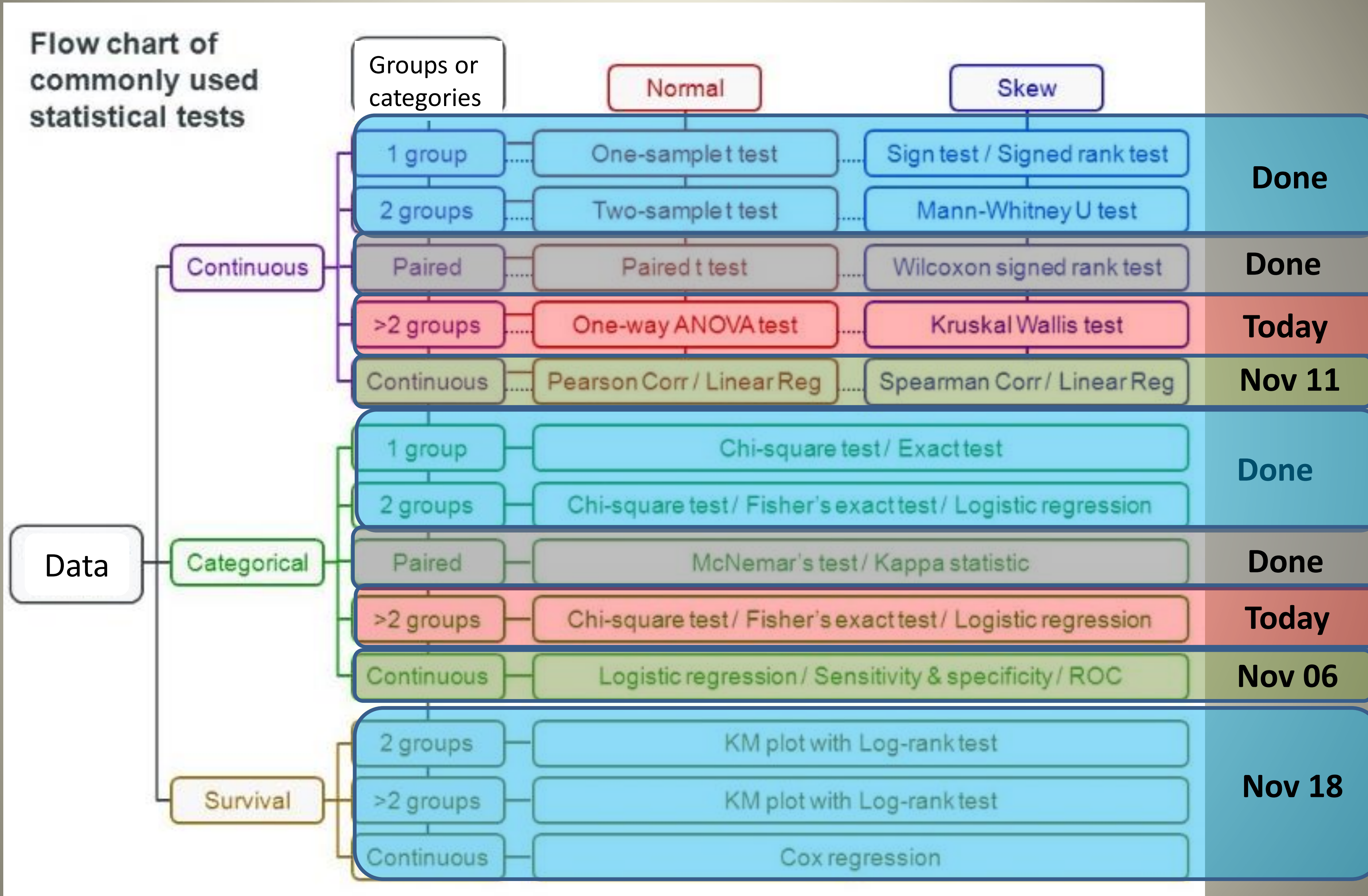
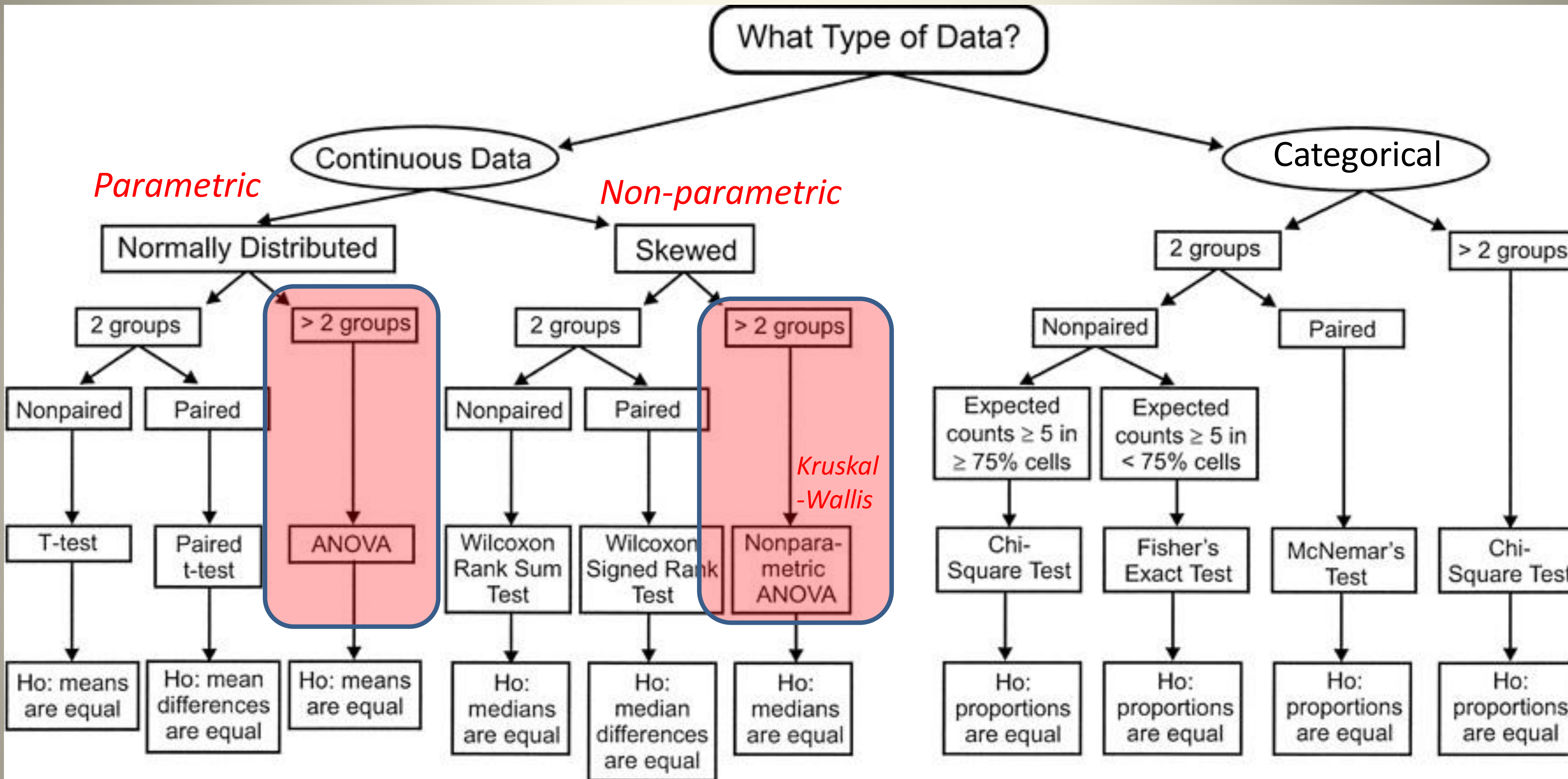
Applied Statistics: Tests on >2 groups One-way and Two-way ANOVAs and the Kruskal-Wallis test

Kathleen Torkko, PhD, MSPH, MS
Associate Professor
Department of Pathology (SOM)
Department of Biostatistics and Bioinformatics (SOPH)


Flow chart of commonly used statistical tests
Groups or categories
Normal
Skew
Data
Continuous
1 group
One-sample t test
Sign test / Signed rank test
2 groups
Two-sample t test
Mann-Whitney U test
Done
Paired
Paired t test
Wilcoxon signed rank test
Done
>2 groups
One-way ANOVA test
Kruskal Wallis test
Today
Continuous
Pearson Corr / Linear Reg
Spearman Corr / Linear Reg
Nov 11
Categorical
1 group
Chi-square test / Exact test
2 groups
Chi-square test / Fisher's exact test / Logistic regression
Done
Paired
McNemar's test / Kappa statistic
Done
>2 groups
Chi-square test / Fisher's exact test / Logistic regression
Today
Continuous
Logistic regression / Sensitivity & specificity / ROC
Nov 06
Survival
2 groups
KM plot with Log-rank test
>2 groups
KM plot with Log-rank test
Continuous
Cox regression
Nov 18



Source: Waning B, Montagne M: *Pharmacoepidemiology: Principles and Practice*: http://www.accesspharmacy.com

Objectives

- Learn the differences between and one-way and a two-way ANOVA
- Learn the assumptions of each ANOVA
- Learn how perform and interpret both ANOVAs using Prism
- Learn to use the non-parametric one-way ANOVA equivalent, the Kruskal-Wallis test

ANOVA Basics
ANalysis Of VAriance (ANOVA)

One categorical variable (“factor”; >2 groups)

AKA the **independent*** variable

Type of diet (diet1, diet2, diet3)

One continuous variable

AKA the **dependent*** variable

the variable we want to know if it differs between groups of the categorical independent variable

Weight loss

*This terminology comes from linear regression where the independent variable “predicts” or determines levels of the dependent variable

weight loss = diet (diet predicts weight loss)

ANOVA Family

- **One way ANOVA**
  - One Categorical (Independent) Variable (AKA one factor)
    - One continuous (dependent) variable measured in different groups
      - Different participants
    - Repeated measures / Within subjects
      - Same participants
- **Factorial ANOVA**
  - More than One Categorical Variable
    - Two way
    - Three way
    - Four way

One-way ANOVA

- Weight loss by type of diet
- Extension of independent (unpaired) t-test for >2 groups

Two-way ANOVA

- Weight loss by type of diet and gender
- Can include tests for *interaction* between factors

One-way ANOVA compared to an independent t-test

A two sample t-test assuming equal variance and an ANOVA comparing only two groups will give you the exact same p-value (for a two-sided hypothesis).

H_0: $\mu_1 = \mu_2$ H_A: $\mu_1 \neq \mu_2$	H_0: $\mu_1 = \mu_2$ H_A: $\mu_1 \neq \mu_2$
One-way ANOVA	t-test assuming equal variance
F-statistic	t-statistic

But the t-test is more flexible

can run a t-test assuming unequal variance if you are not sure that the two populations have the same standard deviations.

One-way ANOVA: Null and Alternative Hypotheses
Example: categorical variable has three groups

H_0: $\mu_1 = \mu_2 = \mu_3$
H_A: at least group one is different from one other

H_A: $\mu_1 = \mu_2 \neq \mu_3$
-- or --
H_A: $\mu_1 \neq \mu_2 = \mu_3$
-- or --
H_A: $\mu_1 \neq \mu_2 \neq \mu_3$
-- or --
H_A: $\mu_1 = \mu_3 \neq \mu_2$

ANOVA Assumptions - basically the same as the t-test
Normality, homoscedasticity, independence

The continuous variable should be approximately normally distributed for *each group* of the categorical variable

- More important if groups have unequal n (unbalanced ANOVA)
    - The more unbalanced, the greater the problem
- If all groups are skewed in the same direction, some skew is OK
- Deviations from normality less important with large sample sizes

Homoscedasticity between the groups of the categorical variable

- The variance (SD) of the continuous variable is equal in each group of the categorical variable (use <2 rule)
- More important if groups have unequal n (unbalanced)

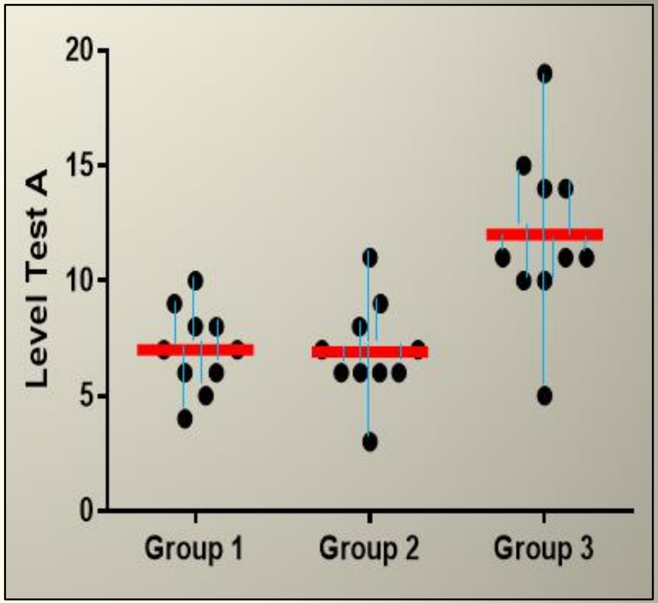
Categorical groups are independent

ANOVA Assumptions: Test on residuals

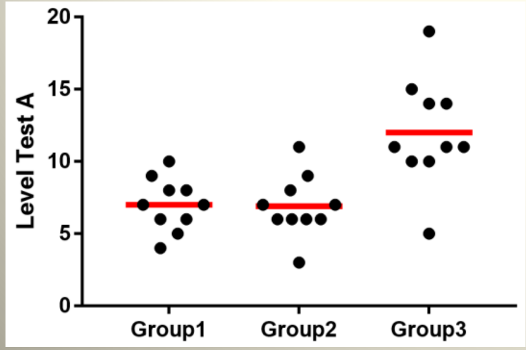
Assumption of normality and homoscedasticity relate to the residuals (AKA errors) of the model and not to the continuous data itself

However, for the ANOVA, they are equivalent

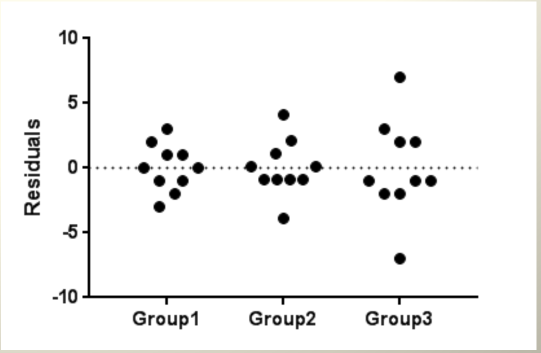
Much easier to test the continuous variable than to have to first compute the residuals and then test them.



Plot of raw data with means



Plot of residuals



Bartlett's and Brown and Forsythe tests for homoscedasticity: *Advice from the creators of Prism*

Bartlett's test works great if the data really are sampled from Gaussian distributions. But if the distributions deviate even slightly from the Gaussian ideal, Bartlett's test may report a small P value even when the differences among standard deviations is trivial.

Conclude there is a difference when there is really not one

For this reason, many do not recommend that test.

The Brown and Forsythe test has the same goal as the Bartlett's test, but is less sensitive to minor deviations from normality.

We suggest that you pay attention to the Brown-Forsythe result, and ignore Bartlett's test (left in to be consistent with prior versions of Prism).

But remember the lack of meaning in small and large sample sizes

If you conclude that the data are not normal or are heteroscedastic, you have some choices:

Transform the data to get a normal distribution and/or equalize the standard deviations, and then rerun the ANOVA.

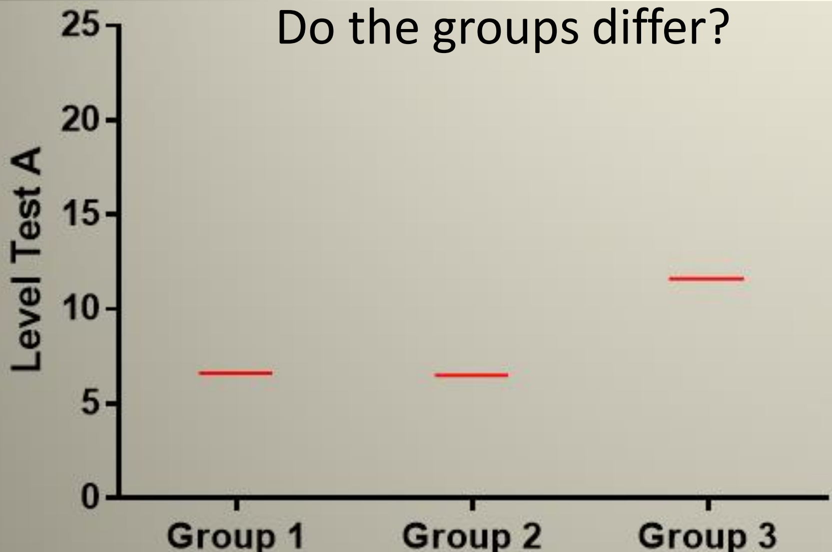
Converting values to their reciprocals or logarithms will often equalize he standard deviations and make the distributions more normal.

Use the nonparametric Kruskal-Wallis test

Or use the ANOVA anyway knowing it is not that sensitive to deviations from the normal distribution (very robust) and unequal variances (not quite as robust), especially if groups have equal n and the sample size is large (~30 or greater).

The Heart of an ANOVA: Between and Within Variability

Between Groups:
Our main question –
Do the groups differ?

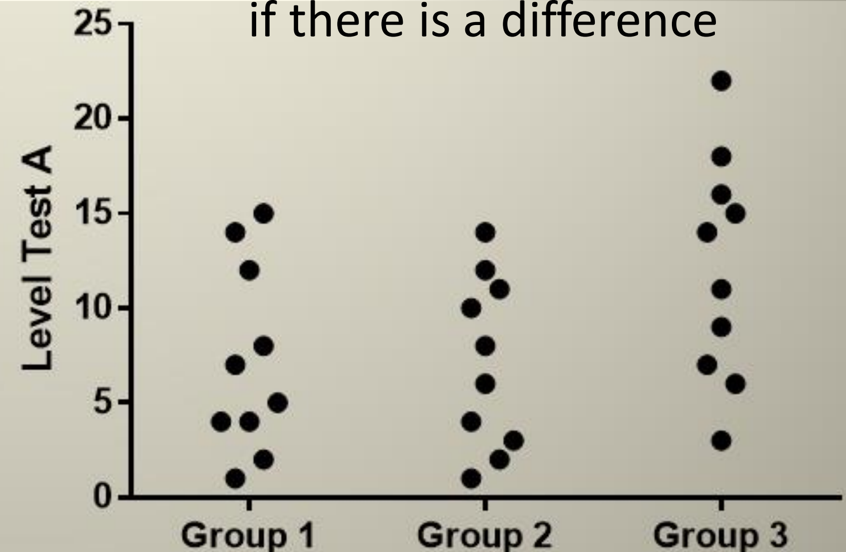


Looks at variability between group means

Are the means different?

Biological variability, error (sampling, measurement)

Within Each Group:
Variability that helps determines if there is a difference



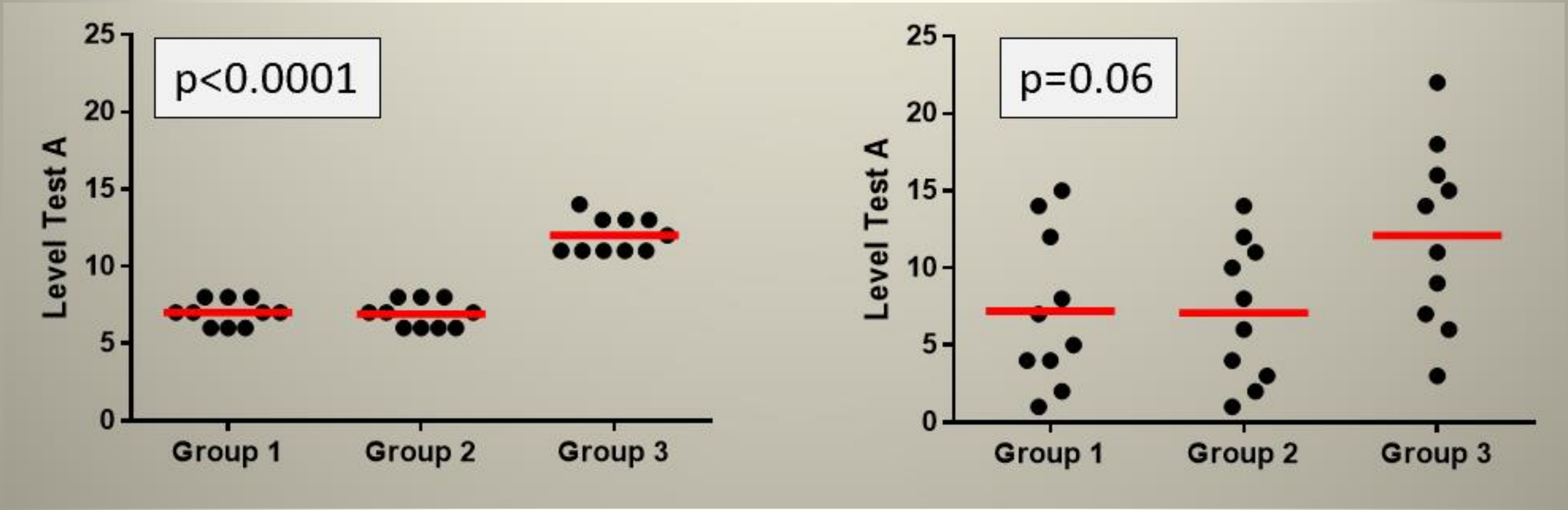
Looks at variability within each group

Is there a lot of spread (variability) in each group? Does the variability differ between groups (AKA heteroscedasticity)?

