

04/04/2025

Portfolio Reflective Report

CHANGEMAKERS: CREATIVITY AND VALUE CREATION
LUCA BURGESS (2315074)

UNIVERSITY OF WALES TRINITY SAINT DAVID | 2,143 Words

Table of Contents

1. T-test	1
1.1 Sample Data.....	1
1.2 Choice of Statistical Test	1
1.3 Assumptions	1
1.4 Calculations.....	2
1.5 Outcomes	3
1.6 Discussion and Evaluation.....	4
2 ANOVA.....	5
2.1 Sample Data.....	5
2.2 Choice of Statistical Test	5
2.3 Assumptions	5
2.4 Calculations.....	5
2.5 Outcomes	6
2.6 Discussion and Evaluation.....	7
3 Chi-squared	8
3.1 Sample Data.....	8
3.2 Choice of Statistical Test	8
3.3 Assumptions	8
3.4 Calculations.....	8
3.5 Outcomes	9
3.6 Discussion and Evaluation.....	10
4 Linear Regression.....	11
4.1 Sample Data.....	11
4.2 Choice of Statistical Test	11
4.3 Assumptions	11
4.4 Calculations.....	12
4.5 Outcomes	12
4.6 Discussion and Evaluation.....	14
5 References.....	15

1. T-test

1.1 Sample Data

Spring		Autumn	
Individual	Weight (g)	Individual	Weight (g)
1	20	1	22
2	18	2	20
3	15	3	19
4	19	4	23
5	20	5	21
6	20	6	17
7	21	7	22
8	20	8	20
9	16	9	23
10	15	10	21
11	16	11	19
12	15	12	17
13	19	13	24
14	18	14	19
15	21	15	23

1.2 Choice of Statistical Test

The chosen statistical test is a paired t-test. This is because the data is paired, as the same robins are being measured at two different seasons (spring and autumn), with the dependent variable being the weight of the birds. A paired t-test is designed to compare the means of two groups, which is ideal for this data analysis. This test will measure whether the difference in average weights is significantly different from zero.

1.3 Assumptions

T-tests are generally quite robust to deviations from assumptions. However, t-test do assume that:

- The data has been randomly sampled from a population.
- The data is continuous.
- The distribution is normal.

1.4 Calculations

*T-test.sav [DataSet1] - IBM SPSS Statistics Data Editor

	Spring_Weight	Autumn_Weight	var	var	var
1	20.00	22.00			
2	18.00	20.00			
3	15.00	19.00			
4	19.00	23.00			
5	20.00	21.00			
6	20.00	17.00			
7	21.00	22.00			
8	20.00	20.00			
9	16.00	23.00			
10	15.00	21.00			
11	16.00	19.00			
12	15.00	17.00			
13	19.00	24.00			
14	18.00	19.00			
15	21.00	23.00			
16					

*T-test.sav [DataSet1] - IBM SPSS Statistics Data Editor

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Spring_Weight	Numeric	8	2		None	None	14	Right	Scale	Input
2	Autumn_Weight	Numeric	8	2		None	None	16	Right	Scale	Input
3											
4											
5											
6											
7											
8											
9											
10											
11											

1.5 Outcomes

Paired Samples Statistics					
		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Spring Weight	18.2000	15	2.24245	.57900
	Autumn Weight	20.6667	15	2.19306	.56625

Paired Samples Correlations					
		N	Correlation	Significance	
				One-Sided p	Two-Sided p
Pair 1	Spring Weight & Autumn Weight	15	.363	.092	.183

Paired Samples Test										
		Paired Differences					t	df	Significance	
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				One-Sided p	Two-Sided p
					Lower	Upper				
Pair 1	Spring Weight – Autumn Weight	-2.46667	2.50333	.64636	-3.85297	-1.08037	-3.816	14	<.001	.002

Paired Samples Effect Sizes						
			Standardizer ^a	Point Estimate	95% Confidence Interval	
					Lower	Upper
Pair 1	Spring Weight – Autumn Weight	Cohen's d	2.50333	-.985	-1.596	-.352
		Hedges' correction	2.64822	-.931	-1.508	-.333
a. The denominator used in estimating the effect sizes.						
Cohen's d uses the sample standard deviation of the mean difference.						
Hedges' correction uses the sample standard deviation of the mean difference, plus a correction factor.						

