

One Sample t-test, Explore, Independent t-test & Paired Samples t-test

One Sample t-test

This test compares the mean value of a sample of data to a known value in the population.

Question 1

The 40-hour work week is an established standard for full time employees. Work less, and you feel guilty for not putting in a full week. Work more, and you feel that you are performing beyond expectations. Use the data below collected from a random sample of 45 graduate employees in a certain week to test, at the 5% significance level, the hypothesis that the average university graduate who is employed full time, works a 40-hour week.

Number of Hours worked for University Graduates								
35	31	45	41	35	46	20	35	60
32	38	26	48	47	40	36	32	40
44	39	28	55	46	36	27	48	35
29	38	42	41	37	37	48	33	55
41	25	51	38	37	52	44	43	40

Select a new datasheet (**week2_work hours**) using **File, New, Data**.

Click on **Variable View** in the bottom left-hand corner of the screen and format a variable for Hours.

Enter **Hours worked per week for University Graduates** under Label, in order to add detail to the information generated in the output.

Click on **Data View** and enter the data as shown below for the 45 cases.

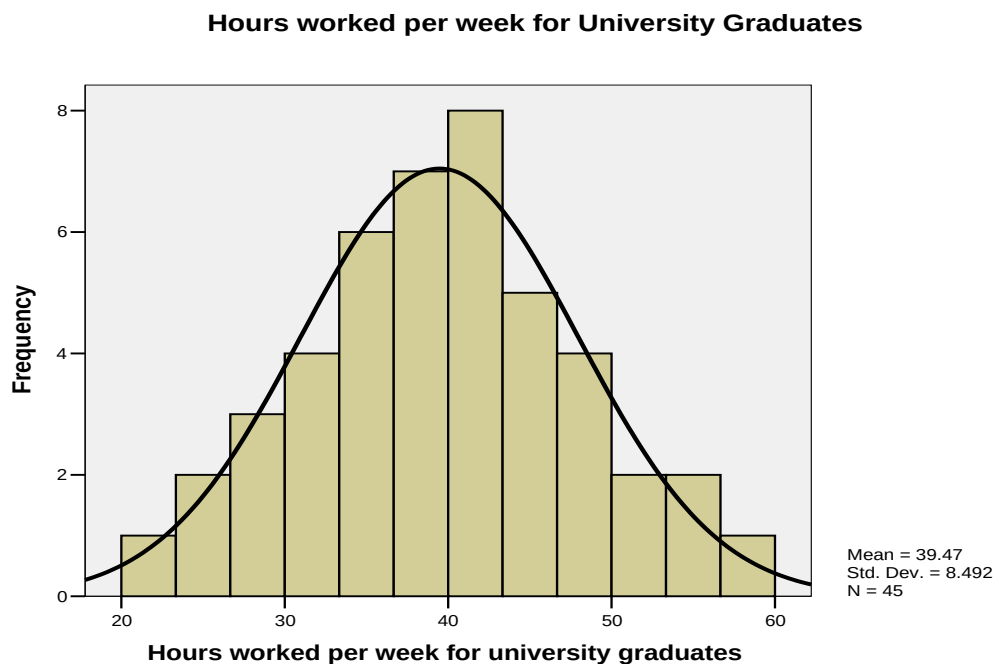
	Hours
1	35
2	32
3	44
4	29

5	41
6	31
7	38
8	39
9	38
10	25

etc.

Save.

To examine the distribution of the data, select **Graphs, Legacy Dialogs, Histogram,** and transfer the variable **Hours worked per week for University Graduates (Hours)** into the Variable dialogue box. Select the check box **Display Normal Curve.** Click on **Titles** and enter **Hours worked per week for University Graduates, Continue, OK.**



The data displays an approximate Normal Distribution, which is a condition for the reliable application of the one sample t-test.

One Sample t-test

Test, at the 5% significance level, the hypothesis that the average university graduate who is employed full time, works a 40 hour week.

Null Hypothesis H_0 : The mean number of hours worked by graduates is 40hrs

Alternative Hypothesis H_1 : The mean number of hours worked by graduates is not 40hrs

Select **Analyze, Compare Means, One-Sample T test**.

