Repeated Measures (RM) ANOVAs

Kathleen Torkko

October 23, 2019

ANOVA Family

One way ANOVA

One Categorical (Independent) Variable (AKA one factor)

- One continuous (dependent) variable measured in different groups
  - Different participants
- One-way Repeated Measures: One continuous variable measured >2 times across time intervals in same person
  - Same participants

Factorial ANOVA

More than One Categorical Variable

- Two way
- Three way
- Four way

Two-way Repeated Measures: One continuous variable measured by two factors

Objectives

Finish Monday’s lecture on the two-way ANOVA

Understand what is meant by repeated measures

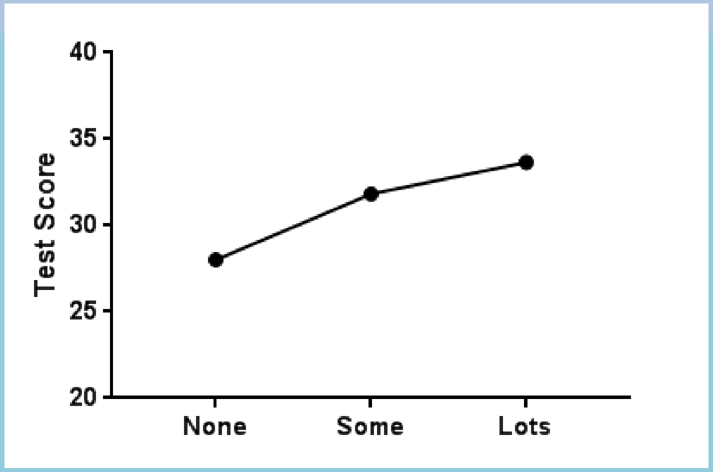
Be able to enter data in required format

Carry out a RM ANOVA in Prism

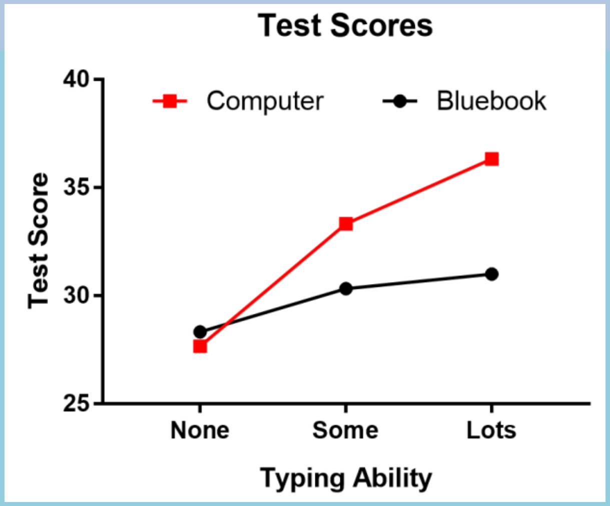
Interpret the output

What is the main question you are trying to answer about your data?

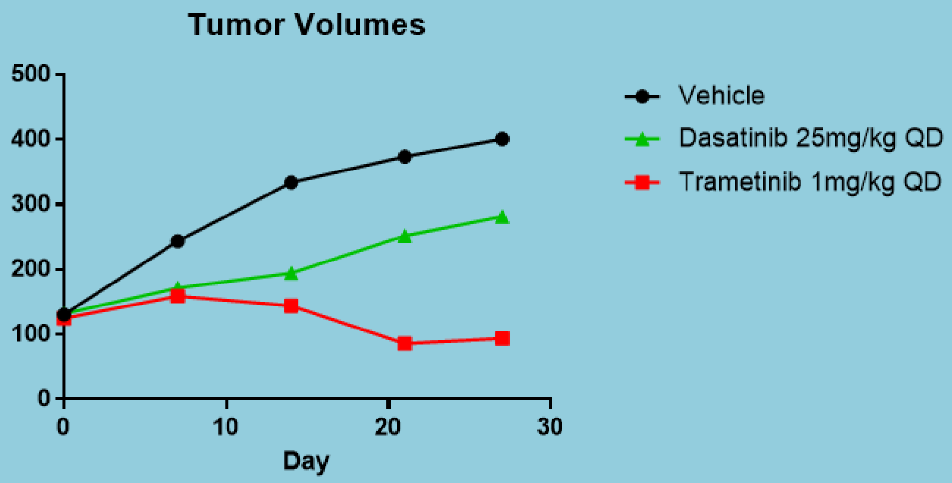
One-Way ANOVA



Two-Way ANOVA



RM ANOVA



Measuring the same variable in the same subject (cell, rat, patient) >2 times

If =2 times, use a paired t-test

What is meant by 'Repeated Measures'?

When there are multiple measurements of the same characteristic from the same individual (or mouse, cell line, etc.)

Subject	Sex	Fed	Fasted 10 days	Fasted 20 days
1	Male	42.8	42.4	38.9
2	Male	43.1	42.2	40.3
3	Male	40.0	40.8	37.5
4	Male	46.6	45.9	42.9
5	Female	42.2	42.4	39.7
6	Female	38.7	38.1	35.8
7	Female	35.3	34.3	32.3
8	Female	40.5	40.1	37.3

Four male and four female turtles had their plasma protein measured when well fed and 10 and 20 days after fasting

Why 'Repeated Measures'?

Increases statistical power (*i.e.*, increases the probability of rejecting a false null hypothesis)

- A subject's repeated measurements will have less variability compared to measurements in multiple subjects
  - Leads to a decrease in sampling error
  - Each subject acts as their own control

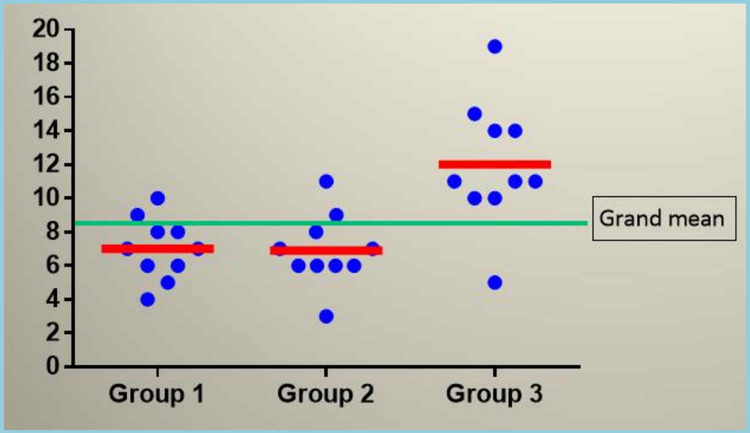
Standard one-way ANOVA assumes independence

To learn something about how the variables change over time or across situations

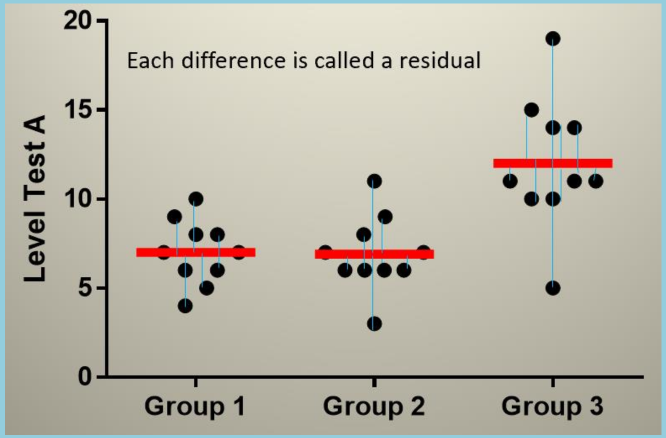
One-way ANOVA review

One-way ANOVA partitions the total variation,
i.e. total sum of squares, into 2 parts

SS_between



SS_within (SS_Residuals)



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

What happens to F-ratio if repeated measures are used?

Numerator (SS_between) includes:

Systematic variation due to difference in treatment

Some unsystematic random variation

Denominator (SS_within) includes:

Variation due to different individuals

Removed by the RM ANOVA since we have the same individual within each group

Denominator for RM ANOVA is smaller than the SS_within for a regular one-way ANOVA

When denominator decreases then F increases leading to higher chance to reject ($p<0.05$) the null hypothesis of equal means.

Turtle	sex	fed	fasted 10 days	fasted 20 days	SD by turtle
1	male	42.8	42.4	38.9	2.15
2	male	43.1	42.2	40.3	1.43
3	male	40.0	40.8	37.5	1.72
4	male	46.6	45.9	42.9	1.97
5	female	42.2	42.4	39.7	1.50
6	female	38.7	38.1	35.8	1.53
7	female	35.3	34.3	32.3	1.53
8	female	40.5	40.1	37.3	1.74
	SD by group	3.4	3.4	3.2	

Limitations of RM ANOVA

Missing data is a problem

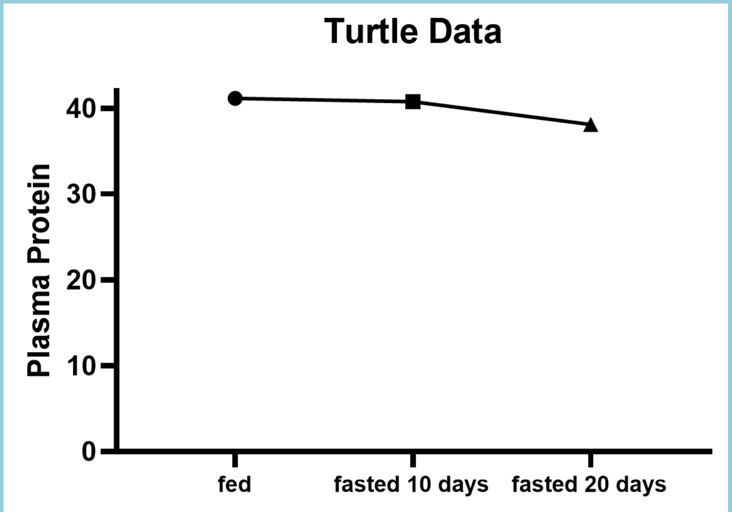
Often tough to complete a study with no missing data.

Involves important assumption of 'Sphericity' that Prism cannot test.

Example

Subject	Sex	Fed	Fasted 10 days	Fasted 20 days
1	Male	42.8	42.4	38.9
2	Male	43.1	42.2	40.3
3	Male	40.0	40.8	37.5
4	Male	46.6	45.9	42.9
5	Female	42.2	42.4	39.7
6	Female	38.7	38.1	35.8
7	Female	35.3	34.3	32.3
8	Female	40.5	40.1	37.3

Eight turtles had their plasma protein measured when well fed and 10 and 20 days after fasting



$$H_o: \mu_{Fed} = \mu_{Fast\ 10\ days} = \mu_{Fast\ 20\ days}$$

A significant finding indicates that at least one group of feeding and fasting is significantly differs from at least one other

RM ANOVA assumptions

Dependent variable should be continuous (plasma protein is continuous)

Independent variable(s) should be categorical with >2 levels (3 fasting times)

Time intervals for repeated measures should be equal (0, 10 20 days)

Subjects are independent (8 different turtles)

Data should be normally distributed and with no outliers by group (see next slide)

Homoscedasticity by group (see next slide)

The data should show 'sphericity' across groups (see following slides)

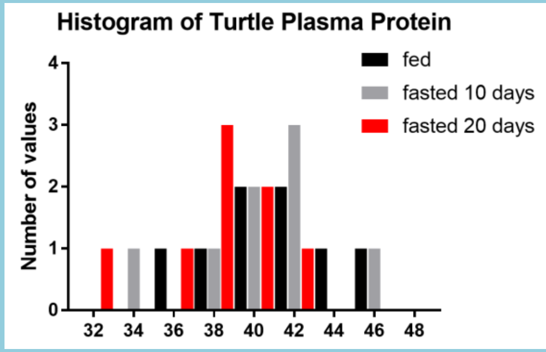
There should be no missing data (OK)

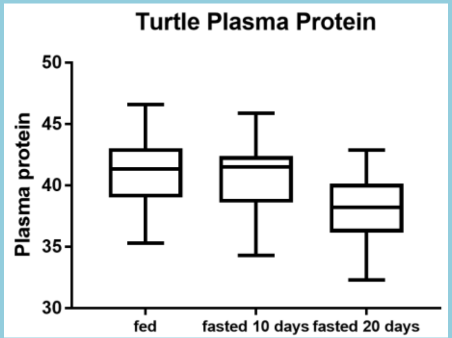
Normality and Homoscedasticity

Not as important when sample size is large unless deviations from normality and homoscedasticity are large

Design has to be balanced for a RM ANOVA

- The test is more robust against deviations from normality and homoscedasticity when balanced





	Col. stats	A	B	C
		fed	fasted 10 days	fasted 20 days
		Y	Y	Y
1	Number of values	8	8	8
2				
3	Minimum	35.3	34.3	32.3
4	25% Percentile	39.03	38.6	36.18
5	Median	41.35	41.5	38.2
6	75% Percentile	43.03	42.4	40.15
7	Maximum	46.6	45.9	42.9
8				
9	Mean	41.15	40.78	38.09
10	Std. Deviation	3.366	3.446	3.185
11	Std. Error of Mean	1.19	1.218	1.126
12				
13	Skewness	-0.2086	-0.6743	-0.4687
14	Kurtosis	0.7489	1.221	0.822
15				

Missing data

Prism does not allow missing data for RM ANOVA

If you have any missing data, you should impute it.

Use classic 'last observation carried forward' strategy for simplicity, but other more complicated imputation strategies may be more appropriate (see your friendly statistician)

Can use a mixed models approach (again see your friendly statistician)

RM ANOVA Key Assumption: Sphericity

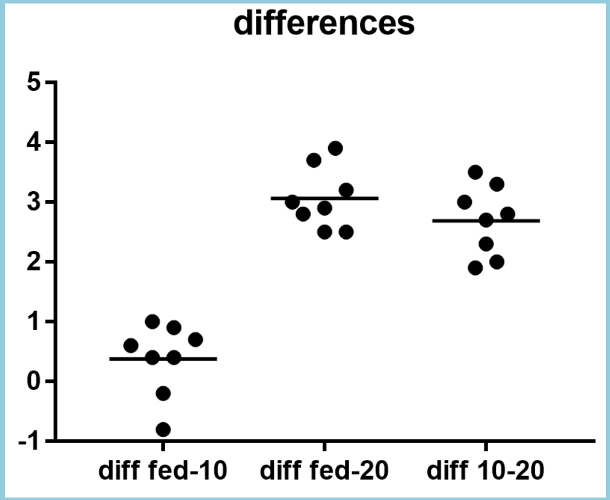
Refers to the equality of variances of the *differences* between the various time groups

Variances of the difference must be the same for each pair of time points

$$\text{Variance}_{t1\text{-}t2} \approx \text{Variance}_{t1\text{-}t3} \approx \text{Variance}_{t2\text{-}t3}$$

Correlation between pairs of time point groups is the same

$$\text{corr}_{t1 \text{ to } t2} \approx \text{corr}_{t2 \text{ to } t3} \approx \text{corr}_{t1 \text{ to } t3}$$



Sphericity (denoted by ε–*epsilon* - and sometimes referred to as *circularity*) is a type of *compound symmetry*.

