

	Interval/Ratio (Normality assumed)	Interval/Ratio (Normality not assumed), Ordinal	Dichotomy (Binomial)
Compare two unpaired groups	Unpaired t test	Mann-Whitney test	Fisher's test
Compare two paired groups	Paired t test	Wilcoxon test	McNemar's test
Compare more than two unmatched groups	ANOVA	Kruskal-Wallis test	Chi-square test
Compare more than two matched groups	Repeated-measures ANOVA	Friedman test	Cochran's Q test
Find relationship between two variables	Pearson correlation	Spearman correlation	Cramer's V
Predict a value with one independent variable	Linear/Non-linear regression	Non-parametric regression	Logistic regression
Predict a value with multiple independent variables or binomial variables	Multiple linear/non-linear regression		Multiple logistic regression

Paired vs. unpaired t-test

There is one important thing you need to understand in a t test: Paired and unpaired. Unpaired means that you simply compare the two groups. So, you will build a model for each group (calculate the mean and variance), and see whether there is a difference. Paired means that you will look at the differences between the two groups. A paired test first calculates the difference from one group to the other, and runs a one-sample t test. So, how should you decide which one to use? It depends on your experimental design. If you use a within-subject design, you should use a paired t test. Because each participant contributed to a data point for each of the two groups (e.g., two interaction techniques), your data have an implicit relationship based on your participants. For example, some participants were slower than other participants in both techniques, but all of the participants were faster with one technique than with the other. paired t test considers such individual differences by taking the differences between the two groups. Otherwise, you can't really have a legitimate way to take the differences, so you need to use an unpaired t test.

A paired t test

You should use a paired t test if you do a within-subject design. What a paired t test does is to take differences between data in the two groups, and see whether the distribution of the differences is too different from the t distribution. Because it uses the differences between the groups, **a paired t test does not assume the variances of the population of the two groups are equal**. But it still assumes the normality. The null hypothesis is there is no significant difference in the means between the two groups. If the p value is less than 0.05, you reject the null hypothesis, and say that you find a significant difference.

An unpaired t test

If you are going to use an unpaired t test, you need to consider another assumption, which is about the homogeneity of variances: **The variances of the population of the two groups are equal**. This is important for an unpaired t test. However, we have a t test which can accommodate the unequal variances, which is called a Welch's t test. Unless you can make sure that the variances of the population of the two groups are equal, you can simply use a Welch's t test without thinking too

much. This is fair because a Welch's t test is generally a more strict test than a standard t test (*i.e.*, your p value with a Welch's t test becomes higher than one with a standard t test).

A t test has a hypothesis, called null hypothesis. The null hypothesis is **there is no significant difference in the means between the two groups**. If the p value is less than 0.05, you reject the null hypothesis, and say that you find a significant difference.

As you can see in the following example with R, the beauty of a t test is its simplicity. This is one reason why I like a t test. It is less likely to mess up an analysis compared to other kinds of statistical tests. I recommend you to think about your experiment so that you can do this nice test.

ANOVA

Analysis of Variance (ANOVA) is one statistical test commonly used in HCI research. Although it says “analysis of variance,” and ANOVA is a method to compare multiple linear models, a very common way to use an ANOVA test is to test the difference in means among more than two groups. So, the intuition is a [t test](#) which accommodates more than two groups to compare. For example, when you want to compare user performance of three interaction techniques, ANOVA is one of the methods you want to consider.

But, unfortunately, ANOVA is not as simple as this intuition, and understanding ANOVA is fairly hard (well, at least I think so. I don't think I fully understand it yet.). **Particularly, doing a two-way repeated-measure ANOVA test is not easy in R, and you may want to use other statistics software, such as SPSS.** It is not like it is impossible, but it is very complicated, and you may have to have some knowledge about linear models (and manipulations for these in R).

As you can see in the table of content, there are many things you need to learn in ANOVA. Nonetheless, you can do an ANOVA test with your statistical software without necessarily knowing all its details, and this is probably why you can see so many different ways to do ANOVA in HCI papers. In this page, I do my best to explain my understandings of ANOVA, and provide the procedures of ANOVA recommended by other literature. But if you have time, I recommend you to read a statistics book to understand ANOVA (and to double-check what I am saying here) as well as this page.

Why not a t test?

One simple question you may have is “why cannot we use multiple t tests instead of ANOVA?” This is a very understandable question, and you could do so with some corrections (and this is what we often do in a post-hoc comparison). However, if you have more than two groups to compare, using ANOVA is a de-facto standard.

There is one clear reason why we cannot use multiple t test instead of ANOVA for saying whether a factor which has more than two groups (e.g., three different techniques) has a significant effect on the dependent variable (e.g., performance time). Let's say, you have Technique A, Technique B, and Technique C, and you have measured performance time of these three techniques. What you want to know here is whether these techniques have a significant effect on performance time, that is whether there exists **at least one significant difference** in performance time against the techniques. In other words, your null hypothesis is that there is no significant difference in the means among three techniques.

If you try to do multiple t tests, you have to do three tests: One test between Technique A and Technique B, one between B and C, and one between A and C. If you test the t test null hypothesis (i.e., there is no difference in the means between between Technique A and Technique B) with 95% confidence (i.e., $p=0.05$) and you cannot reject it, your confidence of that rejection is 95%. Remember that your null hypothesis you really need to test is **there is no significant difference in the means among three techniques**. You need to do two more tests for B and C, and A and C. If all three tests reject the null hypothesis, you can say that there is no significant difference in the means among three techniques. But this also means that your confidence level becomes $95\% * 95\% * 95\% = 85.7\%$ because you do the AND operation on your tests. Obviously, a test with 85.7% confidence is not useful for us. Thus, you cannot simply do multiple t tests if you have more than two groups to compare.

By the way, what if we compare two groups with ANOVA? The results will be the same as what you can get with an unpaired/paired t test. So it is ok to use ANOVA for two groups, but using t tests is more common.

One-way ANOVA

One-way ANOVA means you have one independent variable to use for the comparison of the dependent variables. As we have seen in the examples above, we use one-way ANOVA for testing Value against Group. Thus, what you are looking for is the relationship between Value and Group, and this can be expressed as **Value ~ Group** in R. What one-way ANOVA does is to see whether Group is really important to predict Value.

Two-way ANOVA

Two-way ANOVA means you have two kinds of independent variables to use for your comparison. So it just sound like you have one more independent variable in one-way ANOVA. But it is not quite true, and there are some complicated stuff. One of these stuff is the type of the sum of squares.

Mixed-design ANOVA

Mixed-design means that you have both within-subject and between-subject factors. Let's think about this with the same examples for two-way ANOVA. You have two input devices (Pen and

Touch), and three techniques for each device (A, B, and C). Let's say you recruited different people for the devices: 8 people for Pen and other 8 people for Touch. But within each device, every participant was exposed to the three techniques. In this case, Device is a between-subject factor, and Technique is a within-subject factor.