Move the variable **Hours worked per week for University Graduates (Hours)** into the Test Variable dialogue box. In Test Value type **40, OK**.

The output is shown below.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Hours worked per week for university graduates	45	39.47	8.492	1.266

One-Sample Test

	Test Value = 40					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Hours worked per week for university graduates	-.421	44	.676	-.533	-3.08	2.02

There are two tables of output. The **One Sample Statistics** provides the standard descriptive statistics.

One Sample Test

$t = -0.421$ and the Sig. = 0.676.

Since Sig. > 0.05 , then Accept H_0 at the 5% level of significance.

The sample evidence indicates that the average university graduate who is employed full time, works a 40 hour week.

The mean difference in this sample between the number of hours worked and the test mean (of 40 hrs) is -0.533 hrs, but there is a 95% chance that this mean difference for the population of all graduates could lie between -3.08 hrs and 2.02 hrs.

Note: No difference between the test value and the sample mean would be represented by a mean difference of 0. Since a mean difference of 0 is plausible as it falls within the confidence limits -3.08hrs to 2.02hrs then this information also concludes that there is not enough evidence to suggest a significant difference.

Question 2

Return to the data editor.

Select the data set **World95.sav** which contains demographic data from a sample of countries throughout the world.

From the data set the variable considered is:

Population increase % per year (pop_incr)

Test, at the 5% significance level, whether the mean population increase (%) per year is greater than a previous estimate of 1.5%.

Answer the following from the output obtained:

This is a one-tailed test and hence the significance in the output will need to be halved.

The significance value is _____

Is H_0 accepted or rejected? _____

Is the mean population increase % per year greater than a previous estimate of 1.5%? _____

Explain two factors shown by the 95% confidence interval of the mean difference.

(i) _____

(ii) _____

Examine the distribution of the data by constructing a histogram with a Normal Curve. Comment on your findings.

Explore, Independent t-test & Paired Samples t-test

Independent Samples t-test

This test is used to determine whether there is any difference between two independent groups of data.i.e. investigating whether the data comes from the same or different distributions.

Question 1

A company was piloting a new model of DVD player and was interested in testing the difference between the assessment of the new model between potential male and female customers.

Select the data set **dvdplayer.sav** which contains the results of a survey where an assessment score was evaluated for the product, based on a number of factors including price, functionality, appearance, ease of use and support.

From the data set, the variables considered are:

Male or Female (sex) has labels 1 = Male, 2 = Female

Total DVD Assessment (dvdscore)

Exploring the Data

Investigation of the data is first necessary in order to establish whether the data follows a Normal distribution.

Select **Analyze, Descriptive Statistics, Explore.**

Transfer **Total DVD Assessment (dvdscore)** into the Dependent List, and **Male or Female (sex)** into the Factor list.

Click on **Plots**, deselect **Stem-and-leaf** and select **Histogram** under Descriptive, **Continue, OK.**

The output is shown below.

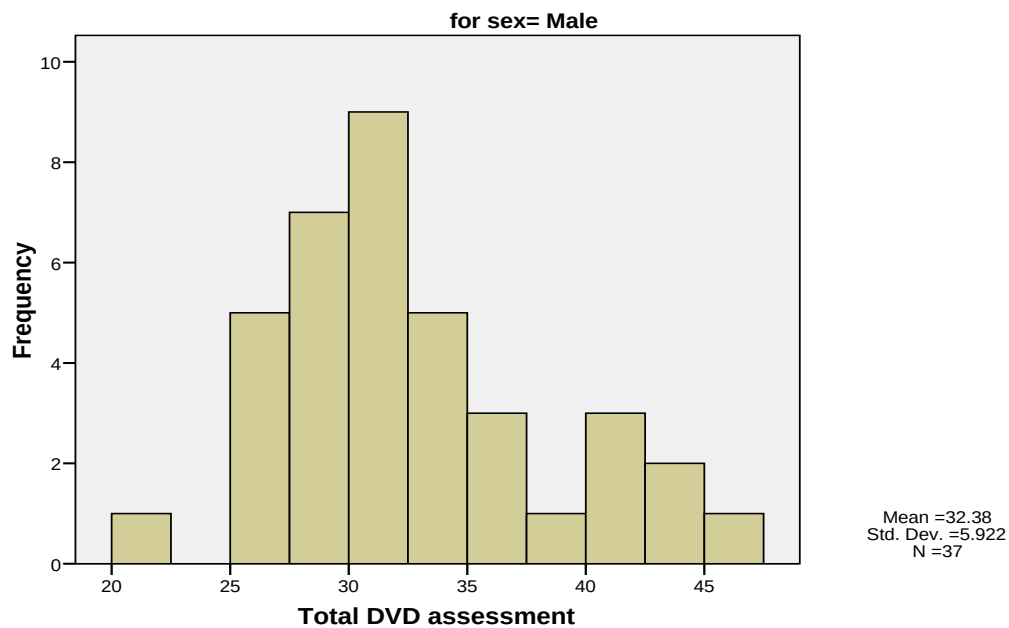
Case Processing Summary							
		Cases					
		Valid		Missing		Total	
		N	Percent	N	Percent	N	Percent
Total DVD assessment	Male	37	100.0%	0	.0%	37	100.0%
	Female	31	100.0%	0	.0%	31	100.0%

