UNIT - IV

F-TEST

F-TEST:

An F-test is any statistical test in which the test statistic has an F-distribution under the null hypothesis. It is most often used when comparing statistical models that have been fitted to a data set, in order to identify the model that best fits the population from which the data were sampled.

General Steps for an F Test

F test by hand, including variances, is tedious and time-consuming. Therefore you'll probably make some errors along the way.

1. State the null hypothesis and the alternate hypothesis.
2. Calculate the F value. The F Value is calculated using the formula F = (SSE1 – SSE2 / m) / SSE2 / n-k, where SSE = residual sum of squares, m = number of restrictions and k = number of independent variables.
3. Find the F Statistic (the critical value for this test). The F statistic formula is:
   F Statistic = variance of the group means / mean of the within group variances.
   You can find the F Statistic in the F-Table.
4. Support or Reject the Null Hypothesis.

F Test to Compare Two Variances

A Statistical F Test uses an F Statistic to compare two variances, s1 and s2, by dividing them. The result is always a positive number (because variances are always positive). The equation for comparing two variances with the f-test is:
F = s21 / s22

If the variances are equal, the ratio of the variances will equal 1. For example, if you had two data sets with a sample 1 (variance of 10) and a sample 2 (variance of 10), the ratio would be 10/10 = 1.

You always test that the population variances are equal when running an F Test. In other words, you always assume that the variances are equal to 1. Therefore, your null hypothesis will always be that the variances are equal.

Assumptions

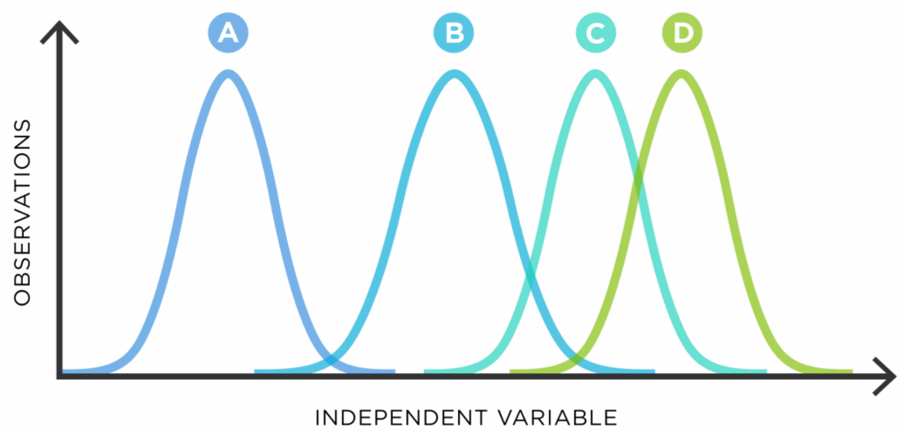
Several assumptions are made for the test. Your population must be approximately normally distributed (i.e. fit the shape of a bell curve) in order to use the test. Plus, the samples must be independent events. In addition, you'll want to bear in mind a few important points:

- The larger variance should always go in the numerator (the top number) to force the test into a right-tailed test. Right-tailed tests are easier to calculate.
- For two-tailed tests, divide alpha by 2 before finding the right critical value.

- If you are given standard deviations, they must be squared to get the variances.
- If your degrees of freedom aren't listed in the F Table, use the larger critical value. This helps to avoid the possibility of Type I errors.

Analysis of Variance (ANOVA):

Analysis of Variance (**ANOVA**) is a statistical formula used to compare variances across the means (or average) of different groups. A range of scenarios use it to determine if there is any difference between the means of different groups.



For example, to study the effectiveness of different diabetes medications, scientists design and experiment to explore the relationship between the type of medicine and the resulting blood sugar level. The sample population is a set of people. We divide the sample population into multiple groups, and each group receives a particular medicine for a trial period. At the end of the trial period, blood sugar levels are measured for each of the individual participants. Then for each group, the mean blood sugar level is calculated. ANOVA helps to compare these group means to find out if they are statistically different or if they are similar.

