|  |
| --- |
| **Assignment on A/B Testing of a Retail Website Page** |
| Course Name – Digital Marketing  Course Code – M411 |
| Submitted By –  Vaskor Roy  Roll - 24  BBA 27th Batch  Institute of Business Administration  University of Dhaka  Submitted By –  Swarup Saha  Lecturer  Institute of Business Administration  University of Dhaka |
| Date of Submission: September 5, 2022 |

Table of Contents

[The Problem 1](#_Toc113058757)

[The Hypotheses and KPIs to be Measured 1](#_Toc113058758)

[Statistical Testing Methodology 2](#_Toc113058759)

[The Results 2](#_Toc113058760)

[Final Recommendation 3](#_Toc113058761)

[References 3](#_Toc113058762)

[Appendix 4](#_Toc113058763)

[Appendix A – SPSS Test Output 4](#_Toc113058764)

[Appendix B – Microsoft Excel 2019 Test Output 7](#_Toc113058765)

# The Problem

The dataset given is a part of an A/B testing online experiment conducted during 83 days by the digital marketing team of hypothetical an online retailer. The online retailer is headquartered in Dhaka, Bangladesh and has gathered accolades as one of the fastest growing online businesses.

The purpose of the A/B testing was to examine whether or not adding a haptic (i.e., touch and feel) description for a product (i.e., Suit) will help improve consumer responses. More specifically, the retailer was interested in comparing e-commerce metrics for two versions of the product page for Men Suit (i.e., the product page that contained a touch and feel description of the suit vs. the product page that did not provide a touch and feel description for that). The objective, thus, is to determine whether there are any differences between the two versions of the website and decide accordingly on which version of the website to adopt for usage.

# The Hypotheses and KPIs to be Measured

In order to determine the better version of the website, statistical hypothesis testing on the dataset of the website for both the versions – the website with haptic feedback and the website without haptic feedback – was conducted. As such, the following hypotheses were chosen –

H0 – There is no statistically significant difference between the efficacy of the versions for any given variable.

H1 – There is statistically significant difference between the efficacy of the versions for any given variable.

Since the hypotheses presented here only predict in the difference between the efficacy of the different versions of the website; in order to provide recommendations on which version of the website should be adopted by the retailer, another set of hypotheses were chosen –

H0 – The version with the haptic feedback performed equally or worse than the version without the haptic feedback (i.e., mean of haptic <= mean of without haptic).

H1 – The version with the haptic feedback performed better than the version without the haptic feedback (i.e., mean of haptic > mean of without haptic).

These set of hypotheses are the primary determiners behind the final recommendation of which version of the website should be adopted by the retailer.

The following key performance indicators were also chosen based on the dataset for the purposes of statistical testing -

1. Sessions
2. Revenues
3. Transactions
4. Average Order Value = Revenues / Transactions
5. Revenue Per Session = Revenues / Sessions
6. Pages Browsed Per Session
7. Average Session Duration
8. Ecommerce Conversion Rate

# Statistical Testing Methodology

In order to conduct statistical tests on the dataset procured for the purpose of A/B testing, two types of statistical tests were chosen –

1. Independent Samples T-Test Assuming Equal Variances
2. Mann – Whitney U Test

The purposes of the two types of tests chosen for the dataset were primarily the comparison between the categories of sample present in the dataset (i.e., data for both the website with haptic feedback and the website without haptic feedback). The independent samples t-test was conducted to compare the mean values between the two groups. The Mann – Whitney u-test was conducted to compare the distribution of the data of the two groups of samples.

For each test, the null hypothesis, H0, as stated previously, was that there was no prevalent statistically significant difference between the efficacy of the two groups of samples for any chosen variable. The alternate hypothesis, H1, accordingly, was that there were statistically significant differences between the efficacy of the two groups of samples for any chosen variable.

In the independent samples t-test, the efficacy of the two versions of the website was compared through the mean values of the versions for any chosen variable or metric. In the Mann-Whitney u-test, the efficacy was measured by the distribution of the values of the versions for any chosen metric, which also represented the consistency of the performance.

After the first set of hypotheses were tested out, the second set of hypotheses were tested out specifically for the comparison of the means of the two groups of samples. In this scenario, the null hypothesis was that the version with haptic feedback achieved means equal or less than the means for the version without the haptic feedback. The alternative hypothesis, on the other hand, was that the version with the haptic feedback achieved means greater than the means for the version without haptic feedback.