Biological variability, error

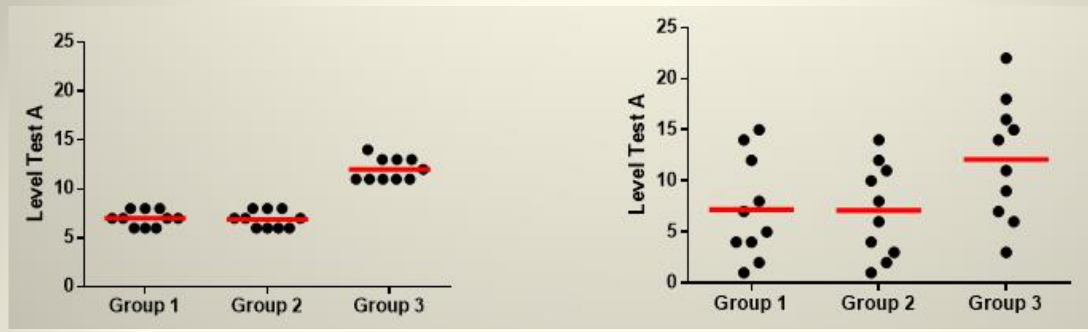
The Heart of an ANOVA: Between Groups and Within Groups

Mean Values for Test A



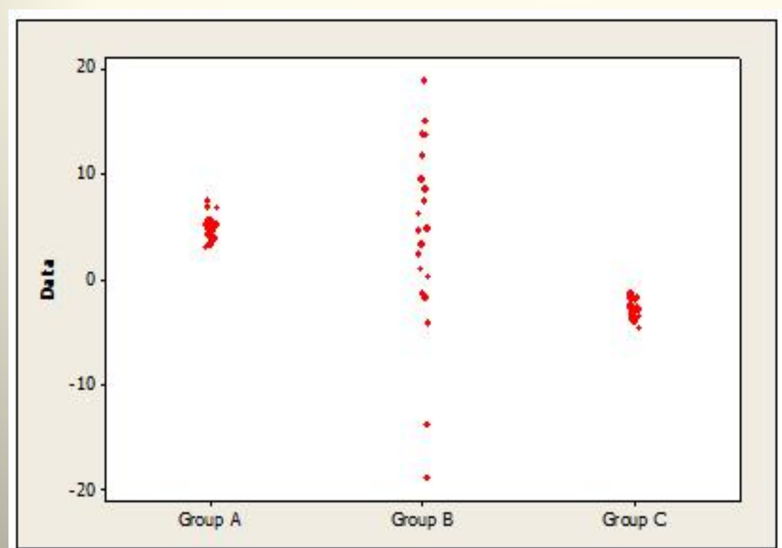
Between groups difference is what you are interested it, but within group variation can get in the way

One-Way ANOVA: test of means for >2 groups



The name analysis of *variance* comes from the way the test uses variability to decide whether the means are different.

This is why homoscedasticity (equality of variances in groups) is important



Group A and Group C exhibit homoscedasticity.

Group A and Group B exhibit heteroscedasticity (literally, "different scatter")

What about Groups B and C?

One-Way ANOVA: test of means for >2 groups

ANOVA tests for one overall effect only (AKA omnibus or global test)

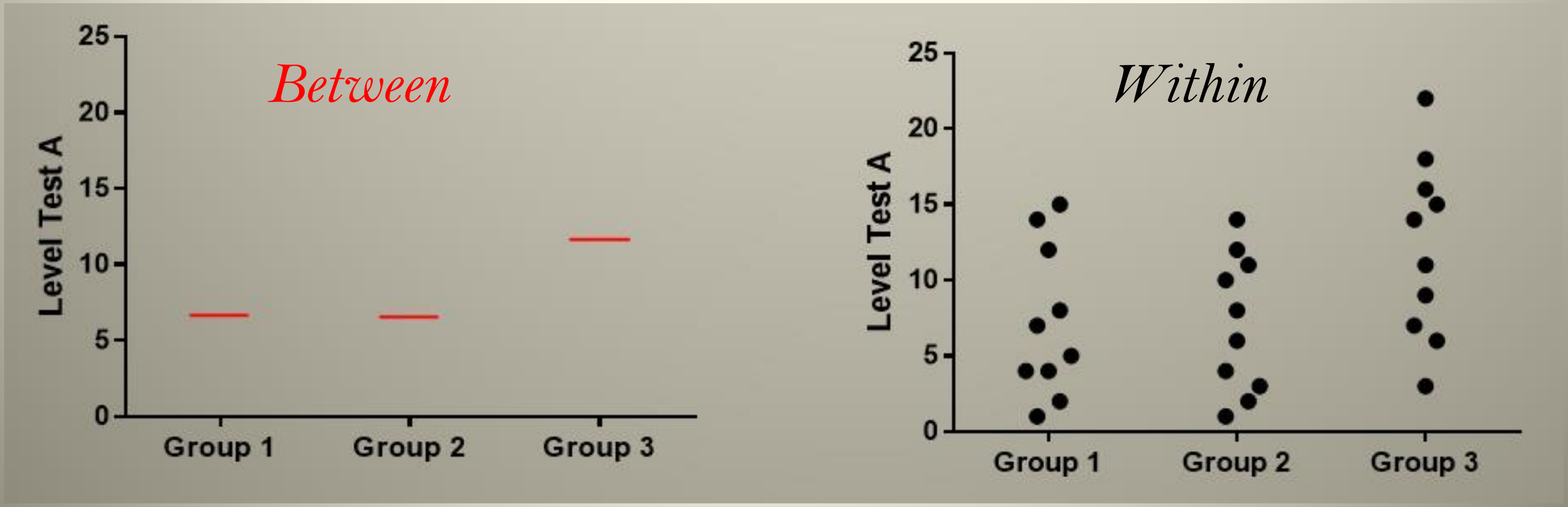
- Can tell you if there is a difference between groups, but not which ones
- Avoids the problem of multiple testing and reduces Type I error
- Uses the entire sample size
- If $p \leq 0.05$ requires post-hoc testing (AKA multiple testing, contrasts) to determine which groups differ from each other

One-Way ANOVA: test of means for >2 groups

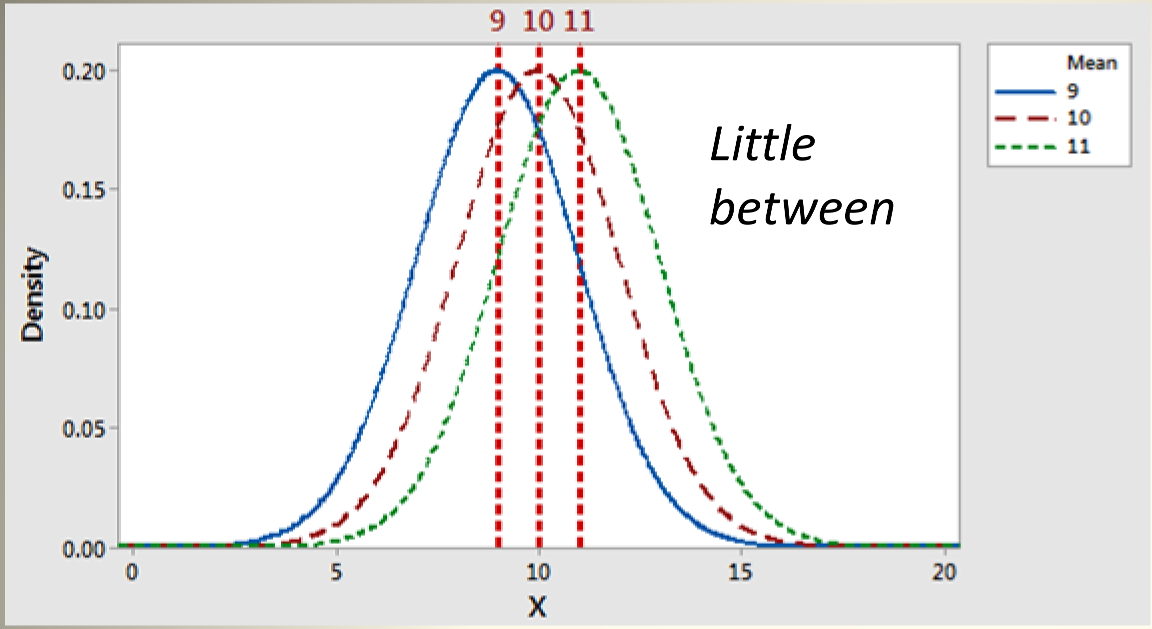
Uses the F statistic (AKA F-ratio; named after Ronald Fisher)

A ratio that compares variability *between* each group mean to variability *within* each group

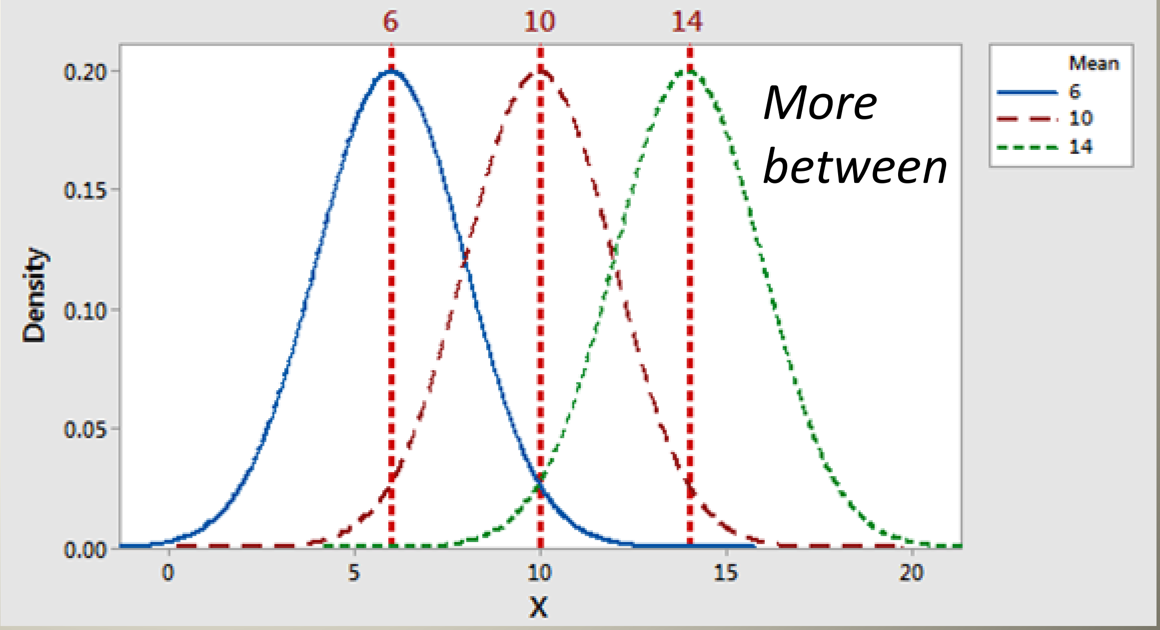
$$F = \frac{\text{Between variability}}{\text{Within variability}}$$



ANOVA: *Between* Variability (with same *within* variability)

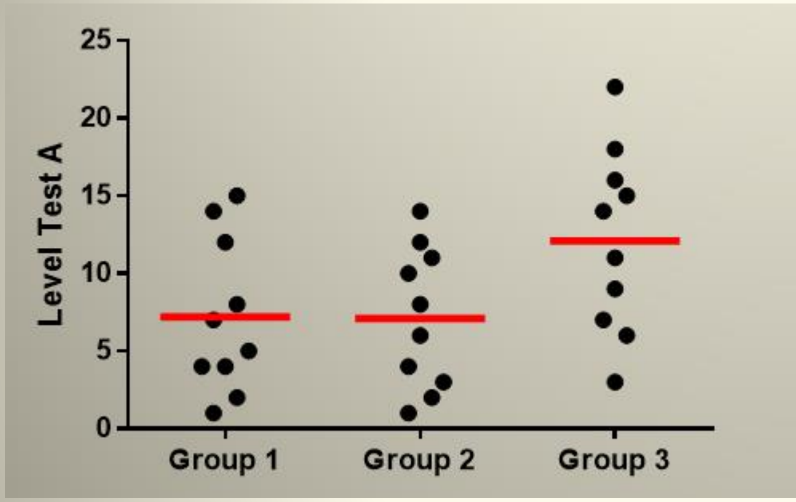


F = $\frac{\text{Between variability}}{\text{Within variability}}$

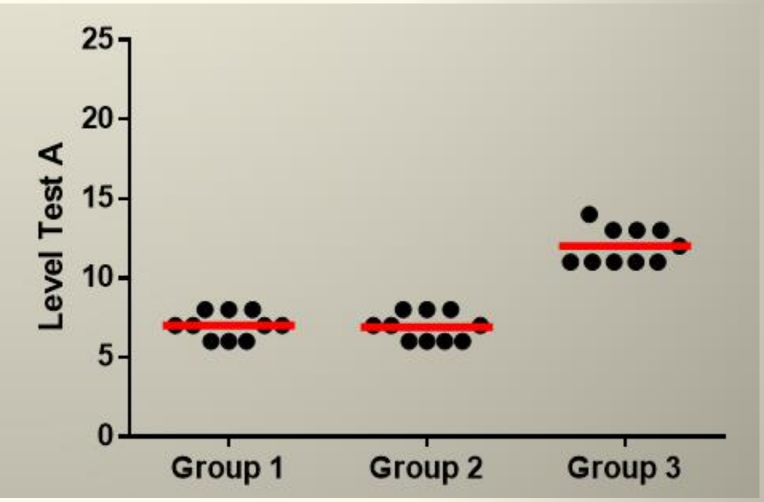


ANOVA: Within Variability

Lots of **within** variability



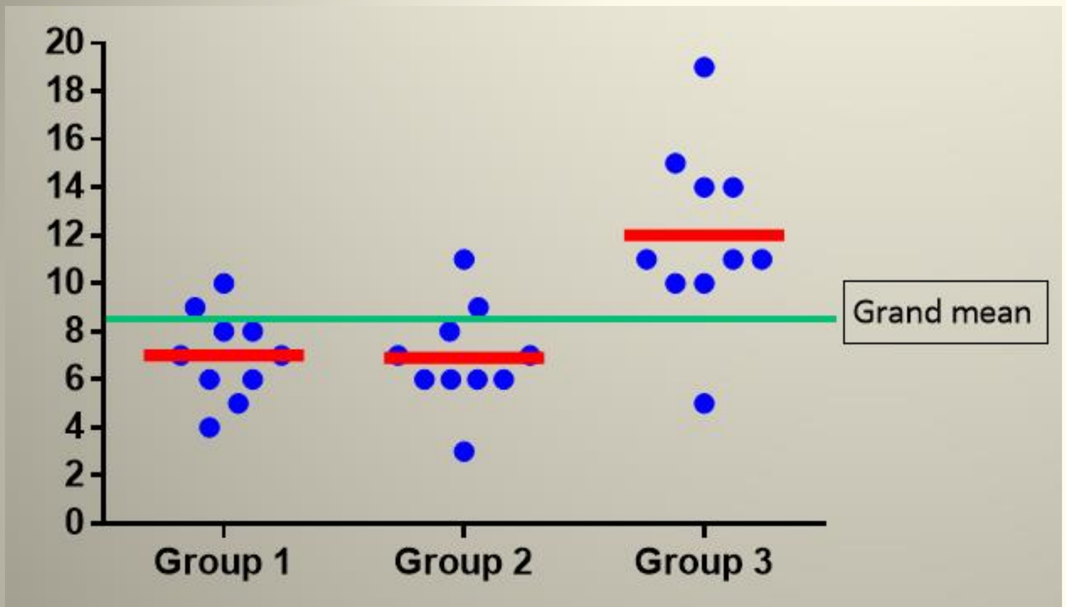
Little **within** variability



$$F = \frac{\text{Between variability}}{\text{Within variability}}$$

Something called “sum of squares” measures both types of variability

Sum of Squares **Between** and Sum of Squares **Within** and the Grand Mean (the mean of all samples regardless of group)



SS_**between** looks at how the means for each group vary from the grand mean

Answers your experimental question, but you need to account for variability **within** each group

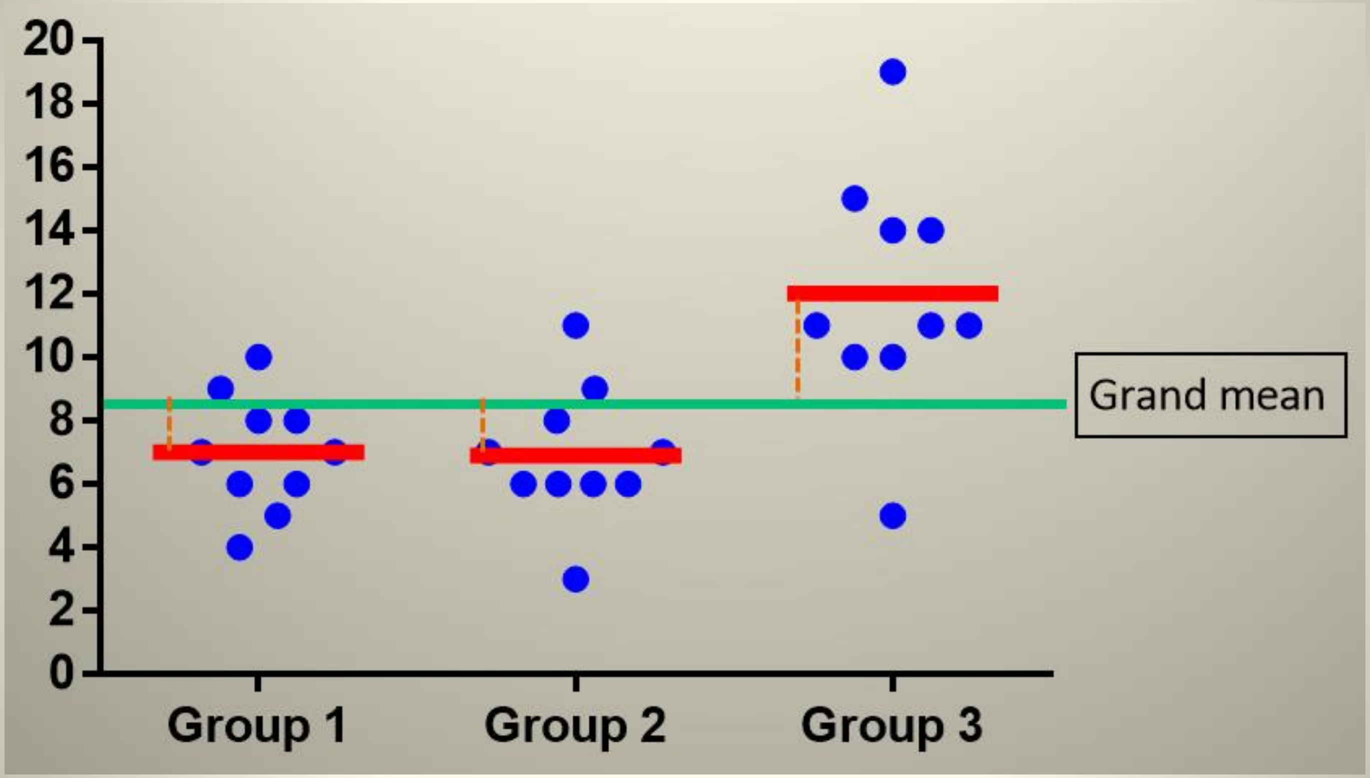
SS_**within** looks at how individual points in each group vary from mean for that group

Measures the "noise" from biological variability and error

The ANOVA test statistic, F, is the ratio of

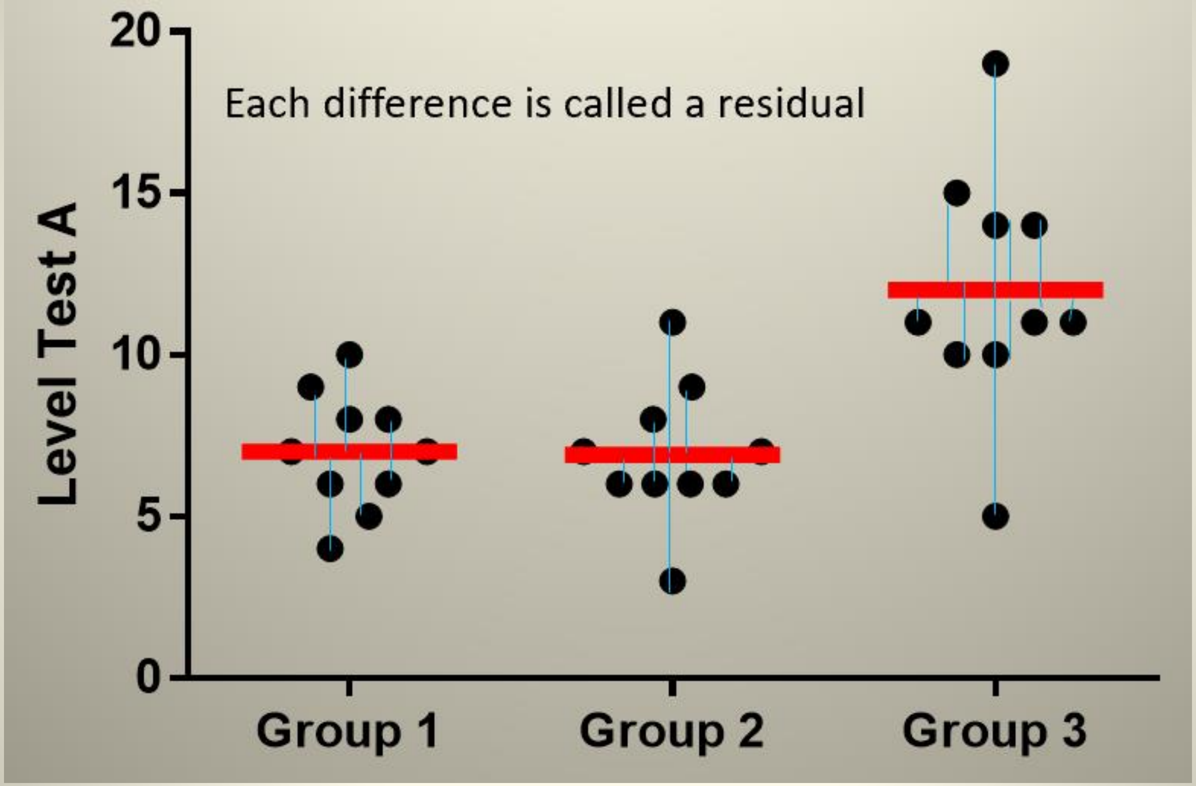
(SS_**between**/df)
―――――――――
(SS_**within**/df)

SS_**between** looks at how the means for each group vary from the grand mean



SS_**between** = $n_1 x(\;)^2 + n_2 x(\;)^2 + n_3 x(\;)^2$

SS_**within** looks at how individual points in each group vary from mean for that group



SS_**within** = sum of each (|)^2

The F Statistic

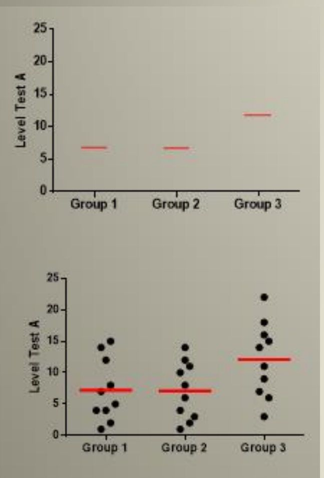
$$F = \frac{\dfrac{SS\_\mathbf{between}}{df\ (=\#groups - 1)}}{\dfrac{SS\_\mathbf{within}}{df\ (=n - \#groups)}} = \frac{\text{Mean of Squares\_}\mathbf{between}}{\text{Mean of Squares\_}\mathbf{within}}$$

#groups = 3 in our example
n = the total sample size, in our example n=30

An F > 1.0 indicates that there is more difference between group means than within groups.

