

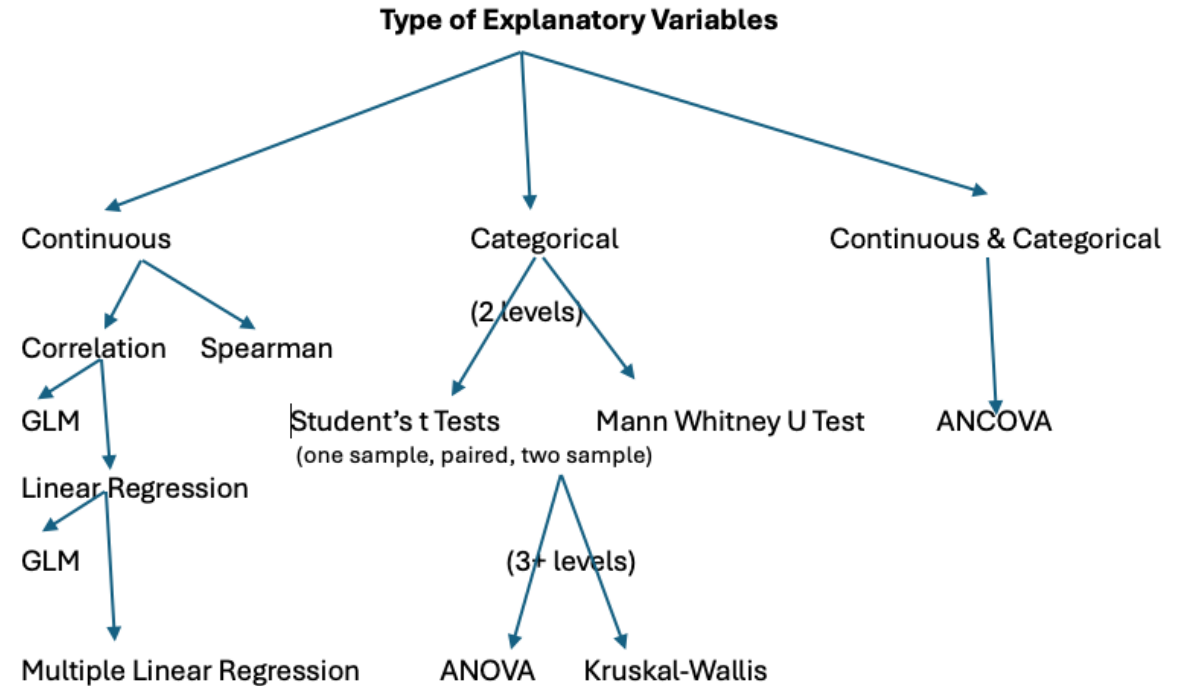
Module 3AB:

ANOVA & Correlation

Assigning signal and noise to variation

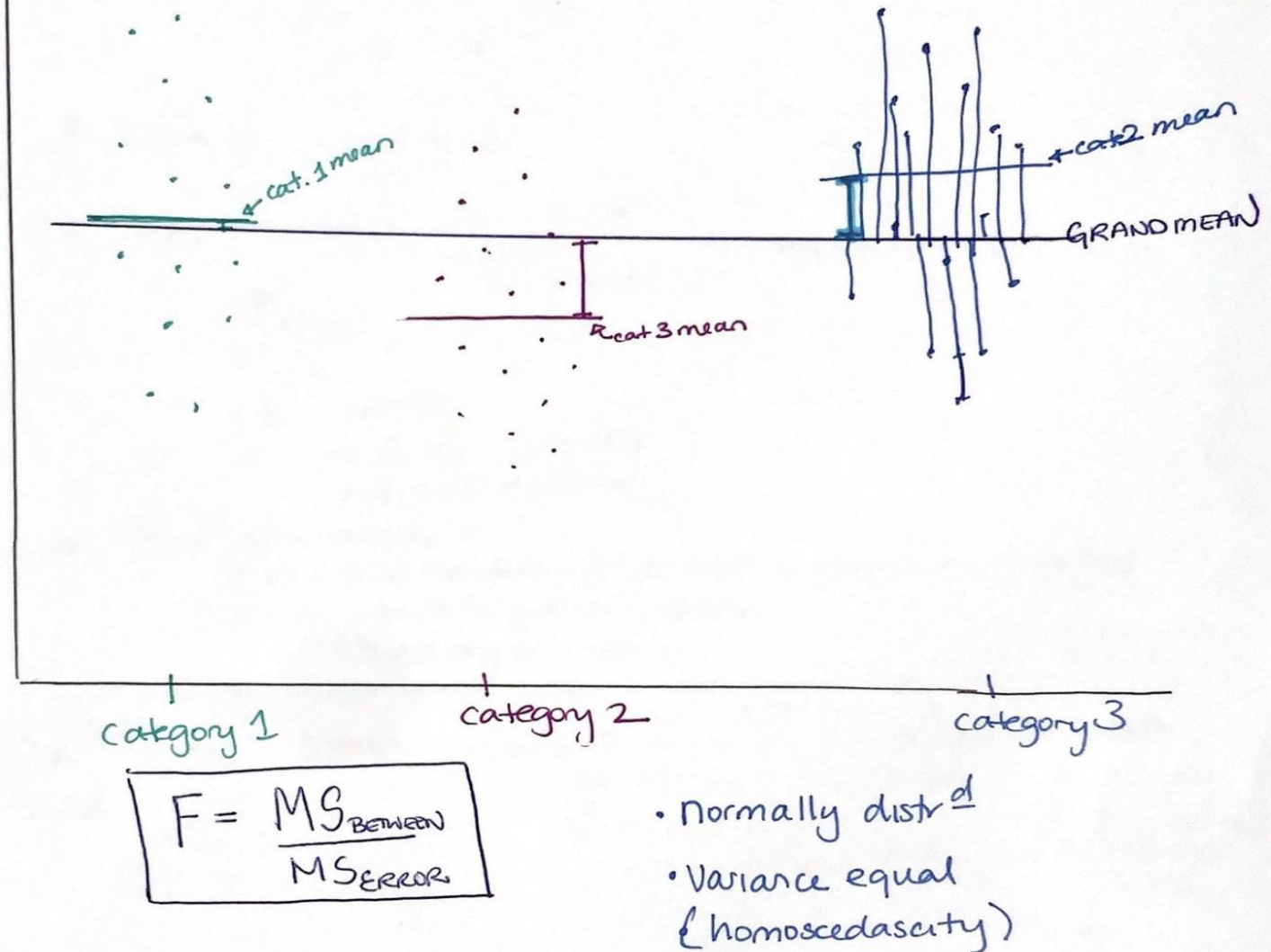
Agenda:

1. ANOVA: Nuts & Bolts
2. Worked Example
 1. One way ANOVA
 2. Post-hoc tests: Tukey-Kramer
 3. Kruskal-Wallis (nonparametric)
3. Linear Correlation
 1. Spearman's rank



Agenda:

1. ANOVA: Nuts & Bolts
2. Worked Example
 - One way ANOVA
 - Post-hoc tests: Tukey-Kramer
 - Kruskal-Wallis
3. Linear Correlation/Regression
 - Spearman's Rank



Analysis of Variance (ANOVA)

Purpose: compare the means of ≥ 2 groups (independent categorical variable) on 1 dependent continuous variable to see if the groups means are different from each other

- Question: Is the variance among groups greater than 0?
 - Method: Allocation of the total variability among different sources

Example:

Three independent categories: current best treatment, control, new treatment

Dependent continuous variable: blood pressure

Analysis of Variance

Purpose: compare the means of ≥ 2 groups (independent categorical variable) on 1 dependent continuous variable to see if the groups means are different from each other

Haven't we already seen a test that compares means?

If there are ≤ 2 groups --> **t-test**

If there are ≥ 2 groups --> **ANOVA**

Why don't we just use multiple t-tests?

$t^2 = F$ when only TWO Categories

$$F = \frac{MSB}{MSW} = \frac{SSB/K-1}{SSW/N-K}$$

When $K=2$

$$F = \frac{MSB}{MSW} = \frac{SSB}{SSW/N-2}$$

$$\frac{(\bar{X} - \bar{Y})^2}{\frac{1}{n_x} + \frac{1}{n_y}} = \frac{(\bar{X} - \bar{Y})^2}{s_p^2 \left(\frac{1}{n_x} + \frac{1}{n_y} \right)} = t^2$$

Remember:

$$t = \frac{(\bar{X} - \bar{Y})}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{(\bar{X} - \bar{Y})}{\sqrt{s_p^2 \left(\frac{1}{n_x} + \frac{1}{n_y} \right)}}$$

s_p^2
Spooled

Analysis of Variance

- o Is the variance among groups greater than 0?
- o *Same question, different metric: **Are the group means significantly different from each other and grand mean?***
 - o Allocation of the total variability among different sources

Why don't we just use multiple t-tests?