ε indicates the degree to which sphericity has been violated

Sphericity

The violation of sphericity is serious for a RM ANOVA model

- Causes the test to become too prone to Type I error (false positive)
- Even more important when sample size is small

Determining if sphericity is violated is important

If a violation occurs, corrections can be made to produce a more valid critical *F*-value (i.e., reduce the increase in Type I error rate).

- Estimating the degree to which sphericity has been violated and apply a correction factor to the df of the *F*-distribution

- Common corrections: Geisser-Greenhouse (Prism) and Huynh-Feldt

How to test sphericity?

Mauchly's Test of Sphericity is a formal way of testing the assumption of sphericity

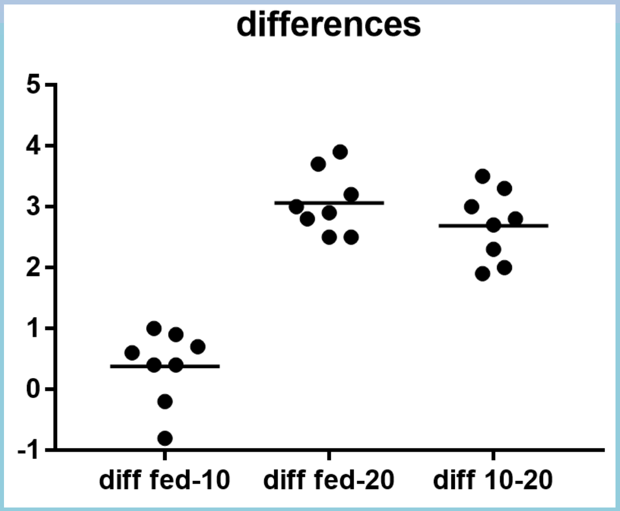
But fails to detect departures from sphericity in small samples and over-detects them in large samples

Also, Prism does not offer this test

Prism can run the test either with a correction (for lack of sphericity) or with no correction (assumes sphericity exists)

Eyeball for sphericity assumptions : Turtle Plasma Protein Data

Col. stats		A	B	C
		diff fed-10	diff fed-20	diff 10-20
		Y	Y	Y
1	Number of values	8	8	8
2				
3	Minimum	-0.8	2.5	1.9
4	25% Percentile	-0.05	2.575	2.075
5	Median	0.5	2.95	2.75
6	75% Percentile	0.85	3.575	3.225
7	Maximum	1	3.9	3.5
8				
9	Mean	0.375	3.063	2.688
10	Std. Deviation	0.6018	0.5153	0.5842
11	Std. Error of Mean	0.2128	0.1822	0.2065



Compound symmetry and sphericity: Look at the correlations:

	Fed	Fasted 10	Fasted 20
Fed	1	0.985	0.989
Fasted 10	0.985	1	0.988
Fasted 20	0.989	0.988	1

Compound symmetry holds if correlations in the blue cells (off diagonal) are similar

Compound symmetry implies sphericity (reverse not true)

An example without compound symmetry

	time1	time2	time3	time4
time1	1.00000	0.94035	-0.14150	0.28445
time2	0.94035	1.00000	-0.02819	0.26921
time3	-0.14150	-0.02819	1.00000	0.27844
time4	0.28445	0.26921	0.27844	1.00000

Certainly do NOT have equal correlations. Time 1 and time 2 are highly correlated (r=0.94), but time 1 and time 3 are *inversely* correlated.

CONCLUSION: Probably no sphericity

When in doubt, assume sphericity assumption is violated and run the test with the correction

Other Things to Consider in a RM ANOVA

Spacing of time intervals

RM ANOVA require that all subjects are measured at equal time intervals

i.e., days 0, 5, 10, 15, 20

NOT days 0 10 15 21 30

Missing Data

Prism will not perform a RM ANOVA if data are missing

What happens if you don’t check the assumptions?

Prism will fail to run an analysis if there is missing data

Inaccurate results

Increased Type-I error – false positive (e.g. if sphericity is assumed but not satisfied)

Decreased power (i.e. increased Type-II error – false negative)

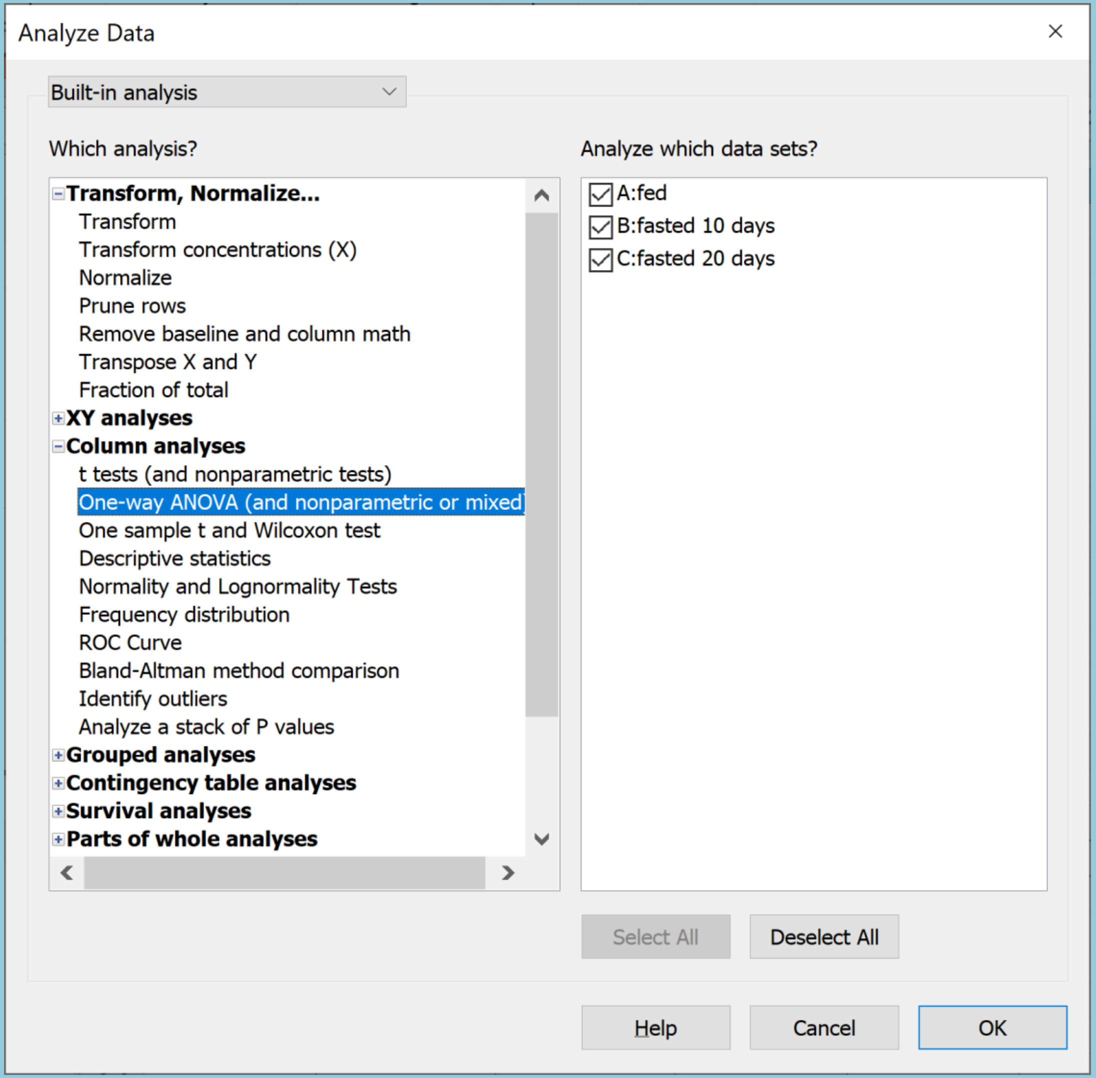
Journals won’t accept the paper

Statistics Instructor will be unhappy

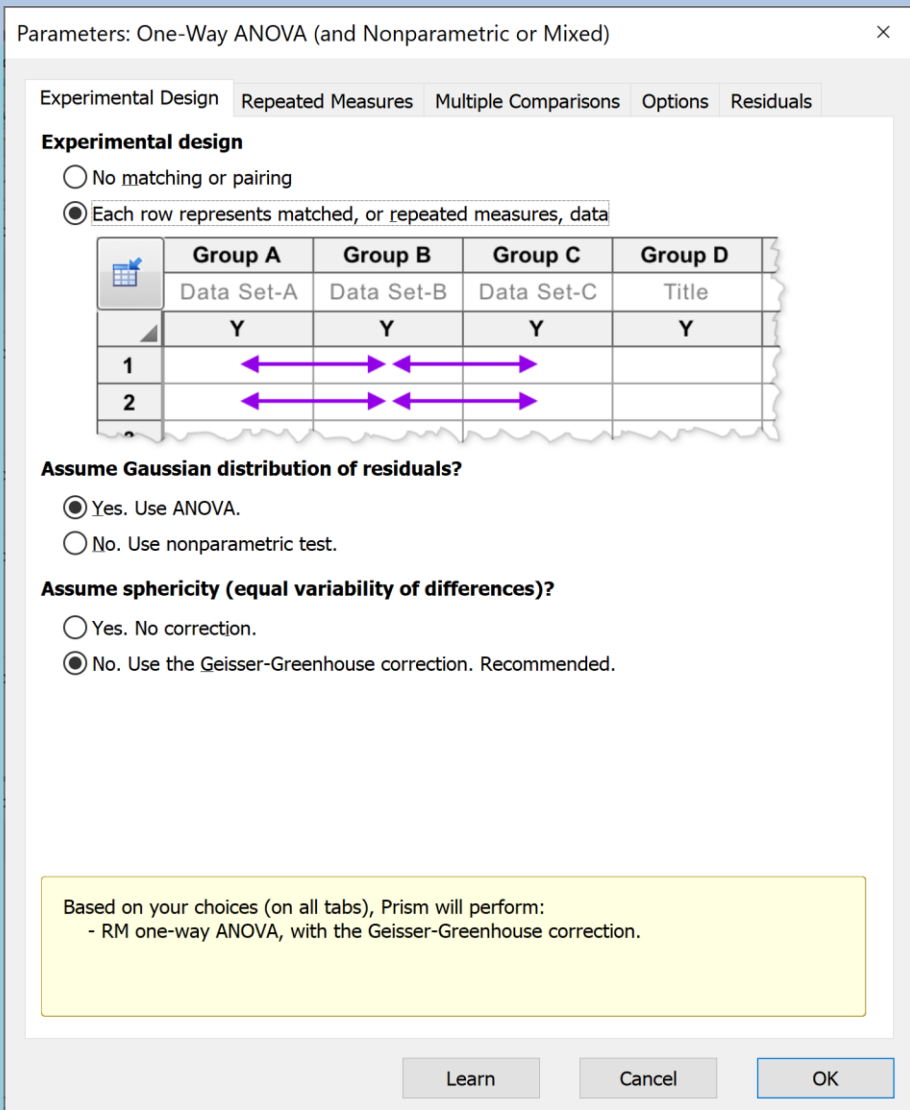
One-way RM ANOVA Example: Turtles, fasting and plasma protein

