

Common Statistical Tests Encountered in Chiropractic Research

Michael T. Haneline, DC, MPH

Introduction:

The purpose of this brief discourse on statistical tests is to enable chiropractors to better understand the mechanisms used by researchers as they evaluate and then draw conclusions from data in scientific articles. The emphasis is on understanding the concepts, so mathematics is purposefully deemphasized.

Statistical tests are utilized in determining the probabilities of relationships in a study. Some of the tests that doctors of chiropractic (DCs) commonly encounter while reading scientific articles are ***t*-tests**, **chi-square** (non-parametric comparisons), **ANOVA**, **contingency tables**, **Pearson's correlation** and **regression**.

- The ***t*-test** (AKA Student's *t*-test) assesses whether the means of two groups are *statistically* different from each other. However, the results of a *t*-test aren't entirely black-and-white. The results actually indicate that the means are probably different, or if the study fails to show an effect, that the means are probably the same.
- **ANOVA** (analysis of variance) is used to compare means when more than two groups are involved. Intuitively, one would think that a *t*-test could be simply be repeated for as many groups as are involved in the study, however this leads to an increased possibility of committing a **type I error** (rejecting a true null hypothesis).
- The **chi-square test** is used with non-parametric data, which includes nominal and ordinal data, but may also include interval and ratio data from a population that does not conform to a normal distribution.
- **Contingency tables** are frequently presented in the literature and are simply tables that represent relationships between categorical or **binomial** (when there are exactly two mutually exclusive outcomes) variables. An example would be the relationship between being adjusted (yes or no) and having a headache (yes or no).
- **Pearson's correlation** (or simply correlation) is a number (***r* = correlation coefficient**) that describes the strength of the relationship between two variables. Relationships can be positive or negative and vary from -1 to +1. A strong positive relationship would be near +1, no relationship would be 0, and a strong negative relationship near -1. An example is the relationship between height and weight, which is strongly positive. Taller people tend to be heavier and shorter people tend to be lighter. **R^2** is the proportion of one variable that can be explained by the other one. Consequently, we can say, in the case of height and weight, 75% of the increase in a person's weight is explained by increased height.
- **Regression** is a method of finding an equation that allows for prediction of the value of one variable that is based on the value of another variable. Regression is similar to correlation, but instead of expressing the strength of relationship, a predictive equation is calculated.

The Tests:

t-test:

A *t*-test is used to determine whether the means of two groups are statistically different from each other. If the means are statistically different, then the null hypothesis (H_0) is rejected and the alternate or research hypothesis (H_1) is accepted. The *t*-test can also be carried out on a single sample by comparing a study's group mean with a known value. For instance, a study involving cervical ROM could compare the resulting mean with a known normal value for cervical flexion.

The *t*-test produces a test statistic, termed the *t*-score or *t*-ratio. This value is used to find the corresponding p-value by means of a *t*-table (see below). The *t*-ratio is comprised of the difference of the means being tested divided by the variability of the data, which is represented by something called the standard error of the difference (you don't have to know this term or the following equation). See equation below:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{\bar{X}_1 - \bar{X}_2}}$$

The following partial reproduction of a *t*-table is used to determine the probability associated with a given *t*-score, taking into consideration the number of subjects in the groups under study (termed degrees of freedom [df], which equals the number of subjects in each group minus 1).