Answer: Like a *t-test* but can compare the means of > 2 groups
without inflating Type I error

Analysis of Variance

Are individuals from different groups ***more different***, on average, than individuals chosen from the same group

- H_0 : population means are equal, and sample means are only different due to random sampling error (noise)
- H_A : ***at least one mean*** is different from the other groups

H_0 : Variance among the groups = 0

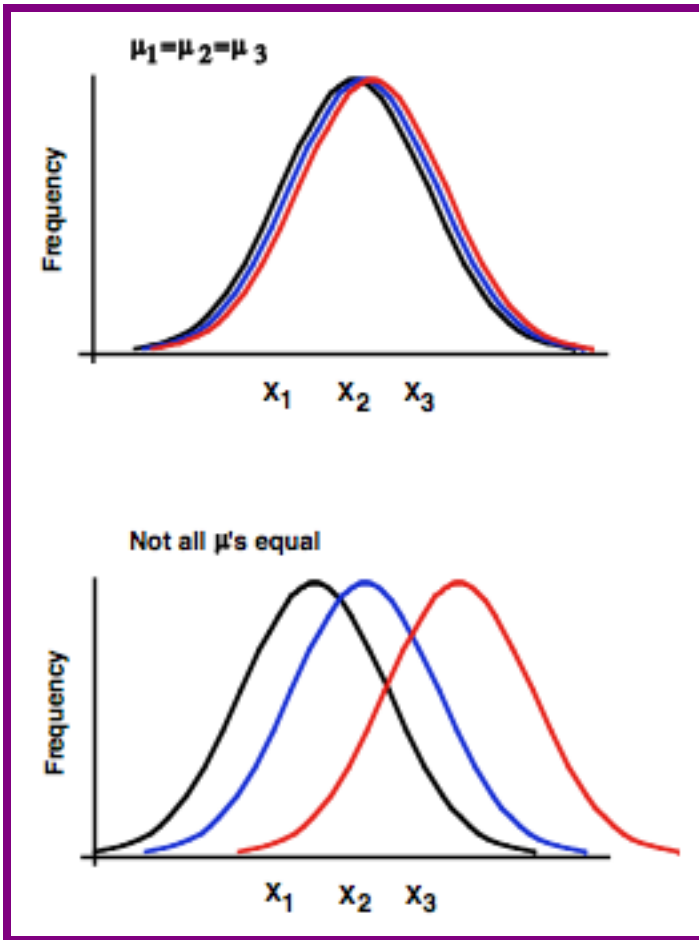
OR

$$H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$$

Analysis of Variance

Are individuals from different groups ***more different***, on average, than individuals chosen from the same group

- H_0 : population means are equal, and sample means only different due to random sampling error
 - Standard error of the null distribution (H_0 is true) is the standard deviation of the group (sample) means so the variance among groups should just be the standard error squared
- H_A : ***at least one mean*** is different from the other groups
 - IF H_0 is NOT true, the variance among groups should be equal to the variance of sample (standard error squared) **PLUS** the real variance among population means



Assumptions:

1. Random samples
2. Normal distribution (each population)
3. Variance among groups is equal
homoscedasticity

- ANOVA is robust to departures from normality
 - especially if n_i is large (Thanks, CTL!)
- If $n_1 = n_2 = n_3$ (and $n = \text{large}$) robust to violations in equal variance (allow up to 10X variance)
- Data transformations can be used if necessary

Analysis of Variance

→ **Even if H_0 is true**, sample means will be different from each other by chance

Question: **Is the variation among sample means *greater than expected by chance alone*?**

- This is evidence that at least one of the population means is different from the others

Assumptions of ANOVA:

- Measurements are random sample
- Variable is normally distributed
- **Variance is the same in all k populations**

How do we handle violations in these assumptions?

1. Robustness (ignore)
 - If data is not normal BUT sample size is large (CLT)
 - variances are not equal, but sample sizes are approximately equal
2. Data Transformation
3. Non-parametric alternative → Kruskal Wallis H test

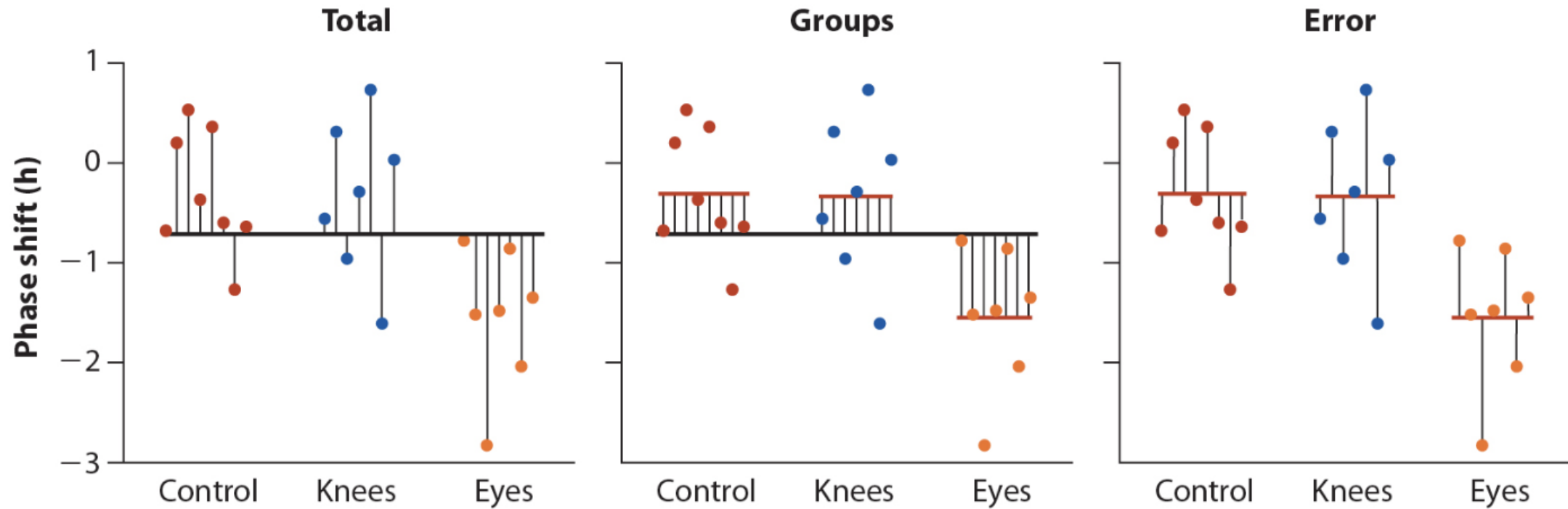


Figure 20.1: Whitlock and Schluter, Fig 15.1.2 – Illustrating the partitioning of sum of squares into MS_{group} and MS_{error} components.

In a one-way ANOVA with 3 groups, a rejection of the null hypothesis implies that:

- A. The 3-population means are equal to each other
- B. The 3-sample means are equal to each other
- C. Each population mean differs significantly from all other population means
- D. Some subset of population means differs from some other subset of population means
- E. Some subset of sample means differs from some other subset of sample means

- Error Mean Square:

- A measure of variability within groups

- Group Mean Square:

- Represents variation among individuals belonging to different groups

Conceptual Crux of ANOVA:

If H_0 is true, then group means should be the same so the two types of mean square should be equal

$$MS_{\text{error}} = MS_{\text{groups}}$$

Under H_0 , the sample mean of each group **should only vary** because of sampling error

The standard deviation of sample means, when the true mean is constant, is just the standard error:

$$\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}}$$

Squaring the standard error, the variance **among** groups due to sampling error is:

$$\sigma_{\bar{X}}^2 = \frac{\sigma_X^2}{n}$$

If H_0 is **not** true, the variance **among** groups should be equal to the variance due to sampling error **plus** the real variance among population means

$$\sigma_{\bar{X}}^2 = \frac{\sigma_X^2}{n} + \text{Variance}(\mu_i)$$

ANOVA tests whether the variance among true group means is **significantly** greater than zero

We do this by asking whether the observed variance among groups is greater than expected by chance

$$\sigma_{\bar{X}}^2 > \frac{\sigma_X^2}{n}$$



$$n\sigma_{\bar{X}}^2 > \sigma_X^2$$

Population Parameters

Estimates from Samples

$$n\sigma_{\bar{X}}^2$$

Estimated by the “mean square group”

