2.1 SPSS exercises

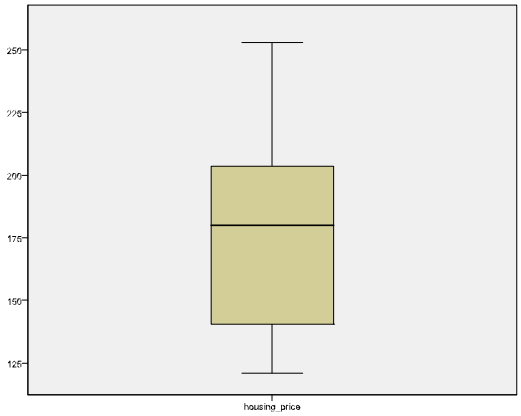
1. A study was conducted to investigate if housing prices have stabilized compared to last year‘s prices, which were in average £195K. Twelve housing prices were collected from the adverts and used for comparison:
£125K, £253K, £207K, £146K, £121K, £135K, £175K, £200K, £210K, £166K, £185K, £192K.
Test at the 5% level if the mean housing price for the current year has changed compared to the last year mean price.

Research hypothesis: Mean housing price for the current year is different from the last year mean price.
Statistical test: One sample t-test
Null and alternative hypotheses:
H0: Mean housing price for the current year is equal to £195K
HA: Mean housing price for the current year is different from £195K
Exploratory data analysis:

Statistics

housing_price

N	Valid	12
	Missing	0
Mean		176.25
Std. Error of Mean		11.437
Median		180.00
Std. Deviation		39.619
Minimum		121
Maximum		253



Sample mean is 176.25 which is lower than 195. At the same time from the boxplot we can see that values span from £121K to £253K, so difficult to say if the true mean is significantly different from £195K.
SPSS test results:

One-Sample Test

	Test Value = 195					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
housing_price	-1.639	11	.129	-18.750	-43.92	6.42

Reporting results from SPSS

Test results: t-test statistics=-1.639, df=11, p-value=0.129

Test decision: p-value=0.129>0.05 – Cannot reject H0

Test conclusion: Insufficient evidence from the data to indicate that the current year mean housing price is significantly different from £195K.

Checking normality assumption:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
housing_price	.114	12	*	.959	12	.774

This is a lower bound of the true significance.

Lilliefors Significance Correction

Shapiro-Wilk test p-value=0.774>0.05. We conclude that the normality assumption is supported by the data and so the results of the one-sample t-test are valid.

2. It is claimed that a 12-week special exercise program will significantly reduce weight. A random sample of 6 people was selected and these people followed the program. The following table gives their weights (in kg) before and after the program.

Subject:	1	2	3	4	5	6
Before	81.8	88.6	80.5	100.5	94.5	90.5
After	83.2	85.0	73.2	92.7	89.5	85.9

Use the 5% significance level to test if the claim is justified by the data.

Research hypothesis: Mean weight before exercise is higher than mean weight after.

Statistical test: Paired samples t-test

Null and alternative hypotheses:

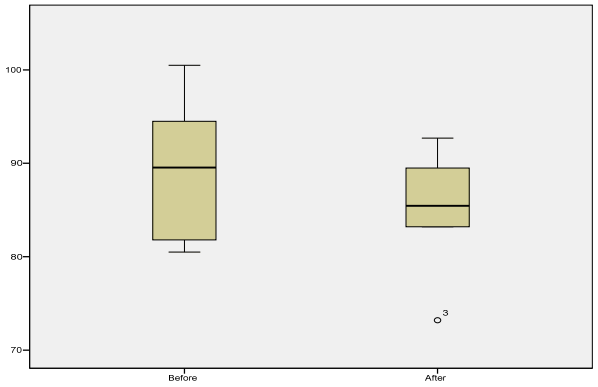
H0: Mean weight before exercise is equal to mean weight after.

HA: Mean weight before exercise is different from mean weight after.

Exploratory data analysis:

Report

	Before	After
Mean	89.400	84.917
N	6	6
Std. Deviation	7.5900	6.6752
Median	89.550	85.450
Minimum	80.5	73.2
Maximum	100.5	92.7
Std. Error of Mean	3.0986	2.7251



Sample mean weight before exercise is higher that sample mean after , however from the boxplot it appears that there is a lot of overlap in values between the two measurements and so difficult to say if results are significant

SPSS test results:

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	Before - After	4.4833	3.3048	1.3492	1.0152	7.9515	3.323	5	.021

Reporting results from SPSS

Test results: t-test statistics=3.323, df=5, p-value=0.021

Test decision: p-value=0.021<0.05 – Reject H0 and accept HA

Test conclusion: Sufficient evidence from the data to indicate that mean weight before is different from mean weight after. Given that the test statistic is positive we conclude that mean weight before is significantly higher than mean weight after and so the physical exercise helps significantly with weight loss.