The outcome of ANOVA is the 'F statistic'. This ratio shows the difference between the within group variance and the between group variance, which ultimately produces a figure which allows a conclusion that the null hypothesis is supported or rejected. If there is a significant difference between the groups, the null hypothesis is not supported, and the F-ratio will be larger.

ANOVA Terminology

Dependent variable: This is the item being measured that is theorized to be affected by the independent variables.

Independent variable/s: These are the items being measured that may have an effect on the dependent variable.

A null hypothesis (H0): This is when there is no difference between the groups or means. Depending on the result of the ANOVA test, the null hypothesis will either be accepted or rejected.

An alternative hypothesis (H1): When it is theorized that there is a difference between groups and means.

Factors and levels: In ANOVA terminology, an independent variable is called a factor which affects the dependent variable. Level denotes the different values of the independent variable that are used in an experiment.

Fixed-factor model: Some experiments use only a discrete set of levels for factors. For example, a fixed-factor test would be testing three different dosages of a drug and not looking at any other dosages.

Random-factor model: This model draws a random value of level from all the possible values of the independent variable.

One-Way ANOVA

The one-way analysis of variance is also known as single-factor ANOVA or simple ANOVA. As the name suggests, the one-way ANOVA is suitable for experiments with only one independent variable (factor) with two or more levels. For instance a dependent variable may be what month of the year there are more flowers in the garden. There will be twelve levels. A one-way ANOVA assumes:

- Independence: The value of the dependent variable for one observation is independent of the value of any other observations.
- Normalcy: The value of the dependent variable is normally distributed
- Variance: The variance is comparable in different experiment groups.
- Continuous: The dependent variable (number of flowers) is continuous and can be measured on a scale which can be subdivided.

Full Factorial ANOVA (also called two-way ANOVA)

Full Factorial ANOVA is used when there are two or more independent variables. Each of these factors can have multiple levels. Full-factorial ANOVA can only be used in the case of a full factorial experiment, where there is use of every possible permutation of factors and their levels. This might be the month of the year when there are more flowers in the garden, and then the number of sunshine hours. This two-way ANOVA not only measures the independent vs the independent variable, but if the two factors affect each other. A two-way ANOVA assumes:

- Continuous: The same as a one-way ANOVA, the dependent variable should be continuous.
- Independence: Each sample is independent of other samples, with no crossover.
- Variance: The variance in data across the different groups is the same.
- Normalcy: The samples are representative of a normal population.
- Categories: The independent variables should be in separate categories or groups.

ANOVA’S working

Some people question the need for ANOVA; after all, mean values can be assessed just by looking at them. But ANOVA does more than only comparing means.

Even though the mean values of various groups appear to be different, this could be due to a sampling error rather than the effect of the independent variable on the dependent variable. If it is due to sampling error, the difference between the group means is meaningless. ANOVA helps to find out if the difference in the mean values is statistically significant.

ANOVA also indirectly reveals if an independent variable is influencing the dependent variable. For example, in the above blood sugar level experiment, suppose ANOVA finds that group means are not statistically significant, and the difference between group means is only due to sampling error. This result infers that the type of medication (independent variable) is not a significant factor that influences the blood sugar level.

Limitations of ANOVA

ANOVA can only tell if there is a significant difference between the means of at least two groups, but it can’t explain which pair differs in their means. If there is a requirement for granular data, deploying further follow up statistical processes will assist in finding out which groups differ in mean value. Typically, ANOVA is used in combination with other statistical methods.

ANOVA also makes assumptions that the dataset is uniformly distributed, as it compares means only. If the data is not distributed across a normal curve and there are outliers, then ANOVA is not the right process to interpret the data.

Similarly, ANOVA assumes the standard deviations are the same or similar across groups. If there is a big difference in standard deviations, the conclusion of the test may be inaccurate.