The larger the F statistic, the more likely $p \leq 0.05$

ANOVA: Sum of Squares to Measure Variability
Example: 3 groups with 10 values in each group



The "grand mean" is the mean for all data combined

SS_**between** = sum of (mean for each group *minus* grand mean)^2

With 3 groups, you will have 3 calculations to sum

This measures the variability of means by group

SS_**within** = sum of (each individual value *minus* mean for their group)^2

With 3 groups of 10 each, you will have 30 calculations

This measures the variability of data points within each group

SS_total = sum of (each individual value *minus* grand mean)^2

Grand mean is the mean of all variables (or of 3 group means)

With 3 groups of 10 each, you will have 30 calculations to sum

This measures total variability

ANOVA Table

Summary ANOVA

Source	Sum of Squares	Degrees of Freedom	Variance Estimate (Mean Square)	F Ratio
Between	SS_B	$K-1$	$MS_B = \dfrac{SS_B}{K-1}$	$\boxed{\dfrac{MS_B}{MS_W}}$
Within	SS_W	$N-K$	$MS_W = \dfrac{SS_W}{N-K}$	
Total	$SS_T = SS_B + SS_W$	$N-1$		

ANOVA F Test Statistic

N = total observations in all groups

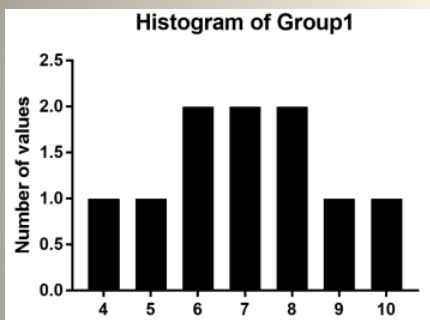
K = number of groups

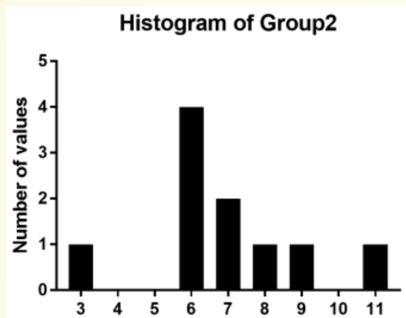
FYI

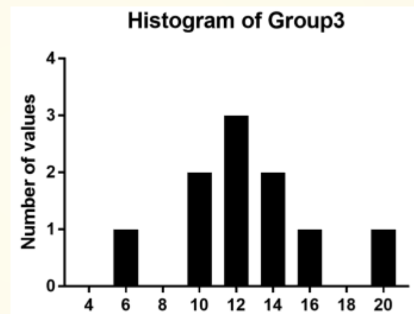
One-Way ANOVA by Hand: Assumptions

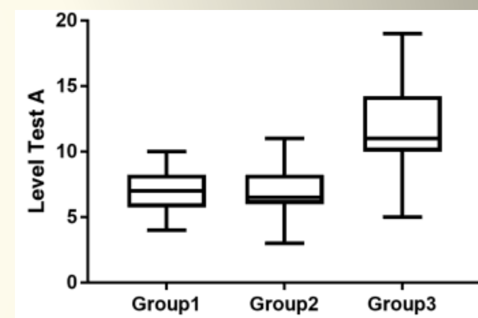
The three groups are <u>independent</u> (measurements done in different people)

The three groups appear to be fairly <u>symmetrical</u> without significant skew or outliers









<u>Are the SD the same?</u>

	Group 1	Group 2	Group3	Ratio largest/smallest
SD	1.83	2.13	3.74	3.74/1.83=2.04

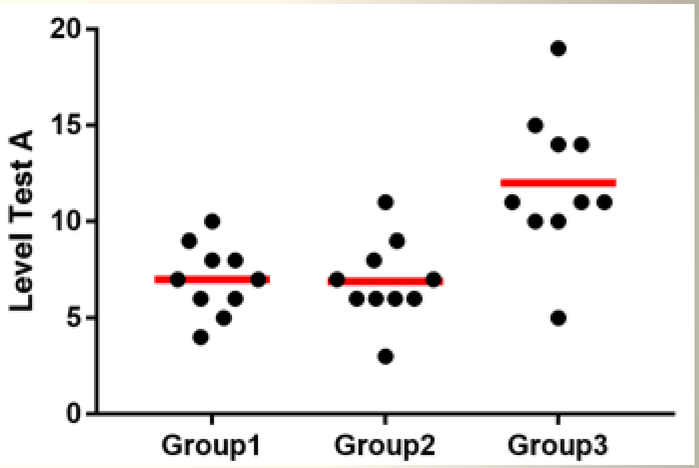
~~Brown-Forsythe test p=0.31 (SD not different)~~

FYI

One-Way ANOVA by Hand (with an assist from Excel)
Are the three groups different?

Number	group1	group2	group3
1	9	7	10
2	8	9	19
3	6	6	14
4	8	6	5
5	10	6	10
6	4	11	11
7	6	6	14
8	5	3	15
9	7	8	11
10	7	7	11
mean	7.0	6.9	12.0
SD	1.83	2.13	3.74

Univariate Scatter Plot with Means



For the purposes of this example, we will assume test assumptions are met

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

H_A: at least one is different from one other

$H_A: \mu_1 = \mu_2 \neq \mu_3$ -- or -- $H_A: \mu_1 \neq \mu_2 = \mu_3$ -- or --

$H_A: \mu_1 \neq \mu_2 \neq \mu_3$ -- or -- $H_A: \mu_1 = \mu_3 \neq \mu_2$

FYI

One-way ANOVA by Hand: Grand Mean

Number	group1	group2	group3
1	9	7	10
2	8	9	19
3	6	6	14
4	8	6	5
5	10	6	10
6	4	11	11
7	6	6	14
8	5	3	15
9	7	8	11
10	7	7	11
mean	7.0	6.9	12.0
SD	1.83	2.13	3.74

Group	All Data
1	9
1	8
1	6
1	8
1	10
1	4
1	6
1	5
1	7
1	7
2	7
2	9
2	6
2	6
2	6
2	11
2	6
2	3
2	8
2	7
3	10
3	19
3	14
3	5
3	10
3	11
3	14
3	15
3	11
3	11
grand mean	8.63

FYI

ANOVA by Hand: Sum of Squares Total

Number	group1	group2	group3
1	9	7	10
2	8	9	19
3	6	6	14
4	8	6	5
5	10	6	10
6	4	11	11
7	6	6	14
8	5	3	15
9	7	8	11
10	7	7	11
mean	7.0	6.9	12.0
SD	1.83	2.13	3.74

SS_total individual values

1. $(x_{11} - \bar{x})^2 = (9 - 8.63)^2 =$ 0.14
2. $(x_{12} - \bar{x})^2 = (8 - 8.63)^2 =$ 0.40
3. $(x_{13} - \bar{x})^2 = (6 - 8.63)^2 =$ 6.92
4. $(x_{14} - \bar{x})^2 = (8 - 8.63)^2 =$ 0.40
5. $(x_{15} - \bar{x})^2 = (10 - 8.63)^2 =$ 1.88

.

.

.

30. $(x_{310} - \bar{x})^2 = (11 - 8.63)^2 =$ 5.62

SS_total = 366.97

Grand mean = 8.63

FYI

One-Way ANOVA by Hand: Sum of Squares Between

Number	group1	group2	group3
1	9	7	10
2	8	9	19
3	6	6	14
4	8	6	5
5	10	6	10
6	4	11	11
7	6	6	14
8	5	3	15
9	7	8	11
10	7	7	11
mean	7.0	6.9	12.0
SD	1.83	2.13	3.74

SS_between individual values

$n_1(\bar{x}_1 - \bar{x})^2 = 10 \times (7.0 - 8.63)^2 = 26.57$

$n_2(\bar{x}_2 - \bar{x})^2 = 10 \times (6.9 - 8.63)^2 = 29.93$

$n_2(\bar{x}_2 - \bar{x})^2 = 10 \times (12.0 - 8.63)^2 = 113.57$

SS_between = 170.07

Grand mean = 8.63

n = 10 (for each group)

FYI

One-Way ANOVA by Hand: Sum of Squares Within

Number	group1	group2	group3
1	9	7	10
2	8	9	19
3	6	6	14
4	8	6	5
5	10	6	10
6	4	11	11
7	6	6	14
8	5	3	15
9	7	8	11
10	7	7	11
mean	7.0	6.9	12.0
SD	1.83	2.13	3.74

SS_Within individual values

1. $(x_{11} - \bar{x}_1)^2 = (9 - 7.0)^2 =$ 4.00
2. $(x_{12} - \bar{x}_1)^2 = (8 - 7.0)^2 =$ 1.00
3. $(x_{13} - \bar{x}_1)^2 = (6 - 7.0)^2 =$ 1.00
4. $(x_{14} - \bar{x}_1)^2 = (8 - 7.0)^2 =$ 1.00
5. $(x_{15} - \bar{x}_1)^2 = (10 - 7.0)^2 =$ 9.00

.

.

.

30. $(x_{310} - \bar{x}_3)^2 = (11 - 12.0)^2 =$ 1.00

SS_within = 196.90

ANOVA by Hand: Put it all together

Summary ANOVA

Source	Sum of Squares	Degrees of Freedom	Variance Estimate (Mean Square)	*F* Ratio
Between	SS_B	$K-1$	$MS_B = \frac{SS_B}{K-1}$	$\frac{MS_B}{MS_W}$
Within	SS_W	$N-K$	$MS_W = \frac{SS_W}{N-K}$	
Total	$SS_T = SS_B + SS_W$	$N-1$		

K = number of groups = 3
N = total sample size = 30

			Variance est	
Source	Sum of squares	df	(mean square)	F ratio
Between	170.07	2	85.04	11.66
Within	196.90	27	7.29	
Total	366.97	29		

$$F = \frac{MS_{Between}}{MS_{Within}}$$

= 85.04/7.29 = 11.66

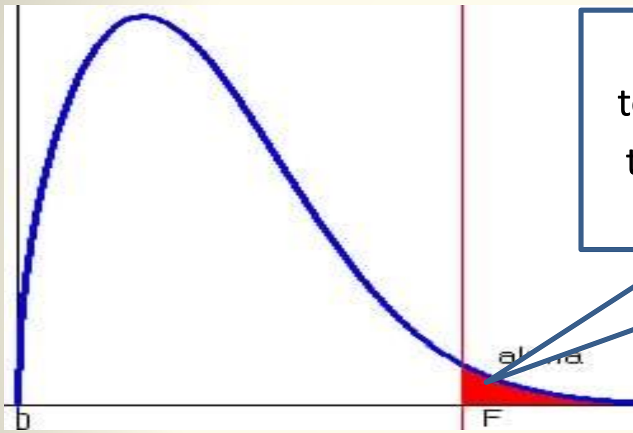
$\boxed{F(2,27) = 11.66}$

The ANOVA F Statistic is one-sided

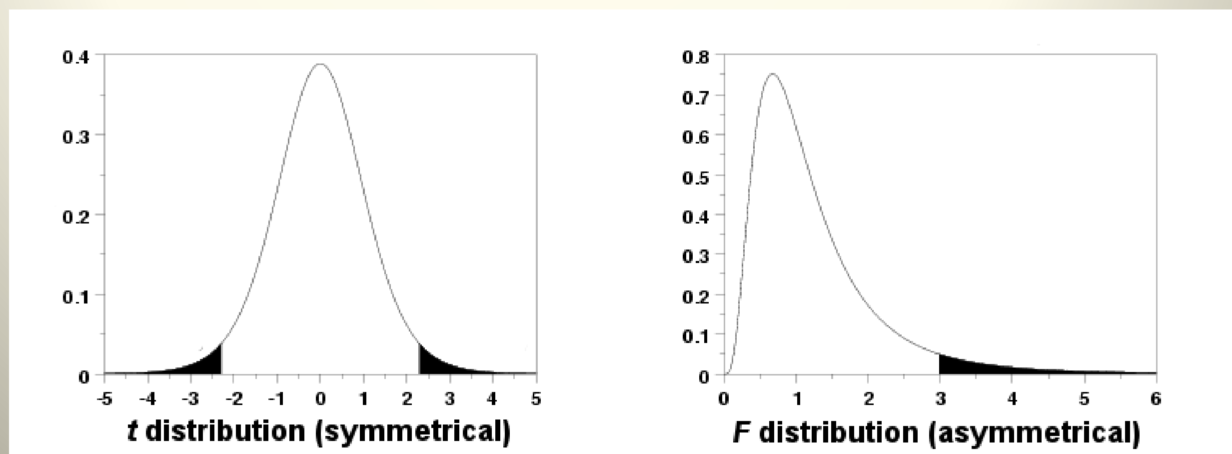
The p-value for ANOVA is calculated using the F-distribution

Because it squares the differences, the F statistic is always positive.

F can be <1.0, but not < 0.



p-value = probability of getting a test statistic at least as extreme as the one found from our sample if the null is true



The ANOVA F Statistic is a one-sided but the test can be two-tailed

The ANOVA F statistic is one-sided, but it is non-directional

a two-tailed test is non-directional, i.e., the difference can be either larger or smaller

As the difference between sample means grows larger (in EITHER direction), the F grows larger.

ANOVA is best thought of as a two-tailed test even though literally only one tail of the F distribution is used

Finding the Critical Value for F(2,27)

F - Distribution (α = 0.05 in the Right Tail)

df_2 \ df_1 (Denominator Degrees of Freedom \ Numerator Degrees of Freedom)	1	2	3	4	5	6	7	8	9
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
2	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385
3	10.128	9.5521	9.2766	9.1172	9.0135	8.9406	8.8867	8.8452	8.8123
4	7.7086	9.9443	6.5914	6.3882	6.2561	6.1631	6.0942	6.0410	6.9988
5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725
6	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2067	4.1468	4.0990
7	5.5914	4.7374	4.3468	4.1203	3.9715	3.8660	3.7870	3.7257	3.6767
8	5.3177	4.4590	4.0662	3.8379	3.6875	3.5806	3.5005	3.4381	3.3881
9	5.1174	4.2565	3.8625	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789
10	4.9646	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355	3.0717	3.0204
11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	2.9480	2.8962
12	4.7472	3.8853	3.4903	3.2592	3.1059	2.9961	2.9134	2.8486	2.7964
13	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153	2.8321	2.7669	2.7144
14	4.6001	3.7389	3.3439	3.1122	2.9582	2.8477	2.7642	2.6987	2.6458
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876
16	4.4940	3.6337	3.2389	3.0069	2.8524	2.7413	2.6572	2.5911	2.5377
17	4.4513	3.5915	3.1968	2.9647	2.8100	2.6987	2.6143	2.5480	2.4943
18	4.4139	3.5546	3.1599	2.9277	2.7729	2.6613	2.5767	2.5102	2.4563
19	4.3807	3.5219	3.1274	2.8951	2.7401	2.6283	2.5435	2.4768	2.4227
20	4.3512	3.4928	3.0984	2.8661	2.7109	2.5990	2.5140	2.4471	2.3928
21	4.3248	3.4668	3.0725	2.8401	2.6848	2.5727	2.4876	2.4205	2.3660
22	4.3009	3.4434	3.0491	2.8167	2.6613	2.5491	2.4638	2.3965	2.3419
23	4.2793	3.4221	3.0280	2.7955	2.6400	2.5277	2.4422	2.3748	2.3201
24	4.2597	3.4028	3.0088	2.7763	2.6207	2.5082	2.4226	2.3551	2.3002
25	4.2417	3.3852	2.9912	2.7587	2.6030	2.4904	2.4047	2.3371	2.2821
26	4.2252	3.3690	2.9752	2.7426	2.5868	2.4741	2.3883	2.3205	2.2655
27	4.2100	3.3541	2.9604	2.7278	2.5719	2.4591	2.3732	2.3053	2.2501
28	4.1960	3.3404	2.9467	2.7141	2.5581	2.4453	2.3593	2.2913	2.2360
29	4.1830	3.3277	2.9340	2.7014	2.5454	2.4324	2.3463	2.2783	2.2229
30	4.1709	3.3158	2.9223	2.6896	2.5336	2.4205	2.3343	2.2662	2.2107
40	4.0847	3.2317	2.8387	2.6060	2.4495	2.3359	2.2490	2.1802	2.1240
60	4.0012	3.1504	2.7581	2.5252	2.3683	2.2541	2.1665	2.0970	2.0401
120	3.9201	3.0718	2.6802	2.4472	2.2899	2.1750	2.0868	2.0164	1.9588
∞	3.8415	2.9957	2.6049	2.3719	2.2141	2.0986	2.0096	1.9384	1.8799

$F(2,27) = 3.3541$

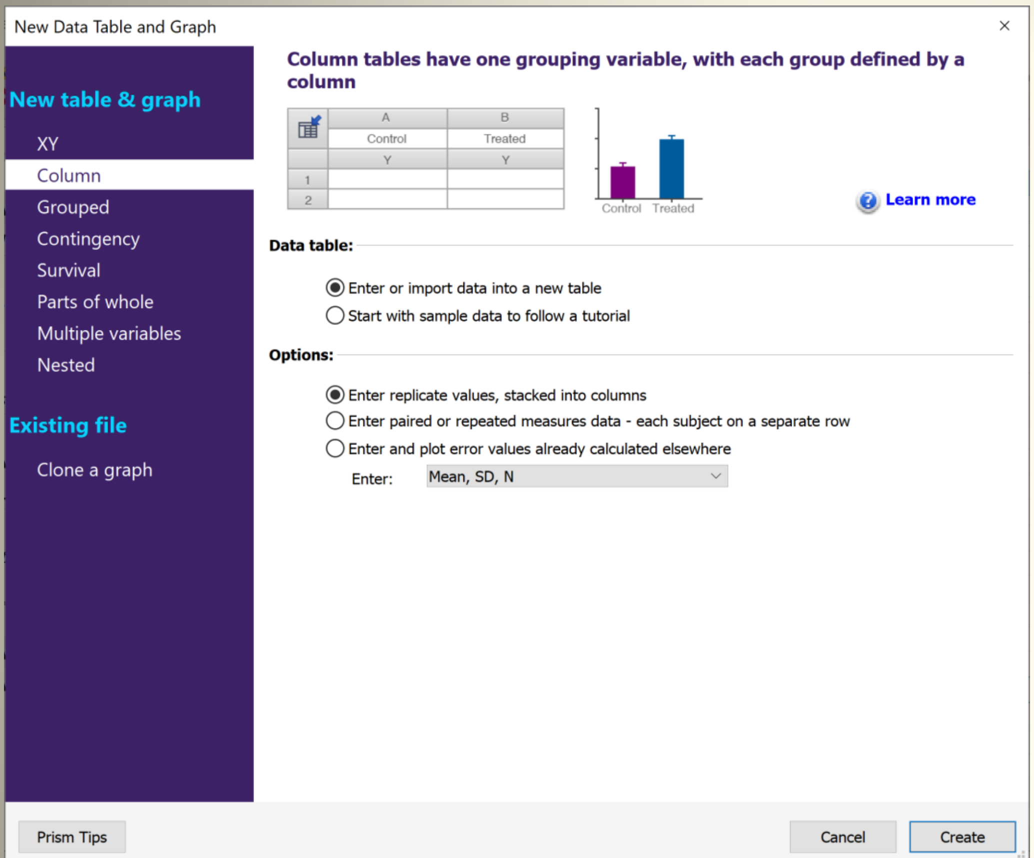
Our F statistic = 11.66

$11.66 > 3.3541$ so $p<0.05$

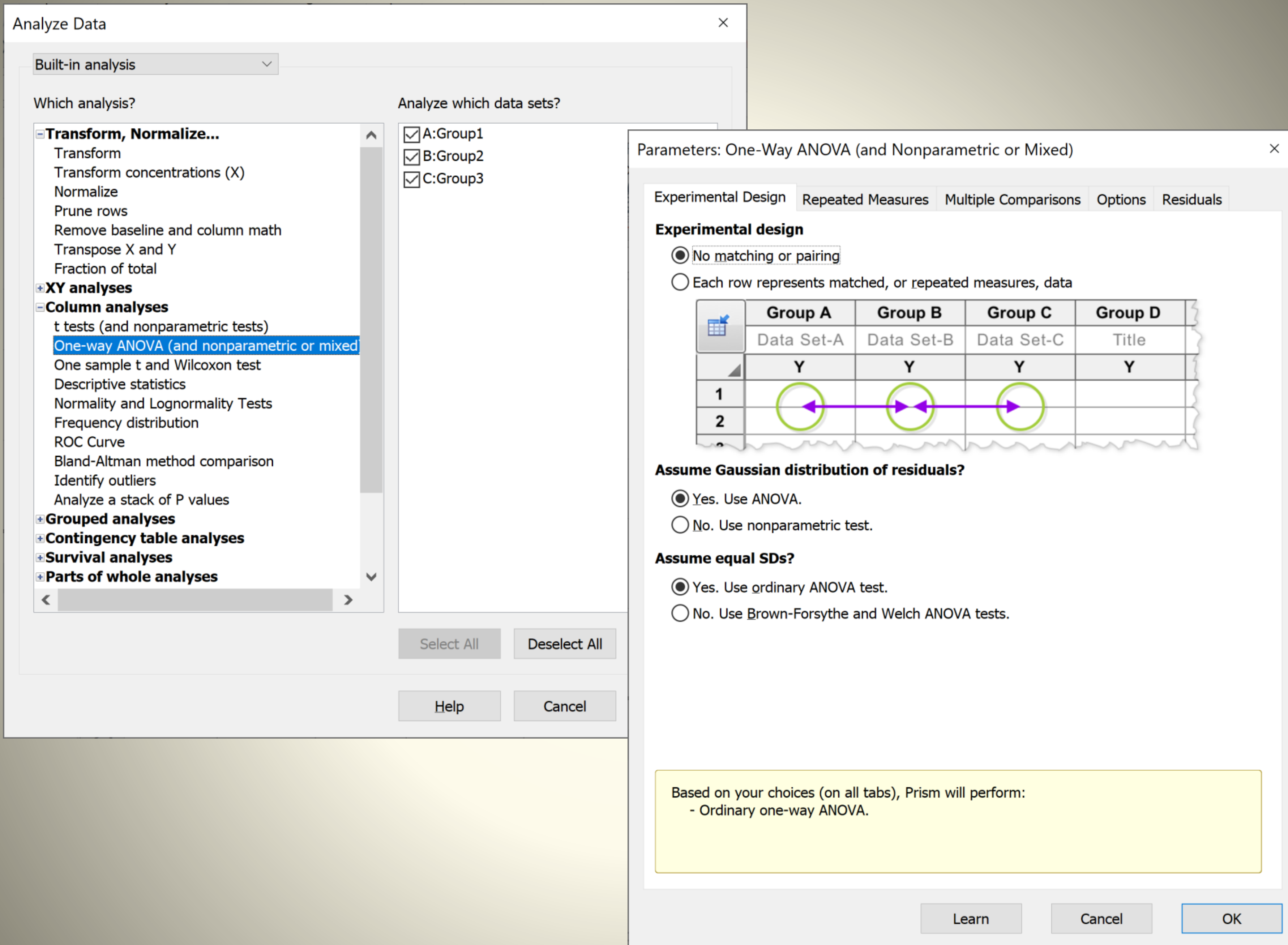
We reject the null hypothesis and conclude that at least one mean is different from one other