Checking normality assumption:

Tests of Normality

	Kolmogorov-Smirnov [a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Before	.175	6	.200 *	.956	6	.787
After	.232	6	.200 *	.929	6	.572

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

a. Lilliefors Significance Correction

Shapiro-Wilk test p-values are p-value=0.787 for before and p-value=0.572 for after. Both p-values are greater than 0.05 and so conclude that the normality assumption is supported by the data and so the results of the paired-samples t-test above are valid.

3. An experiment was conducted to investigate the effect of two types of organic acid pre-treatment on functional properties of banana flour. Ten unripe bananas from the Luvhele cultivar were randomly selected and treated with a dose of citric acid and another ten with same dose of lactic acid. Oil holding capacity of flour obtained from each banana was measured and the results (g/g dry weight) were as follows:

Citric	1.88	2.47	1.83	2.20	1.57	1.63	2.51	1.82	1.66	1.56
Lactic	1.37	1.63	1.50	1.12	1.39	1.29	1.47	1.59	1.13	1.13

At 5% level of significance test whether there is a difference in mean holding capacity between the two treatments.

Research hypothesis: Mean oil holding capacity under citric acid treatment is different from mean holding capacity under lactic acid treatment.

Statistical test: Independent samples t-test

Null and alternative hypotheses:

H0: Mean oil holding capacity under citric acid treatment is equal to mean holding capacity under lactic acid treatment.

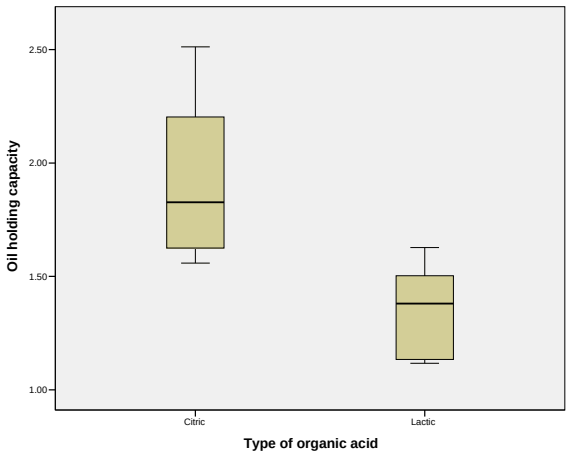
HA: Mean oil holding capacity under citric acid treatment is different from mean holding capacity under lactic acid treatment.

Exploratory data analysis

Report

Oil holding capacity

Type of organic acid	Mean	Std. Error of Mean	N	Std. Deviation	Median	Minimum	Sum
Citric	1.9139	.11363	10	.35933	1.8268	1.56	19.14
Lactic	1.3624	.05981	10	.18912	1.3805	1.12	13.62
Total	1.6381	.08892	20	.39767	1.5786	1.12	32.76



Oil holding capacity sample mean for citric acid treatment is higher than for lactic treatment. At the same time there is only little overlap between the values resulting from the two treatments, which indicate that the two treatments have a different effect. We will formally test that by using the independent samples t-test.

SPSS results

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Oil holding capacity	Equal variances assumed	4.010	.061	4.295	18	.000	.55150	.12841	.28173	.82127
	Equal variances not assumed			4.295	13.631	.001	.55150	.12841	.27539	.82761

Levene's test p-value =0.061>0.05 and therefore equal treatment variances can be assumed. As a result we should report the first line for the independent samples t-test which is:

Test results: t-test statistics=4.295, df=18, p-value=0.000

Test decision: p-value=0.000<0.05 – Reject H0 and accept HA

Test conclusion: Sufficient evidence from the data to indicate that mean holding capacity under citric acid is different from mean holding capacity under lactic acid treatment.

Tests of Normality

		Kolmogorov-Smirnov [a]			Shapiro-Wilk		
	Type of organic acid	Statistic	df	Sig.	Statistic	df	Sig.
Oil holding capacity	Citric	.236	10	.122	.852	10	.062
	Lactic	.186	10	.200 *	.917	10	.331

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

a. Lilliefors Significance Correction

Shapiro-Wilk test p-values are p-value=0.062 for citric acid treatment and p-value=0.331 for lactic acid treatment. Both p-values are greater than 0.05 and so conclude that the normality assumption is supported by the data, and so the results of the independent-samples t-test above are valid.