USABILITY CA

ANALYSIS & REPORT

Jonathan Dempsey – N00112462
Multimedia Systems/Web Engineering – DL824
February 16, 2016

Table of Contents

1 Introduction

The purpose of the project is to answer the research question: **Is there a significant difference in the mean time taken to obtain the 3 star hotel list between the two websites?** An analysis will be carried out on the time taken (in seconds to one decimal place) to source an online list of three star hotels for two cities in Ireland using two different official tourism websites. The two websites to be recorded are as follows:

1. www.discovernorthernireland.com : This website is the official website for the Northern Ireland Tourist Board (NITB). The participants must reach the 3 star hotel list for Belfast City *(fig 1)* from the websites home page.

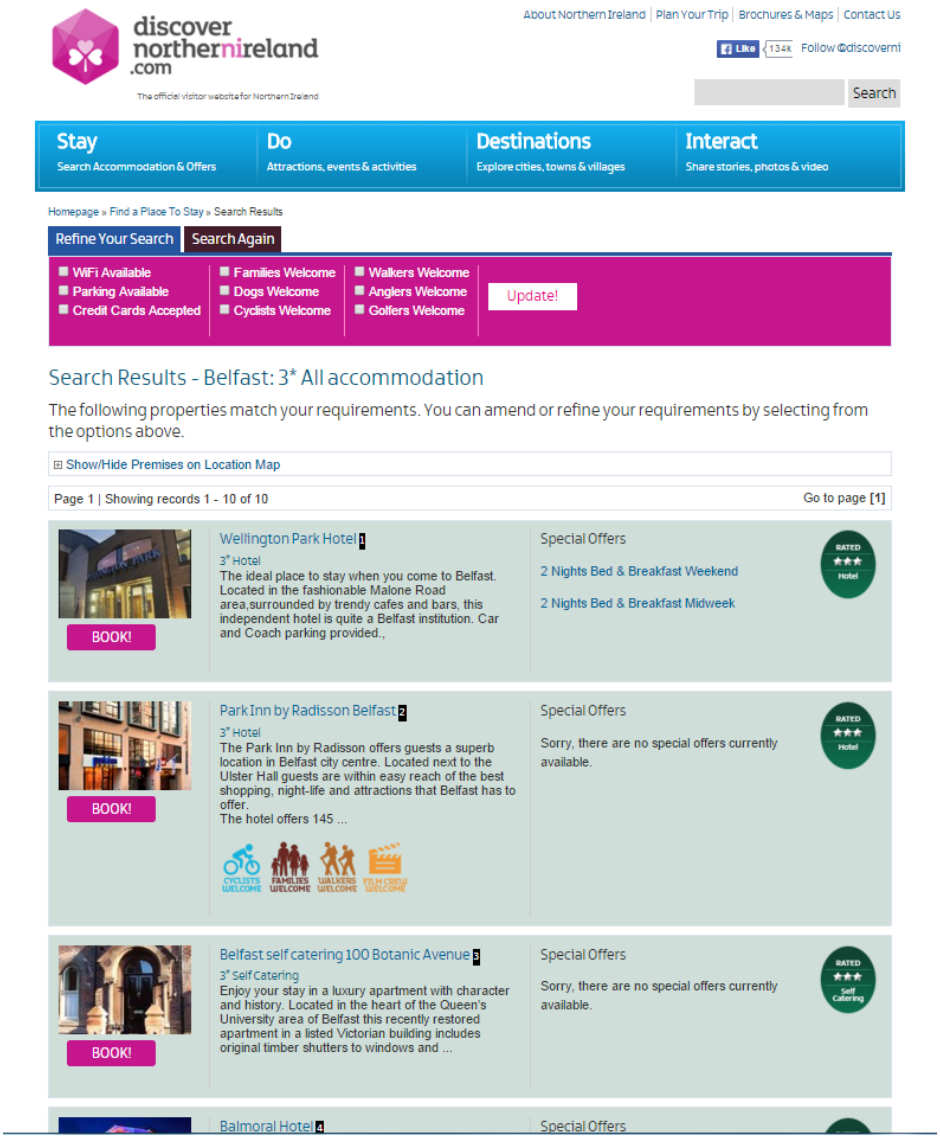

Figure 1: Discover Northern Ireland 3 star hotel list for Belfast City

2. www.discoverireland.ie : This website is the official website for Fáilte Ireland, the national tourism development agency. The participants must reach the 3 star hotel list for Cork City *(fig 2)* from the websites home page.