# The Results

For the first set of hypotheses, two separate statistical tests were conducted to compare the two groups of samples, i.e., two different versions of the website. The independent samples t-test showed that there were statistically significant differences between the two versions of the website when the means of average session duration, average order value and transactions of the two versions were compared. For the rest of the metrics chosen, there were no statistically significant evidence found to reject the null hypothesis that there was no difference between the two versions of the website.

The Mann Whitney u-test showed that statistically significant differences between the two versions of the website were perceived when the consistency of the performance was compared in the following variables – Transactions conducted per day, ecommerce conversion rate per day, Average session duration per day and pages browsed per session. For the rest of the variables, no significant difference was observed.

After the first set of hypotheses were tested out, the second set of hypotheses were tested through a repetition of the independent samples t-test in Microsoft Excel 2019. The second t test showed that the version with haptic feedback had greater means (i.e., performed better) than the version without haptic feedback in the following metrics – transactions, average session durations, ecommerce conversion rate. In all other metrics, there was no significant evidence to reject the null hypothesis that the version with haptic feedback performed equal or worse than the version without the haptic feedback.

# Final Recommendation

Based on the statistical analysis, it can be seen that there are significant differences between the performance of the two versions of the website when the following metrics were compared -

1. Average Session Duration
2. Transactions
3. Average Order Value
4. Pages Browsed per Session
5. Ecommerce Conversion Rate

Among these metrics, for the ecommerce conversion rate metric, while there was no statistically significant difference found between the versions of the website in the first t-test, the second t-test showed that the version with the haptic feedback did perform marginally better than the version without the haptic feedback in that metric. On the contrary, for the average order value and the pages browsed per session metric, while the first t-test showed that there were significant differences between the two versions, the second t-test showed that this difference was in the opposite direction of the alternate hypothesis of the second set, thus showing that the version with haptic feedback may have performed equally or even worse than the version without the haptic feedback. However, in the average session duration and the transactions metrics, the two t-tests showed that the version with the haptic feedback performed better than the version without the haptic feedback.

Therefore, based on these observations, I believe that the retailer should adopt the version with the haptic feedback in order to increase amounts of transactions conducted and engagement in the website.

# References

1. Wright, D. (2019, July 19). *Tail of the Test: Interpreting Excel Data Analysis t-test output - Dawn Wright, Ph.D.* Dawn Wright, Ph.D. https://www.drdawnwright.com/tail-of-the-test-interpreting-excel-data-analysis-t-test-output/

# Appendix

## Appendix A – SPSS Test Output

1. Independent Sample T Test
   1. Test Results

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **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** | |
| Lower | Upper |
| Sessions | Equal variances assumed | 0.103 | 0.749 | 1.212 | 164 | 0.227 | 12.892 | 10.640 | -8.117 | 33.900 |
| Equal variances not assumed |  |  | 1.212 | 163.961 | 0.227 | 12.892 | 10.640 | -8.117 | 33.900 |
| Pages / Session | Equal variances assumed | 0.096 | 0.757 | 1.637 | 164 | 0.104 | 0.219 | 0.134 | -0.045 | 0.483 |
| Equal variances not assumed |  |  | 1.637 | 163.818 | 0.104 | 0.219 | 0.134 | -0.045 | 0.483 |
| Avg. Session Duration | Equal variances assumed | 0.526 | 0.469 | 2.550 | 164 | 0.012 | 26.726 | 10.482 | 6.029 | 47.424 |
| Equal variances not assumed |  |  | 2.550 | 160.525 | 0.012 | 26.726 | 10.482 | 6.025 | 47.428 |
| Ecommerce Conversion Rate | Equal variances assumed | 1.893 | 0.171 | 1.948 | 164 | 0.053 | 0.599% | 0.308% | -0.008% | 1.206% |
| Equal variances not assumed |  |  | 1.948 | 161.110 | 0.053 | 0.599% | 0.308% | -0.008% | 1.207% |
| Transactions | Equal variances assumed | 1.443 | 0.231 | 2.737 | 164 | 0.007 | 3.904 | 1.426 | 1.088 | 6.720 |
| Equal variances not assumed |  |  | 2.737 | 161.891 | 0.007 | 3.904 | 1.426 | 1.087 | 6.720 |
| Revenue (BDT) | Equal variances assumed | 0.211 | 0.646 | 1.267 | 164 | 0.207 | 23955.933 | 18900.460 | -13363.679 | 61275.544 |
| Equal variances not assumed |  |  | 1.267 | 163.943 | 0.207 | 23955.933 | 18900.460 | -13363.775 | 61275.640 |
| Revenue/Session | Equal variances assumed | 4.502 | 0.035 | -0.032 | 164 | 0.974 | -3.348 | 103.773 | -208.251 | 201.556 |
| Equal variances not assumed |  |  | -0.032 | 152.291 | 0.974 | -3.348 | 103.773 | -208.368 | 201.673 |
| Avg. Order Value | Equal variances assumed | 17.944 | 0.000 | -2.320 | 164 | 0.022 | -1692.195 | 729.504 | -3132.625 | -251.764 |
| Equal variances not assumed |  |  | -2.320 | 122.192 | 0.022 | -1692.195 | 729.504 | -3136.297 | -248.092 |