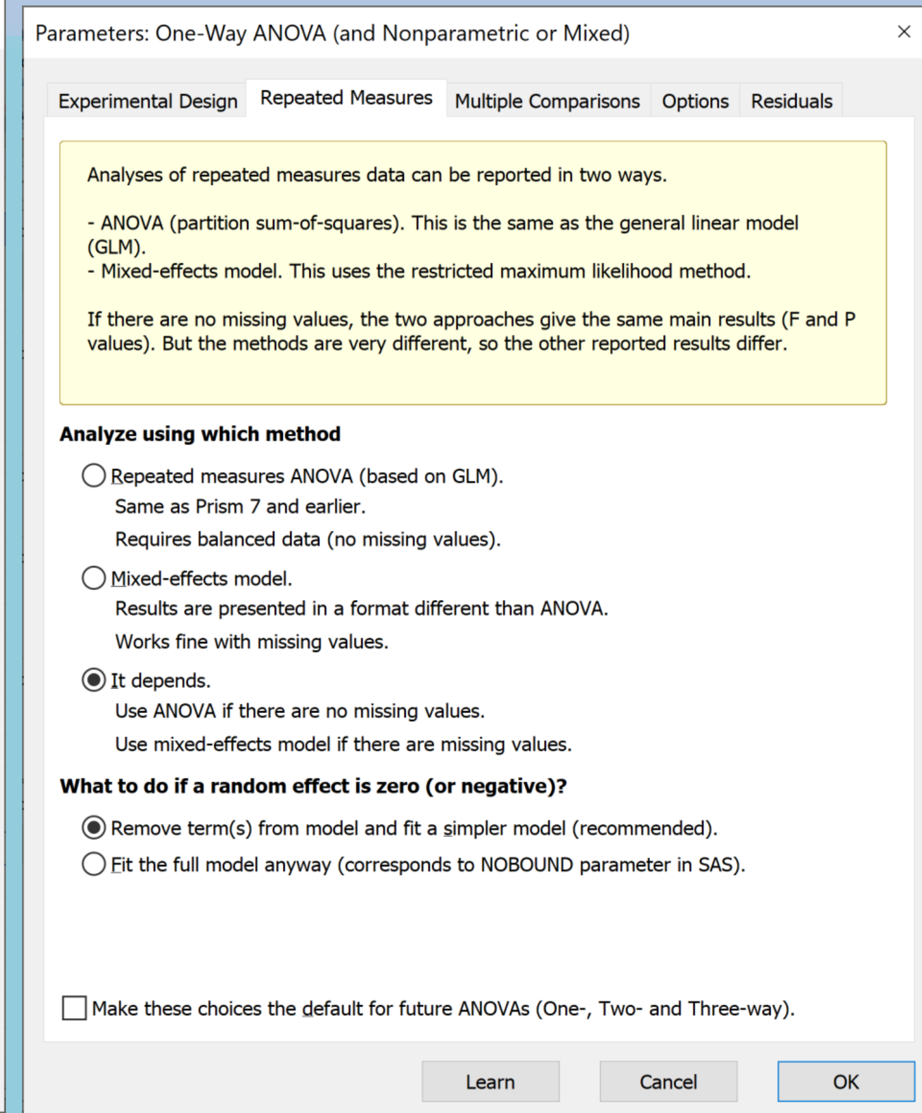
Use a “Columns” table

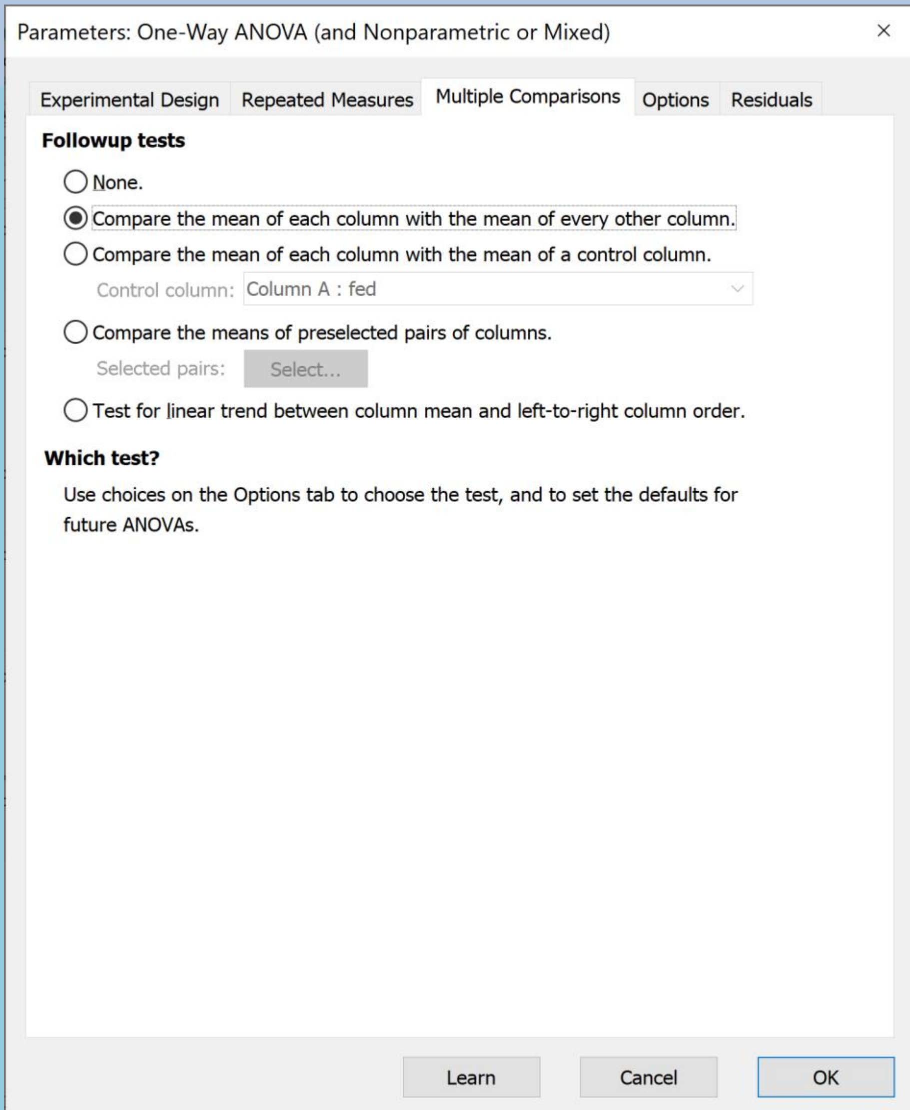
	Group A	Group B	Group C
	fed	fasted 10 days	fasted 20 days
1	42.8	42.4	38.9
2	43.1	42.2	40.3
3	40.0	40.8	37.5
4	46.6	45.9	42.9
5	42.2	42.4	39.7
6	38.7	38.1	35.8
7	35.3	34.3	32.3
8	40.5	40.1	37.3

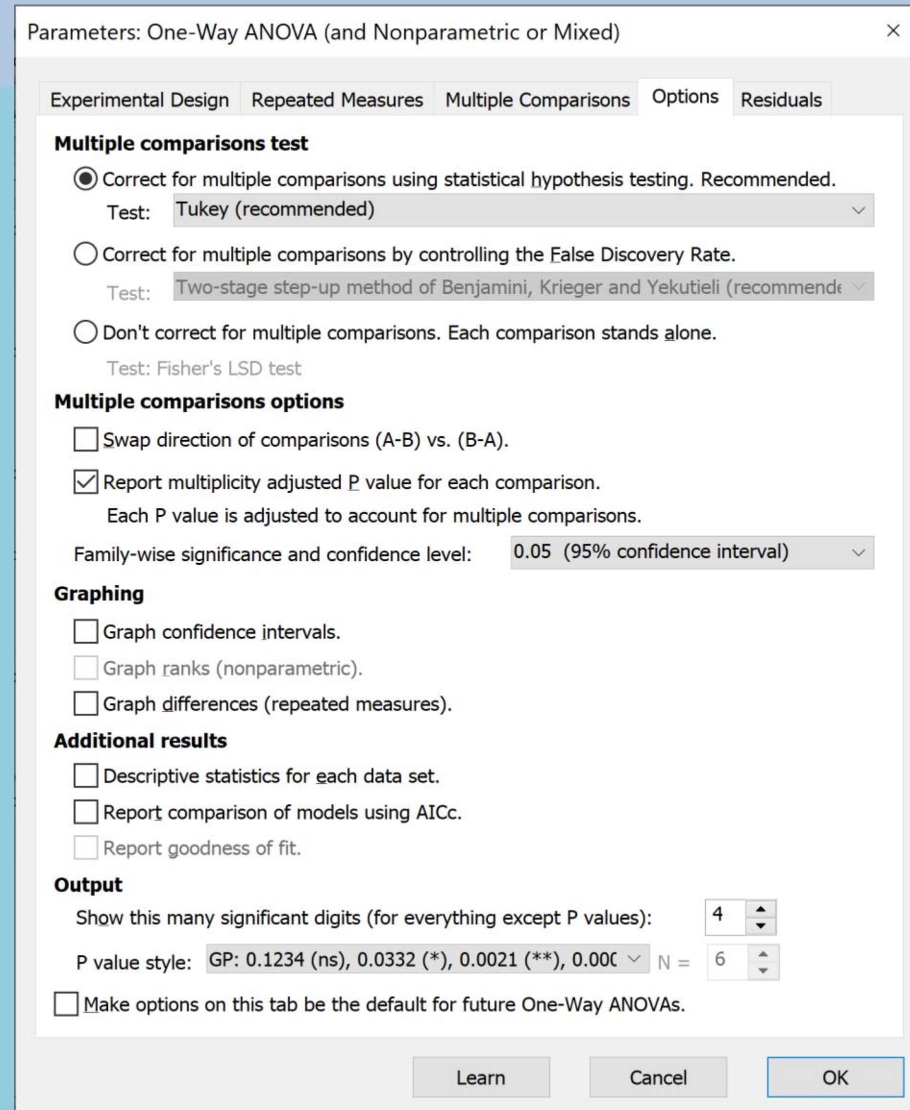


One-way RM ANOVA using prism: Not assuming sphericity (Prism recommended)






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 : fed
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
OK

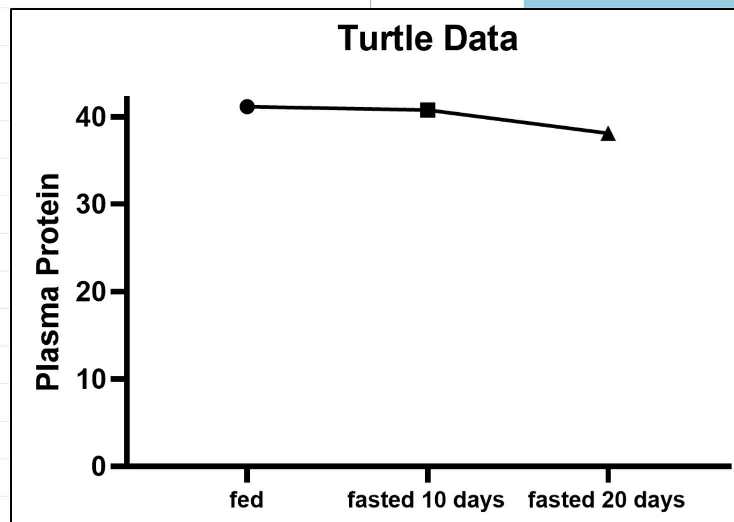

Parameters: One-Way ANOVA (and Nonparametric or Mixed)
Experimental Design
Repeated Measures
Multiple Comparisons
Options
Residuals
Multiple comparisons test
Correct for multiple comparisons using statistical hypothesis testing. Recommended.
Test: Tukey (recommended)
Correct for multiple comparisons 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

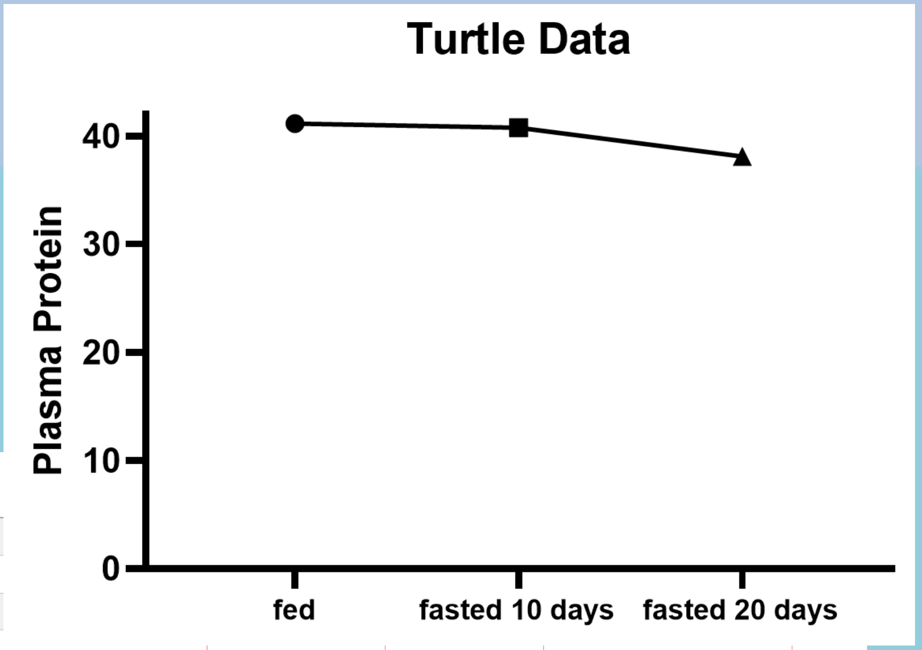
	Group A	Group B	Group C
	fed	fasted 10 days	fasted 20 days
	Y	Y	Y
1	42.8	42.4	38.9
2	43.1	42.2	40.3
3	40.0	40.8	37.5
4	46.6	45.9	42.9
5	42.2	42.4	39.7
6	38.7	38.1	35.8
7	35.3	34.3	32.3
8	40.5	40.1	37.3
9			

ANOVA results | Multiple comparisons

RM one-way ANOVA
ANOVA results

1	Table Analyzed	Turtle Data				
2						
3	**Repeated measures ANOVA summary**					
4	Assume sphericity?	No				
5	F	138.2				
6	P value	<0.0001				
7	P value summary	****				
8	Statistically significant (P < 0.05)?	Yes				
9	Geisser-Greenhouse's epsilon	0.9680				
10	R square	0.9518				
11						
12	**Was the matching effective?**					
13	F	204.5				
14	P value	<0.0001				
15	P value summary	****				
16	Is there significant matching (P < 0.05)?	Yes				
17	R square	0.8313				
18						
19	**ANOVA table**	**SS**	**DF**	**MS**	**F (DFn, DFd)**	**P value**
20	Treatment (between columns)	44.65	2	22.32	F (1.936, 13.55) = 138.2	P<0.0001
21	Individual (between rows)	231.2	7	33.02	F (7, 14) = 204.5	P<0.0001
22	Residual (random)	2.261	14	0.1615		
23	Total	278.1	23			



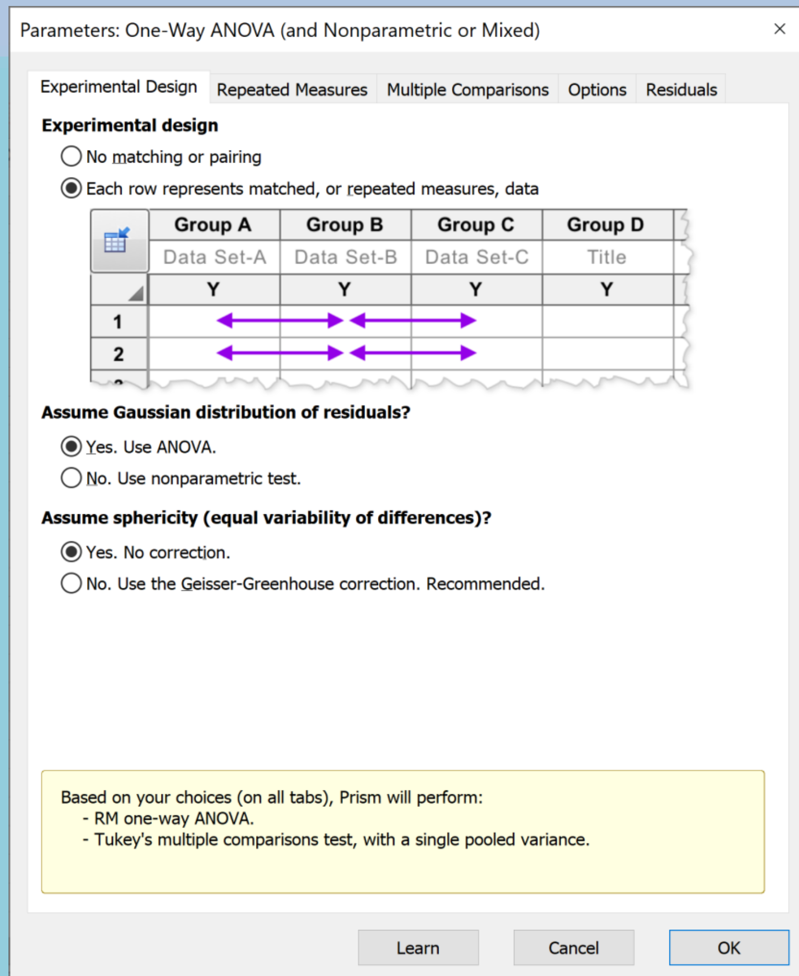


ANOVA results | Multiple comparisons

RM one-way ANOVA
Multiple comparisons

1	Number of families	1					
2	Number of comparisons per family	3					
3	Alpha	0.05					
4							
5	**Tukey's multiple comparisons test**	**Mean Diff.**	**95.00% CI of diff.**	**Significant?**	**Summary**	**Adjusted P Value**	
6	fed vs. fasted 10 days	0.3750	-0.2516 to 1.002	No	ns	0.2496	A-B
7	fed vs. fasted 20 days	3.063	2.526 to 3.599	Yes	****	<0.0001	A-C
8	fasted 10 days vs. fasted 20 days	2.688	2.079 to 3.296	Yes	****	<0.0001	B-C
9							
10	**Test details**	**Mean 1**	**Mean 2**	**Mean Diff.**	**SE of diff.**	**n1**	**n2**
11	fed vs. fasted 10 days	41.15	40.78	0.3750	0.2128	8	8
12	fed vs. fasted 20 days	41.15	38.09	3.063	0.1822	8	8
13	fasted 10 days vs. fasted 20 days	40.78	38.09	2.688	0.2065	8	8

Now assuming sphericity: what is different?