1.6 Discussion and Evaluation

The mean weight is higher in Autumn than in Spring, with birds weighing on average 2.47g more. The standard deviation is similar; 2.24g in Spring and 2.19g in Autumn which indicates proportionate variation in weight across different seasons. The t-value (-3.816) shows that the difference in weight is statistically significant. With the p-value being less than 0.05 (0.002), the null hypothesis is rejected and therefore concludes there is a difference in the robins' weight between the seasons. This can be attributed to the robins needing extra weight to survive a harsh winter where there is less food available (Lövei et al., 1986).

Despite the t-test showing a significant result, some limitations of the data need consideration. With the sample size being 15, this can reduce the validity of the findings, as the sample size is relatively small. Additional factors beyond season change could explain the difference in weight such as migration patterns or food availability in studied area. To conclude, if this study was to be repeated, a larger sample size and increased consideration of differing variables would be highly advised.

2 ANOVA

2.1 Sample Data

Site A	Site B	Site C
38	22	14
47	19	26
39	8	11
25	23	18
42	31	5

2.2 Choice of Statistical Test

For comparing more than two groups, a t-test is insufficient as it can only compare two groups at a time. An ANOVA allows to see whether a mean is significantly different in the groups tested. As opposed to running multiple t-tests would lead to an increase in false positives.

2.3 Assumptions

To maintain the validity of an ANOVA test, certain assumptions must be met. Firstly, the data points must be independent of each other. In this dataset, that means the quadrats were selected randomly, with the measurements from each quadrat not influencing the others. Secondly, the dependent variable must be presented in the same unit e.g. percentages, weight and height. In this case, the dependent variable was percentage cover.

2.4 Calculations

*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor

	Site	Percentage_Cover	Site_num	var
1	A	38.00	1.00	
2	A	47.00	1.00	
3	A	39.00	1.00	
4	A	25.00	1.00	
5	A	42.00	1.00	
6	B	22.00	2.00	
7	B	19.00	2.00	
8	B	8.00	2.00	
9	B	23.00	2.00	
10	B	31.00	2.00	
11	C	14.00	3.00	
12	C	26.00	3.00	
13	C	11.00	3.00	
14	C	18.00	3.00	
15	C	5.00	3.00	
16				
17				

SPSS Data Editor interface showing variable definitions:

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Site	String	8	0	Site of Study	{1, A}...	None	8	Left	Nominal	Input
2	Percentage_Cover	Numeric	8	2		None	None	12	Right	Scale	Input
3	Site_num	Numeric	8	2		None	None	10	Right	Nominal	Input
4											
5											
6											
7											
8											
9											
10											

2.5 Outcomes

ANOVA					
Percentage Cover					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1484.933	2	742.467	11.267	.002
Within Groups	790.800	12	65.900		
Total	2275.733	14			

ANOVA Effect Sizes ^a				
		Point Estimate	95% Confidence Interval	
			Lower	Upper
Percentage Cover	Eta-squared	.653	.174	.782
	Epsilon-squared	.595	.036	.746
	Omega-squared Fixed-effect	.578	.034	.732
	Omega-squared Random-effect	.406	.017	.578

a. Eta-squared and Epsilon-squared are estimated based on the fixed-effect model.

2.6 Discussion and Evaluation

The ANOVA test is statistically significant ($p = 0.002$), meaning there is a significant difference between the three group means. Furthermore, the F-value (11.267) infers the two group means do not differ by chance. Secondly, the ANOVA effect sizes present the degree of variance between the dependent and independent variable. With the dependent variable being Molinia plant growth and the independent variable being the different groups. The Eta-Squared is 0.653, which is regarded as a large size effect, as anything over 0.14 is considered large in ANOVA. The Epsilon-Squared was calculated at 0.595, which points to a strong effect but is a more unbiased estimate of effect size. Next is the Omega-Squared, both the fixed and random effects. The fixed effect is a similar interpretation to Eta-Squared but adjusts for small samples sizes. The fixed effect figure was 0.578, equating to 57.8% variance which is a more conservative effect size. Whereas the random effect was 0.406 (40.6%), based on the assumption that sample groups were randomly selected from a larger population.

To conclude, the ANOVA test has confirmed that there's a statistically significant difference in Molinia cover between sites, however a further post-hoc analysis is needed to determine which site has the highest cover and is therefore most suitable for the butterfly.

3 Chi-squared

3.1 Sample Data

	Farmers	Residents	Total
Gabion Baskets to slow water	23	13	36
Planting more trees to absorb water	15	37	52
Bunding to form a wall	28	26	54
Total	66	76	142

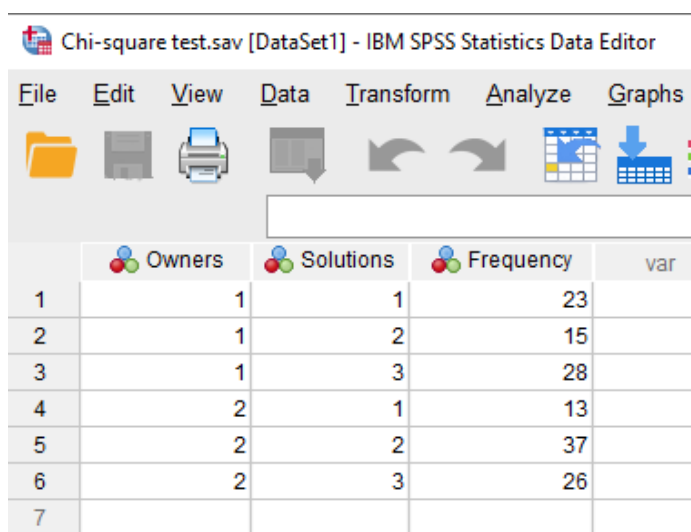
3.2 Choice of Statistical Test

The Chi-squared test was chosen because it is best for analysing categorical data. Furthermore, two independent groups are being compared, farmers and residents. The aim is to determine if there two variables are related. A Chi-squared test is the best statistical test to achieve this aim.

3.3 Assumptions

Like previous statistical tests, assumptions must be checked before calculating. Data must be categorical, which it is as the variables are group types and flood solution preferences. Next, responses must come from different individuals, which also fits as farmers and residents were sampled separately. Lastly, before calculation, expected cell frequency should be ≥ 5 to certify validity of the test.

3.4 Calculations



	Owners	Solutions	Frequency	var
1	1	1	23	
2	1	2	15	
3	1	3	28	
4	2	1	13	
5	2	2	37	
6	2	3	26	
7				

File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help											
											
	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Owners	Numeric	8	0		{1, Farmer}...	None	12	Right	Nominal	Input
2	Solutions	Numeric	8	0		{1, Gabion ...	None	11	Right	Nominal	Input
3	Frequency	Numeric	8	0		None	None	13	Right	Nominal	Input
4											

3.5 Outcomes

Case Processing Summary						
	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Owners * Solutions	142	100.0%	0	0.0%	142	100.0%

Owners * Solutions Crosstabulation						
			Solutions			Total
			Gabion Baskets	Planting more trees	Forming a wall	
Owners	Farmer	Expected Count	16.7	24.2	25.1	66.0
		Residual	6.3	-9.2	2.9	
	Resident	Expected Count	19.3	27.8	28.9	76.0
		Residual	-6.3	9.2	-2.9	
Total		Expected Count	36.0	52.0	54.0	142.0

Chi-Square Tests			
	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	11.512 ^a	2	.003
Likelihood Ratio	11.791	2	.003
Linear-by-Linear Association	.516	1	.473
N of Valid Cases	142		
a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 16.73.			