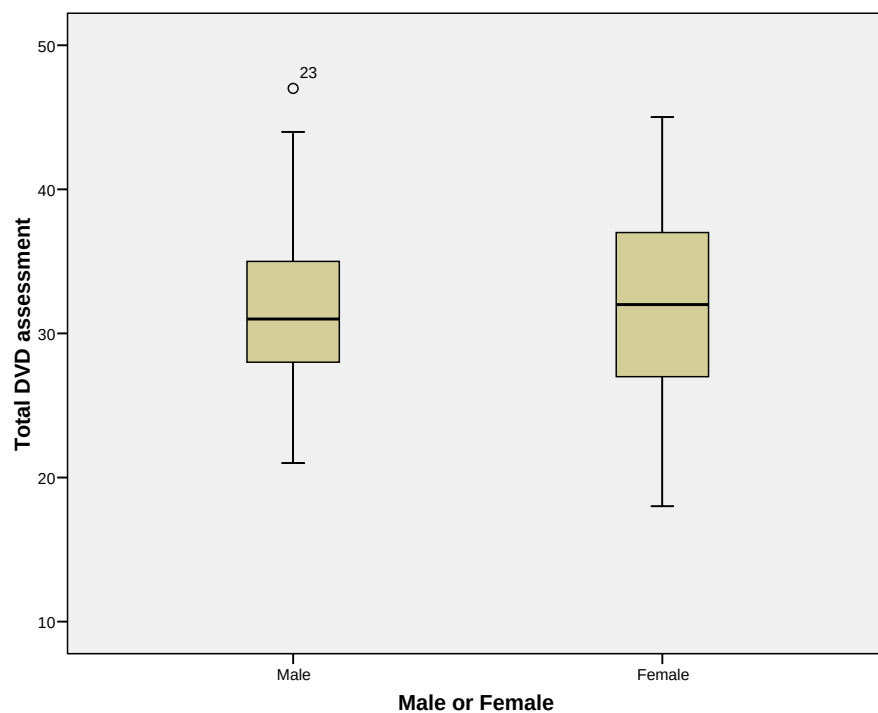
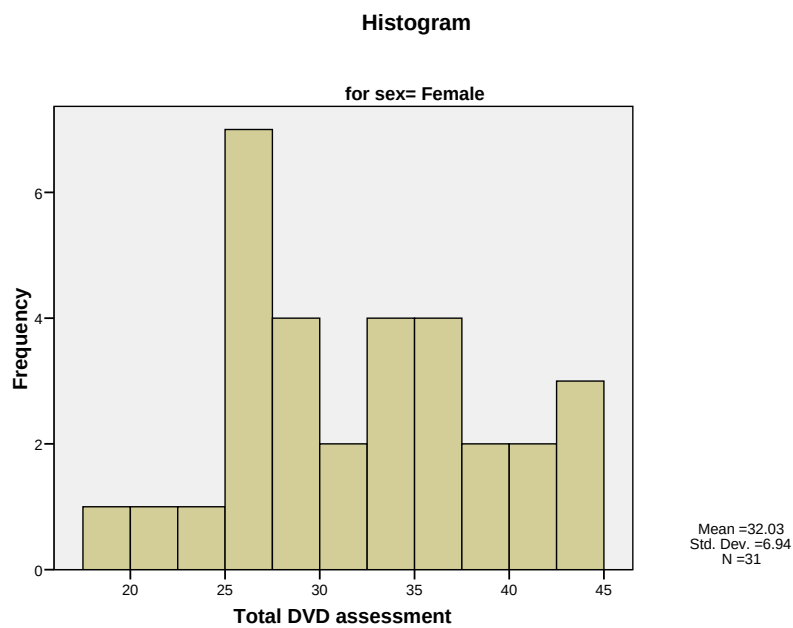
Descriptives

Male or Female				Statistic	Std. Error
Total DVD assessment	Male	Mean		32.38	.974
		95% Confidence Interval for Mean	Lower Bound	30.40	
			Upper Bound	34.35	
		5% Trimmed Mean		32.17	
		Median		31.00	
		Variance		35.075	
		Std. Deviation		5.922	
		Minimum		21	
		Maximum		47	
		Range		26	
		Interquartile Range		7	
		Skewness		.662	
		Kurtosis		.110	
	Female	Mean		32.03	1.246
		95% Confidence Interval for Mean	Lower Bound	29.49	
			Upper Bound	34.58	
		5% Trimmed Mean		32.07	
		Median		32.00	
		Variance		48.166	
		Std. Deviation		6.940	
		Minimum		18	
		Maximum		45	
		Range		27	
		Interquartile Range		10	
		Skewness		.082	
		Kurtosis		-.664	

Total DVD assessment

Histogram





The descriptive statistics show that there is a difference in the male and female total DVD assessment scores. The other statistics suggest that both distributions are reasonably normal, since there is no significant skewness or kurtosis.

The shape of the histograms are approximately normal and the boxplots indicate that the two groups have similar variances.

An overall conclusion from all the evidence implies an approximate normal distribution.

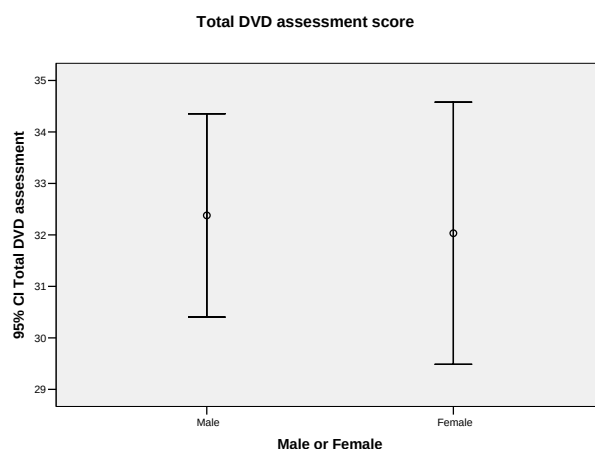
Error bars display the dispersion of the data and show the upper and lower range within which it is 95% likely that the true population mean lies.

Select **Graphs, Legacy dialogs, Error Bar** and with **Simple** error bar selected, click on **Define**. Transfer **Total DVD Assessment (dvdscore)** into the Variable dialogue box and **Male or Female (sex)** into the Category Axis dialogue box.

Click on **Titles** and type **Total DVD assessment Score** in Line 1 of Title, **Continue, OK**.

The output is shown below.

Graph



The error bars indicate complete overlap in the two distributions, showing that the distributions are not very different. This can be confirmed using the t-test.

Independent T-Test

Null Hypothesis

H_0 : There is no difference in the total DVD assessment score between the male and female participants.

Alternative Hypothesis

H_1 : There is a difference in the total DVD assessment score between the male and female participants.

Select **Analyze, Compare Means, Independent-Samples T Test**.

Transfer **Total DVD Assessment (dvdscore)** into the Test Variable dialogue box and **Male or Female (sex)** into the grouping Variable dialogue box.

Click on **Define Groups**. In Group 1: type **1** and in Group 2: type **2**, **Continue**, **OK**.

The output is shown below.