RM one-way ANOVA
ANOVA results

1	Table Analyzed	Turtle Data				
2						
3	**Repeated measures ANOVA summary**					
4	Assume sphericity?	No				
5	F	138.2				
6	P value	<0.0001				
7	P value summary	****				
8	Statistically significant (P < 0.05)?	Yes				
9	Geisser-Greenhouse's epsilon	0.9680				
10	R square	0.9518				
11						
12	**Was the matching effective?**					
13	F	204.5				
14	P value	<0.0001				
15	P value summary	****				
16	Is there significant matching (P < 0.05)?	Yes				
17	R square	0.8313				
18						
19	**ANOVA table**	**SS**	**DF**	**MS**	**F (DFn, DFd)**	**P value**
20	Treatment (between columns)	44.65	2	22.32	F (1.936, 13.55) = 138.2	P<0.0001
21	Individual (between rows)	231.2	7	33.02	F (7, 14) = 204.5	P<0.0001
22	Residual (random)	2.261	14	0.1615		
23	Total	278.1	23			

RM one-way ANOVA
ANOVA results

1	Table Analyzed	Turtle Data				
2						
3	**Repeated measures ANOVA summary**					
4	Assume sphericity?	Yes				
5	F	138.2				
6	P value	<0.0001				
7	P value summary	****				
8	Statistically significant (P < 0.05)?	Yes				
9	R square	0.9518				
10						
11	**Was the matching effective?**					
12	F	204.5				
13	P value	<0.0001				
14	P value summary	****				
15	Is there significant matching (P < 0.05)?	Yes				
16	R square	0.8313				
17						
18	**ANOVA table**	**SS**	**DF**	**MS**	**F (DFn, DFd)**	**P value**
19	Treatment (between columns)	44.65	2	22.32	F (2, 14) = 138.2	P<0.0001
20	Individual (between rows)	231.2	7	33.02	F (7, 14) = 204.5	P<0.0001
21	Residual (random)	2.261	14	0.1615		
22	Total	278.1	23			

Decrease the df and the critical value increases; F statistics must be larger to reach $p \leq 0.05$

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

Let's now do a standard one-way ANOVA
We will pretend the data are not paired

Notice the difference

One-way ANOVA

ANOVA table	SS	DF	MS	F (DFn, DFd)	P value
Treatment (between columns)	44.65	2	22.32	F (2, 21) = 2.008	P=0.1592
Residual (within columns)	233.4	21	11.12		
Total	278.1	23			

One-way RM ANOVA

ANOVA table	SS	DF	MS	F (DFn, DFd)	P value
Treatment (between columns)	44.65	2	22.32	F (2, 14) = 138.2	P<0.0001
Individual (between rows)	231.2	7	33.02	F (7, 14) = 204.5	P<0.0001
Residual (random)	2.261	14	0.1615		
Total	278.1	23			

The residual SS is much smaller for RM ANOVA
The smaller the residual SS, the larger the F-statistic

Another data format

New Data Table and Graph

New table & graph

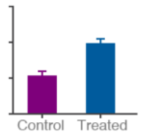
- XY
- Column
- Grouped
- Contingency
- Survival
- Parts of whole
- Multiple variables
- Nested

Existing file

- Clone a graph

Column tables have one grouping variable, with each group defined by a column

	A	B
	Control	Treated
	Y	Y
1		
2		



Learn more

Data table:

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

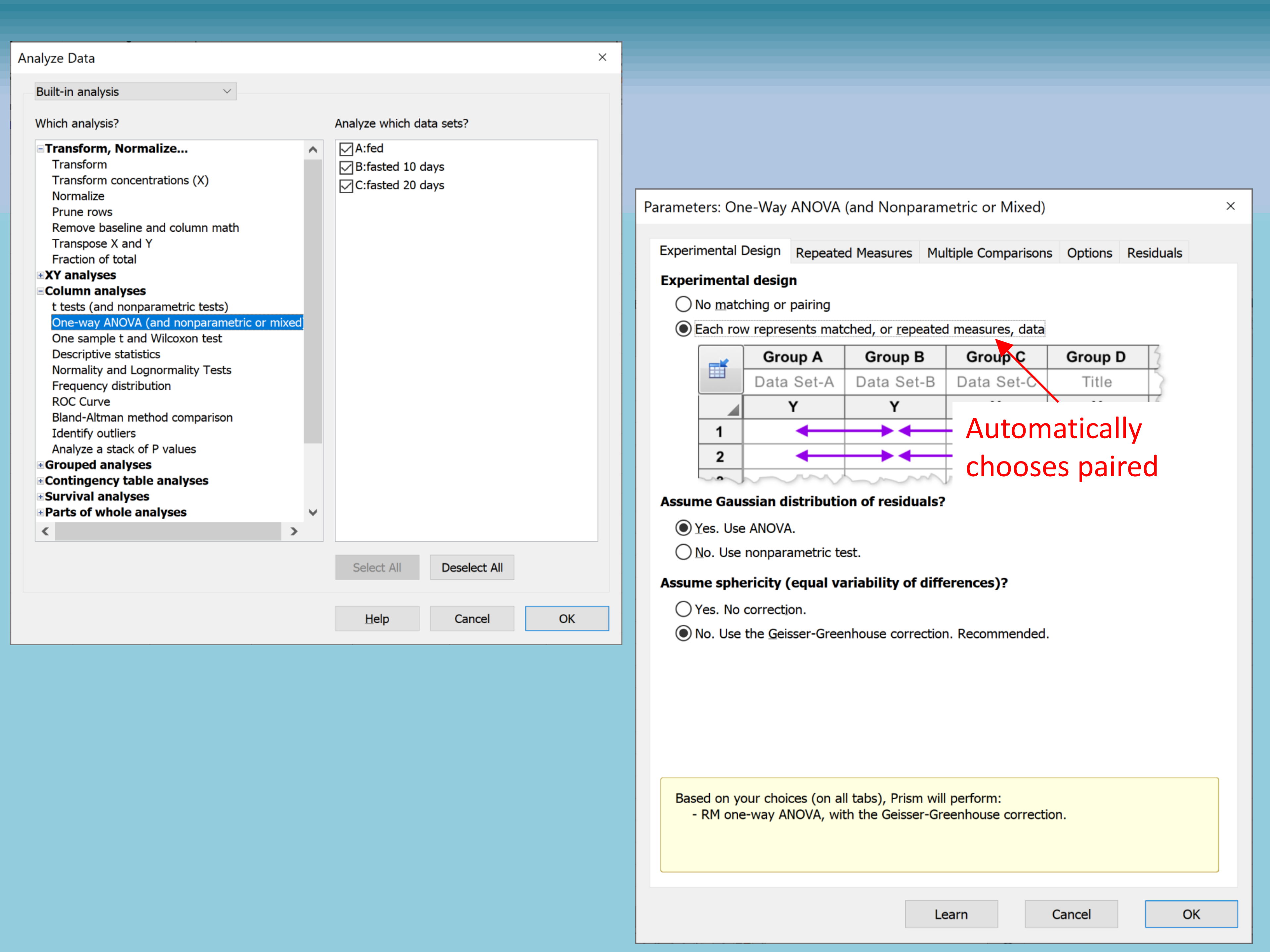
Options:

- ( ) Enter replicate values, stacked into columns
- (•) Enter paired or repeated measures data - each subject on a separate row
- ( ) Enter and plot error values already calculated elsewhere

Enter: Mean, SD, N

Prism Tips

Table format: Column		Group A	Group B	Group C
		fed	fasted 10 days	fasted 20 days
		Y	Y	Y
1	1	42.8	42.4	38.9
2	2	43.1	42.2	40.3
3	3	40.0	40.8	37.5
4	4	46.6	45.9	42.9
5	5	42.2	42.4	39.7
6	6	38.7	38.1	35.8
7	7	35.3	34.3	32.3
8	8	40.5	40.1	37.3


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:fed
B:fasted 10 days
C:fasted 20 days
Select All
Deselect All
Help
Cancel
OK
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
1
2
Automatically chooses paired
Assume Gaussian distribution of residuals?
Yes. Use ANOVA.
No. Use nonparametric test.
Assume sphericity (equal variability of differences)?
Yes. No correction.
No. Use the Geisser-Greenhouse correction. Recommended.
Based on your choices (on all tabs), Prism will perform:
- RM one-way ANOVA, with the Geisser-Greenhouse correction.
Learn
Cancel
OK

Post Hoc tests

Global RM ANOVA ≤ 0.05

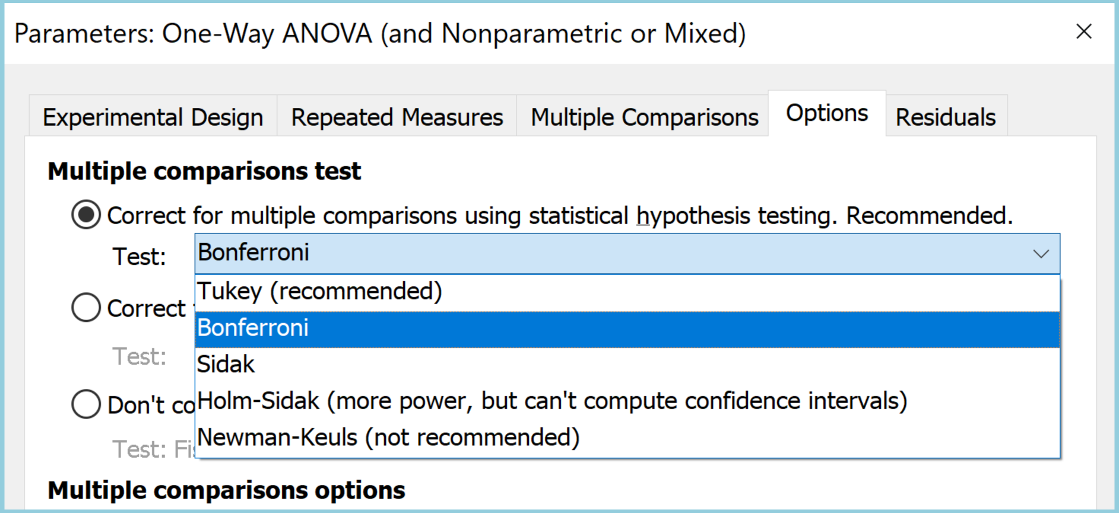
Need Post-Hoc Multiple Comparisons

which groups are different?

If there is sphericity, use standard methods (e.g., Tukey's)

Prism will recommend Tukey's with or without using the correction

Without sphericity better to use conservative approach: Bonferroni test from drop down menu of choice of tests for multiple comparisons



Post hoc tests in Prism

When sphericity is assumed

Tukey's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Significant?	Summary	Adjusted P Value	
fed vs. fasted 10 days	0.375	-0.1509 to 0.9009	No	ns	0.1850	A-B
fed vs. fasted 20 days	3.063	2.537 to 3.588	Yes	****	<0.0001	A-C
fasted 10 days vs. fasted 20 days	2.688	2.162 to 3.213	Yes	****	<0.0001	B-C

When sphericity is not assumed

Tukey's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Significant?	Summary	Adjusted P Value	
fed vs. fasted 10 days	0.375	-0.2516 to 1.002	No	ns	0.2496	A-B
fed vs. fasted 20 days	3.063	2.526 to 3.599	Yes	****	<0.0001	A-C
fasted 10 days vs. fasted 20 days	2.688	2.079 to 3.296	Yes	****	<0.0001	B-C

Write up results (we assume sphericity is present)

The data passed the assumptions for a one-way RM ANOVA (no missing data, independent subjects, normality, homoscedasticity, sphericity, equal time points). With $p < 0.0001$ (two-tailed hypothesis $F(2,14)=138.2$, $\alpha=0.05$), we reject the null hypothesis of no difference in population means by group and conclude there is at least one group different from one other. Tukey post-hoc tests showed that 'Fasted 20' (mean=38.1) was significantly different from 'Fed' (41.2) and 'Fasted 10' (40.8). There was no significant difference between 'Fed' and 'Fasted 10'. We conclude that plasma protein levels decrease with fasting.

	Col. stats	A	B	C
		fed	fasted 10 days	fasted 20 days
		Y	Y	Y
1	Number of values	8	8	8
2				
3	Minimum	35.3	34.3	32.3
4	25% Percentile	39.03	38.6	36.18
5	Median	41.35	41.5	38.2
6	75% Percentile	43.03	42.4	40.15
7	Maximum	46.6	45.9	42.9
8				
9	Mean	41.15	40.78	38.09
10	Std. Deviation	3.366	3.446	3.185
11	Std. Error of Mean	1.19	1.218	1.126

How you might write it if sphericity is not present

The data passed most of the assumptions for a one-way RM ANOVA (no missing data, independent subjects, normality, homoscedasticity by groups, equal time points), except for sphericity. Because of this we will use the one-way RM ANOVA with the Greenhouse-Geisser correction. With $p < 0.0001$ etc…

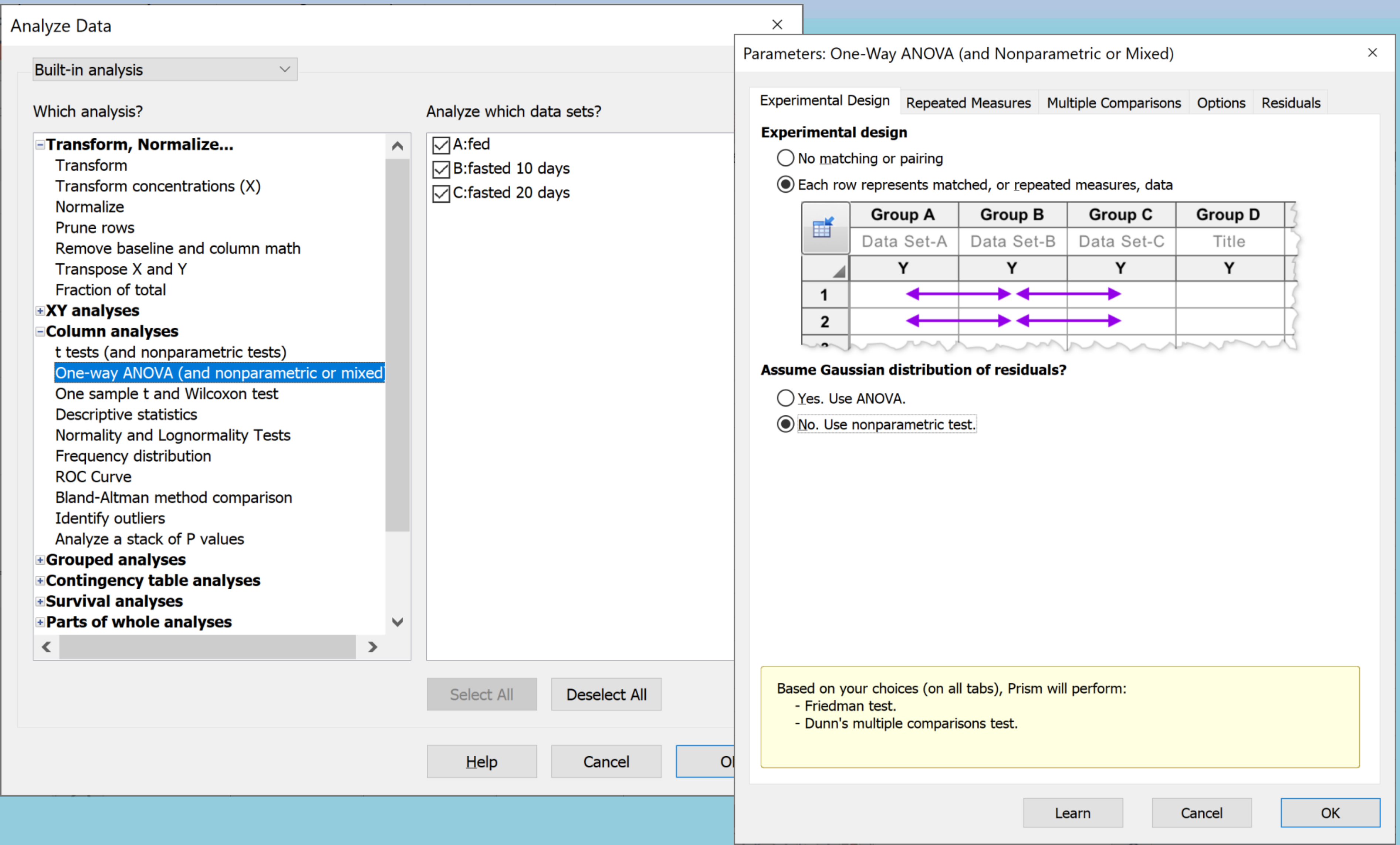
Other options if assumptions for RM ANOVA are not satisfied