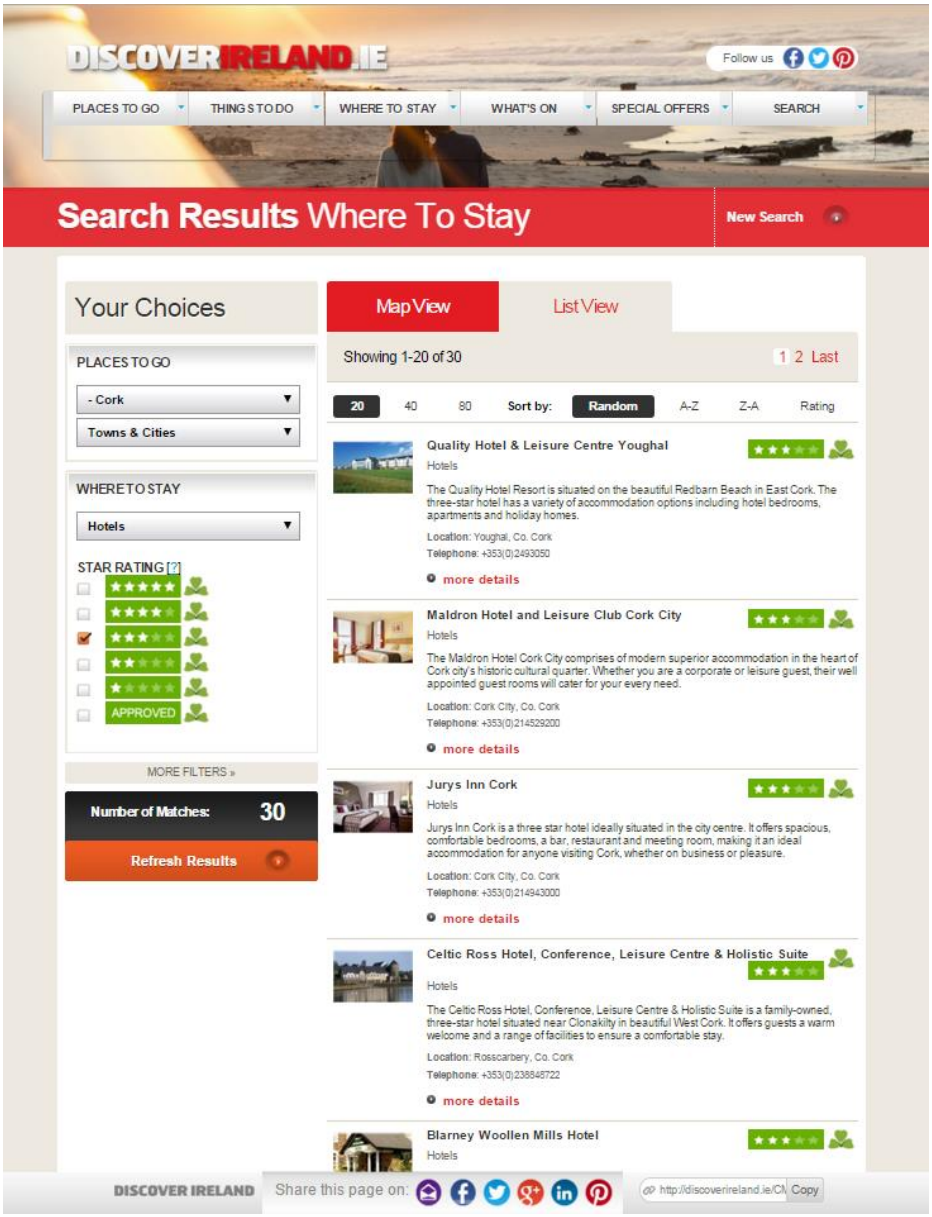

Figure 2: Discover Ireland 3 star hotel list for Cork City

2 Data Collection

The data collection phase will involve each participant recording the time taken to display the list of 3 star hotels for each location. The website to be searched first will be random for each participant. The time will be recorded on a stopwatch when the home page is

fully loaded and will stop when the 3 star hotel list is displayed on the screen. The results will be shown on an excel sheet *(fig 3)* and will record time taken to one decimal place and note the order in which each site was searched.

