Genre vs Song Durations ANOVA Test

To answer the research question 'Do different genres of songs have significantly different durations?', an excerpt of a collection of data on 1 million popular songs compiled by The Echo Nest and a lab at Columbia University will be used to conduct a one-way ANOVA. With the ANOVA results, an omnibus significance test and multiple corrected pairwise t-tests would be conducted and evaluated with respect to the effect size of the effect data and limitations of the tests performed to come to a conclusion to answer the initial research question. Hence, the dependent variable would be a ratio value of each song's duration while the independent variable would be a nominal value of the genres of the songs. Thus, the null hypothesis (H_0) states that the population mean (μ) of song duration for each individual nominal value (level) of genre is equal to one another while the alternative hypothesis (H_a) states that at least one of the population means of song duration for each genre is different from the rest.

One-Way ANOVA

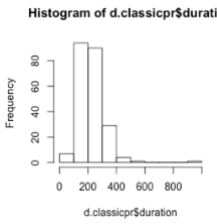
Null Hypothesis (H_0) : All population means of song duration for each genre would be equal
H_0 : $\mu_{classic\ p\&r} = \mu_{classical} = \mu_{d\&e} = \mu_{folk} = \mu_{hiphop} = \mu_{j\&b} = \mu_{metal} = \mu_{pop} = \mu_{punk} = \mu_{s\&r}$

Alternative Hypothesis (H_a) : At least one population mean song duration of a genre is different

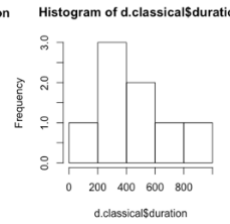
Assumptions for One-Way ANOVA:

- Normality: Tested using histograms for song durations within each genre as shown below

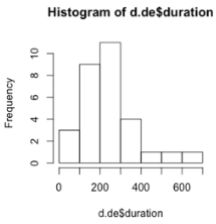
Classic Pop & Rock



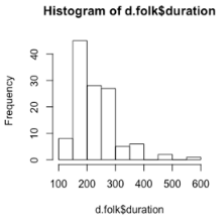
Classical



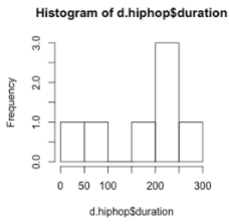
Dance & Electronica



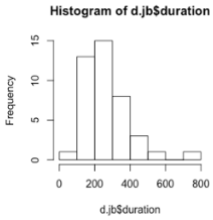
Folk



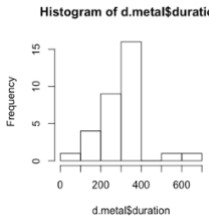
Hip-Hop



Jazz & Blues



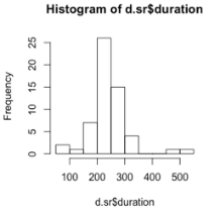
Metal



Pop

Punk

Soul & Reggae



- Homogeneity of Variance: Levene's Test with p-value of 0.0000002246 leads to rejecting the null hypothesis and hence, at least one variance is not equal.
- Independence of Observations: Since each song data recorded is unique to the song, the observations are not dependent on any other variables.

Test Statistic (F): $F = \frac{MS_{between}}{MS_{within}} = \frac{64667}{9254} = 6.988$ p-value: p = 0.00000000143

Degrees of Freedom (df): $df_{between} = 10 - 1 = 9$, $df_{within} = 546 - 9 - 1 = 536$, $df_{total} = 545$

Effect Size: $R^2 = \frac{SS_{between}}{SS_{total}} = \frac{581999}{581999+4960031} = 0.1050154907$

Bonferroni Correction: Pairwise $\alpha = \frac{familywise\ \alpha}{num\ of\ tests} = \frac{0.05}{45} = 0.001111111$

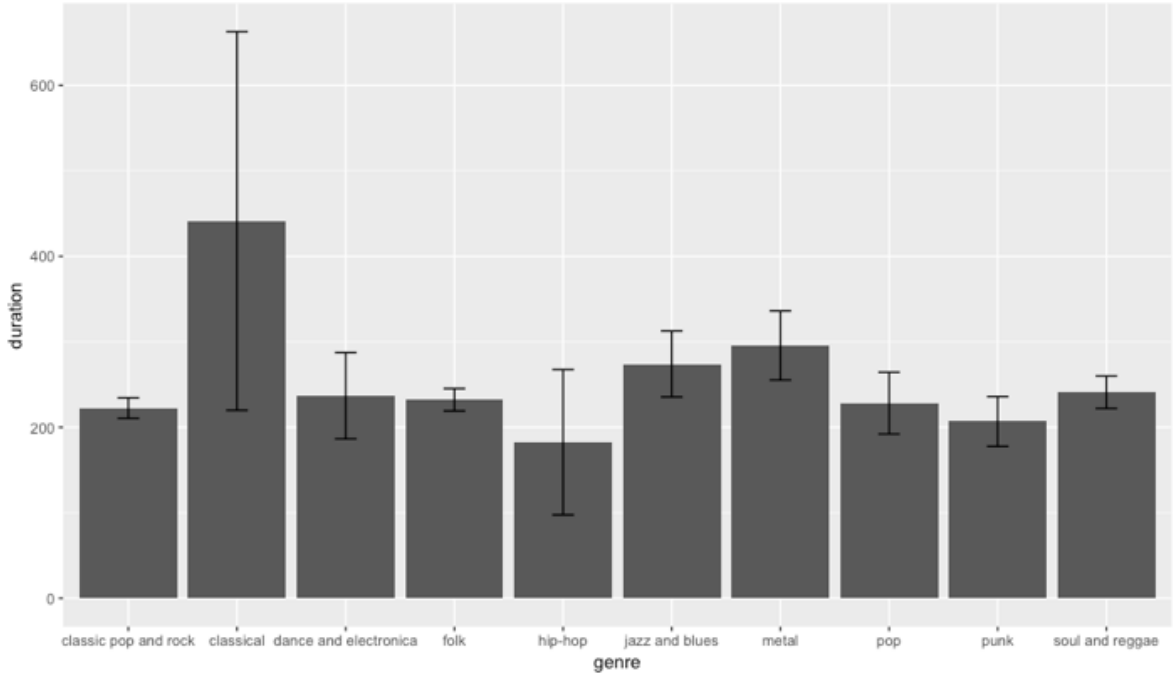
Corrected Pairwise t-tests:

	Classic Pop & Rock	Classical	Dance and Electronica	Folk	Hip-Hop	Jazz & Blues	Metal	Pop	Punk
Classical	5.4e-10	-	-	-	-	-	-	-	-
Dance and Electronica	0.44017	1.4e-07	-	-	-	-	-	-	-
Folk	0.36567	4.8e-09	0.81240	-	-	-	-	-	-
Hip-Hop	0.27953	2.9e-07	0.17845	0.18389	-	-	-	-	-
Jazz & Blues	0.00151	8.1e-06	0.10722	0.01558	0.02017	-	-	-	-
Metal	6.4e-05	0.00014	0.01659	0.00097	0.00500	0.33798	-	-	-
Pop	0.83346	1.1e-06	0.78622	0.88636	0.31099	0.13442	0.03355	-	-
Punk	0.63151	7.2e-07	0.41023	0.44359	0.61681	0.05760	0.01465	0.60710	-
Soul & Reggae	0.19339	5.5e-08	0.84919	0.56955	0.12930	0.09244	0.01042	0.66535	0.32122

95% Confidence Intervals:

Genre	Mean Song Duration	SD Song Duration	Sample Size	$t_{critical}$	Upper Interval	Lower Interval
Classic Pop & Rock	222.4497	91.36827	226	±1.970563	234.4262438	210.4731562
Classical	441.2338	264.86	8	±2.364624	662.6622777	219.8053223
Dance and Electronica	236.8892	135.1696	30	±2.04523	287.362367	186.416033
Folk	232.2343	73.33077	122	±1.979764	245.378064	219.090536
Hip-Hop	182.4872	91.93833	7	±2.446912	267.5159788	97.45842122
Jazz & Blues	273.9929	123.7868	42	±2.019541	312.567583	235.418217
Metal	295.6404	112.0515	32	±2.039513	336.039265	255.241535
Pop	228.2213	59.83761	13	±2.178813	264.3808087	192.0617913
Punk	206.7587	37.84723	9	±2.306004	235.8506546	177.6667454
Soul & Reggae	241.0172	71.6522	57	±2.003241	260.0290904	222.0053096

Bar Graph Visualization of the relationship between Genres and Song Durations:



As per the assumptions for a one-way ANOVA, the song durations are approximately normally distributed with a slight skew amongst most genres except pop, punk and hip-hop, independence of observations exist, but the homogeneity of variances is violated. Since the p-value is 0.00000000143 and below 0.05, the result is statistically significant and hence, the null hypothesis is rejected which means at least one of the population means of song duration for a genre is different. Since F is relatively big with a value of 6.988, the difference in means between genre groups is unlikely to be due to a sampling error. As R^2 is 0.1050154907, the effect size, which is the strength of the phenomenon, is small and shows that genre explains 10.5% of the variance in song duration. Thus, in response to the research question, our results provide substantial evidence to prove that different genres of songs have significantly different durations.

Meanwhile, the probability of getting at least 1 Type I error across all tests is $1 - (1 - 0.05)^{45} = 0.901$ which represents a familywise error rate of 90.1%. However, for the purposes of the Bonferroni correction which resulted in a pairwise α of 0.001111111, familywise α is assumed to be 0.05 since we aim to preserve the Type I error rate across all hypothesis tests. Based on the corrected pairwise t-tests, only the highlighted pairwise t-test results on the table are statistically significant under the pairwise α and reject the H_0. As such, we only entirely reject the null hypothesis for the Classical music genre since its song duration mean is significantly different from every other genre. Apart from that single genre, the mean song durations are significantly different between the genres Metal and Classic Pop and Rock and Folk.

Another interesting pattern that could be seen from the graph of mean song durations by genre is that newer and more modern genres like Punk and Hip Hop have much shorter mean song durations in comparison to older genres such as Classical and Jazz and Blues. This may signify that as we evolve as a society over decades, our musical tastes and popular genres evolve to favor shorter song durations as well. From the 95% confidence intervals table and graph with error bars, the significant difference in population means for song durations for each genre can be observed. However, the error bars for certain genres such as classical and hip-hop have larger error bars signifying greater uncertainty which is a limitation for our hypothesis tests. Hence, more song data resulting in greater and equal sample sizes across genre samples would be required to solve this limitation.