

## Library Impact Data Project: Lincoln data

### Summary of results

- Book loans have small differences in usage between higher and lower level degrees.
- Athens logins indicate a difference in usage between higher and lower level degrees.
- Library entries do not indicate any relationship.

### Introduction

The following report is intended to show the calculation and analysis process for the Library Impact Data Project partner data, with the aim to prove or disprove the project's hypothesis that there is a statistically significant correlation between library use and degree attainment. Due to the nature of the data format, a formal correlation cannot be proven; we can, however, measure whether there are differences between the amount of usage per degree result, and from that where differences lie and their extent.

Several tests have been performed on each partner's data, and on the combined data:

- Kolmogorov-Smirnov test to indicate distribution of the data, and thus whether the following test is an appropriate measure of difference
- Kruskal-Wallis test: measures whether there are differences between degree results, but importantly doesn't tell us where differences lie
- Graphical representation of the data as a boxplot, used to suggest where we may wish to investigate differences or similarities
- Mann-Whitney test to measure whether there are differences between specific degree results
- The effect size: a simple manual calculation using the Mann-Whitney results to measure the size of the effect, in this case the extent of difference. The effect size is calculated by dividing the Mann-Whitney's Z statistic by the square root of the sample size involved *for those degrees to be compared*.

Using these tests, we start with a second, null, hypothesis stating that there is no difference in library usage between degree results. Should the tests demonstrate differences, the null hypothesis can be discarded, and we can use the results as an indicator of a relationship.

### Data analysis

NB – entry figures are only for 2 years as opposed to the 3 years for all other data sets for Lincoln.

Before analysis can commence, data must be tested for normality at a significance level of 5% (0.05) using the Kolmogorov-Smirnov test. If a non-normal distribution is found, the data can be analysed using the Kruskal-Wallis test for differences.

**Tests of Normality**

Final degree		Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Total loans	First class degree	.135	241	.000	.889	241	.000
	Upper second	.141	1290	.000	.847	1290	.000
	Lower second	.171	982	.000	.821	982	.000
	Third class degree	.244	84	.000	.623	84	.000
	Ordinary degree	.243	44	.000	.656	44	.000
Total entries	First class degree	.105	241	.000	.920	241	.000
	Upper second	.081	1290	.000	.918	1290	.000
	Lower second	.081	982	.000	.935	982	.000
	Third class degree	.110	84	.014	.883	84	.000
	Ordinary degree	.154	44	.010	.875	44	.000
Total Athens logins	First class degree	.206	241	.000	.775	241	.000
	Upper second	.174	1290	.000	.815	1290	.000
	Lower second	.206	982	.000	.746	982	.000
	Third class degree	.227	84	.000	.706	84	.000
	Ordinary degree	.249	44	.000	.692	44	.000

a. Lilliefors Significance Correction

Data is non-normal, but the significance values for entries appear to indicate an anomaly in distribution. Boxplots may indicate distribution differences in the post hoc analysis.

**Kruskal-Wallis (KW) test**

As the samples are large, a Monte Carlo estimate is used, repeatedly running simulated samples following the same distribution of the actual data to ensure the analysis is accurate. For 4 degrees of freedom (i.e. the amount of variance between the number of groups analysed), the chi-square value must be at least 13.28 for a significance value of  $p < 0.01$  and 9.49 at  $p < 0.05$ .

### Kruskal-Wallis test for book issues

Ranks			
	Final degree	N	Mean Rank
Total loans	First class degree	241	1420.85
	Upper second	1290	1460.17
	Lower second	982	1179.13
	Third class degree	84	870.38
	Ordinary degree	44	720.36
	Total	2641	

Test Statistics <sup>b,c</sup>			
			Total loans
Chi-square			137.812
df			4
Asymp. Sig.			.000
Monte Carlo Sig.	Sig.		.000 <sup>a</sup>
	99% Confidence Interval	Lower Bound	.000
		Upper Bound	.000