Name	Website (first tested)	Time (seconds)	Website (second tested)	Time (seconds)
Jonathan Browne	discovernorthernireland.com	41.13	discoverireland.ie	51.81
Aaron Power	discovernorthernireland.com	51.81	discoverireland.ie	28.58
David Collins	discovernorthernireland.com	47.09	discoverireland.ie	36.8
Adam Geraghty	discovernorthernireland.com	39.5	discoverireland.ie	22.2
Deri ohUiginn	discovernorthernireland.com	39.92	discoverireland.ie	102.1
Jonathan Dempse	discovernorthernireland.com	150	discoverireland.ie	27.2
Owen Sela	discoverireland.ie	130.3	discovernorthernireland.com	43.6
paul conroy	discovernorthernireland.com	15.63	discoverireland.ie	31.07
Joseph O'Reilly	discoverireland.ie	105.4	discovernorthernireland.com	46.78
Peter Redmond	discovernorthernireland.com	52.45	discoverireland.ie	96.6
Darragh Kenny	discoverireland.ie	42.8	discovernorthernireland.com	76.4
Keith Reilly	discovernorthernireland.com	35.3	discoverireland.ie	48.9
Clara Halpin	discovernorthernireland.com	29.9	discoverireland.ie	33.8
Dominic Rowan	discovernorthernireland.com	55.1	discoverireland.ie	36.04
Michael Kerr	discoverireland.ie	32.29	discovernorthernireland.com	123.19
James O'Donoghu	discovernorthernireland.com	66.8	discoverireland.ie	40.6
jade ernest	discovernorthernireland.com	13.57	discoverireland.ie	67

Figure 3: Results sheet

Based on the data gathered, the majority of the participants searched www.discovernorthernireland.com first. This may have to be taken into consideration towards the end of the experiment as participants may be more comfortable with searching for hotel lists, through the effects of learning, potentially resulting in faster search times for the second searched, www.discoverireland.ie.

Name	Website 1	Time (seconds)	Website 2	Time (seconds)
Jonathan Browne	discovernorthernireland.com	41.1	discoverireland.ie	51.8
Aaron Power	discovernorthernireland.com	51.8	discoverireland.ie	28.6
David Collins	discovernorthernireland.com	47.1	discoverireland.ie	36.8
Adam Geraghty	discovernorthernireland.com	39.5	discoverireland.ie	22.2
Deri ohUiginn	discovernorthernireland.com	39.9	discoverireland.ie	102.1
Jonathan Dempse	discovernorthernireland.com	150	discoverireland.ie	27.2
Owen Sela	discovernorthernireland.com	43.6	discoverireland.ie	130.3
paul conroy	discovernorthernireland.com	15.6	discoverireland.ie	31.1
Joseph O'Reilly	discovernorthernireland.com	46.8	discoverireland.ie	105.4
Peter Redmond	discovernorthernireland.com	52.5	discoverireland.ie	96.6
Darragh Kenny	discovernorthernireland.com	76.4	discoverireland.ie	42.8
Keith Reilly	discovernorthernireland.com	35.3	discoverireland.ie	48.9
Clara Halpin	discovernorthernireland.com	29.9	discoverireland.ie	33.8
Dominic Rowan	discovernorthernireland.com	55.1	discoverireland.ie	36
Michael Kerr	discovernorthernireland.com	123.2	discoverireland.ie	32.3
James O'Donoghu	discovernorthernireland.com	66.8	discoverireland.ie	40.6
jade ernest	discovernorthernireland.com	13.6	discoverireland.ie	67

Figure 4: Results sheet normalized

In figure 4, the results are normalized by website with the time taken reduced to one decimal place.

3 Statistical Analysis

As the experiment has the same individual testing two different websites, the test to be carried out will be a paired test, otherwise known as a same participant design. Compared to unpaired tests, paired tests have a lower variability because it is the same participant testing both the websites. However, paired tests contribute a more accurate estimate of the differences of the usability of the websites. A disadvantage to the paired test is the potential impact of learning effects that could skew the data in favor of the second website. Since there are 17 participants, the first tested website should ideally be split equally between the two websites, this creates a counter balance that minimizes the unfair effects of learning from the second tested site. Unfortunately, as mentioned earlier, the effects of learning will have to be considered as www.discoverireland.ie is only tested first 4 out of 17 times.

3.1 HYPOTHESES

The hypothesis testing requires a null hypothesis and an alternative hypothesis. They are stated in a way that they are mutually exclusive, meaning if one is true, then the other must be false. The null hypotheses *(H0)* is a statement that an intervention has had no impact and/or that the observation result is purely from chance. The alternative hypothesis *(H1)* is a statement that there was some sort of impact and/or that the observation result is affected by a non-random cause.

$$H_0: p = 0.5$$
$$H_a: p <> 0.5$$

Figure 5: Example hypothesis

For example, given the expressed hypothesis shown in figure 5, if a coin was flipped 100 times and the results were 80 heads and 20 tails, the null hypothesis would be rejected. It would be determined that the coin used in the experiment was likely not fair or balanced, as an 80/20 split rejects the null hypothesis.

In the case with the website testing the hypothesis is as follows:

- Null hypothesis:

H0: $\mu_1 - \mu_2 = 0$ for all 17 observations

- Alternative hypothesis:

H1: $\mu_1 \neq \mu_2$

The test will be two-tailed as there is no prior information regarding usability and efficiency of the two websites.

3.2 CALCULATE THE TEST STATISTIC

Name	Website 1	Website 2	Difference	Square of Difference
Jonathan Browne	41.1	51.8	-10.7	114.5
Aaron Power	51.8	28.6	23.2	538.2
David Collins	47.1	36.8	10.3	106.1
Adam Geraghty	39.5	22.2	17.3	299.3
Deri ohUiginn	39.9	102.1	-62.2	3868.8
Jonathan Dempsey	150	27.2	122.8	15079.8
Owen Sela	43.6	130.3	-86.7	7516.9
paul conroy	15.6	31.1	-15.5	240.3
Joseph O'Reilly	46.8	105.4	-58.6	3434.0
Peter Redmond	52.5	96.6	-44.1	1944.8
Darragh Kenny	76.4	42.8	33.6	1129.0
Keith Reilly	35.3	48.9	-13.6	185.0
Clara Halpin	29.9	33.8	-3.9	15.2
Dominic Rowan	55.1	36	19.1	364.8
Michael Kerr	123.2	32.3	90.9	8262.8
James O'Donoghue	66.8	40.6	26.2	686.4
jade ernest	13.6	67	-53.4	2851.6
SUM			**-5.3**	**46637.5**

Figure 6: Calculated difference table

From figure 6, the test statistic is calculated which is a value calculated from the sample data and is used to determine the strength of evidence in support of the null hypothesis. When working out the difference, it is important to keep the calculations consistent. In this case, the difference is calculated as the time of www.discovernorthernireland.com (Website 1) minus the time of www.discoverireland.ie (Website 2). It is also important to keep the direction of the calculations consistent. For example, in the first row the calculated difference is -10.7, this should not be changed to be +10.7 to keep consistency.

The calculations are as follows:

- The sample size:

n = 17

- Degrees of Freedom:

df = 17 – 1

df = 16

- Sum of differences:

Σdi = -5.3

- Sum of squared differences:

Σdi^2 = 46637.5

- Mean of difference:

$$\bar{x} = \frac{\sum x_i}{n}$$

Figure 7: Mean equation (x replaced with d)

$\bar{d}$ = -5.3 / 17

$\bar{d}$ = 0.3

- Standard deviation of differences

$$s_d = \sqrt{\frac{\sum d_i^2 - n\bar{d}^2}{n-1}}$$

Figure 8: Standard deviation equation

s^2d = 46637.5 – 17 × (0.3) ^2 / 17 - 1

s^2d = 46632.4 / 16

s^2d = 2914.5

∴ sd = 54.0

- Test statistic:

$$t = \frac{\bar{d}}{s_d / \sqrt{n}}$$

Figure 9: Test statistic equation

t = 0.3 / 54 / √17

t = 0.3 / 54 / 4.1

t = 0.3 / 13.2

t = 0.0227

3.3 MAKING THE DECISION

To help make the decision, a T-distribution table is used, as shown in figure 10. The test is two-tailed with the decision risk of 5%. The decision risk (α) allots half to testing the statistical significance in one direction and half to testing the other direction, giving 2.5% to each tail in the distribution of the test statistic.

Using the T table with 16 degrees of freedom and the cumulative probabilities 97.5%, the critical values calculated are ± 2.120. A critical value is a value on the test distribution that is compared to the test statistic to determine whether or not to reject the null hypothesis. As the calculated **t** is 0.0227 which appears outside the critical regions, we do not reject the null hypothesis and can determine that the difference between the two websites are **not statistically significant.**

It should be noted that, as mentioned earlier, the effects of learning may have influenced the results of the experiment as well as the small sample size of the experiment. This reduces the tests ability to detect differences giving increased chance to the possibility of a type 2 error (a false negative (β)) to occur. A false negative would indicate that there was no difference between the websites when there actually a difference.

df	75%	80%	85%	90%	95%	97.5%	99%	99.5%	99.75%	99.9%	99.95%
1	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850

Figure 10: T-distribution table

4 Visualized Results

In figure 11, DataDesk is used to accurately visualize the results of the experiment by using a paired t test. To display the results, the user is required to enter the t value (0.0227), the sample size (17) and the amount of simulations to run (in this case, 10'000 simulations). The critical regions are the areas of the bell curve that are beyond the critical values of ±2.120. The table shows the percentage of the tests that are greater than the calculated t-test which in this case is 48.6% and is shown in white. The results given are not within either of the critical values so the null hypothesis is accepted and are consistent with the results gathered from the T-tables.

