

## Inferential Statistics

Conducting the 1-way ANOVA test using the statsmodels python package we will compare grouped reviews based on Product Type (i.e. Instant Videos, Video Games, Android Apps, etc.) and based on wordcount groups.

Inferential Statistics: (1) Hypothesis for the set of test data of overall Star Ratings against the Product Type of the reviewText:

H0: The average number of overall Star Ratings given for customer Reviews for each Product Type is equal.

Ha: The average number of overall Star Ratings given for customer Reviews for each Product Type is not equal.

After running the ANOVA test, the F-statistic is 526.0893902860769 and the p-value is near 0. Since the p-value is near zero we can reject the null hypothesis and conclude that the average number of overall Star Ratings given for customer Reviews for each product category is not equal. This confirms the findings in the boxplot.

Inferential Statistics: (2) Hypothesis for the other set of test data of overall Star Ratings against the wordcount of the reviewText:

H0: The average number of overall Star Ratings given for customer Reviews for each word\_count category is equal.

Ha: The average number of overall Star Ratings given for customer Reviews for each word\_count category is not equal.

After running the ANOVA test, the F-statistic is 0.07203281229366418 and the p-value is 0.9905100882043847. Since the p-value is not near zero we cannot reject the null hypothesis and conclude that the average number of overall Star Ratings given for customer Reviews for each word\_count category is equal.

Inferential Statistics: The test result suggests the groups don't have the same sample means in this case, since the p-value is significant at a 99% confidence level. To check which groups differ after getting a positive ANOVA result, you can perform a follow up test or "post-hoc test". One post-hoc test is to perform a separate t-test for each pair of groups. You can perform a t-test between all pairs by running each pair through the stats.ttest\_ind(). The results below show since the p-values are not below 0.05 we cannot reject the null hypothesis and conclude that the pairs' mean values are equal.

Test Results for F-1-Way ANOVA-Test:

By Product Type:

F\_onewayResult(statistic=526.0893902860769, pvalue=0.0)

## Inferential Statistics

By WordCount:

F\_onewayResult(statistic=0.07203281229366418, pvalue=0.9905100882043847)

Inferential Statistics: This above test compares all possible pairs and we can use it to precisely identify pairs where the difference between two means is greater than the expected standard error.

For each pair of mean values:

H0: The means are equal.

Ha: The means are not equal.

The result below shows each pairs' mean difference. Since the pair's mean values are not statistically significantly different, then we cannot reject the null hypothesis and conclude that the pairs' mean values are equal. In the table below, the 'reject' column has all False values.

Test Results for T-Test:

0 476
Ttest_indResult(statistic=0.5243915355846621, pvalue=0.6004708786574138)
0 999
Ttest_indResult(statistic=0.18242476143756678, pvalue=0.8555851722209963)
0 3048
Ttest_indResult(statistic=0.1211632110752595, pvalue=0.9038203892635411)
0 4848
Ttest_indResult(statistic=nan, pvalue=nan)
476 999
Ttest_indResult(statistic=-0.08187480888206944, pvalue=0.9348403714648184)
476 3048
Ttest_indResult(statistic=0.03744783452639184, pvalue=0.9701752052910082)
476 4848
Ttest_indResult(statistic=nan, pvalue=nan)
999 3048
Ttest_indResult(statistic=0.06909801595573156, pvalue=0.9456734379918976)
999 4848
Ttest_indResult(statistic=nan, pvalue=nan)
3048 4848
Ttest_indResult(statistic=nan, pvalue=nan)

## Inferential Statistics

Multiple Comparison of Means - Tukey  
HSD, FWER=0.05

group1	group2	meandiff	lower	upper	reject
0	443	1.0	-2.3342	4.3342	False
0	639	1.0	-2.3342	4.3342	False
0	803	-0.75	-4.3513	2.8513	False
0	1284	1.25	-2.8335	5.3335	False
443	639	0.0	-3.3342	3.3342	False
443	803	-1.75	-5.3513	1.8513	False
443	1284	0.25	-3.8335	4.3335	False
639	803	-1.75	-5.3513	1.8513	False
639	1284	0.25	-3.8335	4.3335	False
803	1284	2.0	-2.3044	6.3044	False