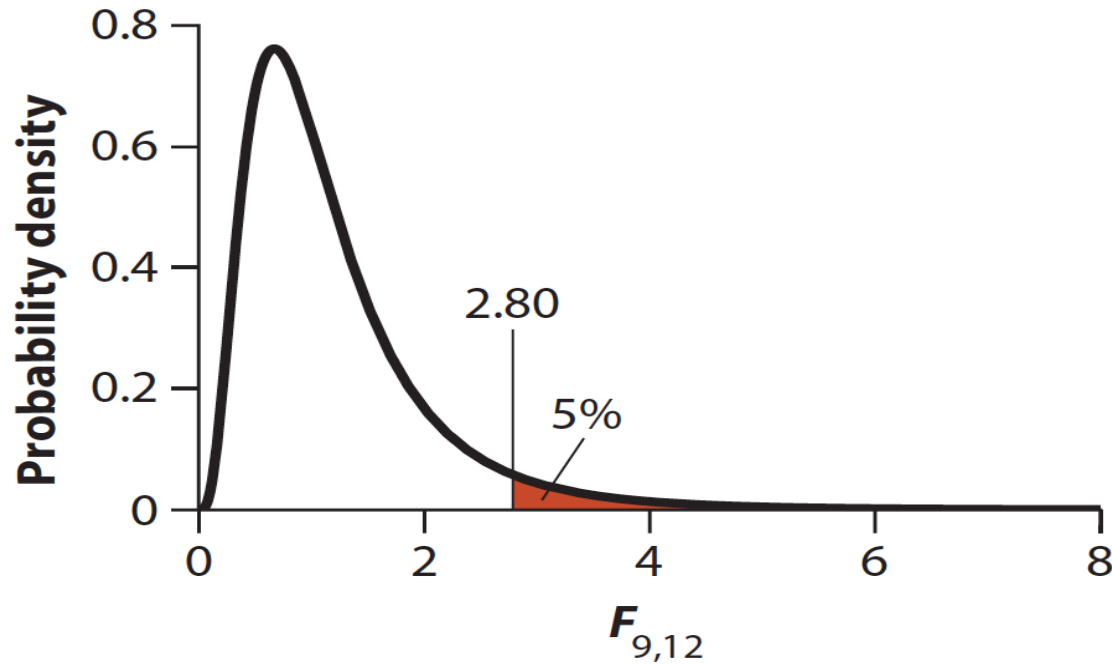
Since it should (almost) **always be the larger value, it is in the NUMERATOR**

MS_{group}

$$\sigma_X^2$$

- The variance within groups
- Estimated by “mean square error”
- One of the assumptions of ANOVA is that this variance is *approximately the same between different groups*

MS_{error}



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

- This is a **one-sided test** 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).

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

- reminder: t-tests also involve a ratio
 - numerator in a t-test is the difference between two sample means
 - numerator in ANOVA is average difference between means squared
- denominator is equivalent in both:
 - t-test: standard error of difference between means
 - ANOVA: average error within groups squared

*summary: just like in the t-test, in ANOVA we are trying to determine the average difference **between** group means relative to the average difference **within** group means*

Conceptual Crux of ANOVA:

If H_0 is true, then group means should be the same so the two types of mean square should be equal

$$MS_{\text{error}} = MS_{\text{groups}}$$

$$F = \frac{MS_{\text{groups}}}{MS_{\text{error}}} \geq 1$$

If $F \approx 1$, we FTR H_0 . If $F \gg 1$, there is enough evidence to reject H_0

$$MS_{error} = \frac{SS_{error}}{df_{error}} = \frac{\sum s_i^2 (n_i - 1)}{N - k}$$

$$SS_{error} = \sum df_i s_i^2 = \sum s_i^2 (n_i - 1)$$

$$df_{error} = \sum df_i = \sum (n_i - 1) = N - k$$

Mean of group i

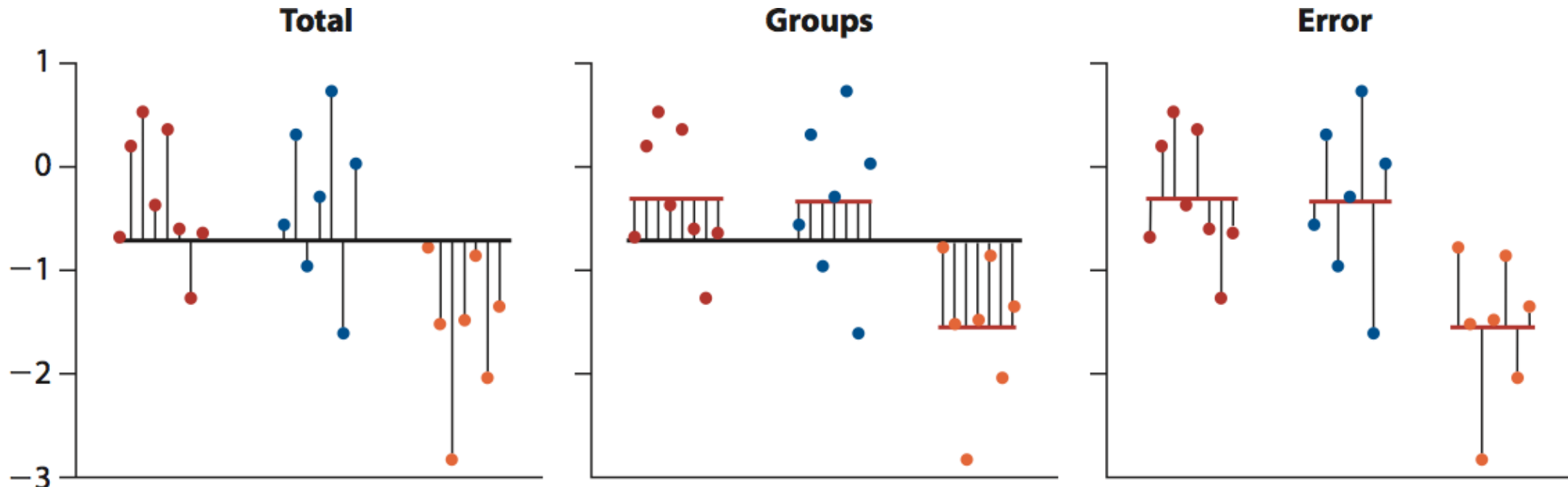
$$MS_{groups} = \frac{SS_{groups}}{df_{groups}} = \frac{\sum n_i (\bar{X}_i - \bar{X}_T)^2}{k - 1}$$

$$\bar{X}_T = \frac{\sum_i \sum_j X_{ij}}{N}$$

$$\bar{X}_T = \frac{\sum_i n_i \bar{X}_i}{N}$$

Results are presented in ANOVA Table:

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



R² value:

- The fraction of variability that is explained by groups
- Measures reduction in scatter around group means compared to the grand mean

$$SS_{\text{Total}} = SS_{\text{groups}} + SS_{\text{error}}$$

$$R^2 = \frac{SS_{\text{groups}}}{SS_{\text{Total}}} \quad ; \quad 0 < R^2 < 1$$

A worked example of ANOVA:

Researchers are investigating the effect of three different diets (A, B, and C) on body weight in genetically modified mice that are prone to obesity. After 8 weeks, the body weights of the mice are measured (in grams). The data is as follows:

Body weights after 8 weeks (grams)

Diet A: 32, 30, 29, 34, 35

Diet B: 40, 42, 43, 45, 41

Diet C: 38, 35, 39, 37, 36

Step 1:

H_0 : There is no difference in the mean body weight after 8 weeks among the diets.

H_A : At least one diet group has a different mean body weight.

Step 1:

H_0 : There is no difference in the mean body weight after 8 weeks among the diets.

H_A : At least one diet group has a different mean body weight.

Step 2:

GROUP	mean	s	n
A	32	2.55	5
B	42.2	1.92	5
C	37	1.58	5

$$N = \sum n = 15$$

Mean square error:

$$SS_{error} = \sum df_i s_i^2 \quad \mathbf{df}_{error} = 4+4+4 = 12$$

$$= 4(2.55)^2 + 4(1.92)^2 + 4(1.58)^2 = 50.74$$

$$\mathbf{MS}_{error} = 50.74/12 = 4.23$$

mean squares groups:

$$\bar{X}_G = \frac{5(32) + 5(42.2) + 5(37)}{15} = 37.07$$

$$\mathbf{df}_{groups} = k - 1 = 3 - 1 = 2$$

$$\mathbf{SS}_{groups} = 5(32.0 - 37.07)^2 + 5(42.2 - 37.07)^2 + 5(37.0 - 37.07)^2 \\ = 260.13$$

$$\mathbf{MS}_{groups} = SS_{groups} / df_{groups} = 260.13/2 = 130.07$$

The test statistic for ANOVA is F:

$$\begin{aligned} F &= MS_{\text{groups}} / MS_{\text{error}} \\ &= 130.07 / 4.23 \\ &= 30.25 \end{aligned}$$

$$F_{0.05(1),2,12} = 3.88$$

Since $30.25 \gg 3.88$, we know that $P < 0.05$ and we can reject H_0 .

The variance between the sample group means is bigger than expected given the variance within sample groups so at least one of the groups has a population mean different from another group

Source of variation	Sum of Squares	df	Mean Squares	F-ratio	P
Groups (treatment)	260.13	2	130.07	30.25	<0.001
Error	50.80	12	4.23		
Total	310.93	14			

$$R^2 = SS_{\text{groups}} / SS_{\text{error}} = 260.13 / 50.80 = 5.12$$

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 $df = k - 1$

Experimental Design:

*How do we identify **which** means are different and the **magnitude** of their difference?*

- Planned comparisons:
 - **A priori** comparison between means of groups that were previously identified as particularly interesting
 - **Baked into the study design**
 - **Determined BEFORE data are examined**
 - Only small number allowed so that α isn't inflated
- Unplanned comparisons:

Experimental Design:

*How do we identify **which** means are different and the magnitude of their difference?*

- Planned comparisons:

- **A priori** comparison between means of groups that were previously identified as particularly interesting
- Only small number allowed so that α isn't inflated

- Method:

- Similar to two-sample t-tests
- Use t distribution

- Different standard error: pooled sample variance (MS_{error}) based on all k groups (i.e. using all the information about variance rather than just a subset)

- df of MS_{error}

$$SE = \sqrt{MS_{error} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$$

What do I mean by inflation of α ?