3.6 Discussion and Evaluation

Since $p < 0.05$, we can reject the null hypothesis and conclude that there's an association between owner and flood solution preference. What the results show is that farmers favoured Gabion Baskets more than expected, with a +6.3 residual. On the other hand, residents preferred planting more trees as a flood solution a lot more than expected (+9.2 residual). Small deviations from the expected values were found for the owners' preference for bunding (forming a wall). Both groups had a relatively similar preference for bunding. The p value (0.003) indicates that occupation does influence flood solution preference.

This test reveals that farmers and residents have differing preferences when it comes to flood defence. Residents may prefer planting more trees to absorb water due to the additional benefits that planting trees has, such as increased biodiversity and improved mental health for residents. Meanwhile, farmers tend to prefer gabion baskets as they're a more immediate form of flood defence, as they take less time to construct. Research shows that integrating green infrastructure into flood strategies can be both effective and socially beneficial (Raymond et al., 2017).

4 Linear Regression

4.1 Sample Data

Respondent number	Knowledge score (Out of 50)	Calcium intake (mg/day)
1	10	450
2	42	1050
3	38	900
4	15	525
5	22	710
6	32	854
7	40	800
8	14	493
9	26	730
10	32	894
11	38	940
12	25	733
13	48	985
14	28	763
15	22	583
16	45	850
17	18	798
18	24	754
19	30	805
20	43	1085

4.2 Choice of Statistical Test

A simple linear regression will be used as there's one independent variable: Calcium intake (mg/day) and one dependent variable: Knowledge score (out of 50). The goal is to see if calcium intake correlates with changes in knowledge scores.

4.3 Assumptions

Before running the test, certain assumptions need to be tested.

- **Linearity:** The relationship between test scores and calcium intake should be linear, this can be checked with a scatter plot.
- **Independence:** Each student's calcium intake should be independent of others.
- **Equal variance:** The spread of residuals should be roughly equal (check residual plots).
- **Normal distribution of residuals:** The residuals should be normally distributed.

4.4 Calculations

*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze

35 : Calcium_Intake

	Knowledge_Score	Calcium_Intake	var	va
1	10.00	450.00		
2	42.00	1050.00		
3	38.00	900.00		
4	15.00	525.00		
5	22.00	710.00		
6	32.00	854.00		
7	40.00	800.00		
8	14.00	493.00		
9	26.00	730.00		
10	32.00	894.00		
11	38.00	940.00		
12	25.00	733.00		
13	48.00	985.00		
14	28.00	763.00		
15	22.00	583.00		
16	45.00	850.00		
17	18.00	798.00		
18	24.00	754.00		
19	30.00	805.00		
20	43.00	1085.00		
21				

File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help

Search application

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Knowledge_Score	Numeric	8	2		None	None	11	Right	Ordinal	Input
2	Calcium_Intake	Numeric	8	2		None	None	9	Right	Scale	Input
3											

4.5 Outcomes

Correlations			
		Knowledge Score	Calcium Intake
Knowledge Score	Pearson Correlation	1	.882**
	Sig. (2-tailed)		<.001
	N	20	20
Calcium Intake	Pearson Correlation	.882**	1
	Sig. (2-tailed)	<.001	
	N	20	20

