

Statistics in Psychology (PSY516)**Table of Content**

Lesson no.	Title	Page no.
1	Introduction to Statistics-I	1
2	Introduction to Statistics-II	4
3	Correlation Method	13
4	Experimental and Non-experimental Method	16
5	Introduction to SPSS	22
6	Descriptive statistics	32
7	Frequency Distribution-I	34
8	Frequency Distribution-II	40
9	Measure of Central Tendency-I	45
10	Measure of Central Tendency-II	49
11	Measure of Variability-I	54
12	Measure of Variability-II	61
13	Z-scores	65
14	Introduction to Probability-I	70
15	Introduction to Probability-II	74
16	Introduction to Probability-III	79
17	Sampling Distribution	84
18	Confidence Interval-I	87
19	Confidence Interval-II	89
20	Hypothesis Testing-I	93
21	Hypothesis Testing-II	101
22	Hypothesis Testing-III	106
23	Hypothesis Testing-IV	108
24	T-Test-I	112
25	T-Test-II	122
26	Independent Sample T-test-I	126
27	Independent Sample T-test-II	134

28	Repeated Measure T-Test	143
29	Analysis of variance (ANOVA)	151
30	One-way ANOVA	156
31	Two-Way ANOVA-I	161
32	Two-Way ANOVA-II	173
33	Correlation Analysis-I	175
34	Correlation Analysis-II	180
35	Correlation Analysis-III	182
36	Types of Correlation-I	188
37	Introduction To Regression	191
38	Simple Linear Regression	197
39	Multiple Linear Regression	204
40	Non-Parametric Test	211
41	Chi-Square Test For Independence-I	213
42	Chi-Square Test For Independence-II	221
43	Mann-Whitney U-Test	225
44	Wilcoxon Signed Rank Test	228
45	Kruskal Wallis & Friedman Test	232

Lesson 23

HYPOTHESIS TESTING-IV

Role Of Sample Size In Statistical Significance

Increasing sample size makes the hypothesis test more sensitive - more likely to reject the null hypothesis when it is, in fact, false. Thus, it increases the power of the test. The effect size is not affected by sample size. Higher sample size allows the researcher to increase the significance level of the findings, since the confidence of the result are likely to increase with a higher sample size. This is to be expected because larger the sample size, the more accurately it is expected to mirror the behavior of the whole group.

Statistical Significance

If 95% confidence intervals do not overlap then we can conclude that the means come from different populations, and, therefore, that they are significantly different.

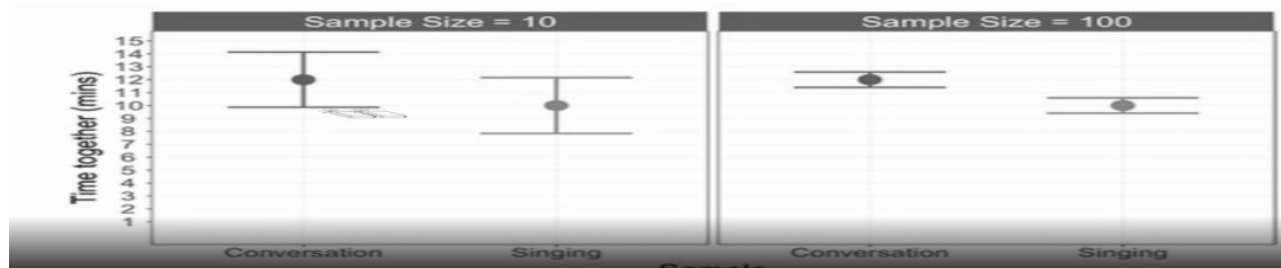
- Whether the results of a study are statistically significant can depend on the sample size.
- A larger sample size tends to yield a statistically significant results.
- On the other hand, if the sample size is too small, an important relationship or difference can go undetected.
- In that case, the power of the test is too low.

Given that power is the ability of a test to find an effect that genuinely exists, and the effect is found by having a statistically significant result (i.e., $p < 0.05$).

Hence, there is also a connection between the sample size and the p-value associated with a test statistic. We clear this concept by giving an example that is study got two groups of 10 heterosexual young men and got them to go up to a woman that they found attractive and either engage them in conversation (group 1) or sing them a song (group 2). Researchers measured how long it was before the woman ran away. Imagine the experiment was also repeated using 100 men in each group.

Results showed that in both cases the singing group had a mean of 10 and a standard deviation of 3, and the conversation group had a mean of 12 and a standard deviation of 3. The only difference between the two experiments is that one collected 10 scores per sample, and the other 100 scores per sample.

The confidence intervals become much narrower when the samples contain 100 scores than when they contain only 10 scores.



The sample size affects whether a difference between samples is deemed significant or not.

In large samples, small differences can be significant, and in small samples large differences can be non-significant. Even a difference of practically zero can be deemed 'significant' if the sample size is big enough.

Calculating Sample Size

We calculate sample size by the following way. The sample size necessary to achieve a given level of power can be calculated by setting the value of α , $1 - \beta$ and effect size. Normally we have alpha value 0.05, power 0.8 and moderate effect size to determine the sample size. Using power to calculate the necessary sample size is the more common and more useful thing to do.

The actual computations are very complicated. Therefore, we usually prefer to use a software for this purpose.

Calculating Sample Size G* Power

G* power is a free to use software or a general power analysis programme, which is used for determining the sample size and analysis power in research studies. To do power analysis to estimate the sample size, we need;

- Hypothesis
- Number of independent variables or number of groups
- What statistical test (one of the inferential statistics) we will be using.

Below picture will show the alpha level (usually .05), power of the test (.80), and effect size (small, moderate or large based on the test being used).

Statistics are showing that total sample size required for the study is 111.

The screenshot shows a statistical software window with the following settings:

- Test family:** t tests
- Statistical test:** Correlation: Point biserial model
- Type of power analysis:** A priori: Compute required sample size - given alpha, power, and effect size
- Input Parameters:**
 - Determine =>** (button)
 - Effect size (p):** 0.3
 - alpha err prob:** 0.05
 - Power (1-beta err prob):** 0.95
 - Tail(s):** One
- Output Parameters:**
 - Noncentrality parameter S:** 3.3133098
 - Critical t:** 1.6589535
 - DF:** 109
 - Total sample size:** 111
 - Actual power:** 0.9503016

Hypothesis Testing Review

Hypothesis testing is a statistical procedure that allows researchers to use sample data to draw inferences about the population of interest.

Hypothesis testing includes:

Stating a hypothesis, either null or alternative. One says that effect does not exist (the null hypothesis, H_0) and the other says that an effect exists (the alternative hypothesis, H_1).

Example:

A researcher begins with a known population, with $\mu=65$ and $\sigma=15$. The researcher suspects that special training in reading skills will produce a change in the scores for individuals in the population. A sample of $n=25$ individuals is selected, and the treatment is given to this sample. Following treatment, the average score for this sample is $M=70$.

We then, (ii) Set criteria for a decision determined by: Alpha level (level of significance that separates the high probability samples from the low-probability samples) and Critical region (extreme sample values that are very unlikely to be obtained if the null hypothesis is true).

Next, (iii) data is collected data and sample statistics are computed

By comparing the data with hypothesis

Comparison is accomplished by computing a z-score

$$z = (M - \mu) / \sigma_M$$

Lastly, (iv) decision is made

A sample value in the critical region is very unlikely to occur if the null hypothesis is true. Therefore, decision is to reject the null hypothesis. If the sample data are not in the critical region then, the sample mean is reasonably close to the population mean (in the center of the distribution) conclusion is to fail to reject the null hypothesis.

Assumptions for hypothesis testing include;

- Random selection of participants

- The values in the sample must consist of independent observations
- The value of population standard deviation (σ) is unchanged by the treatment
- Sample should be normally distributed.

In hypothesis testing a **Type I error** occurs when a researcher rejects a null hypothesis that is actually true. It means that the researcher concludes that a treatment does have an effect when, in fact, it has no effect.

A **Type II error** occurs when a researcher fails to reject a null hypothesis that is really false.

It means that the hypothesis test has failed to detect a real treatment effect.

Hypothesis test is influenced by;

The **variability of the scores** (standard deviation). Higher variability can reduce the chances of finding a significant treatment effect.

The **number of score in the sample**. Increasing the sample size can increase the chances of finding a significant treatment effect.

The statistical power of a test is defined as the probability that the test will reject the null hypothesis if the treatment really has an effect.

In hypothesis testing, the first outcome, failing to reject H_0 when there is a real effect is a Type II error with a probability identified as $p = \beta$. The second outcome have a probability of $1 - \beta$. Hence, rejecting H_0 when there is a real effect this and is the power of the test.

Factors Affecting Power Of The Test:

1. Power of the test is affected by sample size. In general, a larger sample produces greater power for a hypothesis test.
2. Power of the test is affected by alpha level. Reducing the alpha level for a hypothesis test also reduces the power of the test.
3. Power of the test is affected by one-tailed versus two-tailed test. Changing from a regular two-tailed test to a one-tailed test increases the power of the hypothesis test.

Lesson 24

T-TEST-I

The t Statistic: An Alternative To z

1. A sample mean (M) is expected to approximate its population mean (μ). This permits us to use the sample mean to test a hypothesis about the population mean.
2. The standard error provides a measure of how well a sample mean approximates the population mean. Specifically, the standard error determines how much difference is reasonable to expect between a sample mean (M) and the population mean (μ).

$$\sigma_M = \frac{\sigma}{\sqrt{n}} \quad \text{or} \quad \sigma_M = \sqrt{\frac{\sigma^2}{n}}$$

To quantify our inferences about the population, we compare the obtained sample mean (M) with the hypothesized population mean (μ) by computing a z-score test statistic.

$$z = \frac{M - \mu}{\sigma_M} = \frac{\text{obtained difference between data and hypothesis}}{\text{standard distance between } M \text{ and } \mu}$$

The Problem With z-Scores

The shortcoming of using a z-score for hypothesis testing is that the z-score formula requires more information than is usually available. Specifically, a z-score requires that we know the value of the population standard deviation (or variance), which is needed to compute the standard error. In most situations, however, the standard deviation for the population is not known. In fact, the whole reason for conducting a hypothesis test is to gain knowledge about an unknown population. This situation appears to create a paradox, you want to use a z-score to find out about an unknown population, but you must know about the population before you can compute a z-score. Fortunately, there is a relatively simple solution to this problem. When the variability for the population is not known, we use the sample variability in its place.

Introducing The t Statistic

When the variability for the population is not known, we use the sample variability in its place.

Sample variance as an unbiased estimate of the corresponding population variance is

$$s^2 = \sum (X - M)^2 / n - 1$$

$$= SS / n - 1$$

$$= SS / df$$

The sample variance was developed specifically to provide an unbiased estimate of the corresponding population variance. Recall that the formulas for sample variance and sample standard deviation are as follows:

$$\text{sample variance} = s^2 = \frac{SS}{n - 1} = \frac{SS}{df}$$

$$\text{sample standard deviation} = s = \sqrt{\frac{SS}{n - 1}} = \sqrt{\frac{SS}{df}}$$

Using the sample values, we can now *estimate* the standard error. Recall that the value of the standard error can be computed using either standard deviation or variance:

$$\text{standard error} = \sigma_M = \frac{\sigma}{\sqrt{n}} \quad \text{or} \quad \sigma_M = \sqrt{\frac{\sigma^2}{n}}$$

Now we estimate the standard error by simply substituting the sample variance or standard deviation in place of the unknown population value:

$$\text{estimated standard error} = s_M = \frac{s}{\sqrt{n}} \quad \text{or} \quad s_M = \sqrt{\frac{s^2}{n}}$$

Notice that the symbol for the *estimated standard error of M* is s_M instead of σ_M , indicating that the estimated value is computed from sample data rather than from the actual population parameter.

The **estimated standard error (s_M)** is used as an estimate of the real standard error, σ_M , when the value is unknown. It is computed from the sample variance or sample standard deviation and provides an estimate of the standard distance between a sample mean, M , and the population mean, μ .

The estimated standard error of M typically is presented and computed using:

$$s_M = \sqrt{\frac{s^2}{n}}$$

There are two reasons for making this shift from standard deviation to variance:

1. We saw that the sample variance is an *unbiased* statistic; on average, the sample variance (s^2) provides an accurate and unbiased estimate of the population variance (σ^2). Therefore, the most accurate way to estimate the standard error is to use the sample variance to estimate the population variance.
2. In future chapters we encounter other versions of the t statistic that require variance (instead of standard deviation) in the formulas for estimated standard error. To maximize the similarity from one version to another, we use variance in the formula for *all* of the different t statistics. Thus, whenever we present a t statistic, the estimated standard error is computed as:

$$\text{estimated standard error} = \sqrt{\frac{\text{sample variance}}{\text{sample size}}}$$

Now we can substitute the estimated standard error in the denominator of the z score formula. The result is a new test statistic called a t statistics.

$$t = \frac{M - \mu}{s_M}$$

The t statistic is used to test hypotheses about an unknown population mean, μ , when the value of σ is unknown. The formula for the t statistic has the same structure as the z -score formula, except that the t statistic uses the estimated standard error in the denominator.

The only difference between the t formula and the z -score formula is that the z -score uses the actual population variance, σ^2 (or the standard deviation), and the t formula uses the corresponding sample variance (or standard deviation) when the population value is not known.

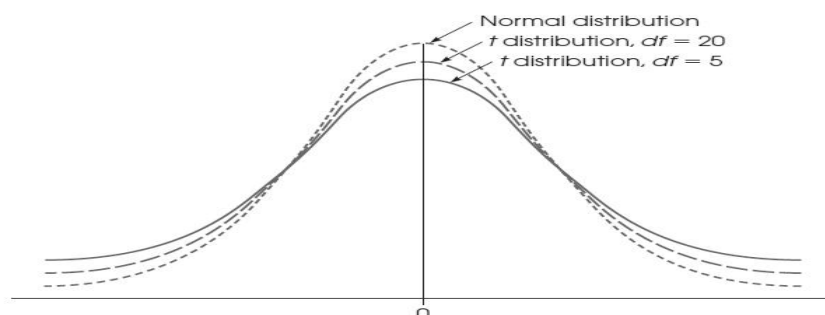
$$z = \frac{M - \mu}{\sigma_M} = \frac{M - \mu}{\sqrt{\sigma^2 / n}} \quad t = \frac{M - \mu}{s_M} = \frac{M - \mu}{\sqrt{s^2 / n}}$$

Reviewing briefly, you must know the sample mean before you can compute sample variance. This places a restriction on sample variability such that only $n - 1$ scores in a sample are independent and free to vary. The value $n - 1$ is called the *degrees of freedom* (or *df*) for the sample variance.

$$\text{degrees of freedom} = df = n - 1$$

Degrees of freedom describe the number of scores in a sample that are independent and free to vary. Because the sample mean places a restriction on the value of one score in the sample, there are $n - 1$ degrees of freedom for a sample with n scores.

A t distribution is the complete set of t values computed for every possible random sample for a specific sample size (n) or a specific degrees of freedom (df). The t distribution approximates the shape of a normal distribution.



The exact shape of a t distribution changes with degrees of freedom. In fact, statisticians speak of a “family” of t distributions. That is, there is a different sampling distribution of t (a distribution of all possible sample t values) for each possible number of degrees of freedom. As df gets very large, the t distribution gets closer in shape to a normal z -score distribution. A quick glance at Figure reveals that distributions of t are bell-shaped and symmetrical and have a mean of zero. However, the t distribution has more variability than a normal z distribution, especially when df values are small. The t distribution tends to be flatter and more spread out, whereas the normal z distribution has more of a central peak. The reason that the t distribution is flatter and more variable than the normal z -score distribution becomes clear if you look at the structure of the formulas for z and t . The t distribution has more variability than a normal z distribution, especially when df values are small. The t distribution tends to be flatter and more spread out, whereas the normal z distribution has more of a central peak.

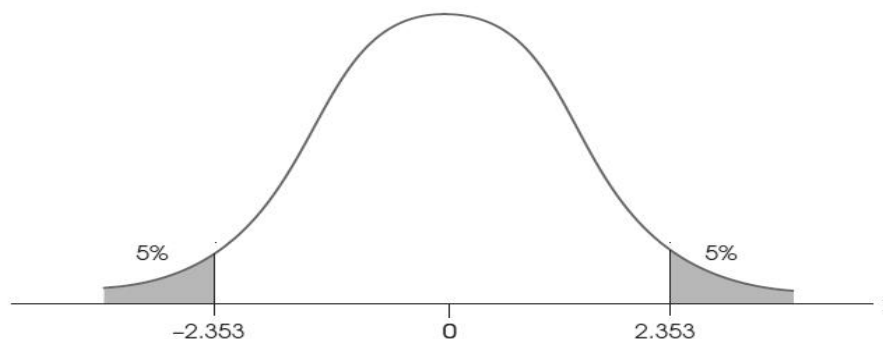
Determining Proportions And Probabilities For t Distributions

Just as we used the unit normal table to locate proportions associated with z -scores, we use a t distribution table to find proportions for t statistics. The complete t distribution table is presented below. The two rows at the top of the table show proportions of the t distribution contained in either one or two tails, depending on which row is used. The first column of the table lists degrees of freedom for the t statistic. Finally, the numbers in the body of the table are the t values that mark the boundary between the tails and the rest of the t distribution.

df	Proportion in One Tail					
	0.25	0.10	0.05	0.025	0.01	0.005
df	Proportion in Two Tails Combined					
	0.50	0.20	0.10	0.05	0.02	0.01
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.816	1.886	2.920	4.303	6.965	9.925
3	0.765	1.638	2.353	3.182	4.541	5.841
4	0.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.707

For example, with $df = 3$, exactly 5% of the t distribution is located in the tail beyond $t = 2.353$. The process of finding this value is highlighted in table above. Begin by locating $df = 3$ in the first column of the table. Then locate a proportion of 0.05 (5%) in the one-tail proportion row. When you line up these two values in the table, you should find $t = 2.353$. Similarly, 5% of the t distribution is located in the tail beyond $t = \pm 2.353$. Finally, notice that a total of 10% (or 0.10) is contained in the two tails beyond $t = \pm 2.353$ (check the proportion value in the "two-tails combined" row at the top of the table).

A close inspection of the t distribution table demonstrates a point we made earlier: As the value for df increases, the t distribution becomes more similar to a normal distribution. For example, examine the column containing t values for a 0.05 proportion in two tails. You will find that when $df = 1$, the t values that separate the extreme 5% (0.05) from the rest of the distribution are $t = \pm 12.706$.



Hypothesis Tests With The t Statistic

As always, the null hypothesis states that the treatment has no effect; specifically, H_0 states that the population mean is unchanged. Thus, the null hypothesis provides a specific value for the unknown population mean. The sample data provide a value for the sample mean. Finally, the variance and estimated standard error are computed from the sample data. When these values are used in the t formula, the result becomes

$$t = \frac{\text{sample mean (from the data)} - \text{population mean (hypothesized from } H_0)}{\text{estimated standard error (computed from the sample data)}}$$

The following research situation demonstrates the procedures of hypothesis testing with the t statistic. Note that this is another example of a null hypothesis that is founded in logic rather than prior knowledge of a population mean.

Constructing Null And Alternative Hypothesis

t statistic permits hypothesis testing in situations for which a population mean is not known

The t test does not require any prior knowledge about the population mean or the population variance.

The null hypothesis would state

$$H_0: \mu = 4$$

The alternative hypothesis would state that

$$H_1: \mu \neq 4$$

Infants, even newborns, prefer to look at attractive faces compared to less attractive faces (Slater, et al., 1998). In the study, infants from 1 to 6 days old were shown two photographs of women's faces. Previously, a group of adults had rated one of the faces as significantly more attractive than the other. The babies were positioned in front of a screen on which the photographs were presented. The pair of faces remained on the screen until the baby accumulated a total of 20 seconds of looking at one or the other. The number of seconds looking at the attractive face was recorded for each infant. Suppose that the study used a sample of $n = 9$ infants and the data produced an average of $M = 13$ seconds for the attractive face with $SS = 72$. Note that all of the available information comes from the sample. Specifically, we do not know the population mean or the population standard deviation.

Step 1

State the hypotheses and select an alpha level. Although we have no information about the population of scores, it is possible to form a logical hypothesis about the value of μ . In this case, the null hypothesis states that the infants have no preference for either face. That is, they should average half of the 20 seconds looking at each of the two faces. In symbols, the null hypothesis states

$$H_0: \mu_{\text{attractive}} = 10 \text{ seconds}$$

The alternative hypothesis states that there is a preference and one of the faces is preferred over the other. A directional, one-tailed test would specify which of the two faces is preferred, but the non-directional alternative hypothesis is expressed as follows:

$$H_1: \mu_{\text{attractive}} \neq 10 \text{ seconds}$$

We set the level of significance at $\alpha = .05$ for two tails.

Step 2

Locate the critical region. The test statistic is a t statistic because the population variance is not known. Therefore, the value for degrees of freedom must be determined before the critical region can be located. For this sample

$$df = n - 1 = 9 - 1 = 8$$

For a two-tailed test at the .05 level of significance and with 8 degrees of freedom, the critical region consists of t values greater than +2.306 or less than -2.306.

Step 3

Calculate the test statistic. The t statistic typically requires much more computation than is necessary for a z -score. Therefore, we recommend that you divide the calculations into a three-stage process as follows:

a. **First, calculate the sample variance.** Remember that the population variance is unknown, and you must use the sample value in its place. (This is why we are using a t statistic instead of a z -score.)

$$s^2 = \frac{SS}{n-1} = \frac{SS}{df} = \frac{72}{8} = 9$$

b. Next, **use the sample variance (s^2) and the sample size (n) to compute the estimated standard error.** This value is the denominator of the t statistic and measures how much difference is reasonable to expect by chance between a sample mean and the corresponding population mean.

$$s_M = \sqrt{\frac{s^2}{n}} = \sqrt{\frac{9}{9}} = \sqrt{1} = 1$$

Finally, compute the t statistic for the sample data.

$$t = \frac{M - \mu}{s_M} = \frac{13 - 10}{1} = 3.00$$

Step 4

Make a decision regarding H_0 . The obtained t statistic of 3.00 falls into the critical region on the right-hand side of the t distribution. Our statistical decision is to reject H_0 and conclude that babies do show a preference when given a choice between an attractive and an unattractive face. Specifically, **the average amount of time that the babies spent looking at the attractive face was significantly different from the 10 seconds that would be expected if there were no preference.**

As indicated by the sample mean, there is a tendency for the babies to spend more time looking at the attractive face.

Assumptions Of The t Test

Two basic assumptions are necessary for hypothesis tests with the t statistic.

1. The values in the sample must consist of *independent* observations. In everyday terms, two observations are independent if there is no consistent, predictable relationship between the first observation and the second. More precisely, two events (or observations) are independent if the occurrence of the first event has no effect on the probability of the second event.
2. The population that is sampled must be normal. This assumption is a necessary part of the mathematics underlying the development of the t statistic and the t distribution table. However, violating this assumption has little practical effect on the results obtained for a t statistic, especially when the sample size is relatively large. With very small samples, a normal population distribution is important. With larger samples, this assumption can be violated without affecting the validity of the hypothesis test. If you have reason to suspect that the population distribution is not normal, use a large sample to be safe.

The Influence Of Sample Size And Sample Variance

The number of scores in the sample and the magnitude of the sample variance both have a large effect on the t statistic and thereby influence the statistical decision.

As the estimated standard error appears in the denominator of the formula, a larger value for s_M produces a smaller value (closer to zero) for t .

The estimated standard error is directly related to the sample variance so that the larger the variance, the larger the error.

The estimated standard error is inversely related to the number of scores in the sample. The larger the sample is, the smaller the error is.

If all other factors are held constant, large samples tend to produce bigger t statistics and therefore are more likely to produce significant results.

The estimated standard error is directly related to the sample variance so that the larger the variance, the larger the error. Thus, large variance means that you are less likely to obtain a significant treatment effect. In general, large variance is bad for inferential statistics. Large variance means that the scores are widely scattered, which makes it difficult to see any consistent patterns or trends in the data. In general, high variance reduces the likelihood of rejecting the null hypothesis.

On the other hand, the estimated standard error is inversely related to the number of scores in the sample. The larger the sample is, the smaller the error is. If all other factors are held constant, large samples tend to produce bigger t statistics and therefore are more likely to produce significant results.

$$t = \frac{M - \mu}{s_M} \quad \text{where } s_M = \sqrt{\frac{s^2}{n}}$$

The effect size for t -test becomes:

$$\text{estimated } d = \frac{\text{mean difference}}{\text{sample standard deviation}} = \frac{M - \mu}{s}$$

$$\text{And where } s = \sqrt{\frac{SS}{n-1}} = \sqrt{\frac{SS}{df}}$$

Lesson 25

T-TEST-II

Calculating T statistics

The one-sample t-test is used to determine whether a sample comes from a population with a specific mean. This population mean is not always known, but is sometimes hypothesized.

For example, you want to show that a new teaching method for pupils struggling to learn English grammar can improve their grammar skills to the national average. Your sample would be pupils who received the new teaching method and your population mean would be the national average score. Alternately, you believe that doctors that work in accident and emergency (A & E) departments work 100 hour per week despite the dangers (e.g., tiredness) of working such long hours. You sample 1000 doctors in A & E departments and see if their hours differ from 100 hours.

In two-tailed test, a sample of $n=4$ individuals is selected from a population with a mean of $\mu=40$. A treatment is administered to the individuals in the sample and, after treatment, the sample has a mean of $M=44$ and a variance of $s^2=16$. In this type of data to know “Is this sample sufficient to conclude that the treatment has a significant effect?” We have to use a two-tailed test with alpha .05. In one-tailed test, in a study, the research question is whether attractiveness affects the behavior of infants looking at photographs of women’s faces. The researcher predicts that the infants will spend more than half of the 20-second period looking at the attractive face.

The researcher tested a sample of $n=9$ infants and obtained a mean of $M=13$ seconds looking at the attractive face with $SS=72$. We use one tail test.

One sample T-test using SPSS**Data Entry**

A random sample of $n=5$ participants with scores on a test (11,13,4,12,10) is obtained from a population with a mean of $\mu=13$. A treatment is administered to the individuals in the sample.

A researcher wants to examine if the sample means differ from population mean after treatment.

Enter all of the scores from the sample in one column of the data editor.

	scores	var	var
1	11.00		
2	13.00		
3	4.00		
4	12.00		
5	10.00		
6			
7			

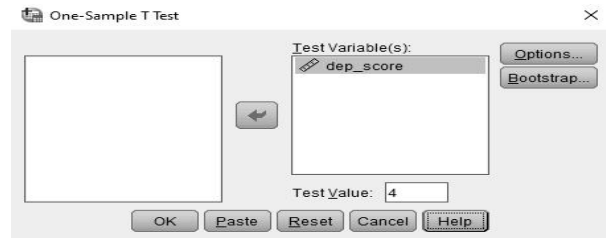
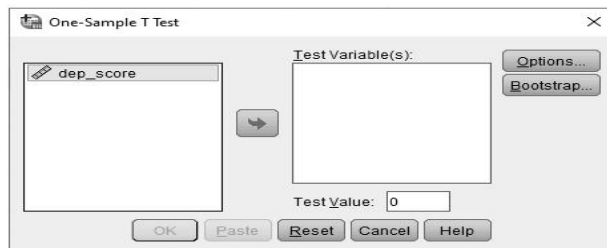
Data Analysis

1. Click **Analyze** on the tool bar, select **Compare Means**, and click on **One-Sample T Test**.

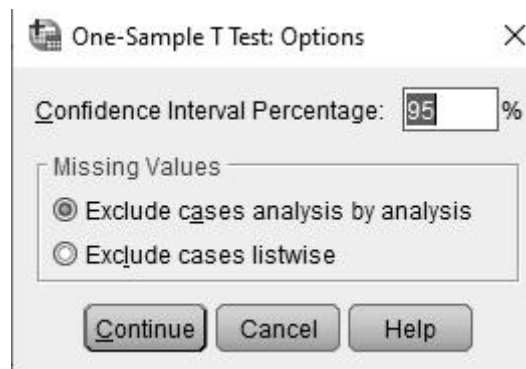
Click **Analyze > Compare Means > One-Sample T Test...** on the main menu:



2. Highlight the column label for the set of scores in the left box and click the arrow to move it into the Test Variable(s) box.
3. In the Test Value box at the bottom of the One-Sample t Test window, enter the hypothesized value for the population mean from the null hypothesis ($\mu=13$)



4. In addition to performing the hypothesis test, the program computes a confidence interval for the population mean difference. **The confidence level is automatically set at 95%**, but you can select Options and change the percentage. Click on the Continue button. You will be returned to the One-Sample T Test dialogue box. Click on the OK button to generate the output. It will generate SPSS output.