- For a two-sample t test, you are dividing up the variance of only **two** groups into the two samples.

$$\underbrace{\frac{(n_1 - 1)s_1^2}{N - k} + \frac{(n_2 - 1)s_2^2}{N - k} + \dots + \frac{(n_k - 1)s_k^2}{N - k}}_{\mathbf{MS_{error}}}$$

s_p^2

- For a planned comparison, you are dividing up **ALL** the variance (all the total deviations of the data points) into **only two** of the **k** groups (note: you can do this because H_0 assumes variance is same in all groups)

Big idea: this means that you have access to all the degrees of freedom provided by the data points even the ones that are in the groups we are not comparing!

- We saw a different test that also ‘absorbed’ inflated error by tweaking df (Welch’s approximate t test, this reduced df instead of expanding it)

Why use MS_{error} instead of a two-sample t-test?

- Increased precision
- Increased power

Assumptions:

- Same as ANOVA but not as robust to violations

Experimental Design:

*How do we identify **which** means are different and the **magnitude** of their difference?*

- Planned comparisons
- Unplanned comparisons:
 - Post hoc
 - Multiple comparisons
 - Determine which means and their magnitude
 - Type **of data dredging** (interleaf) so protect against increasing α
 - Tukey-Kramer procedure tests all pairs of means

Tukey-Kramer test*:

- Already carried out a single-factor ANOVA and rejected H_0
- Compares all group means to all other group means

$$H_0: \mu_1 = \mu_2$$

$$H_0: \mu_1 = \mu_3$$

$$H_0: \mu_2 = \mu_3$$

* Tukey's Honestly Significant difference (HSD) test

So why not just use a series of two-sample t-tests?

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**)

$P(\text{No type I errors}) = (1 - \alpha)^N$, where $N = \text{independent tests}$

$P(\geq 1 \text{ type I error}) = 1 - (1 - \alpha)^N$

Why not use a series of two sample t-tests?

- Multiple comparisons would cause the t-test to reject too many true null hypotheses
- Tukey-Kramer adjusts for the number of tests

Uses larger critical value to limit Type I error

$$P(\geq 1 \text{ Type I error}) = \alpha$$

- Tukey-Kramer also uses information about the variance within groups from all the data, so it has more power than a t-test with a Bonferroni correction (data dredging interleaved): $\alpha^* = \alpha / \# \text{ of tests}$

Tukey-Kramer test:

- Uses **q test statistic**
- Method:
 1. Order group means from smallest to largest
 2. Compare each pair of group means

Ex: First comparison:

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

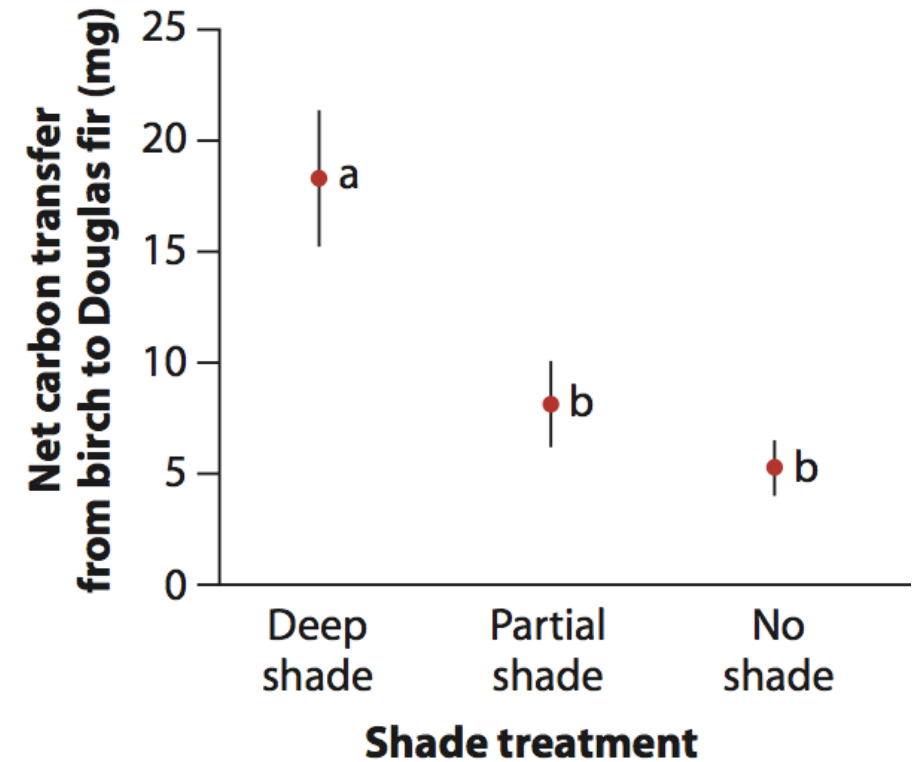
$$H_A: \mu_1 - \mu_2 \neq 0$$

3. Calculate **q** test statistic:

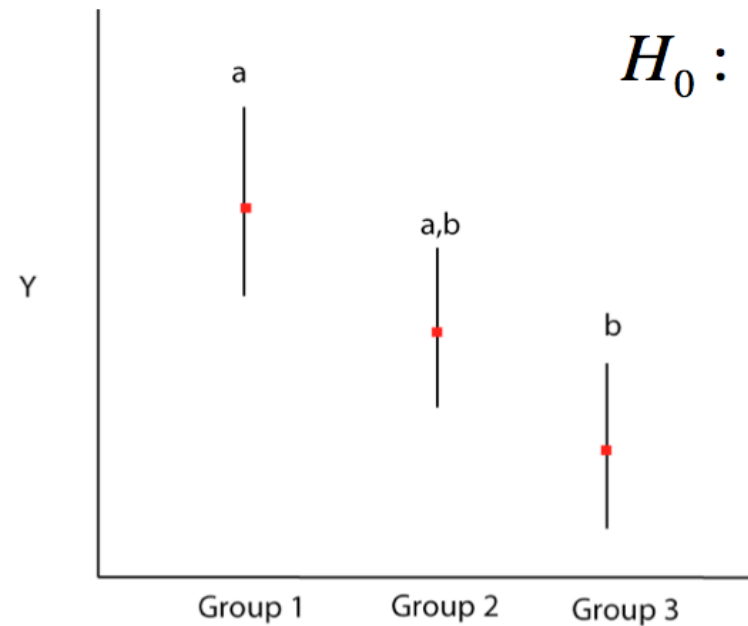
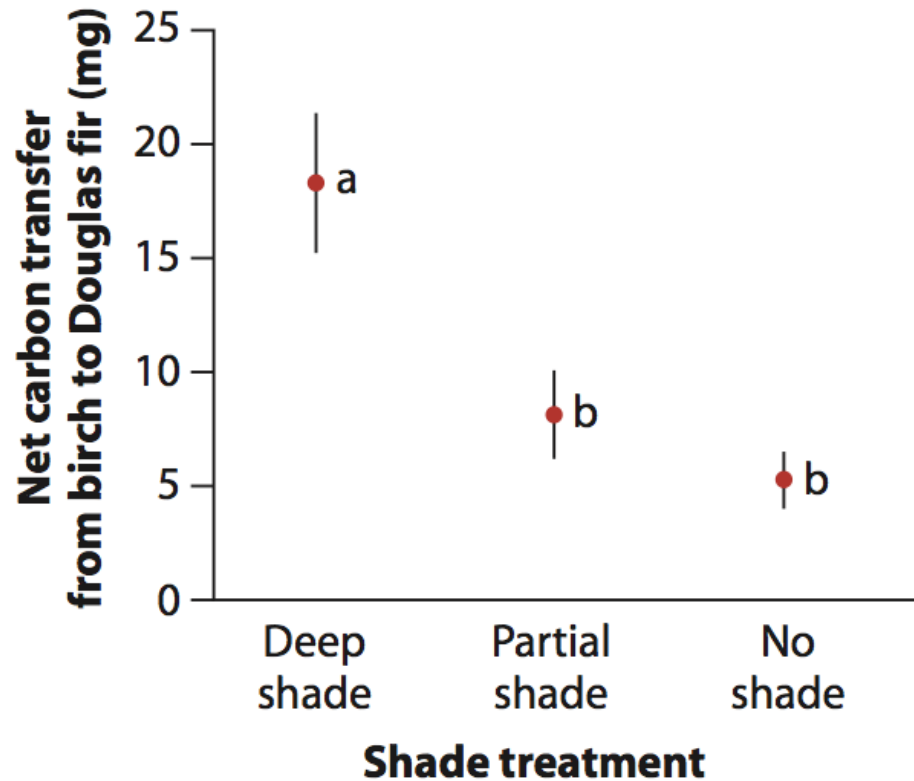
Standard error: MS_{error} df: **k** and **N - k**

Q-distribution (statistical tables for this online)

- Same assumptions as ANOVA but not as robust
- P value is correct when design is balanced (approximately same number of data points in each category) but it is conservative when unbalanced (makes it more difficult to reject the null hypothesis)



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

The Tukey test compares the means between each pair of diets (A, B, C) to see which groups differ significantly:

Group 1	Group 2	Mean Difference	P-Value	95% Confidence Interval	Reject Null Hypothesis
Diet A	Diet B	-10.2	<0.001	[-13.51, -6.88]	yes
Diet A	Diet C	5.0	0.007	[1.69, 8.30]	yes
Diet B	Diet C	5.2	0.006	[1.88, 8.51]	Yes

Fixed Effects:

- 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:

- 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**

How do assumptions for random-effects ANOVA and fixed-effects ANOVA **differ**?