1. Mann-Whitney U Test
   1. Test Summary

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Hypothesis Test Summary** | | | | |
|  | **Null Hypothesis** | **Test** | **Sig.** | **Decision** |
| 1 | The distribution of Sessions is the same across categories of Presence of Haptic Feedback. | Independent-Samples Mann-Whitney U Test | 0.187 | Retain the null hypothesis. |
| 2 | The distribution of Pages / Session is the same across categories of Presence of Haptic Feedback. | Independent-Samples Mann-Whitney U Test | 0.029 | Reject the null hypothesis. |
| 3 | The distribution of Avg. Session Duration is the same across categories of Presence of Haptic Feedback. | Independent-Samples Mann-Whitney U Test | 0.005 | Reject the null hypothesis. |
| 4 | The distribution of Ecommerce Conversion Rate is the same across categories of Presence of Haptic Feedback. | Independent-Samples Mann-Whitney U Test | 0.027 | Reject the null hypothesis. |
| 5 | The distribution of Transactions is the same across categories of Presence of Haptic Feedback. | Independent-Samples Mann-Whitney U Test | 0.008 | Reject the null hypothesis. |
| 6 | The distribution of Revenue (BDT) is the same across categories of Presence of Haptic Feedback. | Independent-Samples Mann-Whitney U Test | 0.119 | Retain the null hypothesis. |
| 7 | The distribution of Revenue/Session is the same across categories of Presence of Haptic Feedback. | Independent-Samples Mann-Whitney U Test | 0.591 | Retain the null hypothesis. |
| 8 | The distribution of Avg. Order Value is the same across categories of Presence of Haptic Feedback. | Independent-Samples Mann-Whitney U Test | 0.345 | Retain the null hypothesis. |
| Asymptotic significances are displayed. The significance level is .050. | | | | |

* 1. Test Results

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Total N | Mann-Whitney U | Wilcoxon W | Test Statistic | Standard Error | Standardized Test Statistic | Asymptotic Sig. (2-sided test) |
| Sessions | 166 | 3853.5 | 7339.5 | 3853.5 | 309.619 | 1.321 | 0.187 |
| Pages Per Session | 166 | 4120 | 7606 | 4120 | 309.632 | 2.182 | 0.029 |
| Avg. Session Duration | 166 | 4311 | 7797 | 4311 | 309.632 | 2.798 | 0.005 |
| Ecommerce Conversion Rate | 166 | 4130 | 7616 | 4130 | 309.629 | 2.214 | 0.027 |
| Transactions | 166 | 4265 | 7751 | 4265 | 308.51 | 2.66 | 0.008 |
| Revenue (BDT) | 166 | 3927 | 7413 | 3927 | 309.632 | 1.558 | 0.119 |
| Revenue Per Session | 166 | 3611 | 7097 | 3611 | 309.632 | 0.538 | 0.591 |
| Avg. Order Value | 166 | 3152 | 6638 | 3152 | 309.632 | -0.945 | 0.345 |

## Appendix B – Microsoft Excel 2019 Test Output

1. Independent Samples T-Test
   1. Revenue

|  |  |  |
| --- | --- | --- |
|  | **With Haptic** | **Without Haptic** |
| Mean | 229514.7349 | 205558.8024 |
| Variance | 14549052308 | 15100820652 |
| Observations | 83 | 83 |
| Pooled Variance | 14824936480 |  |
| Hypothesized Mean Difference | 0 |  |
| df | 164 |  |
| t Stat | 1.267478815 |  |
| P(T<=t) one-tail | 0.103390428 |  |
| t Critical one-tail | 1.654197929 |  |
| P(T<=t) two-tail | 0.206780857 |  |
| t Critical two-tail | 1.974534576 |  |