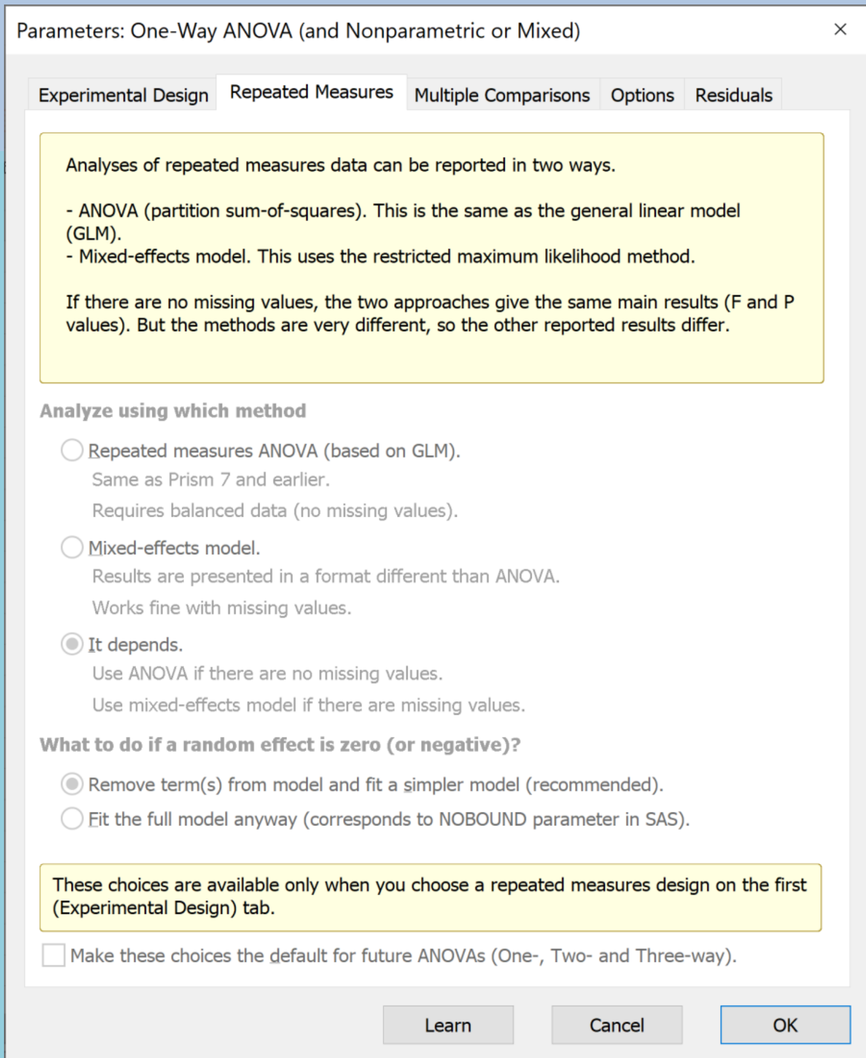
Use mixed models

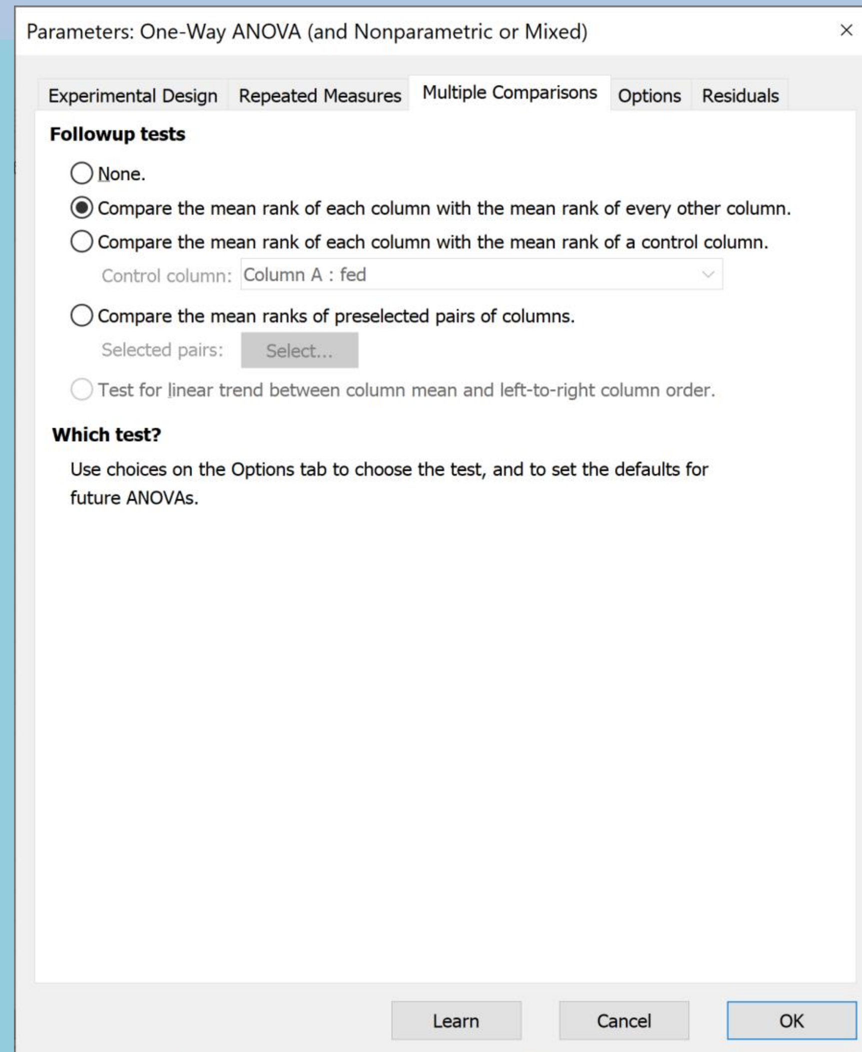
- Mixed models are advanced modelling methods (New in Prism 8)
- Can deal with missing values, uneven time points, and lack of sphericity

Otherwise use the non-parametric Friedman Test

Non-parametric alternative to the one-way RM ANOVA: Freedman Test




Parameters: One-Way ANOVA (and Nonparametric or Mixed)
Experimental Design
Repeated Measures
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 choose a repeated measures design on the first (Experimental Design) tab.
Make these choices the default for future ANOVAs (One-, Two- and Three-way).
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 rank of each column with the mean rank of every other column.
Compare the mean rank of each column with the mean rank of a control column.
Control column: Column A : fed
Compare the mean ranks 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
OK

Friedman Test

	Friedman test ANOVA results	
1	Table Analyzed	Turtle Data Paired Entry
2		
3	**Friedman test**	
4	P value	0.0003
5	Exact or approximate P value?	Exact
6	P value summary	***
7	Are means signif. different? (P < 0.05)	Yes
8	Number of groups	3
9	Friedman statistic	13.00
10		
11	**Data summary**	
12	Number of treatments (columns)	3
13	Number of subjects (rows)	8
14		

One-way RM ANOVA

ANOVA table	SS	DF	MS	F (DFn, DFd)	P value
Treatment (between columns)	44.65	2	22.32	F (2, 14) = 138.2	P<0.0001
Individual (between rows)	231.2	7	33.02	F (7, 14) = 204.5	P<0.0001
Residual (random)	2.261	14	0.1615		
Total	278.1	23			

Friedman Test

ANVOA results | Multiple comparisons

Friedman 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**	**Rank sum diff.**	**Significant?**	**Summary**	**Adjusted P Value**	
6	fed vs. fasted 10 days	4.000	No	ns	0.9519	A-B
7	fed vs. fasted 20 days	14.00	Yes	**	0.0014	A-C
8	fasted 10 days vs. fasted 20 days	10.00	Yes	*	0.0373	B-C
9						
10	**Test details**	**Rank sum 1**	**Rank sum 2**	**Rank sum diff.**	**n1**	**n2**
11	fed vs. fasted 10 days	22.00	18.00	4.000	8	8
12	fed vs. fasted 20 days	22.00	8.000	14.00	8	8
13	fasted 10 days vs. fasted 20 days	18.00	8.000	10.00	8	8

One-way RM ANOVA

ANOVA results | Multiple comparisons

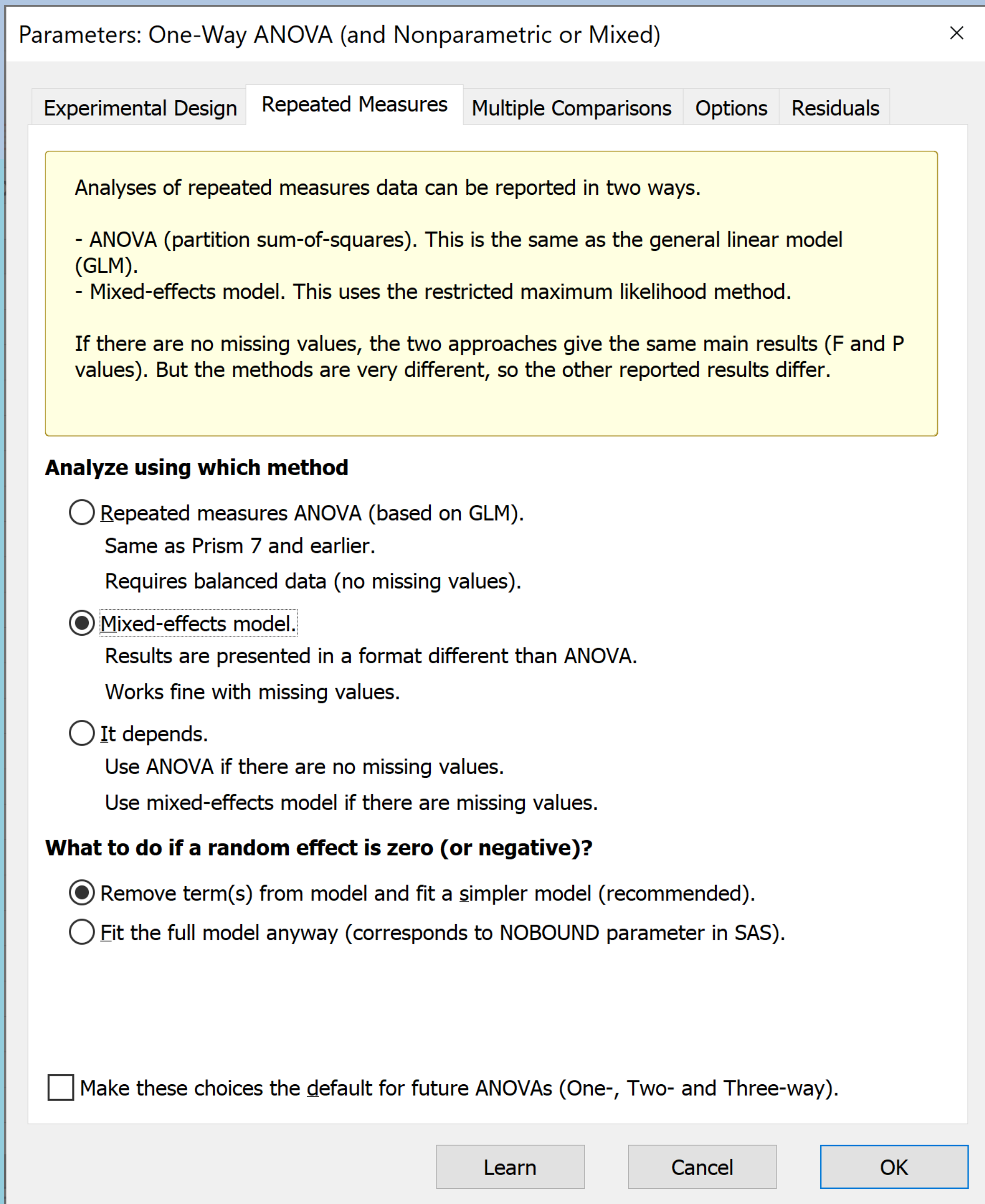
RM one-way ANOVA
Multiple comparisons

1	Number of families	1					
2	Number of comparisons per family	3					
3	Alpha	0.05					
4							
5	**Tukey's multiple comparisons test**	**Mean Diff.**	**95.00% CI of diff.**	**Significant?**	**Summary**	**Adjusted P Value**	
6	fed vs. fasted 10 days	0.3750	-0.2516 to 1.002	No	ns	0.2496	A-B
7	fed vs. fasted 20 days	3.063	2.526 to 3.599	Yes	****	<0.0001	A-C
8	fasted 10 days vs. fasted 20 days	2.688	2.079 to 3.296	Yes	****	<0.0001	B-C
9							
10	**Test details**	**Mean 1**	**Mean 2**	**Mean Diff.**	**SE of diff.**	**n1**	**n2**
11	fed vs. fasted 10 days	41.15	40.78	0.3750	0.2128	8	8
12	fed vs. fasted 20 days	41.15	38.09	3.063	0.1822	8	8
13	fasted 10 days vs. fasted 20 days	40.78	38.09	2.688	0.2065	8	8

An example of how you might write up the results of the Friedman test

The data failed many of the assumptions of a one-way RM ANOVA (normality, homoscedasticity, sphericity) but met the assumptions of independence of groups, paired/correlated data, no missing data, equally spaced time measurements, we will use the non-parametric Friedman test. With p=0.0003 (two-sided hypothesis, a=0.05), we reject the null hypothesis of equal medians and conclude that at least one median is different. Dunn's post-hoc test 'Fasted 20' (median=38.2) was significantly different from 'Fed' (41.35) and 'Fasted 10' (41.5). There was no significant difference between 'Fed' and 'Fasted 10'. We conclude that plasma protein levels decrease with fasting.

What about a mixed model?



THERE IS A BIG BUT….

