The Next Billion-dollar Idea

Analysis of Market of Popular iOS Applications

Alison Hong, Yicheng Shen, Zhihan Yang
Maths 215 - Introduction to Statistics
Carleton College

Introduction

In this age of rapid technological development, applications and mobile devices are changing more quickly than ever. People's work, entertainment, study and communication all rely heavily on various kinds of applications. While we are using those applications in our everyday life, how much do we know about the making and marketing of those products? Have we ever probed into the complicated relationship between different respects of those applications? As companies and organizations have already done many of their products, it is also important for us users to obtain more knowledge on this particular subject which demands increasing attention and further statistical investigation.

This lab report mainly examines various different aspects of a public dataset of applications. For the first question, we are interested in whether it is a wise way for developers to create bigger apps to make more profits. Do bigger sized apps usually mean higher price? Also, as many companies wanted to expand their target users, more languages appeared for the apps. Do the number of languages influence the user rating? Moreover, the third research question inquires about whether the five most popular genres of apps have equal mean prices. We believe that answering this question can help developers choose the most lucrative genre. The last research question is concerned about the association between genre and the number of languages supported. We believe that answering this question can help independent developers select the genre that supports the least number of supported languages in order to reduce their efforts in development. By studying those questions raised and discussed in this report, one will be able to achieve a new and deeper understanding on the making and marketing of popular applications.

Data Description

We obtained this dataset from a website called Kaggle, which is an online community for data scientists and machine learning practitioners to participate in competitions (for detailed url: https://www.kaggle.com/ramamet4/app-store-apple-data-set-10k-apps). This dataset was published by Ramanathan Perumal in July 2017. The publisher randomly gathered more than 7000 highly ranked Apple iOS mobile application details from the iTunes Search API at the Apple Inc website and used R and linux web scraping tools to create the dataset.

In total, there are 7197 apps included in this dataset. We have 16 columns of valid variables, including app ID, app name, size (in bytes), currency type, price (all in U.S. dollars), user rating counts for all version and for current version, average user rating value for all version and for current version, latest version code, content rating, primary genre, number of supporting devices, number of screenshots shown for display, number of supported languages, and Vpp device based licensing enabled (for any app with Vpp license enabled is 1 otherwise 0).

Due to the random sampling and the fairly large size of this available sample, we could use this dataset to infer a wider population of the Apple iOS mobile app market though it has certain limitations that we will discuss later.

Results

Research Question 1: Size and Price of the Non-free Apps

The first research question explores the relationship between the size and the price of iOS applications. In order to better understand this question and eliminate the influences of free apps, we only focused on those 3141 non-free apps in the dataset.

We obtained a five number summary of the prices of all non-free apps described in Table1. The average price is $3.955, which is larger than the median value of all prices ($2.99), indicating that the distribution of prices of these non-free apps is right-skewed. We eliminated two outliers from the dataset with price over $200 because we think these two cases may have to do with copyrights and are therefore not applicable to our analysis. Table 2 shows the five number summary of the rest 3139 non-free apps' prices. The mean price of $3.783 is slightly lower, but it still shows that the distribution is right-skewed, as shown in Figure 1.

	Five Number Summary of Prices of iOS Apps				
Table 1.	Min	1st Quartile	Median	3rd Quartile	Max
Price (U.S. Dollars)	0.99	1.99	2.99	4.99	299.99
	Five Number Summary of Prices of iOS Apps Without Outliers				
Table 2.	Min	1st Quartile	Median	3rd Quartile	Max
Price (U.S. Dollars)	0.99	1.99	2.99	4.99	99.99

	Five Number Summary of Sizes of iOS Apps				
Table 3.	Min	1st Quartile	Median	3rd Quartile	Max
Size (megabytes)	0.59	36.93	92.25	233.27	4025.97

We transformed the unit of measurement of apps' size from bytes to megabytes in order to help us more readily understand the distribution. Similarly, the distribution of sizes of apps is also quite right-skewed as shown in Figure 2. We could also observe this skewness in the five number summary of size in Table 3, from which we could learn that the mean size (265 MB) is much higher than the median size of all non-free apps (92.25 MB).