Effect size for a between groups ANOVA

Calculating effect size for between groups designs is much easier than for within groups. The formula looks like this:

$$\eta^2 = \frac{\text{Treatment Sum of Squares}}{\text{Total Sum of Squares}}$$

So if we consider the output of a between groups ANOVA (using SPSS/PASW)

ANOVA

RECALL

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	31.444	2	15.722	7.447	.006
Within Groups	31.667	15	2.111		
Total	63.111	17			

Looking at the table above, we need the second column (Sum of Squares).

The treatment sum of squares is the first row: Between Groups (31.444)

The total sum of squares is the final row: Total (63.111)

Therefore:

$\eta^2 = \underline{31.444}$

63.111

$\eta^2 = 0.498$

This would be deemed by Cohen's guidelines as a very large effect size; 49.8% of the variance was caused by the IV (treatment).

Effect size for a within subjects ANOVA

The formula is slightly more complicate

d here, as you have to work out the total Sum of Squares yourself:

Total Sum of Squares = Treatment Sum of Squares + Error Sum of Squares + Error (between subjects) Sum of Squares.

Then, you'd use the formula as normal.

$\eta^2 =$ Treatment Sum of Square

Total Sum of Squares

Let's look at an example:

(Again, output 'borrowed' from my lecture slides as PASW is being mean!)

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
SPEED1	Sphericity Assumed	31.444	2	15.722	7.183	.012
	Greenhouse-Geisser	31.444	1.732	18.158	7.183	.017
	Huynh-Feldt	31.444	2.000	15.722	7.183	.012
	Lower-bound	31.444	1.000	31.444	7.183	.044
Error(SPEED1)	Sphericity Assumed	21.889	10	2.189		
	Greenhouse-Geisser	21.889	8.658	2.528		
	Huynh-Feldt	21.889	10.000	2.189		
	Lower-bound	21.889	5.000	4.378		

Tests of Between-Subjects Effects

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Eta Squared	Noncent. Parameter	Observed Power[a]
Intercept	600.889	1	600.889	307.273	.000	.984	307.273	1.000
Error	9.778	5	1.956					

a. Computed using alpha = .05

So, the **total** Sum of Squares, which we have to calculate, is as follows:

31.444 (top table, SPEED 1) + 21.889 (top table, Error(SPEED1)) + 9.778 (Bottom table, Error) = 63.111

As you can see, this value is the same as the last example with between groups – so it works!

Just enter the total in the formula as before:

$\eta^2 = \frac{31.444}{63.111} = 0.498$

Again, 49.8% of the variance in the DV is due to the IV.

TWO FACTORS

Systematic factors

have statistical inflence on the data set.

Random factors

does not have statistical inflence on the data set.

Multiple Comparisons

If our test of the null hypothesis is rejected, we conclude that not all the means are equal: that is, at least one mean is different from the other means. The ANOVA test itself provides only statistical evidence of a difference, but not any statistical evidence as to which mean or means are statistically different.

For instance, using the previous example for tar content, if the ANOVA test results in a significant difference in average tar content between the Perfume brands, a follow up analysis would be needed to determine which brand mean or means differ in tar content. Plus we would want to know if one brand or multiple brands were better/worse than another brand in average tar content. To complete this analysis we use a method called **multiple comparisons.**

Multiple comparisons conducts an analysis of all possible pairwise means. For example, with three brands of Perfume, A, B, and C, if the ANOVA test was significant, then multiple comparison methods would compare the three possible pairwise comparisons:

- Brand A to Brand B
- Brand A to Brand C
- Brand B to Brand C

Repeated Measures ANOVA: Definition, Formula, and Example

A **repeated measures ANOVA** is used to determine whether or not there is a statistically significant difference between the means of three or more groups in which the same subjects show up in each group.

A repeated measures ANOVA is typically used in two specific situations:

1. Measuring the mean scores of subjects during three or more time points.

For example, you might want to measure the resting heart rate of subjects one month before they start a training program, during the middle of the training program, and one month after the training program to see if there is a significant difference in mean resting heart rate across these three time points.

Subject	Resting Heart Rate 1 Month Before Training Program	Resting Heart Rate in Middle of Training Program	Resting Heart Rate 1 Month After Training Program
Michael	65	58	60
Dwight	55	48	49
Andy	58	55	55
Meredith	68	60	64
Angela	47	45	45

Notice how the same subjects show up at each time point. We repeatedly measured the same subjects, hence the reason why we used a repeated measures ANOVA.

2. Measuring the mean scores of subjects under three different conditions. For example, you might have subjects watch three different movies and rate each one based on how much they enjoyed it.

Subject	Movie 1 Rating	Movie 2 Rating	Movie 3 Rating
Michael	88	84	92
Dwight	76	78	90
Andy	78	94	95
Meredith	80	83	88
Angela	82	90	99

Again, the same subjects show up in each group, so we need to use a repeated measures ANOVA to test for the difference in means across these three conditions.

One-Way ANOVA vs. Repeated Measures ANOVA

In a typical one-way ANOVA, different subjects are used in each group. For example, we might ask subjects to rate three movies, just like in the example above, but we use different subjects to rate each movie:

Subject	Movie 1 Rating	Subject	Movie 2 Rating	Subject	Movie 3 Rating
Michael	88	Jim	84	Stanley	92
Dwight	76	Pam	78	Creed	90
Andy	78	Toby	94	Ryan	95
Meredith	80	Kevin	83	Holly	88
Angela	82	Erin	90	Kelly	99

In this case, we would conduct a typical one-way ANOVA to test for the difference between the mean ratings of the three movies.

In real life there are two benefits of using the same subjects across multiple treatment conditions:

1. It's cheaper and faster for researchers to recruit and pay a smaller number of people to carry out an experiment since they can just obtain data from the same people multiple times.

2. We are able to attribute some of the variance in the data to the subjects themselves, which makes it easier to obtain a smaller p-value.

One potential drawback of this type of design is that subjects might get bored or tired if an experiment lasts too long, which could skew the results. For example, subjects might give lower movie ratings to the third movie they watch because they're tired and ready to go home.

Repeated Measures ANOVA: Example

Suppose we recruit five subjects to participate in a training program. We measure their resting heart rate before participating in a training program, after participating for 4 months, and after participating for 8 months.

The following table shows the results:

Subject	Resting Heart Rate Before Training Program	Resting Heart 4 Months Into Training Program	Resting Heart Rate 8 Months Into Training Program
Michael	65	58	60
Dwight	55	48	44
Andy	58	55	55
Meredith	68	60	55
Angela	47	45	45

We want to know whether there is a difference in mean resting heart rate at these three time points so we conduct a repeated measures ANOVA at the .05 significance level using the following steps:

Step 1. State the hypotheses.

The null hypothesis (H0): $\mu1 = \mu2 = \mu3$ (the population means are all equal)

The alternative hypothesis: (Ha): at least one population mean is different from the rest

Step 2. Perform the repeated measures ANOVA.

We will use the Repeated Measures ANOVA Calculator using the following input:

One-Way Repeated Measures ANOVA Calculator

The one-way repeated measures ANOVA calculator compares the means of three or more samples in which each subject shows up in each sample.

Simply enter the values for up to five samples into the cells below, then press the "Calculate" button.

Group 1	Group 2	Group 3	Group 4	Group 5
65	58	60		
55	48	44		
58	55	55		
68	60	55		
47	45	45		

Once we click "Calculate" then the following output will automatically appear:

Source	SS	df	MS	F	P
Between	128.9	2	64.5	9.598	0.00749
Subject	585.1	4	146.3		
Error	53.7	8	6.7		

Step 3. Interpret the results.

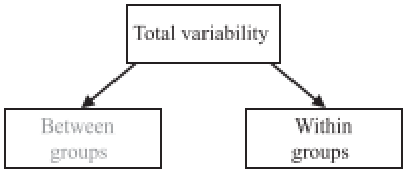
From the output table we see that the F test statistic is **9.598** and the corresponding p-value is **0.00749**.