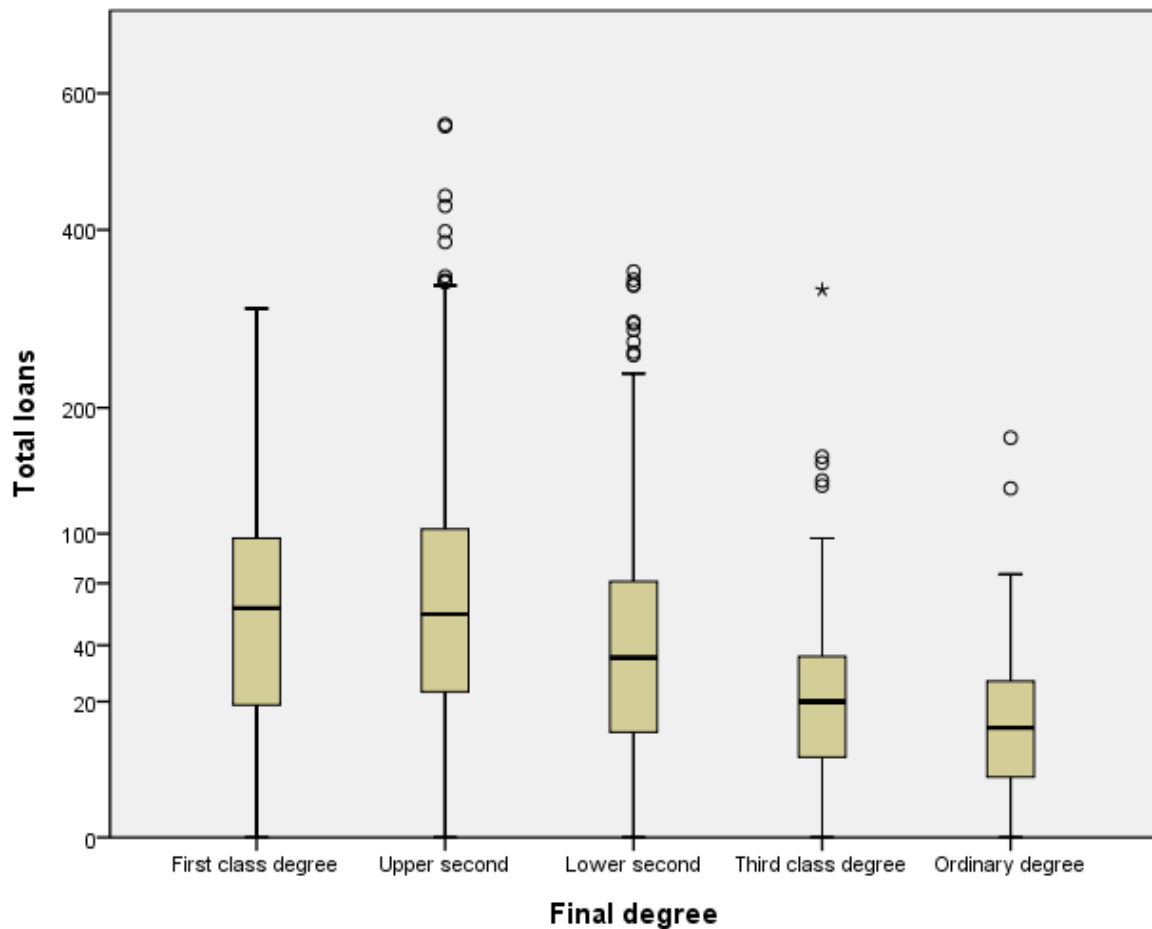
a. Based on 10000 sampled tables with starting seed 1314643744.

b. Kruskal Wallis Test

c. Grouping Variable: Final degree

The KW test results in a very high chi-square value with high significance, and thus the data can be considered to have differences between degree results for book usage.

## Post hoc analysis of book issues



Boxplots are used to identify potential relationship comparisons for significance. Stars indicate extreme outlying data, circles represent outliers in general (i.e. usage that falls outside of the main set of data), the box indicates the central 50% of data with the thick black line indicating the median value. The bar at the top of the vertical line (whisker) is the top of the next 25% of data. Post-hoc analysis is conducted using Mann-Whitney tests, with a limit on the number of tests conducted to ensure the required significance value (0.05 divided by the number of tests, known as the Bonferroni correction) doesn't increase to unachievable levels. Three tests will thus require a significance of  $p < 0.0167$  to ensure they are valid. Data is then tested for effect size in significant results i.e. the amount of difference between groups, reporting the Mann-Whitney U value and  $r$  (the effect size: a negative value indicates a decreasing change) as standard in statistical reports.

