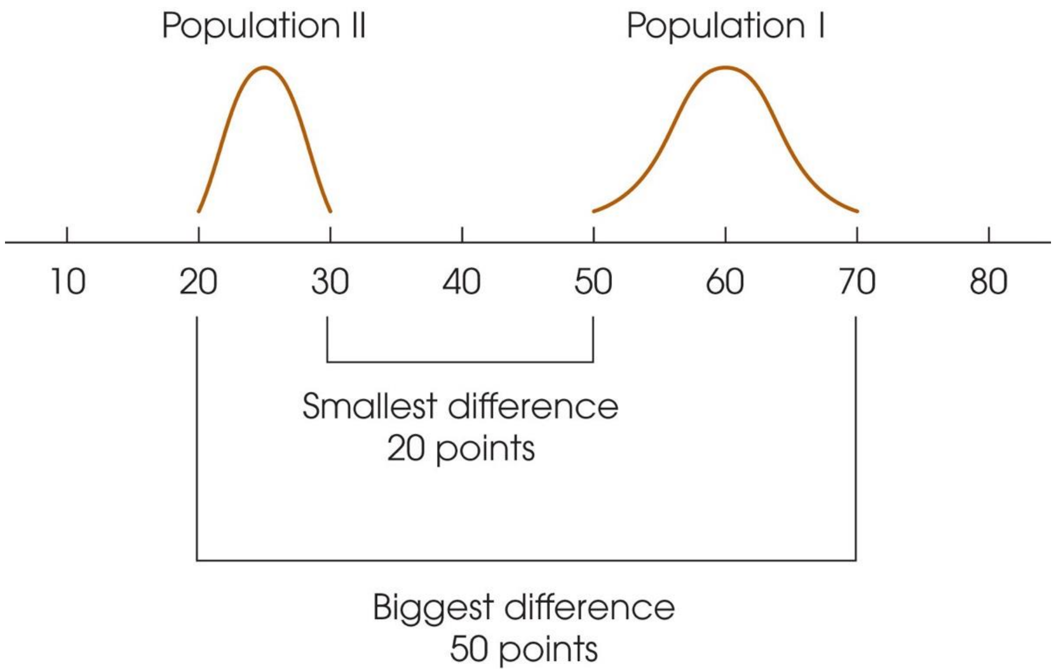


hypothesis testing process

- fill out document before exam
- for conceptual exam, you may bring:
    - flowchart
    - test process document
    - one handwritten help sheet

optional: why add the standard errors?

- range of scores in population I and II?
  - population I: 70-50 = 20
  - population II: 30-20 = 10
- when we take differences between means from population I and II, what is the largest difference possible?
  - largest: 70-20= 50
  - smallest: 50-30 = 20
- standard error of mean differences = range of population I + range of population 2! = 20+10 = 30!



next time

Here are the to-do's for this week:

- Submit Week 11 Quiz
- Submit Problem Set 5
- Complete Practice Midterm 2 (Conceptual)
- Submit any lingering questions here!
- Extra credit opportunities:
    - Submit Exra Credit Questions
    - Submit Optional Meme Submission

Please complete the following BEFORE class

Tuesday

- Complete Practice Midterm 2 (Conceptual)
- Submit any lingering questions here!
- Complete Practice Midterm 2 (Computational) (solutions have been included, but please attempt the practice on your own before you look at the solutions)

Thursday

- Complete Midterm 2 (Conceptual): IN CLASS

After Thursday

- Submit Midterm 2 (Computational): Due Monday midnight, NO late submissions will be graded