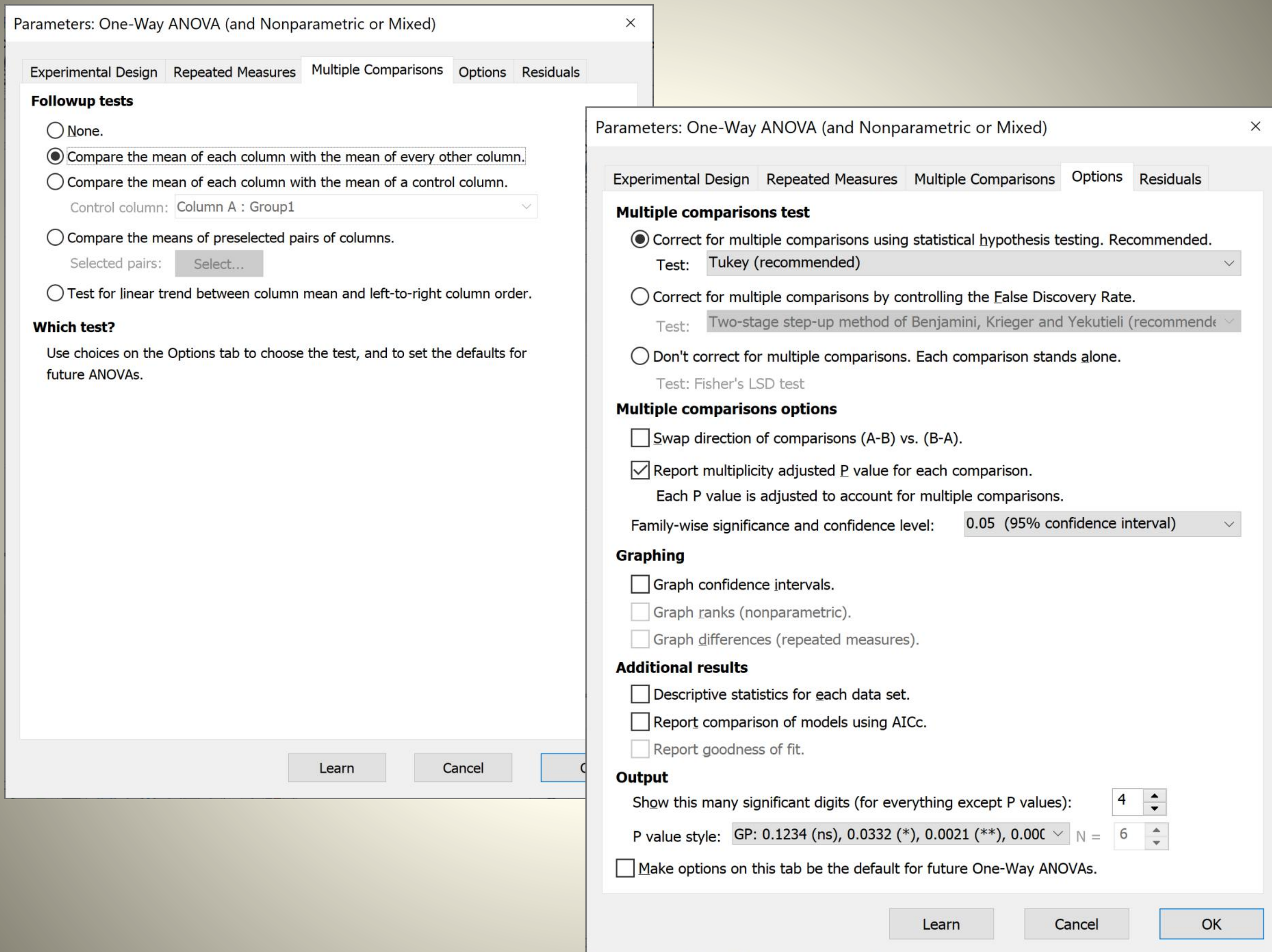
Now, let's use Prism to do the one-way ANOVA

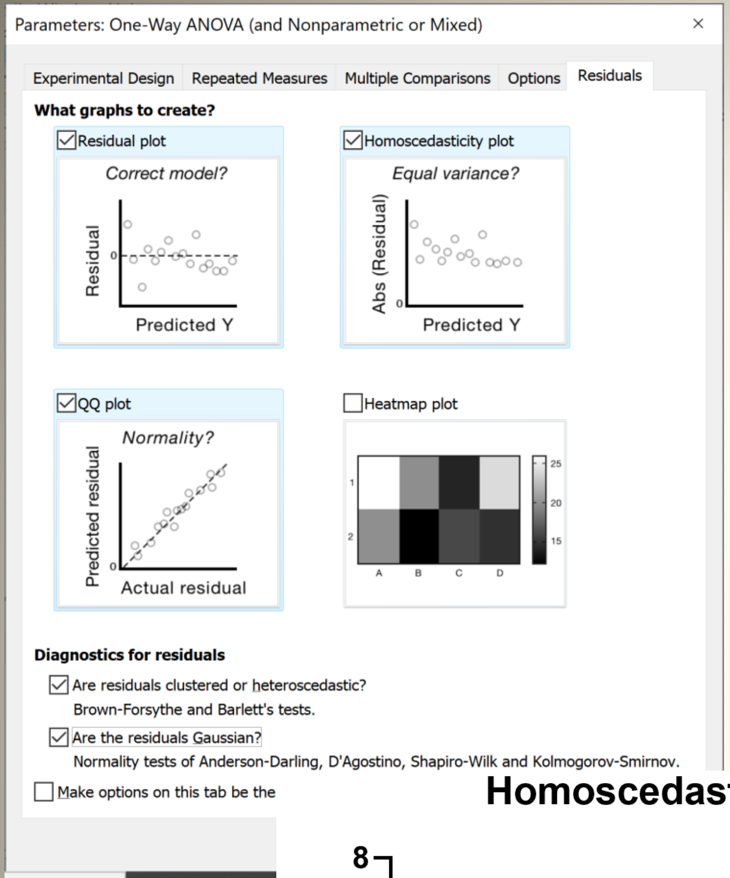
ANOVAbyHand.xlsx

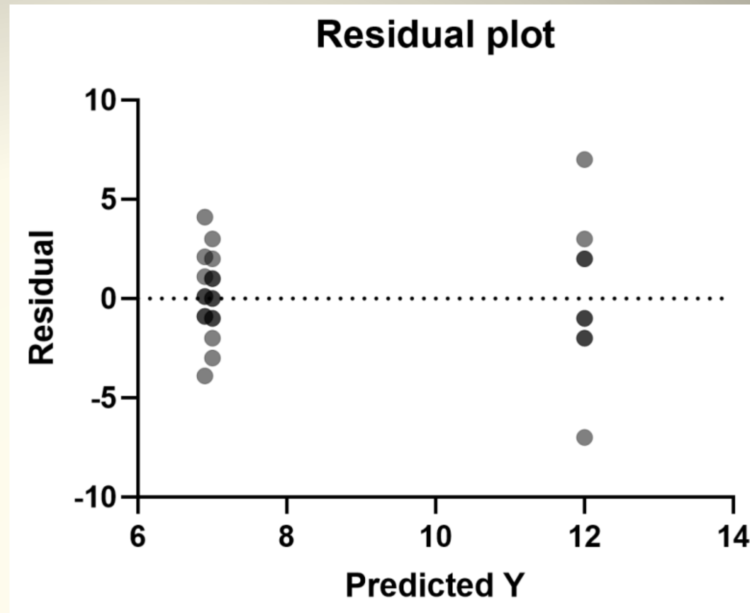


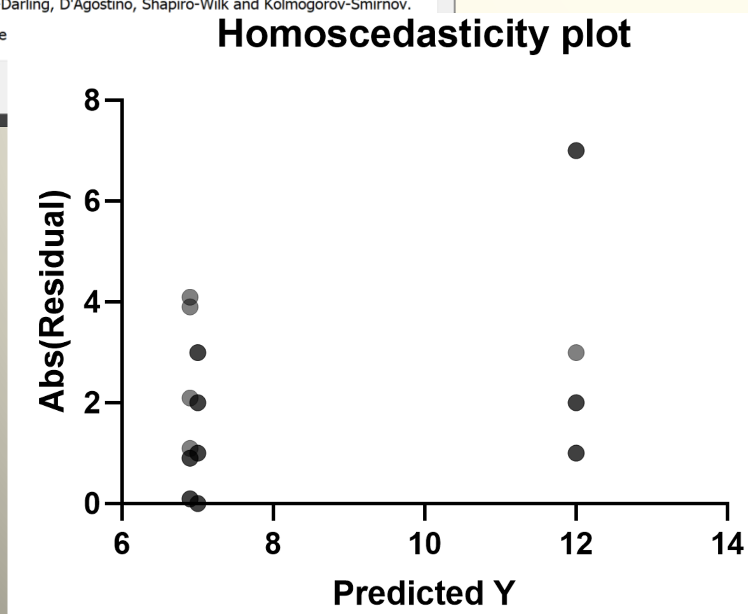
	Group A	Group B	Group C
	Group1	Group2	Group3
1	9	7	10
2	8	9	19
3	6	6	14
4	8	6	5
5	10	6	10
6	4	11	11
7	6	6	14
8	5	3	15
9	7	8	11
10	7	7	11
11			

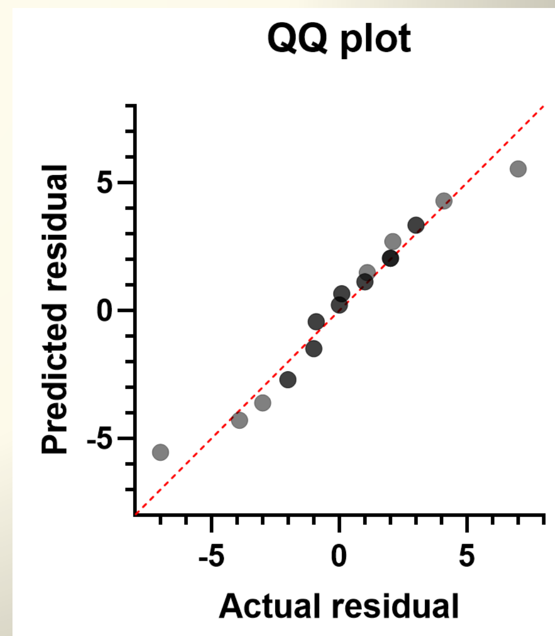

Analyze Data
Built-in analysis
Which analysis?
Transform, Normalize...
Transform
Transform concentrations (X)
Normalize
Prune rows
Remove baseline and column math
Transpose X and Y
Fraction of total
XY analyses
Column analyses
t tests (and nonparametric tests)
One-way ANOVA (and nonparametric or mixed)
One sample t and Wilcoxon test
Descriptive statistics
Normality and Lognormality Tests
Frequency distribution
ROC Curve
Bland-Altman method comparison
Identify outliers
Analyze a stack of P values
Grouped analyses
Contingency table analyses
Survival analyses
Parts of whole analyses
Analyze which data sets?
A:Group1
B:Group2
C:Group3
Select All
Deselect All
Help
Cancel
Parameters: One-Way ANOVA (and Nonparametric or Mixed)
Experimental Design
Repeated Measures
Multiple Comparisons
Options
Residuals
Experimental design
No matching or pairing
Each row represents matched, or repeated measures, data
Group A
Group B
Group C
Group D
Data Set-A
Data Set-B
Data Set-C
Title
Y
Y
Y
Y
1
2
Assume Gaussian distribution of residuals?
Yes. Use ANOVA.
No. Use nonparametric test.
Assume equal SDs?
Yes. Use ordinary ANOVA test.
No. Use Brown-Forsythe and Welch ANOVA tests.
Based on your choices (on all tabs), Prism will perform:
- Ordinary one-way ANOVA.
Learn
Cancel
OK


Parameters: One-Way ANOVA (and Nonparametric or Mixed)
Experimental Design
Repeated Measures
Multiple Comparisons
Options
Residuals
Followup tests
None.
Compare the mean of each column with the mean of every other column.
Compare the mean of each column with the mean of a control column.
Control column: Column A : Group1
Compare the means of preselected pairs of columns.
Selected pairs: Select...
Test for linear trend between column mean and left-to-right column order.
Which test?
Use choices on the Options tab to choose the test, and to set the defaults for future ANOVAs.
Learn
Cancel
Parameters: One-Way ANOVA (and Nonparametric or Mixed)
Experimental Design
Repeated Measures
Multiple Comparisons
Options
Residuals
Multiple comparisons test
Correct for multiple comparisons using statistical hypothesis testing. Recommended.
Test: Tukey (recommended)
Correct for multiple comparisons by controlling the False Discovery Rate.
Test: Two-stage step-up method of Benjamini, Krieger and Yekutieli (recommende
Don't correct for multiple comparisons. Each comparison stands alone.
Test: Fisher's LSD test
Multiple comparisons options
Swap direction of comparisons (A-B) vs. (B-A).
Report multiplicity adjusted P value for each comparison.
Each P value is adjusted to account for multiple comparisons.
Family-wise significance and confidence level: 0.05 (95% confidence interval)
Graphing
Graph confidence intervals.
Graph ranks (nonparametric).
Graph differences (repeated measures).
Additional results
Descriptive statistics for each data set.
Report comparison of models using AICc.
Report goodness of fit.
Output
Show this many significant digits (for everything except P values): 4
P value style: GP: 0.1234 (ns), 0.0332 (*), 0.0021 (**), 0.00(N = 6
Make options on this tab be the default for future One-Way ANOVAs.
Learn
Cancel
OK


Parameters: One-Way ANOVA (and Nonparametric or Mixed)
Experimental Design
Repeated Measures
Multiple Comparisons
Options
Residuals
What graphs to create?
Residual plot
Correct model?
Residual
Predicted Y
Homoscedasticity plot
Equal variance?
Abs (Residual)
Predicted Y
QQ plot
Normality?
Predicted residual
Actual residual
Heatmap plot
Diagnostics for residuals
Are residuals clustered or heteroscedastic?
Brown-Forsythe and Barlett's tests.
Are the residuals Gaussian?
Normality tests of Anderson-Darling, D'Agostino, Shapiro-Wilk and Kolmogorov-Smirnov.
Make options on this tab be the


Residual plot
Residual
10
5
0
-5
-10
6
8
10
12
14
Predicted Y


Homoscedasticity plot
Abs(Residual)
8
6
4
2
0
6
8
10
12
14
Predicted Y


QQ plot
Predicted residual
5
0
-5
-5
0
5
Actual residual

ANOVA results | Multiple comparisons

Ordinary one-way ANOVA
ANOVA results

1	Table Analyzed	One Way ANOVA
2	Data sets analyzed	A-C
3		
4	**ANOVA summary**	
5	F	11.66
6	P value	0.0002
7	P value summary	***
8	Significant diff. among means (P < 0.05)?	Yes
9	R square	0.4634
10		
11	**Brown-Forsythe test**	
12	F (DFn, DFd)	1.212 (2, 27)
13	P value	0.3134
14	P value summary	ns
15	Are SDs significantly different (P < 0.05)?	No
16		
17	**Bartlett's test**	
18	Bartlett's statistic (corrected)	5.177
19	P value	0.0751
20	P value summary	ns
21	Are SDs significantly different (P < 0.05)?	No

23	**ANOVA table**	**SS**	**DF**	**MS**	**F (DFn, DFd)**	**P value**
24	Treatment (between columns)	170.1	2	85.03	F (2, 27) = 11.66	P=0.0002
25	Residual (within columns)	196.9	27	7.293		
26	Total	367.0	29			
27						
28	**Normality of Residuals**					
29	**Test name**	**Statistics**	**P value**	**Passed norm**	**P value summary**	
30	Anderson-Darling (A2*)	0.6244	0.0943	Yes	ns	
31	D'Agostino-Pearson omnibus (K2)	3.481	0.1755	Yes	ns	
32	Shapiro-Wilk (W)	0.9538	0.2136	Yes	ns	
33	Kolmogorov-Smirnov (distance)	0.1506	0.0807	Yes	ns	
34						
35	**Data summary**					
36	Number of treatments (columns)	3				
37	Number of values (total)	30				

R-Square defines the proportion of the total variance explained by the model.
How much of the variance in the variable is explained by group? 46.4%

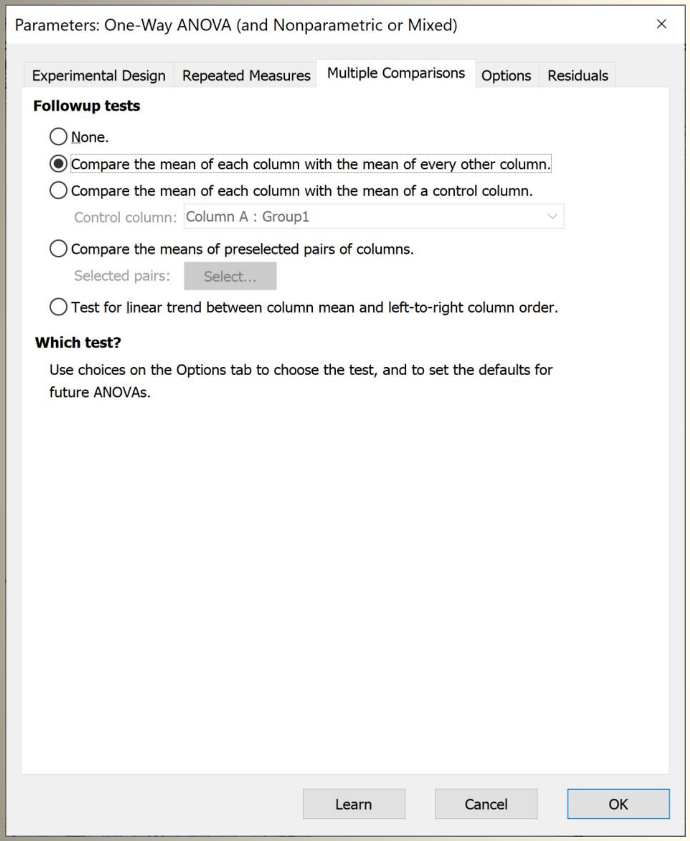
Prism Results

ANOVA table	SS	DF	MS	F (DFn, DFd)	P value
Treatment (between columns)	170.1	2	85.03	F (2, 27) = 11.66	P=0.0002
Residual (within columns)	196.9	27	7.293		
Total	367.0	29			

Our By Hand Results

			Variance est	
Source	Sum of squares	df	(mean square)	F ratio
Between	170.07	2	85.04	11.66
Within	196.90	27	7.29	
Total	366.97	29		

Post hoc tests (AKA Multiple Comparison Tests)



If there is a significant ANOVA result, you need to run post hoc tests to tell which groups are different

Pairwise comparisons are made to determine which groups are different from the others

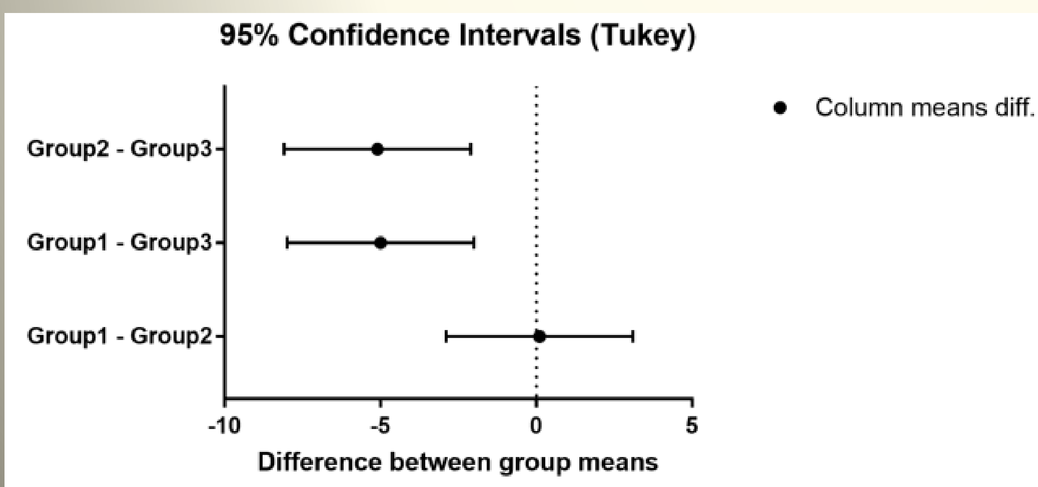
These post-hoc tests are t-tests with p-value adjustments to keep the type 1 error (false positives) to a minimum

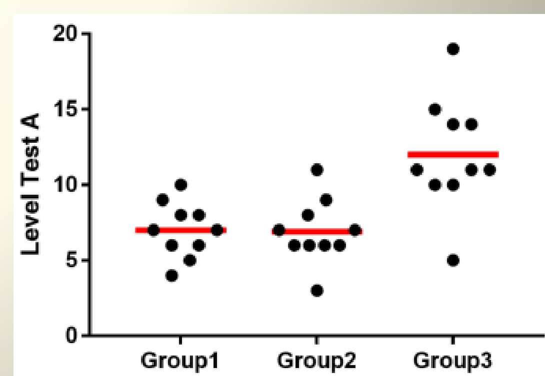
For the ANOVA, Prism recommends the commonly used Tukey's test