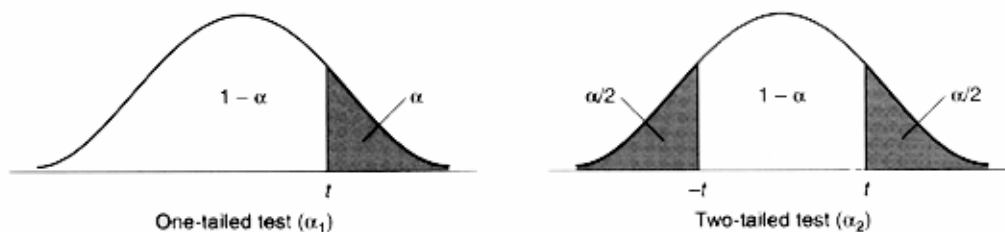


TABLE A.2 CRITICAL VALUES OF *t*

<i>df</i>	α_1	.10	.05	.025	.01	.005	.0005
	α_2	.20	.10	.05	.02	.01	.001
1		3.078	6.314	12.706	31.821	63.657	636.619
2		1.886	2.920	4.303	6.965	9.925	31.598
3		1.638	2.353	3.182	4.541	5.841	12.924
4		1.533	2.132	2.776	3.747	4.604	8.610
5		1.476	2.015	2.571	3.365	4.032	6.859
6		1.440	1.943	2.447	3.143	3.707	5.959
7		1.415	1.895	2.365	2.998	3.499	5.405
8		1.397	1.860	2.306	2.896	3.355	5.041
9		1.383	1.833	2.262	2.821	3.250	4.781
10		1.372	1.812	2.228	2.764	3.169	4.587
11		1.363	1.796	2.201	2.718	3.106	4.437
12		1.356	1.782	2.179	2.681	3.055	4.318
13		1.350	1.771	2.160	2.650	3.012	4.221
14		1.345	1.761	2.145	2.624	2.977	4.140
15		1.341	1.753	2.131	2.602	2.947	4.073

Directional vs. Nondirectional hypotheses: In a nondirectional hypothesis, we would be interested in knowing if the scores differ, but (before the analysis) have not made any statements about whether the score will be higher or lower. The hypothesis would be directional if we had stated that the scores of one group would be higher than the other group. With a nondirectional test alpha is split between the two tails of the distribution and with a directional test alpha is all in one tail. As you will see, it is much easier to reach statistical significance with a directional (one-tail) test than a nondirectional test. However, in order to make a statement that one set of scores will be higher than the other; the other direction must be of no interest. A statement we can rarely make in chiropractic research because it is typically possible that the test could go either way. This topic is somewhat controversial.

Assumptions: There are several assumptions that must be met in order to legitimately perform a *t*-test:

- Samples must be randomly drawn from the parent population
- Data in each group must have a normal distribution
- The variances for each of the groups must be equal (otherwise the *t*-test for unequal variances should be used)
- The samples must be independent. That is, subjects included in the second sample don't depend on those who are included in the first sample (and vice versa)

There are statistical tests that can determine whether or not the data in a given study are normal and whether variances are equal. Most statistical software packages, including Microsoft Excel, are capable of running these tests for you.

An example of how a research paper might report the results of *t*-tests would be something like this: "Chiropractic and PT for lower back pain and disability were compared using an independent *t*-test for the resulting Oswestry Disability Questionnaire scores. A significant difference between the two outcomes was found, with chiropractic being superior ($t(28) = 17.86, p < 0.002$)."

The information that should be presented is: the statistic used, the degrees of freedom of the actual analysis, the obtained value of the statistic (the *t*-score etc.) and the probability of the results (the *p*-value).

$$t^1(28)^2=17.98^3;p<0.002^4$$

1. The statistic used, in this case a *t*-test
2. The degrees of freedom of the analysis
3. The actual obtained value of the statistic
4. The probability associated with that value

The following table contains data from a hypothetical study that compares chiropractic care with physical therapy (two groups) regarding the number of visits required. These data would be analyzed using a *t*-test:

Chiropractic Care		Physical therapy	
<u>Case #</u>	<u>Total Visits</u>	<u>Case #</u>	<u>Total Visits</u>
C1	7	PT1	12
C2	2	PT2	8
C3	2	PT3	4
C4	3	PT4	5
C5	4	PT5	7
C6	3	PT6	5
C7	5	PT7	9
C8	3	PT8	4
C9	4	PT9	4
C10	6	PT10	8
C11	2	PT11	4
C12	3	PT12	8
C13	7	PT13	9
C14	4	PT14	7

This is the print-out from running a *t*-test on the above data using Microsoft Excel:

t-Test: Two-Sample Assuming Unequal Variances		
	<i>PT</i>	<i>Chiro</i>
Mean	6.71	3.93
Variance	6.07	2.99
Observations	14	14
Df	23	
t Stat	3.46	
P one-tail	0.001	
t Critical one-tail	1.71	
P two-tail	0.002	
t Critical two-tail	2.07	

Some researchers use a directional test in similar circumstances because they propose that PT treatment will require more visits. However, since it is possible that PT treatment may require the same or even less visits, a non-directional test is required. In order to use a directional test you must have good reason to propose that the results will only go in one direction.

Another form of the *t*-test, called the **paired *t*-test**, is carried out on dependant samples. This test is used to compare means on the same or related subjects over time or in differing circumstances. For instance, repeated measures studies use the paired *t*-test to compare measurements obtained initially and then at follow-up evaluations. Subjects are often tested in a before-after situation (across time, with some intervention occurring such as chiropractic adjustments), or subjects are paired such as with twins, or with subject as alike as possible. The assumptions for this test are

that the observed data are from the same subject or from a matched subject and are drawn from a population with a normal distribution.

Steps involved in the *t*-test

- After gathering the data, calculate the means and standard deviations.
- Calculate the *t*-ratio.
- Look up the resulting *t*-statistic in a *t*-table.
- If the probability of *t* is less than 0.05 (which is the typical value for alpha), it is significant.
- You can then conclude that the groups are different.

Remember:

- Big *t*-value → Small p-value → SIGNIFICANCE

<i>t</i>-test Summary		
<u>Number of subjects</u>	<u>Assumptions</u>	<u>Results</u>
The more subjects you have-the easier it is to find significance . You will have more power to find differences. The number of subjects determines degrees of freedom: df = N₁ + N₂ - 2 for Independent groups and N_{pairs} - 1 for repeated or matched groups	The two groups should be normal in shape. They should have the same standard deviation . If they are not, the chance of a Type I error increases (incorrectly saying there is a difference between the groups).	A large <i>t</i>-value and a small p-value are needed to reach statistical significance.

ANOVA:

Used when comparing means involving more than two groups.

To carry out ANOVA, the mean and standard deviation for each of the groups involved in the study must be calculated. The analysis of variance will then tell us three things:

- Whether there are any significant differences of means among all of the groups. Provided by the **F -ratio**.
- Whether any of the specific groups differ from each other. There are more than one possible pairs of groups when you have more than two. This is provided by using Comparison Techniques.
- Whether the differences are relatively big or small. Measures of Explained Variance will tell us this.

The F-ratio is similar to the *t*-ratio, however, the F-ratio compares the variance between the groups with the variance within the groups.

$$F = \frac{\text{variance between groups}}{\text{variance within groups}}$$

The within-group variance is related to sampling error, subject differences, etc. It's like the variance we previously discussed. Between-group variance is due to the differences between the groups. If the means of the groups differ significantly, then there will be an associated between groups variance that is high. If the F-ratio is **small**, then groups are probably **not significantly different** from each other. However, if it is **big**, then some (two or more) **might be significantly different** from each other.

We cannot simply calculate repetitive *t*-tests for each group comparison when more than 2 groups are involved because the Type 1 error rate (saying there is a difference between means when there isn't one) will dramatically increase. While we're on the subject, Type 2 error is the opposite, saying there isn't a difference between means when there really is one. The F-Ratio avoids this problem because it is an overall test. A significant F means that there are some differences between some of the groups, but at that point, it is not known which particular groups differ. That's why Comparison Techniques are used, which compare the individual group pairs.

F-tables are available that are similar to *t*- tables, but are specific for the F-ratio. The F-table is consulted to determine if there is a significant difference among the means. Again, computer programs are available to do all this work for you. That's one reason formulas and associated calculations have not been included.

The following table compares chiropractic care, physical therapy, and placebo (three groups) and would be analyzed using ANOVA:

Chiropractic Care		Physical therapy		Acupuncture	
<u>Case #</u>	<u>Total Visits</u>	<u>Case #</u>	<u>Total Visits</u>	<u>Case #</u>	<u>Total Visits</u>
C1	7	PT1	12	A1	8
C2	2	PT2	8	A2	6
C3	2	PT3	4	A3	4
C4	3	PT4	5	A4	10
C5	4	PT5	7	A5	3
C6	3	PT6	5	A6	3
C7	5	PT7	9	A7	12
C8	3	PT8	4	A8	5
C9	4	PT9	4	A9	3
C10	6	PT10	8	A10	5
C11	2	PT11	4	A11	6
C12	3	PT12	8	A12	4
C13	7	PT13	9	A13	7
C14	4	PT14	7	A14	5

In this study, it is anticipated that chiropractors can manage patients with fewer visits than PTs and acupuncturists, so the hypothesis in this study is that patients will require more visits if treated with physical therapy or acupuncture than similar patients under chiropractic care.

