Screen-Time, App Categories in Comparison of Chinese and Indian at Northeastern University, Seattle

Sheng (Jeremy) Jiang & Amal Sharma

Submitted to :Shivani Patel

Course: IE 6200

Table of Contents

Abstract

Background:

With the development and evolution of mobile devices and applications, smartphone has been increasingly crucial as an inevitable tool to everyday life. But how people spent screen-time on smartphone has not comprehensively measured and compared to demographics.

Aims:

The aim of this observational study is to characterize smartphone usage by measuring screen-time, most-used app categories. And find the pattern of smartphone usage with association to demographics in particular to ethnicity. This study particularly focuses on two ethnicity groups: Chinese and Indian. The scope of this observational study will be on Northeastern University, Seattle campus.

Methods:

Collection Method: This study uses survey to randomly collect around 90 data points from respondents.

Sampling Strategy: This study focuses on Chinese and Indian at Seattle campus of Northeastern University who are adults uses smartphone. All 90 collected data points used for exploratory analysis. By setting sample frame, this study filtered out 7 irrelevant data points for statistical analysis. In the next step, analysts use stratified sampling strategy randomly selecting 32 survey respondents from Chinese group and 32 survey respondents from Indian group for two ethnicity groups comparison.

Exploratory Analysis: This study performed plots to visually show the average screen-time distribution in terms of age groups, and ethnicity groups.

Statistical Analysis: By comparing differences in means of screen-time between Chinese and Indian, analysts performed two-sample t-test approach.

Results:

Exploratory Analysis Results: Plots in exploratory analysis indicates that Indian constitutes 54% of sample points, Chinese constitutes 38% of sample points, American constitutes only 8% of sample points. In the box plot, this study indicates Chinese and Indian as our targets have relative close average screen time. Chinese group has an average screen-time 6.09 hours with a median 6.20 hours (IQR 4.487 to 7.082) and Indian group has an average screen-time 5.361 hours with a median 5.10hours (IQR 4.30 to 6.30).

Statistical Analysis Results: Using two sample t-test approach for the true population mean difference in screen-time per day between Chinese and Indian. With p-value =0.3856217, this study suggests that there is no evidence showing the true population mean screen-time for Indian is different than Chinese at the significant level alpha=0.05. With 95% confidence interval, the true difference in mean screen-time between Indian and Chinese is between -0.7069666 and 1.780092.

Conclusion:

These findings on smartphone screen-time, app categories between Chinese and Indian are highly based on survey taken by participants themselves. This study fails to reject that Chinese and Indian has the same average screen- time using their smartphones. Screen-time could differ across age and ethnicity in a larger and less-biased environment setting, but this study cannot support it due to its limitation. These findings cannot support conclusions on causation since there are so many confounding factors exist.

Introduction

With the development and evolution of mobile devices and applications, smartphone has been increasingly crucial as an inevitable tool to everyday life. In the report *US Time Spent with Mobile 2019*, analysts suggest that smartphone will likely remain the dominant device for consumer media and 90% of smartphone time is in apps (Dolliver,2019). But it is still ambiguous if there exists a relationship between screen-time & age group and different ethnicities uses which category of apps more often.

How people spend their time on mobile phone has always been an interest to many fields: app developers seek to build the most profitable and efficient interactive UI to attract phone users; scientists work on to find correlation between smartphone screen-time and productivity in real life; phone users try to balance their time between screen-on and screen-off to optimize their time management.

This observational study is designed to:

A. Characterize smartphone usage by measuring average screen-time for different age groups

B. Find the pattern of smartphone usage with association to ethnicity in two particular groups: Chinese and Indian.

The goal of this observational study is to provide estimations for these questions below:

1. What is average screen-time do people spend on mobile phone?

2. How screen-time changes among different age groups?

3. What type of apps do Chinese and Indian use most often during their screen-time?

4. Is there a difference in the screen-time per day between Chinese and Indian as two ethnicity groups?

Variable Definition:

- **Screen-time (hours. minutes):**

  Definition :Screen time is the direct measurement of phone usage. This study will have two sets of screen-time variable. One is the screen-time over the phone for the last 24 hours, the other is the average screen-time over the phone for the last 10 days. In the exploratory analysis, this study found the latter is more appropriate for analyzing. Because the former set of data may exist biased errors based on one-day movement. These two sets of the variable can be directly found in people's smartphone at Settings tab. We have followed 24-hour clock format. Hours and minutes can be displayed as hours. minutes in decimal format[HH.MM]. For example,1hour and 30 mins can be displayed as1.30 hrs.

  - The type of the variable: Numeric Variable in decimal; Response Variable; Quantitative Continuous Variable.
  - The scale of the variable: Screen-time is ranged from 0 to 24 hours per day.

- **App categories:**

  Definition: This observational study defines types of apps under 7 categories with examples below:

Table 1 : App Category Definition

Category Name	Definition	Example
Utilities	serve as a function to mobile user.	Calendar,Weather, Map, etc.
Entertainment/Games	video games and streaming videos/shows.	Netflix, PubG, Youtube,etc.
News/Information	provide news or information to users.	Buzzfeed, Reddit,Weibo, etc.

Productivity	helps users accomplish a task.	Gmail,Dropbox,Evernote, etc.
Lifestyle	supports your individual facets that define your lifestyle like fitness, dating, food, music, and travel.	Spotify,Uber, Yelp,Tinder, etc.
Social Networking	shares your movement with networks.	Instagram, Facebook, Wechat, LinkedIn, etc.

- Applications : It is the categorical variable which we will use to find out the screen time between two ethnicities i.e. Indian and chinese.
- The type of the variable: Categories Variable; Explanatory Variable; Discrete Variable.
- The scale of the variable: This study set 7 general categories for participants to pick from 2 ethnicities.

- **Age groups:**

  Definition: This observational study focuses on five age groups. People in 18 to 23 year-old as the first group, 24 to 29 year-old as the second group, 30-35 as the third group, 40-47 as fourth group and above 47 as fifth group.
  - The type of the variable: Categories variable; Explanatory variable;Discrete variable.
  - The scale of the variable: Participants has to be in one of these three age groups.
- **Ethnicity:**

  Definition: This observational study narrows ethnicity variable into two groups based on the sampling feedback: Chinese and Indian.

- The type of the variable: Categories variable, Explanatory variable, Discrete variable.
- The scale of the variable: Participants are either Chinese or Indian.

Methods:

- Target population for this study: Adults from China and India who uses smartphone at Northeastern University, Seattle campus.
  - Sampling Size: 90 data points.
  - Sample Frame: Participants need to be over 18 years old, and need to either Chineses or Indian.

Data Collection Methodology:

6-question survey. This study uses survey to collect data from respondents.

- First question is an open-ended question to gather information on respondents' ethnicity.Through this question, analysts can divide respondents into groups based on their ethnicity.

- Second question is a closed-end question about age grouping. Through this question, analysts can divide respondents into different age groups and can find out screen time for these age groups.

- Third and Fourth questions are open-ended questions to gather respondents' screen-time over the past 24 hours and average screen-time over the past 10 days. The purpose of having two sets of screen-time is to gather more accurate screen-time data to avoid some fluctuations from some unusual circumstances. For example, people could use smartphone frequently for navigation purposes during traveling period, but they barely use smartphone for such a long time in their normal days.

- Fifth and Last question is a closed-end multiple choice on most used App name and its categories. Through this question, analysts collect respondents' most-used app categories and

names of specific apps that respondents used most often.

Sampling Strategy:

- First, analysts randomly gathered data for 90 samples of the accessible population as potential participants.
- Secondly, analysts used sample frame to rule out data points from unqualify participants and make sure enough coverage of the sample to the population. From the survey collection, analysts will collect each survey respondent's ethnicity, screen-time for last 24 hours, average screen-time for past 10 days, most-used categories of apps,most-used app name.
- Thirdly, analysts uses stratified sampling strategy for statistical analysis. Stratified sampling divides potential participants into two ethnicity groups. It randomly selects 32 survey respondents from Chinese group and 32 survey respondents from Indian group for two ethnicity groups comparison.

Table 2 : App Category wise Sample distribution

App Category	China	India	US
Game/ Entertainment	3	15	0
Lifestyle	0	2	0
Productivity	1	1	1
Social	27	30	6
Utilities	3	1	0

Randomization:

- In the survey, analysts randomize the order of choices to avoid question order Bias.
- Randomly selects 32 survey respondents from Chinese group and 32 survey respondents from Indian group for two ethnicity groups comparison.

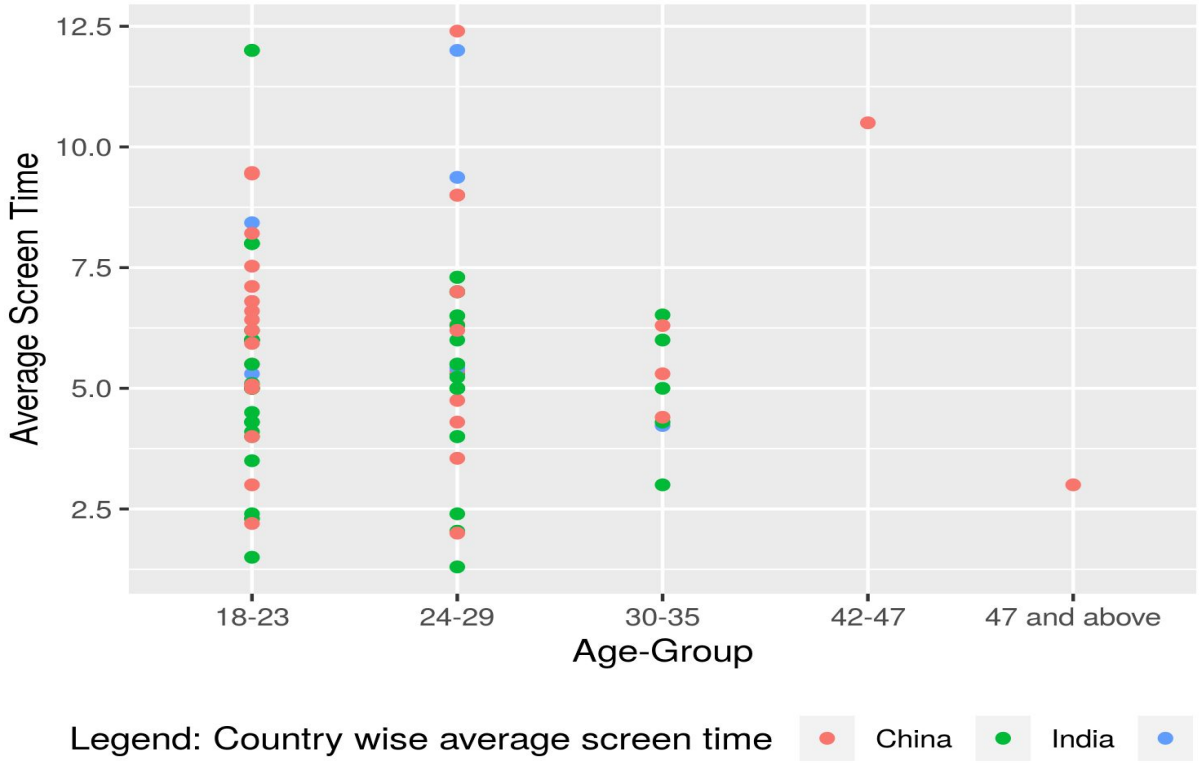
Shortcoming of sample strategy:

Following Bias can be observed:

- Information Bias - People may not take the survey seriously and there are some incorrect information collected.
- Confounding Bias - Other factors like individual behavior, occupation will lead to confounding bias in this study.For example, people will use particular apps most often just because it's job related.
- Response Bias-The survey questions could lead to inaccurately or falsely response. For example, there are a few respondents had wrong understanding on average screen-time over the past 10 days.
- Selection Bias-The sample size is too small for the coverage of population. It may not be a representative sample to the population.

Analysis

Exploratory Analysis



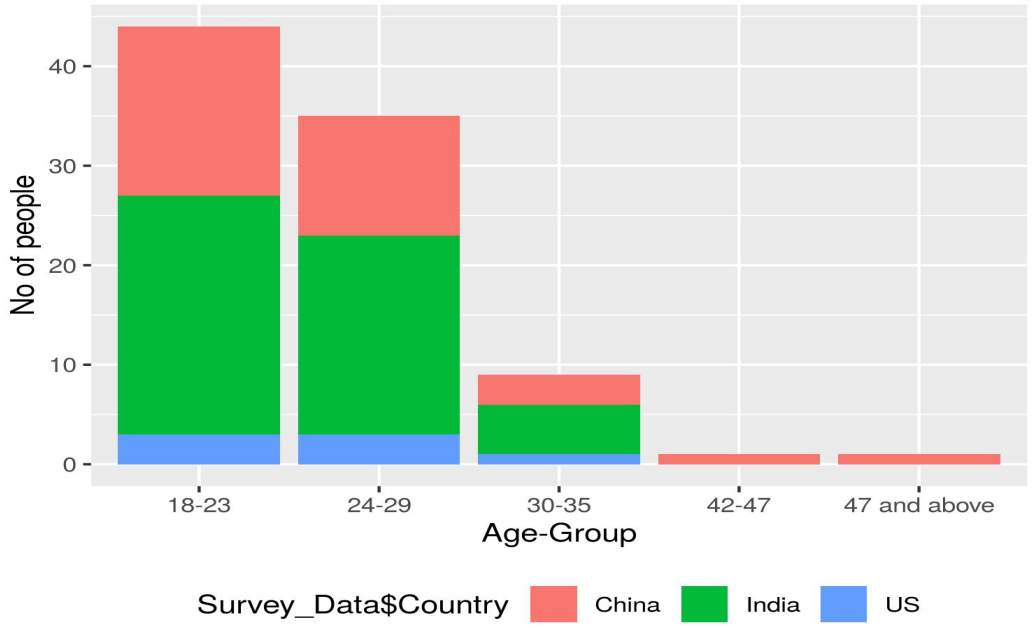
(i) Age Group vs Average Screen Time

We analysed data to check the average screen time among different ethnicities and how data is distributed by plotting dot plot which gives us rough pattern about the sample population screen usage. Above distribution is unexpected in the study as we were expecting a difference in screen time between different age groups. However, Interesting fact is that age group 18-23 and 24-29 constitutes around 87% of the sample. It also shows distribution among different ethnicities present in our sample i.e. Indian constitutes 54%, