# Assignment on Hypothesis Testing

Problem Statement:

You explored the Mashable article popularity dataset in EDA module. You arrived at certain conclusions. You had seen that the articles published on Social Media channel saw on average more number of shares than other channels. You also saw that the average number of shares for articles published over weekend was more than the articles published over weekdays. Now you want to use the concepts in hypothesis testing to confirm these insights and reach at actionable outputs.

Data preparation and cleaning

* Perform outlier treatment over the “Shares” attribute using the techniques learnt in EDA module.
* Figure out a way to extract Day of publishing for each article.

Initial Insight:

Please fill out this table giving out Average number of shares for each article, belonging to various categories.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Weekday | Weekend | Total |
| Social Media | 2448.957 | 2661.786 | 2478.458 |
| Others | 1881.41 | 2248.871 | 1929.369 |
| Total | 1914.374 | 2274.475 |  |

Questions:

Note: For each given problem, you need to mention the Null hypothesis, alternate hypothesis, the critical values, test statistic, p-values ant the final decision to reject or fail to reject the null hypothesis.

1. Confirm at 1% significance level if the average number of shares for each article differ significantly for articles published over weekdays and weekend.

*Null hypothesis: Z1 equal to Z2*

*Alternate hypothesis: Z1 not equal to Z2*

*Critical values: -2.576 < t value < 2.576*

*Test statistic: t test = -18.20548*

*p-value: p-value < 2.2e-16*

*Final decision:* ***Since t value of -18 is way to less than the t value of -2.576 we can reject the Null Hypothesis, So we can say at 99% confidence that the average number of shares for each article differ significantly for articles published over weekdays and weekends***

1. Confirm at 1% significance level if the average number of shares for each article published over weekend differ significantly for articles on Social media channel and other channels.

*Null hypothesis: X12 equal to X22*

*Alternate hypothesis: X12 not equal to X22*

*Critical values: -2.576 < t value < 2.576*

*Test statistic: t test = 5.502432*

*p-value: 7.045e-08*

*Final decision:* **S*ince t value of 5.5024 which is more than the t value of 2.576 we can reject the Null Hypothesis,* *So we can say at 99% confidence that the average number of shares for each article published over weekend differ significantly for articles on Social media channel and other channels***

1. Confirm at 1% significance level if the average number of shares for each article published over weekdays differ significantly for articles on Social media channel and other channels.

*Null hypothesis: X11 equal to X21*

*Alternate hypothesis: X11 not equal to X21*

*Critical values: -2.576 < t value < 2.576*

*Test statistic: t test = 18.85896*

*p-value: p-value < 2.2e-16*

*Final decision:* **S*ince t value of 18.859 which is more than the t value of 2.576 we can reject the Null Hypothesis, so we can say at 99% confidence that the average number of shares for each article published over weekdays differ significantly for articles on Social media channel and other channels***

1. Confirm at 5% significance level if the average number of shares for Social Media articles published over weekdays and weekends differ significantly from each other.

*Null hypothesis: X11 equal to X12*

*Alternate hypothesis: X11 not equal to X12*

*Critical values: -1.96 < t value < 1.96*

*Test statistic: t test = -2.7202*

*p-value:* *0.006789*

*Final decision:* ***Since t value of -2.7202 which is less than the t value of -1.96 we can reject the Null Hypothesis, so we can say at 95% confidence that the average number of shares for Social Media articles published over weekdays and weekends differ significantly from each other****.*