Since this p-value is less than 0.05, we reject the null hypothesis. This means we have sufficient evidence to say that there is a statistically significant difference between the mean resting heart rate at the three different points in time.

THREE *F* TESTS

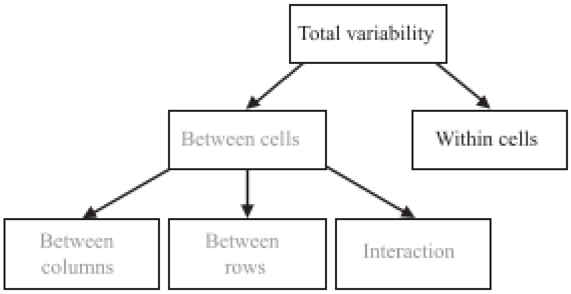
As suggested in diagram, *F* ratios in both a one- and a two-factor ANOVA always consist of a numerator (shaded) that measures some aspect of variability between.

ONE-FACTOR ANOVA



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

TWO-FACTOR ANOVA



$$F_{column} = \frac{\text{Between columns}}{\text{Within cells}}$$

$$F_{row} = \frac{\text{Between rows}}{\text{Within cells}}$$

$$F = \frac{\text{Interaction}}{\text{Within cells}}$$

Sources of variability and F ratios in one- and two-factor ANOVAs.

groups or cells and a denominator that measures variability within groups or cells. In a one-factor ANOVA, a single null hypothesis is tested with one *F* ratio.

In two-factor ANOVA, three different null hypotheses are tested, one at a time, with three *F* ratios: F_{column}, F_{row}, and $F_{interaction}$.

The numerator of each of these three *F* ratios reflects a different aspect of variability between cells: variability between columns (crowd size), variability between rows (degree of danger), and interaction—any *remaining* variability between cells not attributable to either variability between columns (crowd size) or rows (degree of danger).

ESTIMATING EFFECT SIZE

In the previous chapter, a version of the squared curvilinear correlation, η_p^2

to estimate effect size after variance due to individual differences had been removed. Essentially the same type of analysis can be conducted for *each* significant F in a two-factor ANOVA. Each η_p^2 estimates the proportion of the total variance attributable toeither one of the two factors or to the interaction—after excluding from the total known amounts of variance attributable to the remaining treatment components.

,

PROPORTION OF EXPLAINED VARIANCE (TWO-FACTOR ANOVA)

$$\eta_p^2\,(\text{column}) \quad \frac{SS_{column}}{SS_{total} \quad SS_{row} \quad SS_{interaction}} \quad \frac{SS_{column}}{SS_{column} \quad SS_{within}}$$

$$\eta_p^2\,(\text{row}) \quad \frac{SS_{row}}{SS_{row} \quad SS_{within}}$$

$$\eta_p^2\,(\text{interaction}) \quad \frac{SS_{interaction}}{SS_{interaction} \quad SS_{within}}$$

where each η_p^2 is referred to as a *partial* η^2 for that component because the effects of the other two treatment components have been eliminated from the reduced or partial total variance.

Substituting values for the *SS* terms we have

$$\eta_p^2 \text{ (column)} = \frac{72}{72 + 32} = .69$$

$$\eta_p^2 \text{ (row)} = \frac{192}{192 + 32} = .86$$

$$\eta_p^2 \text{ (interaction)} = \frac{56}{56 + 32} = .64$$

OTHER TYPES OF ANOVA

One- and two-factor studies do not exhaust the possibilities for ANOVA. For instance, you could use ANOVA to analyze the results of a three-factor study with three inde- pendent variables, three 2-way interactions, and one 3-way interaction. Furthermore, regardless of the number of factors, each subject might be measured repeatedly along all levels of one or more factors. Although the basic concepts described in this book transfer almost intact to a wide assortment of more intricate research designs, compu- tational procedures grow more complex, and the interpretation of results often is more difficult. Intricate research designs, requiring the use of complex types of ANOVA, provide the skilled investigator with powerful tools for evaluating complicated situa- tions. Under no circumstances, however, should a study be valued simply because of the complexity of its design and statistical analysis. Use the least complex design and analysis that will answer your research questions.