Tabular results | Multiple comparisons

Mixed-effects analysis
Tabular results

1	Table Analyzed	Turtle Data	
2			
3	**Mixed-effects model (REML)**	Matching: Across row	
4	Assume sphericity?	No	
5	Alpha	0.05	
6			
7	**Fixed effect (type III)**	**P value**	**P value summary**
8	Treatment (between columns)	<0.0001	****

Tabular results | Multiple comparisons

Mixed-effects analysis
Multiple comparisons

1	Number of families	1					
2	Number of comparisons per family	3					
3	Alpha	0.05					
4							
5	**Tukey's multiple comparisons test**	**Mean Diff.**	**95.00% CI of diff.**	**Significant?**	**Summary**	**Adjusted P Value**	
6	fed vs. fasted 10 days	0.3750	-0.2516 to 1.002	No	ns	0.2496	A-B
7	fed vs. fasted 20 days	3.063	2.526 to 3.599	Yes	****	<0.0001	A-C
8	fasted 10 days vs. fasted 20 days	2.688	2.079 to 3.296	Yes	****	<0.0001	B-C
9							
10	**Test details**	**Mean 1**	**Mean 2**	**Mean Diff.**	**SE of diff.**	**n1**	**n2**
11	fed vs. fasted 10 days	41.15	40.78	0.3750	0.2128	8	8
12	fed vs. fasted 20 days	41.15	38.09	3.063	0.1822	8	8
13	fasted 10 days vs. fasted 20 days	40.78	38.09	2.688	0.2065	8	8

Running a mixed model is actually rather complex – seek professional help

SAS:
It is very important to explore different variance-covariance structures when running a mixed model on repeated measure.

The output contains fit statistics indicating how well each model with different variance-covariance structures fits the data compared to other models.

Different variance-covariance structures

Compound symmetry
Auto-regressive
Unstructured
etc.

We can do a two-way RM ANOVA to test if there is a difference in plasma protein by fasting *and* turtle sex

Subject	Sex	Fed	Fasted 10 days	Fasted 20 days
1	Male	42.8	42.4	38.9
2	Male	43.1	42.2	40.3
3	Male	40.0	40.8	37.5
4	Male	46.6	45.9	42.9
5	Female	42.2	42.4	39.7
6	Female	38.7	38.1	35.8
7	Female	35.3	34.3	32.3
8	Female	40.5	40.1	37.3

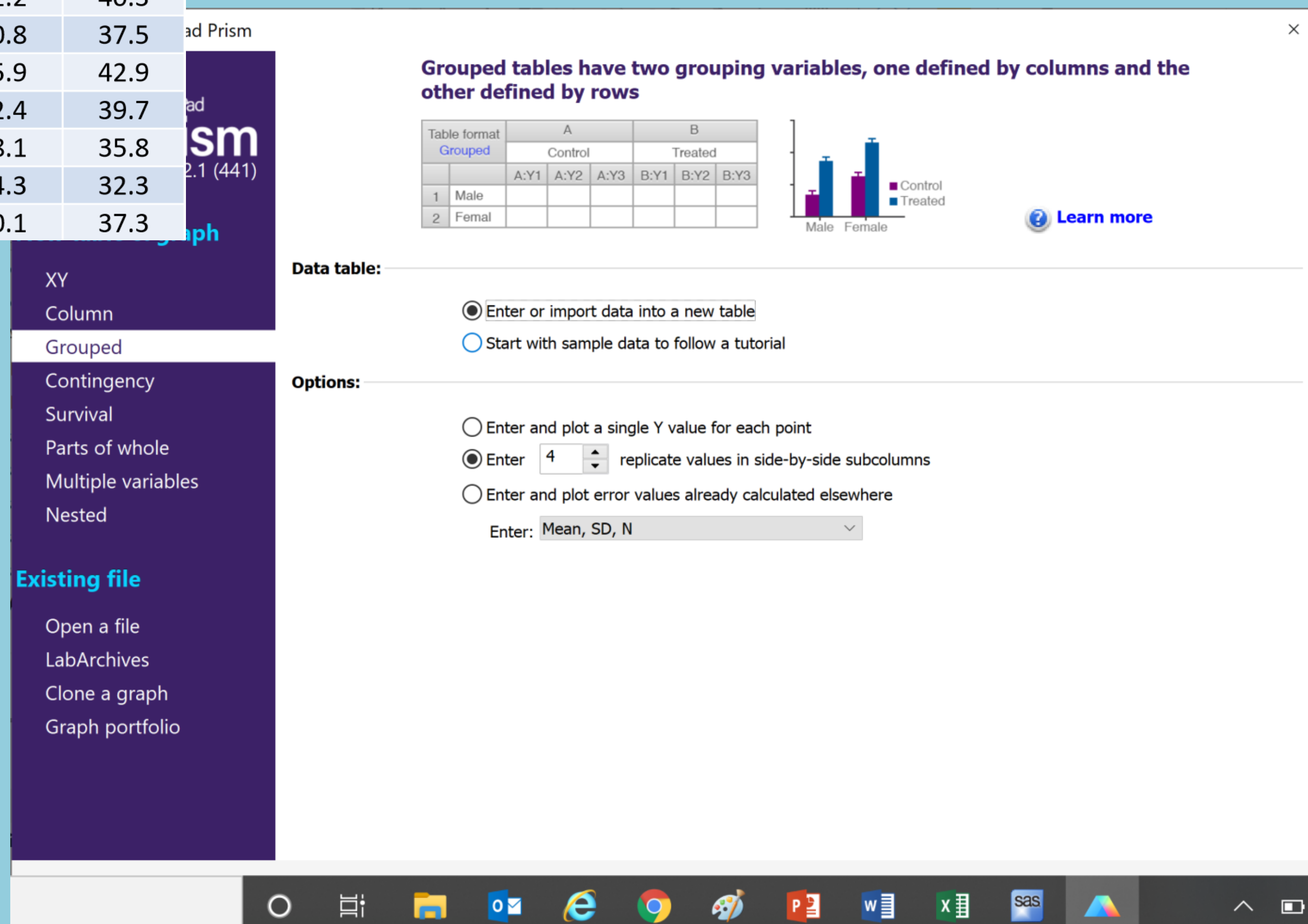
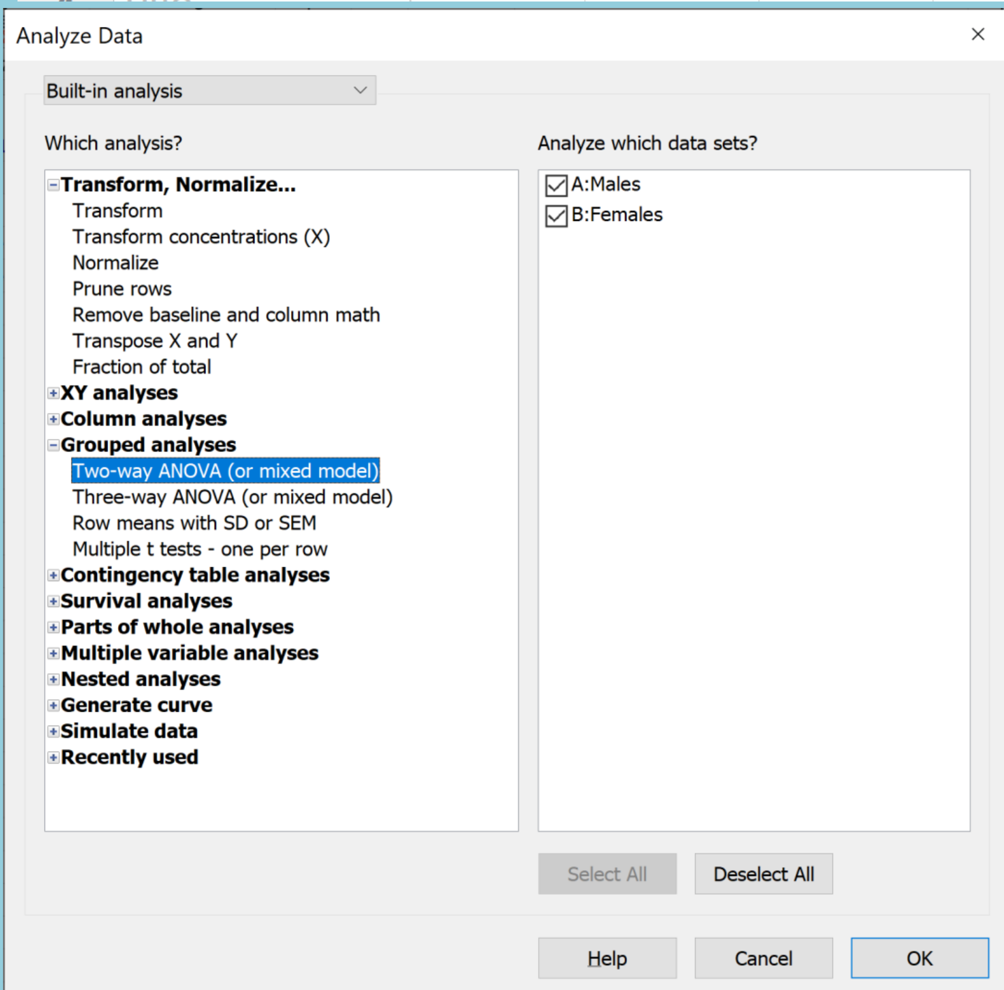
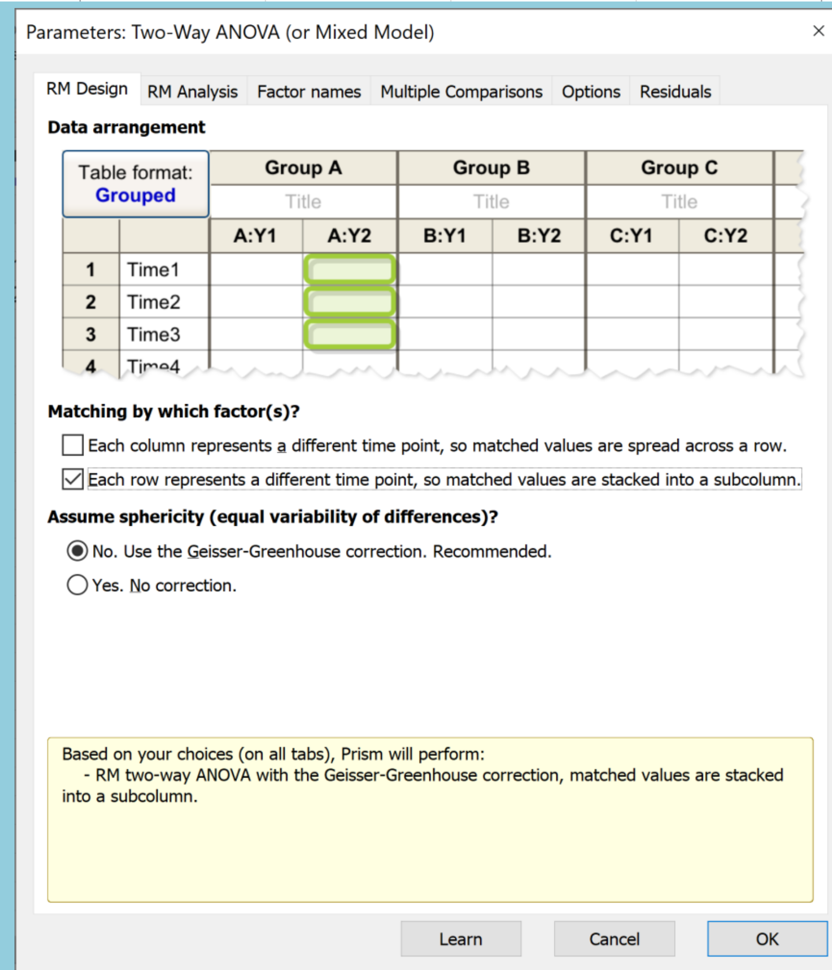
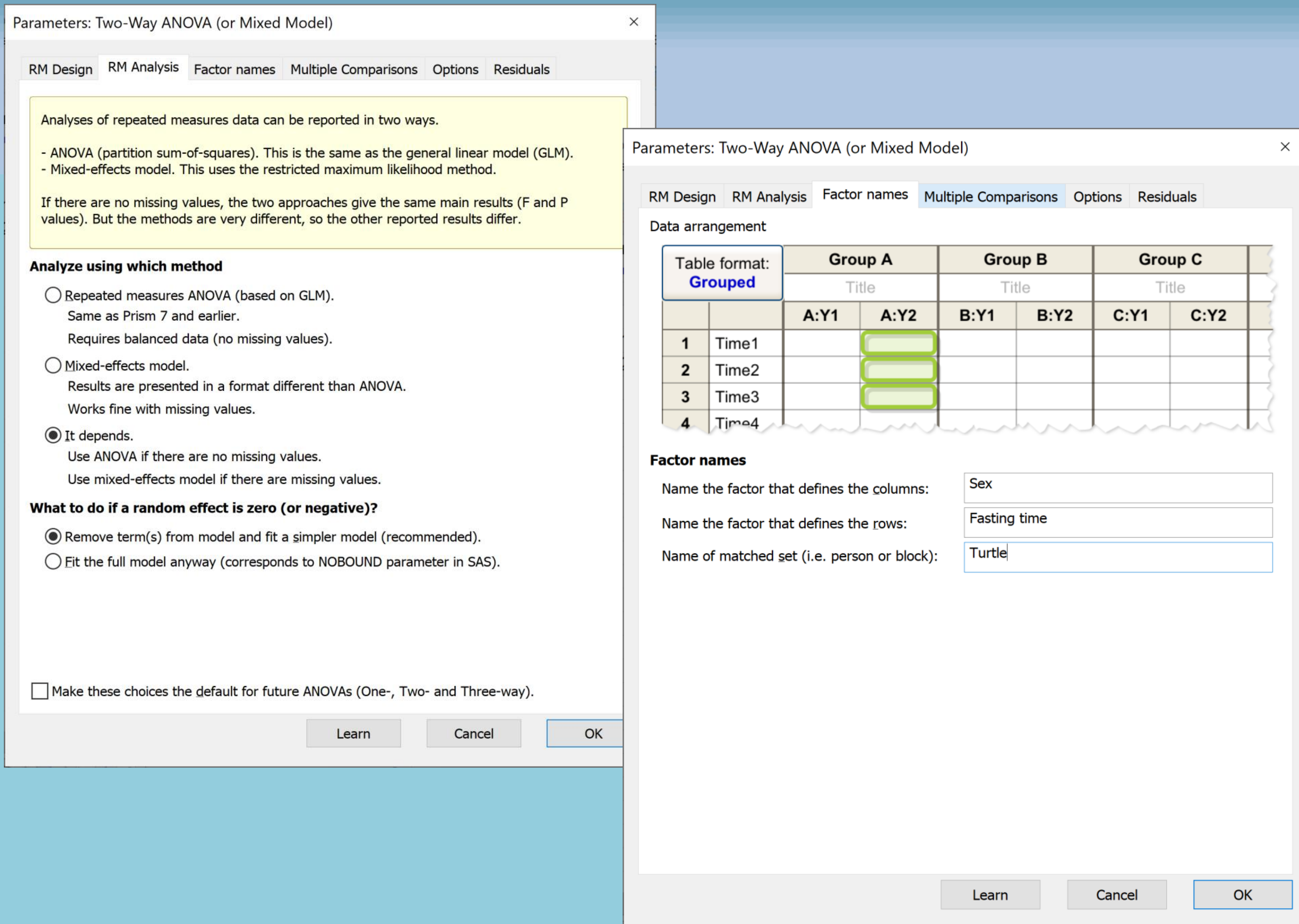