- a) They have the same assumptions
- b) The fixed-effects ANOVA assumes the groups are randomly sampled and the group means have a normal distribution in the population whereas the random-effect ANOVA does not.
- c) The random-effects ANOVA assumes the groups are randomly sampled and the group means have a normal distribution in the population whereas the fixed-effect ANOVA does not.
- d) The fixed-effects ANOVA assumes the variance is the same in all k populations and the variable is normally distributed in each of the k populations whereas the random-effect ANOVA does not.

https://PollEv.com/multiple_choice_polls/rqU7TM0YqDkQDQ59uZP7q/respond

Correlation

Imagine you determine there is a positive correlation between the number of times individuals visit doctors (over the course of their life span) and longevity (how many years the individuals live). Would it be appropriate to conclude that frequent visits to the doctor lead to a longer life span? Why or why not?

Two Variables: Which test?

		Explanatory variable	
		Categorical	Numerical
Response Variable	Categorical	<ul style="list-style-type: none"> Contingency analysis 	<ul style="list-style-type: none"> Logistic regression Survival analysis
	Numerical	<ul style="list-style-type: none"> t-test Analysis of variance 	<ul style="list-style-type: none"> Regression Correlation

- Mathematical relationship between the two methods of analysis are close: share many computational steps which are similar or the same
- Appropriate method depends on the purpose of the investigator and the nature of the variables

Regression:

Intention: describe dependence of a variable Y, on an independent variable X.

Regression equations are employed to support hypotheses regarding possible **causation** of changes in Y by changes in X; to predict Y in terms of X

A line (in linear regression) is fit to the data

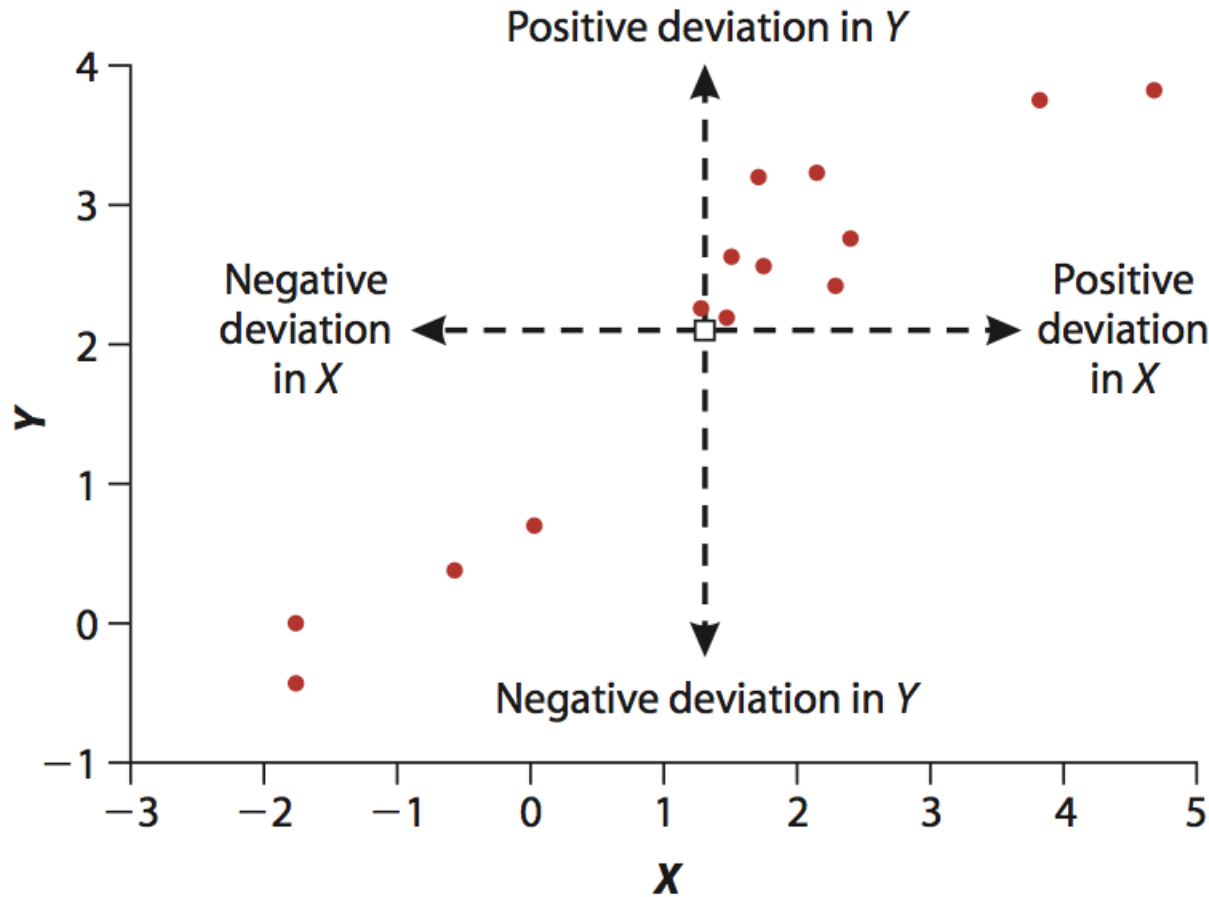
*Study the **effects** of X on Y*

Correlation:

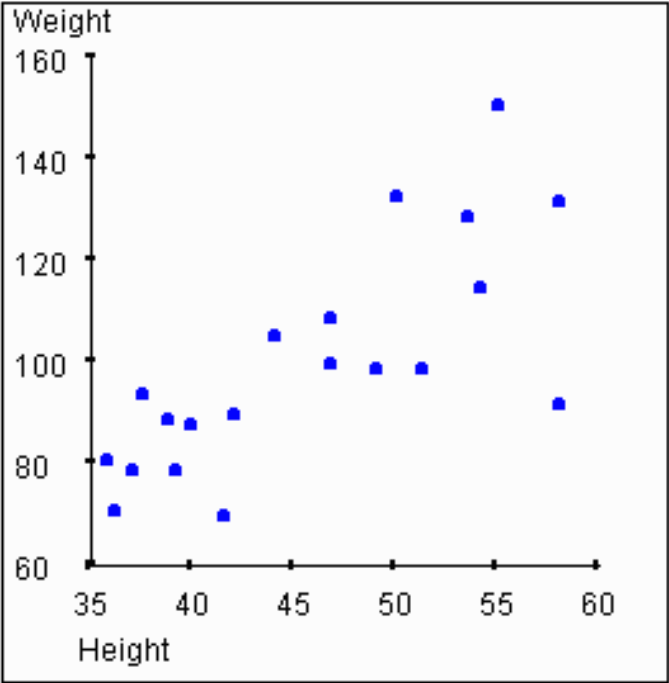
- Measures the amount/degree of linear association between two **numerical** variables
- Estimate the degree to which variables **covary**
- We do not express one variable as a function of the other variable
- No distinction between dependent & independent variables
 - In fact, we usually assume that they both stem from a common cause -- with a pair of variables whose correlation is studied, one may very well be the cause of the other but we neither know nor assume!

(Pearson) Correlation Coefficient

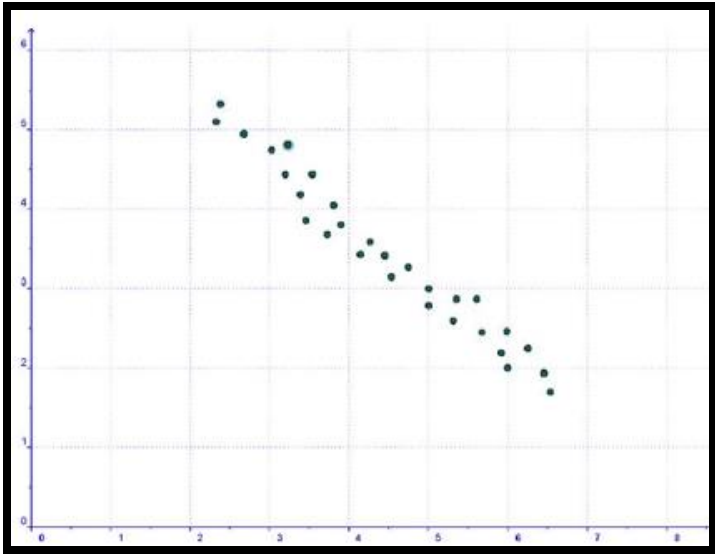
$$r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2} \sqrt{\sum (Y - \bar{Y})^2}} = \frac{\text{Covariance}(X, Y)}{s_x s_y}$$