There appears to be very little difference between higher degrees, with variation beginning to appear at the lower second class level and below. However, upper second class degree results show a number of students lying outside of the main 75% bracket of usage, which while a small number, would potentially skew the median value. Mann-Whitney tests should minimise the risk of skewing. Analysis of difference will be performed on:

- First and lower second class
- Lower second and third class
- Third class and ordinary degree.

**First and lower second class**

**Ranks**

	Final degree	N	Mean Rank	Sum of Ranks
Total loans	First class degree	241	699.77	168645.50
	Lower second	982	590.46	579830.50
	Total	1223		

**Test Statistics<sup>b</sup>**

			Total loans
Mann-Whitney U			97177.500
Wilcoxon W			579830.500
Z			-4.307
Asymp. Sig. (2-tailed)			.000
Monte Carlo Sig. (2-tailed)	Sig.		.000 <sup>a</sup>
	99% Confidence Interval	Lower Bound	.000
		Upper Bound	.000
Monte Carlo Sig. (1-tailed)	Sig.		.000 <sup>a</sup>
	99% Confidence Interval	Lower Bound	.000
		Upper Bound	.000

a. Based on 10000 sampled tables with starting seed 1502173562.

b. Grouping Variable: Final degree

**Lower second and third class**

**Ranks**

	Final degree	N	Mean Rank	Sum of Ranks
Total loans	Lower second	982	543.62	533834.00
	Third class degree	84	415.20	34877.00
	Total	1066		

**Test Statistics<sup>b</sup>**

			Total loans
Mann-Whitney U			31307.000
Wilcoxon W			34877.000
Z			-3.671
Asymp. Sig. (2-tailed)			.000
Monte Carlo Sig. (2-tailed)	Sig.		.000 <sup>a</sup>
	99% Confidence Interval	Lower Bound	.000
		Upper Bound	.001
Monte Carlo Sig. (1-tailed)	Sig.		.000 <sup>a</sup>
	99% Confidence Interval	Lower Bound	.000
		Upper Bound	.001

a. Based on 10000 sampled tables with starting seed 743671174.

b. Grouping Variable: Final degree

**Third class and ordinary degree**

**Ranks**

		N	Mean Rank	Sum of Ranks
Total loans	Third class degree	84	67.85	5699.00
	Ordinary degree	44	58.11	2557.00
	Total	128		

**Test Statistics<sup>b</sup>**

			Total loans
Mann-Whitney U			1567.000
Wilcoxon W			2557.000
Z			-1.414
Asymp. Sig. (2-tailed)			.157
Monte Carlo Sig. (2-tailed)	Sig.		.162 <sup>a</sup>
	99% Confidence Interval	Lower Bound	.152
		Upper Bound	.171
Monte Carlo Sig. (1-tailed)	Sig.		.081 <sup>a</sup>
	99% Confidence Interval	Lower Bound	.074
		Upper Bound	.088

a. Based on 10000 sampled tables with starting seed 957002199.

b. Grouping Variable: Final degree

While there is no significant difference between third class and ordinary degree results, there are differences between higher degrees:

- First and lower second class have a small difference (U=97177.5,  $r=-0.123$ )
- Lower second and third class degrees have a small difference (U=31307,  $r=-0.112$ )

The data thus suggests a relationship between borrowing and degree result.