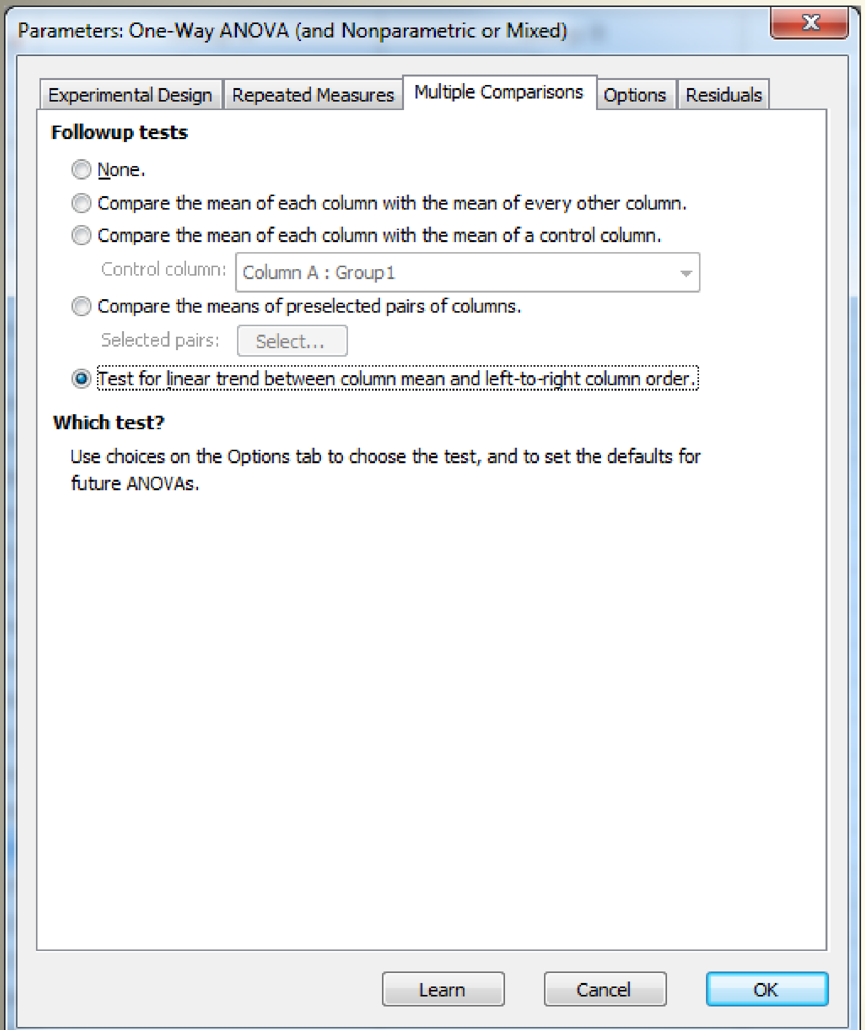
ANOVA results | Multiple comparisons

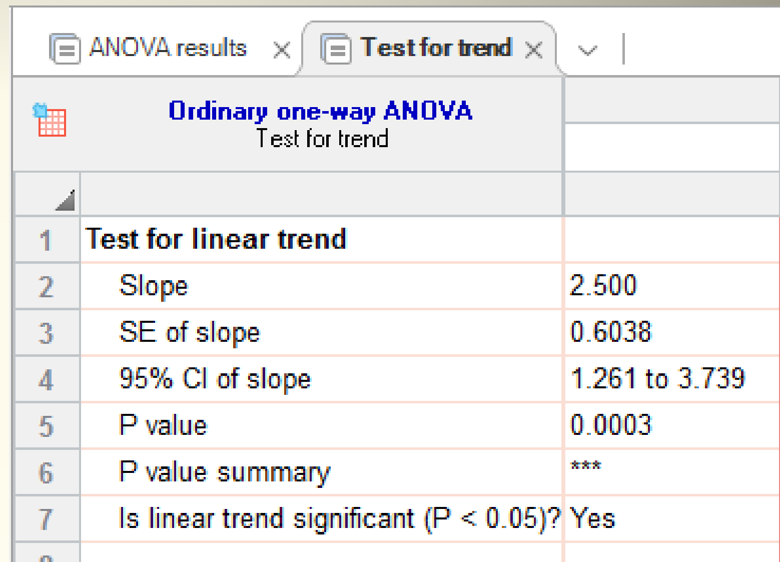
Ordinary one-way ANOVA
Multiple comparisons

1	Number of families							
2	Number of comparisons per family							
3	Alpha							
4								
5	**Tukey's multiple comparisons test**	**95.00% CI of diff.**	**Significant?**	**Summary**	**Adjusted P Value**			
6	Group1 vs. Group2	-2.894 to 3.094	No	ns	0.9962	A-B		
7	Group1 vs. Group3	-7.994 to -2.006	Yes	***	0.0009	A-C		
8	Group2 vs. Group3	-8.094 to -2.106	Yes	***	0.0007	B-C		
9								
10	**Test details**	**Mean 2**	**Mean Diff.**	**SE of diff.**	**n1**	**n2**	**q**	**DF**
11	Group1 vs. Group2	6.900	0.1000	1.208	10	10	0.1171	27
12	Group1 vs. Group3	12.00	-5.000	1.208	10	10	5.855	27
13	Group2 vs. Group3	12.00	-5.100	1.208	10	10	5.972	27

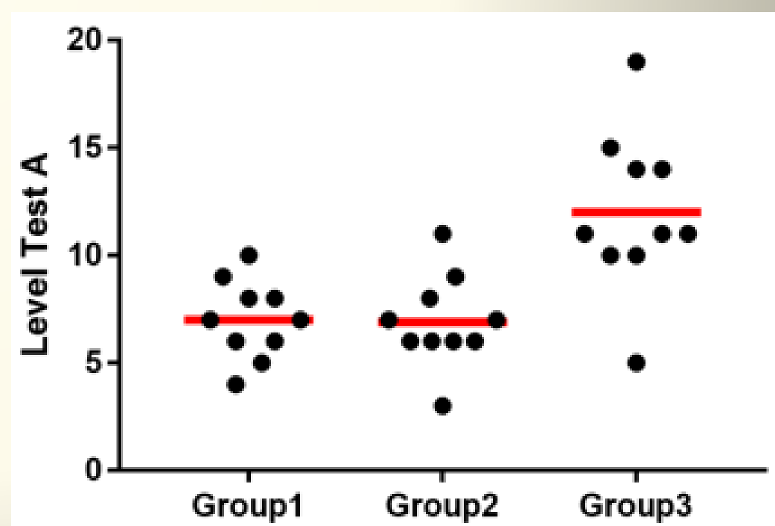








1	Test for linear trend	
2	Slope	2.500
3	SE of slope	0.6038
4	95% CI of slope	1.261 to 3.739
5	P value	0.0003
6	P value summary	***
7	Is linear trend significant (P < 0.05)?	Yes



Write up results

For this class:
The data passed the assumptions for an one-way ANOVA (independence, normality/symmetry, homoscedasticity). With $p=0.0002$ [two-tailed hypothesis, $F(2,27)=11.66$, $\alpha=0.05$), we reject the null hypothesis of no difference in test population means by group and conclude there is at least one group different from one other. Tukey post-hoc tests showed that Group3 (mean=12.0) was significantly different from Group1 (7.0) and Group2 (6.9). There was no difference between Groups 1 and 2.

How you may see it:
We found there was a significant differences between the groups ($p=0.0002$). Post-hoc tests found the differences were between Group3 (mean=12.0) and Group1 (7.0) and Group2 (6.9).

The Kruskal-Wallis Test:
The non-parametric alternative to the one-way ANOVA

Is the Mann-Whitney test extended to >2 groups

Use when the assumptions of one-way ANOVA are not met

Still assumes that observations are independently sampled and the groups are independent

Null hypothesis is similar to the Mann-Whitney test which can be simplified to hypothesis about the medians with additional assumption (same distribution)

Kruskal Wallis test

Is a global test like the ANOVA

tells us if at least one of the groups is different, but not which group is different

Can follow up a significant Kruskal-Wallis test with pairwise Wilcoxon Rank Sum tests to determine which pairs are different from each other.

Prism uses Dunn's method to adjust p-values for multiple comparisons

Kruskal-Wallis Example

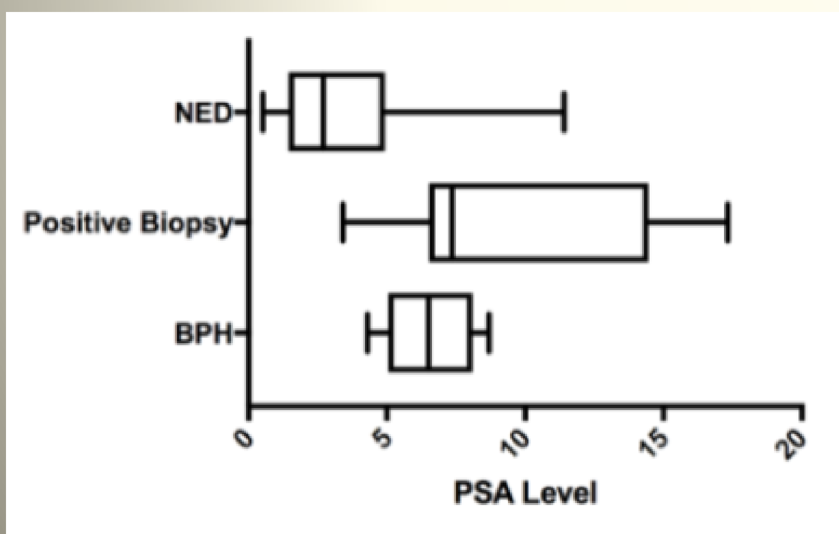
We want to compare PSA levels between 3 groups of patients:

6 with benign prostatic heterotrophy (BPH),

8 with positive biopsy for prostate cancer, and

8 with negative biopsies and no evidence of disease (NED)

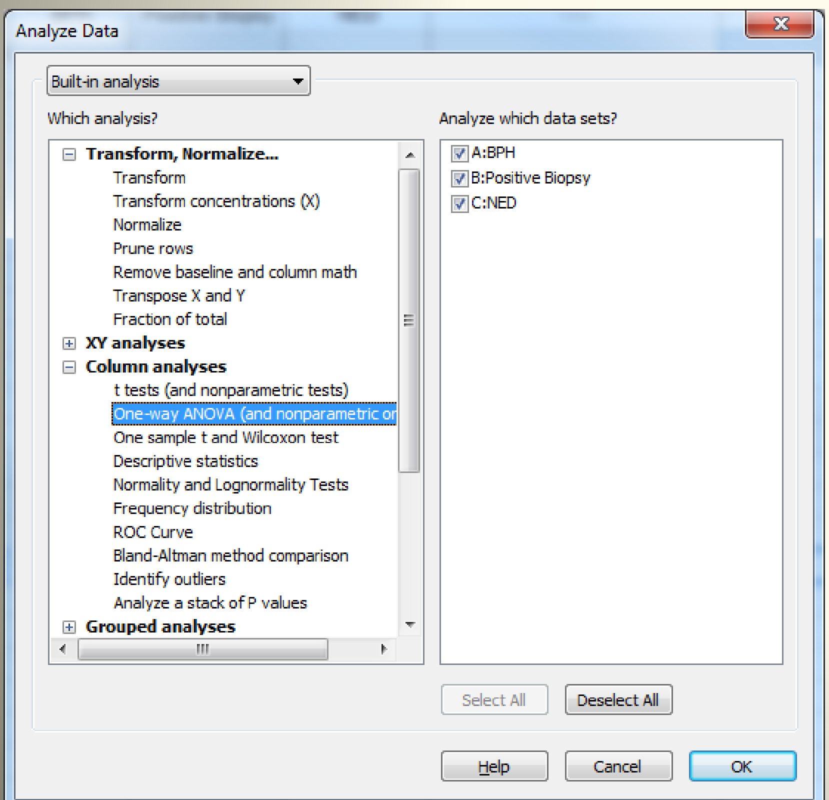
Data are unbalanced, sample sizes are small, and data appear skewed, >2 magnitude difference between SD

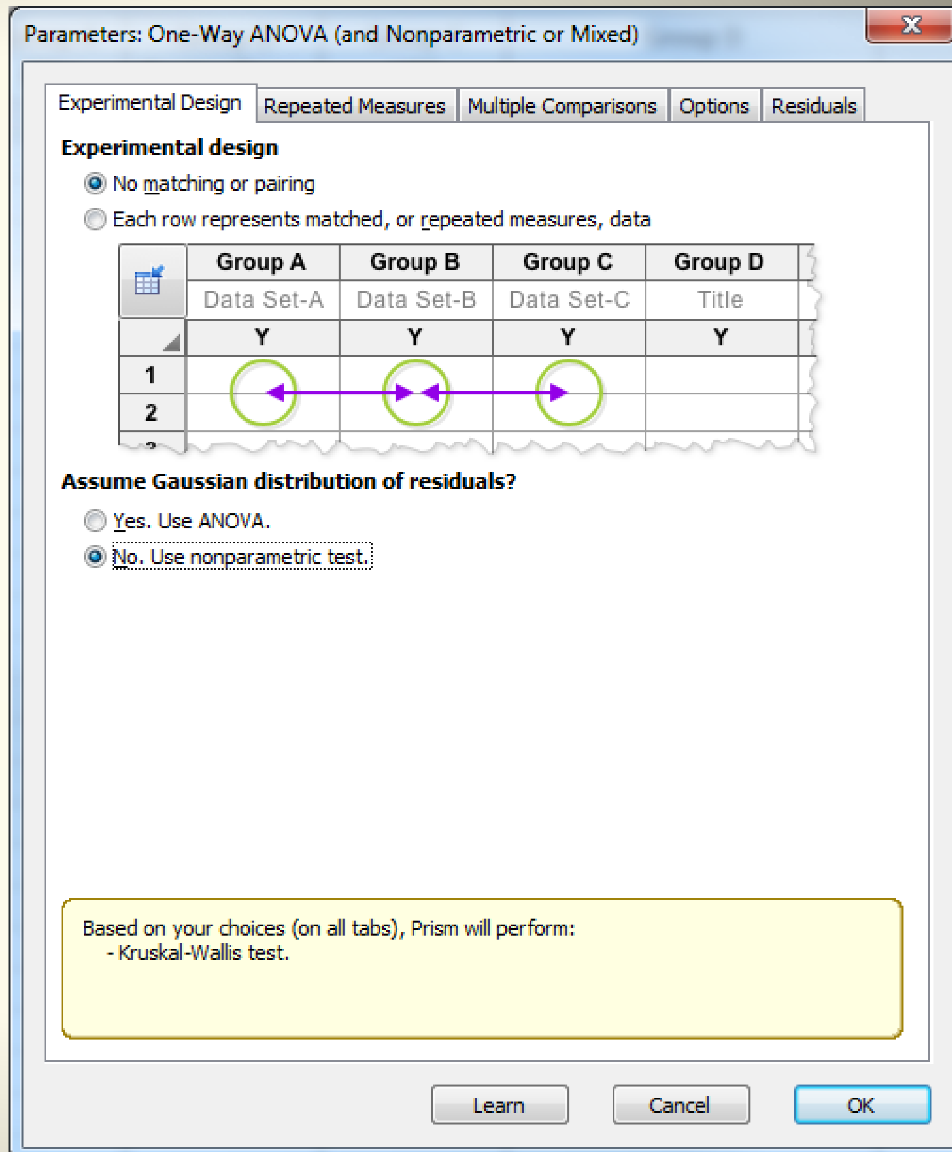


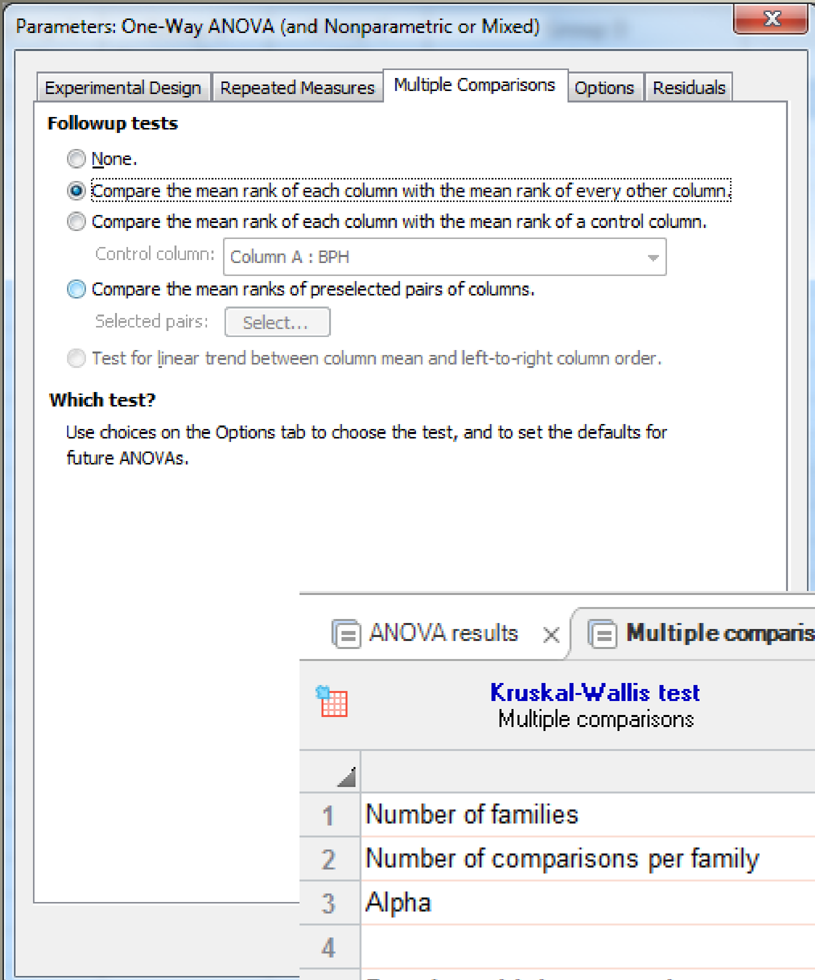
	Col. stats	A BPH	B Positive Biopsy	C NED
1	Number of values	6	8	8
2				
3	Minimum	4.300	3.400	0.5000
4	25% Percentile	5.050	6.525	1.450
5	Median	6.500	7.350	2.750
6	75% Percentile	8.100	14.45	4.925
7	Maximum	8.700	17.30	11.40
8				
9	Mean	6.533	9.588	3.738
10	Std. Deviation	1.618	4.900	3.462
11	Std. Error of Mean	0.6606	1.732	1.224
12				
13	Skewness	-0.02078	0.5498	1.805
14	Kurtosis	-0.8685	-1.226	3.734

Column Table

	Group A	Group B	Group C
	BPH	Positive Biopsy	NED
1	5.3	7.1	11.4
2	7.9	6.6	0.5
3	8.7	6.5	1.6
4	4.3	14.8	2.3
5	6.6	17.3	3.2
6	6.4	3.4	1.4
7		13.4	4.4
8		7.6	5.1



