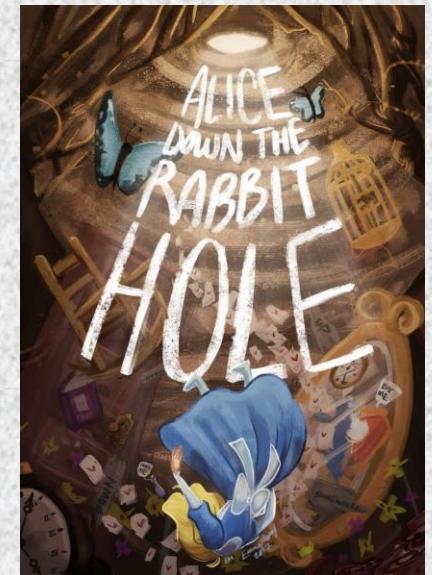


Comparing paired groups of data:  
paired t-tests, Wilcoxon signed rank, and McNemar's test

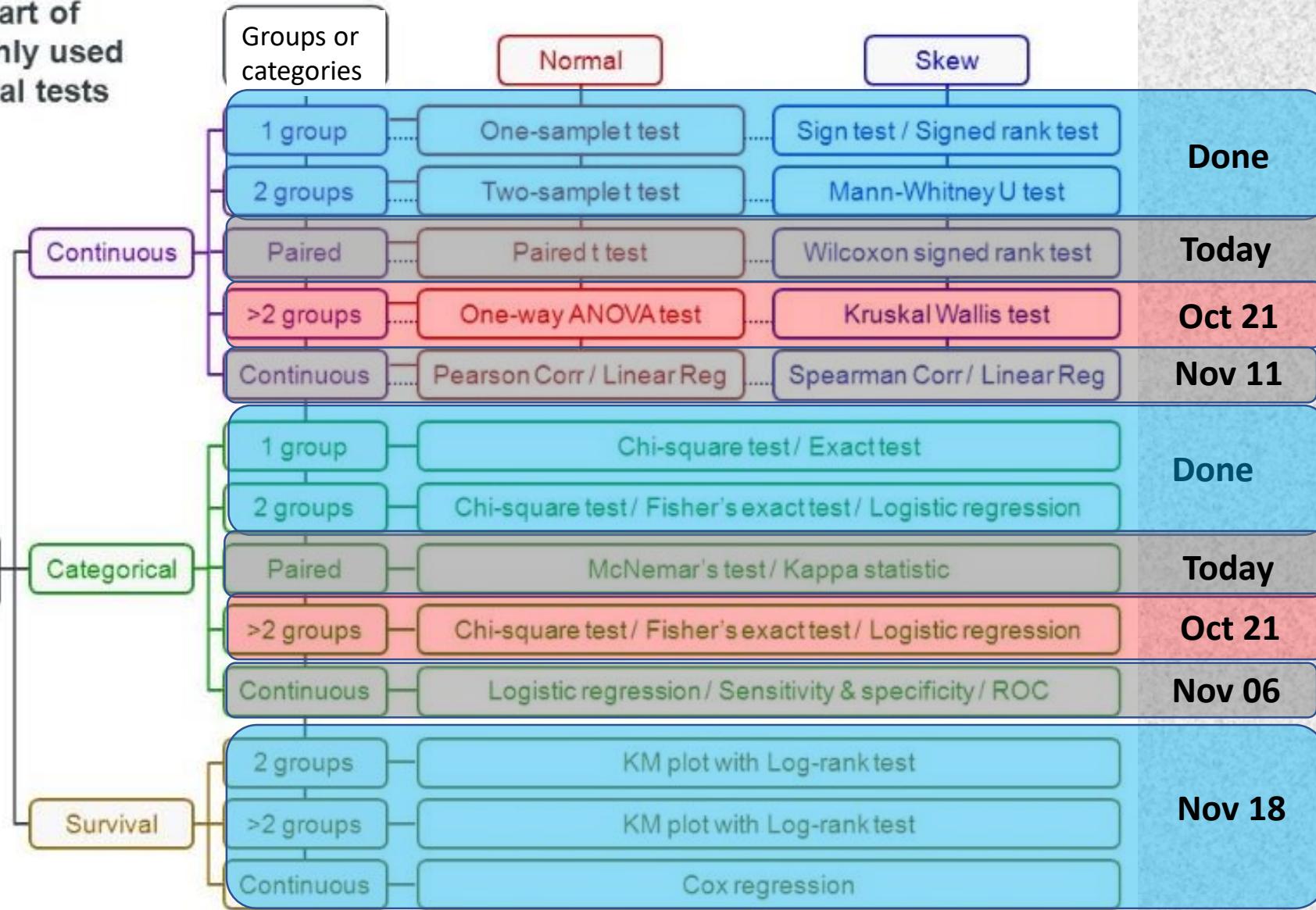
Plus an example that takes us down the rabbit hole...

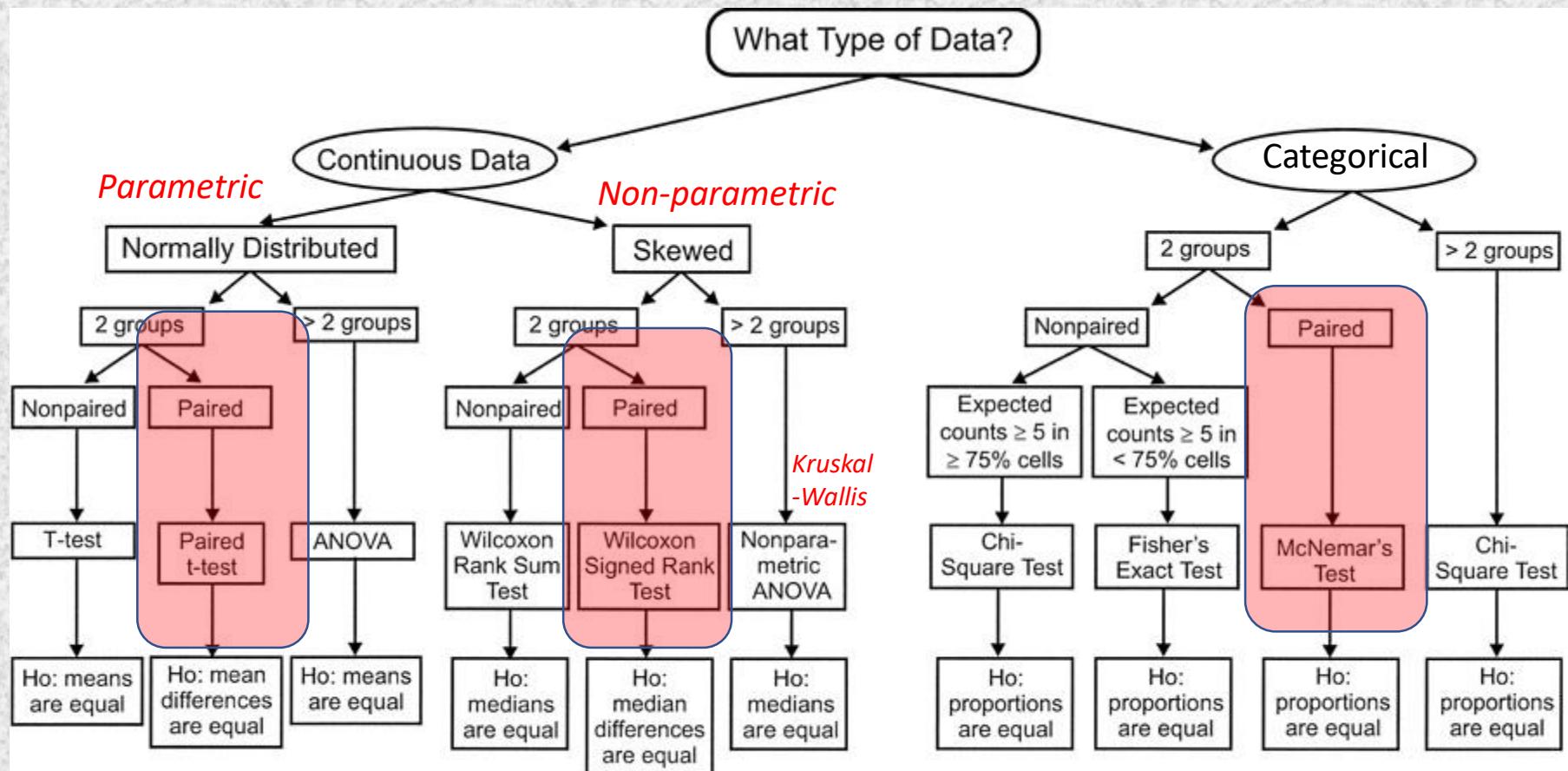
Kathleen Torkko  
October 14, 2019



[www.ellsworthchamber.org](http://www.ellsworthchamber.org)

# Flow chart of commonly used statistical tests





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

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

# The Family of t-tests

t-tests are used to compare ***two*** means

## Independent t-test

data collected from two independent groups

Prism calls these “unpaired” t-tests

## Paired t-test

data are correlated

e.g., same mouse studied at two different times or under  
two conditions

## One-sample t-test

Compare sample mean to a known or standard mean

# The t Statistic

## Independent t-test

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}$$

## Paired t-test

$$t = \frac{\bar{d}}{s_d / \sqrt{n}}$$

## One sample t-test

$$t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

$\bar{x}$  = sample means,  $\mu_0$  = population mean,  $\bar{d}$  = mean of difference between paired means,  $s$  = sample SD,  $s_d$  = SD of the difference,  $n$  = sample size

# Paired t-test Assumptions

Distribution of the *differences* is **normal**

Plot the paired differences

Check mean, median for the difference

Look at skewness of the difference

Normality tests where appropriate

Paired t-test

$$t = \frac{\bar{d}}{S_d / \sqrt{n}}$$

Data are **paired**

two measurements of the same variable in the same person or mouse

Each of the individual differences is **independent**

# Independent or paired t-test? Examples of Paired and Independent Data

**Paired**: Blood pressure measurements in a person before and after exercise

**Independent**: Blood pressure measurements in a group of people at rest and in another group who just completed a workout

**Paired**: Tumor size in a mouse measured before and after treatment

**Independent**: Tumor size measurements in a group of mice on placebo and in another group which received the treatment

**Paired**: Cholesterol measurements in the same patients done last year and now  
*(an example of not paired: cholesterol measurements and foot size in the same person)*

**Independent**: Cholesterol measurement in two groups of patients

**Paired\***: Measurement of expression levels of p53 in tumor and adjacent normal tissue from the same tissue block

**Independent**: Measurement of expression levels of p53 in tumor tissue from a person without cancer and in tumor tissue from a person with cancer

\*many researchers in pathology treat these as independent.

## Comparison of Hours Worked in 1988 and 2014: Paired or Unpaired with Hull Hypotheses

If the same people reported their hours for 1988 and 2014

Paired null hypothesis: The mean of the paired differences = 0

$$H_0 : \mu_d = 0$$

If different people are used in 1988 and 2014.

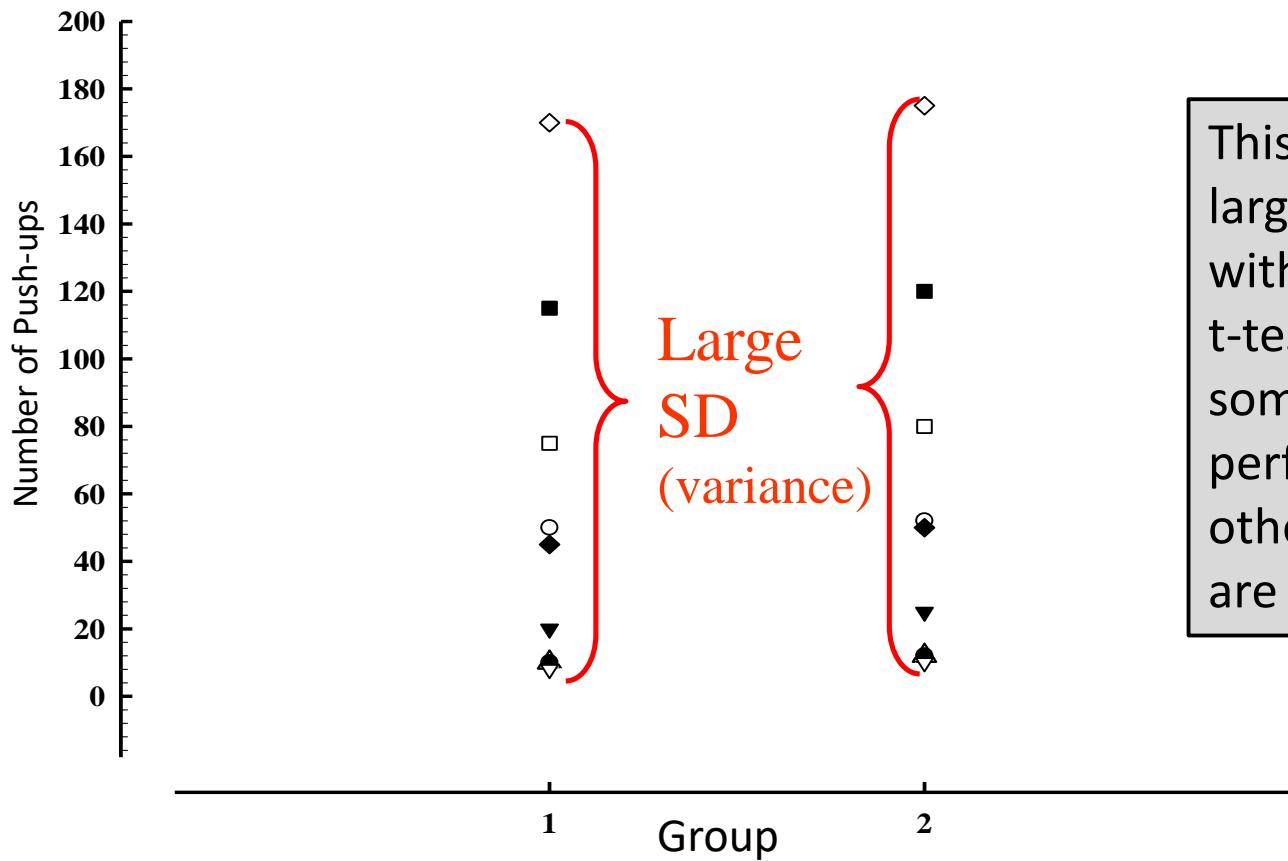
Independent null hypothesis: The mean hours worked in 1988 is equal to the mean for 2014

$$H_0 : \mu_{1988} = \mu_{2014}$$

It is important to remember that hypotheses are never about data, they are about the processes which produce the data. In the formulas above, the value of  $\mu_d$  is unknown. The goal of hypothesis testing is to determine the hypothesis (null or alternative) with which the data are more consistent.

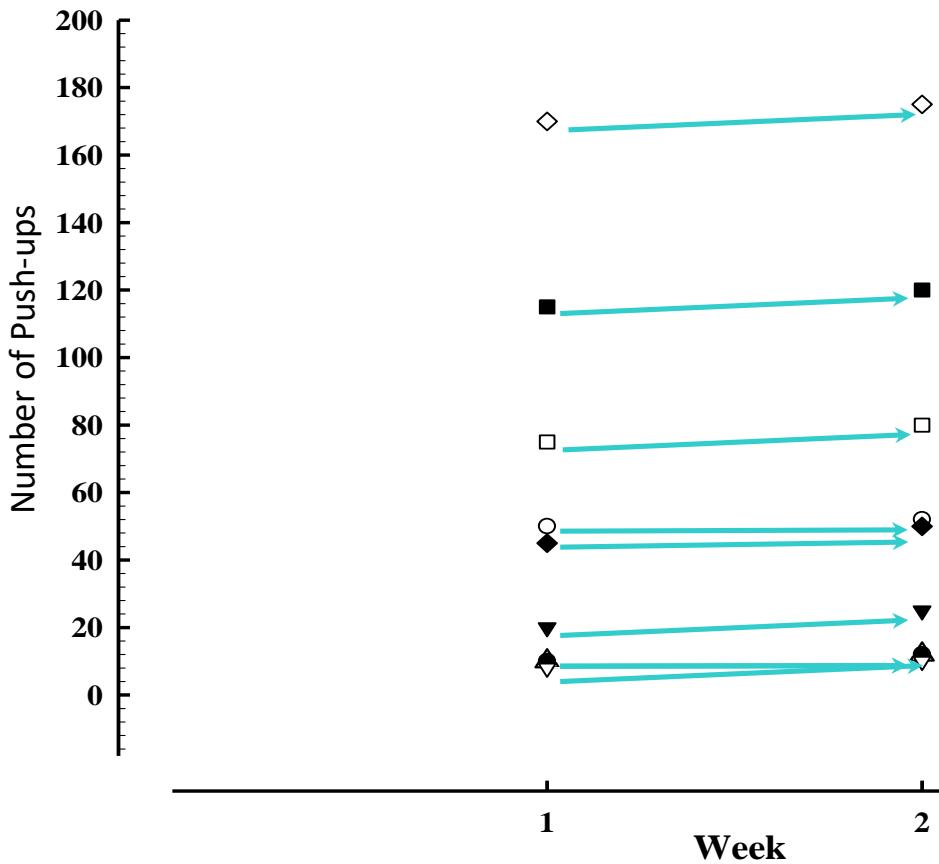
## “Disadvantages” of Independent Data

Data from independent samples are heavily influenced by variation (variance) between *individual* subjects



This data would have a large SD associated with an independent t-test simply because some subjects performed better than others and the data are spread out

# Advantages of Paired Data



The variance **between** individual subjects is removed from the analysis

Each subject serves as its own control

Variability within individuals is often less than between individuals

## A paired t-test can have more power to detect differences, with the right data

Green fluorescent protein (GFP) levels were measured in two distinct cells (left and right neurons) within a single worm. Fluorescence data were collected from 14 wild-type worms and 14 *b* gene inhibited [*b(RNAi)*] worms.

dsRNA will be expressed only in the rightmost cell of one particular neuronal pair, where it is expected to inhibit the expression of gene *b* via the RNAi response. The neuron on the left side will be unaffected. WT cells on both sides will be unaffected.

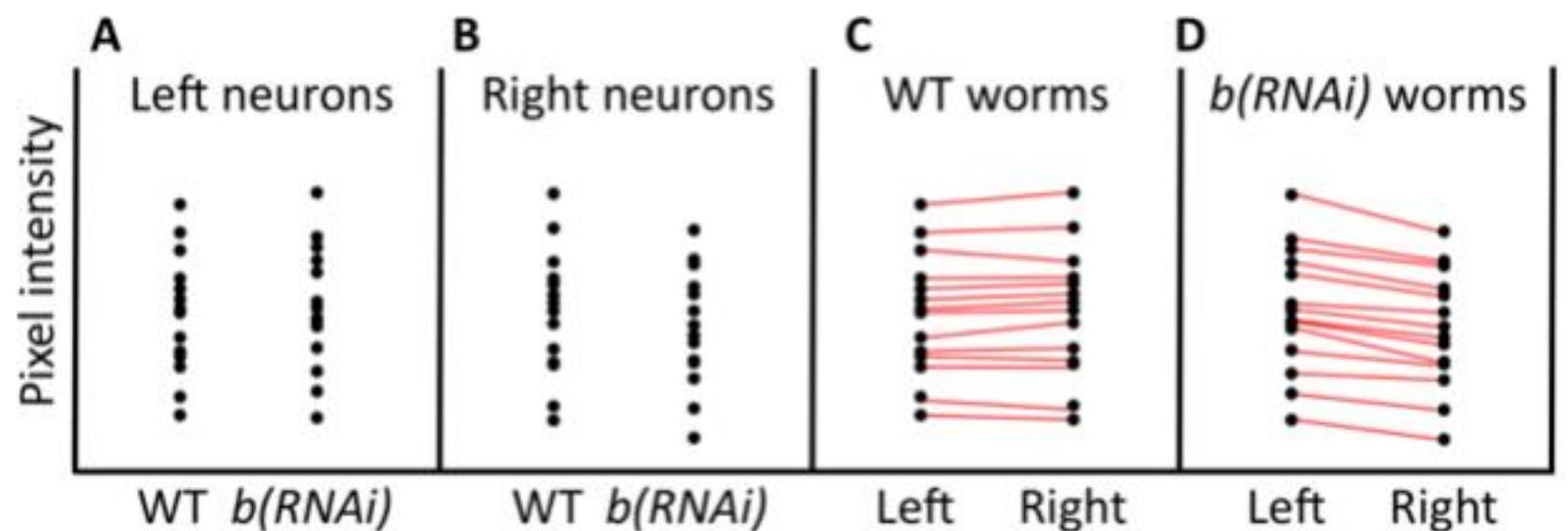
Hypothesis: *b(RNAi)* worms will show reduced fluorescence in the right cell only. WT cells will be unaffected.

For the paired t-test to be valid, data points must be linked in a meaningful way

Assumptions: The measurements are **paired** within each worm, and each worm is **independent** of the other worms

## Independent analysis

## Paired analysis



GFP expression in some worms will happen to be weaker or stronger (resulting in a dimmer or brighter signal) than in other worms.

With the data variability, along with a relatively small mean difference in expression (panel B),  $p$  may be  $>0.05$

In fact for B, two-tail independent t-test  $p>0.05$ .

Worms that are bright in one cell tend to be bright in the other

Panel D, there is a strong tendency to have reduced expression in the right neuron as compared with the left

For D, a two-tailed paired t-test has a  $p=0.01$

## Performing a paired t-test

After meeting the assumptions of paired data and independence of samples:

1. Calculate the difference between each individual pair.  
Test if the differences meet the assumption of normality
2. Calculate the mean and variance for all the differences among the pairs.
3. Finally, perform a one-sample *t*-test where the null hypothesis is the mean difference of the population = zero

Paired t-test

$$H_0 : \mu_d = 0$$
$$H_A : \mu_d \neq 0$$
$$t = \frac{\bar{d}}{S_d / \sqrt{n}}$$

One sample t-test

$$t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

## Paired t-test

$$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}}$$

$$s_d = \sqrt{\frac{\sum_{i=1}^n (d_i - \bar{d})^2}{n-1}}$$

$\bar{d}$  = mean of difference between paired means,  $s_d$  = SD of the difference,  $n$  = sample size (number of pairs)

$$d = x_1 - x_2 \quad \text{Where } x_1 \text{ and } x_2 \text{ are from the same individual}$$

N = number of independent pairs

df = number of pairs – 1

$H_0$ : mean difference of *the population* is zero

$H_A$ : mean difference of *the population* is not zero

From our unpaired t-test analysis

	Group A placebo	Group B treatment
1	18	22
2	21	25
3	16	17
4	22	24
5	19	16
6	24	29
7	17	20
8	21	23
9	23	19
10	18	20
11	14	15
12	16	15
13	16	18
14	19	26
15	18	18
16	20	24
17	12	18
18	22	25
19	15	19
20	17	16

	A	B	C	D
1	Patient	Before Treatment	After Treatment	Difference
2	1	18	22	4
3	2	21	25	4
4	3	16	17	1
5	4	22	24	2
6	5	19	16	-3
7	6	24	29	5
8	7	17	20	3
9	8	21	23	2
10	9	23	19	-4
11	10	18	20	2
12	11	14	15	1
13	12	16	15	-1
14	13	16	18	2
15	14	19	26	7
16	15	18	18	0

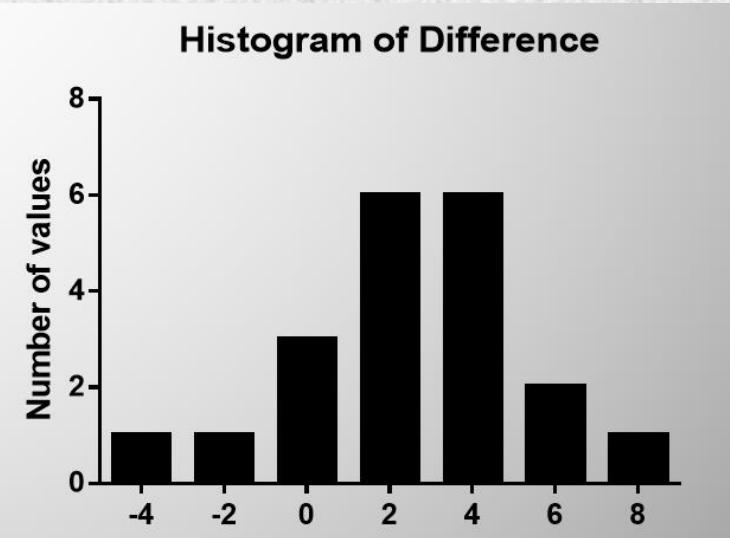
Two methods to do a paired t-test in Prism:

Input 2 columns (Before Rx and After Rx) and choose “paired” in the analysis

Calculate the difference and use a one-sample t-test where the hypothetical value is “0”

# Paired t-test: Normality of differences and independence

		Y
1	Number of values	20
2		
3	Minimum	-4
4	25% Percentile	0.25
5	Median	2
6	75% Percentile	4
7	Maximum	7
8		
9	Mean	2.05
10	Std. Deviation	2.837
11	Std. Error of Mean	0.6344
12		
13	Skewness	-0.4556
14	Kurtosis	-0.01401
15		
16	D'Agostino & Pearson normality test	
17	K2	0.9115
18	P value	0.6340
19	Passed normality test (alpha=0.05)?	Yes
20	P value summary	ns
21		



The paired data are in different individuals

## New Data Table and Graph

## New table &amp; graph

XY

Column

Grouped

Contingency

Survival

Parts of whole

Multiple variables

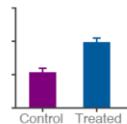
Nested

## Existing file

Clone a graph

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

	A	B
	Control	Treated
Y		
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: 

	Group A	Group B	Group C	G
	Before Treatment	After Treatment	Difference	
1		18	22	4
2		21	25	4
3		16	17	1
4		22	24	2
5		19	16	-3
6		24	29	5
7		17	20	3
8		21	23	2
9		23	19	-4
10		18	20	2
11		14	15	1
12		16	15	-1
13		16	18	2
14		19	26	7
15		18	18	0
16		20	24	4
17		12	18	6
18		22	25	3
19		15	19	4
20		17	16	-1
21				

Prism Tips

Cancel

Create

## 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: Before Treatment

B: After Treatment

C: Difference

Parameters: t tests (and Nonparametric Tests)

Experimental Design Residuals Options

**Experimental design**

Unpaired

Paired

	A	B
	Control	Treated
1	Y	Y
2	Y	Y
3	Y	Y
4	Y	Y
5	Y	Y

Assume Gaussian distribution?

Yes. Use parametric test.

No. Use nonparametric test.

**Choose test**

Paired t test (differences between paired values are consistent)

Ratio paired t test (ratios of paired values are consistent)

Select All

Deselect All

Help

Cancel

Learn

Cancel

OK

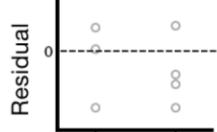
Parameters: t tests (and Nonparametric Tests)

Experimental Design Residuals Options

**What graphs to create?**

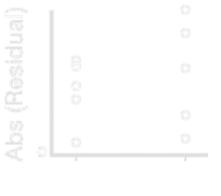
Residual plot

*Correct model?*



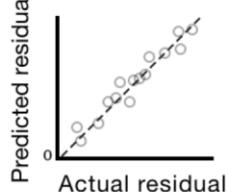
Homoscedasticity plot

*Equal variance?*



QQ plot

*Normality?*



Heatmap plot



**Diagnostics for residuals**

Are the residuals Gaussian?

Normality tests of Anderson-Darling, D'Agostino, Shapiro-Wilk and Kolmogorov-Smirnov.

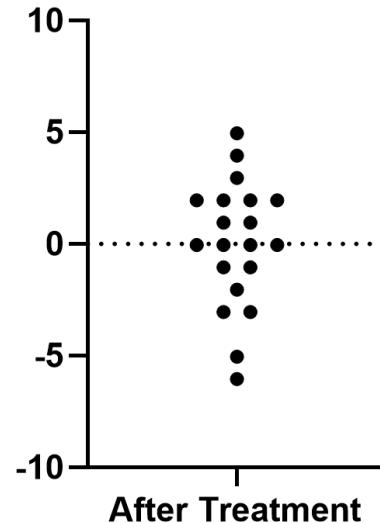
Make options on this tab be the default for future tests.

Learn

Cancel

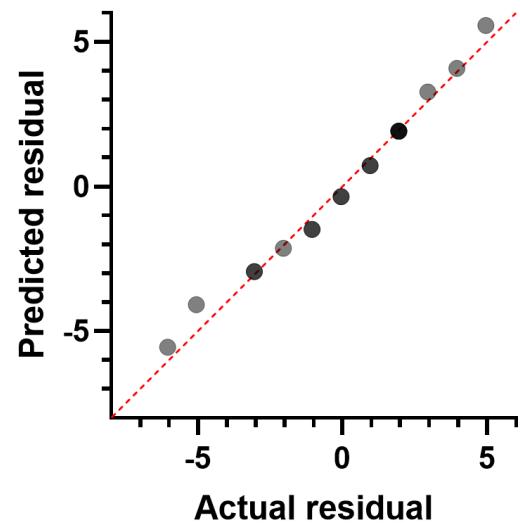
OK

## Residual plot



Before Treatment  
After Treatment

## QQ plot



Parameters: t tests (and Nonparametric Tests)

**Experimental Design** **Residuals** **Options**

**Calculations**

P value:  One-tailed  Two-tailed (recommended)

Report differences as: After Treatment - Before Treatment

Confidence level: 95%

Definition of statistical significance: P < 0.05

**Graphing options**

Graph differences (paired)  
 Graph ranks (nonparametric)  
 Graph correlation (paired)  
 Graph CI of mean of differences

**Additional results**

Descriptive statistics for each data set  
 t test: Also compare models using AICc  
 Wilcoxon. Also compute the CI of the median paired differences  
 Assumes the distribution of paired differences is symmetrical.  
 Wilcoxon: When both values on a row are identical, use method of Pratt  
 If this option is unchecked, those rows are ignored and the results will match prior version of Prism

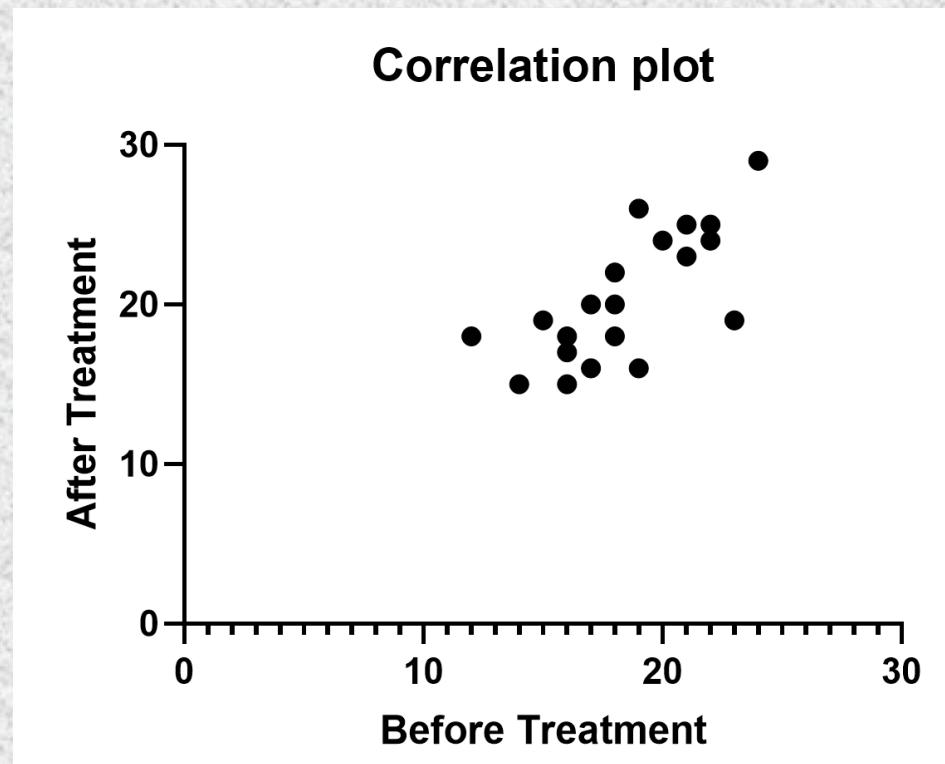
**Output**

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 tests.

**Learn** **Cancel** **OK**



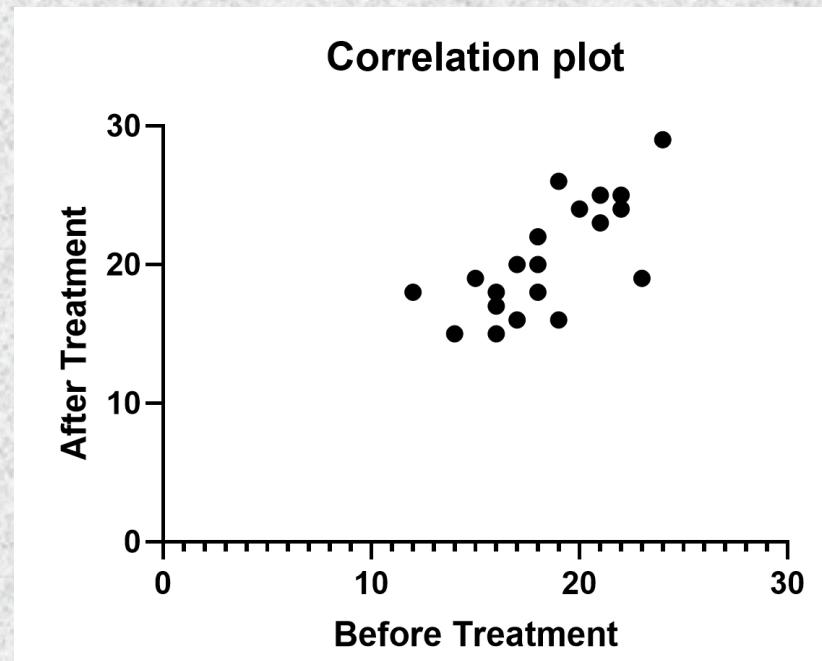
## How effective was the pairing?

Correlation coefficient (r)	0.7175
P value (one tailed)	0.0002
P value summary	***
Was the pairing significantly effective? Yes	

If pairing is effective, you expect the before and after measurements to vary together.

Prism calculates the Pearson correlation coefficient, r. If the pairing was effective, r will be positive and the P value will be small. This means that the two groups are significantly correlated, so it made sense to choose a paired test.

Your choice of whether to use a paired test or not should not be based on this one P value,



Paired t test		
Tabular results		
3	Column B	
4	vs.	
5	Column A	
6		
7	<b>Paired t test</b>	
8	P value	0.0044
9	P value summary	**
10	Significantly different ( $P < 0.05$ )?	Yes
11	One- or two-tailed P value?	Two-tailed
12	t, df	t=3.231, df=19
13	Number of pairs	20
14		
15	<b>How big is the difference?</b>	
16	Mean of differences (B - A)	2.050
17	SD of differences	2.837
18	SEM of differences	0.6344
19	95% confidence interval	0.7221 to 3.378
20	R squared (partial eta squared)	0.3546
21		
22	<b>How effective was the pairing?</b>	
23	Correlation coefficient (r)	0.7175
24	P value (one tailed)	0.0002
25	P value summary	***
26	Was the pairing significantly effective? Yes	

## Write up results

### For this class:

We tested our hypothesis that ..... The data passed the assumptions for a paired t-test (data are paired, pairs are independent, the differences are normal/symmetrical). With  $p=0.004$  (two tailed,  $t=8.031$ ,  $df=45$ ,  $\alpha=0.05$ ), we reject the null hypothesis that the population mean difference of the test before and after treatment is zero and conclude that mean test levels increased (a mean difference of 2.05).

### The way you may read it:

We found that test levels increased after treatment an average of 2.05 in our paired study ( $p=0.004$ )

## 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

X

Analyze which data sets?

- A:Before Treatment
- B:After Treatment
- C:Difference

Parameters: One sample t and Wilcoxon test

Experimental Design Options

#### Choose test

##### One sample t test

Compare the mean of your sample with a hypothetical mean.  
Assumes sampling from a Gaussian distribution.

##### Wilcoxon signed-rank test

Compare the median of your sample with a hypothetical median.  
Nonparametric.

Calculate CI of the discrepancy

#### Hypothetical value

Hypothetical value. Often 0.0, 1.0 or 100:

For the Wilcoxon test, if a value in data set matches the hypothetical value:

Ignore that value entirely

Include that value using method of Pratt (not commonly used)

Learn

Cancel

OK

		A
	Difference	
1	Theoretical mean	0.000
2	Actual mean	2.050
3	Number of values	20
4		
5	<b>One sample t test</b>	
6	t, df	t=3.231, df=19
7	P value (two tailed)	0.0044
8	P value summary	**
9	Significant (alpha=0.05)?	Yes
10		
11	<b>How big is the discrepancy?</b>	
12	Discrepancy	2.050
13	SD of discrepancy	2.837
14	SEM of discrepancy	0.6344
15	95% confidence interval	0.7221 to 3.378
16	R squared (partial eta squared)	0.3546

	Paired t test Tabular results	
	Column B	After Treatment
	vs.	vs.
	Column A	Before Treatment
3		
4		
5		
6		
7	<b>Paired t test</b>	
8	P value	0.0044
9	P value summary	**
10	Significantly different (P < 0.05)?	Yes
11	One- or two-tailed P value?	Two-tailed
12	t, df	t=3.231, df=19
13	Number of pairs	20
14		
15	<b>How big is the difference?</b>	
16	Mean of differences (B - A)	2.050
17	SD of differences	2.837
18	SEM of differences	0.6344
19	95% confidence interval	0.7221 to 3.378
20	R squared (partial eta squared)	0.3546
21		
22	<b>How effective was the pairing?</b>	
23	Correlation coefficient (r)	0.7175
24	P value (one tailed)	0.0002
25	P value summary	***
26	Was the pairing significantly effective? Yes	

Paired t test Tabular results	
3 Column B	After Treatment
4 vs.	vs.
5 Column A	Before Treatment
6	
7 <b>Paired t test</b>	
8 P value	0.0044
9 P value summary	**
10 Significantly different ( $P < 0.05$ )?	Yes
11 One- or two-tailed P value?	Two-tailed
12 t, df	t=3.231, df=19
13 Number of pairs	20
14	
15 <b>How big is the difference?</b>	
16 Mean of differences (B - A)	2.050
17 SD of differences	2.837
18 SEM of differences	0.6344
19 95% confidence interval	0.7221 to 3.378
20 R squared (partial eta squared)	0.3546
21	
22 <b>How effective was the pairing?</b>	
23 Correlation coefficient (r)	0.7175
24 P value (one tailed)	0.0002
25 P value summary	***
26 Was the pairing significantly effective? Yes	

Unpaired t test Tabular results	
3 Column B	treatment
4 vs.	vs.
5 Column A	placebo
6	
7 <b>Unpaired t test</b>	
8 P value	0.0824
9 P value summary	ns
10 Significantly different ( $P < 0.05$ )?	No
11 One- or two-tailed P value?	Two-tailed
12 t, df	t=1.784, df=38
13	
14 <b>How big is the difference?</b>	
15 Mean of column A	18.40
16 Mean of column B	20.45
17 Difference between means (B - A) $\pm$ SEM	2.050 $\pm$ 1.149
18 95% confidence interval	-0.2762 to 4.376
19 R squared (eta squared)	0.07728
20	
21 <b>F test to compare variances</b>	
22 F, DFn, Dfd	1.658, 19, 19
23 P value	0.2795
24 P value summary	ns
25 Significantly different ( $P < 0.05$ )?	No
26	

The SD is smaller for the difference

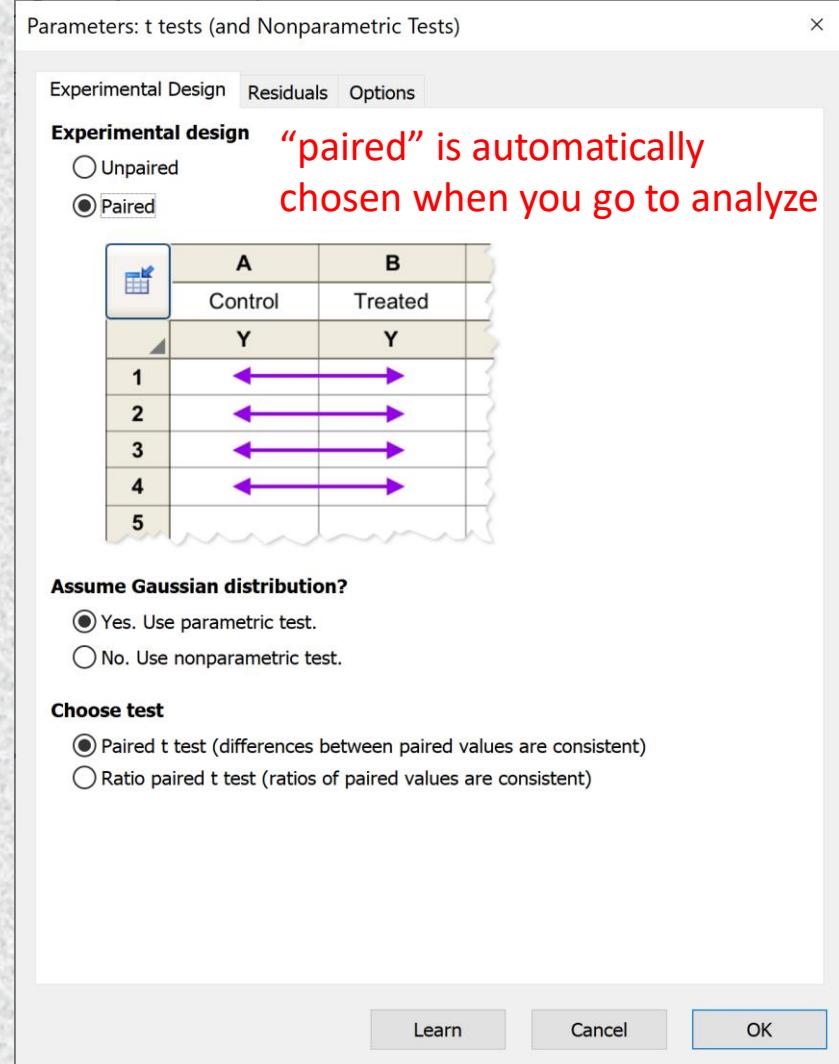
	Col. stats	A Difference
1	Number of values	20
2		
3	Minimum	-4.000
4	25% Percentile	0.2500
5	Median	2.000
6	75% Percentile	4.000
7	Maximum	7.000
8		
9	Mean	2.050
10	Std. Deviation	2.837
11	Std. Error of Mean	0.6344
12		

	Col. stats	A Before Treatment	B After Treatment
1	Number of values	20	20
2			
3	Minimum	12.00	15.00
4	25% Percentile	16.00	17.25
5	Median	18.00	19.50
6	75% Percentile	21.00	24.00
7	Maximum	24.00	29.00
8			
9	Mean	18.40	20.45
10	Std. Deviation	3.152	4.058
11	Std. Error of Mean	0.7049	0.9075
12			

- Enter/import data:**
- 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

Table format: Column	Group A		Group B	
	Before Treatment	After Treatment	Before Treatment	After Treatment
1 Pt1		18		22
2 Pt2	You can use the subjects' IDs	21		25
3 Pt3		16		17
4 Pt4		22		24
5 Pt5		19		16
6 Pt6		24		29
7 Pt7		17		20
8 Pt8		21		23
9 Pt9		23		19
10 Pt10		18		20
11 Pt11		14		15
12 Pt12		16		15
13 Pt13		16		18
14 Pt14		19		26
15 Pt15		18		18
16 Pt16		20		24
17 Pt17		12		18
18 Pt18		22		25
19 Pt19		15		19
20 Pt20		17		16



## Paired t-test example by hand

$$H_0 : \mu_d = 0$$

$$H_A : \mu_d \neq 0$$

$$df = 20 - 1 = 19$$

Mean of difference ( $\bar{d}$ ) = 2.05

SD ( $s$ ) = 2.837

$$t = \frac{\bar{d}}{S_d / \sqrt{n}}$$

$$t = 2.05 / (2.837 / \sqrt{20}) = 2.05 / 0.635 = 3.23$$

df	$\alpha = 0.05$
1	12.706
2	4.303
3	3.182
4	2.776
5	2.571
6	2.447
7	2.365
8	2.306
9	2.262
10	2.228
11	2.201
12	2.179
13	2.160
14	2.145
15	2.131
16	2.120
17	2.110
18	2.101
19	2.093
20	2.086

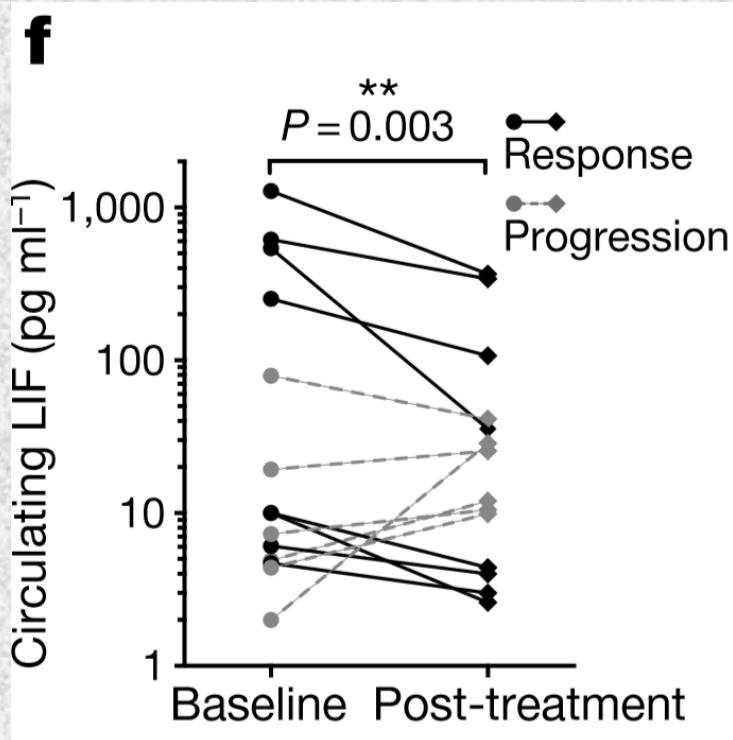
3.23 > 2.093 so  $p < 0.05$

# Targeting LIF-mediated paracrine interaction for pancreatic cancer therapy and monitoring

The rabbit hole...

2 MAY 2019 | VOL 569 | Nature | 131

Fig. 5 | LIF can be a biomarker for PDAC monitoring



f, Correlation of changes in circulating LIF level and therapeutic responses in patients with PDAC. Lines connect paired samples from one patient. n = 14 paired samples.

Statistical significance was determined by **Fisher's exact test** (f).

patients ID	Baseline	Post treatment	Response code
01-010	7.3	10.5	progression
01-011	4.9	12	progression
01-012	10	4.4	response
01-015	4.4	9.9	progression
01-016	10	2.6	response
01-017	79.2	41.2	progression
01-018	4.7	3.0	response
3	616.8	341	response
7	2	28.6	progression
10	19.3	25.5	progression
11	541.9	35.5	response

But, Fisher's Exact test is done on categorical data...  
 I tried to figure out what they did

patients ID	Baseline	Post treatment	Response code	Change
01-010	7.3	10.5	progression	up
01-011	4.9	12	progression	up
01-012	10	4.4	response	down
01-015	4.4	9.9	progression	up
01-016	10	2.6	response	down
01-017	79.2	41.2	progression	down
01-018	4.7	3.0	response	down
3	616.8	341	response	down
7	2	28.6	progression	up
10	19.3	25.5	progression	up
11	541.9	35.5	response	down
13	6.1	4	response	down
14	1280.5	367.3	response	down
30	252.4	107	response	down

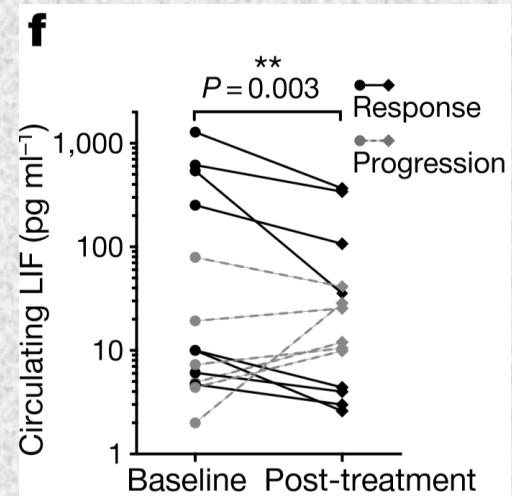
	Progress	Response
Up	5	0
Down	1	8

## Results of Fisher's Exact test

	Progress	Response
Up	5	0
Down	1	8

Contingency

	A	B	C
1 Table Analyzed	Fisher's Exact		
2			
3 P value and statistical significance			
4 Test	Fisher's exact test		
5 P value	0.0030		
6 P value summary	**		
7 One- or two-sided	Two-sided		
8 Statistically significant ( $P < 0.05$ )?	Yes		
9			
10 Data analyzed	Progress	Response	Total
11 Up	5	0	5
12 Down	1	8	9
13 Total	6	8	14



## Categorization hides magnitude of differences

Baseline	Post treatment	Baseline-Post	Response code
7.3	10.5	-3.2	progression
4.9	12	-7.1	progression
4.4	9.9	-5.5	progression
79.2	41.2	38	progression
2	28.6	-26.6	progression
19.3	25.5	-6.2	progression
10	4.4	5.6	response
10	2.6	7.4	response
4.7	3.0	1.7	response
616.8	341	275.8	response
541.9	35.5	506.4	response
6.1	4	2.1	response
1280.5	367.3	913.2	response
252.4	107	145.4	response

They want to know if progression and response values differ between baseline and post treatment.

What test would you use?

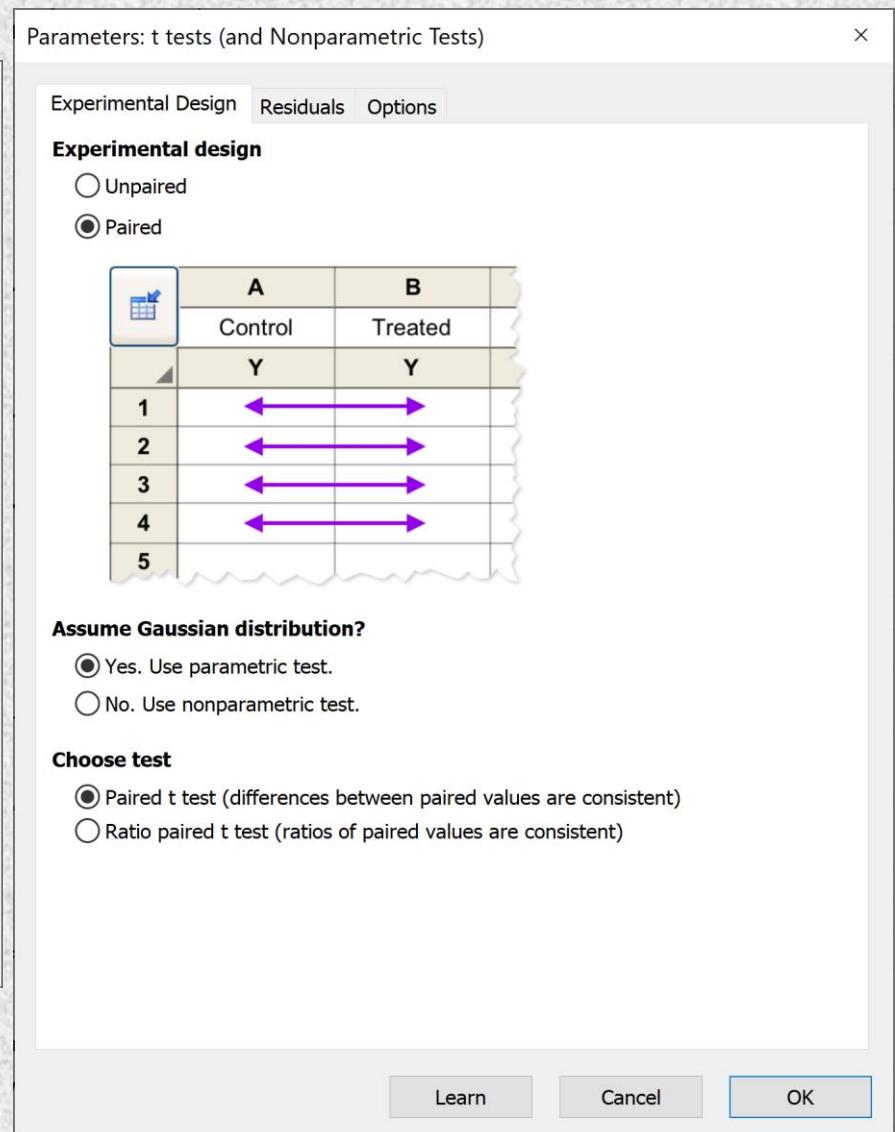
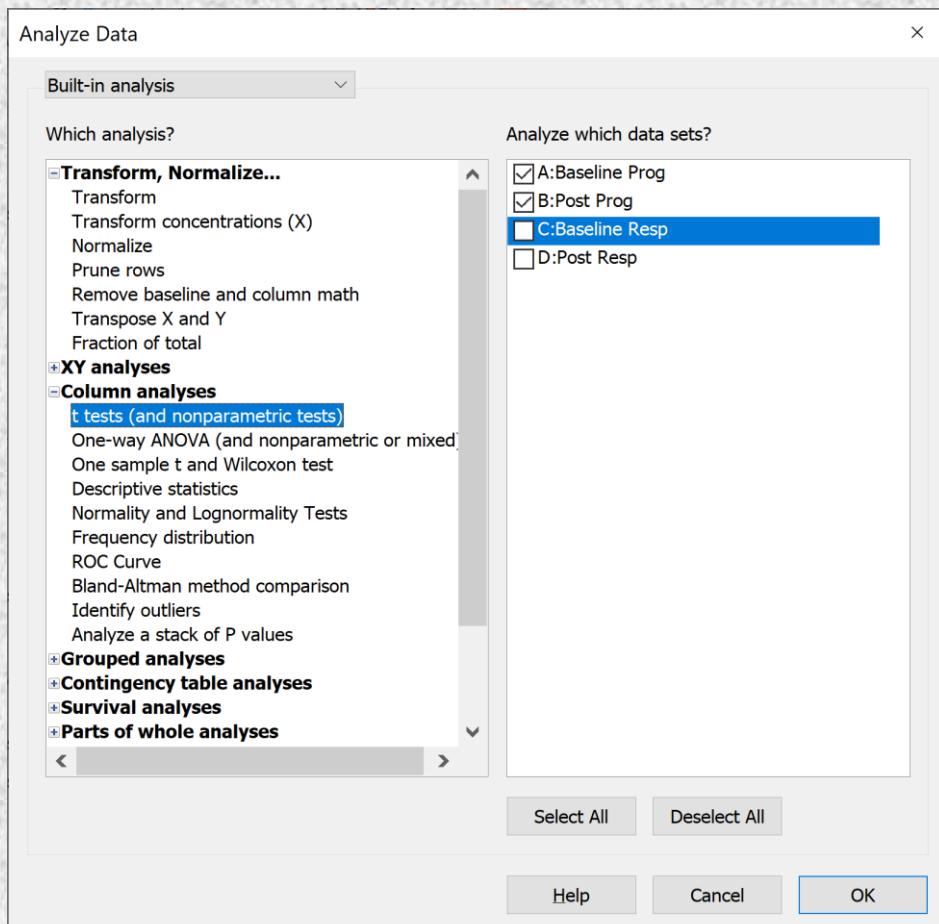
## Let's do paired t-tests for response and progression

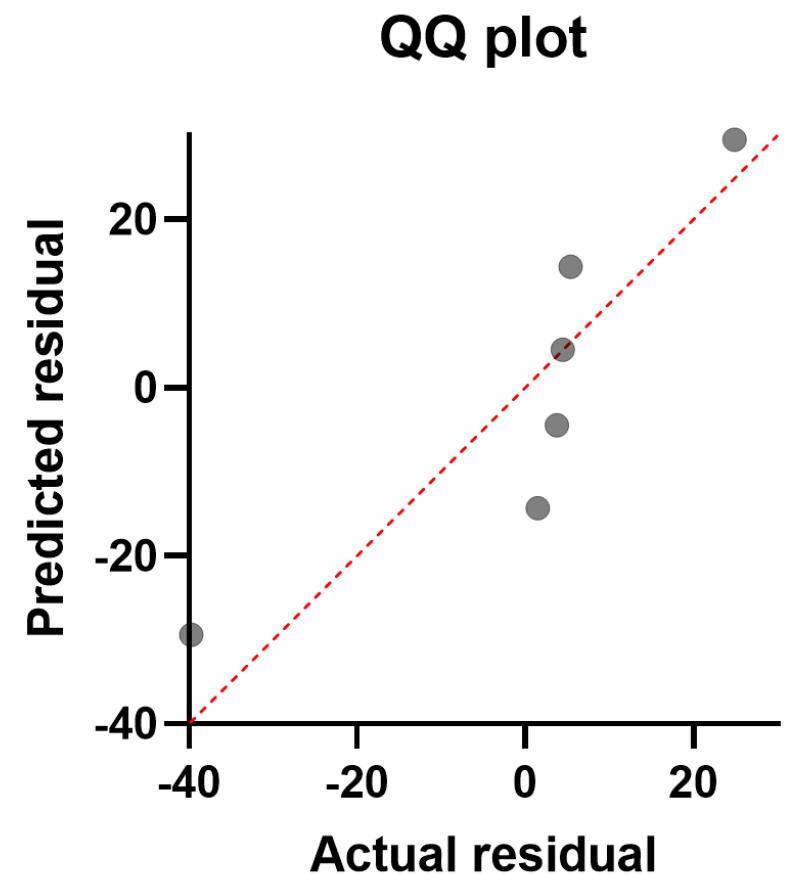
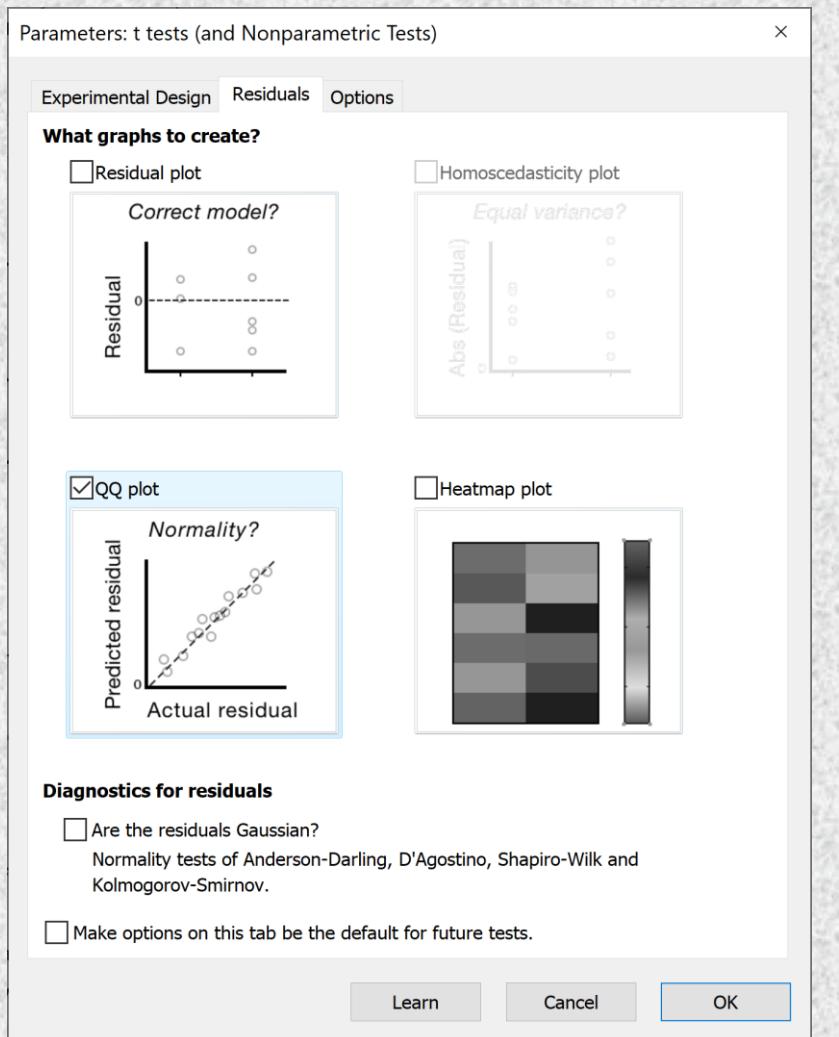
Baseline	Post treatment	Response code
7.3	10.5	progression
4.9	12	progression
4.4	9.9	progression
79.2	41.2	progression
2	28.6	progression
19.3	25.5	progression
10	4.4	response
10	2.6	response
4.7	3.0	response
616.8	341	response
541.9	35.5	response
6.1	4	response
1280.5	367.3	response
252.4	107	response

	Group A	Group B	Group C		Group D	
			Baseline Prog	Post Prog	Baseline Resp	Post Resp
1		7.3	10.5		10.0	4.4
2		4.9	12.0		10.0	2.6
3		4.4	9.9		4.7	3.0
4		79.2	41.2		616.8	341.0
5		2.0	28.6		541.9	35.5
6		19.3	25.5		6.1	4.0
7					1280.5	367.3
8					252.4	107.0
a						

### Baseline – Post treatment

	Group A	Group B
	Progress diff	Response diff
1	-3.2	5.6
2	-7.1	7.4
3	-5.5	1.7
4	38.0	275.8
5	-26.6	506.4
6	-6.2	2.1
7		913.2
8		145.4
9		





Parameters: t tests (and Nonparametric Tests)

Experimental Design Residuals Options

### Calculations

P value:  One-tailed  Two-tailed (recommended)

Report differences as: Baseline Prog - Post Prog

Confidence level: 95%

Definition of statistical significance: P < 0.05

### Graphing options

Graph differences (paired)  
 Graph ranks (nonparametric)  
 Graph correlation (paired)  
 Graph CI of mean of differences

### Additional results

Descriptive statistics for each data set  
 t test: Also compare models using AICc  
 Wilcoxon. Also compute the CI of the median paired differences  
 Assumes the distribution of paired differences is symmetrical.  
 Wilcoxon: When both values on a row are identical, use method of Pratt  
 If this option is unchecked, those rows are ignored and the results will match prior version of Prism

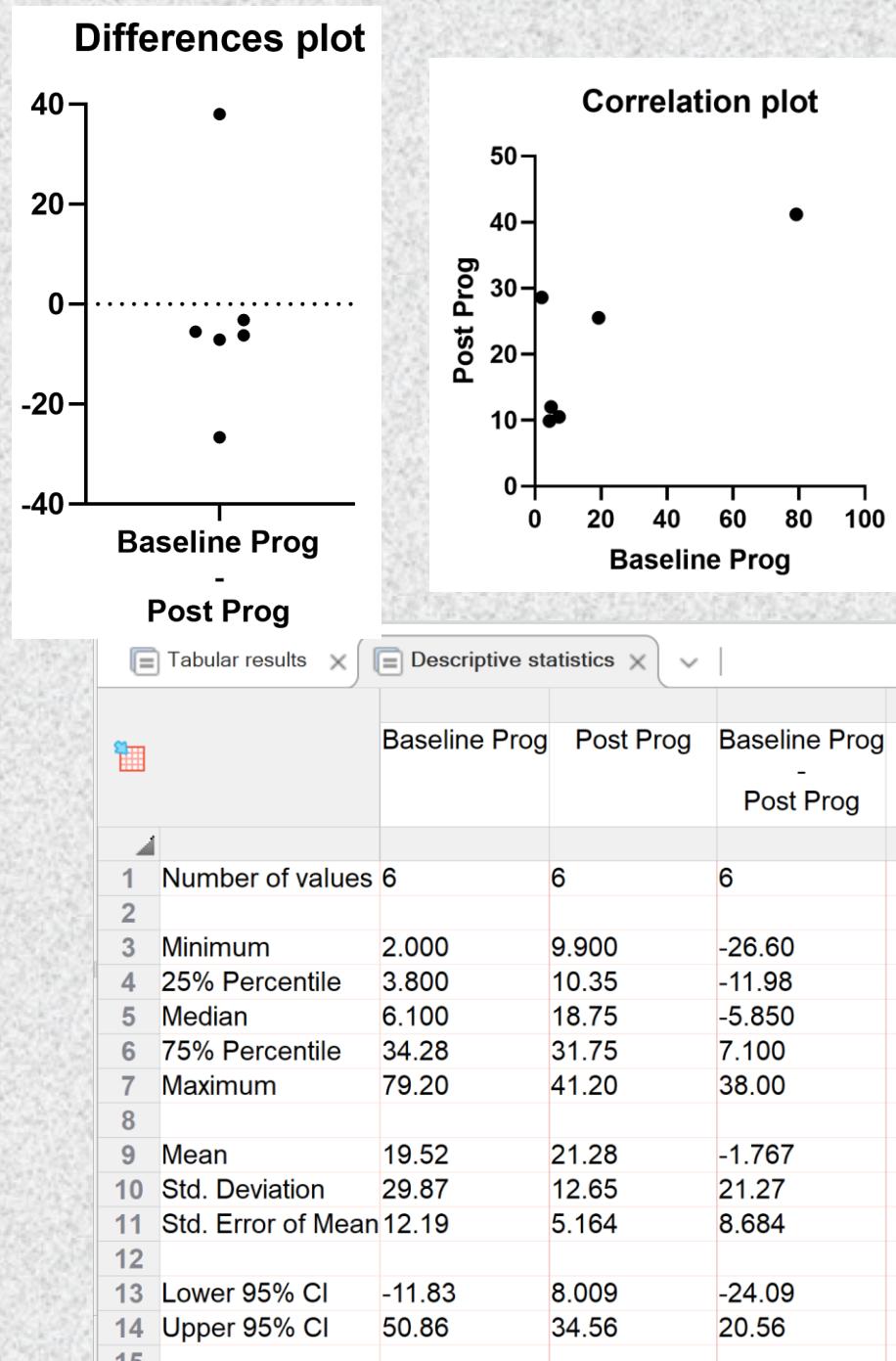
### Output

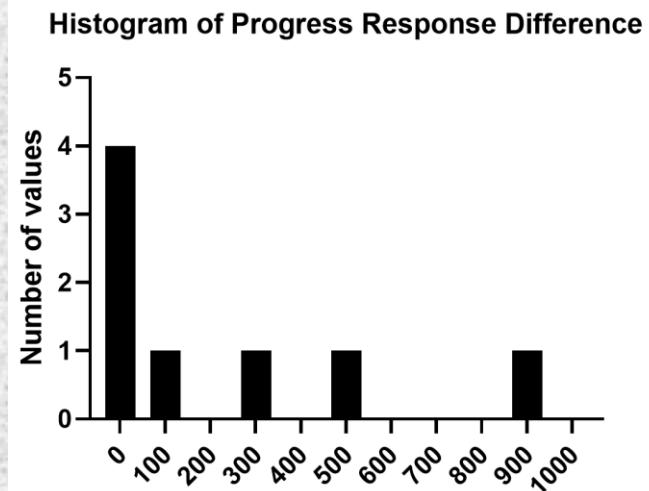
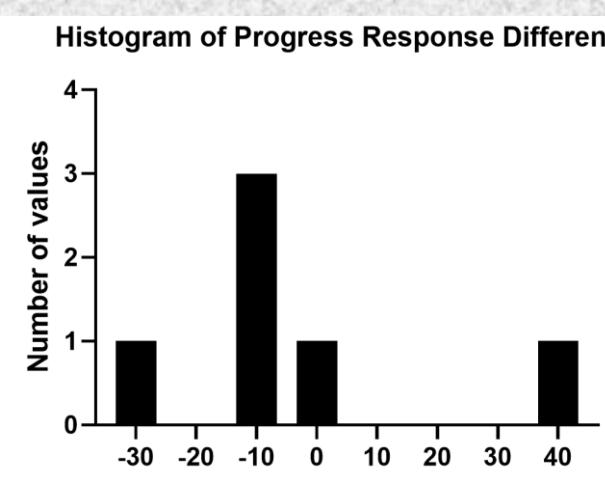
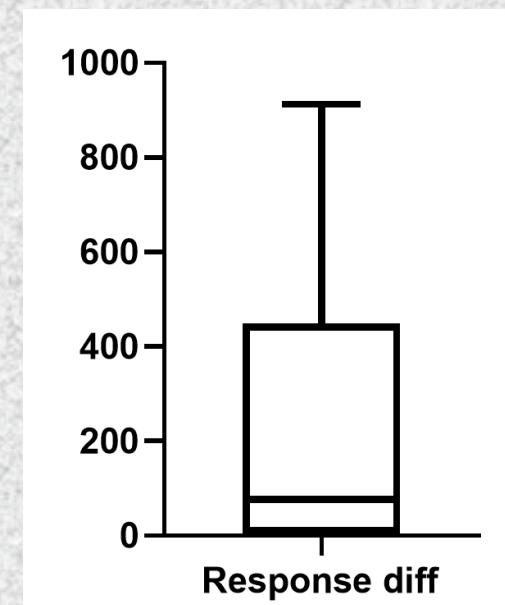
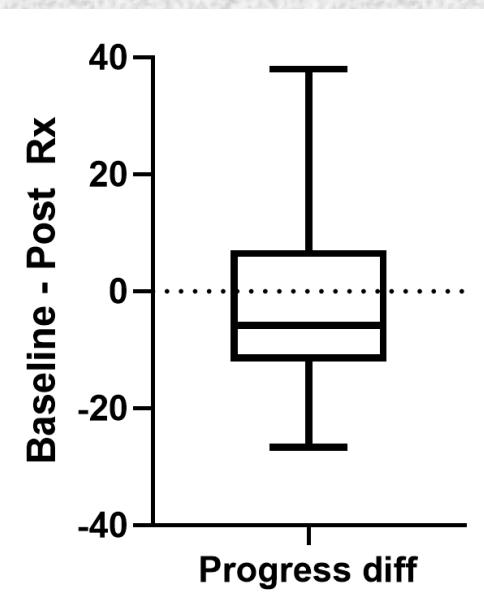
Show this many significant digits (for everything except P values): 4

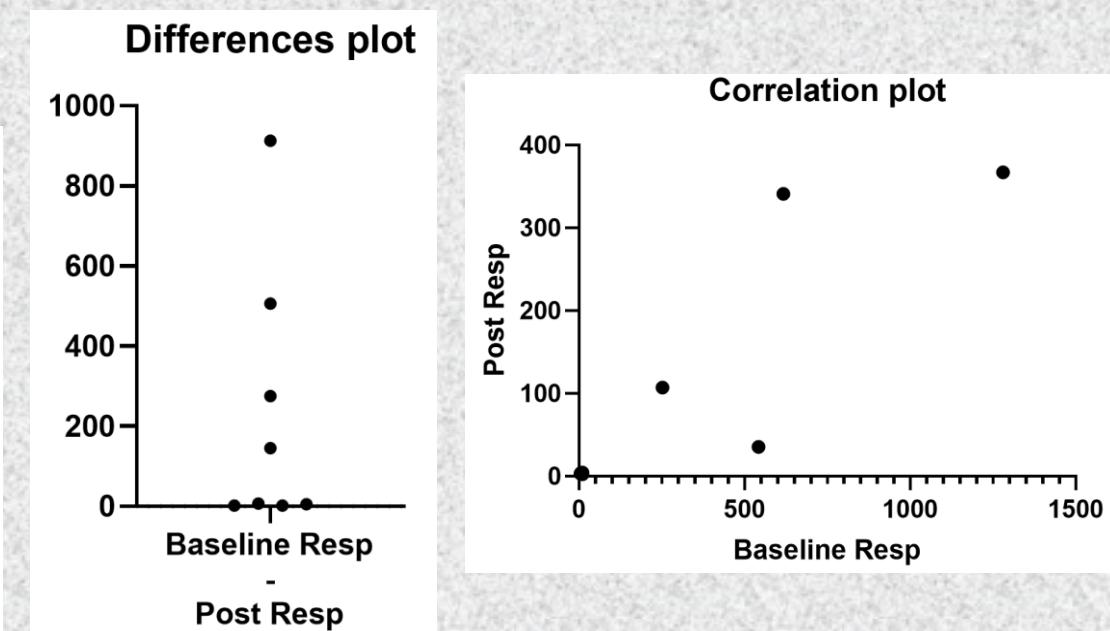
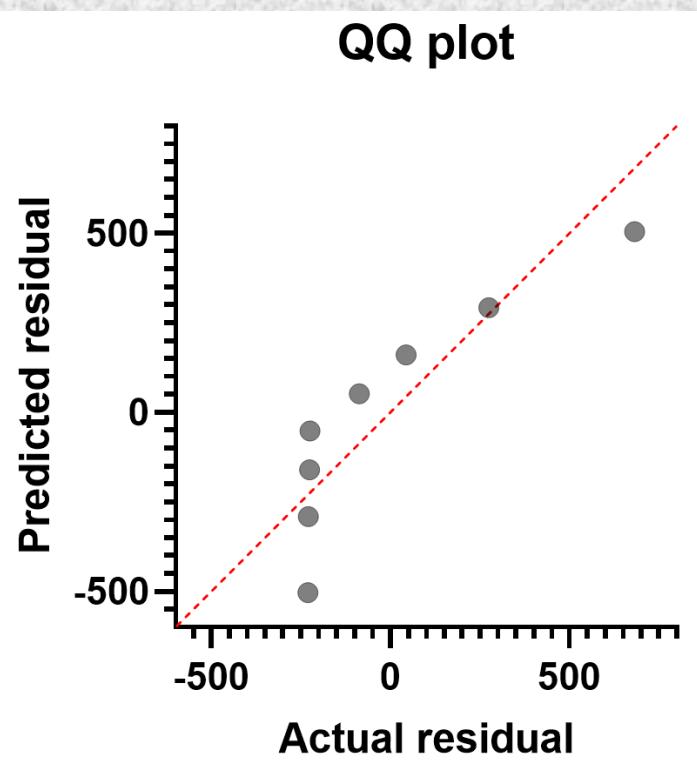
P value style: GP: 0.1234 (ns), 0.0332 (\*), 0.0021 (\*\*), 0.00 N = 6

Make options on this tab be the default for future tests.

Learn Cancel OK







Tabular results   Descriptive statistics

		Baseline Resp	Post Resp	Baseline Resp
				Post Resp
1	Number of values	8	8	8
2				
3	Minimum	4.700	2.600	1.700
4	25% Percentile	7.075	3.250	2.975
5	Median	131.2	19.95	76.40
6	75% Percentile	598.1	282.5	448.8
7	Maximum	1281	367.3	913.2
8				
9	Mean	340.3	108.1	232.2
10	Std. Deviation	455.3	156.1	328.7
11	Std. Error of Mean	161.0	55.17	116.2

Paired t test	
Tabular results	
1 Table Analyzed	Data by Progress and Response
2	
3 Column A	Baseline Prog
4 vs.	vs.
5 Column B	Post Prog
6	
7 <b>Paired t test</b>	
8 P value	0.8468
9 P value summary	ns
10 Significantly different ( $P < 0.05$ )?	No
11 One- or two-tailed P value?	Two-tailed
12 t, df	t=0.2034, df=5
13 Number of pairs	6
14	
15 <b>How big is the difference?</b>	
16 Mean of differences (A - B)	-1.767
17 SD of differences	21.27
18 SEM of differences	8.684
19 95% confidence interval	-24.09 to 20.56
20 R squared (partial eta squared)	0.008210
21	

Paired t test	
Tabular results	
1 Table Analyzed	Data by Progress and Response
2	
3 Column D	Post Resp
4 vs.	vs.
5 Column C	Baseline Resp
6	
7 <b>Paired t test</b>	
8 P value	0.0859
9 P value summary	ns
10 Significantly different ( $P < 0.05$ )?	No
11 One- or two-tailed P value?	Two-tailed
12 t, df	t=1.998, df=7
13 Number of pairs	8
14	
15 <b>How big is the difference?</b>	
16 Mean of differences (D - C)	-232.2
17 SD of differences	328.7
18 SEM of differences	116.2
19 95% confidence interval	-507.0 to 42.61
20 R squared (partial eta squared)	0.3632
21	

*But, did we meet the assumptions of the paired t-test?*

## Non-parametric alternative to the paired t-test: Wilcoxon signed-rank test

The test is based on a difference in rank scores of the differences

The test statistic is  $W$   
defined as the smaller absolute value of either  
 $W_+$  (sum of the positive ranks) or  
 $W_-$  (sum of the negative ranks)

$H_0$ : The distribution of the positive rank differences in the population is the same as the distribution of the negative rank differences in the population (often interpreted as: the median difference in the population is zero)  
 $H_A$ : The rank distributions in the population are not the same (the median difference in the population is not equal to zero)

Assumptions

Paired data

Pairs of data are independent

## Data (n=8)

Child	Before Treatment	After 1 Week of Treatment
1	85	75
2	70	50
3	40	50
4	65	40
5	80	
6	75	
7	55	
8	20	

How it is done

Calculate difference

Child	Before Treatment	After 1 Week of Treatment	Difference (Before-After)
1	85	75	10
2	70	50	20
3	40	50	-10
4	65	40	25
5	80	20	60
6	75	65	10
7	55	40	15
8	20	25	-5

Sort and rank the differences

Observed Differences	Ordered Absolute Values of Differences	Ranks
10	-5	1
20	10	3
-10	-10	3
25	10	3
60	15	5
10	20	6
15	25	7
-5	60	8

The final step is to attach the signs ("+" or "-") of the observed differences to each rank

Observed Differences	Ordered Absolute Values of Difference Scores	Ranks	Signed Ranks
10	-5	1	-1
20	10	3	3
-10	-10	3	-3
25	10	3	3
60	15	5	5
10	20	6	6
15	25	7	7
-5	60	8	8

You sum the positive and negative ranks separately

$$W+ = 32, W- = -4$$

If the null hypothesis is in fact true, the absolute values of both positive ( $W+$ ) and negative ( $W-$ ) should be similar

The smaller absolute  $W$  value is the  $W$  statistic (in this case,  $W=4$ )

## Critical Values for the Wilcoxon Signed-Rank Test

$p \leq 0.05$  if calculated  
 W statistic is  $\leq$  critical  
 value in table

$W_{\text{calc}} = 4$  then  $p=0.05$

	Col. stats	A
		difference
		Y
1	Number of values	8
2		
3	Wilcoxon Signed Rank Test	
4	Theoretical median	0
5	Actual median	12.5
6	Discrepancy	12.5
7	Sum of signed ranks (W)	28
8	Sum of positive ranks	32
9	Sum of negative ranks	-4
10	P value (two tailed)	0.0625
11	Exact or estimate?	Exact
12	Significant (alpha=0.05)?	No
13		

n	$\alpha$			
	.005 (one tail) .01 (two tails)	.01 (one tail) .02 (two tails)	.025 (one tail) .05 (two tails)	.05 (one tail) .10 (two tails)
	*	*	*	1
5	*	*	*	1
6	*	*	1	2
7	*	0	2	4
8	0	2	4	6
9	2	3	6	8
10	3	5	8	11
11	5	7	11	14
12	7	10	14	17
13	10	13	17	21
14	13	16	21	26
15	16	20	25	30
16	19	24	30	36
17	23	28	35	41
18	28	33	40	47
19	32	38	46	54
20	37	43	52	60
21	43	49	59	68
22	49	56	66	75
23	55	62	73	83
24	61	69	81	92
25	68	77	90	101

Are p-values different because Prism's exact p accounted for the 3 ties?

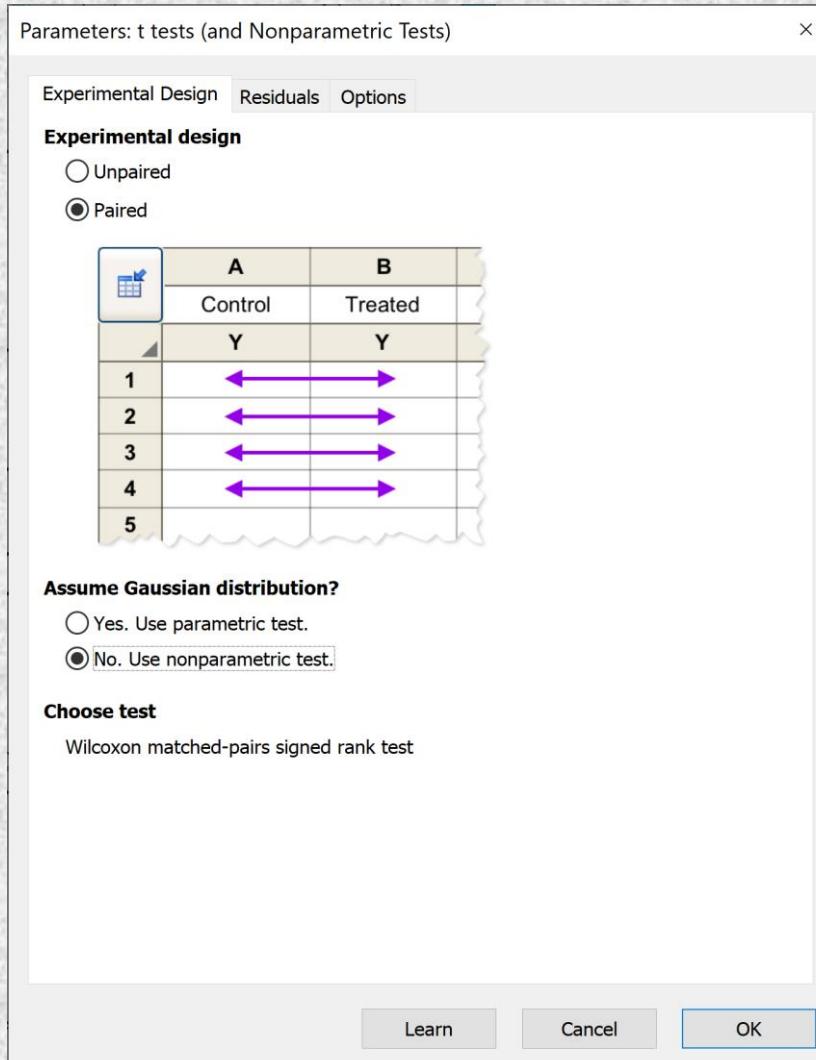
## Write up results (from previous incomplete example)

For this class:

We tested our hypothesis that ..... The data did not pass the normality assumption for a paired t-test so we used the Wilcoxon signed rank test and met the assumptions of the Wilcoxon test (data are paired, pairs are independent). With  $p=0.06$  (two tailed,  $\alpha=0.05$ ), we fail to reject the null hypothesis that the median difference of populations is zero (a median difference of .....).

# Wilcoxon signed rank test on the Progression/Response data in Prism

Prism calls it the Wilcoxon matched pairs signed rank test



Wilcoxon test	
1	Table Analyzed
2	Data by Progress and Response
3	Column A
4	vs.
5	Column B
6	
7	<b>Wilcoxon matched-pairs signed rank test</b>
8	P value
9	Exact or approximate P value?
10	P value summary
11	Significantly different ( $P < 0.05$ )?
12	One- or two-tailed P value?
13	Sum of positive, negative ranks
14	Sum of signed ranks (W)
15	Number of pairs
16	Number of ties (ignored)
17	
18	<b>Median of differences</b>
19	Median

<u>Paired t-test p-values</u>	
Progression	0.85
Response	0.09

*So, which results do you report?  
Paired t-test or Wilcoxon signed rank?*

Wilcoxon test	
1	Table Analyzed
2	Data by Progress and Response
3	Column C
4	vs.
5	Column D
6	
7	<b>Wilcoxon matched-pairs signed rank test</b>
8	P value
9	Exact or approximate P value?
10	P value summary
11	Significantly different ( $P < 0.05$ )?
12	One- or two-tailed P value?
13	Sum of positive, negative ranks
14	Sum of signed ranks (W)
15	Number of pairs
16	Number of ties (ignored)
17	
18	<b>Median of differences</b>
19	Median

*But, we haven't answered the question if the differences are different by progression and response groups.*

How do you do that? Which test should we use?

# Is there a difference between differences?



	Group A	Group B
	Progress diff	Response diff
1	-3.2	5.6
2	-7.1	7.4
3	-5.5	1.7
4	38.0	275.8
5	-26.6	506.4
6	-6.2	2.1
7		913.2
8		145.4

Are these data paired or independent?

Do they meet the assumptions of normality?

What about homoscedasticity?

What test would you use?



	Descriptive statistics	A	B
		Progress diff	Response diff
1	Number of values	6	8
2			
3	Minimum	-26.60	1.700
4	25% Percentile	-11.98	2.975
5	Median	-5.850	76.40
6	75% Percentile	7.100	448.8
7	Maximum	38.00	913.2
8	Range	64.60	911.5
9			
10	Mean	-1.767	232.2
11	Std. Deviation	21.27	328.7
12	Std. Error of Mean	8.684	116.2
13			
17	Skewness	1.474	1.547
18	Kurtosis	3.607	1.902

# I did a Mann Whitney test

Parameters: t tests (and Nonparametric Tests)

Experimental Design Residuals Options

**Calculations**

P value:  One-tailed  Two-tailed (recommended)

Report differences as: Response diff - Progress diff

Confidence level: 95%

Definition of statistical significance: P < 0.05

**Graphing options**

Graph differences (paired)  
 Graph ranks (nonparametric)  
 Graph correlation (paired)  
 Graph CI of difference between medians

**Additional results**

Descriptive statistics for each data set  
 t test: Also compare models using AICc  
 Mann-Whitney: Also compute the CI of difference between medians  
Assumes both distributions have the same shape.  
 Wilcoxon: When both values on a row are identical, use method of Pratt  
If this option is unchecked, those rows are ignored and the results will match prior version of Prism

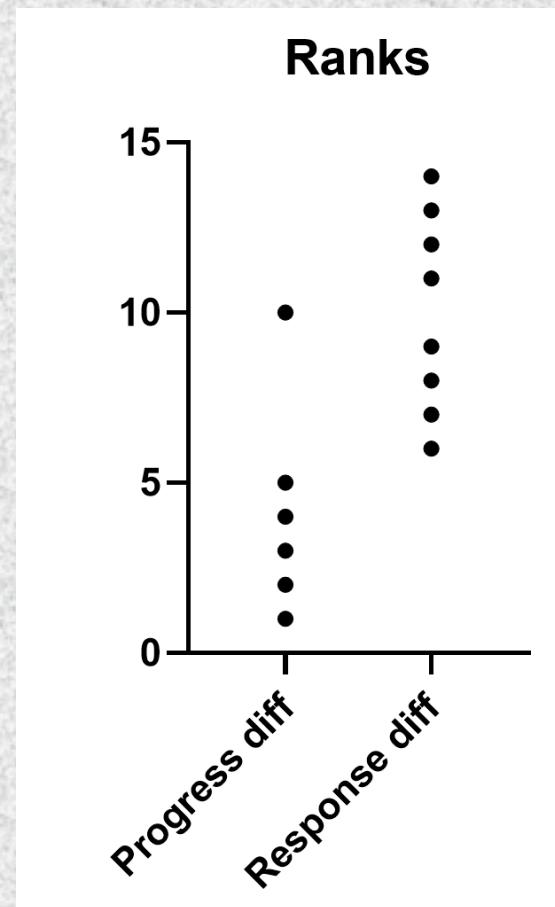
**Output**

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 tests.

Learn Cancel OK





Mann-Whitney test		
Tabular results		
1	Table Analyzed	Progress Response Difference
2		
3	Column B	Response diff
4	vs.	vs.
5	Column A	Progress diff
6		
7	<b>Mann Whitney test</b>	
8	P value	0.0080
9	Exact or approximate P value?	Exact
10	P value summary	**
11	Significantly different ( $P < 0.05$ )?	Yes
12	One- or two-tailed P value?	Two-tailed
13	Sum of ranks in column A,B	25 , 80
14	Mann-Whitney U	4
15		
16	<b>Difference between medians</b>	
17	Median of column A	-5.850, n=6
18	Median of column B	76.40, n=8
19	Difference: Actual	82.25
20	Difference: Hodges-Lehmann	70.70

Mann-Whitney test		
Tabular results		
1	Table Analyzed	Progress Response Difference
2		
3	Column B	Response diff
4	vs.	vs.
5	Column A	Progress diff
6		
7	<b>Mann Whitney test</b>	
8	P value	0.0080
9	Exact or approximate P value?	Exact
10	P value summary	**
11	Significantly different ( $P < 0.05$ )?	Yes
12	One- or two-tailed P value?	Two-tailed
13	Sum of ranks in column A,B	25 , 80
14	Mann-Whitney U	4
15		
16	<b>Difference between medians</b>	
17	Median of column A	-5.850, n=6
18	Median of column B	76.40, n=8
19	Difference: Actual	82.25
20	Difference: Hodges-Lehmann	70.70



What if I did the parametric test...

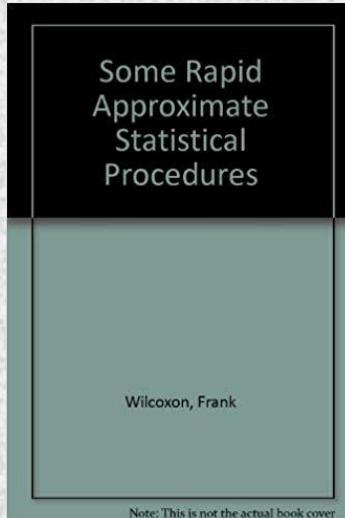
Welch's t test		
1	Table Analyzed	Progress Response Difference
2		
3	Column B	Response diff
4	vs.	vs.
5	Column A	Progress diff
6		
7	<b>Unpaired t test with Welch's correction</b>	
8	P value	0.0842
9	P value summary	ns
10	Significantly different ( $P < 0.05$ )?	No
11	One- or two-tailed P value?	Two-tailed
12	Welch-corrected t, df	t=2.008, df=7.078
13		
14	<b>How big is the difference?</b>	
15	Mean of column A	-1.767
16	Mean of column B	232.2
17	Difference between means (B - A) $\pm$ SEM	234.0 $\pm$ 116.5
18	95% confidence interval	-40.99 to 508.9
19	R squared (eta squared)	0.3628
20		

# McNemar's Test on Paired Categorical Data (2x2 table)

The McNemar test is a non-parametric test for paired categorical data. It's used when you are interested in finding a change in proportion for the paired data.



**McNemar**



Published in 1949

Frank Wilcoxon (1892–1965), chemist and statistician

Created by Quinn McNemar, who was a professor in the Psychology and Statistics department at Stanford University

## McNemar's Test for Paired Samples (2x2 tables only)

		Test 1	
		POS	NEG
Test 2	POS	a	b
	NEG	c	d

McNemar's tests if proportions in cells b and c (marginal cells) are different from each other

$$\chi^2 = \frac{(b - c)^2}{b + c}$$

$H_0$ : marginal homogeneity in the population (cell b = cell c)

$H_A$ : no marginal homogeneity in the population (cell b  $\neq$  cell c)

Assumptions:

Paired data (*i.e.*, test 1 and test 2 done in the same mouse)

Mice are independent (and can only be in one cell of the table)

(Sample is random)

McNemar's is testing for consistency in responses across two variables (e.g., tests)  
A  $p \leq 0.05$  leads to a conclusion that the tests are different

**McNemar's Test:** A comparison between three definitions of diuretic resistance (Tables 3, 4, and 5) found each identified significantly difference on the tests (i.e., urinary sodium and FeNa)

Table 3

Comparison of measured urinary sodium and FeNa as measures of diuretic resistance

		Measured urinary sodium		
		Normal	Diuretic resistant	
FeNa	Normal	154	22	$p < 0.001$
	Diuretic resistant	1	10	

Table 4

Comparison of Na/K ratio and FeNa as measures of diuretic resistance

		Na/K ratio		
		Normal	Diuretic resistant	
FeNa	Normal	166	10	$p = 0.02$
	Diuretic resistant	1	10	

Table 5

Comparison of predicted urinary sodium and measured urinary sodium as measures of diuretic resistance

		Measured urinary sodium		
		Normal	Diuretic resistant	
Na/K ratio	Normal	152	15	$p = 0.01$
	Diuretic resistant	3	17	

## Latex Agglutination Test for Monitoring Antibodies to Avian Influenza Virus Subtype H5N1

Xiaojuan Xu,<sup>1</sup> Meilin Jin,<sup>1,2\*</sup> Zhengjun Yu,<sup>1</sup> Hongchao Li,<sup>1</sup> Dexin Qiu,<sup>1</sup> Yadi Tan,<sup>1,2</sup> and Huanchun Chen<sup>1,2</sup>

<sup>1</sup>Laboratory of Animal Virology, College of Veterinary Medicine, and Unit of Animal Infectious Diseases, National Key Laboratory of Agricultural Microbiology,<sup>2</sup> Huazhong Agricultural University, Wuhan, Hubei Province, People's Republic of China

Received 12 August 2004/Returned for modification 6 November 2004/Accepted 16 December 2004

TABLE 3. Correlation between the LAT and the HI test in detecting antibodies to AIV H5N1<sup>a</sup>

Result	No. of samples with LAT result of:		Total no. (%) of HI or LAT results
	Positive	Negative	
HI positive	664	84	748 (90.1)
HI negative	2	80	82 (9.9)
LAT total (%)	666 (80.2)	164 (19.8)	830

<sup>a</sup> Sensitivity of LAT (664 of 748), 88.8%; specificity of LAT (80 of 82), 97.6%; correlation [(664 + 80)/830], 89.6%.