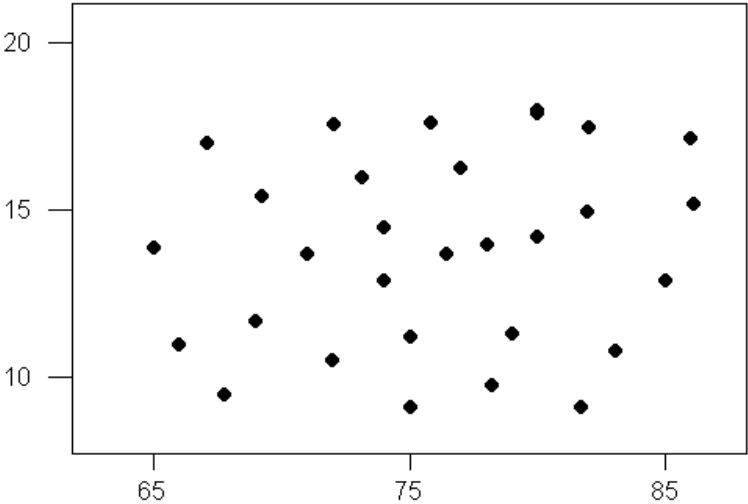
Positive Correlation



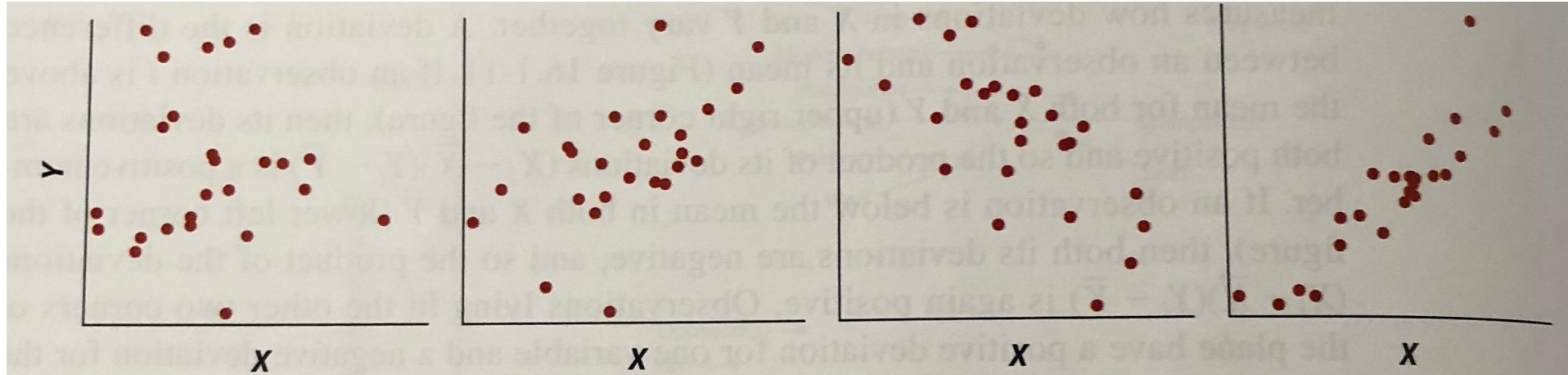
Negative Correlation



No Correlation



Match the Graphs below with the most appropriate r value.
(possible values: -0.7, 0.0, 0.5, 0.9)



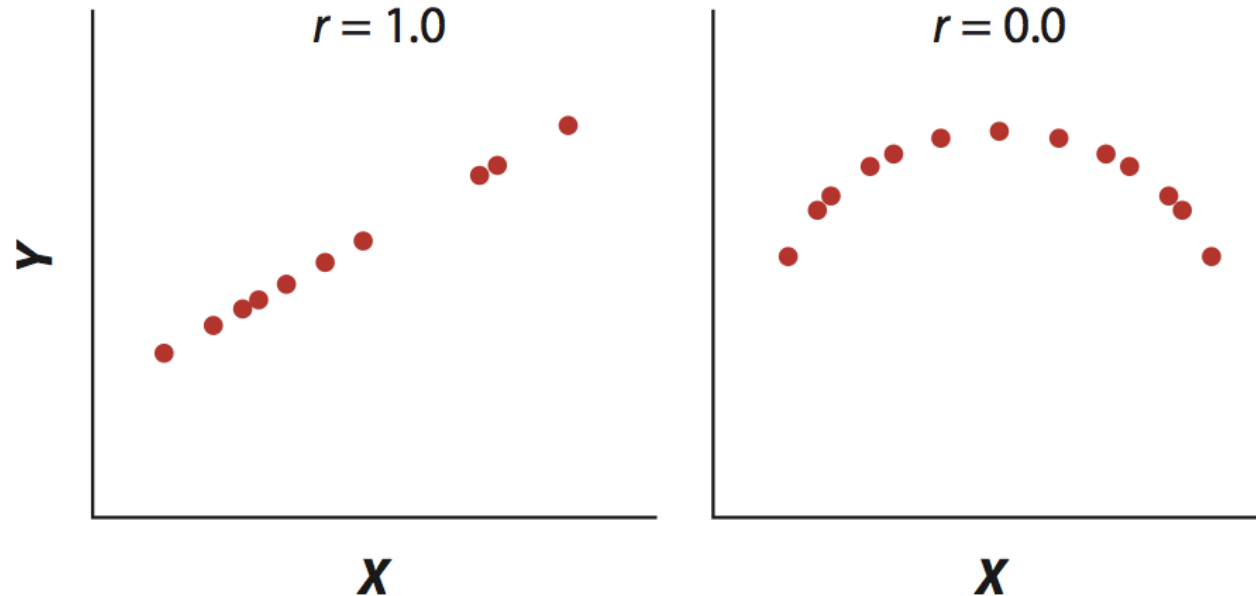
- A) Graph 1: 0.5, Graph 2: 0.9, Graph 3: -0.7, Graph 4: 0.0
- B) Graph 1: 0.0, Graph 2: 0.5, Graph 3: -0.7, Graph 4: 0.9
- C) Graph 1: -0.7, Graph 2: 0.0, Graph 3: 0.5, Graph 4: 0.9
- D) Graph 1: 0.0, Graph 2: 0.5, Graph 3: 0.9, Graph 4: -0.7

https://PollEv.com/multiple_choice_polls/f3KnWzpajzjEoSRTbKaEr/respond

$$r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2} \sqrt{\sum (Y - \bar{Y})^2}} = \frac{\text{Covariance}(X, Y)}{s_x s_y}$$

$$-1 < r, \rho < 1$$

This ONLY describes LINEAR correlation. If two variables are tightly correlated, but nonlinear:



Testing for no correlation:

Step 1: declare null and alternate hypotheses

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

H_A : Some correlation ($\rho \neq 0$)

Testing for no correlation:

Step 1: declare null and alternate

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

H_A : Some correlation ($\rho \neq 0$)

Step 2: test statistic

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

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

Testing for no correlation:

Step 1: declare null and alternate

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

H_A : Some correlation ($\rho \neq 0$)

Step 2: test statistic

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

Step 3: P-value/Critical value

- Null distribution has a **sampling distribution** of Student's t-distribution with $d.o.f. = n - 2 \rightarrow$ use a Student's t table!
- *Why $n - 2$? Use two summaries of data, \bar{X} and \bar{Y}*

Testing for no correlation:

Step 1: declare null and alternate

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

H_A : Some correlation ($\rho \neq 0$)

Step 2: test statistic

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

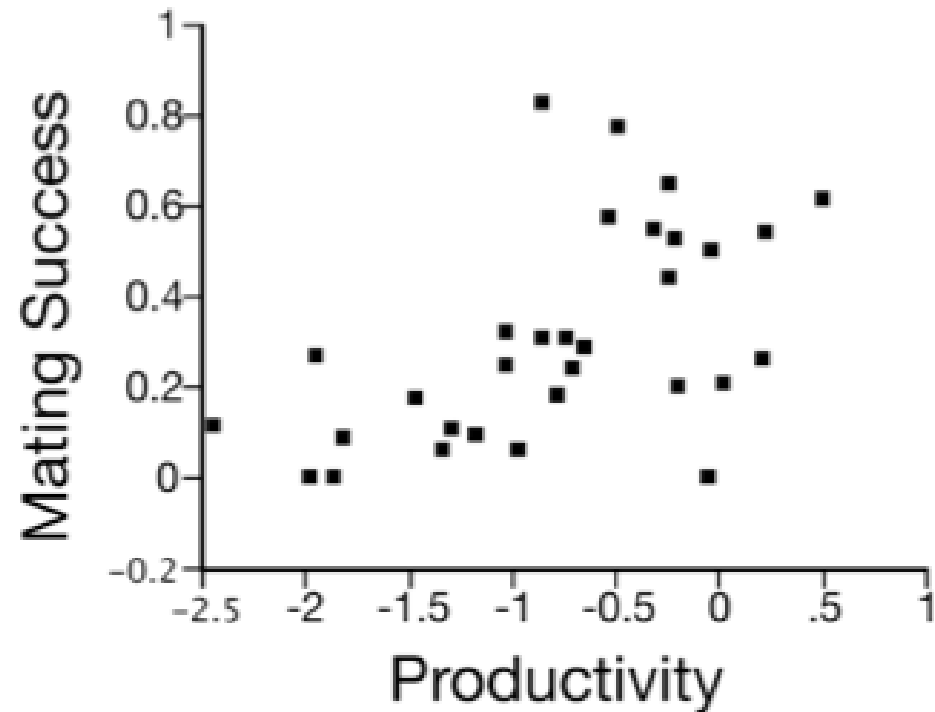
Step 3: P-value/Critical value

Step 4: State conclusion and Confidence interval

Correlation?