* 1. Average Order Value

|  |  |  |
| --- | --- | --- |
|  | **With Haptic** | **Without Haptic** |
| Mean | 11399.69685 | 13091.89137 |
| Variance | 9166773.758 | 35003794.94 |
| Observations | 83 | 83 |
| Pooled Variance | 22085284.35 |  |
| Hypothesized Mean Difference | 0 |  |
| df | 164 |  |
| t Stat | -2.319651987 |  |
| P(T<=t) one-tail | 0.010796252 |  |
| t Critical one-tail | 1.654197929 |  |
| P(T<=t) two-tail | 0.021592504 |  |
| t Critical two-tail | 1.974534576 |  |

* 1. Revenue Per Session

|  |  |  |
| --- | --- | --- |
|  | **With Haptic** | **Without Haptic** |
| Mean | 1211.143495 | 1214.491058 |
| Variance | 322987.6443 | 570826.2464 |
| Observations | 83 | 83 |
| Pooled Variance | 446906.9454 |  |
| Hypothesized Mean Difference | 0 |  |
| df | 164 |  |
| t Stat | -0.032258509 |  |
| P(T<=t) one-tail | 0.487152561 |  |
| t Critical one-tail | 1.654197929 |  |
| P(T<=t) two-tail | 0.974305123 |  |
| t Critical two-tail | 1.974534576 |  |

* 1. Sessions

|  |  |  |
| --- | --- | --- |
|  | **With Haptic** | **Without Haptic** |
| Mean | 193.8433735 | 180.9518072 |
| Variance | 4625.206876 | 4770.363503 |
| Observations | 83 | 83 |
| Pooled Variance | 4697.78519 |  |
| Hypothesized Mean Difference | 0 |  |
| df | 164 |  |
| t Stat | 1.211666655 |  |
| P(T<=t) one-tail | 0.113691456 |  |
| t Critical one-tail | 1.654197929 |  |
| P(T<=t) two-tail | 0.227382912 |  |
| t Critical two-tail | 1.974534576 |  |

* 1. Transactions

|  |  |  |
| --- | --- | --- |
|  | **With Haptic** | **Without Haptic** |
| Mean | 20.21686747 | 16.31325301 |
| Variance | 94.04995592 | 74.77872465 |
| Observations | 83 | 83 |
| Pooled Variance | 84.41434029 |  |
| Hypothesized Mean Difference | 0 |  |
| df | 164 |  |
| t Stat | 2.737050754 |  |
| P(T<=t) one-tail | 0.003441661 |  |
| t Critical one-tail | 1.654197929 |  |
| P(T<=t) two-tail | 0.006883322 |  |
| t Critical two-tail | 1.974534576 |  |

* 1. Average Session Duration

|  |  |  |
| --- | --- | --- |
|  | **With Haptic** | **Without Haptic** |
| Mean | 216.5119625 | 189.7854752 |
| Variance | 5230.903123 | 3889.069511 |
| Observations | 83 | 83 |
| Pooled Variance | 4559.986317 |  |
| Hypothesized Mean Difference | 0 |  |
| df | 164 |  |
| t Stat | 2.549671104 |  |
| P(T<=t) one-tail | 0.005849442 |  |
| t Critical one-tail | 1.654197929 |  |
| P(T<=t) two-tail | 0.011698883 |  |
| t Critical two-tail | 1.974534576 |  |

* 1. Ecommerce Conversion Rate

|  |  |  |
| --- | --- | --- |
|  | **With Haptic** | **Without Haptic** |
| Mean | 0.052690189 | 0.046698908 |
| Variance | 0.000340026 | 0.000445199 |
| Observations | 83 | 83 |
| Pooled Variance | 0.000392612 |  |
| Hypothesized Mean Difference | 0 |  |
| df | 164 |  |
| t Stat | 1.947877875 |  |
| P(T<=t) one-tail | 0.026568484 |  |
| t Critical one-tail | 1.654197929 |  |
| P(T<=t) two-tail | 0.053136968 |  |
| t Critical two-tail | 1.974534576 |  |

* 1. Pages Per Session

|  |  |  |
| --- | --- | --- |
|  | **With Haptic** | **Without Haptic** |
| Mean | 3.425356689 | 3.206435084 |
| Variance | 0.767381192 | 0.717930077 |
| Observations | 83 | 83 |
| Pooled Variance | 0.742655634 |  |
| Hypothesized Mean Difference | 0 |  |
| df | 164 |  |
| t Stat | 1.636511007 |  |
| P(T<=t) one-tail | 0.051825134 |  |
| t Critical one-tail | 1.654197929 |  |
| P(T<=t) two-tail | 0.103650268 |  |
| t Critical two-tail | 1.974534576 |  |