The following Excel printout concerning this data is in the form of a table; aptly called an ANOVA table. Also, automatically included in this printout is a summary table showing some descriptive statistics:

Anova: Single Factor

SUMMARY					
<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>	
Chiro	14	55	3.93	2.99	
PT	14	94	6.71	6.07	
AC	14	81	5.79	7.26	

ANOVA						
<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Between Groups	56.33	2	28.17	5.18	0.01	3.24
Within Groups	212.14	39	5.44			
Total	268.48	41				

Chi-square

The chi-square test allows us to test hypotheses using nominal or categorical data (i.e., classifications or categories). What chi-square does is test whether one set of proportions is different from another by comparing frequencies.

Chi-square goodness of fit:

In a random sample of 100 Americans, we would find (on average) about 50 males and 50 females. However, if a sample found 60 males and 40 females, would that finding be significantly different from normal? Would the observed values in the sample be statistically different from the expected values? Chi-square's purpose is to determine how a set of proportions that are thought to be true (expected) compare with a set of data that have been observed (observed).

The Chi-square formula compares the difference between the observed and expected frequencies, divided by the expected frequencies.

$$X^2 = \frac{\Sigma(O - E)^2}{E}$$

(Where O = the observed frequency and E = the expected frequency)

Considering the sample of 100 Americans with 60 males and 40 females, the chi-square statistic can be calculated as follows:

$$X^2 = \frac{(60 - 50)^2}{50} + \frac{(40 - 50)^2}{50} = 4.0$$

The chi-square statistic is evaluated for statistical significance using a chi-square table, with df equal to the number of categories minus 1. If the tabled value is exceeded, then the test is significant (just like in the *t*-test and F-test). This particular chi-square test is sometimes called the goodness of fit test because it tests whether “Observed” frequencies fit against the “Expected” frequencies.

Chi-square Test of Independence:

This test is done to see whether the breakdown of proportions or frequencies for one category differ significantly from that of another category. The **Chi-square Test of Independence** uses a contingency table (also known as a cross-tabulation table) that lays out the corresponding variables. For instance, in a study where patients were administered Gonstead technique or Diversified, we could measure an outcome variable such as pain or no pain (either the patients end-up in pain or they don't). The table (and this is a **contingency table**) would be set up as follows:

		Pain		Row Total
		Yes	No	
Technique	Gonstead	4	16	20
	Diversified	16	4	20
Column Total		20	20	40

The expected values are calculated as follows:

$$\text{Expected (E)} = \frac{\text{Row Total} \times \text{Column Total}}{\text{Grand Total}}$$

		Pain		Row Total
		Yes	No	
Technique	Gonstead	$E=20*20/40=10$	$E=20*20/40=10$	20
	Diversified	$E=20*20/40=10$	$E=20*20/40=10$	20
Column Total		20	20	40

To calculate chi-square:

$\frac{(4-10)^2}{10}$	$\frac{(16-10)^2}{10}$
$\frac{(16-10)^2}{10}$	$\frac{(4-10)^2}{10}$

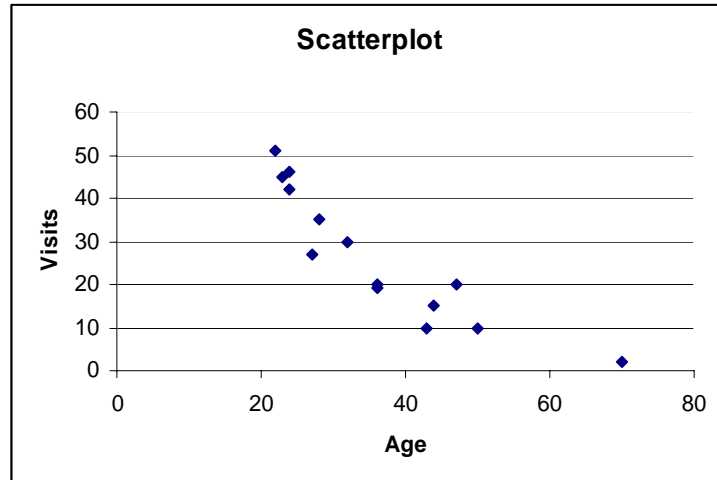
Adding the cells produces the resulting chi-square statistic, which is 14.4 in this case. According to the chi-square table all that is needed for significance at the .05 level is 3.84, so the test is definitely significant. The df for chi-square is equal to the number of categories for variable 1 minus 1 times the number of categories for variable 2 minus 1. So, in this case there is only 1 df.

