Module 3: ANOVA & Correlation

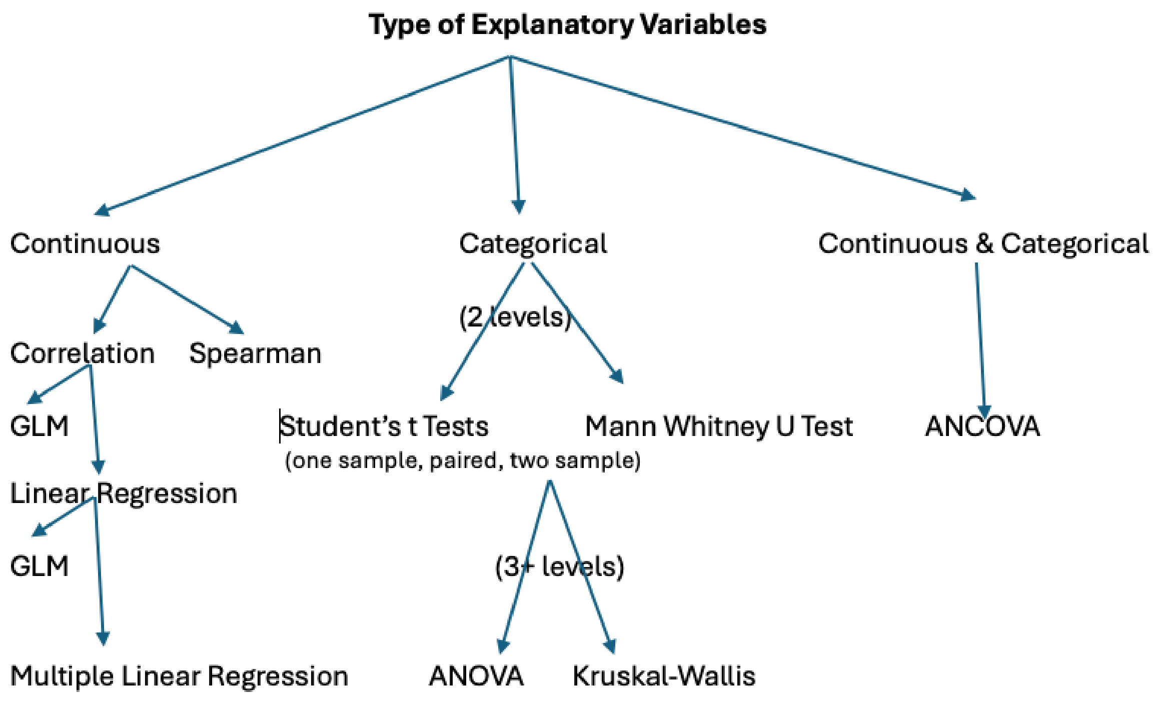
Assigning signal and noise to variation

Agenda:

1. ANOVA: Nuts & Bolts

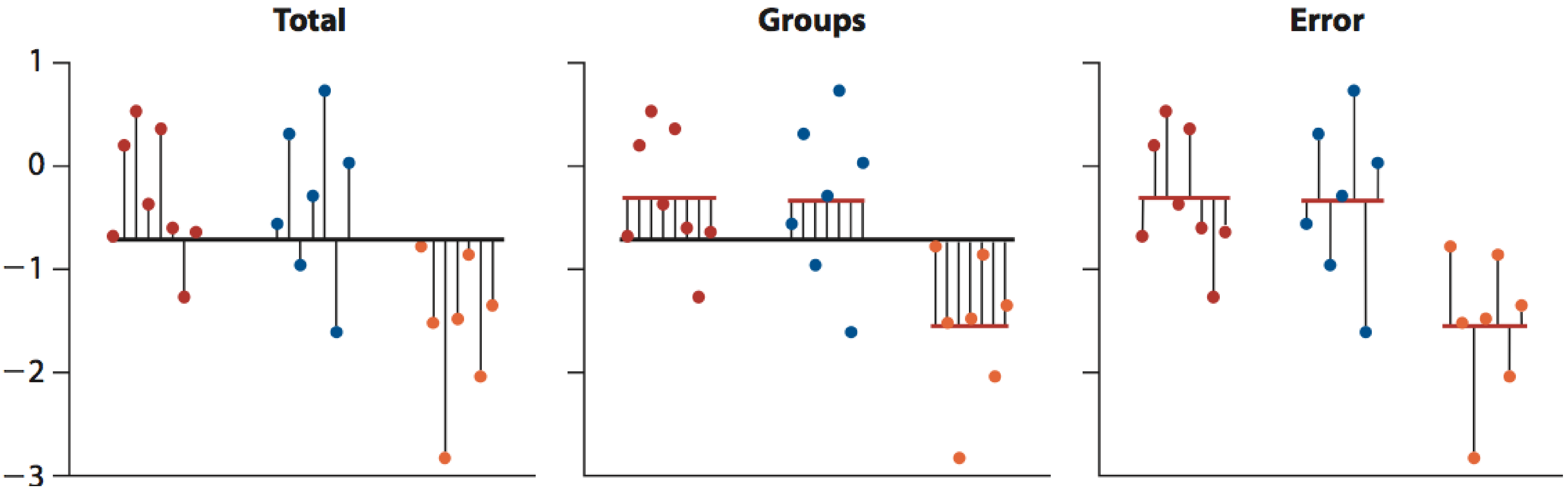
2. Worked Example
   A. One way ANOVA
   B. Post-hoc tests: Tukey-Kramer
   C. Kruskal-Wallis (nonparametric)

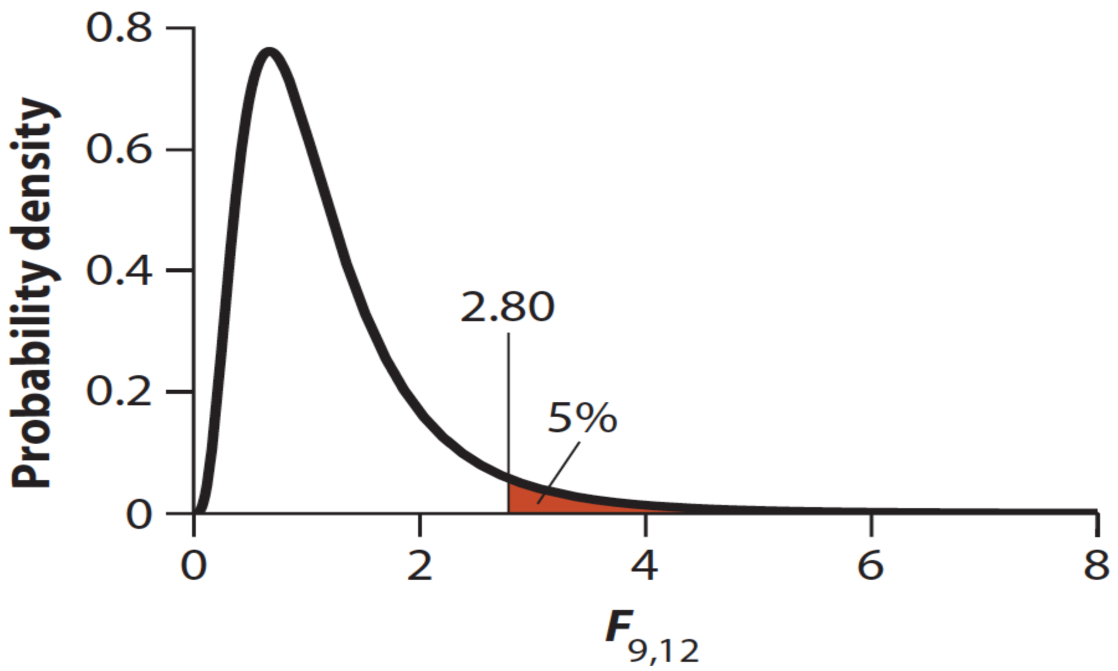
3. Linear Correlation
   A. Spearman’s rank



Results are presented in ANOVA Table:

Source of variation	Sum of Squares	df	Mean Squares	F-ratio	P
Groups (treatment)					
Error					
Total					





$$\text{F-value} = \frac{MS_{group}}{MS_{error}}$$

SIGNAL

NOISE

- This is a <u>**one-sided test**</u> which is different from the F test that we used previously to test variances between populations.