Reporting Results Of T-Test

SPSS Statistics generates two main tables of output for the one-sample t-test that contains all the information you require to interpret the results of a one-sample t-test.

A scientific report typically uses the term significant to indicate that the null hypothesis has been rejected and the term not significant to indicate failure to reject H_0 .

There is a prescribed format for reporting the calculated value of the test statistic, degrees of freedom, and alpha level for a t test.

You can make an initial interpretation of the data using the One-Sample Statistics table, which presents relevant descriptive statistics. It is more common than not to present your descriptive statistics using the mean and standard deviation ("Std. Deviation" column) rather than the standard error of the mean ("Std. Error Mean" column), although both are acceptable.

However, by running a one-sample t-test, you are really interested in knowing whether the sample you have (dep_score) comes from a 'normal' population (which has a mean of 4.0). This is discussed in the next section.

The One-Sample Test table reports the result of the one-sample t-test. The top row provides the value of the known or hypothesized population mean you are comparing your sample data to, as highlighted below:

One-Sample Test						
	Test Value = 4					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
dep_score	-2.381	39	.022	-.27750	-.5132	-.0418

In this example, you can see the 'normal' depression score value of "4" that you entered in earlier. You now need to consult the first three columns of the One-Sample Test table, which provides information on whether the sample is from a population with a mean of 4 (i.e., are the means statistically significantly different).

Moving from left-to-right, you are presented with the observed t-value ("t" column), the degrees of freedom ("df"), and the statistical significance (p-value) ("Sig. (2-tailed)") of the one-sample t-test. In this example, $p < .05$ (it is $p = .022$). Therefore, it can be concluded that the population means are statistically significantly different. If $p > .05$, the difference between the sample-

estimated population mean and the comparison population mean would not be statistically significantly different.

SPSS Statistics also reports that $t = -2.381$ ("t" column) and that there are 39 degrees of freedom ("df" column). You need to know these values in order to report your results, which you could do as follows:

One-Sample Test						
	Test Value = 4					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
dep_score	-2.381	39	.022	-.27750	-.5132	-.0418

Depression score was statistically significantly lower than the population normal depression score, $t(39) = -2.381$, $p = .022$.

One-Sample Test						
	Test Value = 4					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
dep_score	-2.381	39	.022	-.27750	-.5132	-.0418

This section of the table shows that the mean difference in the population means is -0.28 ("Mean Difference" column) and the 95% confidence intervals (95% CI) of the difference are -0.51 to -0.04 ("Lower" to "Upper" columns). For the measures used, it will be sufficient to report the values to 2 decimal places. You could write these results as:

Depression score was statistically significantly lower than the population normal depression score, $t(39) = -2.381$, $p = .022$.

You can also include measures of the difference between the two population means in your written report. This information is included in the columns on the far-right of the One-Sample Test table. This section of the table shows that the mean difference in the population means is -0.28 ("Mean Difference" column) and the 95% confidence intervals (95% CI) of the difference are -0.51 to -0.04 ("Lower" to "Upper" columns). For the measures used, it will be sufficient to report the values to 2 decimal places. You could write these results as:

Depression score was statistically significantly lower by 0.28 (95% CI, 0.04 to 0.51) than a normal depression score of 4.0, $t(39) = -2.381$, $p = .022$

Lesson 26

INDEPENDENT SAMPLE T-TEST-I**The t Test for Two Independent Samples**

Although these single sample techniques are used occasionally in real research, most research studies require the comparison of two (or more) sets of data. For example, a social psychologist may want to compare men and women in terms of their political attitudes, an educational psychologist may want to compare two methods for teaching mathematics, or a clinical psychologist may want to evaluate a therapy technique by comparing depression scores for patients before therapy with their scores after therapy. The two sets of data could come from two completely separate groups of participants. This is called an independent measures research design or between-subject design.

In each case, the research question concerns a mean difference between two sets of data.

The independent-samples t-test (or independent t-test, for short) compares the means between two unrelated groups on the same continuous, dependent variable. For example, you could use an independent t-test to understand whether first year graduate salaries differed based on gender (i.e., your dependent variable would be "first year graduate salaries" and your independent variable would be "gender", which has two groups: "male" and "female"). Alternately, you could use an independent t-test to understand whether there is a difference in test anxiety based on educational level (i.e., your dependent variable would be "test anxiety" and your independent variable would be "educational level", which has two groups: "undergraduates" and "postgraduates").

There are two general research designs that can be used to obtain the two sets of data to be compared:

1. The two sets of data could come from two completely separate groups of participants. For example, the study could involve a sample of men compared with a sample of women. Or the study could compare grades for one group of freshmen who are given laptop computers with grades for a second group who are not given computers.
2. The two sets of data could come from the same group of participants. For example, the researcher could obtain one set of scores by measuring depression for a sample of patients before they begin therapy and then obtain a second set of data by measuring the same individuals after 6 weeks of therapy.

The first research strategy, using completely separate groups, is called an independent measures research design or a between-subjects design. These terms emphasize the fact that the design involves separate and independent samples and makes a comparison between two groups of individuals.

A research design that uses a separate group of participants for each treatment condition (or for each population) is called an independent-measures research design or a between-subjects research design.

Assumptions Underlying The Independent-Measures t Formula

There are three assumptions that should be satisfied before you use the independent measures t formula for hypothesis testing:

1. The observations within each sample must be independent.
2. The two populations from which the samples are selected must be normal.
3. The two populations from which the samples are selected must have equal variances.

The third assumption is referred to as homogeneity of variance and states that the two populations being compared must have the same variance. You may recall a similar assumption for the z -score hypothesis test. For that test, we assumed that the effect of the treatment was to add a constant amount to (or subtract a constant amount from) each individual score. As a result, the population standard deviation after treatment was the same as it had been before treatment. We now are making essentially the same assumption, but phrasing it in terms of variances.

The t statistic for an independent-measures research design involves two separate samples, we need some special notation to help specify which data go with which sample. This notation involves the use of subscripts, which are small numbers written beside a sample statistic. For example, the number of scores in the first sample would be identified by n_1 ; for the second sample, the number of scores is n_2 . The sample means would be identified by M_1 and M_2 . The sums of squares would be SS_1 and SS_2 .

Hypothesis Testing

The goal of an independent-measures research study is to evaluate the mean difference between two populations (or between two treatment conditions). Using subscripts to differentiate the two populations, the mean for the first population is μ_1 , and the second population mean is μ_2 . The difference between means is simply $\mu_1 - \mu_2$. As always, the null hypothesis states that there is

no change, no effect, or, in this case, no difference. Thus, in symbols, the null hypothesis for the independent-measures test is

$H_0: \mu_1 - \mu_2 = 0$ (No difference between the population means)

The alternative hypothesis states that there is a mean difference between the two populations

$H: \mu_1 - \mu_2 \neq 0$

Calculation Of Independent Sample t-Test

1. The basic structure of the t statistic is the same for both the independent-measures and the single-sample hypothesis tests. The overall t formula is given below:

$t = \frac{\text{sample mean difference} - \text{population mean difference}}{\text{estimated standard error}}$

$$t = \frac{(M_1 - M_2) - (\mu_1 - \mu_2)}{S(M_1 - M_2)}$$

2. The independent-measures t is basically a *two-sample t that doubles all the elements of the single-sample t formulas*. To demonstrate the second point, we examine the two t formulas piece by piece.

The single-sample t uses one sample mean to test a hypothesis about one population mean. The sample mean and the population mean appear in the numerator of the t formula, which measures how much difference there is between the sample data and the population hypothesis.

$$t = \frac{\text{sample mean} - \text{population mean}}{\text{estimated standard error}} = \frac{M - \mu}{S_M}$$

The independent measure t test uses difference between two sample means to evaluate a hypothesis about the difference between population means. Thus, the independent measure t formula is:

$t = \frac{\text{sample mean difference} - \text{population mean difference}}{\text{estimated standard error}}$

$$t = \frac{(M_1 - M_2) - (\mu_1 - \mu_2)}{S(M_1 - M_2)}$$

In this formula, the value of $M_1 - M_2$ is obtained from the sample data and the value of $\mu_1 - \mu_2$ comes from the null hypothesis.

The overall t formula The single-sample t uses one sample mean to test a hypothesis about one population mean. The sample mean and the population mean appear in the numerator of the t formula, which measures how much difference there is between the sample data and the population hypothesis.

Interpreting The Estimated Standard Error

The estimated standard error of $M_1 - M_2$ that appears in the bottom of the independent-measures t statistic can be interpreted in two ways. First, the standard error is defined as a measure of the standard, or average, distance between a sample statistic ($M_1 - M_2$) and the corresponding population parameter ($\mu_1 - \mu_2$). As always, samples are not expected to be perfectly accurate and the standard error measures how much difference is reasonable to expect between a sample statistic and the population parameter.

This produces a second interpretation for the estimated standard error. Specifically, the standard error can be viewed as a measure of how much difference is reasonable to expect between two sample means if the null hypothesis is true. The second interpretation of the estimated standard error produces a simplified version of the independent-measures t statistic.

For the independent-measures t statistic, we want to know the total amount of error involved in using two sample means to approximate two population means. To do this, we find the error from each sample separately and then add the two errors together. The resulting formula for standard error is

For the independent-measures t statistic, there are two SS values and two df values (one from each sample). The values from the two samples are combined to compute what is called the pooled variance.

With one sample, the variance is computed as SS divided by df . With two samples, the pooled variance is computed by combining the two SS values and then dividing by the combination of the two df values. As we mentioned earlier, the pooled variance is actually an average of the two sample variances, but the average is computed so that the larger sample carries more weight in determining the final value. The following examples demonstrate this point.

When computing the pooled variance, the weight for each of the individual sample variances is determined by its degrees of freedom. Because the larger sample has a larger df value, it carries more weight when averaging the two variances. This produces an alternative formula for computing *pooled variance*.

For the independent-measures t formula, the standard error measures the amount of error that is expected when you use a sample mean difference ($M_1 - M_2$) to represent a population mean difference ($\mu_1 - \mu_2$). The standard error for the sample mean difference is represented by the symbol $S(M_1 - M_2)$.

Each of the two sample means represents its own population mean, but in each case there is some error.

M_1 approximates μ_1 with some error.

M_2 approximates μ_2 with some error.

For the independent-measures t statistic, we want to know the total amount of error involved in using two sample means to approximate two population means.

$$S(M_1 - M_2) = \sqrt{(s_1^2/n_1) + (s_2^2/n_2)}$$

The *estimated standard error of $M_1 - M_2$* that appears in the bottom of the independent-measures t statistic can be interpreted in two ways. First, the standard error is defined as a measure of the standard, or average, distance between a sample statistic ($M_1 - M_2$) and the corresponding population parameter ($\mu_1 - \mu_2$). As always, samples are not expected to be perfectly accurate and the standard error measures how much difference is reasonable to expect between a sample statistic and the population parameter.

This produces a second interpretation for the estimated standard error. Specifically, the standard error can be viewed as a measure of how much difference is reasonable to expect between two sample means if the null hypothesis is true. The second interpretation of the estimated standard error produces a simplified version of the independent-measures t statistic.

For the independent-measures t statistic, we want to know the total amount of error involved in using *two* sample means to approximate *two* population means. To do this, we find the error from each sample separately and then add the two errors together.

Pooled Variance

For correcting the bias in the standard error is to combine the two sample variances into a single value called the pooled variance.

The pooled variance is obtained by averaging or “pooling” the two sample variances using a procedure that allows the bigger sample to carry more weight in determining the final value.

For the independent-measures t statistic, there are two SS values and two df values. The values from the two samples are combined to compute what is called the pooled variance.

The pooled variance is identified by the symbol S^2_p

$$S^2_p = (SS_1 + SS_2) / (df_1 + df_2)$$

With one sample, the variance is computed as SS divided by df . With two samples, the pooled variance is computed by combining the two SS values and then dividing by the combination of the two df values. As we mentioned earlier, the pooled variance is actually an average of the two sample variances, but the average is computed so that the larger sample carries more weight in determining the final value. The following examples demonstrate this point.

Note that the pooled variance is exactly halfway between the two sample variances. Because the two sample are exactly the same size, the pooled variance is simply the average of the two sample variance.

This time the pooled variance is not located halfway between the two sample variance. Instead, the pooled value is closer to the variance for the larger sample ($n=9$ and $s^2=6$) than to the variance for the smaller sample ($n=3$ and $s^2=10$). the larger sample carries more weight when the pooled variance is computed.

When computing the pooled variance, the weight for each of the individual sample variances is determined by its degrees of freedom. Because the larger sample has a larger df value, it carries more weight when averaging the two variances. This produces an alternative formula for computing pooled variance.

Estimated Standard Error

Using the pooled variance in place of the individual sample variances, we can now obtain an unbiased measure of the standard error for a sample mean difference. The resulting formula for the independent-measures estimated standard error is

$$M_1 - M_2 \pm S(M_1 - M_2) = \sqrt{(S^2_p/n_1) + (S^2_p/n_2)}$$

Confidence intervals for estimating $\mu_1 - \mu_2$

For the independent-measures t , we use a sample mean difference, $M_1 - M_2$, to estimate the population mean difference, $\mu_1 - \mu_2$.

$$\mu_1 - \mu_2 = M_1 - M_2 \pm t \times S(M_1 - M_2)$$

Conceptually, this standard error measures how accurately the difference between two sample means represents the difference between the two population means. In a hypothesis test, H_0 specifies that $\mu_1 - \mu_2 = 0$, and the standard error also measures how much difference is expected, on average, between the two sample means. In either case, the formula combines the error for the first sample mean with the error for the second sample mean. Also note that the pooled variance from the two samples is used to compute the standard error for the sample mean difference.

The basic structure of the t statistic is the same for both the independent-measures and the single-sample hypothesis tests.

$t = \text{sample statistic} - \text{hypothesized population mean} / \text{estimated standard error}$

The overall t formula is given below:

$t = \text{sample mean difference} - \text{population mean difference} / \text{estimated standard error}$

$$t = [(M_1 - M_2) - (\mu_1 - \mu_2)] / S^{(M_1 - M_2)}$$

The degrees of freedom for the independent-measures t statistic are determined by the df values for the two separate samples:

$df \text{ for the } t \text{ statistic} = df \text{ for the first sample} + df \text{ for the first sample}$

$$= df_1 + df_2$$

$$= (n_1 - 1) + (n_2 - 1)$$

Equivalently, the df value for the independent measure t statistic can be expressed as:

$$df = n_1 + n_2 - 2$$

	Sample Data	Hypothesized Population Parameter	Estimated Standard Error	Sample Variance
Single-sample t statistic	M	μ	$\sqrt{\frac{s^2}{n}}$	$s^2 = \frac{SS}{df}$
Independent-measures t statistic	$(M_1 - M_2)$	$(\mu_1 - \mu_2)$	$\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$	$s_p^2 = \frac{SS_1 + SS_2}{df_1 + df_2}$

Reporting The Results

SPSS Statistics generates two main tables of output for the independent t -test. Group statistics table provides useful descriptive statistics for the two groups that you compared, including the

mean and standard deviation. independent samples test table provides the actual results from the independent t-test.

The participants who were informed about the withheld evidence gave significantly longer sentences than those participants who were not informed about the withhold evidence. The mean difference was significant, $t(14) = -3.53$, $p < .05$.

If an exact probability is available from a computer analysis, it should be reported.

If a confidence interval is reported to describe effect size, it appears immediately after the results from the hypothesis test.

Lesson 27

INDEPENDENT SAMPLE T-TEST-II

Independent Sample t-Test In SPSS

The independent-samples t-test (or independent t-test, for short) compares the means between two unrelated groups on the same continuous, dependent variable. For example, you could use an independent t-test to understand whether first year graduate salaries differed based on gender (i.e., your dependent variable would be "first year graduate salaries" and your independent variable would be "gender", which has two groups: "male" and "female"). Alternately, you could use an independent t-test to understand whether there is a difference in test anxiety based on educational level (i.e., your dependent variable would be "test anxiety" and your independent variable would be "educational level", which has two groups: "undergraduates" and "postgraduates").

Independent sample t-test is used when the mean scores of two different groups of people or conditions are compared. It is also called between-subject design. This is a parametric test because it requires assumptions about parameters. Parametric tests require a numerical score for each individual in the sample. The scores then are added, squared, averaged, and otherwise manipulated using basic arithmetic.

This test tells whether there is a statistically significant difference in the mean scores for the two groups, for instance, whether males and females differ significantly in terms of their self-esteem levels. In statistical terms, it tests the probability that the two sets of scores (for males and females) came from the same population.

To run independent sample t-test in SPSS, we need two set of variables;

1. The **Continuous /Dependent Variable**, whose mean score is compared between the two groups. SPSS defines it as the test Variable having scale measurement.
2. The **Categorical/Independent Variable** with two categories that defines which two samples will be compared in the t-test. In SPSS, for t-test it is defined as the Grouping Variable.

The grouping variable must have at least two categories (groups); it may have more than two categories but a t-test can only compare two groups, so there is need to specify which two groups to compare. We can also use a continuous variable by specifying a cut point to create two groups (i.e., values at or above the cut point and values below the cut point).

Downs and Abwender (2002) evaluated athletes (soccer players and swimmers) to determine whether the blows to the head experienced by both groups produced long-term neurological deficits. In the study, neurological tests were administered to mature soccer players and swimmers and a researcher obtained the following data presented in the table.

Are there any significant differences in the neurological test scores?

Swimmers	Soccer Players
10	7
8	4
7	9
8	3
13	7
7	
6	
12	

Running Independent Sample t-Test In SPSS

Entering Data

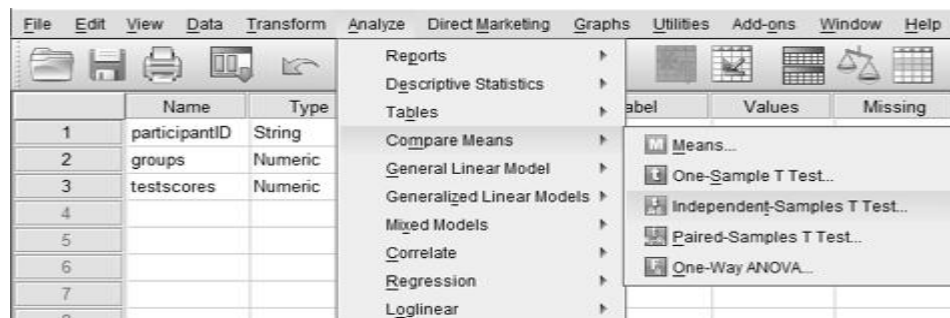
- In keeping the above mentioned example, first of all, the variables must be setup in the variable view tab of SPSS data editor, although it is not necessary but it is good practice to give each person a unique ID.
- We also need to provide codes that distinguish between the two groups (i.e., swimmers and soccer players).
- By clicking in the box of the 'values' field in the row containing the groups variable, we get a pop-up dialog that allows us to code the categorical variable. The swimmer is coded as '1' and the soccer player as '2'.

	participantID	groups	testscores	var
1	1	1.00	10.00	
2	2	1.00	8.00	
3	3	1.00	7.00	
4	4	1.00	8.00	
5	5	1.00	13.00	
6	6	1.00	7.00	
7	7	1.00	6.00	
8	8	1.00	12.00	
9	9	2.00	7.00	
10	10	2.00	4.00	
11	11	2.00	9.00	
12	12	2.00	3.00	
13	13	2.00	7.00	
14				

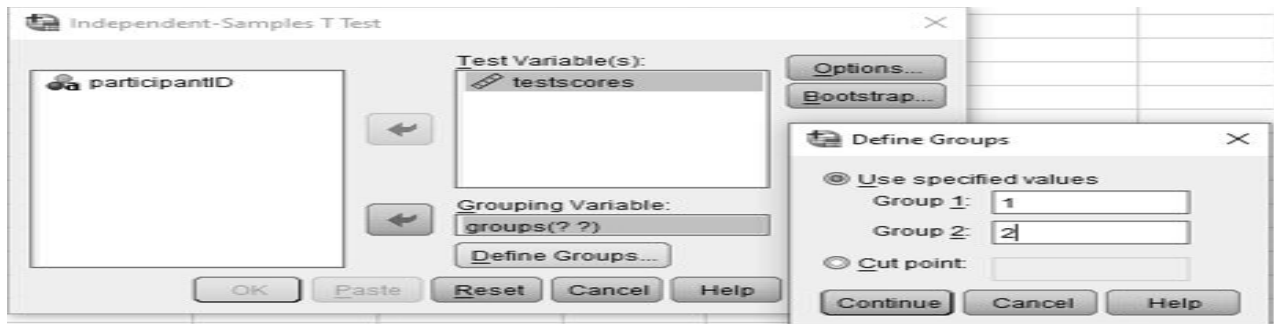
- The data is entered in the data view tab of the SPSS editor.
- We need to enter data for both the independent(categorical, i.e., blows to the head experienced by swimmers and soccer players)and dependent (continuous, i.e., test scores) variables.
- A code is entered for the both groups as they were on (1 = swimmer, 2 = soccer player) and the test score of each is entered.

Running Analysis

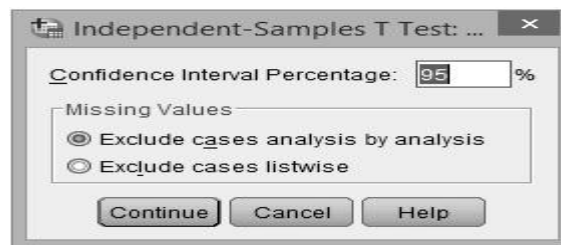
- ◆ To run independent sample t-test in SPSS, Click **Analyze > Compare Means > Independent-Samples T Test**, on the top menu, as shown below:



- ◆ The Independent-Samples T Test window opens where we transfer the dependent variable, tests cores, into the Test Variable(s): box, and the independent variable, groups, into the Grouping Variable: box, by highlighting the relevant variables and pressing the SPSS right arrow button.
- ◆ We then need to define the groups (athletes, swimmers and soccer players).
- ◆ By clicking on the button we will be presented with the Define Groups dialogue box, as shown below.
- ◆ Enter 1 into the Group 1: box and enter 2 into the Group 2: box.
- ◆ Remember that we labeled the swimmer group as 1 and the soccer player group as 2.
- ◆ Click the button Continue.



- ◆ The Options section is where we can set our desired confidence interval/level for the mean difference, and specify how SPSS should handle missing values. In SPSS by default **confidence interval is set at 95%** and we do not need to change it.
- ◆ The Missing Values section is not relevant if we have only specified one dependent variable; it only matters if we are entering more than one dependent (continuous numeric) variable. By clicking to the exclude values case by case.



- ◆ Then, click Continue and OK when finished.

Interpreting Results:

In the Group Statistics box, SPSS gives us the mean and standard deviation for each of your groups (in this case, swimmers and soccer players).

Group Statistics					
groups		N	Mean	Std. Deviation	Std. Error Mean
testscores	swimmers	8	8.8750	2.53194	.89518
	soccer players	5	6.0000	2.44949	1.09545

It also gives the number of people in each group (N).

The second section, Independent Samples Test, displays the results most relevant to the Independent Samples t-test.

Independent Samples Test									
		Levene's Test for Equality of Variances		t-test for Equality of Means					
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference
test scores	Equal variances assumed	.022	.886	2.015	11	.069	2.87500	1.42651	Lower: -.26474 Upper: 6.01474
	Equal variances not assumed			2.032	8.867	.073	2.87500	1.41469	Lower: -.33260 Upper: 6.08260

There are two parts that provide different pieces of information:

(A) Levene's Test for Equality of Variances

(B) t-test for Equality of Means.

In B part (t-test for Equality of Means) if the value in the Sig. (2-tailed) column is equal to or less than .05, there is a significant difference in the mean scores on dependent variable for each of the two groups.

The independent t-test assumes the variances of the two groups you are measuring are equal in the population. If your variances are unequal, this can affect the Type I error rate. The assumption of homogeneity of variance can be tested using Levene's Test of Equality of Variances, which is produced in SPSS Statistics when running the independent t-test procedure. If you have run Levene's Test of Equality of Variances in SPSS Statistics. This test for homogeneity of variance provides an F-statistic and a significance value (p-value). We are primarily concerned with the significance value – if it is greater than 0.05 (i.e., $p > .05$), our group variances can be treated as equal. However, if $p < 0.05$, we have unequal variances and we have violated the assumption of homogeneity of variances.

If the value is above .05, there is no significant difference between the two groups.

In the example presented here, the Sig. (2-tailed) value is .069. As this value is above the required cut-off of .05, we can conclude that there is not a statistically significant difference in the mean neurological test scores for Swimmers and soccer players.

Mean Difference is the difference between the sample means; it also corresponds to the numerator of the test statistic. Std. Error Difference is the standard error; it also corresponds to the denominator of the test statistic. Confidence interval of the Difference part of the t-test output complements the significance test results.

Typically, if the CI for the mean difference contains 0, the results are not significant at the chosen significance level. In this example, the CI is $[-.265, +6.015]$, which does contain zero; this agrees with the p-value of the significance test.

Assumptions Of Independent Sample t-Test

There are some general assumptions that apply to all of the parametric tests (e.g. t-tests, analysis of variance), and additional assumptions associated with specific techniques.

These general assumption include;

1. Level of Measurement

Each of the parametric approaches assumes that the dependent variable is measured at the interval or ratio level; that is, using a continuous scale rather than discrete categories.

2. Random Sampling

The parametric techniques assume that the scores are obtained using a random sample from the population.

3. Independence of Observations

The observations that make up data must be independent of one another; that is, each observation or measurement must not be influenced by any other observation or measurement.