	Positive	Negative
	664	84
	2	80

$$\chi^2 = \frac{(b-c)^2}{b+c} = \frac{(84-2)^2}{84+2} = 78.19, 1 \text{ df (2x2 table)}$$

Table of the chi square distribution

df	Level of Significance $\alpha$								
	0.200	0.100	0.075	0.050	0.025	0.010	0.005	0.001	0.0005
1	1.642	2.706	3.170	3.841	5.024	6.635	7.879	10.828	12.116
2	3.219	4.605	5.181	5.991	7.378	9.210	10.597	13.816	15.202
3	4.642	6.251	6.905	7.815	9.348	11.345	12.838	16.266	17.731
4	5.989	7.779	8.496	9.488	11.143	13.277	14.860	18.467	19.998
5	7.289	9.236	10.008	11.070	12.833	15.086	16.750	20.516	22.106
6	8.558	10.645	11.466	12.592	14.449	16.812	18.548	22.458	24.104
7	9.803	12.017	12.883	14.067	16.013	18.475	20.278	24.322	26.019
8	11.030	13.362	14.270	15.507	17.535	20.090	21.955	26.125	27.869
9	12.242	14.684	15.631	16.919	19.023	21.666	23.589	27.878	29.667
10	13.442	15.987	16.971	18.307	20.483	23.209	25.188	29.589	31.421
11	14.631	17.275	18.294	19.675	21.920	24.725	26.757	31.265	33.138
12	15.812	18.549	19.602	21.026	23.337	26.217	28.300	32.910	34.822
13	16.985	19.812	20.897	22.362	24.736	27.688	29.820	34.529	36.479
14	18.151	21.064	22.180	23.685	26.119	29.141	31.319	36.124	38.111
15	19.311	22.307	23.452	24.996	27.488	30.578	32.801	37.698	39.720
16	20.465	23.542	24.716	26.296	28.845	32.000	34.267	39.253	41.309
17	21.615	24.769	25.970	27.587	30.191	33.409	35.719	40.791	42.881
18	22.760	25.989	27.218	28.869	31.526	34.805	37.157	42.314	44.435
19	23.900	27.204	28.458	30.144	32.852	36.191	38.582	43.821	45.974
20	25.038	28.412	29.692	31.410	34.170	37.566	39.997	45.315	47.501

78.19 > 3.841

P < 0.05

Reject null hypothesis

*<https://www.graphpad.com/quickcalcs/>*

The screenshot shows the GraphPad QuickCalcs website. At the top, there's a header bar with the URL "graphpad.com/quickcalcs/" in the address bar. The header includes the GraphPad logo, navigation links for Prism, Customers, Resources, Support, Pricing, a shopping cart icon, Sign In, a search icon, and a "Free Trial" button. Below the header, a large purple section features the title "QuickCalcs" in white, bold letters. A subtitle below it reads "Run statistical analyses quickly and directly in your browser". The main content area has a white background and displays a card for "Categorical data" which lists statistical tests like Fisher's, Chi square, McNemar's, Sign test, CI of proportion, NNT, and kappa. There's also a right-pointing arrow icon.

graphpad.com/quickcalcs/

GraphPad Prism Customers Resources Support Pricing Cart Sign In Free Trial

# QuickCalcs

Run statistical analyses quickly and directly in your browser

Categorical data

Fisher's, Chi square, McNemar's, Sign test, CI of proportion, NNT (number needed to treat), kappa.

# QuickCalcs

[1. Select category](#)

**2. Choose calculator**

[3. Enter data](#)

[4. View results](#)

Analyze categorical data

- Confidence interval of a proportion or count.
- Chi-square. Compare observed and expected frequencies.
- Fisher's and chi-square. Analyze a 2x2 contingency table.
- McNemar's test to analyze a matched case-control study.
- Binomial and sign test. Compare observed and expected proportions.
- NNT (Number Needed to Treat) with confidence interval.
- Predictive values from sensitivity, specificity, and prevalence.
- Kappa. Quantify interrater agreement.

[CONTINUE >](#)

Prism interprets the result like you had conducted a matched case-control study so the interpretation from the Prism results can be a bit confusing

## McNemar's test to analyze a matched case-control study

McNemar's test is used to compare paired proportions. It can be used to analyze retrospective case-control studies, where each case is matched to a particular control. Or it can be used to analyze experimental studies, where the two treatments are given to matched subjects. [Read an example with explanation.](#)

Risk Factor?

Control Case # of pairs

No	Yes	84
Yes	No	2
Yes	Yes	664
No	No	80

*It doesn't matter if you enter 2 then 84. Just be careful how you interpret the percentages calculated by Prism.*

Calculate

Use McNemar's test (and this calculator) only when you are analyzing matched pairs. Each value you enter above represents a number of PAIRS. The total number of subjects in the study is twice the total of the values you enter above.

Note that the calculations are based entirely on the first two numbers you enter. Enter the remaining two numbers in order to document your full results.

TABLE 3. Correlation between the LAT and the HI test in detecting antibodies to AIV H5N1<sup>a</sup>

Result	No. of samples with LAT result of:		Total no. (%) of HI or LAT results
	Positive	Negative	
HI positive	664	84	748 (90.1)
HI negative	2	80	82 (9.9)
LAT total (%)	666 (80.2)	164 (19.8)	830

<sup>a</sup> Sensitivity of LAT (664 of 748), 88.8%; specificity of LAT (80 of 82), 97.6%; correlation [(664 + 80)/830], 89.6%.

## Results of McNemar's test for a case-control study

**Summary:** *Or in our example, the two tests produce the same results*

If there were no association between the risk factor and the disease, you'd expect the number of pairs where cases was exposed to the risk factor but control was not to equal the number of pairs where the control was exposed to the risk factor but the case did not. In this study, there were 86 discordant pairs (case and control had different exposure to the risk factor). There were 2 ( 2.326%) pairs where the control was exposed to the risk factor but the case was not, and 84 ( 97.674%) pairs where the case was exposed to the risk factor but the control was not.

*Or in our example, 2.3% of discordant pairs had a positive LAT test and a negative HI test*

**P Value:**

The two-tailed P value is less than 0.0001

*i.e., strong evidence against the null*

By conventional criteria, this difference is considered to be extremely statistically significant.

The P value was calculated with McNemar's test with the continuity correction.

Chi squared equals 76.291 with 1 degrees of freedom.

The P value answers this question: If there is no association between risk factor and disease, what is the probability of observing such a large discrepancy (or larger) between the number of the two kinds of discordant pairs? A small P value is evidence that there is an association between risk factor and disease.

*Or in our example, the two tests have different performances*

# Interpreting a McNemar's test result

If the two tests were working about the same, you would expect the disagreement between the two tests to be similar (cells b (84) and c (2) would have similar counts).

There were 84 cases (pairs) where the HI test was positive but the LAT test was not (97.7% of the discordant pairs)

There were 2 cases (pairs) where the LAT test was positive but the HI test was not (2.3% of the discordant pairs)

What is the probability of observing such a large discrepancy (or larger) between the number of the two kinds of discordant pairs?

A small P value is evidence that the tests are not measuring the same thing.

		LAT test	
		Pos	Neg
HI test	Pos	664	84
	Neg	2	80

p<0.0001 (chi-square=76.3, 1 df)

The LAT test is positive 80.2% (666/830) of the time

The HI test is positive 90.1% (748/830) of the time

## Write up Results

We tested the hypothesis that the HI test detected a larger population of positive tests compared to the LAT test. We performed a McNemar's test on the paired data (each person had both tests performed on their samples). The data met the assumptions of paired data and independent pairs. With  $p<0.0001$  (two-sided, chi-square=76.3, 1 df,  $\alpha=0.05$ ), we reject the null hypothesis that the marginal cells are equivalent and conclude that the HI test is finding a higher number of positive results (90.1%) compared to the LAT test (80.2%)



Research paper

## Quantification of B-cell maturation antigen, a target for novel chimeric antigen receptor T-cell therapy in Myeloma



Dalia A. Salem<sup>a,b,\*</sup>, Irina Maric<sup>c</sup>, Constance M. Yuan<sup>a</sup>, David J. Liewehr<sup>d</sup>, David J. Venzon<sup>d</sup>, James Kochenderfer<sup>c</sup>, Maryalice Stetler-Stevenson<sup>a</sup>

**Table 3**

Comparison between FC and IHC in qualitative detection of BCMA in Plasma Cell Myeloma.

	FC +	FC -	Total
IHC +	28	0	28
IHC -	10	1	11
IHC na	3	1	4
Total	41	2	43

FC; Flow Cytometry. IHC; Immunohistochemistry, BCMA; B-cell maturation antigen, na; not applicable (could not be assessed).

Within the 39 specimens assessed successfully by both IHC and FC there was higher positivity rate achieved by FC (38/39; 97%) than IHC (28/39; 72%) (Table 3) indicating that FC is more sensitive than IHC in detection of BCMA ( $p = 0.002$ ; McNemar's test,  $n = 39$  specimens).

# Another way to test the discordant cells: the Binomial test

## McNemar's test

Prev  Ne 

### Computing the P value with Prism using the Binomial test

When you read about McNemar's test, most books explain how to do a chi-square calculation. Prism won't do that, but we offer a [free web calculator](#) that does. The binomial test asks the same question, but is more accurate, especially with small studies. Follow these steps with Prism:

1. Create a parts-of-whole data table.
2. Enter the numbers of discordant pairs in the first two rows of column A. For the example, enter 25 and 4.
3. Click Analyze and choose the analysis that compares observed and expected counts.
4. Choose to enter the expected values as percentages, and enter 50 as both expected percentages.
5. Choose the binomial test, rather than the chi-square test.
6. For the sample data, the P value is less than 0.0001. The P value answers this question: If there really were no association between disease and risk factor, what is the chance that the two values entered into this analysis would be as far apart as they are, or even further?

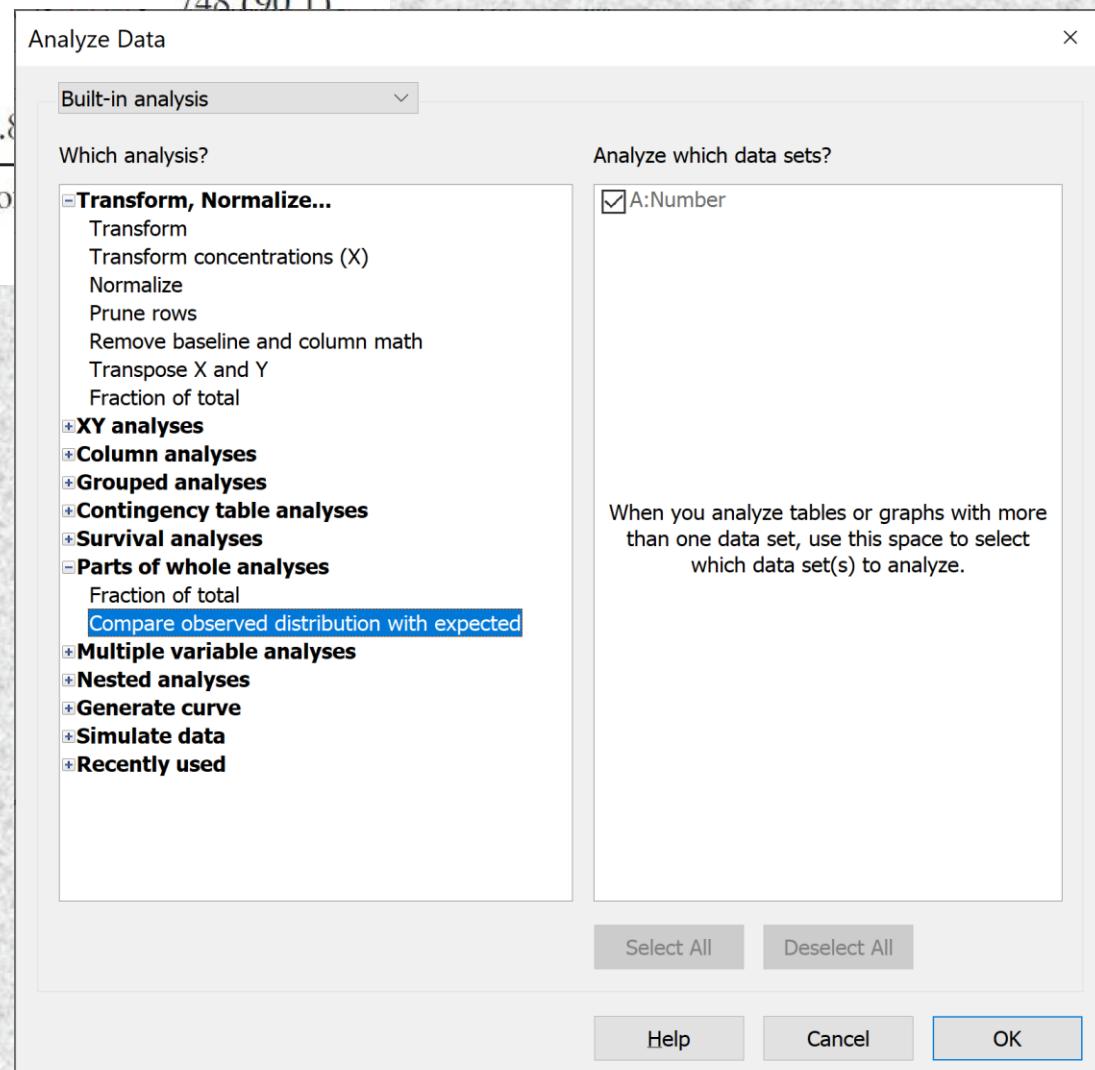
TABLE 3. Correlation between the LAT and the HI test in detecting antibodies to AIV H5N1<sup>a</sup>

Result	No. of samples with LAT result of:		Total no. (%) of HI or LAT results
	Positive	Negative	
HI positive	664	84	748 (90.1)
HI negative	2	80	
LAT total (%)	666 (80.2)	164 (19.8)	

<sup>a</sup> Sensitivity of LAT (664 of 748), 88.8%; specificity of LAT (80 of 164), 49.4%; correlation [(664 + 80)/830], 89.6%.

Table format:  
Parts of whole

	A	
	Number	
1	Cell b	84
2	Cell c	2
3	Title	



### Parameters: Compare observed distribution with expected

This analysis expects that each value in the data table is an actual number of events or items, and is not normalized in any way.

#### Data set to analyze:

A : Number

#### Enter expected values as:

- Actual numbers of objects or events
- Percentages

#### With two rows, perform:

- Binomial test (recommended)
- Chi-square test for goodness of fit

#### Expected distribution

Row	Outcome	Observed %	Expected %
1	Cell b	97.67	50
2	Cell c	2.33	50

#### Output

Method to calculate CI: Wilson/Brown (recommended)

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

Learn

Cancel

OK

We are testing if the proportions of Cell b and Cell a are the same

Null: proportion in population are the same  
 Alternative: proportions in population are different

O vs. E		A
1	Table analyzed	Data 1
2	Column analyzed	Column A
3		
4	Binomial test	
5	P (one-tailed)	<0.0001
6	P (two-tailed)	****
7	P value summary	
8	Is discrepancy significant (P < 0.05)? Yes	