Chi-Square:

A chi-square (χ^2) statistic is a test that measures how a model compares to actual observed data. The data used in calculating a chi-square statistic must be random, raw, mutually exclusive, drawn from independent variables, and drawn from a large enough sample. For example, the results of tossing a fair coin meet these criteria.

Chi-square tests are often used in hypothesis testing. The chi-square statistic compares the size of any discrepancies between the expected results and the actual results, given the size of the sample and the number of variables in the relationship.

For these tests, degrees of freedom are utilized to determine if a certain null hypothesis can be rejected based on the total number of variables and samples within the experiment. As with any statistic, the larger the sample size, the more reliable the results.

The Formula for Chi-Square Is

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Where,

c=Degrees of freedom

O=Observed value(s)

E=Expected value(s)

There are two main kinds of chi-square tests: the test of independence, and the goodness-of-fit test, which asks something like "How well does the coin in my hand match a theoretically fair coin?"

Chi-square analysis is applied to categorical variables and is especially useful when those variables are nominal (where order doesn't matter, like marital status or gender).

Independence

When considering student sex and course choice, a χ^2 test for independence could be used. To do this test, the researcher would collect data on the two chosen variables (sex and courses picked) and then compare the frequencies at which male and female students select among the offered classes using the formula given above and a χ^2 statistical table.

If there is no relationship between sex and course selection (that is, if they are independent), then the actual frequencies at which male and female students select each offered course should be expected to be approximately equal, or conversely, the proportion of male and female students in any selected course should be approximately equal to the proportion of male and female students in the sample.

A χ^2 test for independence can tell us how likely it is that random chance can explain any observed difference between the actual frequencies in the data and these theoretical expectations.

Goodness-of-Fit

χ^2 provides a way to test how well a sample of data matches the (known or assumed) characteristics of the larger population that the sample is intended to represent. This is known as goodness of fit. If the sample data do not fit the expected properties of the population that we are interested in, then we would not want to use this sample to draw conclusions about the larger population.

Example

For example, consider an imaginary coin with exactly a 50/50 chance of landing heads or tails and a real coin that you toss 100 times. If this coin is fair, then it will also have an equal probability of landing on either side, and the expected result of tossing the coin 100 times is that heads will come up 50 times and tails will come up 50 times.

In this case, χ^2 can tell us how well the actual results of 100 coin flips compare to the theoretical model that a fair coin will give 50/50 results. The actual toss could come up 50/50, or 60/40, or even 90/10. The farther away the actual results of the 100 tosses is from 50/50, the less good the fit of this set of tosses is to the theoretical expectation of 50/50, and the more likely we might conclude that this coin is not actually a fair coin.

When to Use a Chi-Square Test

A chi-square test is used to help determine if observed results are in line with expected results, and to rule out that observations are due to chance. A chi-square test is appropriate for this when the data being analyzed is from a random sample, and when the variable in question is a categorical variable. A categorical variable is one that consists of selections such as type of car, race, educational attainment, male vs. female, how much somebody likes a political candidate (from very much to very little), etc.

These types of data are often collected via survey responses or questionnaires. Therefore, chi-square analysis is often most useful in analyzing this type of data.

chi-square test used for

Chi-square is a statistical test used to examine the differences between categorical variables from a random sample in order to judge goodness of fit between expected and observed results.

Who uses chi-square analysis?

Since chi-square applies to categorical variables, it is most used by researchers who are studying survey response data. This type of research can range from demography to consumer and marketing research to political science and economics.

Is chi-square analysis used when the independent variable is nominal or ordinal?

A nominal variable is a categorical variable that differs by quality, but whose numerical order could be irrelevant. For instance, asking somebody their favorite color would produce a nominal variable. Asking somebody's age, on the other hand, would produce an ordinal set of data. Chi-square can be best applied to nominal data.