Figure 1. Prices of Non-free Apps without Outliers

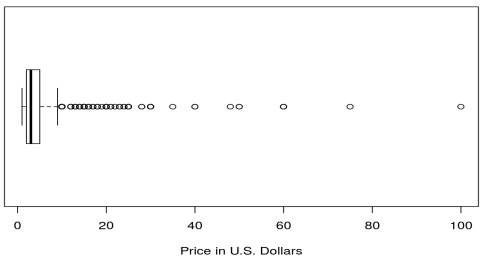
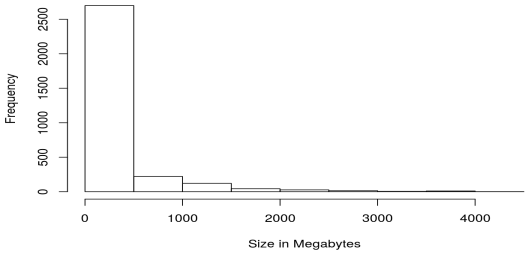
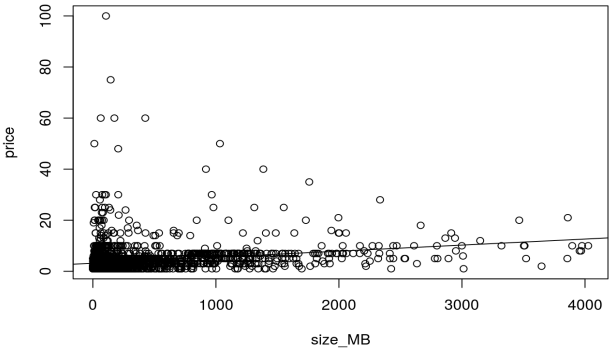


Figure 2. Sizes of Non-free Apps



We constructed a regression slope of the scatterplot between the size and the price of apps in Figure 3. With a sample slope correlation r value of 0.243, the linear relation between the two quantitative variables is positive and very weak. The size of apps can only explain 5.92% of the observed sample variations in prices and most data are clustered near the origin (r = 0.243, R-squared = 0.0592).

Figure 3. Scatter plot of Size and Price with regression line



We then constructed a regression slope of the scatterplot of between the size and log(price) of apps in Figure 4, which gave us a stronger and more positive linear relation with a sample correlation r value of 0.332. Now The size of apps can explain 11.05% of the observed sample variation in log(price) (r = 0.332, R-squared = 0.1105). According to the residuals plot in Figure 5, as the size increases, we observe from the pattern that the residuals are

not symmetrically distributed, which means that the linear model might still not be a very perfect way to capture the relationship between the log(price) and size of an app in this sample.

Figure 4. Scatter plot of Size and log(Price) with regression line

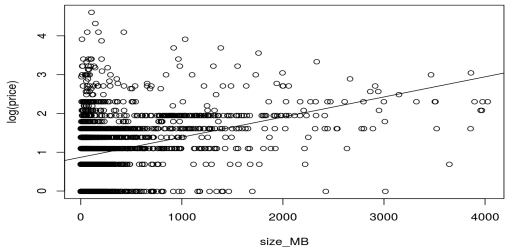
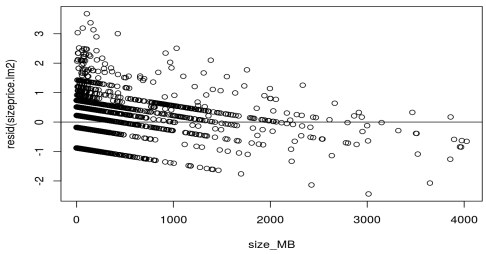


Figure 5. Residuals plot of Size and log(Price)



To determine whether the size is an effective linear predictor of the log(price) for the app at 5% significance level, we first set up our hypothesis.

Our null hypothesis (Ho) is : the slope between log(price) and size is zero;

Alternative hypothesis (Ha) is : the slope between log(price) and size is not zero.

From the R output for the least squares line to predict log(price) using the app's size, we could observe that the standardized test statistic is 19.74, with a p-value of approximately 0. This gives strong evidence that the size is an effective predictor of the log(price) for a popular non-free app (t*=19.74, df=3137, p-value<2e-16).

The 95% confidence interval for the slope of log(price) and size in MB is from 0.000467 to 0.00057. To interpret it more understandably, we are 95% confident that every 100 MB increase in size will result in the multiplicate increase from e^0.0467 to e^ 0.057 in median price of the app.

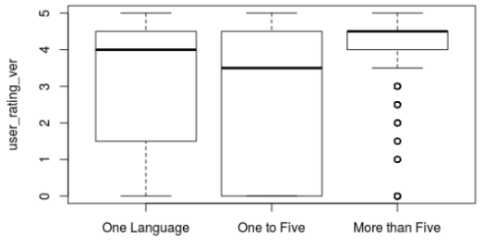
Research Question 2: Language Number and User Rating

Now we know that there is a weak linear correlation between the sizes of the applications and their prices, we are now exploring whether the mean user ratings are equal for apps that support 1 language, apps that support 1 (exclusive) to 5 languages and apps that support more than 5 languages. We partitioned the data in this way so that each level reflects a different magnitude of effort put into the app development. For a given app, its number of supported languages may indicate the amount of effort its developers put into it; therefore, we expect that mean user rating is highest for apps that support more than 5 languages.