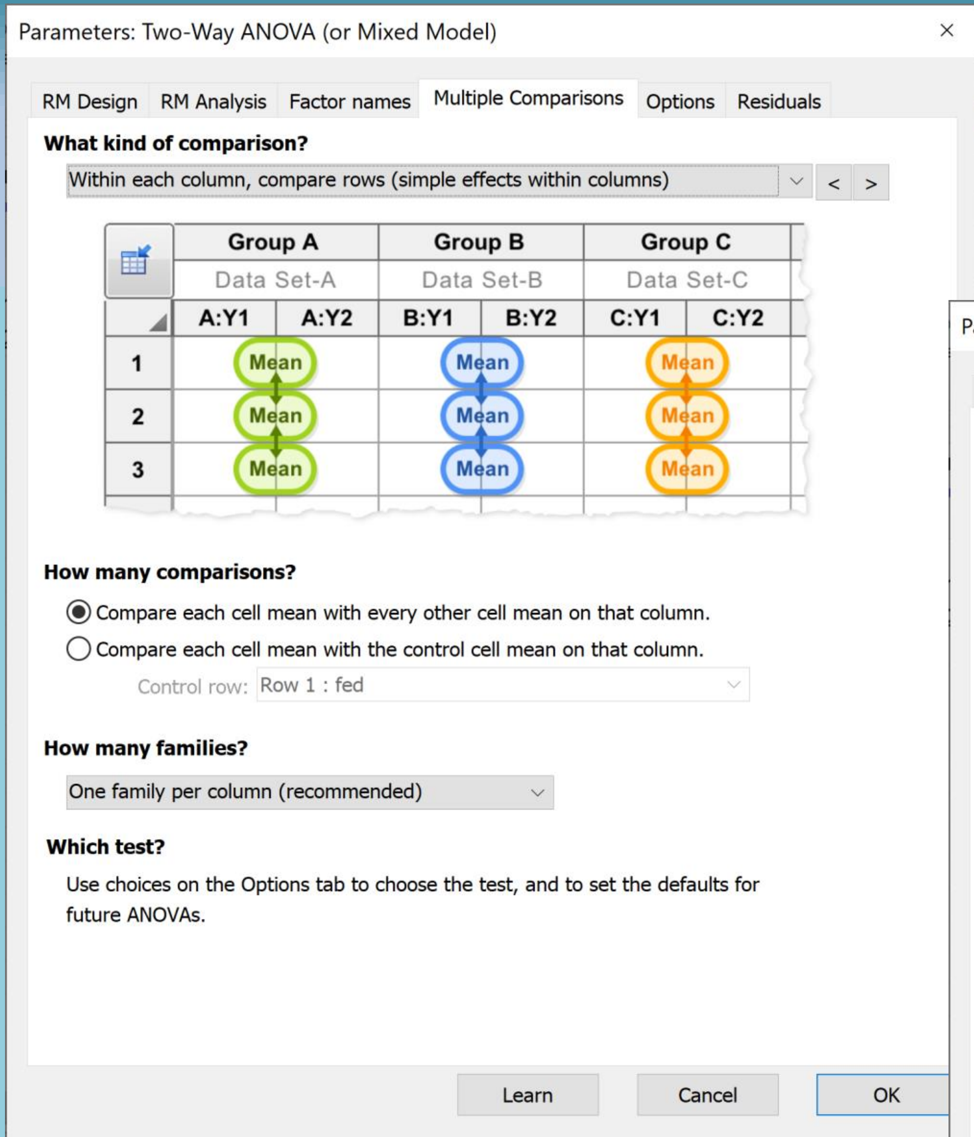
Table format: Grouped	Group A				Group B			
	Males				Females			
	A:1	A:2	A:3	A:4	B:1	B:2	B:3	B:4
1 fed	42.8	43.1	40.0	46.6	42.2	38.7	35.3	40.5
2 fasted 10 days	42.4	42.2	40.8	45.9	42.4	38.1	34.3	40.1
3 fasted 20 days	38.9	40.3	37.5	42.9	39.7	35.8	32.3	37.3

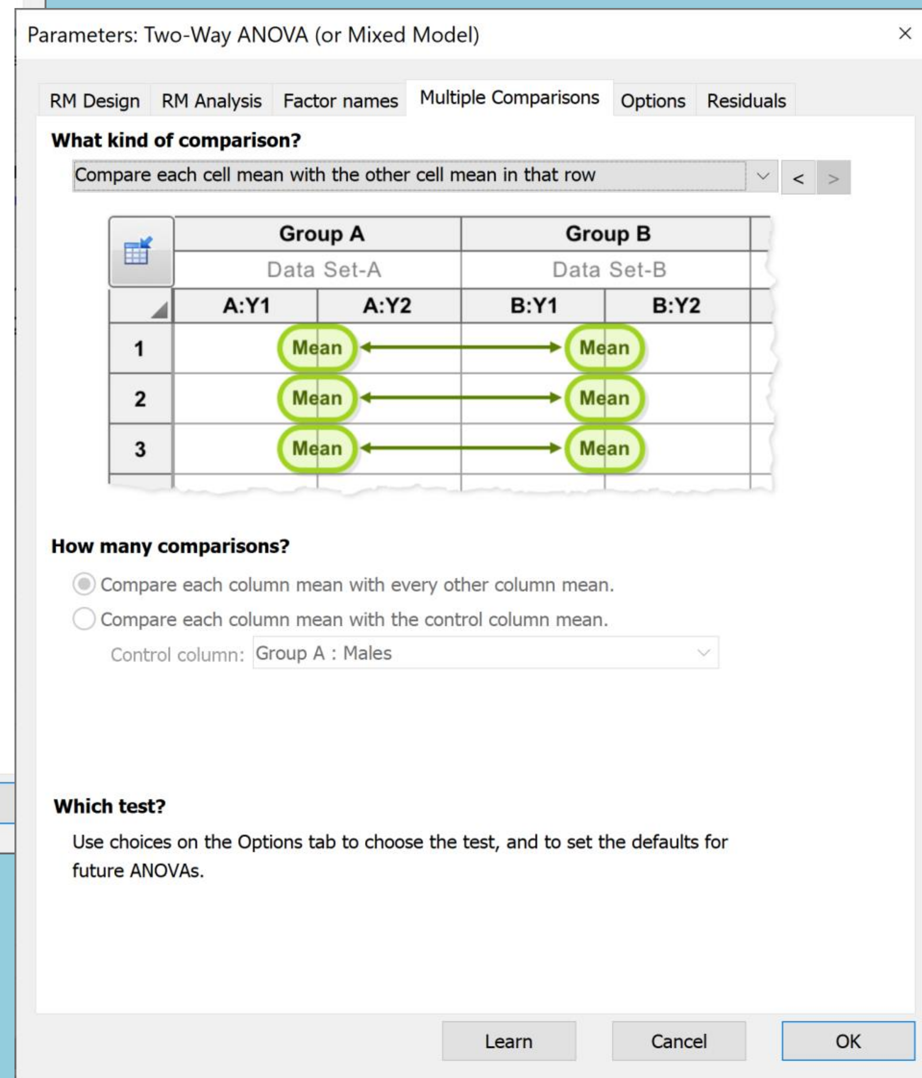


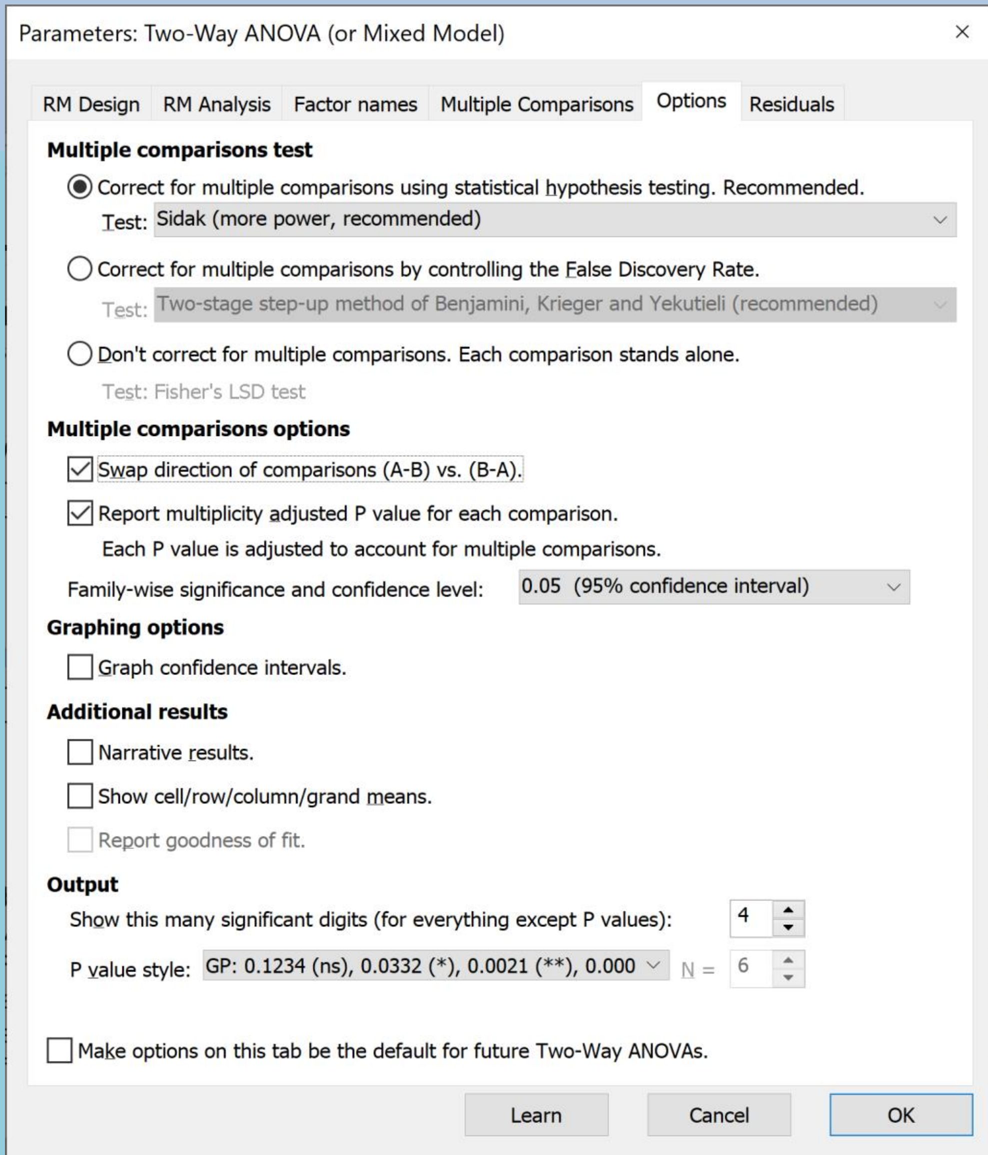


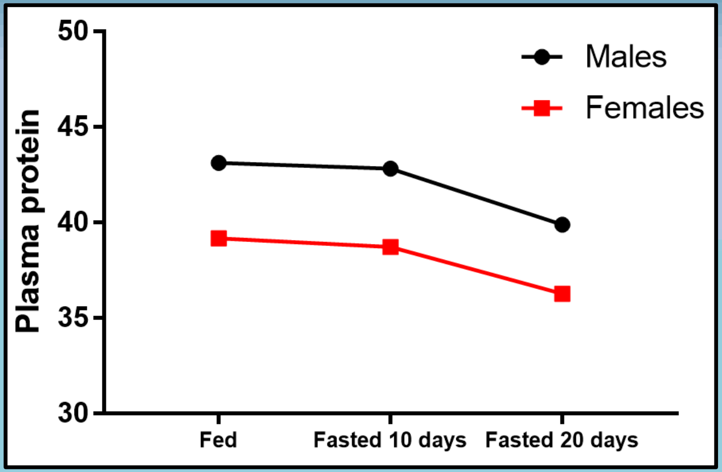


Parameters: Two-Way ANOVA (or Mixed Model)
RM Design
RM Analysis
Factor names
Multiple Comparisons
Options
Residuals
Analyses of repeated measures data can be reported in two ways.
- ANOVA (partition sum-of-squares). This is the same as the general linear model (GLM).
- Mixed-effects model. This uses 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).
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
Title
Title
A:Y1
A:Y2
B:Y1
B:Y2
C:Y1
C:Y2
1 Time1
2 Time2
3 Time3
4 Time4
Factor names
Name the factor that defines the columns:
Sex
Name the factor that defines the rows:
Fasting time
Name of matched set (i.e. person or block):
Turtle
Learn
Cancel
OK


Parameters: Two-Way ANOVA (or Mixed Model)
RM Design
RM Analysis
Factor names
Multiple Comparisons
Options
Residuals
What kind of comparison?
Within each column, compare rows (simple effects within columns)
Group A
Group B
Group C
Data Set-A
Data Set-B
Data Set-C
A:Y1
A:Y2
B:Y1
B:Y2
C:Y1
C:Y2
1
2
3
Mean
How many comparisons?
Compare each cell mean with every other cell mean on that column.
Compare each cell mean with the control cell mean on that column.
Control row: Row 1 : fed
How many families?
One family per column (recommended)
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
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 : Males
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



Evidence of interaction and main effects?

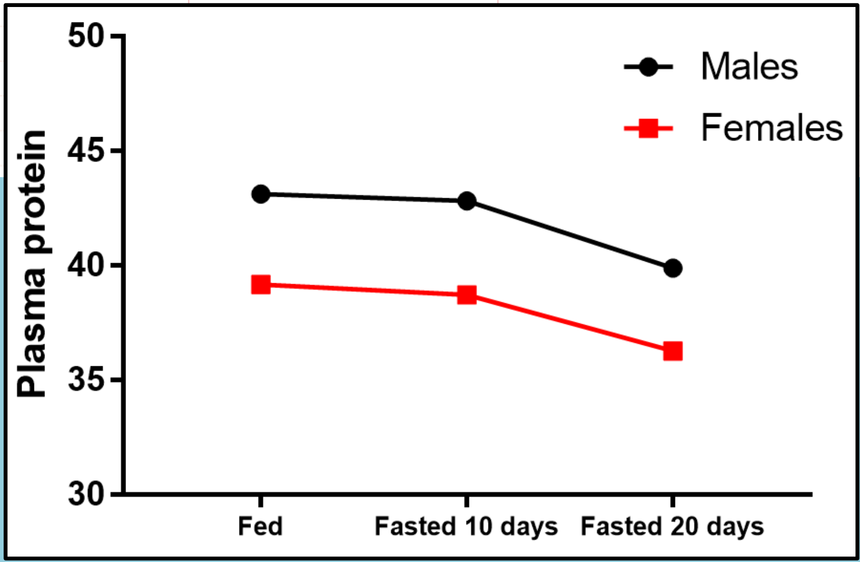
<u>Evidence of main effect by fasting time?</u>
Are there differences between fasting time points, i.e., are values increasing or decreasing over time (combining males and females)?

<u>Evidence of main effect by fasting time?</u>
Do males and females differ at any time points, i.e., is there a main effect where one sex group has higher values compared to the other group?