T-Test

Group Statistics

	Male or Female	N	Mean	Std. Deviation	Std. Error Mean
Total DVD assessment	Male	37	32.38	5.922	.974
	Female	31	32.03	6.940	1.246

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	
Total DVD assessment	Equal variances assumed	1.815	.183	.222	66	.825	.346	1.560	-2.768	3.460
	Equal variances not assumed			.219	59.361	.828	.346	1.582	-2.818	3.511

For Levene's Test for the Equality of Variances, $F = 1.815$, $\text{Sig.} = 0.183 > 0.05$

Since the variances are not significantly different, at the 5% level, then the top row of the t-test should be used.

For the t-test, with equal variances assumed, $t = 0.222$ with 66 degrees of freedom, $\text{Sig.} = 0.825 > 0.05$. Accept H_0 at the 5% significance level.

The sample evidence indicates that there is no significant difference in the total DVD assessment score between the males and the females.

Question 2

Return to the data editor. Select the data set **Graduate earnings.sav** which contains earnings data for recent graduates and non-graduates (those without a degree).

Examine whether there is any difference in the mean earnings of the two groups of individuals.

Graduates

Do the mean and trimmed means have similar values? _____

Is the skewness significantly different from a normal distribution? _____

Is the kurtosis significantly different from a normal distribution? _____

Non-Graduates (those without a degree)

Do the mean and trimmed means have similar values? _____

Is the skewness significantly different from a normal distribution? _____

Is the kurtosis significantly different from a normal distribution? _____

Does the shape of the histograms indicate a normal distribution? _____

State two factors that can be seen by the shape of the boxplots.

(i) _____

(ii) _____

Would it be suitable to use the independent t-test to examine any difference in the attention?
remaining between the pilots flying manual and automated cockpits? _____

Draw the error bars. What information is shown from the error bars?

Question 3

Return to the data editor.

Select the data set **Employee data.sav** which contains a sample of a company's records on its employees.

From the data set the variables considered are:

Previous experience in months (prevexp)

Gender (gender)

Use the Explore procedure to examine the distribution of the data in the variable **prevexp** for the male and female employees.

Answer the following from the output obtained.

Female

Do the mean and trimmed means have similar values? _____

Is the skewness significantly different from a normal distribution? _____

Is the kurtosis significantly different from a normal distribution? _____

Male

Do the mean and trimmed means have similar values? _____

Is the skewness significantly different from a normal distribution? _____

Is the kurtosis significantly different from a normal distribution? _____

Does the shape of the histograms indicate a normal distribution? _____

State two factors that can be seen by the shape of the boxplots.

(i) _____

(ii) _____

Would it be suitable to use the independent t-test to examine any difference in the previous experience between the female and male employees? _____

Question 4 (use sales.xlsx)

An industrial psychologist wishes to study the effects of motivation on sales in a particular firm. Of 24 new salespersons, 12 are paid an hourly rate and 12 are paid a commission. The 24 individuals are randomly assigned to the two groups. The following data represents the sales volume (in thousands of pounds) achieved during the first month on the job.

	Sales (£000)					
Hourly Rate	256	239	222	207	228	241
	212	216	236	219	225	230
Commission	224	273	237	261	254	262
	254	285	277	258	225	245

Test, at the 5% significance level, whether there is any evidence that wage incentives (through commission) yield a greater average sales volume.

Return to the data editor.

Select File, New, Data.

Click on **Variable View** in the bottom left-hand corner of the screen and format a variable for Method of Payment (code 1 = Hourly Rate and 2 = Commission), and a variable for Sales.

Click on **Data View** and enter the data as shown overleaf.

	method	sales
1	1	256
2	1	239
3	1	222
4	1	207
5	1	228
6	1	241
7	1	212
8	1	216
9	1	236
10	1	219
11	1	225
12	1	230
13	2	224
14	2	273
15	2	237
16	2	261
16	2	254
18	2	262
19	2	254
20	2	285
21	2	277
22	2	258
23	2	225
24	2	245

Use the Explore procedure to examine the distribution of the data in the sales variable for the hourly rate and the commission paid employees.

Answer the following from the output obtained.

Hourly Rate

Do the mean and trimmed means have similar values? _____

Is the skewness significantly different from a normal distribution? _____

Is the kurtosis significantly different from a normal distribution? _____

Commission

Do the mean and trimmed means have similar values? _____

Is the skewness significantly different from a normal distribution? _____

Is the kurtosis significantly different from a normal distribution? _____

Does the shape of the histograms indicate a normal distribution? _____

Do the boxplots indicate similar variances? _____

Does the data satisfy the requirements to use the Independent t-test? _____

Draw a graph of the Error bars of the data for the hourly paid and commission paid salespersons.

Do the error bars indicate a significant difference? _____

Perform the Independent T-test to examine whether there is any evidence that wage incentives (through commission) yield a greater average sales volume.

From the output answer the following:

Does Levene's Test indicate similar variances? _____

What is the significance of this one-tailed Independent t-test? _____

Is H_0 accepted or rejected? _____

Is there any evidence that wage incentives (through commission) yield a greater average sales volume? _____

Explain two factors shown by the 95% confidence interval of the mean difference.

(i) _____

(ii) _____

Paired Samples T-Test

The paired t-test is used to examine the differences between two groups of data which are directly related.

Question 5

Select the data set **Salaries.sav** which contains records of a sample of employees of a large organisation.

Test, at the 5% significance level, whether there is any increase in the proposed productivity-based salary as compared with the present salary of the employees.

Exploring the Data

Investigation of the data is first necessary in order to establish whether the distribution of the differences between the two sets of data are normal.

The differences between the paired data must first be evaluated.

It is necessary to create a variable called **diff** by subtracting the present salary (salnow) from the initial salary (salbeg).

Select **Transform, Compute** and type in **diff** in Target Variable.

Transfer **productivity-based salary (prodsal)** into the Numeric expression dialogue box, click on **the minus sign** on the keypad, transfer **present salary (salary)** into the Numeric expression, **OK**, then **OK**.

Select **Analyze, Descriptive Statistics, Explore**.

Transfer **diff** into the Dependent List.

Click on **Plots**, deselect **Stem-and-leaf** and select **Histogram** under Descriptive.

The output is shown below.

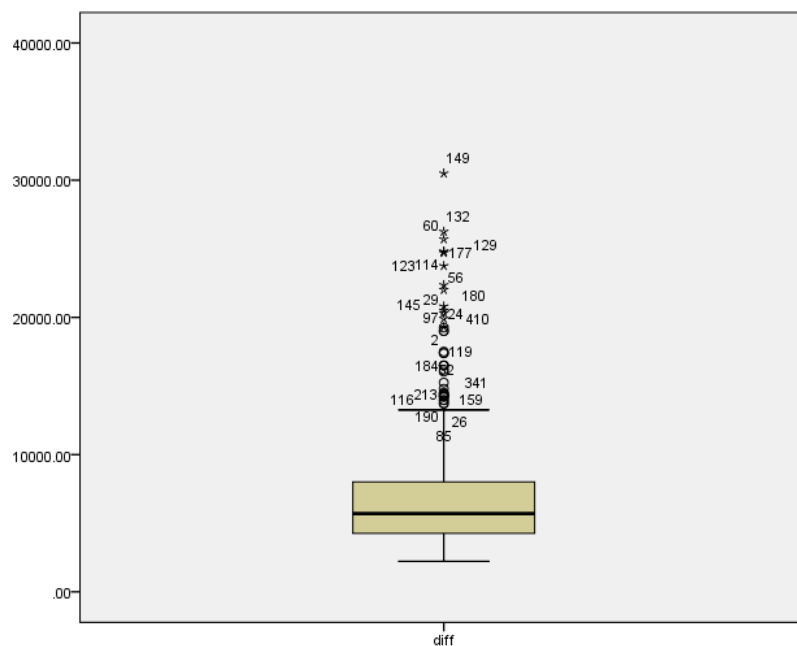
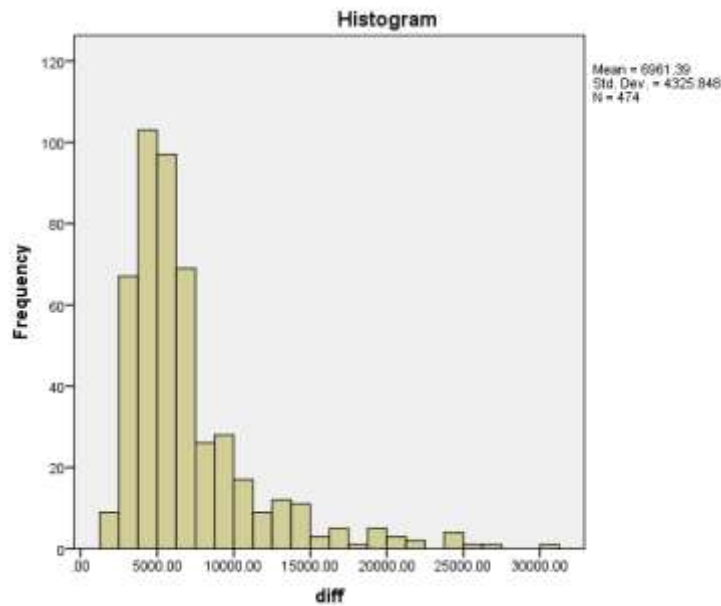
Explore