4. Normal Distribution

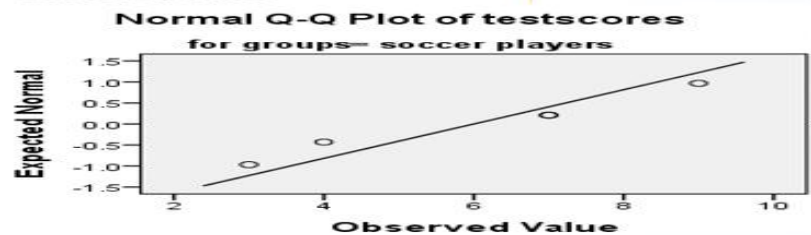
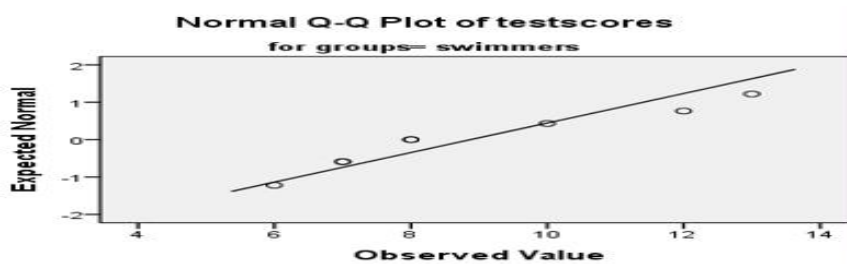
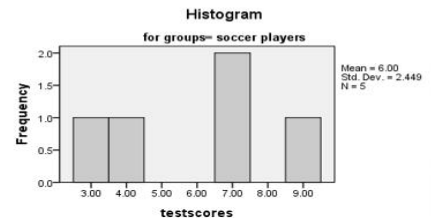
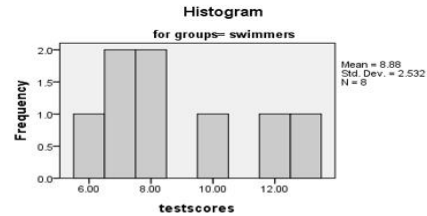
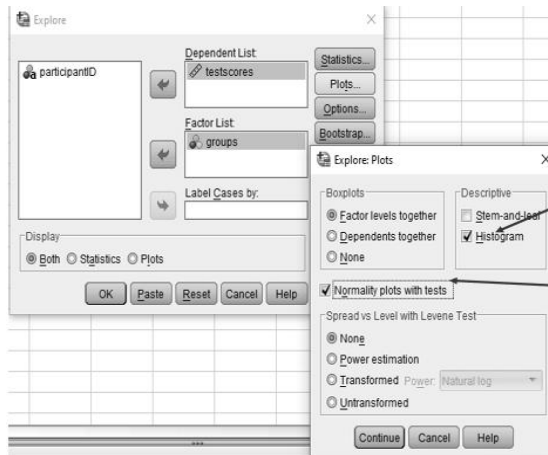
For parametric techniques, it is assumed that the populations from which the samples are taken are normally distributed.

The technique available in SPSS is using the Explore option of the Descriptive Statistics menu.

- Click on the variable(s) (test score) and move it into the Dependent List box.
- In the Factor list section add groups.
- In the Display section, make sure that Both is selected.
- Click on the Statistics button and click on Descriptives and Outliers. Click on Continue.



- Click on the Plots button. Under Descriptive, click on Histogram to select it.



- Click on Normality plots with tests. Click on Continue.
- Click on the Options button. In the Missing Values section, click on Exclude cases pairwise.
- The actual shape of the distribution for each group can be seen in the Histograms.
- This is also supported by an inspection of the normal probability plots (labeled Normal Q-Q Plot).
- In this plot, the observed value for each score is plotted against the expected value from the normal distribution. A reasonably straight line suggests a normal distribution. The Skewness value provides an indication of the symmetry of the distribution. Positive skewness values suggest that scores are clustered to the left at the low values. Negative skewness values indicate a clustering of scores at the high end (right-hand side of a graph). Kurtosis, on the other hand, provides information about the 'peakedness' of the distribution. Positive kurtosis values indicate that the distribution is rather peaked (clustered in the center), with

long thin tails. Kurtosis values below 0 indicate a distribution that is relatively flat (too many cases in the extremes).

- In the table labeled Tests of Normality, we are given the results of the Kolmogorov-Smirnov statistic. This assesses the normality of the distribution of scores. A non-significant result (Sig. value of more than .05) indicates normality. Other than mean and SD, descriptive statistics also provide some information concerning the distribution of scores on continuous variables (skewness and kurtosis). This information is needed if the variables are to be used in parametric statistical techniques (e.g. t-tests, analysis of variance).

Tests of Normality

groups		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
testscores	swimmers	.260	8	.118	.896	8	.268
	soccer players	.258	5	.200*	.925	5	.563

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

5. Homogeneity of Variance

This assumption states that samples are obtained from populations of equal variances. This means that the variability of scores for each of the groups is similar. To test this, SPSS performs *Levene's test for equality of variances* as part of the t-test and analysis of variance analyses.

The first section of the Independent Samples Test output box gives the results of Levene's test for equality of variances.

This tests whether the variance (variation) of scores for the two groups (swimmers and soccer players) is the same. The outcome of this test determines which of the t-values that SPSS provides is the correct one to use.

If the Sig. value for Levene's test is larger than .05 we should use the first line in the table, which refers to Equal variances assumed. If the significance level of Levene's test is $p \leq .05$ or less, this means that the variances for the two groups are not the same. Therefore, refer to second line equal variances not assumed.

Effect Size

Effect size is the strength of the difference between groups, or the influence of the independent variable.

There are a number of different effect size statistics. The most commonly used to compare groups are partial eta squared and Cohen's d. Partial eta squared measures effect size for analysis of variance. Cohen's d, presents difference between groups for t-test.

SPSS calculates partial eta squared as part of the output (e.g. analysis of variance). However, it does not provide effect size statistics for t-tests.

We can use the information provided in the SPSS to calculate whichever effect size statistic we need preferably when the results are significant. For the independent samples T-test, Cohen's d is determined by calculating the mean difference between two groups, and then dividing the result by the pooled standard deviation. The magnitude of the differences in the means as calculated by The online Chen's d calculator was 1.105 as shown in the figure.

Group 1

Mean (M): 8.8

Standard deviation (s): 2.53

Sample size (n): 8

Group 2

Mean (M): 6.00

Standard deviation (s): 2.54

Sample size (n): 5

Calculate Reset

Success!

Cohen's $d = (6 - 8.8) / 2.535005 = 1.104534$.

Reporting Results In Apa Format

An independent-samples t-test was conducted to compare the neurological test scores for swimmers and soccer players. There were no significant differences in scores for swimmers ($M = 8.8$, $SD = 2.53$) and soccer players ($M = 6.00$, $SD = 2.54$); $t(11) = 2.02$, $p = .069$, two-tailed). The magnitude of the differences in the means (mean difference = 2.86, 95% CI: -.265 to 6.015) was (Cohen's $d = 1.105$).

Lesson 28

REPEATED MEASURE T-TEST

The independent-measures design is characterized by the fact that two separate samples are used to obtain the two sets of scores that are to be compared. In this lesson, we examine an alternative strategy known as a repeated-measures design, or a within-subjects design. With a repeated-measures design, two separate scores are obtained for each individual in the sample.

For example, a group of patients could be measured before therapy and then measured again after therapy. Or, response time could be measured in a driving simulation task for a group of individuals who are first tested when they are sober and then tested again after two alcoholic drinks. In each case, the same variable is being measured twice for the same set of individuals; that is, we are literally repeating measurements on the same sample.

A repeated-measures design, or a within-subject design, is one in which the dependent variable is measured two or more times for each individual in a single sample. The same group of subjects is used in all of the treatment conditions.

In a matched-subjects design, each individual in one sample is matched with an individual in the other sample. The matching is done so that the two individuals are equivalent (or nearly equivalent) with respect to a specific variable that the researcher would like to control.

Repeated measure can be made understandable by this example. A research indicates that the color red increases men's attraction to women (Elliot & Niesta, 2008). In the original study, men were shown women's photographs presented on either a white or a red background. Photographs presented on red were rated significantly more attractive than the same photographs mounted on white. In a similar study, a researcher prepares a set of 30 women's photographs, with 15 mounted on a white background and 15 mounted on red. One picture is identified as the test photograph, and appears twice in the set, once on white and once on red. Research question is "are the ratings for the test photograph significantly different when it is presented on a red background compared to a white background?" Each male participant looks through the entire set of photographs and rates the attractiveness of each woman on a 12-point scale.

The T Statistic For A Repeated-Measures Research Design

In repeated-measures research design we use difference scores. Table below shows an example of data from a study that examines this phenomenon. Note that there is one sample of $n = 4$ participants, and that each individual is measured twice. The first score for each person (X_1) is a measurement of reaction time before the medication was administered. The second score (X_2) measures reaction time 1 hour after taking the medication. Because we are interested in how the medication affects reaction time, we have computed the difference between the first score and the second score for each individual. The difference scores, or D values, are shown in the last column of the table.

Person	Before Medication (X_1)	After Medication (X_2)	Difference D
A	215	210	-5
B	221	242	21
C	196	219	23
D	203	228	25
			$\Sigma D = 64$
$M_D = \frac{\Sigma D}{n} = \frac{64}{4} = 16$			

Typically, the difference scores are obtained by subtracting the first score (before treatment) from the second score (after treatment) for each person:

$$\text{Difference score} = D = X_2 - X_1$$

Hypothesis Testing

In repeated-measures research design, the researcher's goal is to use the sample of difference scores to answer questions about the general population. In particular, the researcher would like to know whether there is any difference between the two treatment conditions for the general population. Note that we are interested in a population of *difference scores*.

That is, we would like to know what would happen if every individual in the population were measured in two treatment conditions (X_1 and X_2) and a difference score (D) were computed for everyone. Specifically, we are interested in the mean for the population of difference scores. We identify this population mean difference with the symbol μ_D (using the subscript letter D to indicate that we are dealing with D values rather than X scores). As always, the null hypothesis states that, for the general population, there is no effect, no change, or no difference.

For a repeated-measures study, the null hypothesis states that the mean difference for the general population is zero.

$$H_0: \mu D = 0$$

The alternative hypothesis states that there is a treatment effect that causes the scores in one treatment condition to be systematically higher (or lower) than the scores in the other condition.

$$H_1: \mu D \neq 0$$

For the repeated-measures design, the sample data are difference scores and are identified by the letter *D*, rather than *X*. Therefore, we use *D*s in the formula to emphasize that we are dealing with difference scores instead of *X* values. Also, the population mean that is of interest to us is the population mean difference (the mean amount of change for the entire population), and we identify this parameter with the symbol μD . With these simple changes, the *t* formula for the repeated-measures design becomes:

$$t = (MD - \mu D) / SMD$$

To calculate the estimated standard error, the first step is to compute the variance (or the standard deviation) for the sample of *D* scores.

$$s = \sqrt{SS/n-1} = \sqrt{SS/df}$$

After calculating the variance, calculate the SMD just dividing the variance by taking under root of the sample size of the sample.

$$SMD = s/\sqrt{n} = \sqrt{s^2/n}$$

Running Repeated Measure T-test in SPSS

We will run SPSS repeated measure T-test in SPSS by considering the following example.

Researcher wants to examine the effect of a treatment on depression by measuring a group of $n=8$ participants before and after they receive the treatment. Is there a significant treatment effect?

Use alpha .05, two tails.

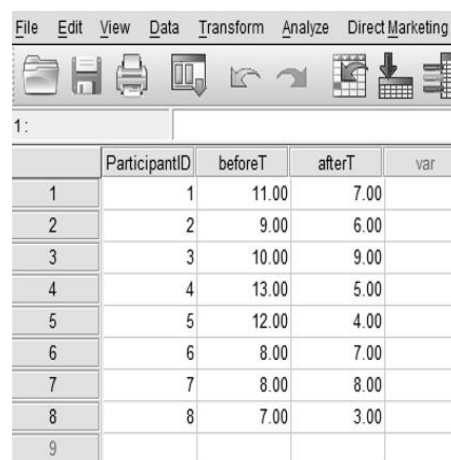
Participants	Before Treatment	After Treatment
1	11	7
2	9	6
3	10	9
4	13	5
5	12	4
6	8	7
7	8	8
8	7	3

We need **two set of variables** to run paired sample t-test in SPSS:

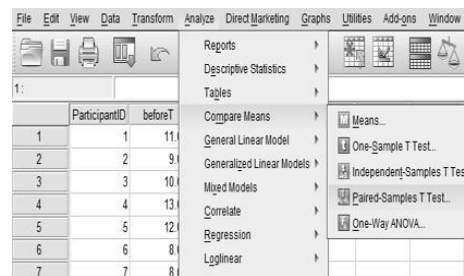
- **One categorical independent variable** (in this case it is treatment with two different levels before treatment, after treatment)
- **One continuous, dependent variable** (e.g. Fear of Statistics Test scores measured on two different occasions or under different conditions).

Entering Data

- First of all, the variables must be setup in the **variable view tab of SPSS data editor such as before treatment and after treatment.**
- We use **participant ID as identifier instead of participant names, as this allows us to collect data while keeping the participants anonymous.**
- The data is entered in the **data view tab of the SPSS editor.**
- **Before T variable represents the depression scores of participants before treatment.**
- **After T variable represents the depression scores of participants after treatment.**
- To run repeated measure t-test in SPSS, Click **Analyze > Compare Means > Paired-Samples T Test on the top menu.**

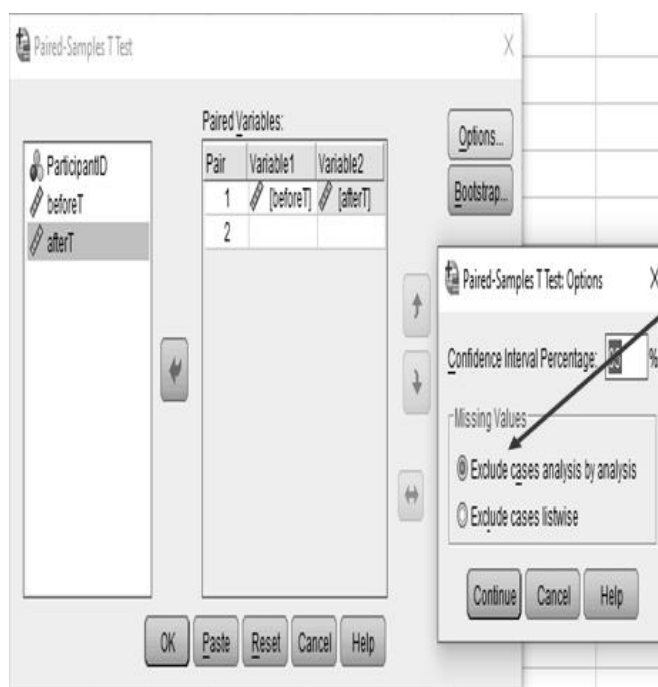


	ParticipantID	beforeT	afterT	var
1	1	11.00	7.00	
2	2	9.00	6.00	
3	3	10.00	9.00	
4	4	13.00	5.00	
5	5	12.00	4.00	
6	6	8.00	7.00	
7	7	8.00	8.00	
8	8	7.00	3.00	
9				



Running Analysis

1. The Paired-Samples T Test window opens where we select the both of these conditions one at a time and click the arrow to move them into the Paired Variables box.
2. The Options section is where we can set our desired confidence interval/level for the mean difference, and specify how SPSS should handle missing values.
3. In SPSS by default confidence interval is set at 95% and we do not need to change it.
4. The Missing Values section is not relevant if we have only specified one dependent variable; it only matters if we are entering more than one dependent (continuous numeric) variable.
5. Then, click Continue and OK button to run the analysis.



Assumptions of Repeated Measure T-test

The related-samples t statistic requires following basic assumptions:

1. Level of Measurement

Just like independent sample t-test, repeated measures also assumes that the dependent variable is measured at the interval or ratio level; that is, using a continuous scale.

2. Random Sampling

Another assumption includes that the scores are obtained using a random sample from the population.

3. Independence of Observations

The observations within each treatment condition must be independent.

This assumption of independence refers to the scores within each treatment.

Inside each treatment, the scores are obtained from different individuals and should be independent of one another.

4. Normal Distribution

This assumption states that the population distribution of difference scores (D values) must be normal.

The normality assumption is not a cause for concern unless the sample size is relatively small.

To test this assumption in SPSS we need to create a new variable that will represent the difference between before treatment and after treatment scores.

- To do this we will go to Transform> Compute Variable.
- Name the target variable (i.e., difference D). The numeric expression will be before treatment – after treatment then click OK.
- Go to Explore option of the Descriptive Statistics menu.
- Click on the newly generated variable (difference D) and move it into the Dependent List box.
- Click on the Statistics button and click on *Descriptives* and Outliers. Click on Continue.
- Click on the Plots button. Under Descriptive, click on Histogram to select it.
- Click on Normality plots with tests. Click on Continue and OK.
- The positive skewness value is suggesting the data is slightly skewed.
- Kurtosis values below 0 indicate a distribution that is relatively flat (too many cases in the extremes).
- The actual shape of the distribution for each group can be seen in the Histograms.
- In Q-Q plot, a reasonably straight line suggests a normal distribution.
- In the table labelled *Tests of Normality*, we are given the results of the Kolmogorov-Smirnov and Shapiro-Wilk statistics.
- The p value is 0.2 a non-significant result (Sig. value of more than .05) would indicate the data is normally distributed.

Interpreting The Output

The Paired Samples Statistics box displays the descriptive statistics for two conditions.

Paired Samples Statistics

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1 before T	9.7500	8	2.12132	.75000
after T	6.1250	8	2.03101	.71807

We can see from the two means that participants have lower depression scores after treatment (M=6.13) than before treatment (M=9.75).

SD also indicates the scores in both conditions are similarly dispersed.

The second table in the SPSS Output tells us how the two conditions relate to one another. This table, however, is not necessary for the interpretation of the t-test results.

Paired Samples Correlation

	N	Correlation	Sig
Pair 1 before T & after T	8	-.091	.830

The Paired Samples Test box helps us decide whether there is a statistically significant difference between the conditions.

We need to look in the final column, labelled Sig. (2-tailed). this is the probability (p) value. **If this value is less than .05 (in this case .012), we can conclude that there is a significant difference between two scores.**

Paired Samples

	Paired Differences					t	df	Sig.(2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95%CI of the difference				
				Lower	Upper			
Pair1 before T& after T	3.6350	3.06769	1.08459	1.06035	6.18965	3.342	7	.012

Effect size

In case of repeated measures t-test, we calculate effect size from Cohen's d (measure of effect size). Effect size will be calculated from simply divides the sample mean difference by the sample standard deviation.

Cohen's d = mean difference / SD

$$d = 3.63 / 3.07 = 1.18$$

Reporting results of repeated measure T-test in APA format

A paired-samples t-test was conducted to evaluate the impact of the intervention on participants' scores on depression. There was a statistically significant decrease in depression scores from before treatment ($M = 9.75$, $SD = 2.12$) to after treatment ($M = 6.13$, $SD = 2.03$), $t(7) = 3.34$, $p < .05$ (two-tailed) and Cohen's $d = 1.18$. The mean decrease in depression scores was 3.62 with a 95% confidence interval ranging from 1.06 to 6.19.

Lesson 29

ANALYSIS OF VARIANCE (ANOVA)

Analysis of variance (ANOVA) is a hypothesis-testing procedure that is used to evaluate mean differences between two or more treatments (or populations).

While running ANOVA we must decide between two interpretations:

1. There really are no differences between the populations (or treatments).
2. The populations (or treatments) really do have different means.

When a researcher manipulates a variable to create the treatment conditions in an experiment, the variable is called an independent variable. When a researcher uses a non-manipulated variable to designate groups, the variable is called a quasi-independent variable.

In ANOVA, the variable (independent or quasi-independent) that designates the groups being compared is called a factor. The individual groups or treatment conditions that are used to make up a factor are called the levels of the factor. ANOVA can be used to evaluate the results from a research study that involves more than one factor.

For example, a researcher may want to compare two different therapy techniques, examining their immediate effectiveness as well as the persistence of their effectiveness over time.

A study that combines two factors, is called a two-factor design or a factorial design (Two way ANOVA). A study that examines only one independent variable is called single-factor design (or One way ANOVA).

Hypotheses for (ANOVA)

In analysis of variance the hypothesis is as follows:

The null hypothesis states that there is no treatment effect.

$$H_0: \mu_1 = \mu_2 = \mu_3$$

The alternative hypothesis states that there is at least one mean difference among the populations.

$$H_1: \mu_1 \neq \mu_2 \neq \mu_3$$

OR

$$H_1: \mu_1 = \mu_3 \text{ but } \mu_2 \text{ is different}$$

In ANOVA we use variance to measure sample mean differences when there are two or more samples. The test statistic for ANOVA uses this fact to compute an F-ratio with the following structure:

$$F = \text{Variance between treatments} / \text{Error Variance (within group variance)}$$

Logic Behind Anova And Its Assumptions

The first step in ANOVA is to **determine the total variability** for the entire set of data. (one general measure of variability for the complete experiment).

The word analysis means **dividing into smaller parts**. Because we are going to analyze variability, the process is called analysis of variance.

The analysis process divides the total variability into two basic components.

1. *Between-Treatments Variance*

This simply measures how much difference exists between the treatment conditions.

There are two possible explanations for these between-treatment differences:

The differences are the **result of sampling error**.

The differences between treatments have been **caused by the treatment effects**.

2. *Within-Treatment Variance*

Inside each treatment condition, a set of individuals receive exactly the same treatment but their scores differ.

The differences that exist within a treatment represent random and unsystematic differences when there are no treatment effects

Thus, the within-treatments variance provides a measure of how big the differences are when H_0 is true.

The F-ratio: The Test Statistic For ANOVA

Once the total variability is analyzed into two basic components (between treatments and within treatments), we simply compare them.

The comparison is made by computing an F-ratio as:

$F = \text{Variance between treatments} / \text{Variance within treatments}$

For ANOVA, the denominator of the F-ratio is called the error term. The error term provides a measure of the variance caused by random, unsystematic differences. Because ANOVA typically is used to examine data from more than two treatment conditions (and more than two samples), we need a notational system to keep track of all the individual scores and totals.

Assumptions For The ANOVA

- The observations within each sample **must be independent.**
- The populations from which the samples are selected **must be normal.**
- The populations from which the samples are selected **must have equal variances** (homogeneity of variance).
- **IV is always categorical and DV is continues**
- The assumption of **homogeneity of variance** is an important one.
- If a researcher suspects that **it has been violated**, it can be tested by **Welch and Brown-Forsythe tests.**
- If one of the assumptions for the independent-measures ANOVA has been violated, an **alternative statistical analysis known as the Kruskal-Wallis test can be used.**

Calculating ANOVA by Hand

In calculating ANOVA, the **letter k is used to identify the number of treatment conditions, that is, the number of levels of the factor.** For an independent-measures study, k also specifies the number of separate samples. **The number of scores in each treatment is identified by a lowercase letter n .**

For the final F-ratio we need an MS (variance) between treatments for the numerator and an MS (variance) within treatments for the denominator. In each case

$$F = s^2_{\text{between}} / s^2_{\text{within}}$$

$$F = MS_{\text{between}} / MS_{\text{within}}$$

$$MS = s^2 = SS / df$$

In ANOVA, each of the two variances in the F-ratio is calculated using the basic formula for sample variance.

$s^2 = SS / df$. SS total is the sum of squares for the entire set of N scores and calculated as:

$$SS_{\text{total}} = \sum X^2 - (\sum X)^2 / N$$

Firstly we will calculate within-treatments sum of squares (SS within treatments). SS within is the variability inside each treatment condition. To find the overall within-treatment sum of squares, we simply add these values together:

$$SS_{\text{within treatment}} = \sum SS_{\text{inside each treatment}}$$

Secondly we will calculate between-treatments sum of squares (SS between treatments). The value for SS between treatments can be found simply by subtraction:

SS between treatments = SS_{total} - SS_{within}

After calculating the sum of squares, let calculate the total degrees of freedom (*df* total). To find the *df* associated with SS total we compute *df* as:

When all of these individual treatment values are added together, we obtain within-treatments degrees of freedom (*df* within)

$df_{\text{within}} = \sum(n-1) = \sum df \text{ in each treatment}$

OR

$df_{\text{within}} = N - k$

To find *df* between, simply count the number of treatments and subtract 1.

$df_{\text{between}} = K - 1$

ANOVA Summary Tables

It is useful to organize the results of the analysis in one table called an ANOVA summary table.

The table shows the source of variability (between treatments, within treatments, and total variability), SS, *df*, MS, and F and it is constructed as shown:

Source	SS	df	MS	
Between treatments	30	2	15	F=11.28
Within treatments	16	12	1.33	
Total	46	14		

Effect Size

For ANOVA, the simplest and most direct way to measure effect size is to compute the percentage of variance accounted for by the treatment conditions. We determine how much of the total SS is accounted for by the SS between treatments. The percentage of variance accounted for by the treatment effect is usually called η^2 (the Greek letter eta squared) instead of using r^2 .

For this example:

$\eta^2 = \text{SS between treatments} / \text{SS total}$

Post Hoc Tests

Post hoc tests are additional hypothesis tests that are done after an ANOVA to determine exactly which mean differences are significant and which are not. As the name implies, post hoc tests are done after an ANOVA. More specifically, these tests are done after ANOVA when, you reject H_0 and there are three or more treatments.

In general, a post hoc test enables us to go back through the data and compare the individual treatments two at a time. In statistical terms, this is called *making pairwise comparisons*. **Tukey's HSD** test is a most commonly used post hoc test in psychological research. Tukey's test allows to compute a single value that determines the minimum difference between treatment means that is necessary for significance.

Lesson 30

ONE-WAY ANOVA**One-Way ANOVA in SPSS**

The one-way analysis of variance (ANOVA) is used to determine whether there are any statistically significant differences between the means of two or more independent (unrelated) groups (although you tend to only see it used when there are a minimum of three, rather than two groups).

For example, you could use a one-way ANOVA to understand whether exam performance differed based on test anxiety levels among students, dividing students into three independent groups (e.g., low, medium and high-stressed students). Also, it is important to realize that the one-way ANOVA is an omnibus test statistic and cannot tell you which specific groups were statistically significantly different from each other; it only tells you that at least two groups were different. Since you may have three, four, five or more groups in your study design, determining which of these groups differ from each other is important. You can do this using a post hoc test.

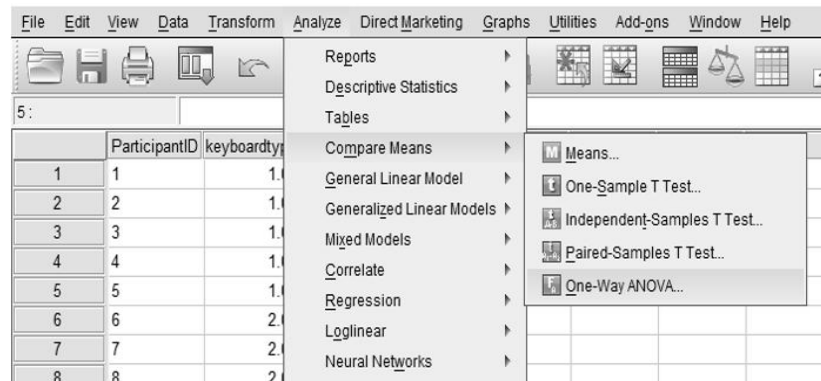
Entering Data in SPSS

Let understand the entering data into SPSS through keyboard example. We separate the groups for analysis by creating a grouping variable called Keyboard type (i.e., the independent variable. Number of errors committed is entered under the variable name errors (i.e., the dependent variable).

- By clicking in the box of the 'values' field in the row containing the groups variable, we get a pop-up dialog that allows us to code the categorical variable. Three types of keyboard are coded.
- The data is entered in the data view tab of the SPSS editor.
- A code is entered for the categories of independent variable as (1 = keyboard A, 2 = keyboard B, and 3 = keyboard C) and the number of committed errors in errors column.

Running One-Way ANOVA in SPSS

- To run one-way ANOVA in SPSS, Click Analyze > Compare Means > One-Way ANOVA on the top menu.



- The One-Way ANOVA window opens.
- Click on your dependent (continuous) variable (e.g. errors). Move this into the box marked **Dependent List** by clicking on the arrow button.
- Click on your independent, categorical variable (e.g. keyboard type). Move this into the box labelled **Factor**.
- Click the Options button and click on Descriptive, Homogeneity of variance test, Brown-Forsythe, Welch and Means Plot.
- For Missing values, make sure there is a dot in the option marked Exclude cases analysis by analysis. Click on Continue.
- Click on the button marked Post Hoc. Click on Tukey.
- Click on Games-Howell under the section Equal Variances Not Assumed
- Click on Continue and then OK.

Assumptions of One-Way ANOVA

ANOVA has the same general assumptions that apply to all of the parametric tests with an additional assumption of homogeneity of variances.

The general assumptions include;

1. **Level of measurement** (DV measured on interval or ratio scale, i.e., continuous)
2. **Random sampling**
3. **Independence of observations**
4. **Normal distribution** (by assessing skewness, kurtosis, histogram, Q-Q plots, Shapiro-Wilk and Kolmogorov-Smirnov tests)

5. Homogeneity of Variance

The homogeneity of variance option gives us the *Levene's test* for homogeneity of variances, which tests whether the variance in scores is the same for each of the three groups.

Check the significance value (Sig.) for Levene's test. If this number is greater than .05, the assumption is not violated.

If the assumption is found to be violated then the F test of the ANOVA is not robust enough to be used. In this case, there is need to consult the table in the output headed Robust Tests of Equality of Means.

The two tests there (*Welch and Brown-Forsythe*) are preferable when the assumption of the homogeneity of variance is violated.

Interpreting the Output

The descriptive table gives us the information about each group (number in each group, means, standard deviation, minimum and maximum, etc.)

The ANOVA table gives both between-groups and within-groups sums of squares, degrees of freedom etc.

The main thing is the Sig. (the p value). If the Sig. value is *less than or equal to .05*, there is a significant difference somewhere among the mean scores on dependent variable for the three groups.

ANOVA

typing performance in terms of errors

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	70.000	2	35.000	9.130	.004
Within Groups	46.000	12	3.833		
Total	116.000	14			

In this example the overall Sig. value is .004, which is less than .05, indicating a statistically significant result somewhere among the groups with an F-ratio of 9.13.

ANOVA

typing performance in terms of errors

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	70.000	2	35.000	9.130	.004
Within Groups	46.000	12	3.833		
Total	116.000	14			

The post-hoc tests table tells us exactly where the differences among the groups occur.

In the column labelled *Mean Difference* look for any asterisks (*) next to the values listed. If there is an (*), this means that the two groups being compared are significantly different from one another at the $p < .05$ level.

The exact significance value is given in the column labelled Sig.

There is a significant difference in errors (typing performance) between the groups that used the keyboard A and keyboard B ($p = .018$), as well as between the keyboard A and keyboard C ($p = .004$).

However, there were no differences between the groups that used keyboard B and keyboard C ($p = .706$).

Post Hoc Tests**Multiple Comparisons**

Dependent Variable: typing performance in terms of errors

		Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Tukey HSD	(I) keyboardtype (J) keyboardtype					
	keyboard A keyboard B	-4.00000*	1.23828	.018	-7.3036	-.6964
	keyboard A keyboard C	-5.00000*	1.23828	.004	-8.3036	-1.6964
	keyboard B keyboard A	4.00000*	1.23828	.018	.6964	7.3036
	keyboard B keyboard C	-1.00000	1.23828	.706	-4.3036	2.3036
	keyboard C keyboard A	5.00000*	1.23828	.004	1.6964	8.3036
Games-Howell	keyboard A keyboard B	-4.00000*	1.26491	.034	-7.6623	-.3377
	keyboard A keyboard C	-5.00000*	1.14018	.006	-8.2620	-1.7380
	keyboard B keyboard A	4.00000*	1.26491	.034	.3377	7.6623
	keyboard B keyboard C	-1.00000	1.30384	.733	-4.7501	2.7501
	keyboard C keyboard A	5.00000*	1.14018	.006	1.7380	8.2620
	keyboard C keyboard B	1.00000	1.30384	.733	-2.7501	4.7501

*. The mean difference is significant at the 0.05 level.

Effect Size

Although SPSS does not generate effect size for this analysis, it is possible to determine the effect size for this result.

The information that is needed to calculate eta squared, is provided in the ANOVA table.

$$\eta^2 = SS_{\text{between}}/SS_{\text{total}}$$

$$\eta^2 = 70/116 = 0.603$$

Reporting Results in APA

A one-way between-groups analysis of variance was conducted to explore the impact of keyboard type on typing performance, as measured by the number of errors committed. Participants were divided into three groups according to their use of keyboard (Group 1: keyboard A; Group 2: keyboard B; Group 3: keyboard C). The analysis of variance indicates that there are significant differences among the three keyboard types, $F(2,12) = 9.13$, $p = .004$, $\eta^2 = 0.60$.

Post-hoc comparisons using the Tukey HSD test indicated that the number of errors were statistically significantly higher after using the keyboard B ($M = 5.00$, $SD = 2.24$, $p = .018$) and keyboard C ($M = 6.00$, $SD = 1.87$, $p = .004$) as compared to keyboard A ($M = 1.00$, $SD = 1.73$). There were no statistically significant differences between the keyboard B and keyboard C ($p = .706$).

Lesson 31

TWO-WAY ANOVA-I

Introduction to Two-Way ANOVA

Previously we discussed one way ANOVA which is an experimental design having only one independent variable. In this chapter, we will extend the analysis of variance to cover experimental designs involving two or more independent variables.

In real life, variable rarely exists or relate to other variables in isolation. To examine more complex situations, researchers often design research studies that include more than one independent variable. Researchers systematically change two (or more) variables and then observe how the changes influence another (dependent) variable. When a research study involves more than one factor, it is called a *factorial design*. And when a research study examines ANOVA having exactly two factors, it is called Two-Way ANOVA. Two-way ANOVA means groups are defined by two independent variables. The two independent variables in a two-factor experiment are identified as factor A and factor B. the goal of the factorial design is to evaluate the mean differences that may be produced by either of the factors acting independently or by the two factors working together. This technique allows us to look at the individual and joint effect of two independent variables on one dependent variable

Main Effects and Interactions

The mean differences among the levels of one factor are referred to as the main effect of that factor. When the design of the research study is represented as a matrix with one factor determining the rows and the second factor determining the columns, then the mean differences among the rows describe the main effect of one factor, and the mean differences among the columns describe the main effect for the second factor.

For factorial ANOVA, we took an example of Shrauger's study. Shrauger (1972) tested participants on a concept-formation task. Half of the participants worked alone (no audience), and half worked with an audience of people who claimed to be interested in observing the experiment. Shrauger also divided the participants into two groups on the basis of personality: those high in

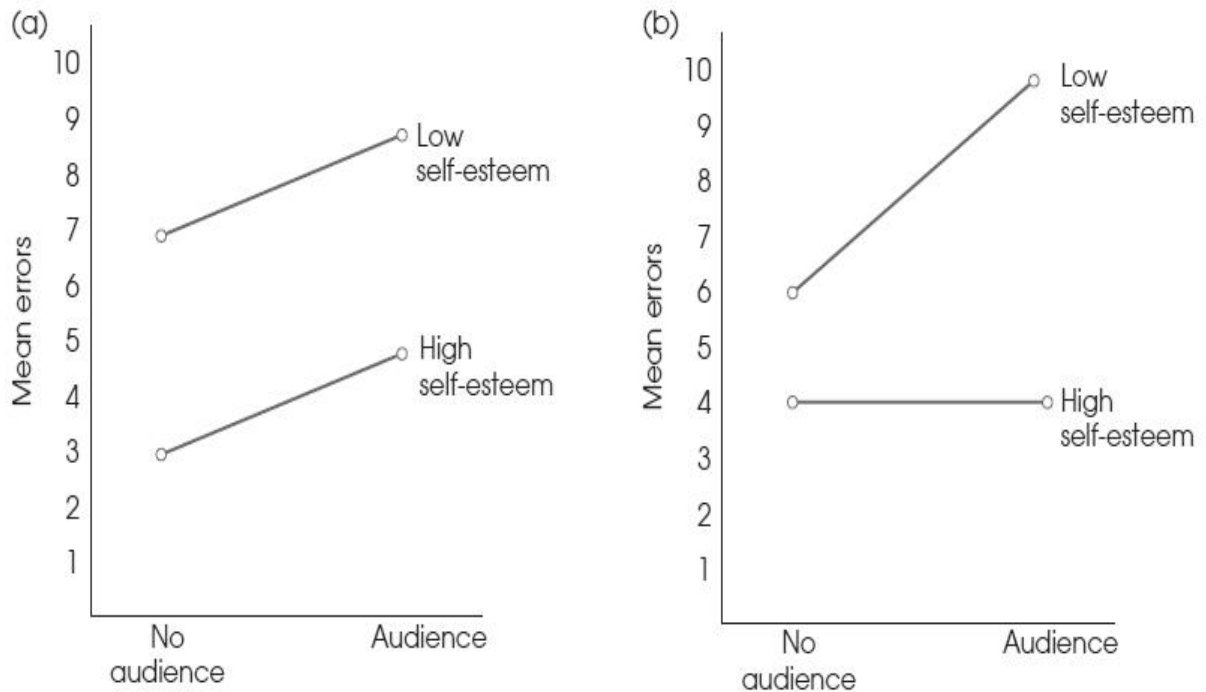
Self-esteem and those low in self-esteem. The dependent variable for this experiment was the number of errors on the concept formation task. Data similar to those obtained by Shrauger are

shown in Figure below. Notice that the audience had no effect on the high-self-esteem participants. However, the low-self-esteem participants made nearly twice as many errors with an audience

as when working alone. Note that the study involves two separate factors: One factor is manipulated by the researcher, changing from no-audience to audience, and the second factor is self-esteem, which varies from high to low. The two factors are used to create a *matrix* with the different levels of self-esteem defining the rows and the different audience conditions defining the columns. The resulting two-by-two matrix shows four different combinations of the variables, producing four different conditions. Thus, the research study would require four separate samples, one for each *cell*, or box, in the matrix. The dependent variable for the study is the number of errors on the concept formation task for people observed in each of the four conditions.

	No Audience	Audience	
Low	$M = 7$	$M = 9$	$M = 8$
High	$M = 3$	$M = 5$	$M = 4$
	$M = 5$	$M = 7$	

The real advantage of combining two factors within the same study is the ability to examine not only the main effects but the unique effects caused by an interaction. The concept of interaction can be defined in terms of the pattern displayed in the graph.



To construct this figure, we select one of the factors to be displayed on the horizontal axis; in this case, the different levels of the audience factor. The dependent variable, the number of errors, is shown on the vertical axis. The figure actually contains two separate graphs: The top line shows the relationship between the audience factor and errors for the low-self-esteem and the bottom line shows the relationship for the high-self-esteem participants.

The $A \times B$ interaction typically is called the “ A by B ” interaction. If there is an interaction between an audience and self-esteem, it may be called the “audience by self-esteem” interaction.

Hypotheses for Two-Way (ANOVA)

In ANOVA, the variable (independent or quasi-independent) that designates the groups being compared is called a **factor**. The individual conditions or values that make up a factor are called **the levels of the factor**. It consists of three hypothesis tests, two for main effects and one for interaction.

- H_0 : No main effect of A
- H_1 : There exists a main effect of factor A
- H_0 : No main effect of B
- H_1 : There exists a main effect of factor B

- H_0 : No interaction
- H_1 : There exists an interaction between factor A and factor B

Degrees of Freedom (df)

For two-way ANOVA, we calculate three kinds of degrees of freedom.

1. Total Degrees of Freedom (df_{total})
2. Within-Treatments Degrees of Freedom (df_{within})
3. Between-Treatments Degrees of Freedom ($df_{between}$)

Assumptions for the Independent- Measures ANOVA

The independent-measures ANOVA requires the same three assumptions that were necessary for the independent-measures t hypothesis test:

1. The observations within each sample **must be independent**.
2. The populations from which the samples are selected **must be normal**.
3. The populations from which the samples are selected **must have equal variances** (homogeneity of variance).

Two-Way ANOVA in SPSS

To run two-way ANOVA we need three variables:

- Two categorical independent variables (e.g. sex: males/females; age group: young, middle, old)
- One continuous dependent variable (e.g. optimism).

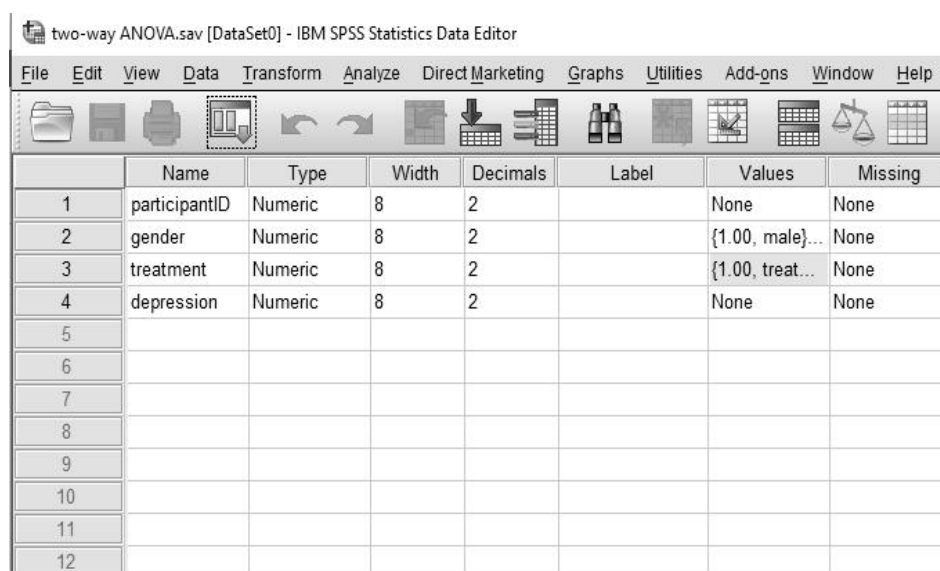
Factor A Gender	Factor B Treatments		
	Male	I	II
		3	2
		8	8
		9	7
		4	7
		0	12
		0	6

	Female	2	9
		6	13

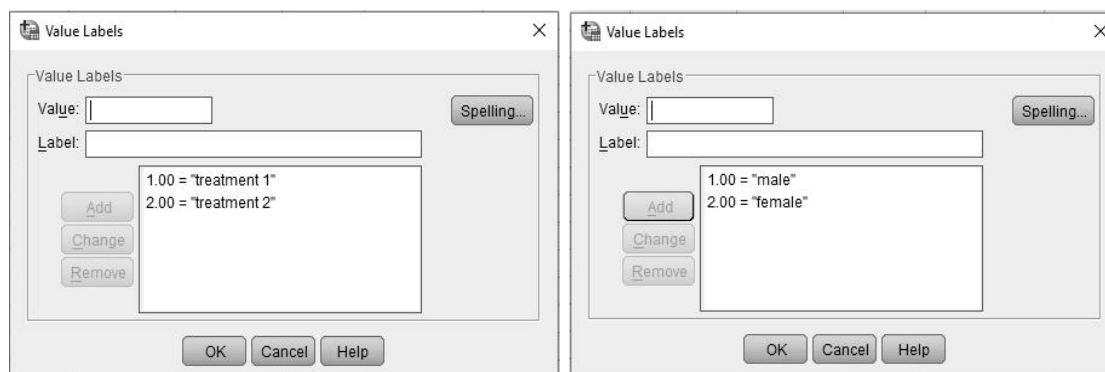
The following data are from a two-factor study examining the depression scores across gender and two treatment conditions. Use an ANOVA with $\alpha=.05$ for all tests to evaluate the significance of the main effects and the interaction.

We need to create two grouping variables in the variable view of SPSS as, **treatment and gender** (i.e., the independent variables).

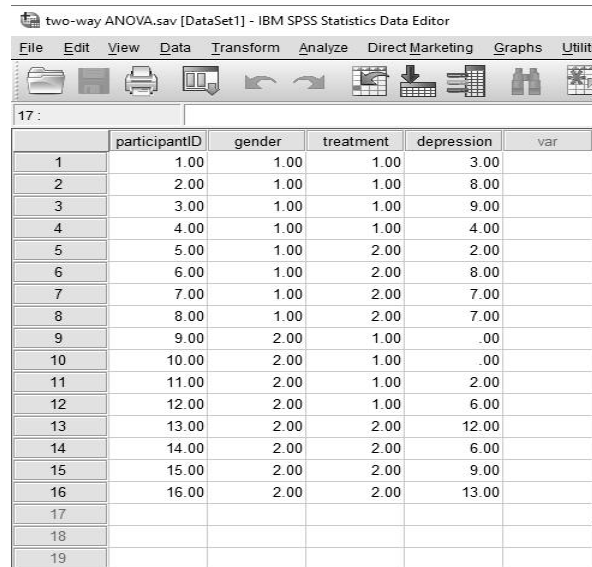
Depression scores are entered under the variable name **depression** (i.e., the dependent variable).



In the box of the 'values' we get a pop-up dialog that allows us to code the categorical variable. Gender and Treatment conditions are coded as shown in the value labels window.

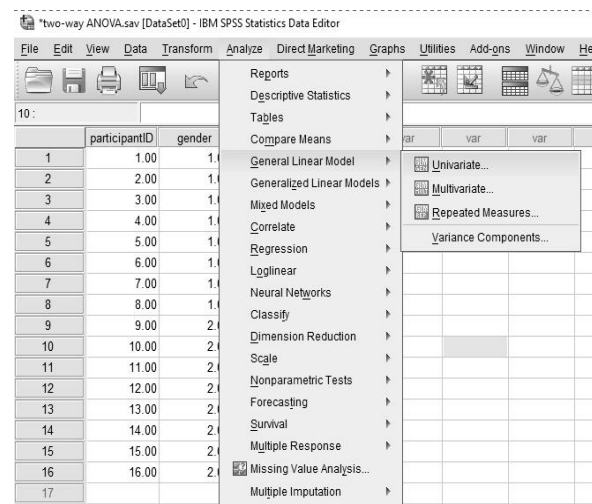


While entering data A code is entered for the categories of independent variables as (**Gender:** 1=male, 2=female; **Treatment:** 1=treatment1, 2=treatment2) and the **depression** scores in the depression column.



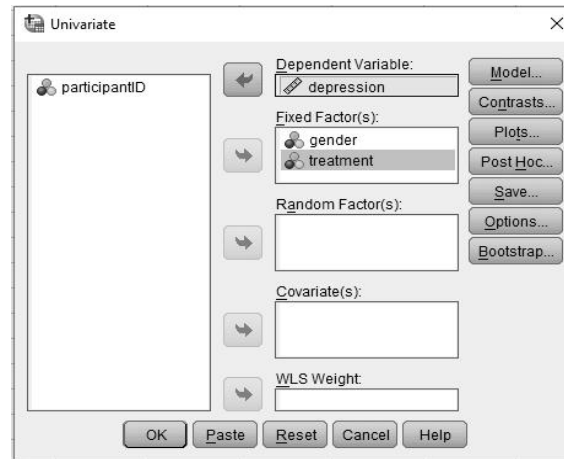
	participantID	gender	treatment	depression	var
1	1.00	1.00	1.00	3.00	
2	2.00	1.00	1.00	8.00	
3	3.00	1.00	1.00	9.00	
4	4.00	1.00	1.00	4.00	
5	5.00	1.00	2.00	2.00	
6	6.00	1.00	2.00	8.00	
7	7.00	1.00	2.00	7.00	
8	8.00	1.00	2.00	7.00	
9	9.00	2.00	1.00	.00	
10	10.00	2.00	1.00	.00	
11	11.00	2.00	1.00	2.00	
12	12.00	2.00	1.00	6.00	
13	13.00	2.00	2.00	12.00	
14	14.00	2.00	2.00	6.00	
15	15.00	2.00	2.00	9.00	
16	16.00	2.00	2.00	13.00	
17					
18					
19					

To run one-way ANOVA in SPSS, from the menu at the top of the screen, click on **Analyze > General Linear Model > Univariate.**

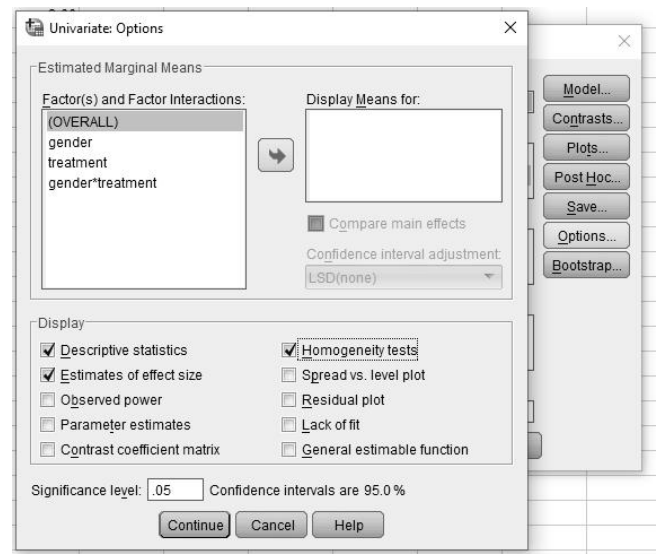


- A **Univariate** window opens.
- Click on the dependent, continuous variable (e.g. depression) and move it into the box labelled **Dependent Variable**.

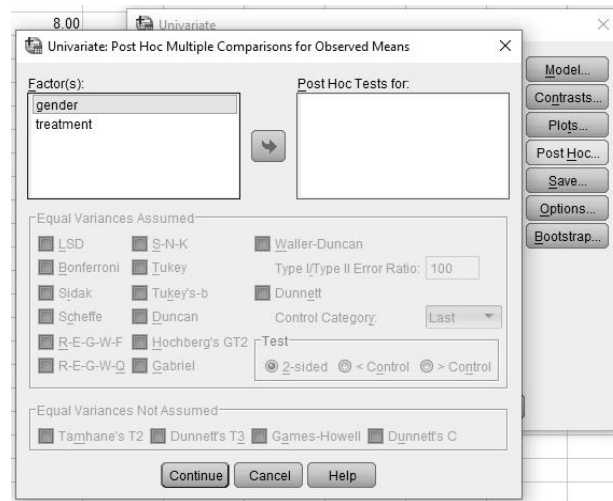
- Click on two independent, categorical variables (gender, treatment) and move these into the box labelled **Fixed Factors**.
-



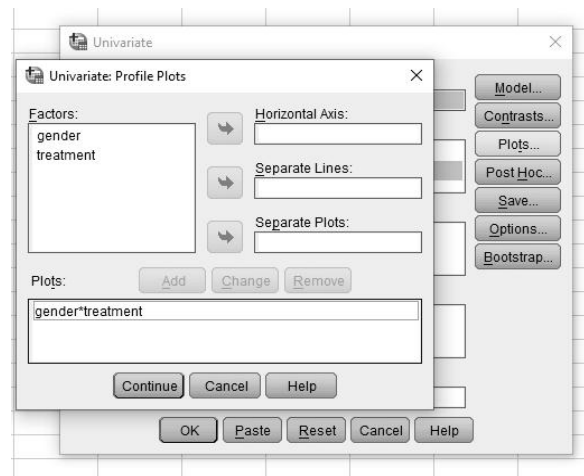
- Click on the **Options** button. Click on **Descriptive Statistics**, **Estimates of effect size** and **Homogeneity tests** and then **Continue** when finished.



- Click on the **Post Hoc** button.
- From the **Factors** listed on the left-hand side, choose the independent variable(s) you are interested in (only if the variables have more than two levels).
- Click on the arrow button to move it into the **Post Hoc Tests for** section.
- Choose the test **Tukey** and click on **Continue**.



- Click on the **Plots** button.
- In the **Horizontal Axis** box, put the independent variable that has the most groups (e.g. gender).
- In the box labelled **Separate Lines**, put the other independent variable (e.g. treatment).
- Click on **Add**.
- In the section labelled **Plots**, you will see the two variables listed as gender*treatment
- Click on **Continue** and then **OK**.



Assumptions of Two-Way ANOVA

The general assumptions include;

1. Random sampling
2. Independence of observations

3. **Normal distribution** (by assessing skewness, kurtosis, histogram, Q-Q plots, Shapiro-Wilk and Kolmogorov-Smirnov tests).

Homogeneity of Variance

- **Levene's test for homogeneity of variances** allows us to use the sample variances from data to determine whether there is evidence for any differences among the population variances.
- The **Levene's test for homogeneity of variances** tests whether the variance in scores is the same for each of the groups.
- For the assumption to be fulfilled the **Sig. value should be greater than .05** and therefore *non-significant* results. A significant result (Sig. value less than .05) suggests that the variance of dependent variable across the groups is not equal.

Levene's Test of Equality of Error Variances^a

Dependent Variable: depression

F	df1	df2	Sig.
.200	3	12	.894

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + gender + treatment + gender * treatment

- The **Descriptive Statistics** table provides the **Mean** scores, **Std. deviations** and **N** for each subgroup.
- The main output from two-way ANOVA is a table labelled **Tests of Between-Subjects Effects**.
- This table gives us the *main effects* and *interaction effects*.
- The first thing to do is to check for the possibility of an interaction effect (e.g., that the influence of treatments on depression scores depends on whether the individual is male or female).

Tests of Between-Subjects Effects

Dependent Variable: depression

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	128.000 ^a	3	42.667	5.020	.018	.557
Intercept	576.000	1	576.000	67.765	.000	.850
gender	.000	1	.000	.000	1.000	.000
treatment	64.000	1	64.000	7.529	.018	.386
gender * treatment	64.000	1	64.000	7.529	.018	.386
Error	102.000	12	8.500			
Total	806.000	16				
Corrected Total	230.000	15				

a. R Squared = .557 (Adjusted R Squared = .446)

- If the interaction effect is significant, the main effects cannot be interpreted easily.
- This is because, in order to describe the influence of one of the independent variables, there is need to specify the level of the other independent variable.
- The output we need to look at is labelled **gender*treatment**.

Tests of Between-Subjects Effects

Dependent Variable: depression

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	128.000 ^a	3	42.667	5.020	.018	.557
Intercept	576.000	1	576.000	67.765	.000	.850
gender	.000	1	.000	.000	1.000	.000
treatment	64.000	1	64.000	7.529	.018	.386
gender * treatment	64.000	1	64.000	7.529	.018	.386
Error	102.000	12	8.500			
Total	806.000	16				
Corrected Total	230.000	15				

a. R Squared = .557 (Adjusted R Squared = .446)

- In this example, the interaction effect is significant ($F = 7.529$, $p = .018$).
- This indicates that there is significant difference in the effect of treatment on depression males and females.

Tests of Between-Subjects Effects

Dependent Variable: depression

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	128.000 ^a	3	42.667	5.020	.018	.557
Intercept	576.000	1	576.000	67.765	.000	.850
gender	.000	1	.000	.000	1.000	.000
treatment	64.000	1	64.000	7.529	.018	.386
gender*treatment	64.000	1	64.000	7.529	.018	.386
Error	102.000	12	8.500			
Total	806.000	16				
Corrected Total	230.000	15				

a. R Squared = .557 (Adjusted R Squared = .446)

- To determine whether there is a main effect for each independent variable, in the left-hand column, find the variable (e.g., gender) and check in the column marked Sig. next to each variable. **If the value is less than or equal to .05 there is a significant main effect for that independent variable.**

Tests of Between-Subjects Effects

Dependent Variable: depression

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	128.000 ^a	3	42.667	5.020	.018	.557
Intercept	576.000	1	576.000	67.765	.000	.850
gender	.000	1	.000	.000	1.000	.000
treatment	64.000	1	64.000	7.529	.018	.386
gender*treatment	64.000	1	64.000	7.529	.018	.386
Error	102.000	12	8.500			
Total	806.000	16				
Corrected Total	230.000	15				

a. R Squared = .557 (Adjusted R Squared = .446)

- In the output shown in the table, there is no significant main effect for gender ($p = 1.00$) but a significant main effect for treatment ($p = .018$).
- This means that males and females do not differ in terms of their depression scores, but there is a difference in scores for treatment I and treatment II.