Five Number Summary / Mean / SD for Ratings of Applications w/ Different Numbers of Language

	Min.	1st. Quantile	Median	Mean	3rd Quantile	Max.	SD
1 Language	0	1.5	4	3.111	4.5	5	1.836
1-5 Languages	0	0	3.5	2.618	4.5	5	2.104
>5 Languages	0	4	4.5	3.911	4.5	5	1.288

Table 4



User Rating and Language Boxplot

Figure 6

From Table 4 and Figure 6, we are able to see that all the means are all smaller than the medians, indicating left-skewness for all three distributions. Noticeably, ratings for applications supporting more than five languages are more concentrated in the 3-5 range and have a significantly larger mean user rating (3.911) than the other two groups (2.618 and 3.111).

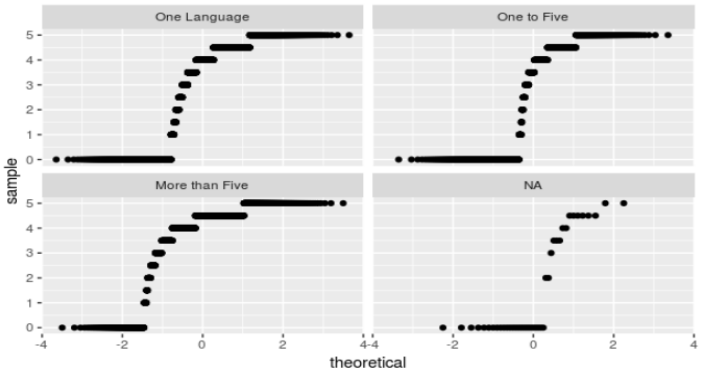
To further investigate whether the mean user ratings are the same for the three groups, we can conduct an ANOVA test to find out. The null hypothesis is that $\mu 1=\mu 1\text{-}5=\mu 5$; the alternative hypothesis is that the three means are not equal. $\mu 1$ denotes for mean user rating of Apps with only one language available, $\mu 1\text{-}5$ for that of Apps with 1 (exclusive) to 5 languages, $\mu 5$ for that of Apps with more than 5 languages.

The value of the F test statistics is 247.1 and the p-value is extremely small and close to 0, suggesting a statistically significant result. Therefore, we have strong evidence to reject the null hypothesis and conclude that at least one of the three mean user ratings is not equal to others. The result also corresponds to the boxplot in Figure 6, which shows that the rating of Apps supporting more than 5 languages appear to be the higher one among the three levels.

Count for Each Category of Language Number (Table 5)		
1 Language	1-5 Languages	More than 5
3767	1253	2136

To check if we can conduct the ANOVA test on the data, we can see that the SD rule is met since the largest SD (2.104) is smaller than twice the smallest SD (2*1.288 = 2.576). In addition, the sample size large enough for each category (Table 5). The ratings within each group are not exactly normally distributed according to Figure 7. However, according to McDonald (2014), since the p-value is extremely small here, we can still conduct ANOVA tests.

The QQ-plot of Language Number (Figure 7)



The 95% confidence interval of the difference between the mean user rating for applications with one to five languages and that for applications with five languages is from -1.42 and -1.16. This means that the mean user rating for applications with one to five languages is from 1.42 dollars to 1.16 dollars less expensive than the mean rating for applications with more than five languages.

Research Question 3: Genre and Price

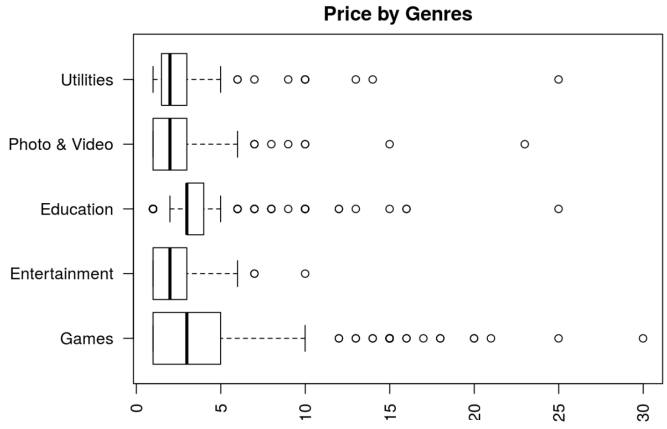


Figure 8. Boxplot of Price by Genre

This research question is concerned about whether the five most popular genres of apps have equal mean prices. The genres analyzed include Game (3862 apps), Entertainment (535 apps), Education (453 apps), Photo & Video (349 apps) and Utilities (248 apps).