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Diff	474	100.0%	0	0.0%	474	100.0%

Descriptives

	Statistic	Std. Error
Mean	6961.3924	198.69278
95% Confidence Interval for Mean	Lower Bound 6570.9627 Upper Bound 7351.8221	
5% Trimmed Mean	6430.4079	
Median	5700.0000	
Variance	18712960.780	
Std. Deviation	4325.84798	
Minimum	2220.00	
Maximum	30496.00	
Range	28276.00	
Interquartile Range	3756.00	
Skewness	2.182	.112
Kurtosis	5.764	.224



The descriptive statistics suggest that the difference between the two variables is not a normally distributed variable since there is a significant skewness and kurtosis. But, we can still proceed with the Paired T test.

Paired T-Test

Null Hypothesis

H_0 : There is no difference between the initial salary and the present salary.

Alternative Hypothesis

H_1 : The present salary is greater than the initial salary.

Return to the data editor.

Select **Analyze, Compare Means, Paired-Samples T-Test.**

Click on both **present salary(salnow)** and **initial salary (salbeg)** then select them for Paired Variables, **OK.**

The output is shown below.

T-Test

Paired Samples Statistics

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1 BEGINNING SALARY	6806.43	474	3148.255	144.604
CURRENT SALARY	13767.83	474	6830.265	313.724

Paired Samples Correlations

	N	Correlation	Sig.
Pair 1 BEGINNING SALARY & CURRENT SALARY	474	.880	.000

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	BEGINNING SALARY - CURRENT SALARY	-6961.392	4325.848	198.693	-7351.822	-6570.963	-35.036	473	.000

$t = -35.036$ and the Sig. = 0.000 for a two-tailed test.

However, this is a one-tailed test and it is necessary to halve the significance value, but, of course, $.000/2$ still is .000. Reject H_0 at the 5% significance level.

There is evidence to suggest that the current salary is greater than the beginning salary.

Question 6

Return to the data editor.

Select the data set **World95.sav** which contains demographic data from a sample of countries throughout the world.

From the data set the variables considered are:

Average female life expectancy (lifeexpf)

Average male life expectancy (lifeexpm)

Test, at the 5% significance level, whether there is any difference in the average life expectancy between the males and the females for the countries of the world.

Evaluate and explore the differences between the paired data

Answer the following from the output obtained.

Do the mean and trimmed means have similar values?_____

Is the skewness significantly different from a normal distribution?_____

Is the kurtosis significantly different from a normal distribution?_____

Does the shape of the histogram and boxplot indicate an approximately normal distribution?_____

Does the data satisfy the requirements to use the Paired t-test?_____

Perform the paired t-test in order to examine whether there is any difference in the average life expectancy between the males and the females for the countries of the world.

Answer the following from the output obtained.

Is there significant correlation between the average male life expectancy and the average female life expectancy?_____

What is the significance of the paired t-test?_____

Is H_0 accepted or rejected?_____

Is there any evidence that there is any difference in the average life expectancy between the males and the females for the countries of the world?_____

Explain two factors shown by the 95% confidence interval of the mean difference.

(i) _____

(ii) _____