- ANOVA F test is one-sided because MS_{group} is ALWAYS in the numerator (there isn't a 50:50 chance like in the F test for equal variances).

Data Dredging:

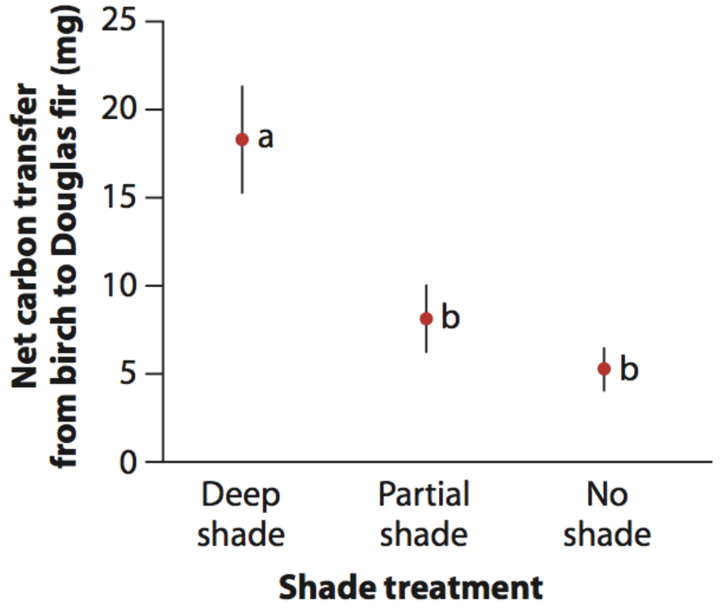
When you use multiple tests on a data set, the **actual** probability of making **at least one** type I error, α, is larger than the significance level states

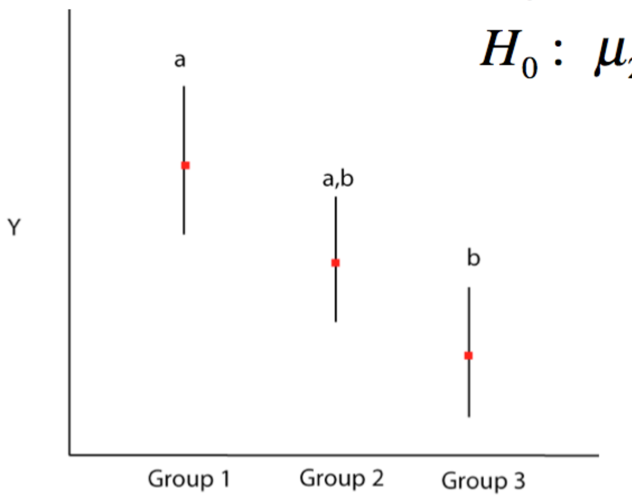
- each hypothesis test has a probability of error and these errors compound as more tests are conducted

- Example: two independent studies are performed to test the same null hypothesis. What is the probability that at least one study obtains a significant result and rejects the null hypothesis **even if the null hypothesis is true**? Assume that in each study there is a **0.05** probability of rejecting the null hypothesis (Answer is **0.0975**)

P(No type I errors)= $(1-\alpha)^N$*, where N = independent tests*
P(≥1 type I error)= $1-(1-\alpha)^N$

How Tukey-Kramer results are displayed:

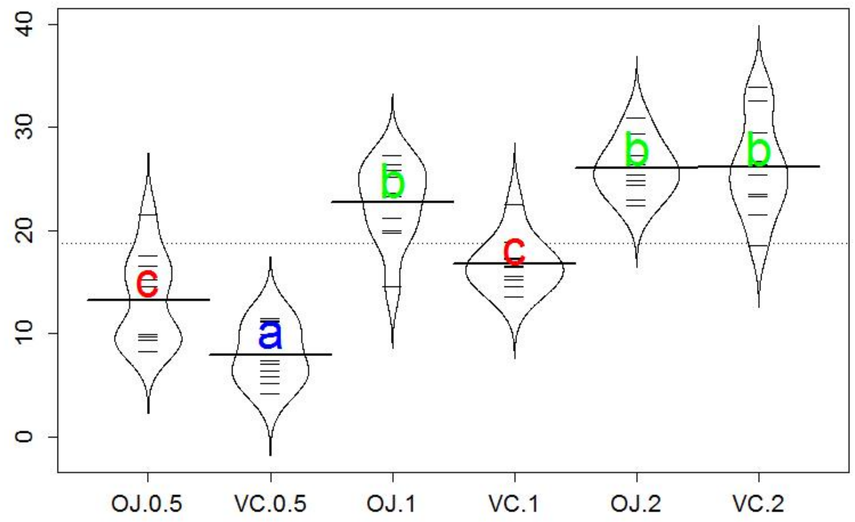




$H_0 : \mu_1 = \mu_2$ Cannot reject

$H_0 : \mu_1 = \mu_3$ Reject

$H_0 : \mu_2 = \mu_3$ Cannot reject



Compact Letter Display

Kruskal-Wallis Test:

o A non-parametric test similar to a single factor ANOVA

o Uses the **ranks** of the data points; tests **medians** not means

- Data points are not compared, their ranks are!

*Using **ranks** is what frees us from having to assume normality since all distributions have similar predictions about ranks*

- All group samples are random samples
- Distribution of the variable has the same shape in every population
- Small samples lead to little power but when n is large, Kruskal-Wallis has the same power as ANOVA

o ***H***, sampling distribution is χ^2 with df = k - 1

https://datatab.net/tutorial/kruskal-wallis-test#examples

Fixed Effects: The groups *are* the question

- Also called Model 1 ANOVA
  - What we have been using so far
- Different categories of explanatory variable are predetermined and repeatable
  - **Results cannot be generalizable**
  - Example: specific drug treatment, specific diets, specific season

Random Effects: The groups are a source of noise in the system you are modeling

- Also called Model 2 ANOVA
- Different categories of explanatory variable are *randomly sampled from a larger population of groups*
  - **Results are generalizable;** conclusions reached about difference among groups can be generalized to the whole population
  - Example: family in a study about resemblance of IQ
    - Chose a random family in a population of families
    - Family: group
    - Replicates: different children within each family
  - **The population and not the particular families involved is the target of study**

Quick heuristic:

Ask: “Do I want to estimate the mean of each group?”

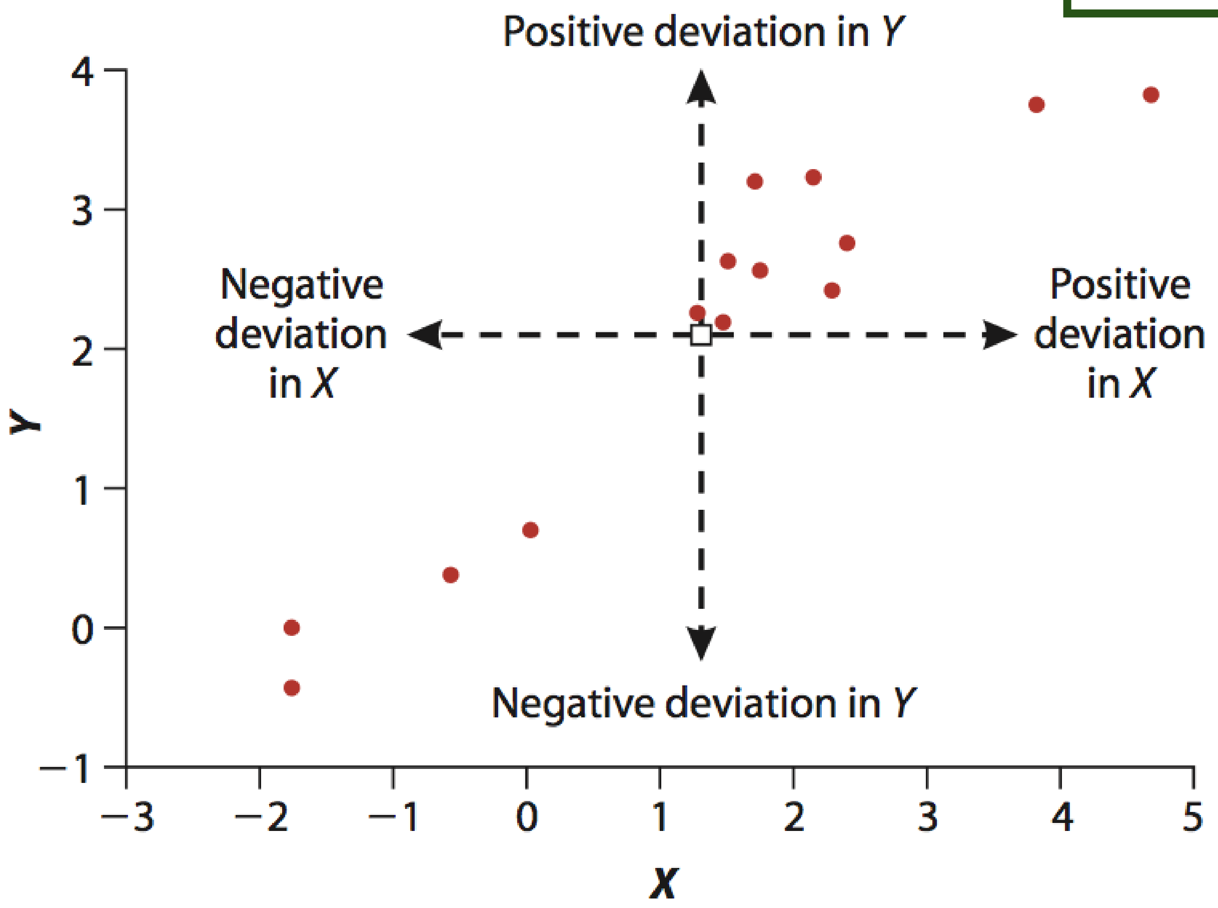
Yes → Fixed effect
No → Random effect

Ask: “Do these group levels represent all the possibilities or just a sample?”

All (or all that matter) →Fixed effect
Just a sample of many possible levels →Random effect

(Pearson) Correlation Coefficient

$$r = \frac{\overbrace{\square\,(X-\bar{X})(Y-\bar{Y})}^{\text{SIGNAL}}}{\underbrace{\sqrt{\square\,(X-\bar{X})^2}\sqrt{\square\,(Y-\bar{Y})^2}}_{\text{NOISE}}} = \frac{Co\,\mathrm{var}\,iance(X,Y)}{s_x s_y}$$



Testing for no correlation:

Step 1: declare null and alternate

H_0: Zero correlation ($\rho=0$)

H_A: Some correlation ($\rho\neq0$)

Step 2: test statistic

$$t = \frac{r - \rho}{SE_r}$$

$$\boxed{SE_r = \sqrt{\frac{1-r^2}{n-2}}}$$

Spearman's rank correlation:

Test for correlation in the normal way….

Step 1: declare null and alternate

H_0: Zero correlation (ρ_s=0)
H_A: Some correlation (ρ_s≠0)

Step 2: test statistic

$$\boxed{r_s = \frac{\sum (R-\bar{R})(S-\bar{S})}{\sqrt{\sum (R-\bar{R})^2}\sqrt{\sum (S-\bar{S})^2}}}$$

Step 3:State α/P-value/Critical value

Table or computer!

Step 4: State conclusion

Add correlation and ANOVA to your flowchart