Project Proposal

Leader: Zhiying Xie
Members: Yifei Chen, Seoyoung Park, Ziqi (Katie) Chen

Section 1

- **Description:**

Currently, mobile apps are experiencing a booming period. Android takes over 53.2% of the smartphone market and iOS system takes 43%. More and more apps are competing with each other, which begs the questions: what are the factors that navigate users to find iOS apps easily? And what are the primary factors that could be used to predict the success of an app?

In this study we will present an effective resampling tool, bootstrap, to provide good estimation of the ratings from the users based on different characteristics of the app. The dataset was obtained from Kaggle (https://www.kaggle.com/ramamet4/app-store-apple-data-set-10k-apps) where it was originally extracted from iTunes Search API at the Apple Inc website, which contains more than 7000 iOS application system details. The dataset contains 7198 apps and 16 variables. The data was collected in July 2017.

- Variables:
    - "id" : App ID
    - "track_name": App Name
    - "size_bytes": Size (in Bytes)
    - "currency": Currency Type
    - "price": Price amount
    - "rating_count_tot": User Rating counts (for all version)
    - "rating_count_ver": User Rating counts (for current version)
    - "user_rating" : Average User Rating value (for all version)
    - "user_rating_ver": Average User Rating value (for current version)
    - "ver" : Latest version code
    - "cont_rating": Content Rating
    - "prime_genre": Primary Genre
    - "sup_devices.num": Number of supporting devices
    - "ipadSc_urls.num": Number of screenshots shown for display
    - "lang.num": Number of supported languages
    - "vpp_lic": Vpp Device Based Licensing Enabled

Section 2

1. Q: What could be some important factors to predict the success (rating value) of an APP. (i.e. user rating, version number, number of supporting device/language/screenshot, size)
    - Q: What are the three most important factors affecting the average rating of an App?
    - Q: What are the most important factors affecting the total rating counts of an App?
    - Q: Will the genre/ category affect the rating?

2. Q: Is it more profitable to develop a free app (more user) or paid app.
    - How do you visualize price distribution of paid apps?
    - How does the price distribution get affected by category?
    - As the size of the app increases do they get pricier?
    - Does the range of price differs depending on the category?

Approaches:

Groups the apps into paid and free app and see how the rating value changes.

1. Possible resampling methods we are going to use:
   Bootstrap
    a. Get the sampling distribution of users rating score based on free (price = 0) and paid (price > 0) apps.
    b. Get the sampling distribution of total users rating counts based on the number given.
    c. Predict the population mean and variance of the rating score from different groups.
2. Possible dependent variable: user ratings (user_rating), rating count (rating_count_tot)
3. Possible independent variables to consider:
    a. Price
    b. Primary genre (prime_genre)
    c. Supporting device number (sup_devices.num)
    d. Version (ver)
    e. Content rating (cont_rating)
    f. Supporting languages (lang.num)