Confidence Interval needs to use transformation since SE_r is not normally distributed

Example: Are the effects of new mutations on mating success and productivity correlated? 31 data points from various visible mutations in *Drosophila melanogaster*.



Example: Are the effects of new mutations on mating success and productivity correlated? Data from various visible mutations in *Drosophila melanogaster*.

H_0 : Mating success and productivity are not related

H_A : Mating success and productivity are related

Example: Are the effects of new mutations on mating success and productivity correlated? Data from various visible mutations in *Drosophila melanogaster*.

H_0 : Mating success and productivity are not related

H_A : Mating success and productivity are related

X is productivity, Y is the mating success

$$\sum X = -24.228$$

$$\sum X^2 = 35.1808$$

$$\sum XY = -4.62741$$

$$\sum Y = 9.498$$

$$\sum Y^2 = 4.5391$$

$$n = 31$$

$$r = \frac{2.796}{\sqrt{(16.245)(1.6289)}} = 0.535$$

$$SE_r = \sqrt{\frac{1 - r^2}{n - 2}} = \sqrt{\frac{0.7045}{29}} = 0.1558$$

$$t = \frac{0.5435}{0.1558} = 3.49$$

Example: Are the effects of new mutations on mating success and productivity correlated? Data from various visible mutations in *Drosophila melanogaster*.

H_0 : Mating success and productivity are not related

H_A : Mating success and productivity are related

$t = 3.49$

$df = 29$

This is greater than $t_{0.05(2), 29} = 2.045$, so we can reject the null hypothesis and say that productivity and male mating success are correlated ($\rho \neq 0$).

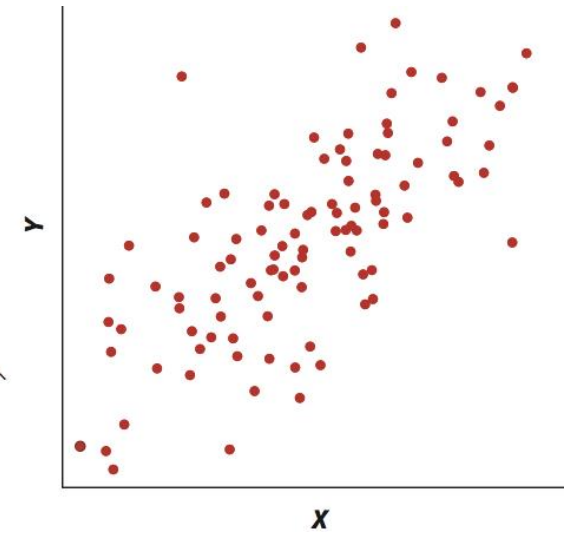
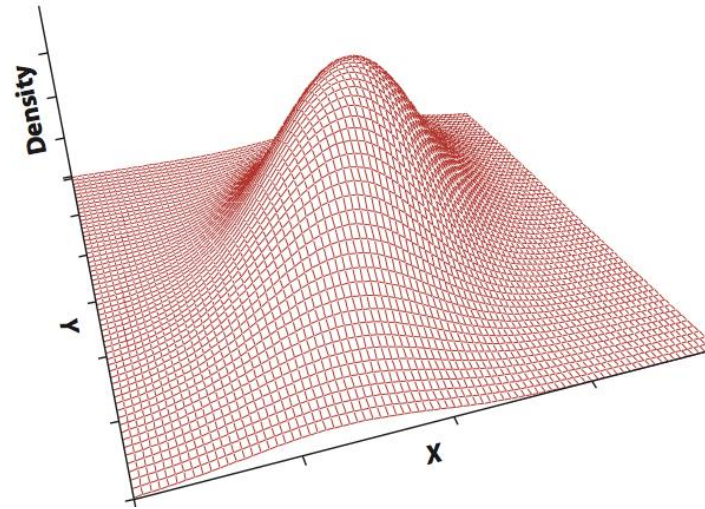
The 95% confidence interval for this parameter is:

$0.22 < \rho < 0.747$

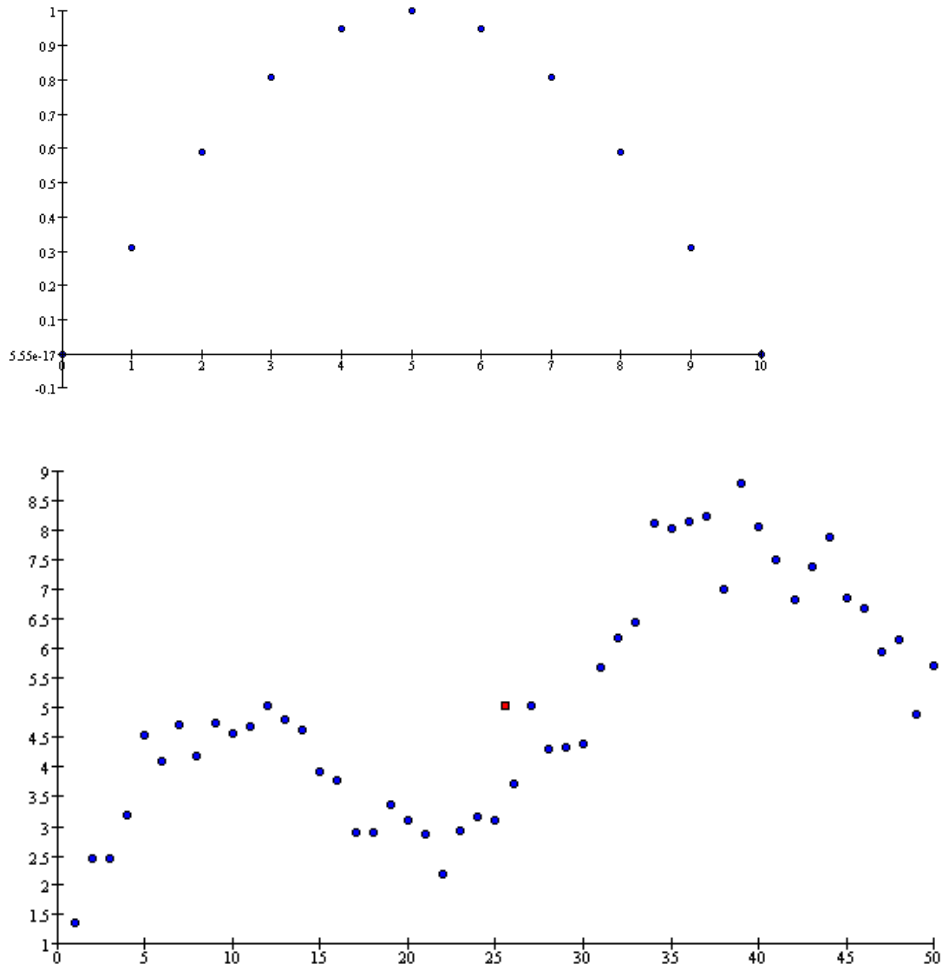
* I used a horrible conversion (Fisher's Z transformation) to get this confidence interval that is not intuitive at all.

Assumptions:

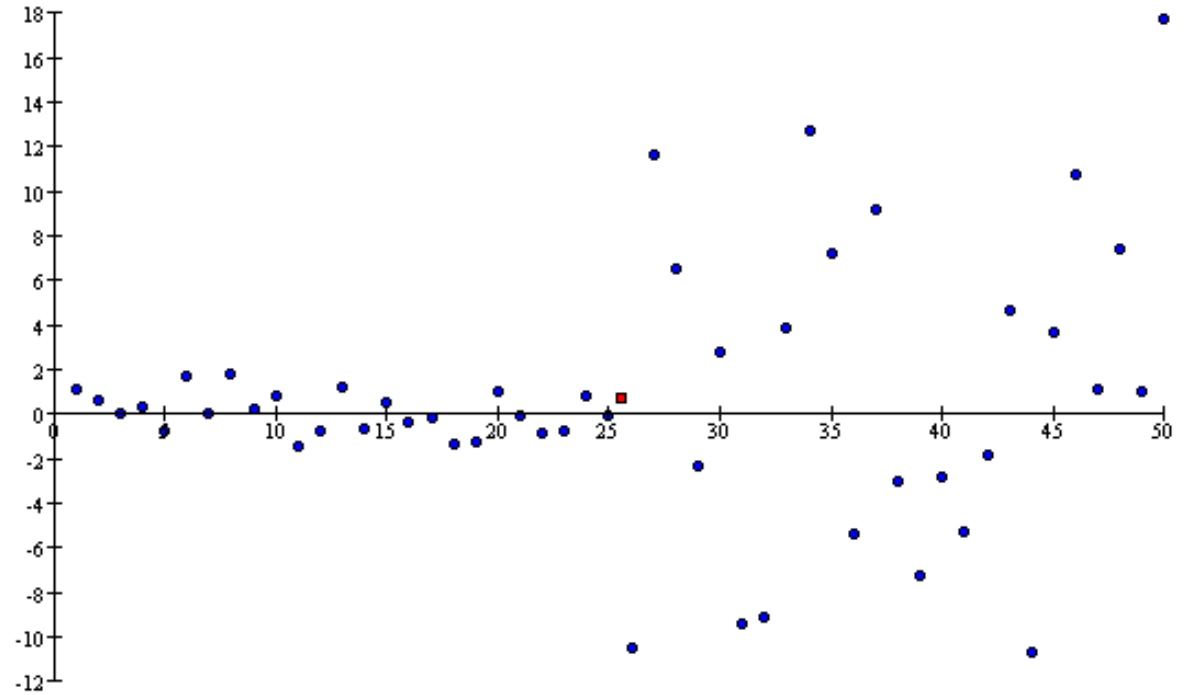
- Random sample
- Linearity
- Correlation depends on range of values
- Homoscedastic variances
- **Bivariate Normal Distribution**
 - X is normally distributed
 - Y is normally distributed
 - X and Y have linear relationship