EDA using boxplot of raw data highlighted three notable outliers (all found in the genre "Education") with prices higher than 50 U.S. dollars. Since these expensive prices are largely due to copyright reasons that are unlikely to be shared by other apps, the three outliers were excluded from our analysis. The boxplot of cleaned data is shown in Figure 8.

Since we are testing for whether or not the five genres have equal means, the ANOVA test is a reasonable choice. The ratio of the largest standard deviation of price to the smallest one is less than 2 (Appendix), suggesting that the *equal-variance* assumption is satisfied. In addition, the number of data points per group is sufficient for analysis with ANOVA.

Although the data distributions are quite right skewed, McDonald (2014) suggested that ANOVA only fails, i.e., results in high false-positive rates, when the data distributions are both extremely non-normal and p-value is smaller than but close to 0.05. The p-value, as we will see later, is much smaller than 0.05. Therefore, although the assumption that data are drawn from normal distributions is not satisfied, ANOVA is still valid in context.

The result of the F test is 31.6. Results show that, if all genres have equal mean prices, we would see the observed F-statistic test statistic, or one even larger, less than 2e-14% (< 0.01%) of the time. Therefore, we have very strong evidence that at least one genre has a mean price that is different from other genres.

We constructed a 95% confidence interval (1.41, 1.96) for the largest observed difference between two mean prices (1.68 dollars), which occurred between Education and Entertainment. Therefore, we are 95% confident that the true mean price of Education is 1.41 dollars to 1.96 dollars higher than that of Entertainment.

Research Question 4: Genre and Number of Supported Languages

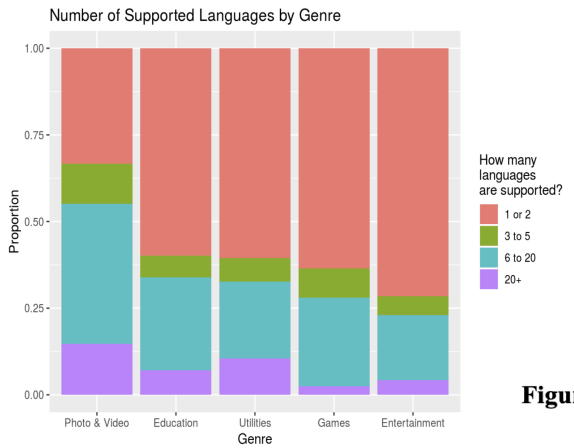


Figure 9 shows that Photo & Video has the lowest proportion of apps supporting 1 or 2 languages (33.3%) and Entertainment has the highest proportion of apps supporting 1 or 2 languages (71.5%). The exact proportions are shown in Table 6. There appears to be an

Figure 9. Stacked bar graph of number of supported languages by genre.

association between the two variables, but we would like to quantify the exact degree of statistical significance.

Genre	Photo & Video	Education	Utilities	Games	Entertainment
"1 or 2" Language Prop.	33.3%	59.9%	60.5%	63.5%	71.5%

Table 6. Proportion of apps supporting 1 or 2 languages for the five most popular genres.

Since we are testing for association between two categorical variables, the chi-square test is a reasonable choice. All expected counts are greater than 5 (Appendix), indicating that the data satisfies the assumption needed for inference.

The chi-square test statistic is 258.45. Results show that, if there is no association between the two categorical variables, we would see the observed chi-square test statistic, or one even larger, only about 2.2e-14 % (< 0.01%) of the time. Therefore, we have very strong evidence that there is an association between the two categorical variables.

To quantify the observed association, we constructed a 95% confidence interval for the largest observed difference in sample proportions (38.2% points), which occurred between the observed proportion of apps supporting 1 or 2 languages in Entertainment and that in Photo & Video. The 95% confidence interval for this difference is (31.8%, 44.6%). Therefore, we are 95% confident that the true proportion of apps supporting 1 or 2 languages in Entertainment is 31.8% pts to 44.6% higher than that in Photo & Video.

Discussion:

In general, our analysis of the dataset is able to provide us with many interesting discoveries about the iOS market as well as evidence to support our hypotheses. Most of our hypotheses are supported by statistically significant results.