The bottom line in this hypothetical example is that the presence of pain is dependant upon the technique that was used. In other words resulting pain is not independent of the technique used.

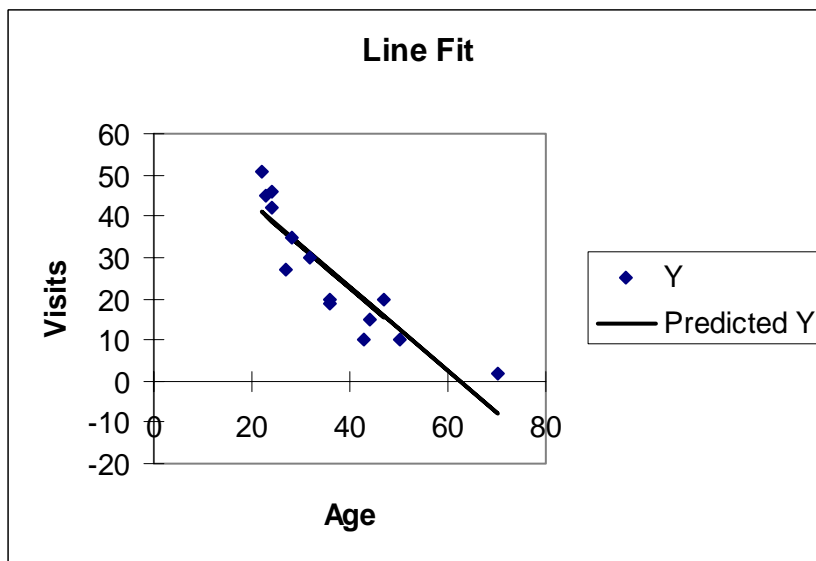
Pearson's correlation and Regression

The purpose of a **Correlation Coefficient** (r) is to show the strength of a linear relationship between two variables. In other words, if you have two sets of scores: Are they related? If so, how strong is the relationship? Does one score predict the other? For example, the relationship between patient age and the number of visits they received during a study can be examined as follows:

Age	Visits
23	45
24	46
47	20
28	35
36	20
36	19
43	10
32	30
24	42
22	51
50	10
70	2
44	15
27	27



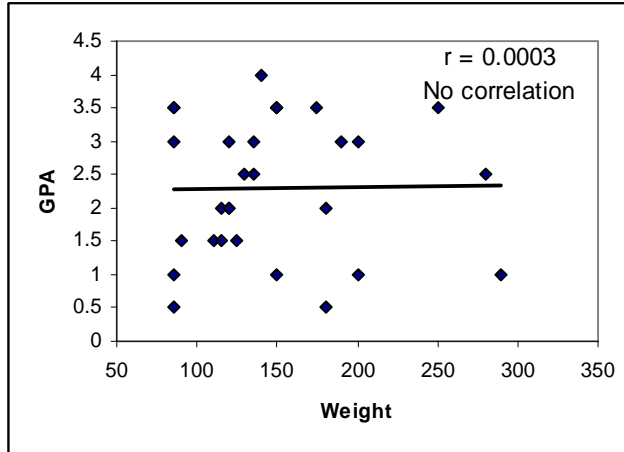
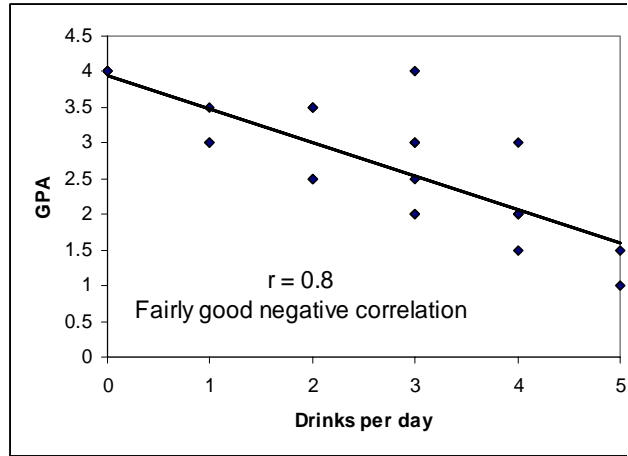
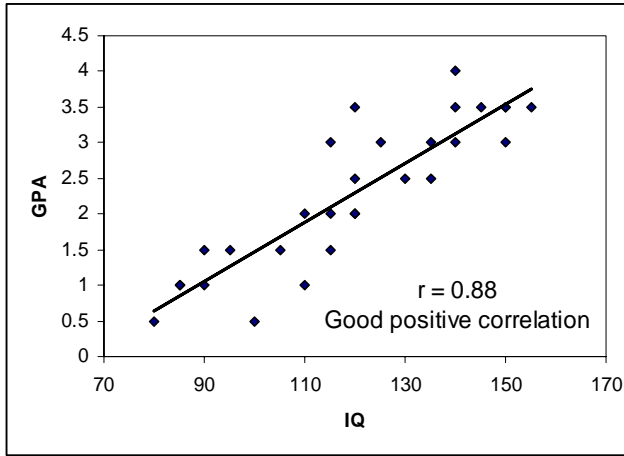
There is a negative relationship between age and the number of visits. As age increases, visits decrease.
 $r = -0.897$



This is a computer generated **regression** line that best predicts where a visit should appear on the plot for a given age. The regression line is an average of all the graphed data points.

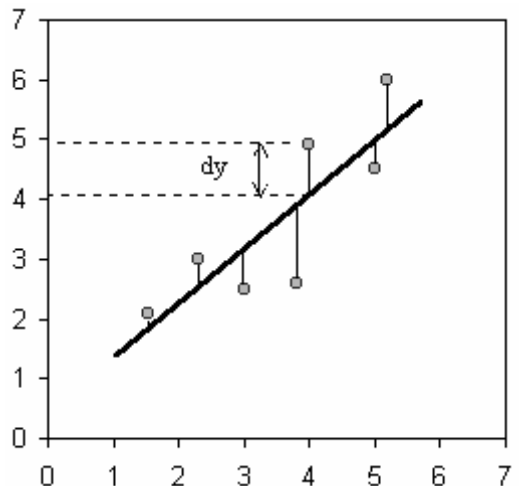
Values for r range from -1 to $+1$. A value of zero (0.0) specifies no relationship between the variables, while values approaching -1 or $+1$ indicate predicative relationships. ± 1 is the strongest possible relationship, with the sign indicating the direction of relationship.

Following are examples of positive correlation, no correlation, and negative correlation:



R^2 can be calculated by squaring the correlation coefficient. This provides the proportion of explained variance; in other words, how much of the variation of the Y variable can be explained by variation of the X variable. In the previous example of age and visits $r = -0.897$ and R^2 (the proportion of explained variance) = 0.805. R^2 could be multiplied by 100 to produce the percent explained variance and we could say “80% of the variation of visits can be explained by age.”

A **regression line** is calculated mathematically to represent a straight line represented by the formula of a line $Y=a+bx$, where a is the y-intercept and b is the slope (similar to the equation of a sloped line you may remember from algebra $Y=mX+b$). The line is fitted to a set of data by minimizing the sum of squared deviations (dy) of the observed data points from the line. Conceptually, it’s almost like finding the average of these deviations, but involving fairly complicated math. Again, let’s leave it up to the computer to calculate as I did up above.



Statistical Test Decision Table (from Intuitive Biostatistics by Harvey Motulsky)

Goal	Type of Data		
	Measurement from Gaussian Population*	Rank, Score, or Measurement from Non-Gaussian Population	Binomial (Two Possible Outcomes)
Describe one group	Mean, SD	Median, interquartile range	Proportion
Compare one group to a hypothetical value	One-sample <i>t</i> test	Wilcoxon test	Chi-square or Binomial test
Compare two unpaired groups	Unpaired <i>t</i> test	Mann-Whitney test	Fisher's test (chi-square for large samples)
Compare two paired groups	Paired <i>t</i> test	Wilcoxon test	McNemar's test
Compare three or more unmatched groups	One-way ANOVA	Kruskal-Wallis test	Chi-square test
Compare three or more matched groups	Repeated-measures ANOVA	Friedman test	Cochrane Q
Quantify association between two variables	Pearson correlation	Spearman correlation	Contingency coefficients
Predict value from another measured variable	Simple linear regression or Nonlinear regression	Nonparametric regression	Simple logistic regression
Predict value from several measured or binomial variables	Multiple linear regression or Multiple nonlinear regression		Multiple logistic regression

* Gaussian Population = normally distributed population