### Kruskal-Wallis test for Athens logins

Ranks			
Final degree		N	Mean Rank
Total Athens logins	First class degree	241	1424.23
	Upper second	1290	1446.51
	Lower second	982	1181.50
	Third class degree	84	1079.80
	Ordinary degree	44	649.75
	Total	2641	

Test Statistics <sup>b,c</sup>			
			Total athens logins
Chi-square			114.820
df			4
Asymp. Sig.			.000
Monte Carlo Sig.	Sig.		.000 <sup>a</sup>
	99% Confidence Interval	Lower Bound	.000
		Upper Bound	.000

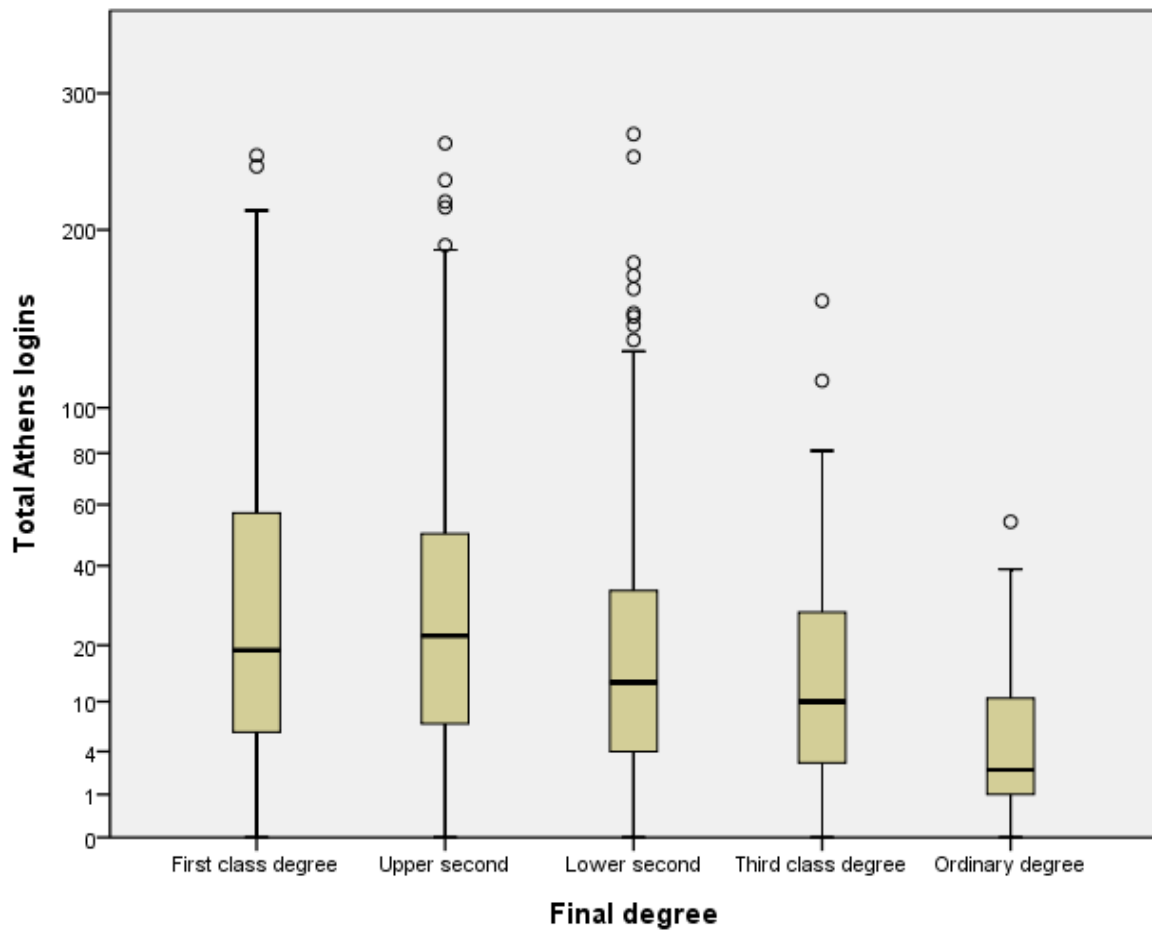
a. Based on 10000 sampled tables with starting seed 334431365.

b. Kruskal Wallis Test

c. Grouping Variable: Final degree

The KW test indicates a strong difference with high chi-square and significance values.

## Post hoc analysis of Athens logins



Athens logins appear to indicate some difference at lower degree result usage, but again the median value for upper second class degree results may be skewed slightly by the values outside the 75% level. Analysis will be performed on:

- First and upper second class
- First and lower second class
- Lower second class and ordinary degree.

## Mann-Whitney tests for Athens logins

### First and upper second class

		Ranks		
	Final degree	N	Mean Rank	Sum of Ranks
Total Athens logins	First class degree	241	760.17	183200.50
	Upper second	1290	767.09	989545.50
	Total	1531		

**Test Statistics<sup>b</sup>**

			Total Athens logins
Mann-Whitney U			154039.500
Wilcoxon W			183200.500
Z			-.223
Asymp. Sig. (2-tailed)			.823
Monte Carlo Sig. (2-tailed)	Sig.		.823 <sup>a</sup>
	99% Confidence Interval	Lower Bound	.813
		Upper Bound	.833
Monte Carlo Sig. (1-tailed)	Sig.		.408 <sup>a</sup>
	99% Confidence Interval	Lower Bound	.395
		Upper Bound	.420

a. Based on 10000 sampled tables with starting seed 1335104164.

b. Grouping Variable: Final degree

**First and lower second class**

**Ranks**

		Final degree	N	Mean Rank	Sum of Ranks
Total Athens logins	First class degree		241	698.72	168390.50
	Lower second		982	590.72	580085.50
	Total		1223		

**Test Statistics<sup>b</sup>**

			Total Athens logins
Mann-Whitney U			97432.500
Wilcoxon W			580085.500
Z			-4.256
Asymp. Sig. (2-tailed)			.000
Monte Carlo Sig. (2-tailed)	Sig.		.000 <sup>a</sup>
	99% Confidence Interval	Lower Bound	.000
		Upper Bound	.000
Monte Carlo Sig. (1-tailed)	Sig.		.000 <sup>a</sup>
	99% Confidence Interval	Lower Bound	.000
		Upper Bound	.000

a. Based on 10000 sampled tables with starting seed 329836257.

b. Grouping Variable: Final degree

**Lower second class and ordinary degree**

**Ranks**

	Final degree	N	Mean Rank	Sum of Ranks
Total Athens logins	Lower second	982	522.86	513452.00
	Ordinary degree	44	304.52	13399.00
	Total	1026		

**Test Statistics<sup>b</sup>**

			Total Athens logins	
Mann-Whitney U			12409.000	
Wilcoxon W			13399.000	
Z			-4.785	
Asymp. Sig. (2-tailed)			.000	
Monte Carlo Sig. (2-tailed)	Sig.		.000 <sup>a</sup>	
		99% Confidence Interval	Lower Bound	.000
			Upper Bound	.000
Monte Carlo Sig. (1-tailed)	Sig.		.000 <sup>a</sup>	
		99% Confidence Interval	Lower Bound	.000
			Upper Bound	.000

a. Based on 10000 sampled tables with starting seed 1535910591.

b. Grouping Variable: Final degree

There is no significant difference between first and upper second class degree usage, but other degrees do show a difference:

- First class and lower second class have a small difference (U=97432.5,  $r=-0.122$ )
- Lower second class and ordinary degrees have a small difference (U=12409,  $r=-0.149$ ).

We can therefore reasonably assume that there is some relationship between Athens usage and degree result.

**Kruskal-Wallis test on library entries**

**Ranks**

	Final degree	N	Mean Rank
Total entries	First class degree	241	1198.10
	Upper second	1290	1368.72
	Lower second	982	1301.75
	Third class degree	84	1363.49
	Ordinary degree	44	943.61
	Total	2641	

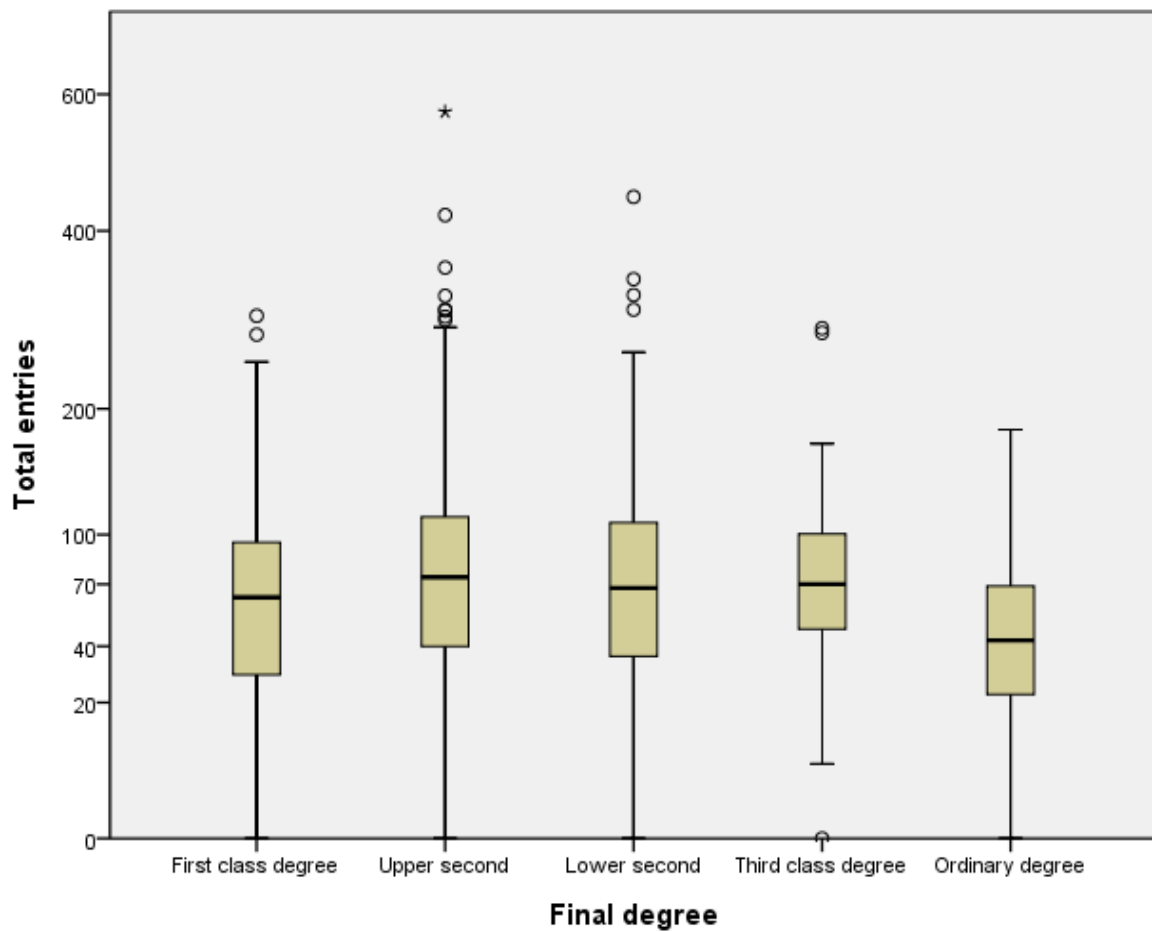
**Test Statistics<sup>b,c</sup>**

			Total entries
Chi-square			22.983
df			4
Asymp. Sig.			.000
Monte Carlo Sig.	Sig.		.000 <sup>a</sup>
	99% Confidence Interval	Lower Bound	.000
		Upper Bound	.001

- a. Based on 10000 sampled tables with starting seed 624387341.
- b. Kruskal Wallis Test
- c. Grouping Variable: Final degree

While the KW test shows a difference at a significant level, the chi-square value is comparatively small to other data analysed for Lincoln. Further analysis will still use the boxplot as guidance for comparison, if appropriate.

**Post hoc analysis of entries**



The boxplot indicates that there is little difference between library entries and final degree. It does appear that ordinary degree entries have skewed the KW test to report a difference. On removal of the ordinary degree from the KW test, the following result is achieved:

Ranks			
	Final degree	N	Mean Rank
Total entries	First class degree	241	1172.02
	Upper second	1290	1339.67
	Lower second	982	1273.80
	Third class degree	84	1333.35
	Total	2597	

Test Statistics <sup>b,c</sup>			
			Total entries
Chi-square			11.996
df			3
Asymp. Sig.			.007
Monte Carlo Sig.	Sig.		.007 <sup>a</sup>
	99% Confidence Interval	Lower Bound	.005
		Upper Bound	.009

a. Based on 10000 sampled tables with starting seed 112562564.

b. Kruskal Wallis Test

c. Grouping Variable: Final degree

This result still indicates some difference between library entries and result, with a chi-square statistic of significance at  $p < 0.01$ . Some comparison will be run, but it is not expected to produce significant results. Analysis will be run on:

- First and upper second class
- First and lower second class
- First and third class.

#### First and upper second class

Ranks				
	Final degree	N	Mean Rank	Sum of Ranks
Total entries	First class degree	241	683.27	164667.50
	Upper second	1290	781.46	1008078.50
	Total	1531		

**Test Statistics<sup>b</sup>**

			Total entries
Mann-Whitney U			135506.500
Wilcoxon W			164667.500
Z			-3.165
Asymp. Sig. (2-tailed)			.002
Monte Carlo Sig. (2-tailed)	Sig.		.002 <sup>a</sup>
	99% Confidence Interval	Lower Bound	.001
		Upper Bound	.002
Monte Carlo Sig. (1-tailed)	Sig.		.001 <sup>a</sup>
	99% Confidence Interval	Lower Bound	.000
		Upper Bound	.002

a. Based on 10000 sampled tables with starting seed 221623949.

b. Grouping Variable: Final degree

**First and lower second class**

**Ranks**

		Final degree	N	Mean Rank	Sum of Ranks
Total entries	First class degree		241	573.29	138162.00
	Lower second		982	621.50	610314.00
	Total		1223		

**Test Statistics<sup>b</sup>**

			Total entries
Mann-Whitney U			109001.000
Wilcoxon W			138162.000
Z			-1.899
Asymp. Sig. (2-tailed)			.058
Monte Carlo Sig. (2-tailed)	Sig.		.056 <sup>a</sup>
	99% Confidence Interval	Lower Bound	.050
		Upper Bound	.062
Monte Carlo Sig. (1-tailed)	Sig.		.027 <sup>a</sup>
	99% Confidence Interval	Lower Bound	.023
		Upper Bound	.031

a. Based on 10000 sampled tables with starting seed 303130861.

b. Grouping Variable: Final degree

## First and third class

		Ranks		
Final degree		N	Mean Rank	Sum of Ranks
Total entries	First class degree	241	157.47	37949.50
	Third class degree	84	178.88	15025.50
Total		325		

			Test Statistics <sup>b</sup>
			Total entries
Mann-Whitney U			8788.500
Wilcoxon W			37949.500
Z			-1.799
Asymp. Sig. (2-tailed)			.072
Monte Carlo Sig. (2-tailed)	Sig.		.080 <sup>a</sup>
	99% Confidence Interval	Lower Bound	.073
		Upper Bound	.087
Monte Carlo Sig. (1-tailed)	Sig.		.040 <sup>a</sup>
	99% Confidence Interval	Lower Bound	.035
		Upper Bound	.045

a. Based on 10000 sampled tables with starting seed 92208573.

b. Grouping Variable: Final degree

As expected, there is no significant level of difference between degree attainment and library entries, in spite of what the KW test suggested. There is a very small difference between first and upper second class results, but it is below the usual accepted base allowance (0.08, where small is considered to start at 0.1), and so given other data analysed in the process of the project, it is recommended that this difference be ignored, particularly given the anomalies showing in the Kolmogorov-Smirnov test. It is difficult to know whether there would have been any difference in results had there been a full 3 year set of data, but as other institutions do not appear to have a link between library entries and attainment, for this case we will assume the same.

### Course averages

Some specific courses suggest a relationship between library use and attainment, such as in the sciences and some of the arts subjects (including textiles, design and drama courses). Some anomalies appear due to the number of students on a particular course or achieving a specific degree result, which may explain why differences on a university-wide level show comparatively small differences.

### Conclusion

There is some level of relationship between library usage and degree result, although not connected to library entries. Further investigation at course level is recommended.