Tests of Between-Subjects Effects

Dependent Variable: depression

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	128.000 ^a	3	42.667	5.020	.018	.557
Intercept	576.000	1	576.000	67.765	.000	.850
gender	.000	1	.000	.000	1.000	.000
treatment	64.000	1	64.000	7.529	.018	.386
gender * treatment	64.000	1	64.000	7.529	.018	.386
Error	102.000	12	8.500			
Total	806.000	16				
Corrected Total	230.000	15				

a. R Squared = .557 (Adjusted R Squared = .446)

- The effect size for the independent variables is provided in the column labelled **Partial Eta Squared** (e.g., treatment, $p = .386$).

Tests of Between-Subjects Effects

Dependent Variable: depression

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	128.000 ^a	3	42.667	5.020	.018	.557
Intercept	576.000	1	576.000	67.765	.000	.850
gender	.000	1	.000	.000	1.000	.000
treatment	64.000	1	64.000	7.529	.018	.386
gender * treatment	64.000	1	64.000	7.529	.018	.386
Error	102.000	12	8.500			
Total	806.000	16				
Corrected Total	230.000	15				

a. R Squared = .557 (Adjusted R Squared = .446)

- ANOVA output also provides Post-hoc tests if the post hoc option is checked in the analysis procedure since these are relevant only if we have more than two levels (groups) of an independent variable.
- However, there is no need to look at them until you find a significant main effect or interaction effect in the overall (omnibus) analysis of variance test.

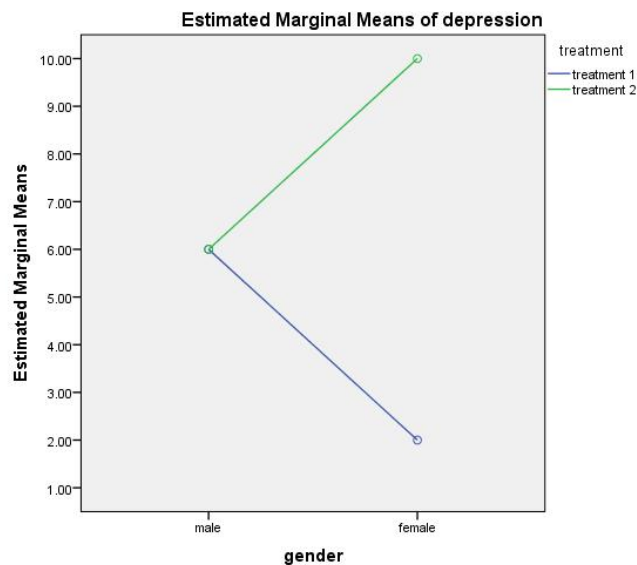
Lesson 32

TWO-WAY ANOVA-II

Interpreting the Plots

The output gives a plot of the depression scores for males and females, across the treatments. This plot is very useful for allowing us to visually inspect the relationship among variables. Although presented last, the plots are often useful to inspect first to help us better understand the impact of two independent variables. **The main effect for gender is non-significant and the plot indicates that the mean score of depression for both males and females is same.** The main effect for treatment is based on the tendency for the depression scores to increase from treatment I to treatment II as shown in graph.

Profile Plots



The plots show the interdependence of both factors. Treatment I is likely to produce a decrease in depression scores for females than males. On the other hand, treatment II influences females more than males. That is, females are more likely to have increase in depression scores after treatment II.

Anova Is Robust Test

Each time we do a hypothesis test, we select an alpha level that determines the risk of a Type I error. With $\alpha=.05$, for example, there is a 5%, or a 1-in-20, risk of a Type I error.

Often a single experiment requires several hypothesis tests to evaluate all the mean differences.

However, each test has a risk of a Type I error, and the more tests we do, the more risk there is.

For this reason, researchers often make a distinction between the *test wise alpha level* and the

experiment wise alpha level. The test wise alpha level is simply the alpha level that is selected

select for each individual hypothesis test. The experiment wise alpha level is the total probability of a Type I error accumulated from all of the separate tests in the experiment.

For example, an experiment involving three treatments would require three separate t tests to compare all of the mean differences:

- Test 1 compares treatment I with treatment II.
- Test 2 compares treatment I with treatment III.
- Test 3 compares treatment II with treatment III.

The advantage of ANOVA is that it performs all three comparisons simultaneously in one

hypothesis test. Thus, no matter how many different means are being compared, ANOVA uses one test with one alpha level to evaluate the mean differences, and thereby avoids the problem of an inflated experiment wise alpha level. ANOVA or F-test is also robust to moderate departures from normality when sample sizes are reasonably large and are equal.

The ANOVA, therefore, can tolerate data that is non-normal (skewed or kurtotic distributions) with only a small effect on the Type I error rate.

Reporting Results

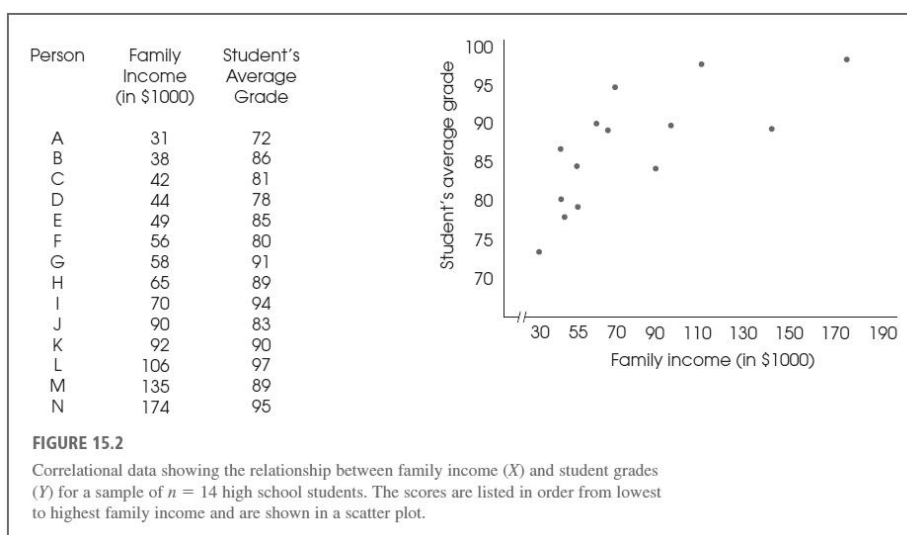
A two-way between-groups analysis of variance was conducted to explore the impact of gender and treatment on depression. The interaction effect between gender and treatment was statistically significant, $F(1, 12) = 7.53, p < .05$. There was a statistically significant main effect for treatment (Treatment I, $M=4, SD=3.42$; Treatment II, $M=8, SD=3.46$), $F(1, 12) = 7.53, p < .05, \eta^2 = 0.37$. The main effect for gender was not significant (Male, $M=6, SD=2.61$; Female, $M=6, SD=5.09$), $F(1, 12) = .00, p = 1.00$.

Lesson 33

CORRELATION ANALYSIS-I

Correlation is a statistical technique that is used to measure and describe the relationship between two variables. Usually the two variables are simply observed as they exist naturally in the environment—there is no attempt to control or manipulate the variables. For example, a researcher could check high school records (with permission) to obtain a measure of each student's academic performance, and then survey each family to obtain a measure of income. The resulting data could be used to determine whether there is relationship between high school grades and family income. Notice that the researcher is not manipulating any student's grade or any family's income, but is simply observing what occurs naturally.

You also should notice that a correlation requires two scores for each individual (one score from each of the two variables). These scores normally are identified as X and Y . The pairs of scores can be listed in a table, or they can be presented graphically in a scatter plot (Figure 15.2). In the scatter plot, the values for the X variable are listed on the horizontal axis and the Y values are listed on the vertical axis. Each individual is then represented by a single point in the graph so that the horizontal position corresponds to the individual's X value and the vertical position corresponds to the Y value. The value of a scatter plot is that it allows you to see any patterns or trends that exist in the data.



Applications of Correlation

Although correlations have a number of different applications, a few specific examples include:

1. **Correlation is used for prediction:** In general, the squared correlation (r^2) measures the gain in accuracy that is obtained from using the correlation for prediction. The squared correlation measures the proportion of variability in the data that is explained by the relationship between X and Y . It is sometimes called the *coefficient of determination*.
2. Correlation is used to demonstrate the **validity** of the test. Validity refers to whether or not the test measures what it claims to measure.
3. Correlations are also used to determine **reliability**. A measurement procedure is considered reliable to the extent that it produces stable, consistent measurements.

Furthermore, the prediction of the theory could be tested by determining the correlation between the two variables. The value of a correlation can be affected greatly by the range of scores represented in the data. The value of a correlation can be affected greatly by the range of scores represented in the data. One or two extreme data points, often called outliers, can have a dramatic effect on the value of a correlation.

The Characteristics of a Relationship

A correlation is a numerical value that describes and measures three characteristics of the relationship between X and Y . These three characteristics are as follows:

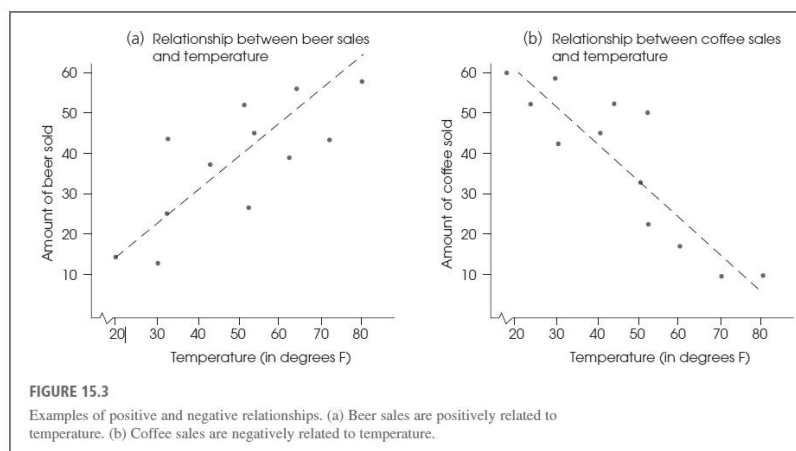
1. The Direction of the Relationship. The sign of the correlation, **positive or negative**, describes the direction of the relationship.

In a **positive correlation**, the two variables tend to change in the same direction: As the value of the X variable increases from one individual to another, the Y variable also tends to increase; when the X variable decreases, the Y variable also decreases.

In a **negative correlation**, the two variables tend to go in opposite directions. As the X variable increases, the Y variable decreases. That is, it is an inverse relationship.

Suppose you run the drink concession at the football stadium. After several seasons, you begin to notice a relationship between the temperature at game time and the beverages you sell. Specifically, you have noted that when the temperature is low, you sell relatively little beer. However, as the temperature goes up, beer sales also go up (Figure 15.3). This is an example of a positive correlation. You also have noted a relationship between temperature and coffee sales:

On cold days you sell a lot of coffee, but coffee sales go down as the temperature goes up. This is an example of a negative relationship.

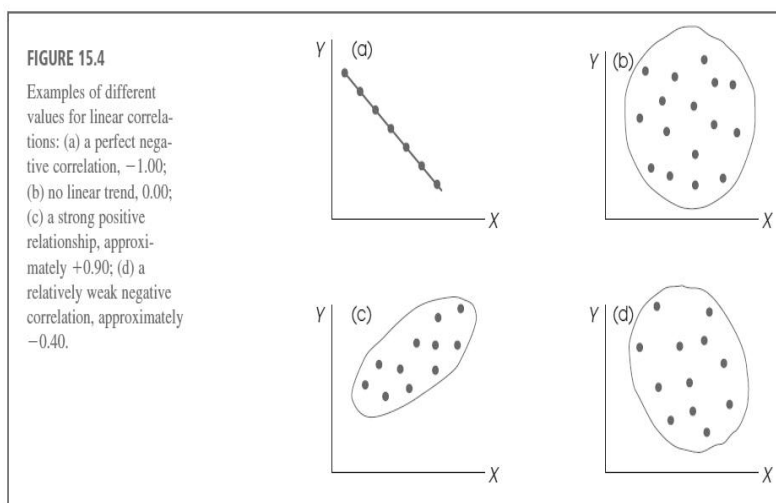


2. The Form of the Relationship. In the preceding coffee and beer examples, the relationships tend to have a **linear form**; that is, the points in the scatter plot tend to cluster around a straight line. We have drawn a line through the middle of the data points in each figure to help show the relationship. **The most common use of correlation is to measure straight-line relationships.** However, other forms of relationships do exist and there are special correlations used to measure them.

3. The Strength or Consistency of the Relationship. Finally, the correlation measures the consistency of the relationship. **For a linear relationship, for example, the data points could fit perfectly on a straight line. Every time X increases by one point, the value of Y also changes by a consistent and predictable amount.**

Figure 15.4(a) shows an example of a perfect linear relationship. However, relationships are usually not perfect. Although there may be a tendency for the value of Y to increase whenever X increases, the amount that Y changes is not always the same, and occasionally, Y decreases when X increases. In this situation, the data points do not fall perfectly on a straight line. **The consistency of the relationship is measured by the numerical value of the correlation. A perfect correlation** always is identified by a correlation of 1.00 and indicates a perfectly consistent relationship. **For a correlation of 1.00 (or -1.00), each change in X is accompanied by a perfectly predictable change in Y .** At the other extreme, a correlation of 0 indicates no consistency at all.

For a correlation of 0, the data points are scattered randomly with no clear trend [see Figure 15.4(b)]. Intermediate values between 0 and 1 indicate the degree of consistency.



- At the other extreme, a correlation of 0 indicates no consistency at all. For a correlation of 0, the data points are scattered randomly with no clear trend. The sign (+ or -) and the strength of a correlation are independent. A correlation of 1.00 indicates a **perfectly consistent relationship** whether it is positive (+1.00) or negative (-1.00). Similarly, correlations of +0.80 and -0.80 are **equally consistent relationships**. One of the most common errors in interpreting correlations is to assume that a **correlation necessarily implies a cause-and-effect relationship** between the two variables.

Correlation and Causation

Although there may be a causal relationship between two variables, the simple existence of a correlation does not prove it. To establish a cause-and-effect relationship, it is necessary to conduct a true experiment in which one variable is manipulated by a researcher and other variables are rigorously controlled. Although there may be a causal relationship between two variables, the simple existence of a correlation does not prove it. To establish a cause-and-effect relationship, it is necessary to conduct a true experiment in which one variable is manipulated by a researcher and other variables are rigorously controlled. A study shows a positive correlation between number of churches and number of crimes in towns and cities. It is reasonable that small towns would have less crime and fewer churches and that large cities would have large values for

both variables. Although a strong correlation exists between number of churches and crime, the real cause of the relationship is the size of the population.

Lesson 34

CORRELATION ANALYSIS-II

The Pearson Correlation

By far the most common correlation is the *Pearson correlation* (or the Pearson product– moment correlation) which measures the degree of straight-line relationship. The **Pearson correlation** measures the degree and the direction of the linear relationship between two variables. The

Pearson correlation is identified by the letter r .

Conceptually, this correlation is computed by:

$$r = \frac{\text{degree to which } X \text{ and } Y \text{ vary together}}{\text{degree to which } X \text{ and } Y \text{ vary separately}}$$

$$= \frac{\text{covariability of } X \text{ and } Y}{\text{variability of } X \text{ and } Y \text{ separately}}$$

When there is a perfect linear relationship, every change in the X variable is accompanied by a corresponding change in the Y variable. In this case, the covariability (X and Y together) is identical to the variability of X and Y separately, and the formula produces a correlation with a magnitude of +1.00 or –1.00. When there is no linear relationship, a change in the X variable does not correspond to any predictable change in the Y variable. In this case, there is no covariability, and the resulting correlation is zero.

1

EXAMPLE 15.2 The same set of $n = 4$ pairs of scores is used to calculate SP , first using the definitional formula and then using the computational formula.

For the definitional formula, you need deviation scores for each of the X values and each of the Y values. Note that the mean for the X s is $M_X = 3$ and the mean for the Y s is $M_Y = 5$. The deviations and the products of deviations are shown in the following table:

Caution: The signs (+ and –) are critical in determining the sum of products, SP .

Scores		Deviations		Products
X	Y	$X - M_X$	$Y - M_Y$	$(X - M_X)(Y - M_Y)$
1	3	–2	–2	+4
2	6	–1	+1	–1
4	4	+1	–1	–1
5	7	+2	+2	+4
				+6 = SP

For these scores, the sum of the products of the deviations is $SP=6$. For the computational formula, you need the X value, the Y value, and the XY product for each individual. Then you

find the sum of the X s, the sum of the Y s, and the sum of the XY products. These values are as follows:

X	Y	XY	
1	3	3	
2	6	12	
4	4	16	
<u>5</u>	<u>7</u>	<u>35</u>	
12	20	66	Totals

Substituting the totals in the formula gives

$$\begin{aligned}
 SP &= \sum XY - \frac{\sum X \sum Y}{n} \\
 &= 66 - \frac{12(20)}{4} \\
 &= 66 - 60 \\
 &= 6
 \end{aligned}$$

Both formulas produce the same result, $SP = 6$.

As noted earlier, the Pearson correlation consists of a ratio comparing the covariability of X and Y (the numerator) with the variability of X and Y separately (the denominator). In the formula for the Pearson r , we use SP to measure the covariability of X and Y . The variability of X is measured by computing SS for the X scores and the variability of Y is measured by SS for the Y scores. With these definitions, the formula for the Pearson correlation becomes

$$r = \frac{SP}{\sqrt{SS_X SS_Y}}$$

Lesson 35

CORRELATION ANALYSIS-III

Hypothesis Testing with the Pearson Correlation

The Pearson correlation is generally computed for sample data. A sample correlation is often used to answer questions about the corresponding population correlation. The basic question for this hypothesis test is whether a correlation exists in the population.

The null hypothesis is ‘No. There is no correlation in the population,’ or ‘The population correlation is zero.’

The alternative hypothesis is ‘Yes. There is a real, nonzero correlation in the population.’

The population correlation is traditionally represented by ρ (the Greek letter rho). Hypotheses would be stated in symbols as,

$$H_0: \rho = 0 \text{ (There is no population correlation)}$$

$$H_1: \rho \neq 0 \text{ (There is real correlation)}$$

When there is a specific prediction about the direction of the correlation, it is possible to do a directional, or one-tailed, test. For example,

$$H_0: \rho \leq 0 \text{ (The population correlation is not positive)}$$

$$H_1: \rho < 0 \text{ (The population correlation is positive)}$$

Samples are not expected to be identical to the populations from which they come; there is some discrepancy (sampling error) between a sample statistic and the corresponding population parameter.

Degrees of freedom

The hypothesis test for the Pearson correlation has degrees of freedom defined by $df = n - 2$. An intuitive explanation for this value is that a sample with only $n = 2$ data points has no degrees of freedom. Specifically, if there are only two points, they will fit perfectly on a straight line, and the sample produces a perfect correlation of $r = +1.00$ or $r = -1.00$. Because the first two points always produce a perfect correlation, the sample correlation is free to vary only when the data set contains more than two points. Thus, $df = n - 2$. The computations for evaluating r have already been completed and are summarized in **Table B.6 in Appendix B**. The table is based on the

concept that a sample is expected to be representative of the population from which it was obtained.

	Level of Significance for One-Tailed Test			
	.05	.025	.01	.005
<i>df = n - 2</i>	Level of Significance for Two-Tailed Test			
	.10	.05	.02	.01
1	.988	.997	.9995	.9999
2	.900	.950	.980	.990
3	.805	.878	.934	.959
4	.729	.811	.882	.917
5	.669	.754	.833	.874
6	.622	.707	.789	.834
7	.582	.666	.750	.798
8	.549	.632	.716	.765
9	.521	.602	.685	.735
10	.497	.576	.658	.708
11	.476	.553	.634	.684
12	.458	.532	.612	.661
13	.441	.514	.592	.641

Correlation in SPSS

The Pearson correlation (r) is designed for **interval level (continuous) variables**. It can also be used if you have **one continuous variable** (e.g. scores on a measure of self-esteem) and **one dichotomous variable** (e.g. gender: Male/Female)-that is the point-biserial correlation. **To run correlation in SPSS we need two variables: both continuous, or one continuous and the other dichotomous (two values).**

Assumptions

- The **scale of measurement** for the variables should be **interval or ratio** (continuous).
- One exception to this is if we have **one dichotomous independent variable** (with only two values e.g., gender) **and one continuous dependent variable**, we can still run correlation.
- **Each subject must provide a score on both variable X and variable Y (related pairs).**
- Both pieces of information must be from the **same subject**.
- The **observations that make up the data must be independent of one another**. That is, each observation or measurement must not be influenced by any other observation or measurement.

- The observations that make up the data must be independent of one another. That is, each observation or measurement must not be influenced by any other observation or measurement.
- Scores on each variable should be **normally distributed**
- The **relationship** between the two variables **should be linear**.

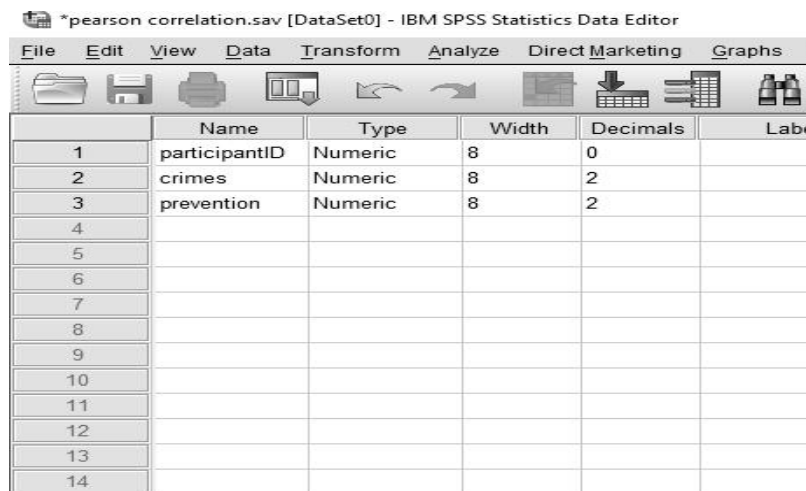
This means that when we look at a scatterplot of scores there should be see a straight line (roughly), not a curve.

For Example, a researcher records the annual number of serious crimes and the amount spent on crime prevention. The data are given in table. Compute the Pearson correlation to measure the degree of relationship between the two variables. Is the correlation statistically significant? Use a two-tailed test with alpha .01.

Number of crimes	Amount spent for prevention
3	6
4	7
6	3
7	4
8	11
9	12
11	8
12	9
13	16
14	17
16	13
17	14

Data Entry

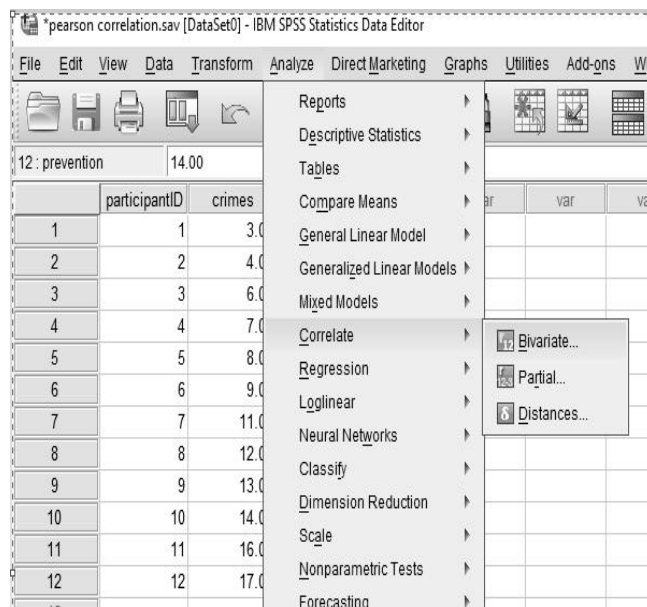
First of all, we need to create two variables on scale measurement namely crime (independent variable) and prevention (dependent variable) in variable view. The data are entered into two columns in the data editor, one for the X values (variable crimes) and one for the Y values (Variable prevention), with the two scores for each individual in the same row.



	Name	Type	Width	Decimals	Label
1	participantID	Numeric	8	0	
2	crimes	Numeric	8	2	
3	prevention	Numeric	8	2	
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					

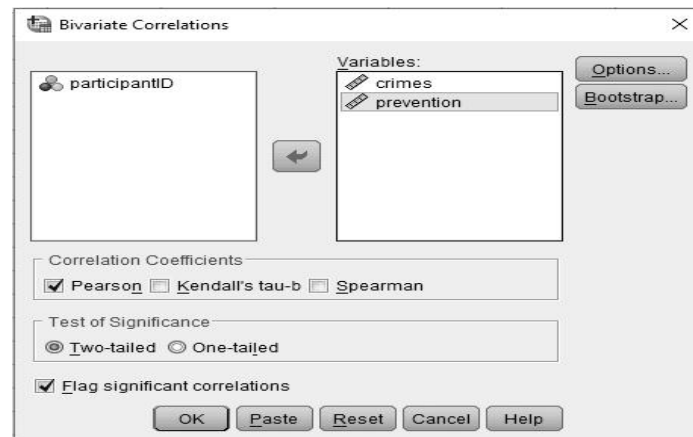
Running Analysis

- From the menu at the top of the screen, click on **Analyze**, then select **Correlate**, then **Bivariate**.

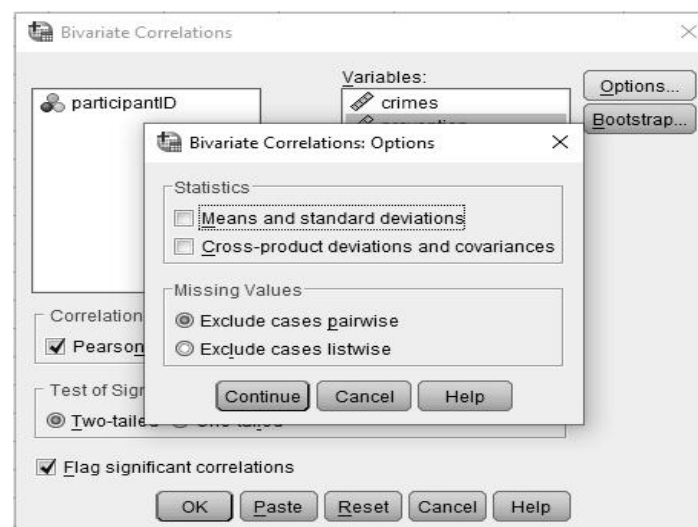


	participantID	crimes	prevention
1	1	3.0	14.00
2	2	4.0	
3	3	6.0	
4	4	7.0	
5	5	8.0	
6	6	9.0	
7	7	11.0	
8	8	12.0	
9	9	13.0	
10	10	14.0	
11	11	16.0	
12	12	17.0	

- Select two variables and move them into the box marked **Variables** (e.g., crimes and prevention).
- In the **Correlation Coefficients** section, the **Pearson** box is the default option.



- Click on the **Options** button. For **Missing Values**, click on the **Exclude cases pairwise** box.
- Under **Options**, we can also obtain means and standard deviations if we wish.
- Click on **Continue** and then on **OK**



Interpreting Results

SPSS provides a table giving the correlation coefficients between each pair of variables listed, the significance level and the number of cases. In the example given here, the Pearson

correlation coefficient ($r=.77$, $p<.01$) indicating a significant positive correlation between number of crimes and the amount spent on prevention. The more number of crimes people commit the more time is spent to prevent these crimes.

Correlations

		crimes	prevention
crimes	Pearson Correlation	1	.765**
	Sig. (2-tailed)		.004
	N	12	12
prevention	Pearson Correlation	.765**	1
	Sig. (2-tailed)	.004	
	N	12	12

** . Correlation is significant at the 0.01 level (2-tailed).

The size of the value of the correlation coefficient ranges from -1 to 1 . This value indicates the strength of the relationship between your two variables. Cohen (1988) suggests the following guidelines;

Small $r=.10$ to $.29$

Medium $r=.30$ to $.49$

Large $r=.50$ to 1.0

The relationship between number of crimes and amount spent on prevention was investigated using Pearson product-moment correlation coefficient. Preliminary analyses were performed to ensure no violation of the assumptions. There was a strong significant positive correlation between the two variables, $r = .77$, $n = 12$, $p < .01$, with more number of crimes associated with more time spent on prevention

Lesson 36

TYPES OF CORRELATION-I

Partial Correlation

Sometimes a researcher may suspect that the relationship between two variables is being distorted by the influence of a third variable. A statistical technique, known as *partial correlation*, allows a researcher to measure the relationship between two variables while eliminating the influence of a third variable.

For example, a study showed a positive correlation between number of churches and number of crimes in towns and cities. However, it was unlikely that there was a direct relationship between churches and crime. Instead, both variables were influenced by population. If population were controlled, there probably would be no real correlation between churches and crime.

In a situation with three variables, X , Y , and Z , it is possible to compute three individual Pearson correlations:

1. r_{XY} measuring the correlation between X and Y
2. r_{XZ} measuring the correlation between X and Z
3. r_{YZ} measuring the correlation between Y and Z

The statistical significance of a partial correlation is determined using the same procedure as is used to evaluate a regular Pearson correlation. For a partial correlation, however, we must use $df = n - 3$ instead of the $n - 2$ value that is used for the Pearson correlation. A significant correlation means that it is very unlikely ($p < \alpha$) that the sample correlation would occur without a corresponding relationship in the population.

The Spearman Correlation

When the Pearson correlation formula is used with **data from an ordinal scale (ranks)**, the result is called the *Spearman correlation*. The Spearman correlation is used in two situations. It is identified by the **symbol r_s** to differentiate it from the Pearson correlation. The Spearman correlation is used to measure the relationship between X and Y when both variables are measured on ordinal scales. Additionally, it can be used as a valuable alternative to the Pearson correlation, even when the original raw scores are on an interval or a ratio scale.

As we have noted, the **Pearson correlation measures the degree of *linear relationship*** between two variables—that is, how well the data points fit on a straight line. However, a researcher often

expects the data to show a consistently one-directional relationship but not necessarily a linear relationship. If the Pearson correlation were computed for these data, it would not produce a correlation of 1.00 because the data do not fit perfectly on a straight line. In a situation like this, the Spearman correlation can be used to measure the consistency of the relationship, independent of its form.

The Spearman correlation measures consistency, rather than form. When two variables are consistently related, their ranks are linearly related. For example, a perfectly consistent positive relationship means that every time the X variable increases, the Y variable also increases. When there is a consistently one-directional relationship between two variables, the relationship is said to be *monotonic*. Thus, the Spearman correlation measures the degree of monotonic relationship between two variables.

The Point-Biserial Correlation

The point-Biserial correlation is used to measure the relationship between two variables in situations in which one variable consists of regular, numerical scores, But the second variable has only two values. A variable with only two values is called a *dichotomous variable* or a *binomial variable*. For example, Male versus female, College graduate versus not a college graduate, Older than 30 years versus younger than 30 years.

To compute the point-biserial correlation, the dichotomous variable is first converted to numerical values by assigning a value of zero (0) to one category and a value of one (1) to the other category. Then the regular Pearson correlation formula is used with the converted data. Squaring the value of the point-biserial correlation produces r^2 which is exactly the value of r^2 we obtain measuring effect size of independent sample t-test. As with most correlations, the *strength* of the relationship is best described by the value of r^2 . r^2 is the coefficient of determination, which measures how much of the variability in one variable is predicted or determined by the association with the second variable.

The Phi-Coefficient

When both variables (X and Y) measured for each individual are dichotomous, the correlation between the two variables is called the *phi-coefficient*. It is represented by the symbol (ϕ). A two-step procedure is followed:

- Convert each of the dichotomous variables to numerical values by assigning a 0 to one category and a 1 to the other category for each of the variables.

- Use the regular Pearson formula with the converted scores.
- The phi-coefficient can be used to assess the relationship between two dichotomous variables.
- The more common statistical procedure is a chi-square statistic.

Lesson 37

INTRODUCTION TO REGRESSION

LINEAR EQUATIONS

In general, a *linear relationship* between two variables X and Y can be expressed by the Equation

$$Y = bX + a \quad (16.1)$$

where a and b are fixed constants.

For example, a local video store charges a membership fee of \$5 per month, which allows you to rent videos and games for \$2 each. With this information, the total cost for 1 month can be computed using a *linear equation* that describes the relationship between the total cost (Y) and the number of videos and games rented (X).

$$Y = 2X + 5$$

In the general linear equation, the value of b is called the *slope*. The slope determines how much the Y variable changes when X is increased by 1 point. For the video store example, the slope is $b = 2$ and indicates that your total cost increases by \$2 for each video you rent. The value of a in the general equation is called the *Y-intercept* because it determines the value of Y when $X = 0$. (On a graph, the a value identifies the point where the line intercepts the Y -axis.) For the video store example, $a = 5$; there is a \$5 membership charge even if you never rent a video.

Because a straight line can be extremely useful for describing a relationship between two variables, a statistical technique has been developed that provides a standardized method for determining the best-fitting straight line for any set of data. The statistical procedure is *regression*, and the resulting straight line is called the *regression line*.

The goal for regression is to find the best-fitting straight line for a set of data. To accomplish this goal, however, it is first necessary to define precisely what is meant by “best fit.” For any particular set of data, it is possible to draw lots of different straight lines that all appear to pass through the center of the data points. Each of these lines can be defined by a linear equation of the form

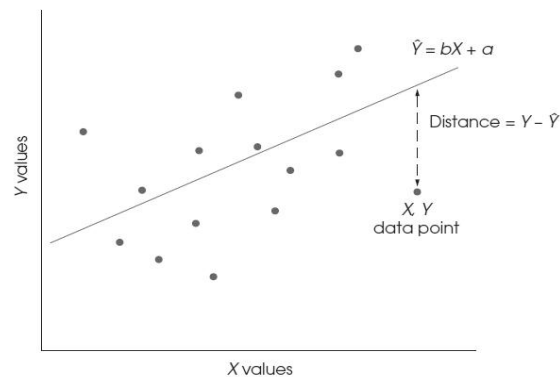
$$Y = bX + a$$

Where b and a are constants that determine the slope and Y -intercept of the line, respectively. Each individual line has its own unique values for b and a . The problem is to find the specific line that provides the best fit to the actual data points.

The Least-Squares Solution

To determine how well a line fits the data points, the first step is to define mathematically the distance between the line and each data point. For every X value in the data, the linear equation determines a Y value on the line. This value is the predicted Y and is called $\hat{Y}$ (“ Y hat”). The distance between this predicted value and the actual Y value in the data is determined by

$$\text{distance} = Y - \hat{Y}$$



Because some of these distances are positive and some are negative, the next step is to square each distance to obtain a uniformly positive measure of error. Finally, to determine the total error between the line and the data, we add the squared errors for all of the data points. The result is a measure of overall squared error between the line and the data:

$$\text{total squared error} = \sum (Y - \hat{Y})^2$$

Now we can define the *best-fitting* line as the one that has the smallest total squared error. For obvious reasons, the resulting line is commonly called the *least-squared-error solution*. In symbols, we are looking for a linear equation of the form

$$\hat{Y} = bX + a$$

For each value of X in the data, this equation determines the point on the line ($\hat{Y}$) that gives the best prediction of Y . The problem is to find the specific values for a and b that make this the best-fitting line.

The calculations that are needed to find this equation require calculus and some sophisticated algebra, so we do not present the details of the solution. The results, however, are relatively straightforward, and the solutions for b and a are as follows:

$$b = \frac{SP}{SS_X} \quad (16.2)$$

where SP is the sum of products and SS_X is the sum of squares for the X scores.

A commonly used alternative formula for the slope is based on the standard deviations for X and Y . The alternative formula is

$$b = r \frac{s_Y}{s_X} \quad (16.3)$$

where s_Y is the standard deviation for the Y scores, s_X is the standard deviation for the X scores, and r is the Pearson correlation for X and Y . The value of the constant a in the equation is determined by

$$a = M_Y - bM_X \quad (16.4)$$

Note that these formulas determine the linear equation that provides the best prediction of Y values. This equation is called the *regression equation for Y*. The **regression equation for Y is the linear equation**

$$\hat{Y} = bX + a$$

For example, the scores in the following table are used to demonstrate the calculation and use of the regression equation for predicting Y .

X	Y	$X - M_X$	$Y - M_Y$	$(X - M_X)^2$	$(Y - M_Y)^2$	$(X - M_X)(Y - M_Y)$
2	3	-2	-5	4	25	10
6	11	2	3	4	9	6
0	6	-4	-2	16	4	8
4	6	0	-2	0	4	0
7	12	3	4	9	16	12
5	7	1	-1	1	1	-1
5	10	1	2	1	4	2
3	9	-1	1	1	1	-1
				$SS_X = 36$	$SS_Y = 64$	$SP = 36$

For these data, $\Sigma X = 32$, so $M_X = 4$. Also, $\Sigma Y = 64$, so $M_Y = 8$. These values have been used to compute the deviation scores for each X and Y value. The final three columns show the squared deviations for X and for Y , and the products of the deviation scores.

Our goal is to find the values for b and a in the regression equation. Using Equations 16.2 and 16.4, the solutions for b and a are

$$b = \frac{SP}{SS_X} = \frac{36}{36} = 1.00$$

$$a = M_Y - bM_X = 8 - 1(4) = 4.00$$

The resulting equation is

$$\hat{Y} = X + 4$$

In general, there is some error between the predicted Y values and the actual data. Although the amount of error varies from point to point, on average the errors are directly related to the magnitude of the correlation. With a correlation near 1.00 (or 1.00), the data points generally are clustered close to the line and the error is small. As the correlation gets nearer to zero, the points move away from the line and the magnitude of the error increases.

The Standard Error of Estimate

The **standard error of estimate** gives a measure of the standard distance between the predicted Y values on the regression line and the actual Y values in the data. Conceptually, the standard error of estimate is very much like a standard deviation: Both provide a measure of standard distance. Also, the calculation of the standard error of estimate is very similar to the calculation of standard deviation. To calculate the standard error of estimate, we first find the sum of squared deviations (SS). Each deviation measures the distance between the actual Y value (from the data) and the predicted Y value (from the regression line). This sum of squares is commonly

called SS_{residual} because it is based on the remaining distance between the actual Y scores and the predicted values.

$$SS_{\text{residual}} = \sum (Y - \hat{Y})^2 \quad (16.8)$$

The obtained SS value is then divided by its degrees of freedom to obtain a measure of variance. This procedure should be very familiar:

$$\text{Variance} = \frac{SS}{df}$$

The degrees of freedom for the standard error of estimate are $df = n - 2$. The reason for having $n - 2$ degrees of freedom, rather than the customary $n - 1$, is that we now are measuring deviations from a line rather than deviations from a mean. To find the equation for the regression line, you must know the means for both the X and the Y scores. Specifying these two means places two restrictions on the variability of the data, with the result that the scores have only $n - 2$ degrees of freedom. (*Note:* the $df = n - 2$ for SS_{residual} is the same $df = n - 2$ that we encountered when testing the significance of the Pearson correlation on page 529.)

The final step in the calculation of the standard error of estimate is to take the square root of the variance to obtain a measure of standard distance. The final equation is

$$\text{standard error of estimate} = \sqrt{\frac{SS_{\text{residual}}}{df}} = \sqrt{\frac{\sum (Y - \hat{Y})^2}{n - 2}} \quad (16.9)$$

Assumptions of Regression

- **Data should be linear.**
- For any two observations the **residual terms should be uncorrelated** (i.e., **independent**). This eventuality is sometimes described as a lack of autocorrelation.
- **The assumption of autocorrelation can be tested with the Durbin–Watson test**, which tests for serial correlations between errors. Specifically, it tests whether adjacent residuals are correlated. The test statistic varies between 0 and 4, with a value of 2 meaning that the residuals are uncorrelated.
- At each level of the predictor variable(s), the variance of the **residual terms should be constant**. This assumption means that **the residuals at each level of the predictor(s) should have the same variance (homoscedasticity)**; **when the variances are very unequal there is said to be heteroscedasticity**.
- Normally distributed errors: It can be helpful if the **residuals in the model are random**, normally distributed variables with a mean of 0.

- This assumption means that the differences between the predicted and observed data are most frequently zero or very close to zero, and that differences much greater than zero happen only occasionally.
- All predictor variables must be quantitative (continuous) or categorical (with two categories).
- The outcome variable must be quantitative (continuous).
- No perfect multicollinearity: If the model has more than one predictor then there should be no perfect linear relationship between two or more of the predictors. So, the predictor variables should not correlate too highly.
- To identify multicollinearity we need to scan the correlation matrix for predictor variables that correlate very highly (values of r above 0.80 or 0.90).
- SPSS Statistics also compute the variance inflation factor (VIF), which indicates whether a predictor has a strong linear relationship with the other predictor(s).
- SPSS also gives the tolerance statistic, which is the reciprocal of VIF, $(1/VIF)$.

Some general guidelines for interpreting the VIF include:

1. If the largest VIF is greater than 10 (or the tolerance is below 0.1) then this indicates a serious problem.
2. If the average VIF is substantially greater than 1 then the regression may be biased
3. Tolerance below 0.2 indicates a potential problem

Lesson 38

SIMPLE LINEAR REGRESSION

Multiple regression is a family of techniques that can be used to explore the relationship between one continuous dependent variable and a number of independent variables or predictors. It allows a more sophisticated exploration of the interrelationship among a set of variables. Multiple regression can be used to address a variety of research questions. It can tell you how well a set of variables is able to predict a particular outcome. For example, you may be interested in exploring how well a set of subscales on an intelligence test is able to predict performance on a specific task. Multiple regression will provide you with information about the model as a whole (all subscales) and the relative contribution of each of the variables that make up the model (individual subscales). As an extension of this, multiple regression will allow you to test whether adding a variable (e.g. motivation) contributes to the predictive ability of the model, over and above those variables already included in the model. Multiple regression can also be used to statistically control for an additional variable (or variables) when exploring the predictive ability of the model. Some of the main types of research questions that multiple regression can be used to address are:

- how well a *set of variables* is able to predict a particular outcome
- *which variable* in a set of variables is the best predictor of an outcome
- Whether a particular predictor variable is still able to predict an outcome when the effects of another variable are *controlled for* (e.g. socially desirable responding).

Major Types of Multiple Regression

1. **Multiple linear regression** also called standard or simultaneous regression: All the independent (or predictor) variables are entered into the model simultaneously. Each independent variable is evaluated in terms of its predictive power, over and above that offered by all the other independent variables.
2. **Hierarchical multiple regression** (also called sequential regression): Variables or sets of variables are entered in steps (or blocks), with each independent variable being assessed in terms of what it adds to the prediction of the dependent variable after the previous variables have been controlled for.

3. **Stepwise multiple regression:** The researcher provides a list of independent variables and then allows the program to select which variables it will enter and in which order they go into the equation, based on a set of statistical criteria. There are three different versions of this approach: **forward selection**, **backward deletion** and **stepwise regression**.

The **forward selection method** is often used to provide an initial screening of the variables when a large group of variables exists. SPSS selects the variable that has the highest R-Squared. At each step, it selects the predictor variable that increases R-Squared the most. Finally, it stops adding variables when none of the remaining variables are significant.

The **backward deletion procedure** begins with a model that includes all the independent variables. It deletes one variable at a time by determining whether the least significant variable currently in the model can be removed because its p-value is less than the user-specified or default value.

Stepwise regression is a modification of the forward selection so that after each step in which a variable was added, all variables in the model are checked to see if their significance has been reduced below the specified tolerance level. If a non-significant variable is found, it is removed from the model.

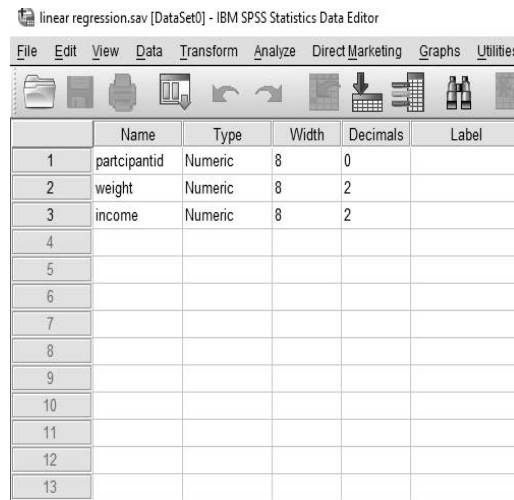
Simple Linear Regression Through SPSS-I

Linear regression is the next step up after correlation. It is used when we want to predict the value of one variable based on the value of another variable. The variable we want to predict is called the dependent variable (or sometimes, the outcome variable). The variable we are using to predict the other variable's value is called the independent variable (or sometimes, the predictor variable). For simple linear regression dependent variable should be measured at the continuous level (i.e., it is either interval or ratio variable). The independent variable could be categorical (having only two categories) or continuous. If the dependent variable is categorical, that is having two levels, the appropriate procedure would be a logistic regression (non-parametric test) instead of linear regression.

For example, for the following set of data for income (DV) and weight (IV). To simplify the weight variable, the women are classified into five categories that measure actual weight relative to height, from 1=thinnest to 5=heaviest. find the linear regression. Does the regression account for a significant portion of the variance? Use alpha level .05 to evaluate the F -ratio.

Data Entry

We need to create two variables namely weight (IV) and income (DV), in the SPSS data editor. For predictor variable (weight), enter the values in one column and for outcome variable (income), enter the values in a second column of the SPSS data editor as shown.

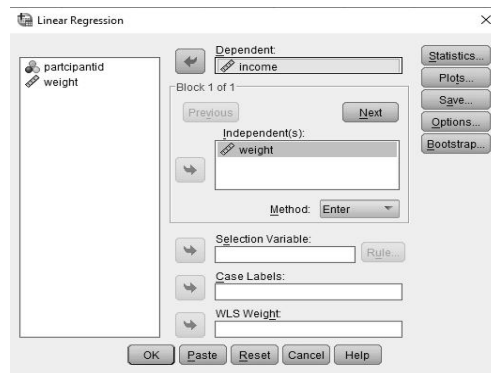


For predictor variable (weight), enter the values in one column and for outcome variable (income), enter the values in a second column of the SPSS data editor as shown.

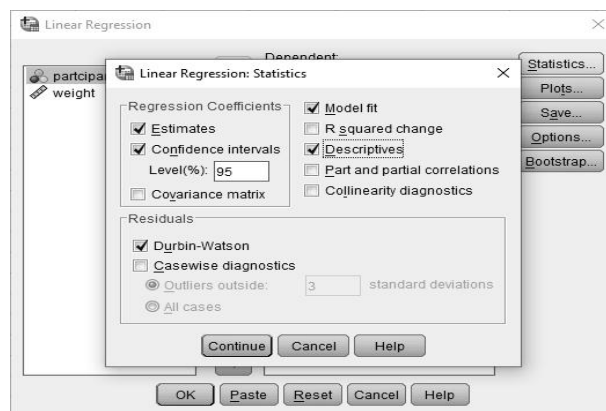
	participantid	weight	income	var
1	1	1.00	125.00	
2	2	2.00	78.00	
3	3	4.00	49.00	
4	4	3.00	63.00	
5	5	5.00	35.00	
6	6	2.00	84.00	
7	7	5.00	38.00	
8	8	3.00	51.00	
9	9	1.00	93.00	
10	10	4.00	44.00	
11				
12				
13				
14				
15				
16				
17				

Running Analysis

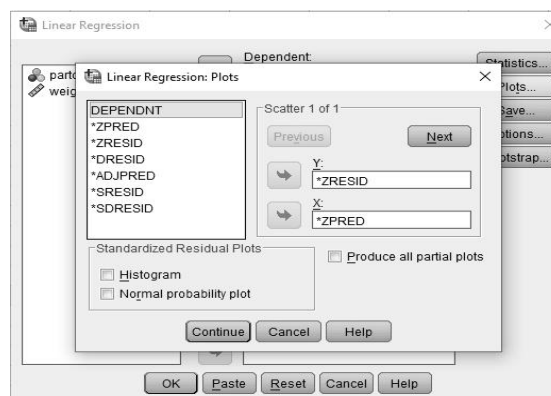
- Click Analyze on the tool bar, select Regression, and click on Linear.
- In the left-hand box, highlight the column label for the outcome variable, then click the arrow to move the column label into the Dependent Variable box.
- For the predictor variable, highlight the column label for the weight and click the arrow to move it into the Independent Variable(s) box.



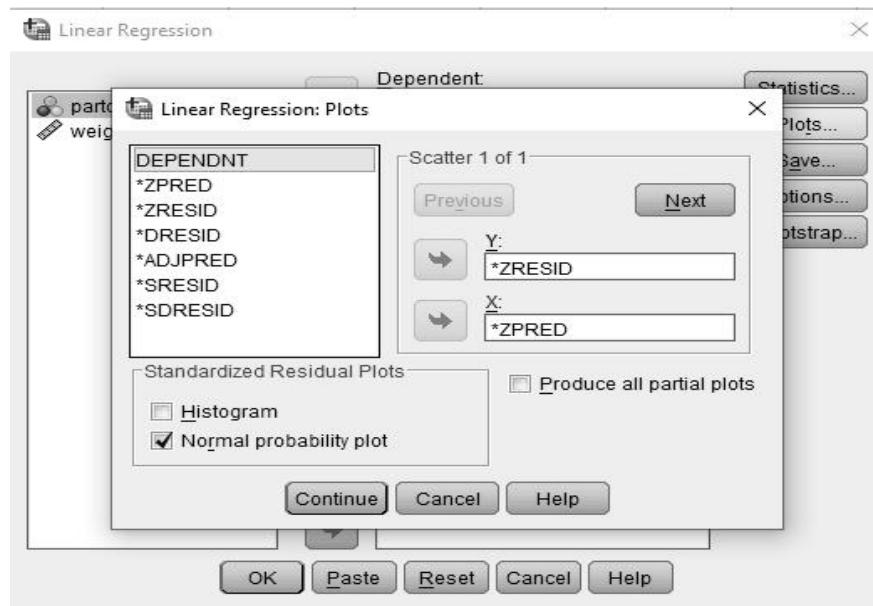
- From the Statistics section, check the Estimates, Confidence intervals, Descriptives and model fit boxes.
- To assess the assumption of independence of observations (no autocorrelation) check the Durbin-Watson box.
- Click Continue.



- In Plots section, move the ZRESID (standardized residuals) and ZPRED (standardized predicted values) in Y and X boxes respectively.
- The scatter plot will assess the homoscedasticity in the data.

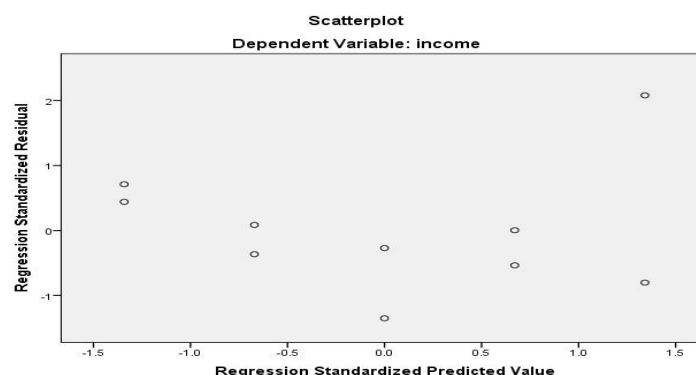


- In the section headed Standardized Residual Plots, tick the Normal probability plot option to assess the normality of the data.
- Click on Continue and OK.

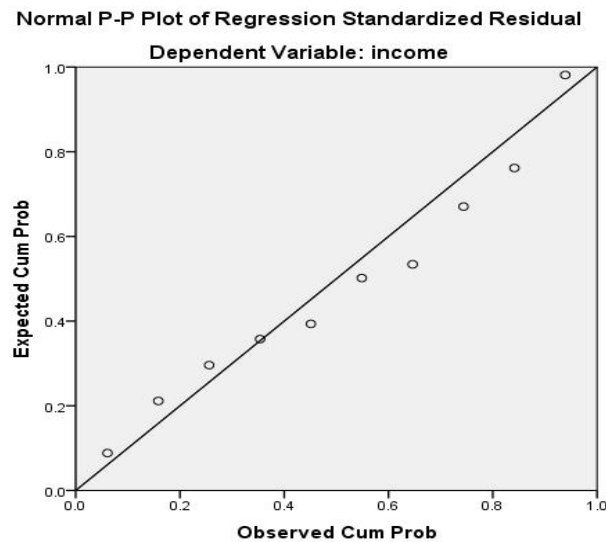


Interpreting Results

In the Scatterplot of the standardized residuals we are hoping that the residuals will be roughly rectangularly distributed, with most of the scores concentrated in the center (along the 0 point). There should not be curvilinear pattern in the data. Deviations from a centralized rectangle suggest some violation of the assumptions.



In the Normal P-P Plot, we check whether the data points lie in a reasonably straight diagonal line from bottom left to top right. This would suggest no major deviations from normality.



Now we need to look at the Model Summary box. This table provides R and R square values. The R value represents the correlation and is ($R=.931$) which indicates a high correlation. The R square value tells us how much of the variance in the DV (income) is explained by the IV (weight).

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.931 ^a	.867	.851	11.10152	1.663

a. Predictors: (Constant), weight

b. Dependent Variable: income

The table also gives the Durbin-Watson value which is 1.6 and is under acceptable range. This indicates that there is no autocorrelation and the assumption is satisfied.

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.931 ^a	.867	.851	11.10152	1.663

a. Predictors: (Constant), weight

b. Dependent Variable: income

The next table is the ANOVA table, which reports how well the regression equation fits the data (i.e., predicts the dependent variable). This tests the null hypothesis that R in the population equals 0. The model in this example reaches statistical significance ($F=52.29$, $p<.001$).

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6444.050	1	6444.050	52.287	.000 ^b
	Residual	985.950	8	123.244		
	Total	7430.000	9			

a. Dependent Variable: income

b. Predictors: (Constant), weight

The Coefficients table provides the necessary information to predict income from weight. The weight is significantly negatively predicting income ($B=-17.95$, $p<.001$). This means that for every 1-unit increase in the weight, the income will decrease by 17.95 units.

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	119.850	8.233		14.557	.000
	weight	-17.950	2.482	-.931	-7.231	.000

a. Dependent Variable: income

Simple linear regression was run to assess whether weight predicts income in women. Preliminary analyses were conducted to ensure no violation of the assumptions of normality, linearity, and homoscedasticity. The total variance explained by the model as a whole was 86.7%, $F(1, 8) = 52.28$, $p < .001$. Weight was found to be significant negative predictor of income ($B=-17.95$, $p<.001$).

Lesson 39

MULTIPLE LINEAR REGRESSION

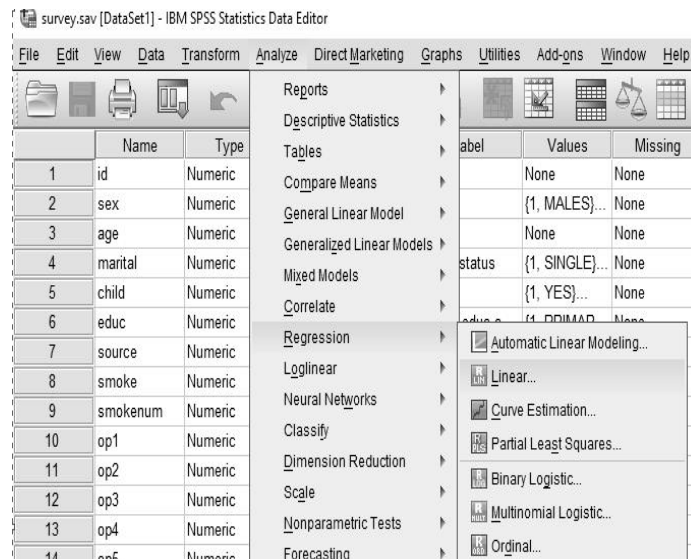
Multiple linear regression tells us how much of the variance in dependent variable can be explained by the independent variables. Tests allow us to determine the statistical significance of the results, in terms of both the model itself and the individual independent variables. Let's take an example of the impact of respondents' perceptions of control on their levels of perceived stress. In this example, we are interested in exploring how well the Mastery Scale and the PCOISS are able to predict scores on a measure of perceived stress. In the questionnaire, the Mastery Scale, measured the degree to which people feel they have control over the events in their lives; and the Perceived Control of Internal States Scale (PCOISS) measured the degree to which people feel they have control over their internal states (their emotions, thoughts and physical reactions).

Multiple Linear Regression in SPSS

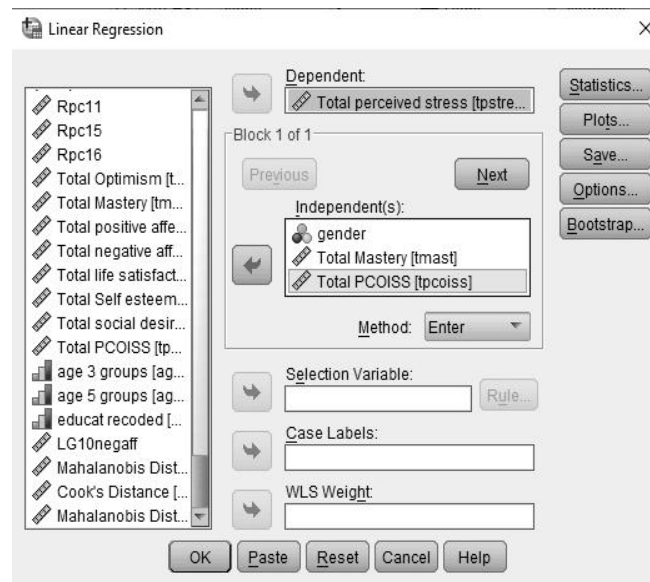
- To run the analysis we need,
- One continuous dependent variable (Total perceived stress)
- Two or more continuous independent variables (mastery, PCOISS)
- We can also use dichotomous independent variables having two categories, (e.g., gender: males=1, females=2).

Running Analysis

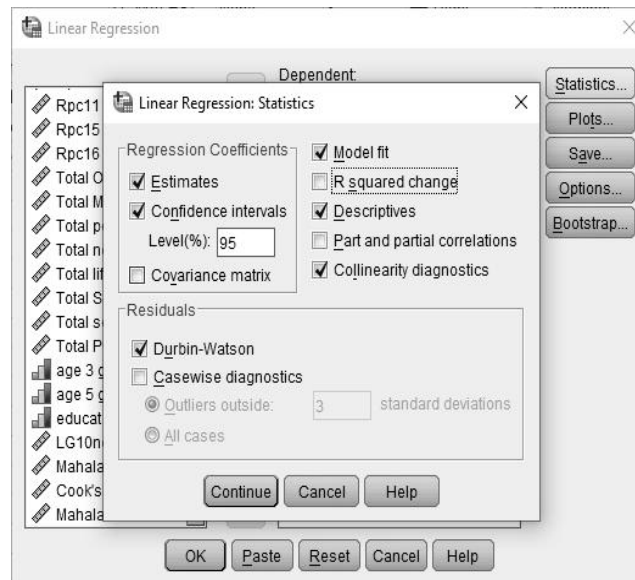
- From the menu at the top of the screen, click on **Analyze**, then select **Regression**, then **Linear**.
- Click on the continuous dependent variable (Total perceived stress: tpstress) and move it into the **Dependent** box.
- Click on the independent variables (Total Mastery: tmast; Total PCOISS: tpcoiss; Gender) and click on the arrow to move them into the **Independent(s)** box.



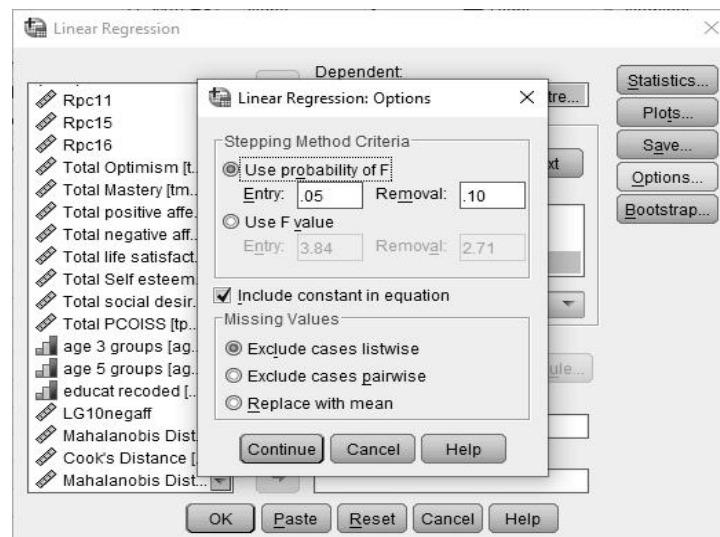
- Click on the continuous dependent variable (Total perceived stress: tpstress) and move it into the **Dependent** box.
- Click on the independent variables (Total Mastery: tmast; Total PCOISS: tpcoiss; Gender) and click on the arrow to move them into the **Independent(s)** box.



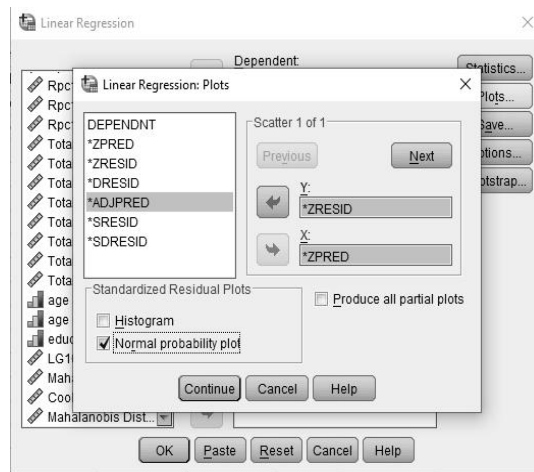
- Click on the **Statistics** button.
- Select the following: **Estimates, Confidence Intervals, Model fit, Descriptives, Collinearity diagnostics and Durbin-Watson.**



- Click on the **Options** button. In the **Missing Values** section, select **Exclude cases list wise**.
- Click on **Continue**.



- Click on the **Plots** button.
- Click on ***ZRESID** and the arrow button to move this into the **Y** box.
- Click on ***ZPRED** and the arrow button to move this into the **X** box.
- In the section headed **Standardized Residual Plots**, tick the **Normal probability plot** option.
- Click on **Continue** and **OK**.



Assumptions of Multiple Linear Regression

1. Multicollinearity

- The correlations between the variables in the model are provided in the table labelled **Correlations**.
- Check that if independent variables show at least some relationship with the dependent variable (above .3 preferably).

Correlations

		Total perceived stress	gender	Total Mastery	Total PCOISS
Pearson Correlation	Total perceived stress	1.000	.147	-.611	-.581
	gender	.147	1.000	-.115	-.102
	Total Mastery	-.611	-.115	1.000	.527
	Total PCOISS	-.581	-.102	.527	1.000
Sig. (1-tailed)	Total perceived stress	.	.001	.000	.000
	gender	.001	.	.009	.018
	Total Mastery	.000	.009	.	.000
	Total PCOISS	.000	.018	.000	.
N	Total perceived stress	426	426	426	426
	gender	426	426	426	426
	Total Mastery	426	426	426	426
	Total PCOISS	426	426	426	426

- The other value given is the **VIF** (Variance inflation factor), which is just the inverse of the Tolerance value (1 divided by Tolerance).
- VIF values above 10 would be a concern here, indicating multicollinearity.

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	49.387	1.510		32.704	.000	46.419	52.356		
	gender	.740	.422	.063	1.755	.080	-.089	1.568	.985	1.016
	Total Mastery	-.613	.061	-.417	-9.973	.000	-.734	-.492	.718	1.392
	Total PCOISS	-.173	.020	-.354	-8.484	.000	-.213	-.133	.720	1.388

a. Dependent Variable: Total perceived stress

2. Independence of Observations

- The **Model summary** table gives the Durbin-Watson value which is 1.91 for this example.
- This indicates that there is no autocorrelation and the assumption is satisfied.

Model Summary^b

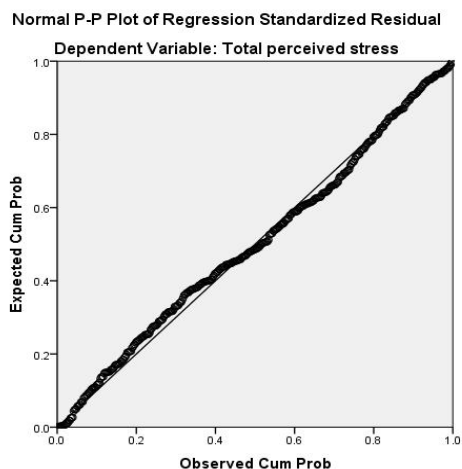
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.685 ^a	.470	.466	4.267	1.906

a. Predictors: (Constant), Total PCOISS, gender, Total Mastery

b. Dependent Variable: Total perceived stress

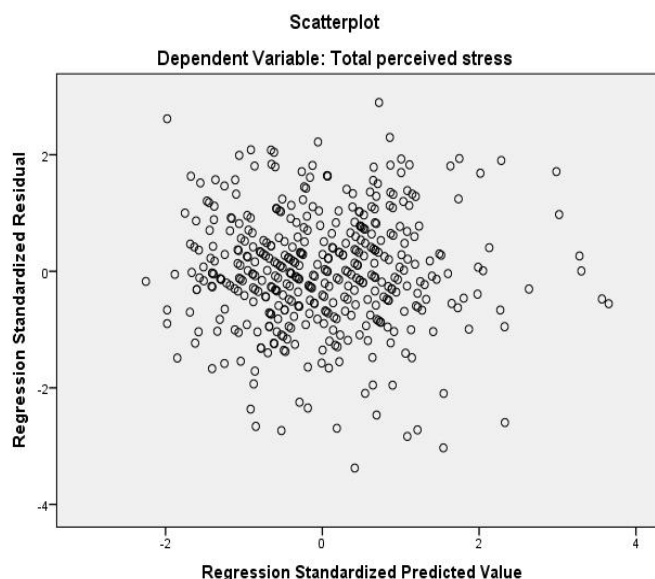
3. Normality

- In the **Normal P-P Plot**, the data points lie along the line from bottom left to top right.
- In the current example, the data suggest no deviations from normality, hence, the assumption is fulfilled.



4. Linearity and Homoscedasticity

- The **Scatterplot** indicates that the residuals are roughly distributed, with most of the scores concentrated in the center.
- the pattern of residuals in the scatter plot suggests that the assumptions of homoscedasticity and linearity are satisfied.



Interpreting the Results

- First of all, the **Model Summary** box gives the **R Square**.
- This tells you how much of the variance in the dependent variable (perceived stress) is explained by the model (which includes the variables of gender, Total Mastery and Total PCOISS).
- In this case, the value is .468.

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.685 ^a	.470	.466	4.267	1.906

a. Predictors: (Constant), Total PCOISS, gender, Total Mastery

b. Dependent Variable: Total perceived stress

- To assess the statistical significance of the result, it is necessary to look in the table labelled **ANOVA**.

- The model in this example gives the statistical significance (Sig. = .000 or $p < .001$).

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6807.615	3	2269.205	124.623	.000 ^b
	Residual	7684.009	422	18.209		
	Total	14491.624	425			

a. Dependent Variable: Total perceived stress

b. Predictors: (Constant), Total PCOISS, gender, Total Mastery

- The next thing we want to know is which of the variables included in the model contributed to the prediction of the dependent variable.
- We find this information in the output box labelled Coefficients.**
- Look in the column labelled Beta under Standardized Coefficients**

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	49.387	1.510		32.704	.000	46.419	52.356		
	gender	.740	.422	.063	1.755	.080	-.089	1.568	.985	1.016
	Total Mastery	-.613	.061	-.417	-9.973	.000	-.734	-.492	.718	1.392
	Total PCOISS	-.173	.020	-.354	-8.484	.000	-.213	-.133	.720	1.388

a. Dependent Variable: Total perceived stress

- In this case the largest beta coefficient is $-.42$ (ignoring the sign), which is for Total Mastery. This means that this variable makes the strongest unique contribution to explaining the dependent variable, when the variance explained by all other variables in the model is controlled for.
- In this case the largest beta coefficient is $-.42$ (ignoring the sign), which is for Total Mastery. This means that this variable makes the strongest unique contribution to explaining the dependent variable, when the variance explained by all other variables in the model is controlled for.

Lesson 40

NON-PARAMETRIC TEST

Often, researchers are confronted with experimental situations that do not conform to the requirements of parametric tests. When the assumptions of a test are violated, the test may lead to an erroneous interpretation of the data. In these situations, it may not be appropriate to use a parametric test. There are several hypothesis-testing techniques that provide alternatives to parametric tests. These alternatives are called nonparametric tests. Nonparametric tests sometimes are also called *assumption-free tests*. The main reasons to apply the nonparametric test include:

- The underlying data do not meet the assumptions about the population sample
- The population sample size is too small
- The analyzed data is ordinal or nominal
- Nonparametric statistics do not assume that data is drawn from a normal distribution, therefore, called distribution free tests.
- Nonparametric statistics includes nonparametric descriptive statistics, statistical models, inference, and statistical tests.
- The model structure of nonparametric models is not specified *a priori* but is instead determined from data.
- Nonparametric test usually involves qualitative data or Nominal / Ordinal data
- Most of the time even quantitative data is converted to nominal/ordinal data for use with non-parametric tests
- For example, numerical scores measuring self-esteem can be converted to categories like high, medium, and low self-esteem etc.

Chi-Square Test

The name of the test comes from the Greek letter (chi, pronounced “kye”), which is used to identify the test statistic. The most obvious difference between the chi-square tests and the other hypothesis tests we have considered (t and ANOVA) is the nature of the data. For chi-square, the data are in frequencies rather than numerical scores. The chi-square test can be defined as the frequency data from a sample to evaluate the relationship between two variables in the population. Each individual in the sample is classified on both of the two variables, creating a

two-dimensional frequency-distribution matrix. The frequency distribution for the sample is then used to test hypotheses about the corresponding frequency distribution for the population. In this situation, each individual in the sample is measured or classified on two separate variables. In general terms, this **chi-square test uses the proportions obtained for sample data to test hypotheses about the corresponding proportions in the population**

Assumptions of Chi-Square Test

To use a chi-square test for goodness of fit or a test of independence, several conditions must be satisfied. For any statistical test, violation of assumptions and restrictions casts doubt on the results. Some important assumptions and restrictions for using chi-square tests are the following:

1. Independence of Observations

This is not to be confused with the concept of independence between variables, as seen in the chi-square test for independence. One consequence of independent observations is that each observed frequency is generated by a different individual. A chi-square test would be inappropriate if a person could produce responses that can be classified in more than one category or contribute more than one frequency count to a single category.

2. Size of Expected Frequencies

A chi-square test should not be performed when the expected frequency of any cell is less than 5. The chi-square statistic can be distorted when f_e is very small. Consider the chi-square computations for a single cell. Suppose that the cell has values of f_o 1 and f_o 5. Note that there is a 4-point difference between the observed and expected frequencies.

Chi-square statistic involves two tests:

- **The chi-square test for goodness of fit**
- **The chi-square test for independence**

Lesson 41

CHI-SQUARE TEST FOR INDEPENDENCE-I

The chi-square statistic may also be used to test whether there is a relationship between two variables. In this situation, each individual in the sample is measured or classified on two separate variables. For example, a group of students could be classified in terms of personality (introvert, extrovert) and in terms of color preference (red, yellow, green, or blue). Usually, the data from this classification are presented in the form of a matrix, where the rows correspond to the categories of one variable and the columns correspond to the categories of the second variable. For example, in table below a hypothetical data for a sample of $n = 200$ students who have been classified by personality and color preference. The number in each box, or cell, of the matrix indicates the frequency, or number of individuals in that particular group. In Table, for example, there are 10 students who were classified as introverted and who selected red as their preferred color. To obtain these data, the researcher first selects a random sample of $n = 200$ students. Each student is then given a personality test and is asked to select a preferred color from among the four choices.

Note that the classification is based on the measurements for each student; the researcher does not assign students to categories. Also, note that the data consist of frequencies, not scores, from a sample. The goal is to use the frequencies from the sample to test a hypothesis about the population frequency distribution. Specifically, are these data sufficient to conclude that there is a significant relationship between personality and color preference in the population of students?

	Red	Yellow	Green	Blue	
Introvert	10	3	15	22	50
Extrovert	90	17	25	18	150
	100	20	40	40	$n = 200$

The null hypothesis for the chi-square test for independence states that the two variables being measured are independent; that is, for each individual, the value obtained for one variable is not related to (or influenced by) the value for the second variable. This general hypothesis can be expressed in two different conceptual forms, each viewing the data and the test from slightly

different perspectives. The null hypothesis for the chi-square test for independence states that the two variables being measured are independent.

H_0 : For the general population of students, there is no relationship between color preference and personality.

The alternative hypothesis, H_1 , states that there is a relationship between the two variables.

Calculating Chi-square test for independence

The chi-square test for independence uses the same basic logic that was used for the goodness-of-fit test. First, a sample is selected and each individual is classified or categorized. Because the test for independence considers two variables, every individual is classified on both variables, and the resulting frequency distribution is presented as a two-dimensional matrix. As before, the frequencies in the sample distribution are called observed frequencies and are identified by the symbol f_o .

The next step is to find the expected frequencies, or f_e values, for this chi-square test. As before, the expected frequencies define an ideal hypothetical distribution that is in perfect agreement with the null hypothesis. Once the expected frequencies are obtained, we compute a chi-square statistic to determine how well the data (observed frequencies) fit the null hypothesis (expected frequencies). Although you can use either version of the null hypothesis to find the expected frequencies, the logic of the process is much easier when you use H_0 stated in terms of equal proportions. To find the expected frequencies, we first determine the overall distribution of color preferences and then apply this distribution to both categories of personality. For these data, 100 people selected red as their preferred color. Because the total sample consists of 200 people, the proportion selecting red is 100 out of 200, or 50%. The complete set of color preference proportions is as follows: 100 out of 200 50% preferred 20 out of 200 10% prefer yellow 40 out of 200 20% prefer green 40 out of 200 20% prefer blue. The row totals in the matrix define the two samples of personality types. For example, the matrix in Table shows a total of 50 introverts (the top row) and a sample of 150 extroverts (the bottom row). According to the null hypothesis, both personality groups should have the same proportions for color preferences.

Beginning with the sample of 50 introverts in the top row, we obtain expected frequencies of:

50% preferred: f_e 50% of 50 $0.50(50)$ 25

10% prefer yellow: f_e 10% of 50 $0.10(50)$ 5

20% prefer green: f_e 20% of 50 $0.20(50)$ 10

20% prefer blue: f_e 20% of 50 $0.20(50)$ 10

Using exactly the same proportions for the sample of n 150 extroverts in the bottom row, we obtain expected frequencies of:

50% preferred: f_e 50% of 150 $0.50(50)$ 75

10% prefer yellow: f_e 10% of 150 $0.10(50)$

15 20% prefer green: f_e 20% of 150 $0.20(50)$ 30

20% prefer blue: f_e 20% of 150 $0.20(50)$ 30

A simple formula for determining expected frequencies

Although expected frequencies are derived directly from the null hypothesis and the sample characteristics, it is not necessary to go through extensive calculations to find f_e values. In fact, there is a simple formula that determines f_e for any cell in the frequency distribution matrix:

$$f_e = fc f_r / n$$

where fc is the frequency total for the column (column total), f_r is the frequency total for the row (row total), and n is the number of individuals in the entire sample. The chi-square test for independence uses exactly the same chi-square formula as the test for goodness of fit:

$$\chi^2 = \sum (f_o - f_e)^2 / f_e$$

Thus, the total number of f_e values that you can freely choose is $(R - 1)(C - 1)$, and the degrees of freedom for the chi-square test of independence are given by the formula $df (R - 1)(C - 1)$ (17.5) Also note that once you calculate the expected frequencies to fill the smaller matrix, the rest of the f_e values can be found by subtraction.

Chi-square test for independence in SPSS

Entering Data

- In the SPSS data editor, create three variables, one for frequencies and two categorical variables (e.g., personality and color).

chi-square for independence.sav [DataSet0] - IBM SPSS Statistics Data Editor

	Name	Type	Width	Decimals	Label
1	personality	Numeric	8	2	
2	color	Numeric	8	2	
3	frequencies	Numeric	8	2	
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					

- In Value Labels section, assign codes to the levels of the variables.

Value Labels

Value Labels

Value:

Label:

Spelling...

Add

Change

Remove

1.00 = "introvert"
2.00 = "extrovert"

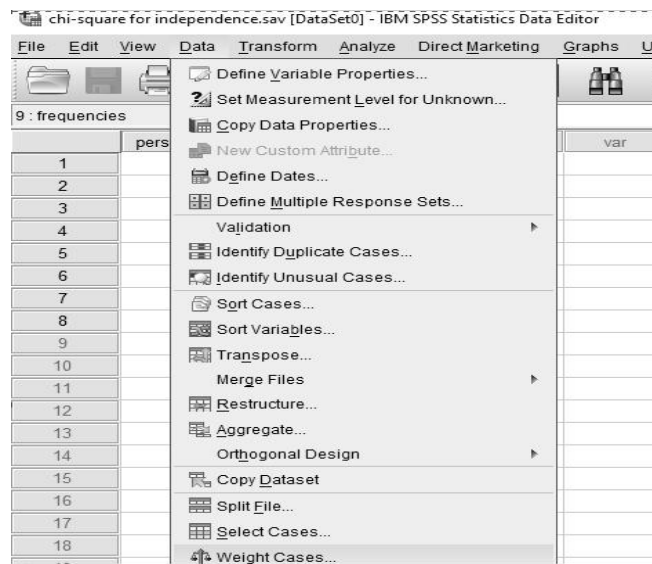
OK Cancel Help

- In a first column, enter a code (e.g., 1 and 2) that identifies the row corresponding to each observed frequency.
- In the second column, enter a code (e.g., 1, 2, 3 and 4) that identifies the column corresponding to each observed frequency.
- Enter the complete set of observed frequencies in a column named frequencies.

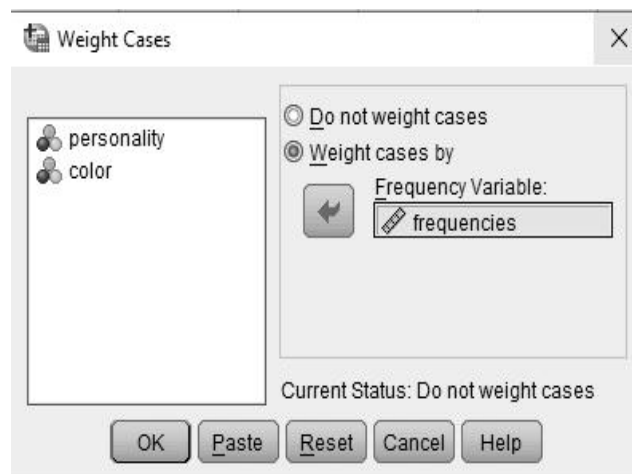
chi-square for independence.sav [DataSet0] - IBM SPSS Statistics Data Editor

	personality	color	frequencies	var
1	1.00	1.00	10.00	
2	1.00	2.00	3.00	
3	1.00	3.00	15.00	
4	1.00	4.00	22.00	
5	2.00	1.00	90.00	
6	2.00	2.00	17.00	
7	2.00	3.00	25.00	
8	2.00	4.00	18.00	
9				
10				
11				
12				
13				
14				
15				

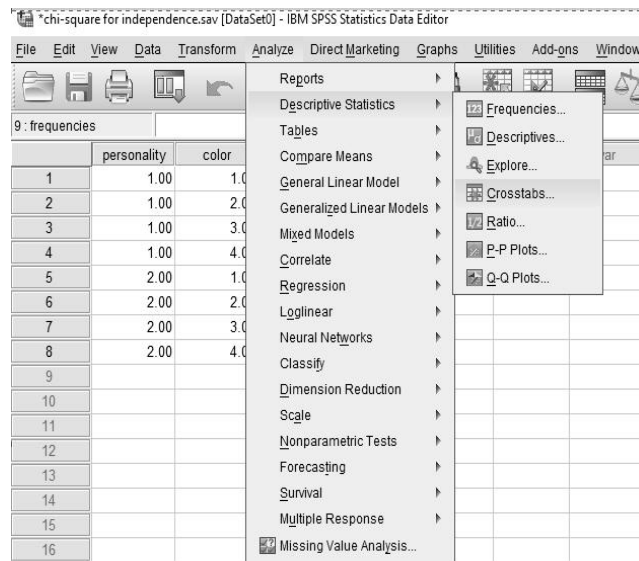
- We need to Weight our cases to tell the SPSS Statistics that we have summated the categories.
- Click Data in the top menu and go to Weight Cases.



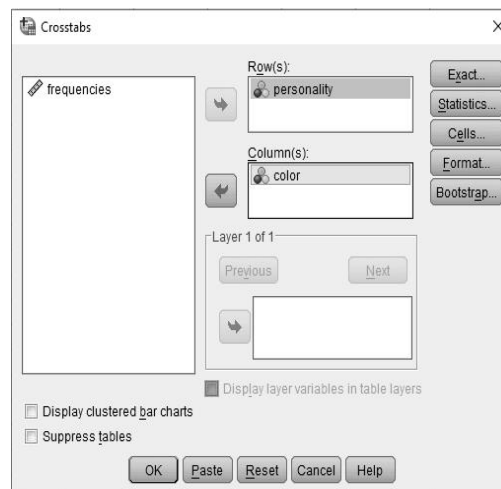
- Click on the circle Weight cases by and drag the frequencies and variable into the Frequency Variable box.
- Click Ok.



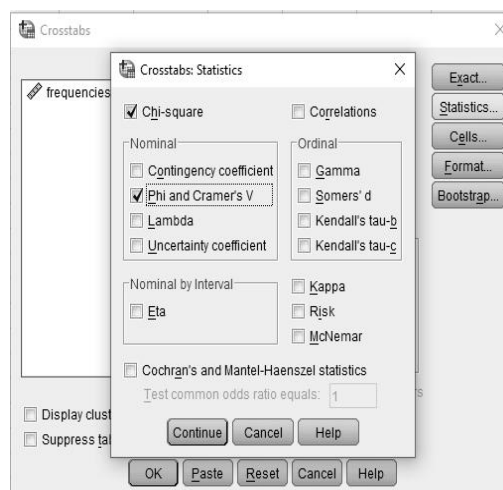
- Click Analyze on the tool bar at the top of the page, select Descriptive Statistics, and click on Crosstabs.



