# 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 | 2150.972 | 2424.713 | 2189.37 |
| Others | 1598.041 | 1936.077 | 1640.895 |
| Total | 1628.914 | 1966.528 | 1671.973 |

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:*  **Average number of shares for articles published on the weekday is same as those published over weekends.**

*Alternate hypothesis:* **Average number of shares for articles published on the weekday is NOT equal to average number of shares for articles published on weekend.**

*Critical values:* **Zc = -2.58 , 2.58**

*Test statistic: t-statistic - -18.181*

*p-value: 2.2e-16*

*Final decision: Reject the NULL hypothesis, the test statistic is much higher than critical value*

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: mean of shares (weekend) for soc\_med is equal to others*

*Alternate hypothesis: mean of shares (weekend) for soc\_med is NOT equal to others*

*Critical values:* **Zc = -2.58 ,2.58**

*Test statistic: t= 6.4013*

*p-value:5.664e-10*

*Final decision:* *Reject the NULL hypothesis, the test statistic is much higher than critical value*

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: mean of shares (weekday) for soc\_med is same as others*

*Alternate hypothesis: mean of shares (weekday) for soc\_med is NOT same as others*

*Critical values:* **Zc = -2.58 , 2.58**

*Test statistic: t=18.946*

*p-value:2.26e-16*

*Final decision: Reject the NULL hypothesis, the test statistic is much higher than critical value*

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: mean of shares(weekday) for soc\_med is equal to mean of shares (weekend) soc\_med*

*Alternate hypothesis: mean of shares(weekday) for soc\_med is NOT equal to mean of shares (weekend) soc\_med*

*Critical values: Zc= -1.96, 1.96*

*Test statistic: t=3.4438*

*p-value:0.000642*

*Final decision: Reject the NULL hypothesis, the test statistic is much higher than critical value*