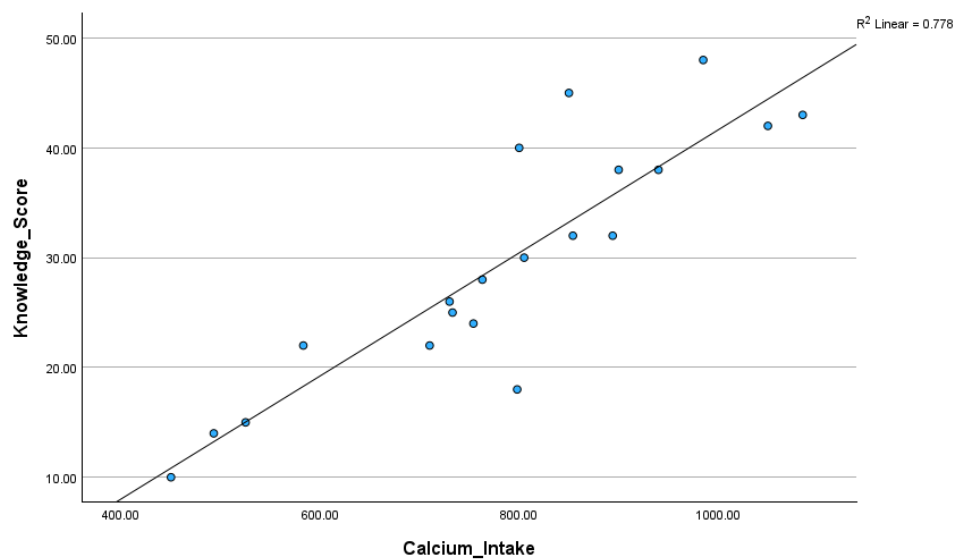
** . Correlation is significant at the 0.01 level (2-tailed).

Regression

Variables Entered/Removed ^a			
Model	Variables Entered	Variables Removed	Method
1	Knowledge Score ^b	.	Enter
a. Dependent Variable: Calcium Intake			
b. All requested variables entered.			

Coefficients ^a								
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	373.743	55.067		6.787	<.001	258.051	489.435
	Knowledge Score	13.897	1.748	.882	7.951	<.001	10.225	17.569
a. Dependent Variable: Calcium Intake								

Graph



4.6 Discussion and Evaluation

The correlation table shows there's a very strong correlation between Calcium Intake and Knowledge Score. This is shown by The Pearson correlation coefficient ($r = 0.882$, $p < 0.001$). With p being < 0.01 the correlation is significant as opposed to being up to chance.

The constant (373.743) is the estimated calcium intake when the Knowledge Score is 0. The B coefficient for Knowledge Score is 13.897, means that for every point increase in Knowledge Score, Calcium Intake increases by 13.897 mg/day.

The scatter plot shows a strong positive linear relationship between calcium intake and knowledge score. Also, the line of best fit follows the data points closely, which suggests a strong correlation. The R^2 value is 0.778, which means that 77.8% of the variation in knowledge score can be explained by calcium intake.

This observed relationship is not that surprising considering that studies have found higher calcium intake to be associated with better cognitive performance (Kravchenko et al., 2024). Calcium's benefits extend beyond healthy teeth and bones, extending also to cognitive health.

5 References

- Lövei, G, Scebba, S, Minichiello, F & Milone, M 1986, 'Seasonal activity, wing shape, weights and fat reserve variation of Robins (*Erithacus rubecula*) in southern Italy', *Supplementi alle Ricerche di Biologia della Selvaggina*, vol. 10, pp. 229-242.
- Kravchenko, G., Stephenson, S. S., Gutowska, A., Klimek, K., Chrząstek, Z., Pięłowska, M., Kostka, T., & Sołtysik, B. K. (2024). The Concurrent Association of Magnesium and Calcium Deficiencies with Cognitive Function in Older Hospitalized Adults. *Nutrients* , 16(21). <https://doi.org/10.3390/nu16213756>
- Raymond, C. M., Frantzeskaki, N., Kabisch, N., Berry, P., Breil, M., Nita, M. R., Geneletti, D., & Calfapietra, C. (2017). A framework for assessing and implementing the co-benefits of nature-based solutions in urban areas. *Environmental Science and Policy*, 77, 15–24. <https://doi.org/10.1016/j.envsci.2017.07.008>