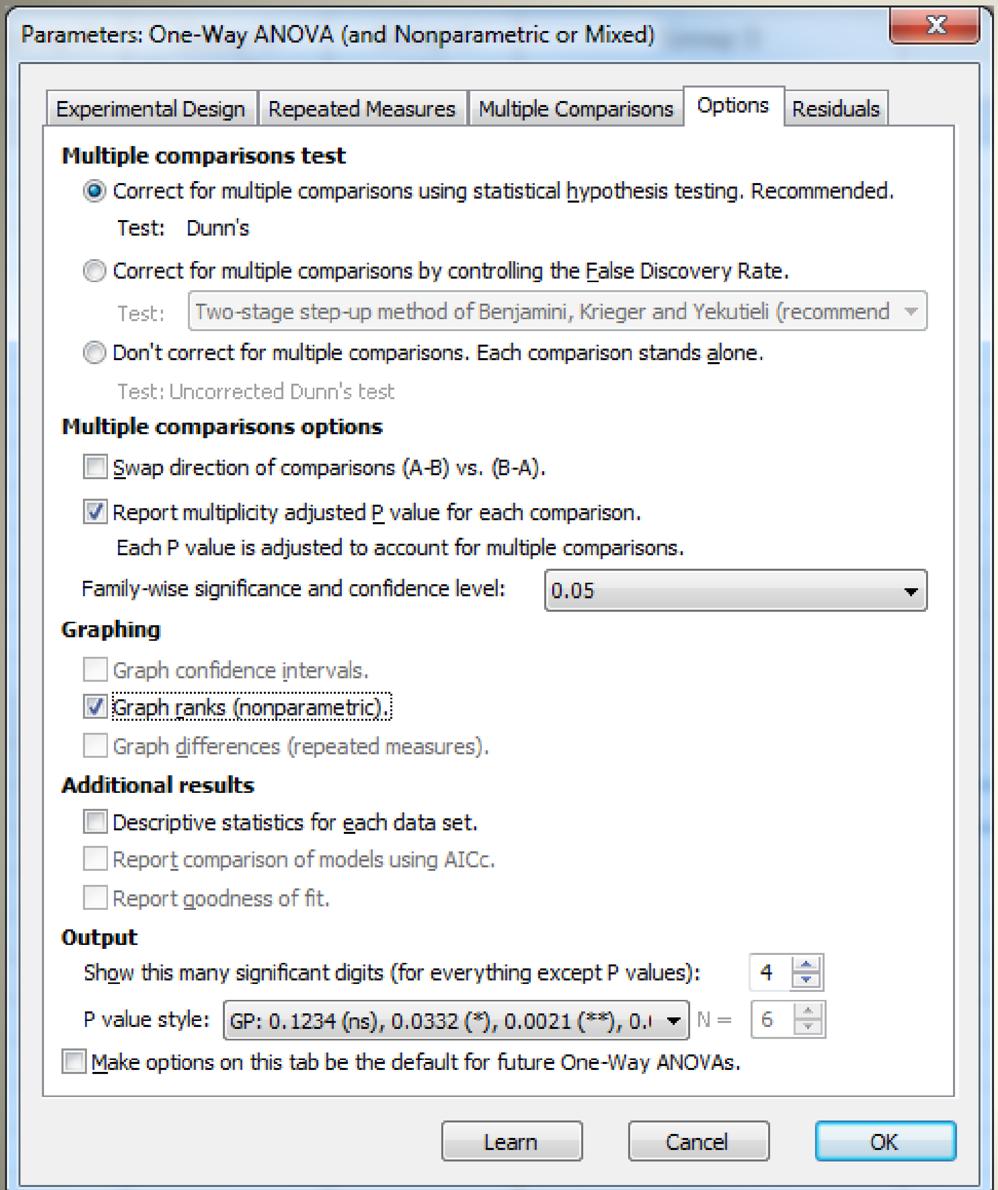


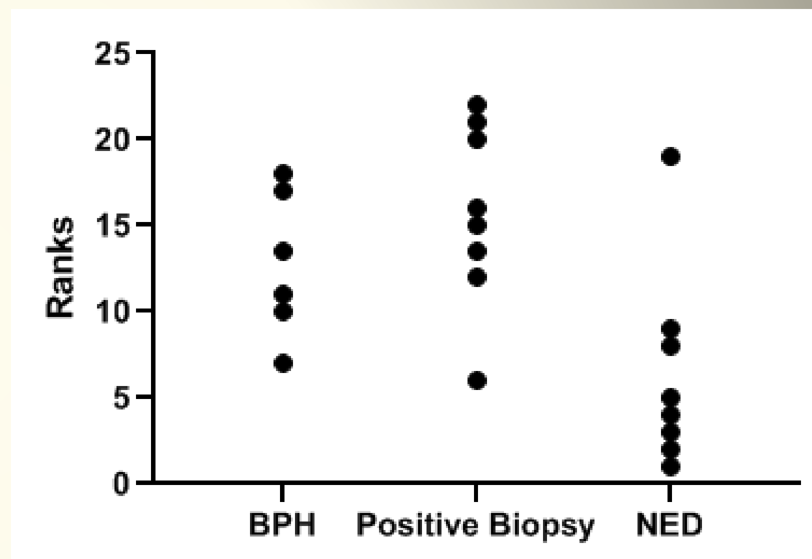


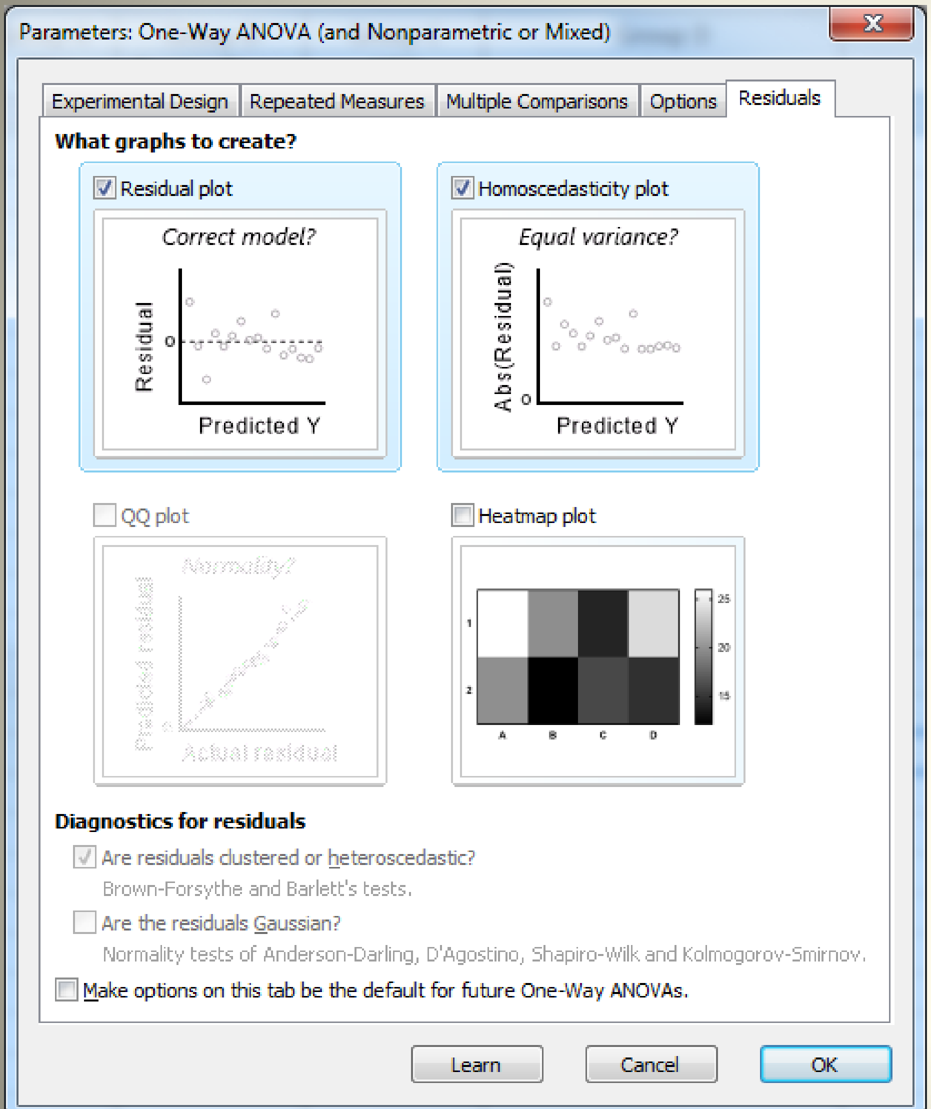
ANOVA results | Multiple comparisons

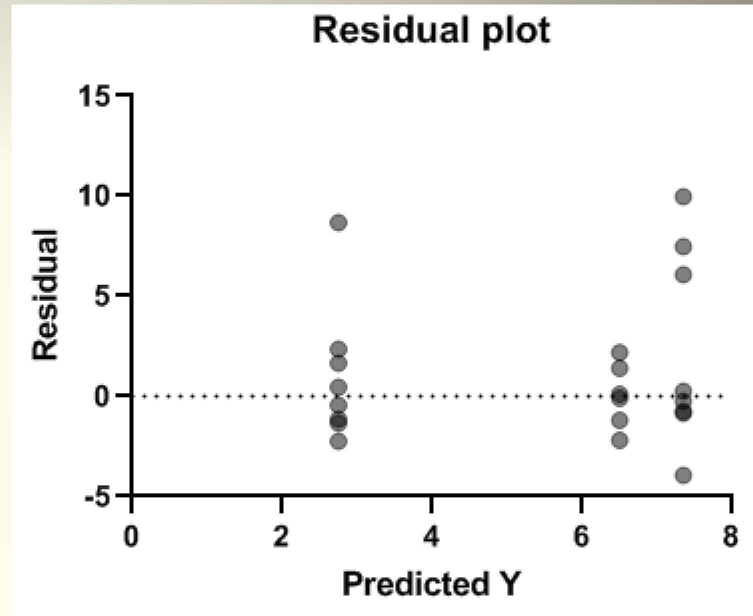
Kruskal-Wallis test
Multiple comparisons

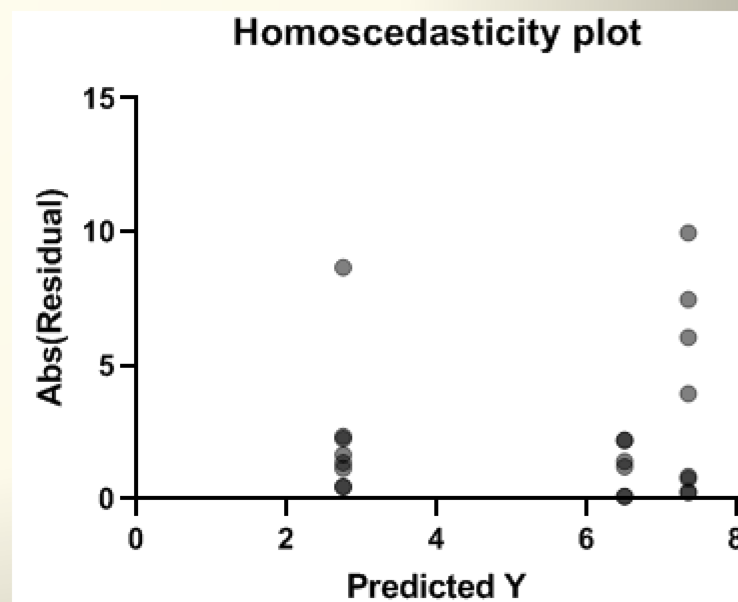
1	Number of families	1					
2	Number of comparisons per family	3					
3	Alpha	0.05					
4							
5	**Dunn's multiple comparisons test**	**Mean rank diff.**	**Significant?**	**Summary**	**Adjusted P Value**		
6	BPH vs. Positive Biopsy	-2.938	No	ns	>0.9999	A-B	
7	BPH vs. NED	6.375	No	ns	0.2070	A-C	
8	Positive Biopsy vs. NED	9.313	Yes	*	0.0124	B-C	
9							
10	**Test details**	**Mean rank 1**	**Mean rank 2**	**Mean rank diff.**	**n1**	**n2**	**Z**
11	BPH vs. Positive Biopsy	12.75	15.69	-2.938	6	8	0.8379
12	BPH vs. NED	12.75	6.375	6.375	6	8	1.818
13	Positive Biopsy vs. NED	15.69	6.375	9.313	8	8	2.869


Parameters: One-Way ANOVA (and Nonparametric or Mixed)
Experimental Design
Repeated Measures
Multiple Comparisons
Options
Residuals
Multiple comparisons test
Correct for multiple comparisons using statistical hypothesis testing. Recommended.
Test: Dunn's
Correct for multiple comparisons by controlling the False Discovery Rate.
Test: Two-stage step-up method of Benjamini, Krieger and Yekutieli (recommend
Don't correct for multiple comparisons. Each comparison stands alone.
Test: Uncorrected Dunn's test
Multiple comparisons options
Swap direction of comparisons (A-B) vs. (B-A).
Report multiplicity adjusted P value for each comparison.
Each P value is adjusted to account for multiple comparisons.
Family-wise significance and confidence level:
0.05
Graphing
Graph confidence intervals.
Graph ranks (nonparametric).
Graph differences (repeated measures).
Additional results
Descriptive statistics for each data set.
Report comparison of models using AICc.
Report goodness of fit.
Output
Show this many significant digits (for everything except P values):
4
P value style:
GP: 0.1234 (ns), 0.0332 (*), 0.0021 (**), 0.(
N =
6
Make options on this tab be the default for future One-Way ANOVAs.
Learn
Cancel
OK


Ranks
25
20
15
10
5
0
BPH
Positive Biopsy
NED


Parameters: One-Way ANOVA (and Nonparametric or Mixed)
Experimental Design
Repeated Measures
Multiple Comparisons
Options
Residuals
What graphs to create?
Residual plot
Correct model?
Residual
0
Predicted Y
Homoscedasticity plot
Equal variance?
Abs(Residual)
0
Predicted Y
QQ plot
Heatmap plot
Diagnostics for residuals
Are residuals clustered or heteroscedastic?
Brown-Forsythe and Barlett's tests.
Are the residuals Gaussian?
Normality tests of Anderson-Darling, D'Agostino, Shapiro-Wilk and Kolmogorov-Smirnov.
Make options on this tab be the default for future One-Way ANOVAs.
Learn
Cancel
OK

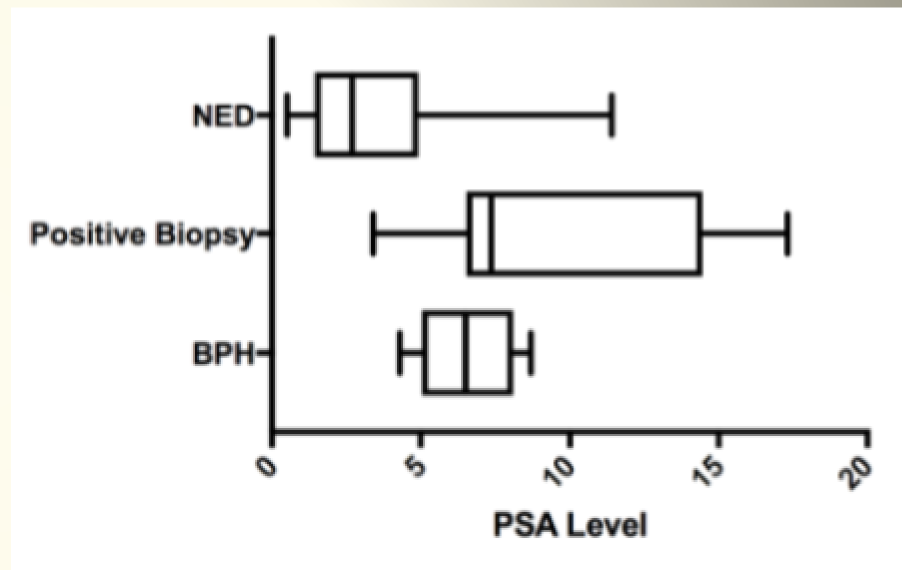

Residual plot
Residual
15
10
5
0
-5
0
2
4
6
8
Predicted Y


Homoscedasticity plot
Abs(Residual)
15
10
5
0
0
2
4
6
8
Predicted Y

ANOVA results | Multiple comparisons

Kruskal-Wallis test
ANOVA results

1	Table Analyzed	PSA Values
2		
3	**Kruskal-Wallis test**	
4	P value	0.0089
5	Exact or approximate P value?	Exact
6	P value summary	**
7	Do the medians vary signif. (P < 0.05)?	Yes
8	Number of groups	3
9	Kruskal-Wallis statistic	8.537
10		
11	**Data summary**	
12	Number of treatments (columns)	3
13	Number of values (total)	22



5	Dunn's multiple comparisons test	Mean rank diff.	Significant?	Summary	Adjusted P Value	
6						
7	BPH vs. Positive Biopsy	-2.938	No	ns	>0.9999	A-B
8	BPH vs. NED	6.375	No	ns	0.2070	A-C
9	Positive Biopsy vs. NED	9.313	Yes	*	0.0124	B-C

ANOVA results | Multiple comparisons

Kruskal-Wallis test
ANOVA results

1	Table Analyzed	PSA Values
2		
3	**Kruskal-Wallis test**	
4	P value	0.0089
5	Exact or approximate P value?	Exact
6	P value summary	**
7	Do the medians vary signif. (P < 0.05)?	Yes
8	Number of groups	3
9	Kruskal-Wallis statistic	8.537
10		
11	**Data summary**	
12	Number of treatments (columns)	3
13	Number of values (total)	22

ANOVA results | Multiple comparisons

Ordinary one-way ANOVA
ANOVA results

1	Table Analyzed	PSA Values
2	Data sets analyzed	A-C
3		
4	**ANOVA summary**	
5	F	4.909
6	P value	0.0191
7	P value summary	*
8	Significant diff. among means (P < 0.05)?	Yes
9	R square	0.3407
10		

Write up results

For this class:
The data did not pass the assumptions for an one-way ANOVA (independence, normality/symmetry, homoscedasticity) as the data were unbalanced, group distributions were skewed, and there was evidence of heteroscedasticity. Using the Kruskal-Wallis test, with p=0.001 (two-tailed hypothesis, α=0.05), we reject the null hypothesis of no difference in PSA medians by group and conclude there is at least one group different from one other. Dunn’s post-hoc tests showed that Positive Biopsy (median PSA=7.35) was significantly different from NED (2.75). BPH PSA (6.5) was not different from either of the other groups.

Two-Way ANOVA

AKA Two Factor ANOVA, Factorial ANOVA

A two-way ANOVA has 2 categorical (independent) factors

Gender and diet type are both factors (a total of 4 groups)

Is there was a gender effect on weight lost as well as type of diet effect and do they combine to increase (or decrease) weight loss?

Similar results as separate one-way ANOVAs (or in this case t-tests) on each factor, but **interaction** between factors can be tested with a two way ANOVA

The categorical variables Diet and Gender are factors (independent) variables

Between groups factor

Between groups factor

Continuous (Dependent) variable

	WeightLOST	Diet	gender
16	1.1	1	Male
17	1.5	1	Female
18	1.7	2	Female
19	1.7	2	Male
20	1.9	1	Female
21	2.0	2	Female
22	2.0	2	Female
23	2.0	1	Female
24	2.0	1	Female

Two-way ANOVA

Continuous (Dependent) variable: Weight Loss

Categorical (Independent) variables (factors): Diet type and Gender

A two-way ANOVA tests three hypotheses, the null's are:

1. H_0: Mean weight loss in the population does not differ by diet type

    Test of *main* effects for one categorical variable (factor)

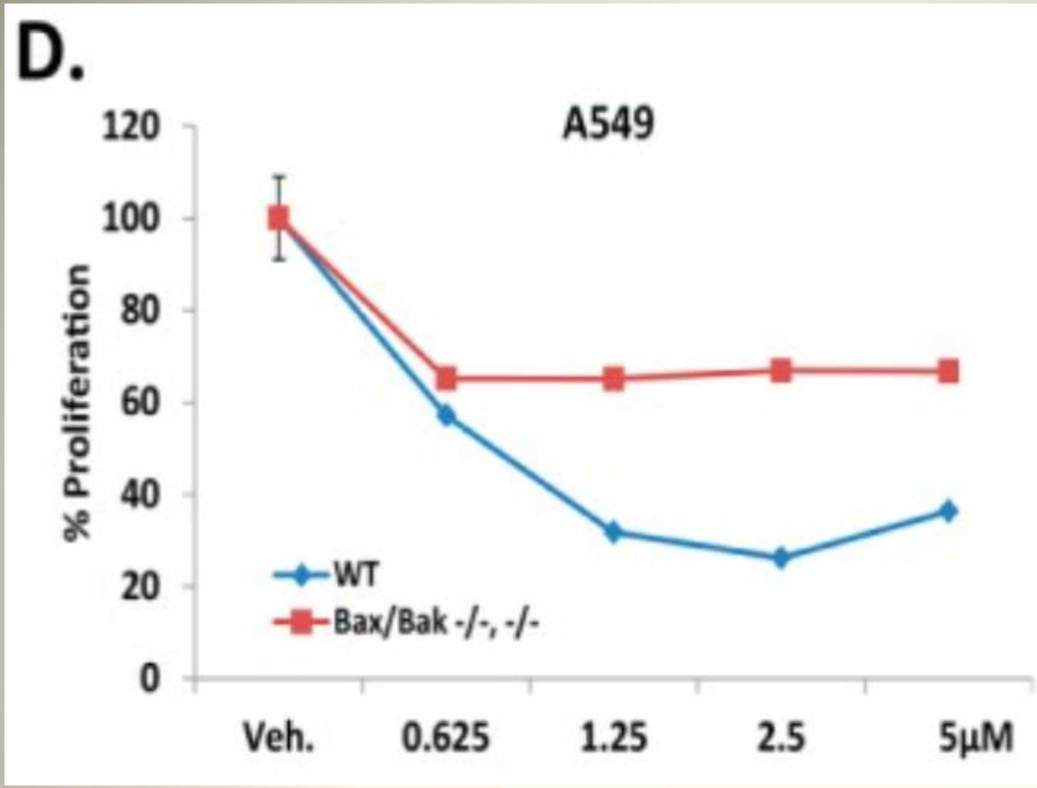
2. H_0: Mean weight loss in the population does not differ by gender

    Test of *main* effects for the other categorical variable (factor)

3. H_0: There is no interaction between diet type and gender

    Test of interaction between diet and gender

Example of a Two-Way ANOVA plot



Continuous variable:
% Proliferation

Factor1: Borrelidin dose
(5 groups)

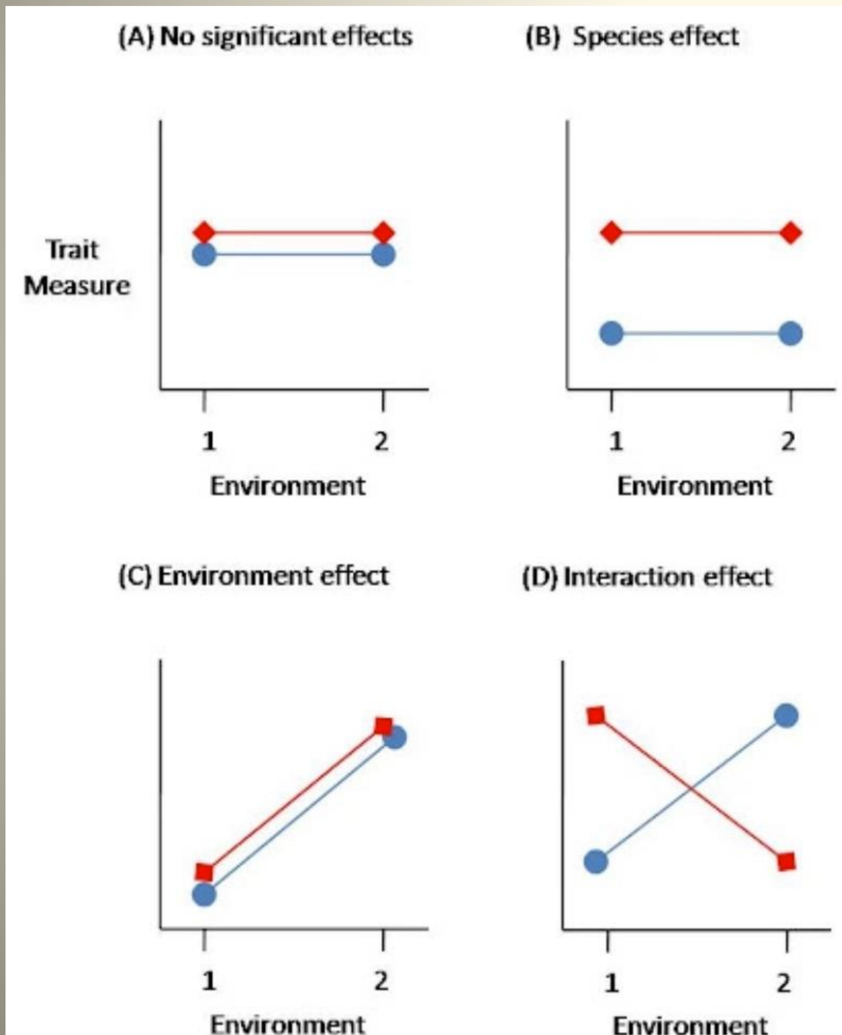
Factor2: cell type
(2 groups)

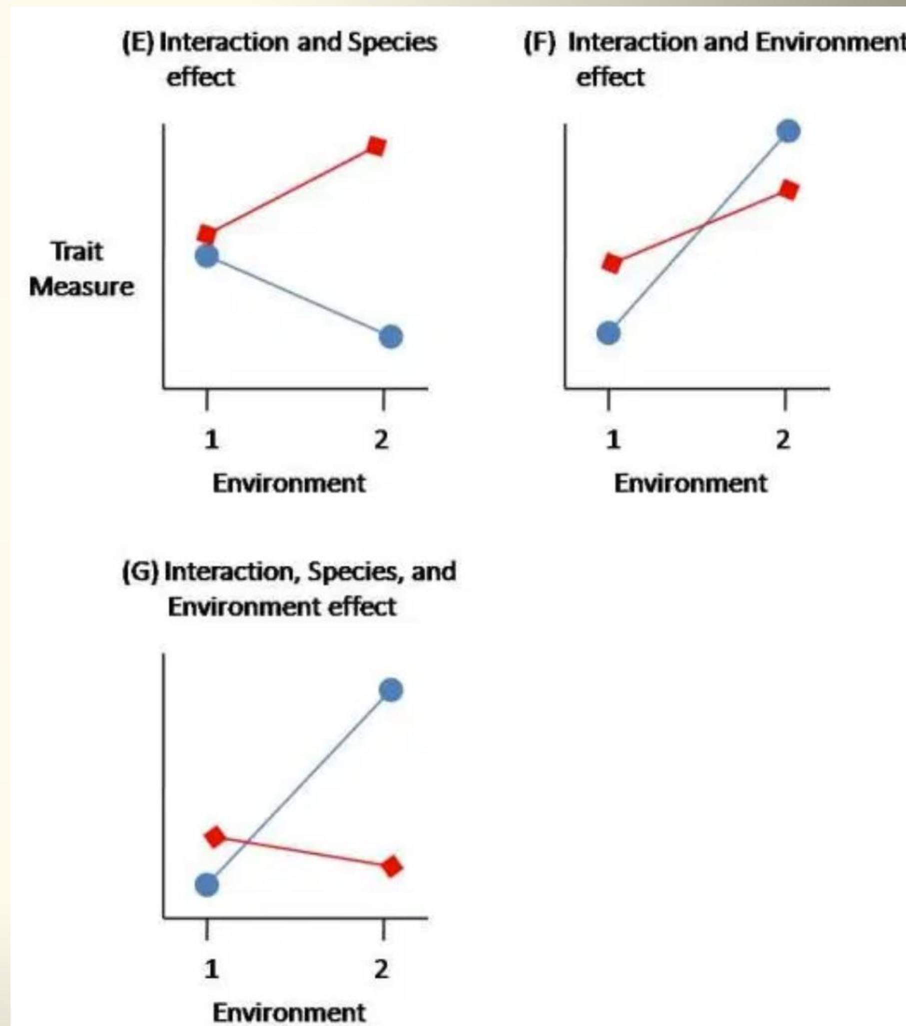
(D) Proliferation assay with BAX$^{(-/-/-/-/)}$ BAK$^{(-/-)}$ A549 cells at 16 h
two-way ANOVA, dose $p < 0.0001$, between wildtype and knockout $p < 0.0001$, a and interaction $p = 0.0016$

Borrelidin Induces the Unfolded Protein Response in Oral Cancer Cells and Chop-Dependent Apoptosis
ACS Med Chem Lett. 2015 Nov 12; 6(11): 1122–1127.