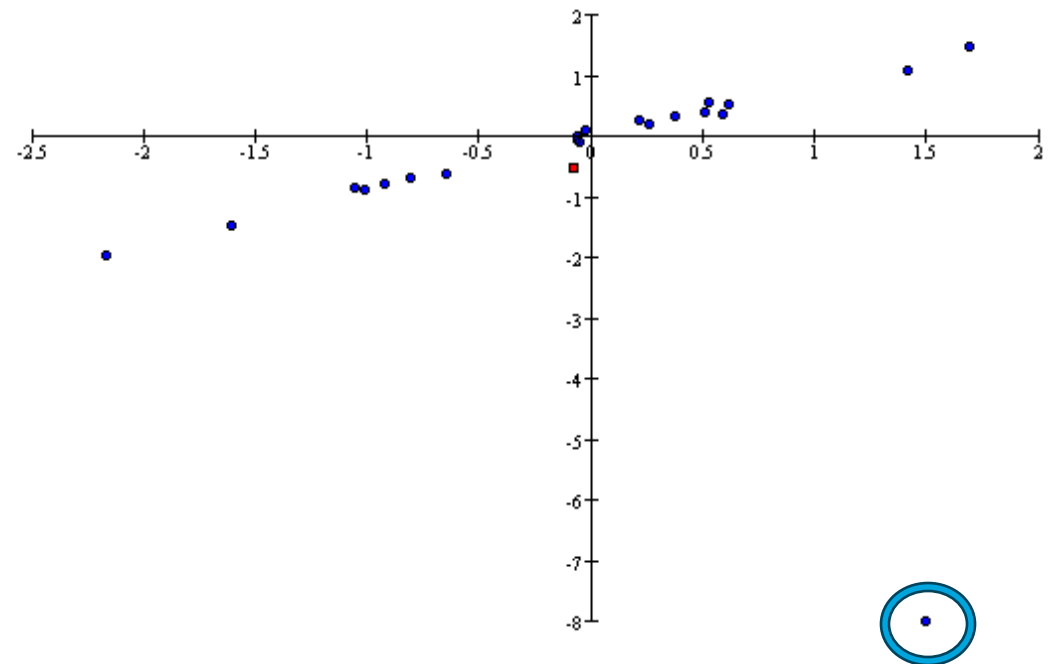
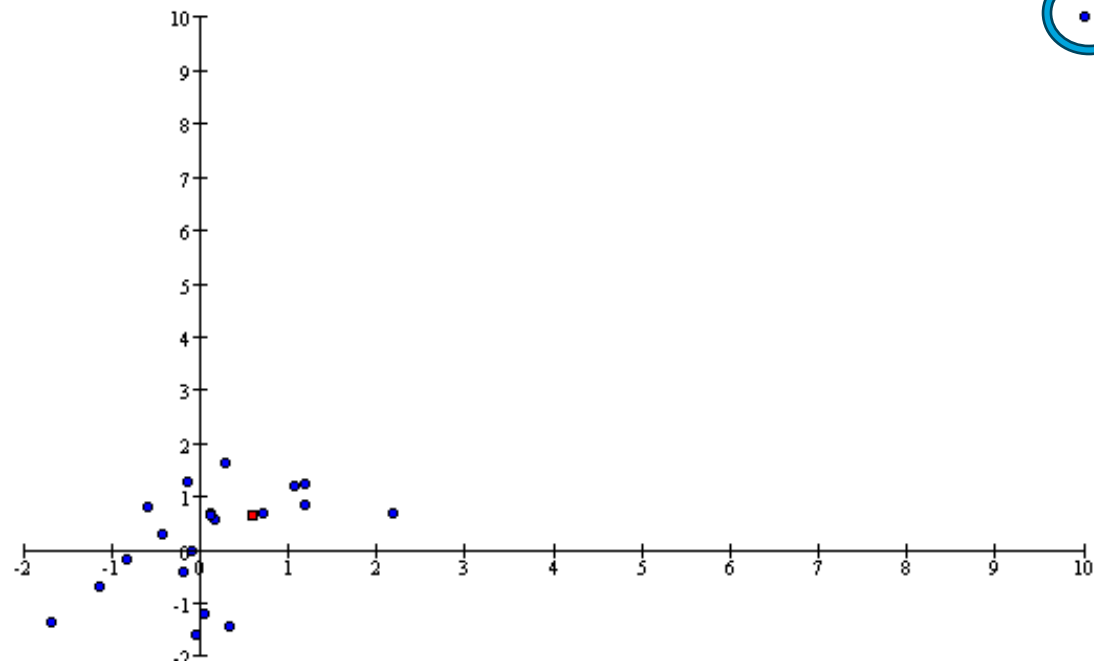
Non-linearity



Heteroscedascity



Outliers:



- If data are not bivariate or are not linearly related, try **transformation**
- If data are heteroscedastic or have outliers... try a **non-parametric method**..... But, remember non-parametric methods are more conservative (they have less power) than parametric metrics.

Spearman's rank

assumes:

- * random sample
- * linear relationship

Spearman's rank correlation:

- Measures strength and direction of linear association between the **ranks** of two variables
- Two variables are ranked separately

Parameter: ρ_s ; sample estimate: r_s

Spearman's rank correlation:

Test for correlation in the normal way....

Step 1: declare null and alternate

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

H_A : Some correlation ($\rho_s \neq 0$)

Step 2: test statistic

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

Step 3: State α /P-value/Critical value

Table or computer!

Step 4: State conclusion

If $n > 100$:

$$t = \frac{r_s - \rho_s}{SE_{r(s)}}$$

This is the same procedure as linear
Pearsons correlation, but for ranks!

where:

$$SE_{r_s} = \sqrt{\frac{1 - r_s^2}{n - 2}}$$

t is $\sim t$ -distributed with $n - 2$ degrees of freedom

Tricky part: reject null hypothesis if

$$t \geq t_{0.05(2), n-2}$$

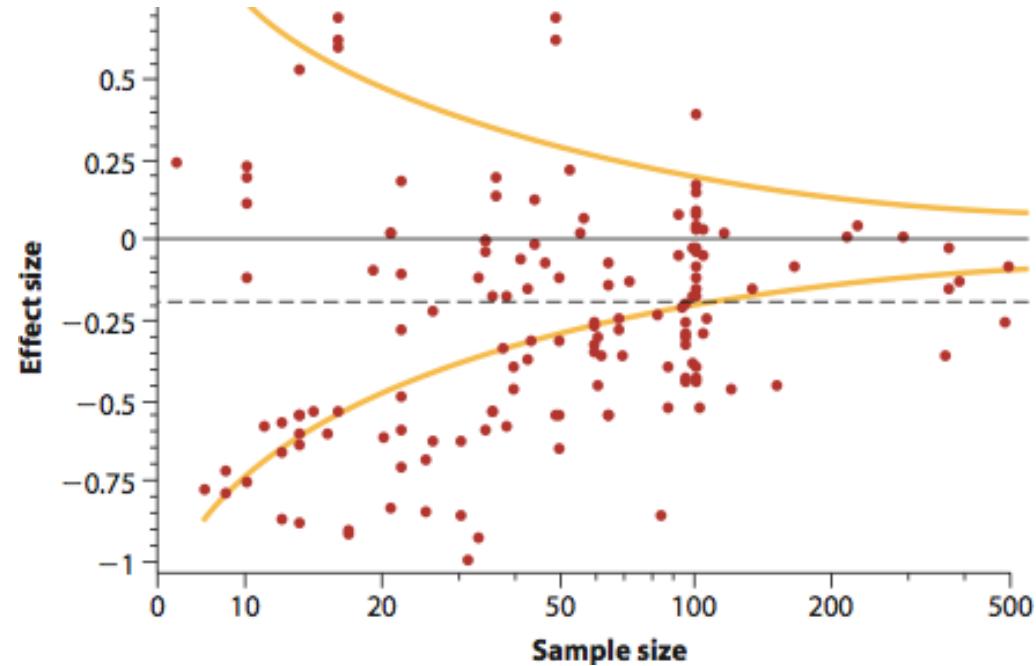
$$t \leq -t_{0.05(2), n-2}$$

Publication Bias

Papers that:

- Reject null
- Have large effect

tend to be published



<http://www.badsience.net/about-dr-ben-goldacre/>