- Highlight the label for the column containing the rows (personality) and move it into the Rows box by clicking on the arrow.
- Highlight the label for the column containing the columns (color) and move it into the Columns box by clicking on the arrow.



- Click on the Statistics button. Tick Chi-square and Phi and Cramer's V. Click on Continue.
- Click OK.



- The first thing we should check is whether we have violated one of the assumptions of chi-square concerning the 'minimum expected cell frequency', which should be 5 or greater. The main value that we are interested in is the Pearson Chi-Square value, which is presented in the Chi-Square Tests.
- In the example the chi-square value is $\chi^2(3)=35.6$, with an associated significance level of $p<.001$, presented in the column labelled Asymptotic Significance (2-sided).
- This indicates that our result is significant and there appears to be an association between personality type and color preference.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	35.600 ^a	3	.000
Likelihood Ratio	35.033	3	.000
Linear-by-Linear Association	34.476	1	.000
N of Valid Cases	200		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 5.00.

- For 2 by 2 tables the most commonly used effect size is the phi coefficient, which is a correlation coefficient and can range from 0 to 1, with higher values indicating a stronger association between the two variables.
- For tables larger than 2 by 2 the value to report is Cramer's V, which takes into account the degrees of freedom.

- In this example the Cramer's V value (shown in the table Symmetric Measures) is .422 indicating a large effect size.
- For tables larger than 2 by 2 the value to report is Cramer's V, which takes into account the degrees of freedom.
- In this example the Cramer's V value (shown in the table Symmetric Measures) is .422 indicating a large effect size.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.422	.000
	Cramer's V	.422	.000
N of Valid Cases		200	

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Lesson 42

CHI-SQUARE TEST FOR INDEPENDENCE-II

The Pearson correlation measures the degree of linear relationship between two variables when the data (X and Y values) consist of numerical scores from an interval or ratio scale of measurement. However, other correlations have been developed for nonlinear relationships and for other types of data. When the Pearson correlation formula is used with data from an ordinal scale (ranks), the result is called the Spearman correlation.

Pearson's r and Spearman's ρ are very similar. They are both correlation coefficients, interpreted in the same way. Pearson's r is used when our data meet the assumptions for a parametric test. Spearman's ρ (ρ) is used when the data do not conform to these assumptions.

To summarize, the Spearman correlation measures the relationship between two variables when both are measured on ordinal scales (ranks). There are two general situations in which the Spearman correlation is used:

1. Spearman is used when the original data are ordinal; that is, when the X and Y values are ranks. In this case, you simply apply the Pearson correlation formula to the set of ranks.
2. Spearman is used when a researcher wants to measure the consistency of a relationship between X and Y, independent of the specific form of the relationship.

In this case, the original scores are first converted to ranks, then the Pearson correlation formula is used with the ranks. Because the Pearson formula measures the degree to which the ranks fit on a straight line, it also measures the degree of consistency in the relationship for the original scores. Incidentally, when there is a consistently one-directional relationship between two variables, the relationship is said to be monotonic. Thus, the Spearman correlation measures the degree of monotonic relationship between two variables. In either case, the Spearman correlation is identified by the symbol r_s to differentiate it from the Pearson correlation. When we have small participant numbers and are uncertain as to whether we meet the assumptions for Pearson's r , it is advised to use the Spearman's ρ .

Spearman's ρ transforms the original scores into ranks before performing further calculations. This test is less sensitive to bias due to the effect of outliers. This can be used to reduce the weight of outliers (large distances get treated as a one-rank difference). It does not require assumption of normality.

Look at the following data. Nine people were asked to rate the attractiveness of a target person, and then rate themselves (myattract) on a 10-point scale from 1 (awful) to 10 (wonderful). Small participant numbers, the nature of the data and the fact that many participants rated themselves as near-wonderful should make you suspicious about such data conforming to the assumptions for a parametric test.

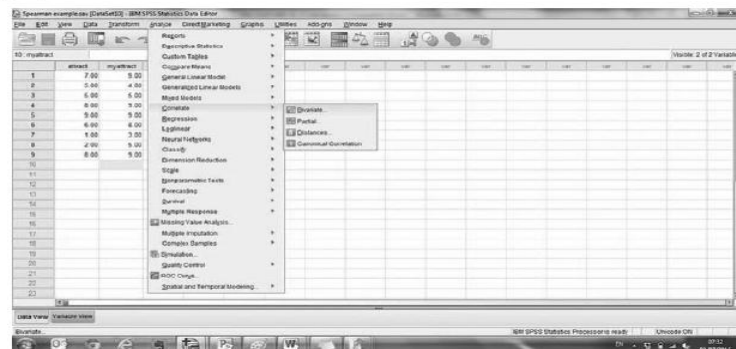
Hypotheses

The hypothesis in this example are given below:

- H_0 : There is no correlation between the X (person attractiveness) and the Y (self-attractiveness).
- H_1 : There is a correlation between the X and the Y.

attract	myattract	attract	myattract
7.00	9.00	6.00	8.00
5.00	4.00	1.00	3.00
5.00	5.00	2.00	5.00
8.00	9.00	8.00	9.00
9.00	9.00		

Open your datafile. Choose *Analyze, Correlate, Bivariate*:



Entering Data

- Create two variables (i.e., personattract and myattract) in the SPSS data editor as shown.

spearman correlaton.sav [DataSet0] - IBM SPSS Statistics Data Editor

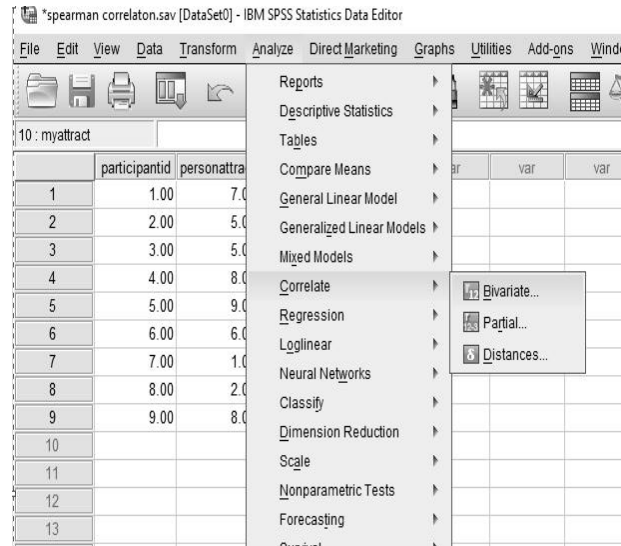
	Name	Type	Width	Decimals	Lab
1	participantid	Numeric	8	2	
2	personattract	Numeric	8	2	
3	myattract	Numeric	8	2	
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					

- Enter data for personal attract and my attract in two columns respectively.

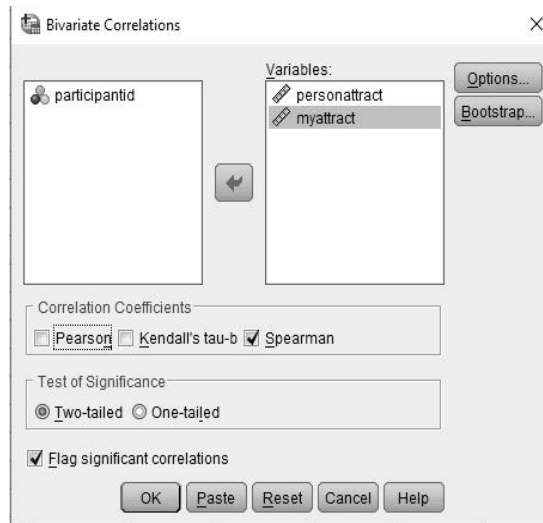
*spearman correlaton.sav [DataSet0] - IBM SPSS Statistics Data Editor

	participantid	personattract	myattract	var
1	1.00	7.00	9.00	
2	2.00	5.00	4.00	
3	3.00	5.00	5.00	
4	4.00	8.00	9.00	
5	5.00	9.00	9.00	
6	6.00	6.00	8.00	
7	7.00	1.00	3.00	
8	8.00	2.00	5.00	
9	9.00	8.00	9.00	
10				
11				
12				
13				
14				
15				
16				
17				
18				

- In the top menu, go to **Analyze > Correlate > Bivariate**.



- This brings us to the bivariate correlations dialogue box.
- We need to uncheck the default Pearson option and select Spearman.
- Move the variables of interest from the left-hand side to the variable list on the right.
- Click OK.



- As in the matrix from Pearson's r , we only need to focus on one half of the matrix.
- The results indicate that there is significant positive correlation between the ratings given to a target person and oneself ($\rho = .92, p < .001$).

Nonparametric Correlations

[DataSet0] C:\Users\Peaceful Tears\Desktop\spearman correlaton.sav

Correlations				
			personattract	myattract
Spearman's rho	personattract	Correlation Coefficient	1.000	.921**
		Sig. (2-tailed)	.	.000
		N	9	9
	myattract	Correlation Coefficient	.921**	1.000
		Sig. (2-tailed)	.000	.
		N	9	9

** . Correlation is significant at the 0.01 level (2-tailed).

Reporting Results

The relationship between person attractiveness and self-attractiveness was investigated using Spearman correlation coefficient. There was a strong, positive correlation between the two variables, $\rho = .92, n = 9, p < .001$.

Lesson 43

MANN-WHITNEY U-TEST

The Mann-Whitney U Test is used to test for differences between two independent groups on a continuous measure. The Mann-Whitney test is designed to use the data from two separate samples to evaluate the difference between two treatments (or two populations). The calculations for this test require that the individual scores in the two samples be rank-ordered. For example, do males and females differ in terms of their self-esteem? This test is the non-parametric alternative to the t-test for independent samples. Instead of comparing means of the two groups, as in the case of the t-test, the Mann-Whitney U Test actually compares medians.

It converts the scores on the continuous variable to ranks across the two groups. It then evaluates whether the ranks for the two groups differ significantly. As the scores are converted to ranks, the actual distribution of the scores does not matter. This test is far simpler than the t-tests in that it does not involve calculations of means, standard deviations and standard errors.

The mathematics of the Mann-Whitney test are based on the following simple observation: A real difference between the two treatments should cause the scores in one sample to be generally larger than the scores in the other sample. If the two samples are combined and all the scores are ranked, then the larger ranks should be concentrated in one sample and the smaller ranks should be concentrated in the other sample. For example, for the Mann Whitney U test, we will be exploring the impact of gender on participants' levels of self-esteem using Survey.sav SPSS data file. For this purpose, research questions are given below:

Research Question

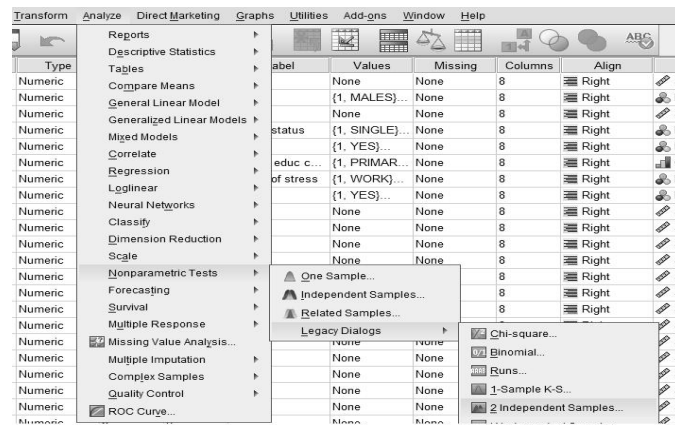
- Do males and females differ in terms of their levels of self-esteem?
- Do males have higher levels of self-esteem than females?

Mann Whitney U Test in SPSS

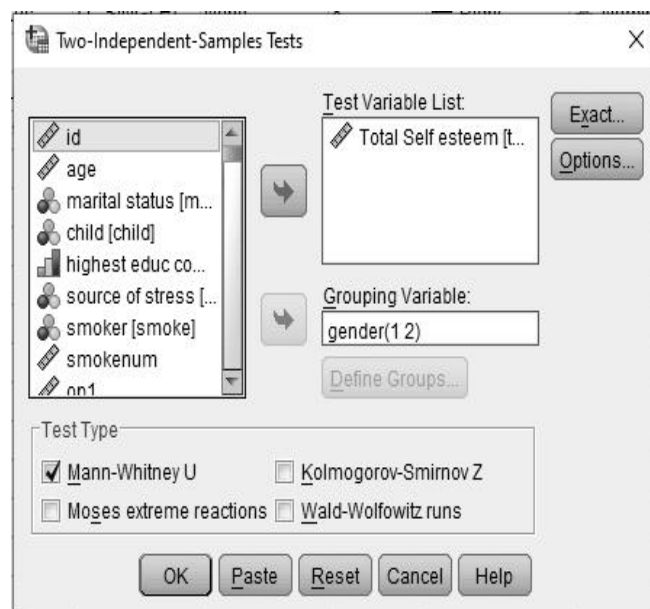
To run this test in SPSS we need two variables: One categorical variable with two groups (e.g., gender) and one continuous variable (e.g., total self-esteem).

Running Analysis

- Choose Analyze, Nonparametric Tests, Legacy Dialogs and 2 Independent Samples as shown.



- Move the DV to the Test variable list and IV to the grouping variable boxes.
- Click on define groups as we did with the t-test. Make sure to define them.
- Also check Mann-Whitney U option. Press OK.



- The first section of the output shows the mean ranks of both groups, and the sum of ranks.
- We can also check the number of participants in each condition.

Ranks				
gender		N	Mean Rank	Sum of Ranks
Total Self esteem	MALES	184	227.14	41794.00
	FEMALES	252	212.19	53472.00
Total		436		

Test Statistics ^a	
	Total Self esteem
Mann-Whitney U	21594.000
Wilcoxon W	53472.000
Z	-1.227
Asymp. Sig. (2-tailed)	.220

a. Grouping Variable: gender

- The next section shows the value of the test statistic, in this case the Mann–Whitney U, which is 21594.0.
- In this case the **Sig. value is P=.220.**
- This is not less than .05, therefore the result is not significant.
- In other words the distribution of self-esteem scores is same across both categories of gender.
- SPSS does not provide an effect size statistic for Mann-Whitney U test, however, we can calculate an approximate value of r using the z value (called **standardized test statistic**) which is available in the Test statistic table (z= -1.23).
- $r = z / \text{square root of } N$ where N = total number of cases.
- In this example, $z = -1.23$ and $N = 436$; $-1.23/\sqrt{436}$
- **Therefore the r value is .06.**
- This would be considered a very small effect size using Cohen (1988) criteria of .1=small effect, .3=medium effect, .5=large effect.

Reporting Results

A Mann-Whitney U Test revealed no significant difference in the self-esteem levels of males (Mean Rank = 227.1, n =184) and females (Mean Rank = 212.9, n = 252), $U = 21594$, $z = -1.23$, $p = .22$, $r = .06$.

Lesson 44

WILCOXON SIGNED RANK TEST**Alternative to the paired t-test: the Wilcoxon**

The Wilcoxon Signed Rank Test (also referred to as the Wilcoxon matched pairs signed ranks test) is designed for use with repeated measures; that is, when the participants are measured on two occasions, or under two different conditions. The Wilcoxon test is designed to evaluate the difference between two treatments, using the data from a repeated-measures experiment. Recall that a repeated-measures study involves only one sample, with each individual in the sample being measured twice. The difference between the two measurements for each individual is recorded as the score for that individual. The Wilcoxon test requires that the differences be rank-ordered from smallest to largest in terms of their absolute magnitude, without regard for sign or direction.

For the Wilcoxon test, there are two possibilities for tied scores:

1. A participant may have the same score in treatment 1 and in treatment 2, resulting in a difference score of zero.
2. Two (or more) participants may have identical difference scores (ignoring the sign of the difference).

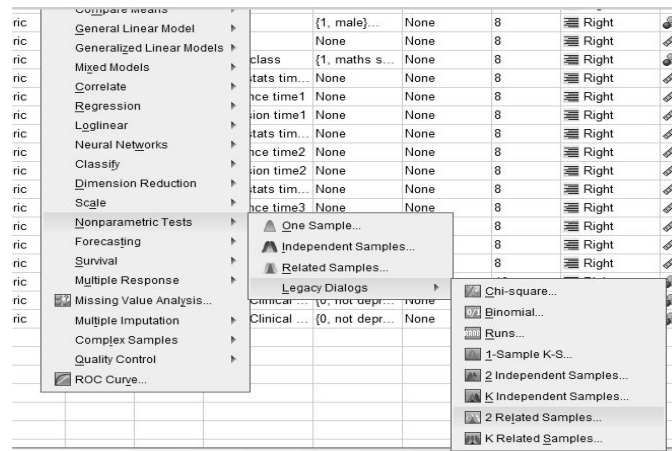
Imagine the psychologist is interested in the change in depression levels. She wants to compare the BDI scores on Sunday to those on Wednesday. Suppose the distributions of scores for drugs is not normal on one of the two days. So the psychologist would have to use a non-parametric test. The Wilcoxon signed-rank test is based on ranking the differences between scores in the two conditions we're comparing. Once these differences have been ranked, the sign of the difference (positive or negative) is assigned to the rank.

To illustrate the use of this test, we will be using data from the `experim.sav` file. In this example, in which we are studying "Is there a change in the scores on the Fear of Statistics Test from Time 1 to Time 2?" We will compare the scores on a Fear of Statistics Test administered before and after an intervention.

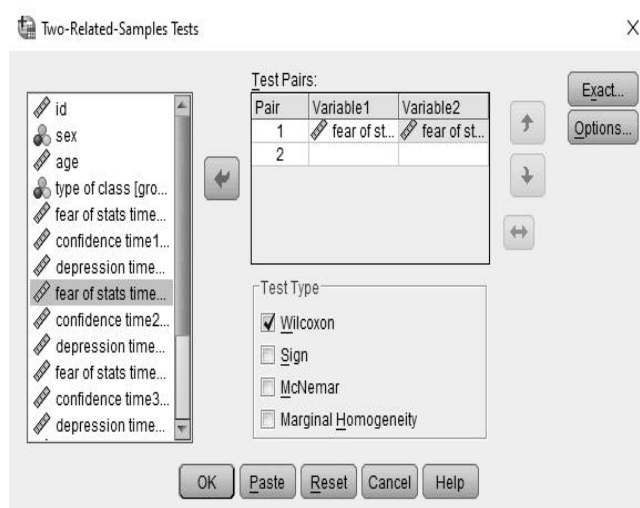
Running Analysis

- To run this test in SPSS, there is need of one group of participants measured on the same continuous scale or measured on two different occasions.

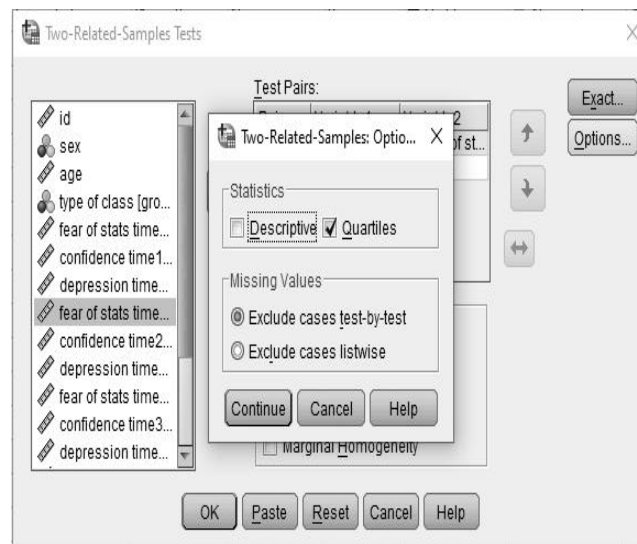
- The variables involved are scores at Time 1 or Condition 1, and scores at Time 2 or Condition 2.
- To follow along with this example, open the **experim.sav** data file.
- From the menu at the top of the screen, click on **Analyze**, then select **Nonparametric Tests**, then **Legacy Dialogs** and then **2 Related Samples**.



- Click on the variables that represent the scores at Time 1 and at Time 2 (e.g., fear of stats time1: fost1, fear of stats time2: fost2).
- Click on the arrow to move these into the Test Pairs box.
- Make sure that the Wilcoxon box is ticked in the Test Type section.



- Click on the Options button. Choose Quartiles (this will provide the median scores for each time point).
- Click on Continue and then on OK.



- The first table in the output gives descriptive statistics (Median values) for time 1 and time 2.

Descriptive Statistics

	N	Percentiles		
		25th	50th (Median)	75th
fear of stats time1	30	37.00	40.00	44.00
fear of stats time2	30	34.50	38.00	40.00

Interpreting Results

- The two things of interest are the Z value and the associated significance levels, presented as Asymp. Sig. (2-tailed).
- If the significance level is equal to or less than .05. We can conclude that the difference between the two scores is statistically significant.

Test Statistics^a

	fear of stats time2 - fear of stats time1
Z	-4.180 ^b
Asymp. Sig. (2-tailed)	.000

a. Wilcoxon Signed Ranks Test

b. Based on positive ranks.

- The effect size for this test can be calculated using the same procedure as described for the Mann-Whitney U Test; that is, by dividing the z value by the square root of N.
- In this situation, N = the number of observations over the two time points, not the number of cases.
- For this calculation we can ignore any negative sign out the front of the z value.
- In this example, $Z = 4.18$, $N = 60$ (cases $\times$ 2); therefore $r = .54$, indicating a large effect size using Cohen (1988) criteria of .1 = small effect, .3 = medium effect, .5 = large effect.

Test Statistics^a

	fear of stats time2 - fear of stats time1
Z	-4.180 ^b
Asymp. Sig. (2-tailed)	.000

a. Wilcoxon Signed Ranks Test

b. Based on positive ranks.

Reporting Results

A Wilcoxon Signed Rank Test revealed a statistically significant reduction in fear of statistics following participation in the training program, $z = -4.18$, $p < .001$, with a large effect size ($r = .54$). The median score on the Fear of Statistics Scale decreased from pre-program ($Md = 40$) to post-program ($Md = 38$).

Lesson 45

KRUSKAL WALLIS & FRIEDMAN TEST**The Kruskal Wallis Test: An Alternative to the Independent-Measure ANOVA**

The Kruskal-Wallis Test (sometimes referred to as the Kruskal-Wallis H Test) is the non-parametric alternative to a one-way between-groups analysis of variance. It allows you to compare the scores on some continuous variable for *three or more groups*. It is similar in nature to the Mann-Whitney U Test, but it allows you to compare more than just two groups. Scores are converted to ranks and the mean rank for each group is compared. This is a ‘*between groups*’ analysis, so different people must be in each of the different groups.

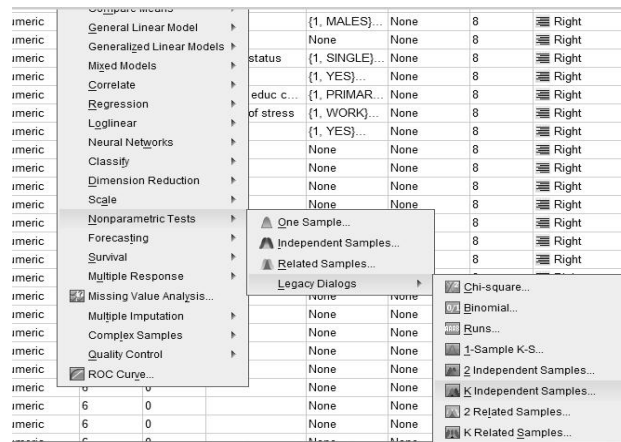
To begin with, scores are ordered from lowest to highest, ignoring the group to which the score belongs. The lowest score is assigned a rank of 1, the next highest a rank of 2 and so on. Once ranked, the scores are collected back into their groups and their ranks are added within each group. The Kruskal–Wallis test tells us that, overall, groups come from different populations. However, it doesn’t tell us which groups differ. The simplest way to break down the overall effect is to compare all pairs of groups (known as pairwise comparisons). Let’s take an example to run this Kruskal Wallis Test in SPSS. In this example, we will explore levels of optimism across three age groups. The research question is: Is there a difference in optimism levels across three age levels?

We need two variables:

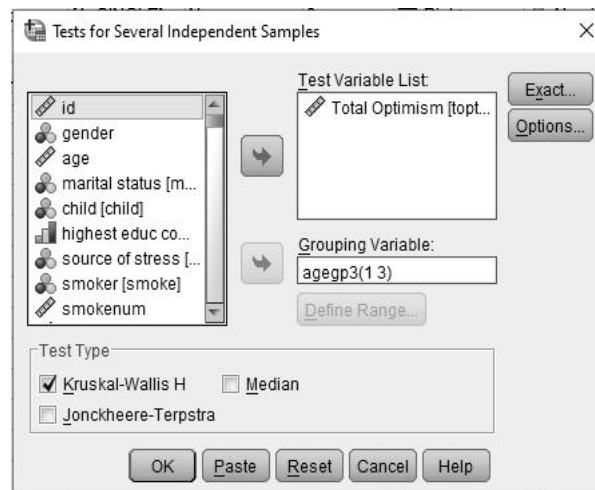
- One categorical independent variable with three or more categories (e.g. age 3: 18–29, 30–44, 45+)
- One continuous dependent variable (e.g. total optimism).

Running Analysis

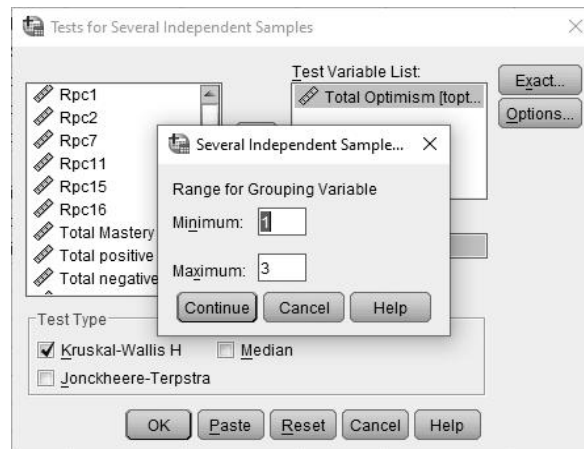
- To follow along with this example, open the survey.sav data file.
- From the menu at the top of the screen click on Analyze, select Nonparametric Tests, Legacy Dialogs and K Independent samples.



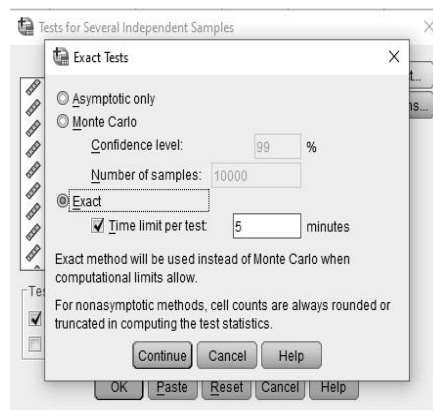
- Select the continuous dependent variable (e.g., Total optimism) and move it into the **Test Variable List** box.
- Click on the categorical independent variable (e.g., age 3 groups; agegp3) and move it into the **Grouping Variable** box.



- Then press the **Define Range** button. This enables us to define the range of the groups.
- Here, our groups are coded 1, 2, and 3, so the range is 1-3.
- Enter the range into the dialogue box and press **Continue**.



- Select the **Exact** button and check the **Exact** option.
- Click **Continue**.



- Ensure that the **Kruskal–Wallis H** box is checked.
- For Descriptives, press the **Options** button and select descriptive option.
- Press **OK**.