Interpreting Effects from a Two-way ANOVA

measuring a biological trait from two species raised in two environments





Assumptions

Means	Male	Female
Diet1	10.2	9.8
Diet2	12.5	8.7

All cells (groups) of the study design should be normal

- Less important for large sample sizes

The samples must be independent

- Observations or participants must be independent
- Participants in one group cannot influence participants in another

Homoscedasticity in every cell.

The cells should have the same sample size

The Limitations of Prism

Prism does not test for violations of assumption of normality for a two-way ANOVA

Prism advises that if you don't think your data are sampled from a normal distribution (and no transformation will make the distribution normal), you should consider performing nonparametric two-way ANOVA.

But Prism does not offer this test

Prism also cannot test the assumption of equal variances

But you can eyeball the data as we have done before

Two-Way ANOVA Example

	ABILITY	METHOD	X
1	none	blue-book	23
2	none	blue-book	32
3	none	blue-book	25
4	some	blue-book	29
5	some	blue-book	30
6	some	blue-book	34
7	lots	blue-book	31
8	lots	blue-book	36
9	lots	blue-book	33
10	none	computer	32
11	none	computer	26
12	none	computer	26
13	some	computer	34
14	some	computer	41
15	some	computer	35
16	lots	computer	23
17	lots	computer	26
18	lots	computer	32
	18	18	18

A statistics teacher is concerned that allowing some students to use a computer for the final exam might give them an unfair advantage if their data entry (aka typing) skill are good (a person can usually type faster than they can write). She is interested in the effect the of type of medium (bluebook or computer) and typing ability on test results.

Her main question is about how typing affects test scores for people using computers or bluebooks

For a practice exam, she randomly divided the classes in half: 50% wrote in a bluebook, 50% used computers. In addition the students were partitioned into three groups, no typing ability, some typing ability, and lots of typing ability.

The continuous (dependent) variable is the final exam score (X).

Modified from http://www.psychstat.missouristate.edu/multibook/mlt09.htm

Assumptions for this example

Assumption of independence seems to be OK

However, Prism can't test the assumptions of normality and homoscedasticity by cells (it assumes these to be met)

- Cell # = 3 (levels of typing ability) x 2 (exam methods) = 6

We could restructure the data and test the 6 cells in a column (ANOVA) analysis

- But each cell only has an n of 3
  - Too small for statistical tests and what would graphing tell us?

What do we do?

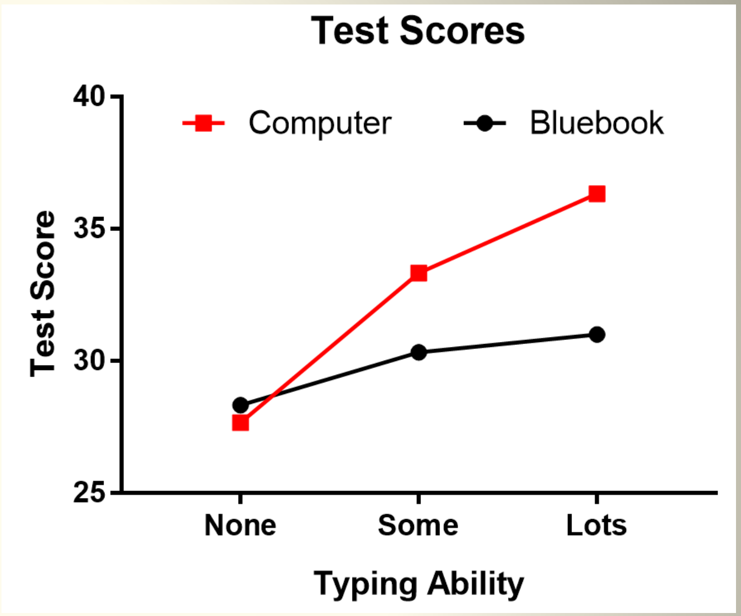
- Run the example as it is for now.

- If n's are small, seek professional statistical help before you publish

A first look at the data

If lines are not parallel, there could be interaction

If scores for medium factor (computer vs. Bluebook) have different levels for different categories of the other factor (typing ability), there could be interaction



	Computer			Bluebook		
	n	mean	SD	n	mean	SD
None	3	27.7	2.08	3	28.3	1.53
Some	3	33.3	1.15	3	30.3	1.53
Lots	3	36.3	1.53	3	31.0	2.00

New Data Table and Graph

New table & graph

- XY
- Column
- Grouped
- Contingency
- Survival
- Parts of whole
- Multiple variables
- Nested

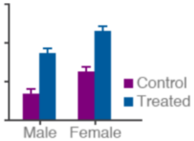
Existing file

- Clone a graph

Prism Tips

Grouped tables have two grouping variables, one defined by columns and the other defined by rows

Table format Grouped	A			B		
	Control			Treated		
	A:Y1	A:Y2	A:Y3	B:Y1	B:Y2	B:Y3
1 Male						
2 Femal						



Learn more

Data table:

- (•) Enter or import data into a new table
- ( ) Start with sample data to follow a tutorial

Options:

- ( ) Enter and plot a single Y value for each point
- (•) Enter 3 replicate values in side-by-side subcolumns
- ( ) Enter and plot error values already calculated elsewhere

Enter: Mean, SD, N

Table format: Grouped	Group A			Group B		
	Bluebook			Computer		
	A:1	A:2	A:3	B:1	B:2	B:3
1 None	30	28	27	30	26	27
2 Some	29	30	32	34	32	34
3 Lots	31	33	29	38	36	35

Graphing the data differently by different data structures

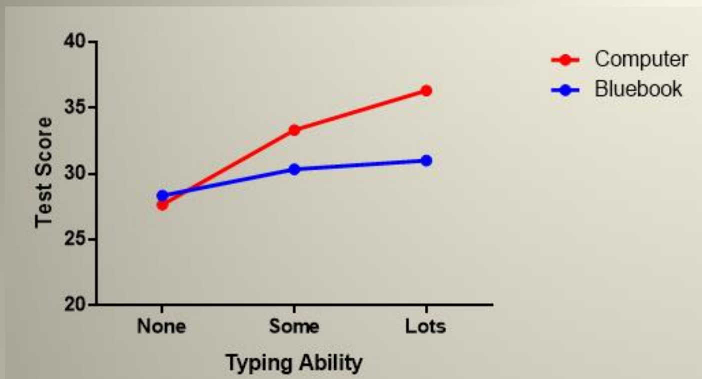


Table format: Grouped		Group A			Group B		
		Bluebook			Computer		
		A:Y1	A:Y2	A:Y3	B:Y1	B:Y2	B:Y3
1	None	30	28	27	30	26	27
2	Some	29	30	32	34	32	34
3	Lots	31	33	29	38	36	35
4	Title						

This set-up better answers her main question

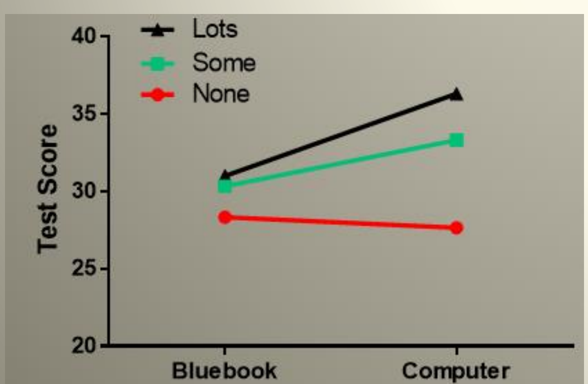
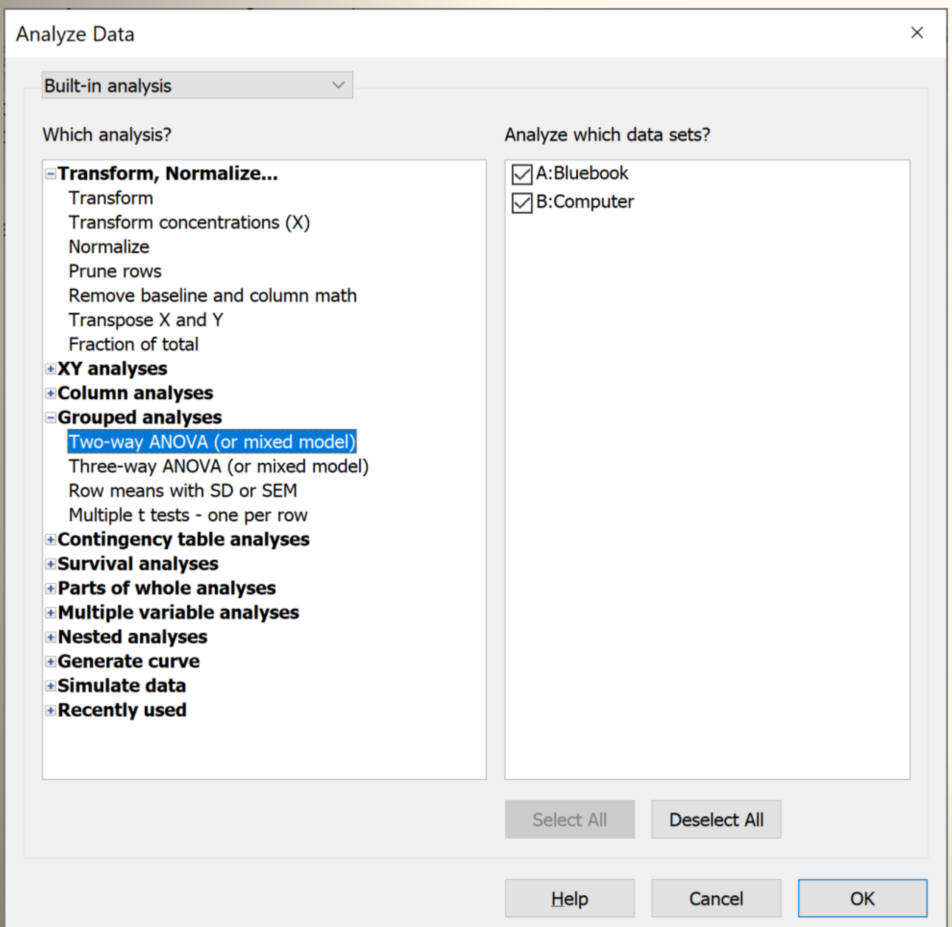
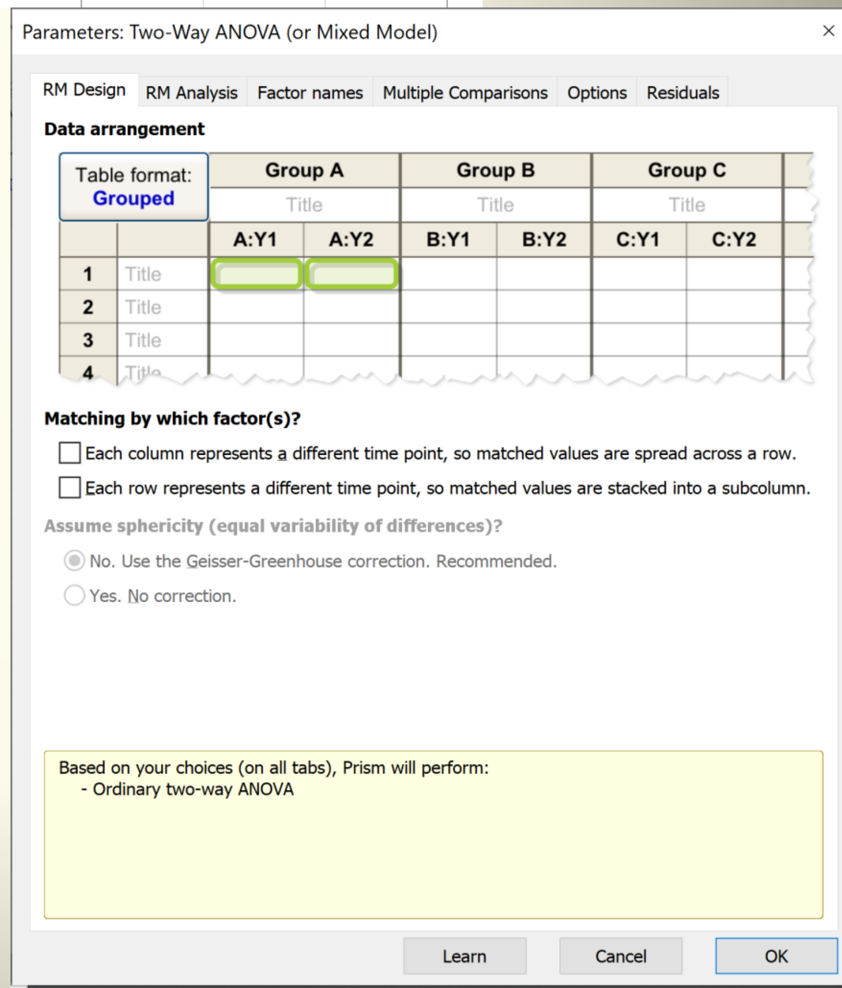
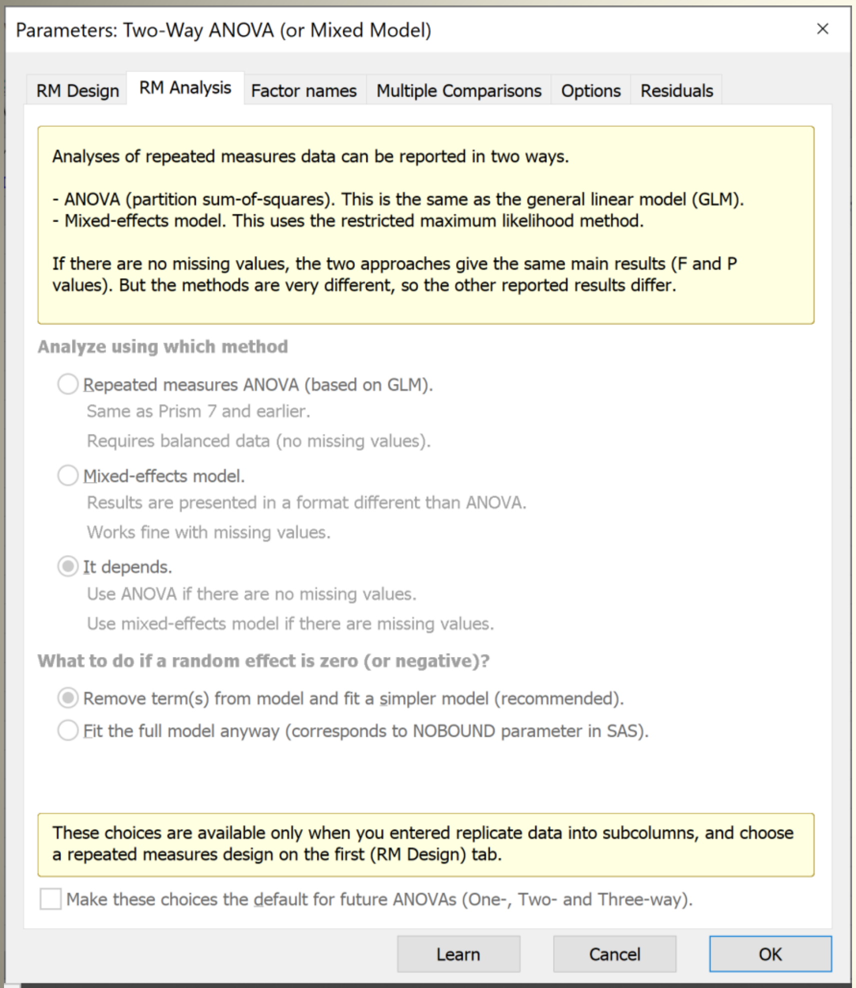


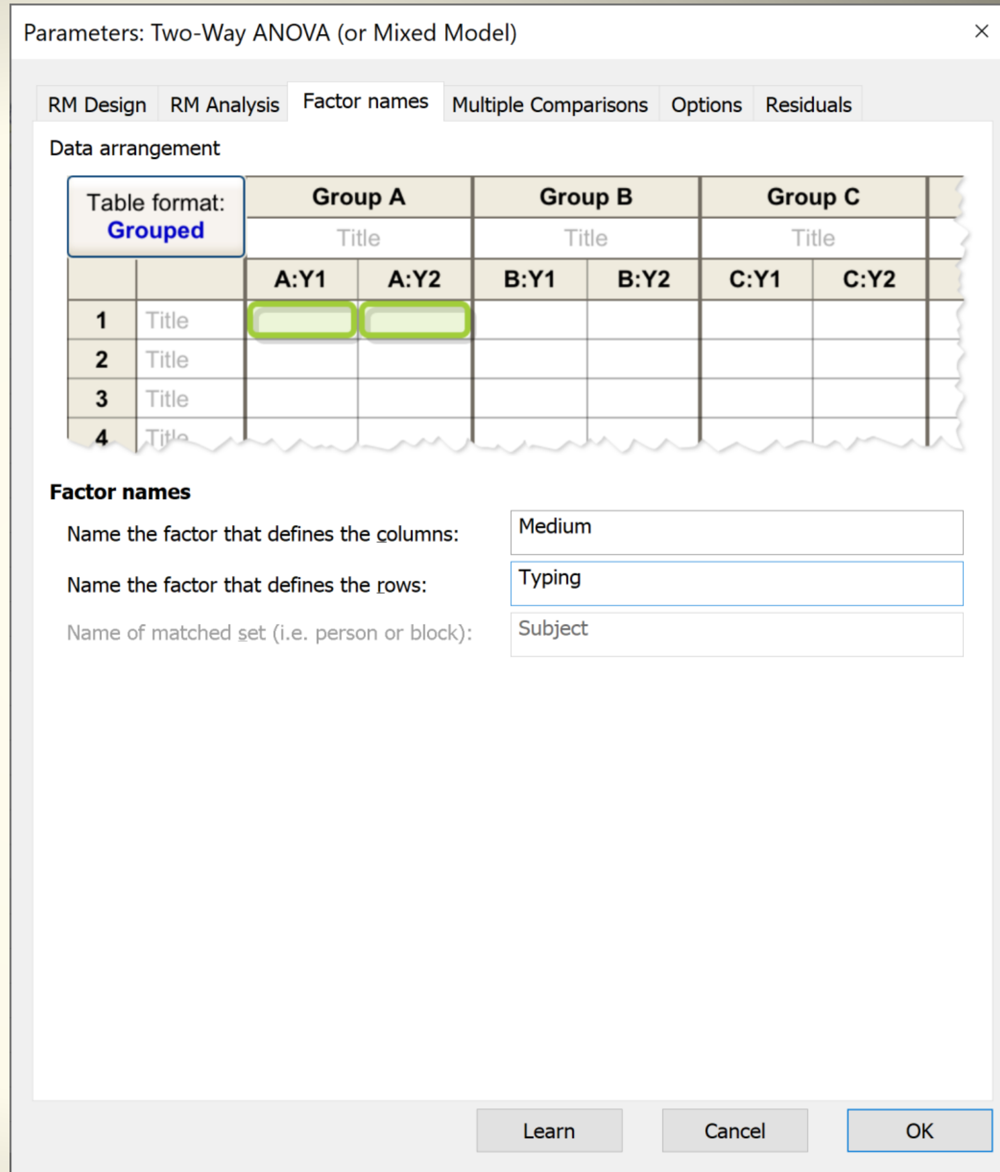
Table format: Grouped		Group A			Group B			Group C		
		None			Some			Lots		
		A:Y1	A:Y2	A:Y3	B:Y1	B:Y2	B:Y3	C:Y1	C:Y2	C:Y3
1	Bluebook	30	28	27	29	30	32	31	33	29
2	Computer	30	26	27	34	32	34	38	36	35
3	Title									

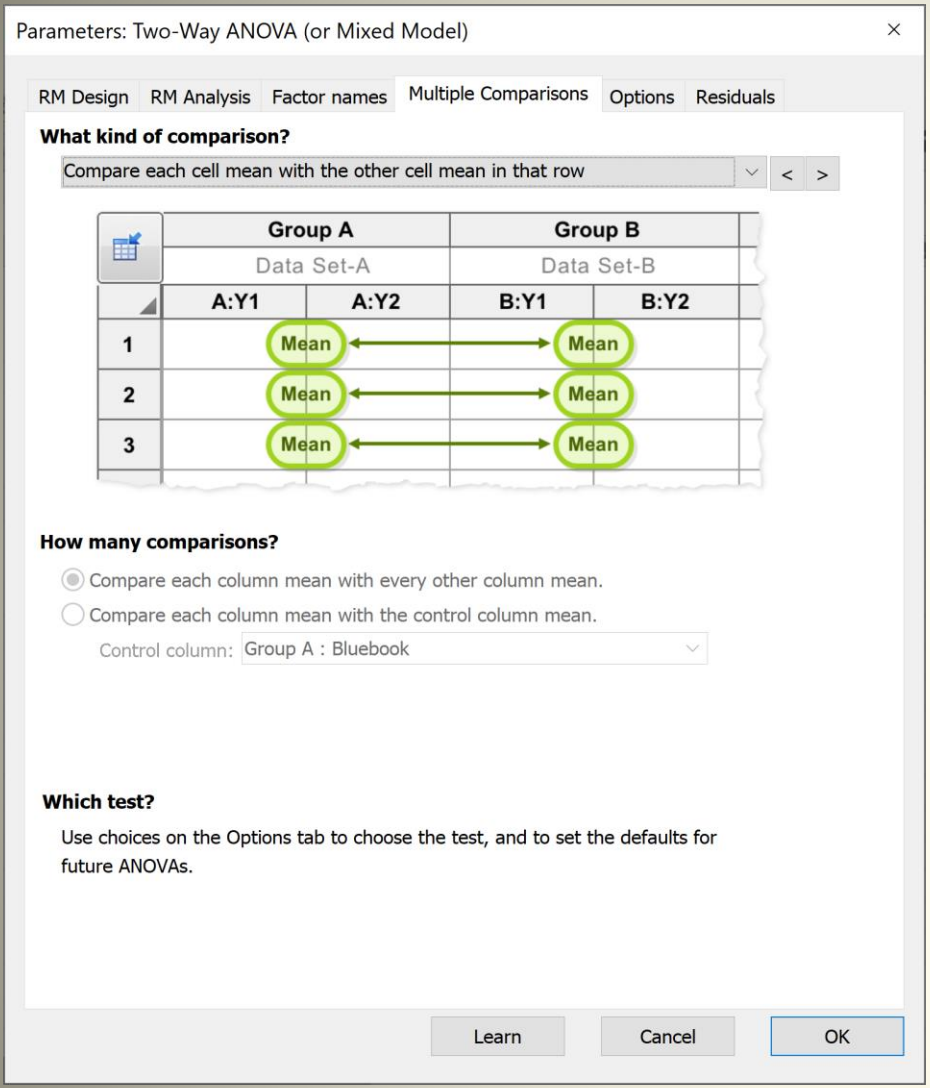
Table format: Grouped		Group A Bluebook			Group B Computer		
		A:1	A:2	A:3	B:1	B:2	B:3
1	None	30	28	27	30	26	27
2	Some	29	30	32	34	32	34
3	Lots	31	33	29	38	36	35