<u>Evidence of interaction between sex and fasting time?</u>
Do the sex groups differ in their responses over fasting time, i.e., is there a time*group interaction (where one sex group changes differently over time compared to the other)?

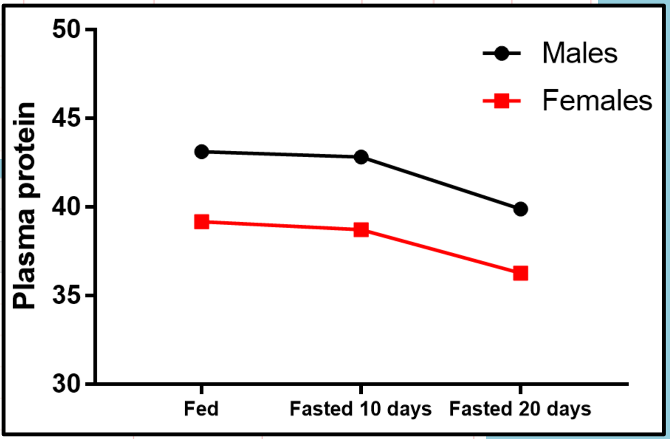
2way ANOVA
ANOVA results

3	**Two-way RM ANOVA**	Matching: Stacked				
4	Assume sphericity?	No				
5	Alpha	0.05				
6						
7	**Source of Variation**	**% of total variation**	**P value**	**P value summary**	**Significant?**	**Geisser-Greenhouse's epsilon**
8	Fasting time x Sex	0.08481	0.5163	ns	No	
9	Fasting time	16.06	<0.0001	****	Yes	0.9438
10	Sex	32.68	0.0962	ns	No	
11	Turtle	50.45	<0.0001	****	Yes	
12						
13	**ANOVA table**	**SS**	**DF**	**MS**	**F (DFn, DFd)**	**P value**
14	Fasting time x Sex	0.2358	2	0.1179	F (2, 12) = 0.6988	P=0.5163
15	Fasting time	44.65	2	22.32	F (1.888, 11.33) = 132.3	P<0.0001
16	Sex	90.87	1	90.87	F (1, 6) = 3.886	P=0.0962
17	Turtle	140.3	6	23.38	F (6, 12) = 138.6	P<0.0001
18	Residual	2.025	12	0.1688		
19						
20	**Difference between column means**					
21	Mean of Males	41.95				
22	Mean of Females	38.06				
23	Difference between means	3.892				
24	SE of difference	1.974				
25	95% CI of difference	-0.9387 to 8.722				



2way ANOVA
Multiple comparisons

1	Compare each cell mean with the other cell mean in that row								
2									
3	Number of families	1							
4	Number of comparisons per family	3							
5	Alpha	0.05							
6									
7	**Sidak's multiple comparisons test**	**Mean Diff.**	**95.00% CI of diff.**	**Significant?**	**Summary**	**Adjusted P Value**			
8									
9	Females - Males								
10	fed	-3.950	-10.52 to 2.622	No	ns	0.2620			
11	fasted 10 days	-4.100	-11.20 to 3.004	No	ns	0.2680			
12	fasted 20 days	-3.625	-10.13 to 2.885	No	ns	0.3035			
13									
14									
15	**Test details**	**Mean 1**	**Mean 2**	**Mean Diff.**	**SE of diff.**	**N1**	**N2**	**t**	**DF**
16									
17	Females - Males								
18	fed	39.18	43.13	-3.950					
19	fasted 10 days	38.73	42.83	-4.100					
20	fasted 20 days	36.28	39.90	-3.625					



2way ANOVA
Multiple comparisons

1	Within each column, compare rows (simple effects within columns)					
2						
3	Number of families	2				
4	Number of comparisons per family	3				
5	Alpha	0.05				
6						
7	**Tukey's multiple comparisons test**	**Mean Diff.**	**95.00% CI of diff.**	**Significant?**	**Summary**	**Adjusted P Value**
8						
9	Males					
10	fed vs. fasted 10 days	0.3000	-1.291 to 1.891	No	ns	0.7351
11	fed vs. fasted 20 days	3.225	1.804 to 4.646	Yes	**	0.0051
12	fasted 10 days vs. fasted 20 days	2.925	1.434 to 4.416	Yes	**	0.0078
13						
14	Females					
15	fed vs. fasted 10 days	0.4500	-0.5947 to 1.495	No	ns	0.3096
16	fed vs. fasted 20 days	2.900	2.285 to 3.515	Yes	***	0.0006
17	fasted 10 days vs. fasted 20 days	2.450	1.678 to 3.222	Yes	**	0.0019

How you might write this up

<you checked all assumptions and you conclude that you can use the two-way RM ANOVA>

We used a two-way RM ANOVA to study differences in plasma protein by time fasting and sex of turtle. Interaction between gender and time fasting was not significant (we fail to reject the null hypothesis of no interaction (two sided hypothesis, $F(2,12)=0.70$, $p=0.53$, $\alpha=0.05$). Gender was also not significant ($F(2,12=3.886$, $p=0.10$). With $p<0.0001$ (two-sided hypothesis, $F(2,12)=132.3$, $\alpha=0.05$), we reject the null hypothesis that mean plasma protein levels are the same across groups of time fasting and conclude that at least one is different. Tukey's post hoc tests showed that 'Fasted 20' (mean=38.1) was significantly different from 'Fed' (41.2) and 'Fasted 10' (40.8). There was no significant difference between 'Fed' and 'Fasted 10'. We conclude that plasma protein levels decrease with fasting and that gender has no effect on this finding.

Note: because gender was not significant, report the results for time fasting for males and females combined.

Using a one-way ANOVA when a RM ANOVA would be better

OPEN ACCESS Freely available online PLoS One 2014 Apr 23;9(4):e95964 PLOS ONE

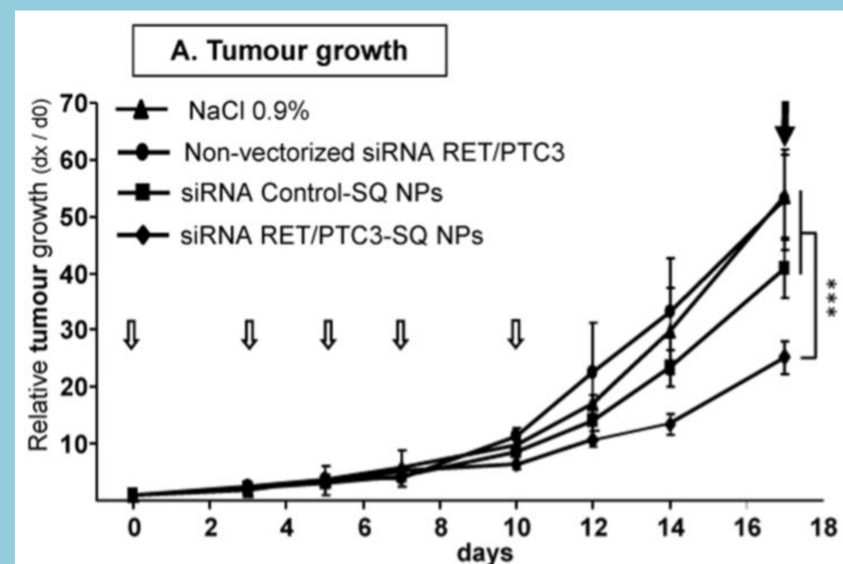
Effects of siRNA on RET/PTC3 Junction Oncogene in Papillary Thyroid Carcinoma: From Molecular and Cellular Studies to Preclinical Investigations

Hafiz Muhammad Ali[1,2,3], Giorgia Urbinati[1,2,3], Hubert Chapuis[4], Didier DesmaEle[4], Jean-Rémi Bertrand[1,2,3], Patrick Couvreur[4], Liliane Massaad-Massade[1,2,3]*

1 Université Paris-Sud 11, Laboratoire de Vectorologie et Thérapeutiques Anticancéreuses, UMR 8203, Villejuif, France, **2** CNRS, Villejuif, Laboratoire de Vectorologie et Thérapeutiques Anticancéreuses, UMR 8203, Villejuif, France, **3** Gustave Roussy, Laboratoire de Vectorologie et Thérapeutiques Anticancéreuses, UMR 8203, Villejuif, France, **4** Institut Galien, UMR CNRS 8612, Université Paris-Sud 11, Faculté de pharmacie, Châtenay-Malabry, France

Statistical analysis

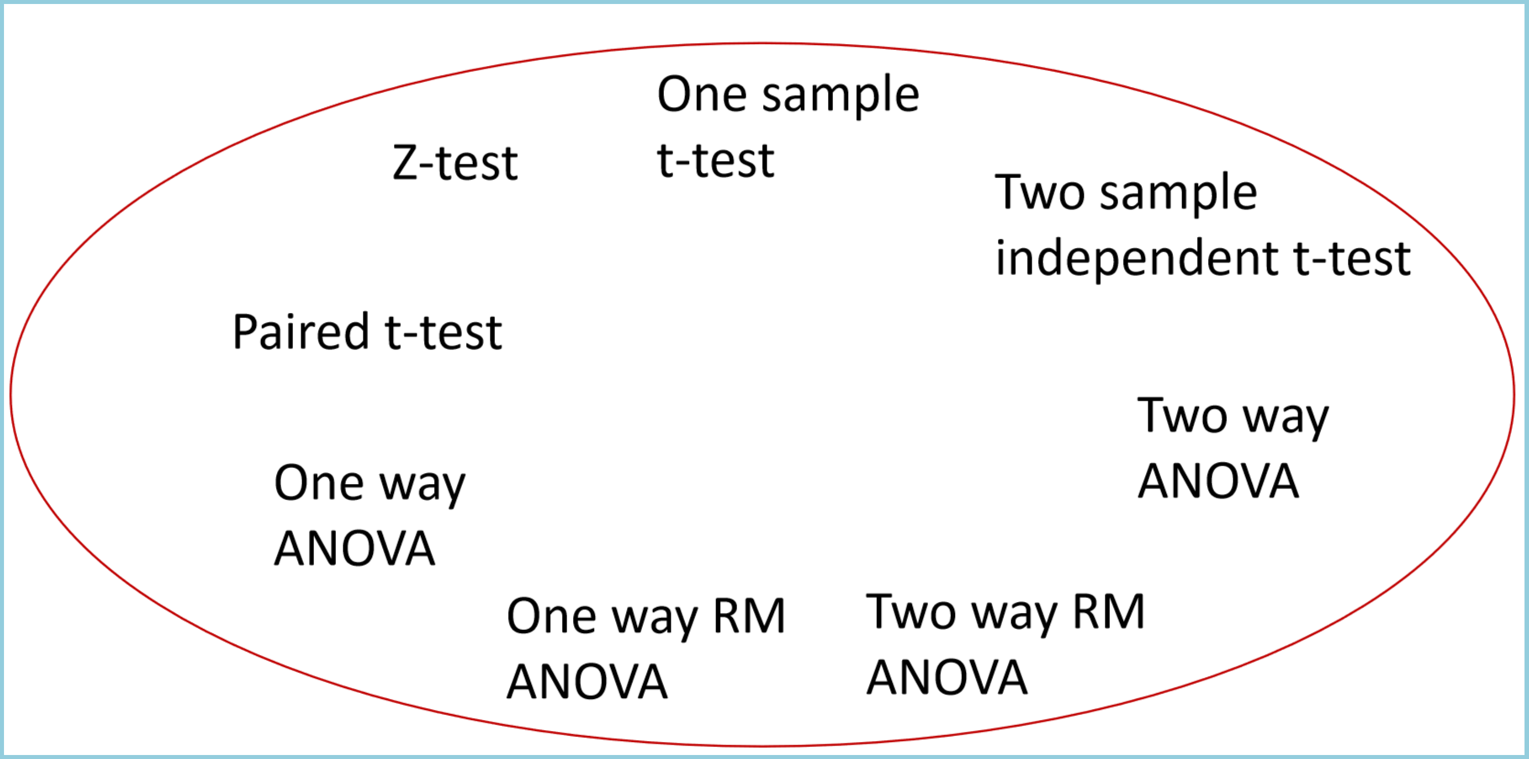
All data are presented as mean ± SD (Standard deviation). One-way analysis of variance (ANOVA) was employed to compare multiple treatments. Using a linear regression model, doubling time was calculated by GraphPad Prism 4 software. While all pair wise comparisons between different treatment groups were done by least significant difference (LSD) Post-hoc test by using "InVivoStat" software. $p<0.05$ was considered as statistically significant level.



To compare difference between treatments, a statistical analysis was performed by using ANOVA followed by LSD Post-hoc test (*** = $p<0.001$)

Applied statistics so far: Which test should you use?

You have learned many statistical tools. Given a real problem, can you decide which one to use?



Example 1

A principal at a certain school claims that the students in his school are above average intelligence. A random sample of thirty students IQ scores have a mean score of 112. Is there sufficient evidence to support the principal's claim? It is known that the mean population IQ is 100 with a standard deviation of 15.

Example 2

A study was designed to compare the effect of two soporific drugs. 100 volunteers were chosen and they were randomly assigned either drug A or drug B. Number of hours of increased sleep was recorded for each patient. Use the data to test if Drug A is more effective than Drug B at increasing sleep.

Example 3

A sample of 50 students were given a diagnostic test before studying a particular module and then again after completing the module. Analyze the data to test if studying the module leads to improvements in students' knowledge/skills (i.e. test scores).

Example 4

A study was designed to combine the effect of two soporific drugs. 150 volunteers were chosen and they were randomly assigned either drug A, drug B or drug A+B. Each drug was administered to 50 participants. Number of hours of increased sleep was recorded for each patient. Use the data to test if there is any difference among effectiveness of the drugs to increase sleep.

Example 5

A researcher was interested in whether an individual's interest in politics was influenced by their level of education and gender. In particular, the researcher wanted to know if the effect of level of education on interest in politics is different for males and females? To answer this question, a random sample of 60 participants were recruited to take part in the study – 30 males and 30 females – equally split by level of education: school, college and university (i.e., 10 participants in each group). Each participant in the study completed a questionnaire that scored their interest in politics on a scale of 0 to 100, with higher scores indicating a greater interest in politics.

How can these be used to answer the research questions?

Example 6

This is a study to test the effect of a specific diet program proposed to reduce weight. Ten people in need of losing weight agree to try the diet for two months. Their weight was measured in Kilograms at the start of the diet and then after 1 month and 2 months. Did the diet work?

Before Diet	After 1 Month	After 2 Months
63.75	65.38	81.34
62.98	66.24	69.31
65.98	67.70	77.89
107.27	102.72	91.33
66.58	69.45	72.87
120.46	119.96	114.26
62.01	66.09	68.01
71.87	73.62	55.43
83.01	75.81	71.63
76.62	67.66	68.60

Example 7

Severe atheroma of the aortic arch has now been established as an important risk factor and mechanism for stroke. Two drugs (A) a statin (B) an anticoagulant were given to stroke patients and followed up every 6 months for 3 years. At every follow up, the thickness of plaque was measured. The scientific questions are:

(i) Does the thickness of the plaque change over time post medication?

(ii) Is the any difference between the change in the thickness of plaque between patients who received drug A and patients who received drug B?