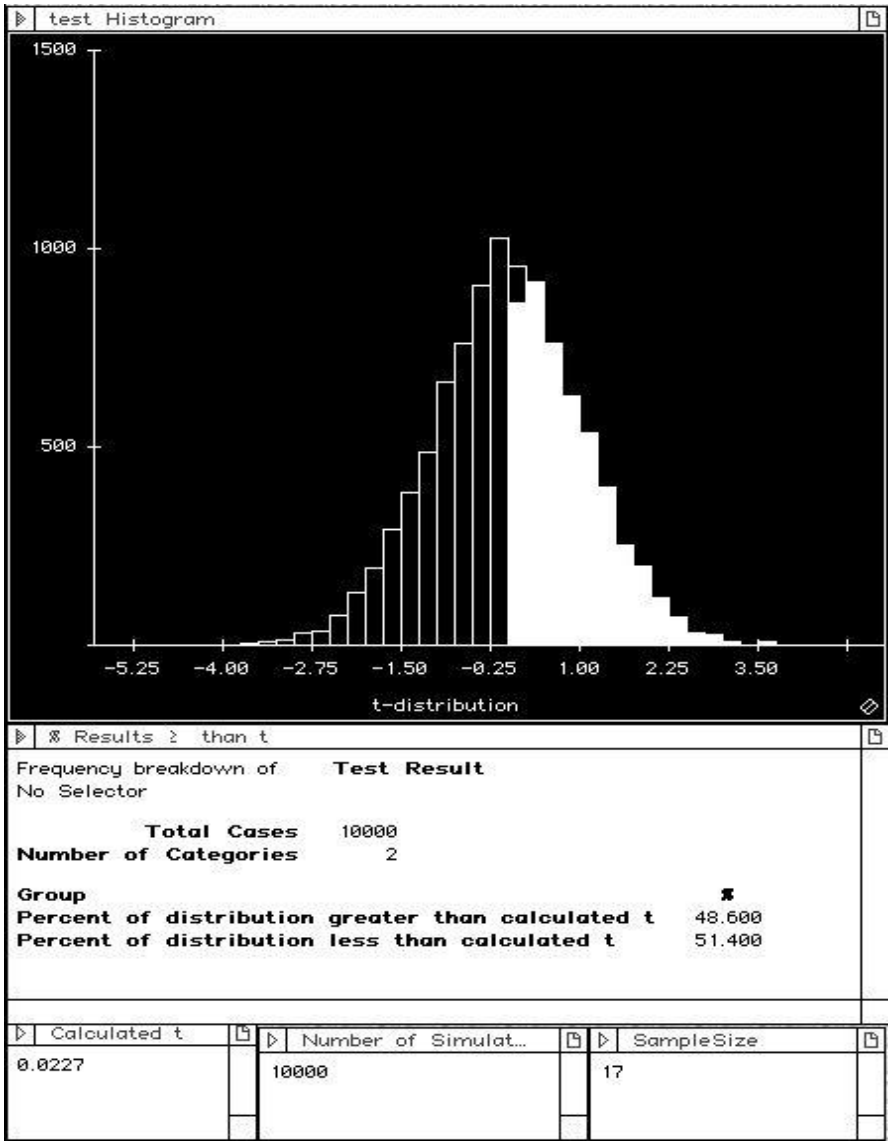

Figure 11: DataDesk paired t-test

5 References

[1]. Connolly, C. “Statistical Inference for Usability Experiments”, BSc in Multimedia Systems/Web Engineering, Year 4 Course Handbook.

[2]. Stattrek.com (2016). “Hypothesis Test: Paired Means”. Retrieved: Feb 11th, 2016 from http://stattrek.com/hypothesis-test/paired-means.aspx?Tutorial=AP

[3]. Support.minitab.com (2016). “Why Should I Use a Paired T-Test?” Retrieved: Feb 11th, 2016 from http://support.minitab.com/en-us/minitab/17/topic-library/basic-statistics-and-graphs/hypothesis-tests/tests-of-means/why-use-paired-t/