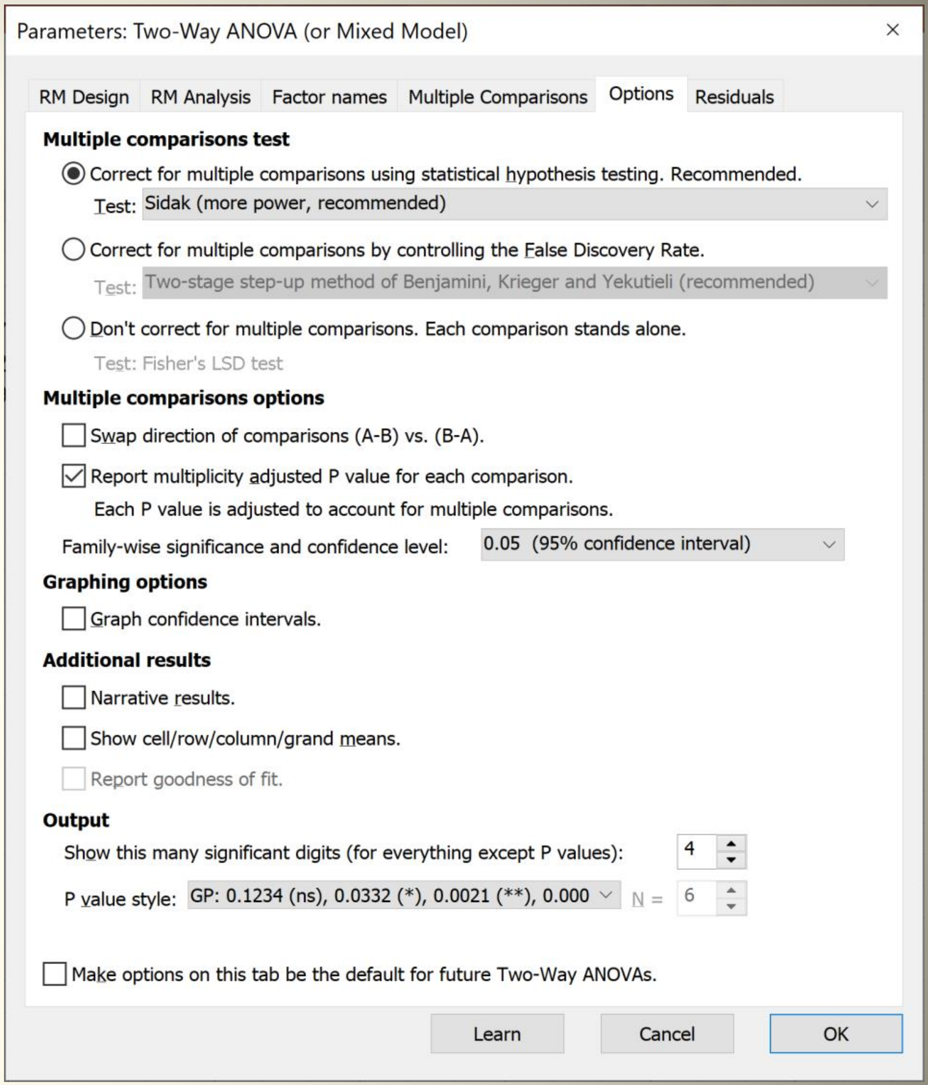


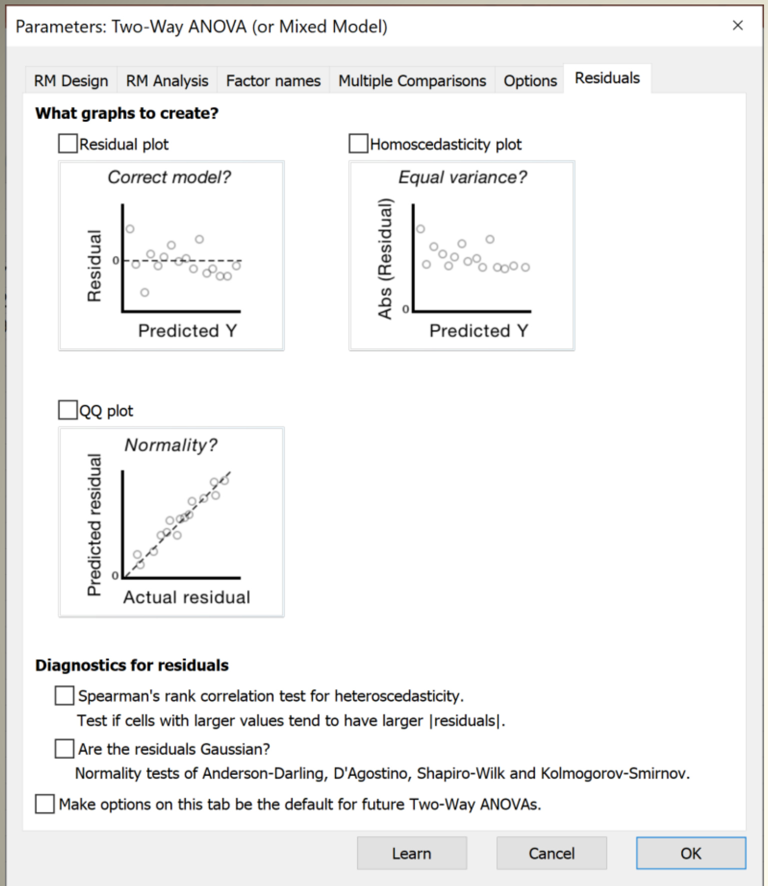


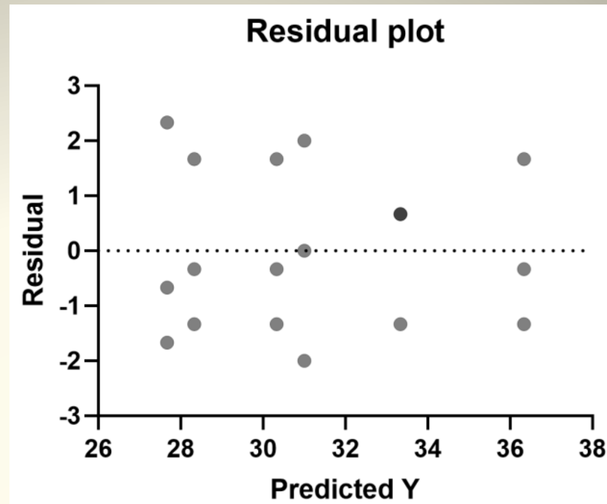

Parameters: Two-Way ANOVA (or Mixed Model)
RM Design
RM Analysis
Factor names
Multiple Comparisons
Options
Residuals
Analyses of repeated measures data can be reported in two ways.
- ANOVA (partition sum-of-squares). This is the same as the general linear model (GLM).
- Mixed-effects model. This uses the restricted maximum likelihood method.
If there are no missing values, the two approaches give the same main results (F and P values). But the methods are very different, so the other reported results differ.
Analyze using which method
Repeated measures ANOVA (based on GLM).
Same as Prism 7 and earlier.
Requires balanced data (no missing values).
Mixed-effects model.
Results are presented in a format different than ANOVA.
Works fine with missing values.
It depends.
Use ANOVA if there are no missing values.
Use mixed-effects model if there are missing values.
What to do if a random effect is zero (or negative)?
Remove term(s) from model and fit a simpler model (recommended).
Fit the full model anyway (corresponds to NOBOUND parameter in SAS).
These choices are available only when you entered replicate data into subcolumns, and choose a repeated measures design on the first (RM Design) tab.
Make these choices the default for future ANOVAs (One-, Two- and Three-way).
Learn
Cancel
OK

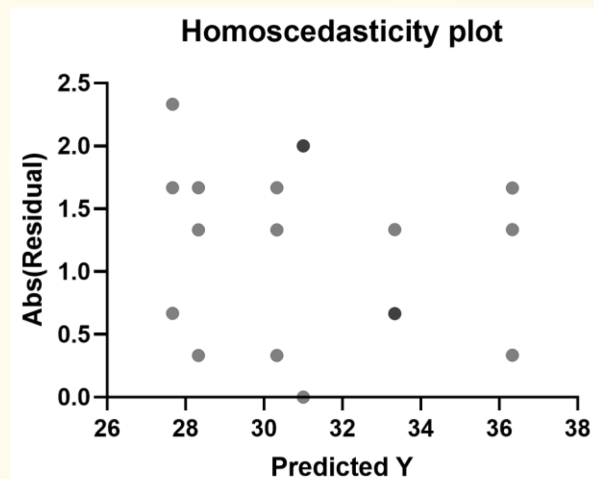

Parameters: Two-Way ANOVA (or Mixed Model)
RM Design
RM Analysis
Factor names
Multiple Comparisons
Options
Residuals
Data arrangement
Table format:
Grouped
Group A
Group B
Group C
Title
A:Y1
A:Y2
B:Y1
B:Y2
C:Y1
C:Y2
1 Title
2 Title
3 Title
Factor names
Name the factor that defines the columns:
Medium
Name the factor that defines the rows:
Typing
Name of matched set (i.e. person or block):
Subject
Learn
Cancel
OK

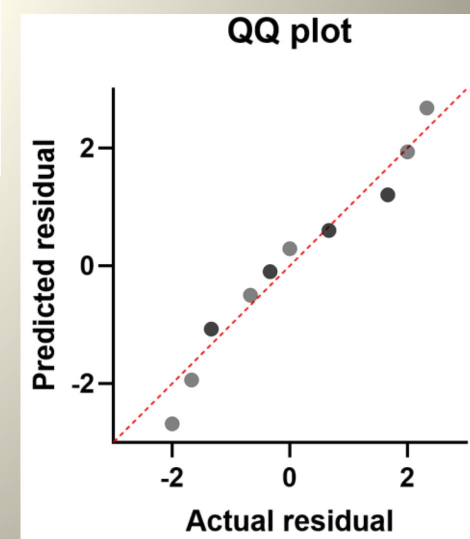

Parameters: Two-Way ANOVA (or Mixed Model)
RM Design
RM Analysis
Factor names
Multiple Comparisons
Options
Residuals
What kind of comparison?
Compare each cell mean with the other cell mean in that row
Group A
Group B
Data Set-A
Data Set-B
A:Y1
A:Y2
B:Y1
B:Y2
1
2
3
Mean
How many comparisons?
Compare each column mean with every other column mean.
Compare each column mean with the control column mean.
Control column: Group A : Bluebook
Which test?
Use choices on the Options tab to choose the test, and to set the defaults for future ANOVAs.
Learn
Cancel
OK

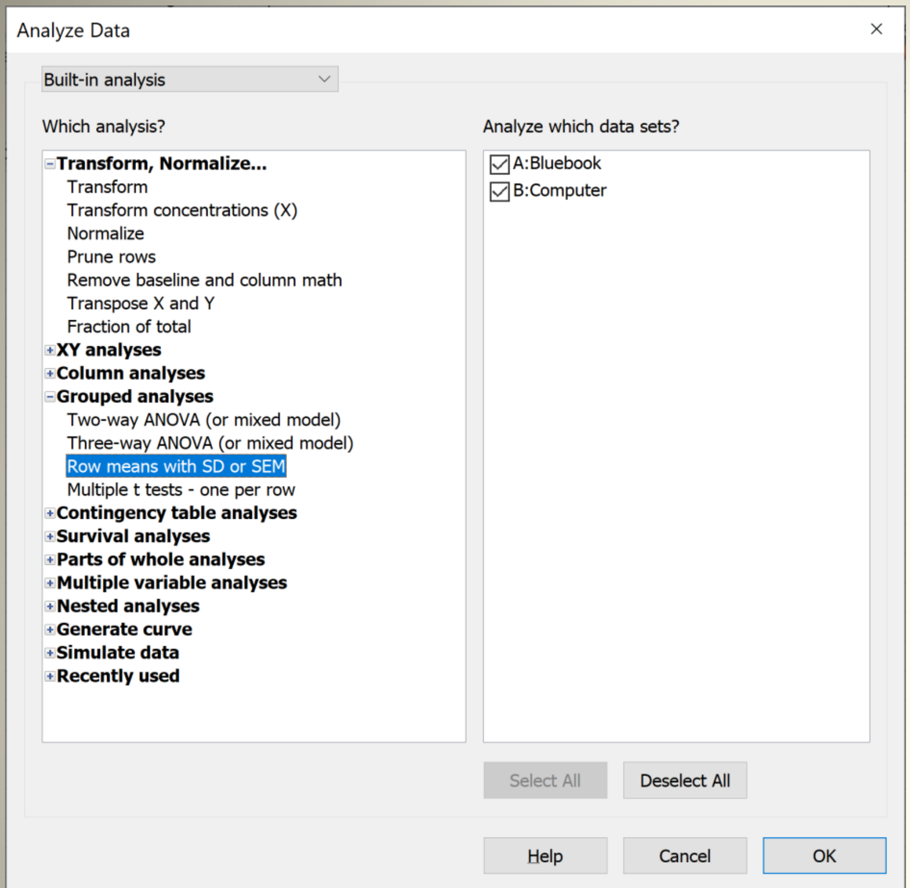

Parameters: Two-Way ANOVA (or Mixed Model)
RM Design
RM Analysis
Factor names
Multiple Comparisons
Options
Residuals
Multiple comparisons test
Correct for multiple comparisons using statistical hypothesis testing. Recommended.
Test: Sidak (more power, recommended)
Correct for multiple comparisons by controlling the False Discovery Rate.
Test: Two-stage step-up method of Benjamini, Krieger and Yekutieli (recommended)
Don't correct for multiple comparisons. Each comparison stands alone.
Test: Fisher's LSD test
Multiple comparisons options
Swap direction of comparisons (A-B) vs. (B-A).
Report multiplicity adjusted P value for each comparison.
Each P value is adjusted to account for multiple comparisons.
Family-wise significance and confidence level: 0.05 (95% confidence interval)
Graphing options
Graph confidence intervals.
Additional results
Narrative results.
Show cell/row/column/grand means.
Report goodness of fit.
Output
Show this many significant digits (for everything except P values): 4
P value style: GP: 0.1234 (ns), 0.0332 (*), 0.0021 (**), 0.000
N = 6
Make options on this tab be the default for future Two-Way ANOVAs.
Learn
Cancel
OK

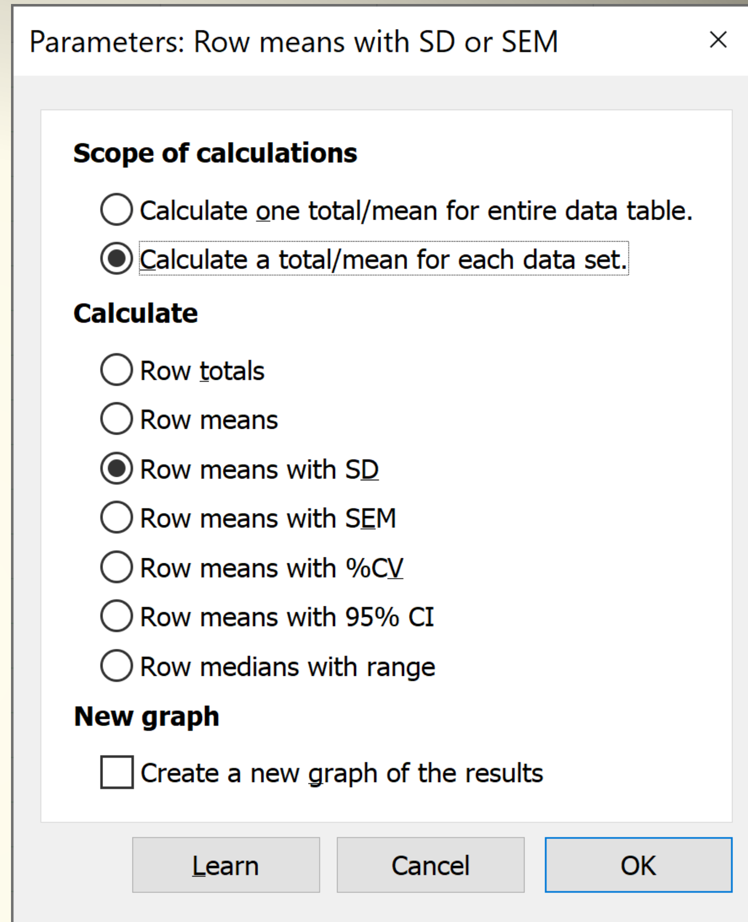

Parameters: Two-Way ANOVA (or Mixed Model)
RM Design
RM Analysis
Factor names
Multiple Comparisons
Options
Residuals
What graphs to create?
Residual plot
Correct model?
Residual
Predicted Y
Homoscedasticity plot
Equal variance?
Abs (Residual)
Predicted Y
QQ plot
Normality?
Predicted residual
Actual residual
Diagnostics for residuals
Spearman's rank correlation test for heteroscedasticity.
Test if cells with larger values tend to have larger |residuals|.
Are the residuals Gaussian?
Normality tests of Anderson-Darling, D'Agostino, Shapiro-Wilk and Kolmogorov-Smirnov.
Make options on this tab be the default for future Two-Way ANOVAs.
Learn
Cancel
OK


Residual plot
Residual
Predicted Y


Homoscedasticity plot
Abs(Residual)
Predicted Y


QQ plot
Predicted residual
Actual residual





Row stats	A			B		
	Bluebook			Computer		
	Mean	SD	N	Mean	SD	N
1 None	28.333	1.528	3	27.667	2.082	3
2 Some	30.333	1.528	3	33.333	1.155	3
3 Lots	31.000	2.000	3	36.333	1.528	3
4						

Two-Way ANOVA: Summary Table
(Don't worry, I won't ask you to describe this one)

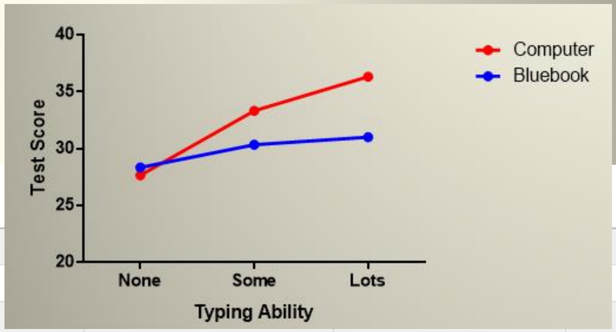
Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F
A (Row)	a - 1	SS(A)	MS(A)	$\frac{MS(A)}{MSE}$
B (Column)	b - 1	SS(B)	MS(B)	$\frac{MS(B)}{MSE}$
AB (Interaction)	(a-1)(b-1)	SS(AB)	MS(AB)	$\frac{MS(AB)}{MSE}$
Error	n - ab	SSE	MSE	
Total	n - 1	SS(Total) ← *Same as Other Designs*		

Remember the 3 null hypotheses about

Interaction

Main effect for factor1

Main effect for factor2



ANOVA results | Multiple comparisons

2way ANOVA
ANOVA results

1	Table Analyzed	Test Scores				
2						
3	**Two-way ANOVA**	Ordinary				
4	Alpha	0.05				
5						
6	**Source of Variation**	**% of total variation**	**P value**	**P value summary**	**Significant?**	
7	Interaction	14.41	0.0272	*	Yes	
8	Typing	52.67	0.0002	***	Yes	
9	Medium	15.43	0.0069	**	Yes	
10						
11	**ANOVA table**	**SS**	**DF**	**MS**	**F (DFn, DFd)**	**P value**
12	Interaction	27.44	2	13.72	F (2, 12) = 4.940	P=0.0272
13	Typing	100.3	2	50.17	F (2, 12) = 18.06	P=0.0002
14	Medium	29.39	1	29.39	F (1, 12) = 10.58	P=0.0069
15	Residual	33.33	12	2.778		

Interaction

Main effect 1

Main effect 2

ANOVA results | Multiple comparisons

2way ANOVA
Multiple comparisons

1	Compare each cell mean with the other cell mean in that row						
2							
3	Number of families						
4	Number of comparisons per family						
5	Alpha						
6							
7	**Sidak's multiple comparisons test**	**Significant?**	**Summary**	**Adjusted P Value**			
8							
9	Bluebook - Computer						
10	None	No	ns	0.9506			
11	Some	No	ns	0.1365			
12	Lots	Yes	**	0.0061			
13							
14							
15	**Test details**	**Mean Diff.**	**SE of diff.**	**N1**	**N2**	**t**	**DF**
16							
17	Bluebook - Computer						
18	None	0.6667	1.361	3	3	0.4899	12.00
19	Some	-3.000	1.361	3	3	2.205	12.00
20	Lots	-5.333	1.361	3	3	3.919	12.00

Write up results

For this class:

The data passed the assumptions for an two-way ANOVA (independence, homoscedasticity for each cell by both factors). With only and n of three in each cell, it was difficult to assess normality. Given that the ANOVA is robust to some deviation from normality, we will proceed with the two-way anov. Both main effects were significant so we reject the null hypotheses of no difference in score by medium used (computer vs. bluebook) or typing ability in the population. People using the computer scored higher [$F(1.12)=10.58$, $p=0.01$, $\alpha=0.05$] as did those with higher typing ability [$F(2.12)=18.06$, $p=0.002$; see table 1 for mean values]. The interaction term had a p-value of 0.03 [$F(2,12)=4.94$] meaning that we reject the null hypothesis of no interaction and conclude that that people who used the computer and had better typing skills had the highest test scores (see Figure 1 and Table 1).

Table 1.

	Computer			Bluebook		
	n	mean	SD	n	mean	SD
None	3	27.7	2.08	3	28.3	1.53
Some	3	33.3	1.15	3	30.3	1.53
Lots	3	36.3	1.53	3	31.0	2.00

Figure 1.