Previously, people might think that the app's price is directly associated with how large it is since more efforts were put on developing these apps. Our analysis suggests that the linear correlation between price and size or log(price) and size in this sample is not very strong, and the residuals plot suggests that the linear regression model systematically overestimates the log(price) of larger apps. We do have strong evidence in our analysis that the size is an effective predictor of the log(price) for the app and we are 95% confident that every 100 MB increase in

size of an app will result in the multiplicate increase from e^0.0467 to e^ 0.057 in median of price of the app.

While whether there is a relationship between user rating and language number is not a subject on which one can make an easy assumption, the test result suggests that there is a statistically significant association between these two variables. We are 95% confident that the mean user rating for applications with one to five languages is from 1.42 dollars to 1.16 dollars less expensive than the mean rating for applications with more than five languages.

We cannot tell for certain which of the 3 groups (group of Apps with 1, 1 to 5 , and more than 5 languages) has or have unequal mean(s) with the rest of the groups. While the distribution is not strictly normally distributed, there do exist some questions on the ANOVA test.

From the conclusion we have drawn from the statistics, there emerges the question of how the number of languages influences user ratings. As it is not a randomized experiment, we cannot predict causal relationships between the two variables, and one possible confounding variable is the total number of ratings. Does more available languages lead to more raters? Does more raters result in higher ratings? Further studies will be needed to investigate the subject.

The third research question inquired about whether the five most popular genres of apps have equal mean prices. We hypothesized that genres have different mean prices because developers of different genres have different motivations and different genres have different audience bases. Analysis has shown that, indeed, it is highly unlikely that all genres have equal mean prices. Specifically, we are 95% confident that the mean price of the Education genre is 1.41 dollars to 1.96 dollars higher than that of the Entertainment genre. This is reasonable given that apps in Education usually have copyright affiliations while apps in Entertainment usually profit from users' subscriptions rather than one-time payments.

The fourth research question is concerned about the association between genre and the number of languages supported. Analysis has shown that there is a strong association between genre and the number of supported languages. Specifically, we are 95% confident that the proportion of apps supporting 1 or 2 languages in Entertainment is from 31.8% pts to 44.6% pts higher than that in Photo & Video. Entertainment had the highest proportion of apps having only 1 or 2 languages, and we were not surprised because mainstream media is dominated by U.S.

produced movies. However, we were surprised to find out that Education and Gaming, two genres with huge user bases in multiple languages, have a large proportion of apps supporting only 1 or 2 languages. We were also surprised that Photo & Video apps support more languages than education and gaming apps.

One problem that we experienced during hypothesis testing is that the distributions of numerical variables are often severely right-skewed, challenging the normality assumption underlying the ANOVA test. Fortunately, MacDonald (2014) ran numerous experiments with skewed distributions and discovered that false positive rates of ANOVA stayed well below 5% when the corresponding p-values were far away from 5%. Since the p-values of our results are often negligible (< 0.001), our conclusions are deemed to be statistically significant.

Overall, our generalization primarily reflects the Apple iOS mobile app market prior to July 2017. Since our dataset largely consists of more frequently ranked apps, our findings should be mainly and only applied to the population of popular apps in the market. Future studies can collect user-side and developer-side variables and use them in combination with variables we used to answer questions about app-user interaction. For example, given that different genres have different mean prices, why are users happy to afford some genres but not others? Is the number of supported languages a good reflection of developmental effort? Such questions will undoubtedly lead to more interesting insights of the market.

6. Conclusion

This study is intended to help users to gain more knowledge of the patterns emerged in a popular application Market, the iOS App Store. Our investigation has led to interesting conclusions: size is positively but weakly correlated with price; applications that support more languages tend to have higher user ratings; different genres have different mean prices and different levels of language support.

Reference

McDonald, J.H. 2014. Handbook of Biological Statistics (3rd ed.). Sparky House Publishing, Baltimore, Maryland.

Title: The Next Billion-dollar Idea -- Analysis of Market of Popular iOS Applications

Names: Allison Hong, Yicheng Shen, Zhihan Yang

Abstract

Nowadays, people's work, entertainment and communication rely heavily on various kinds of applications. While we are using these applications in our everyday lives, how much do we know about the trends in their market? Since app-development companies have conducted countless studies on this topic, it is important for us users to gain more knowledge in this subject as well. The data that we investigated was published by Ramanathan Perumal on Kaggle. It contains variables including size, number of supported languages, user rating and price of 7197 applications, which were randomly sampled from the highly ranked applications on Apple App Store in July 2017. Our investigation has led to several interesting insights: size is positively but weakly correlated with price; applications that support more languages tend to have higher user ratings; different genres have different mean prices and different levels of language support.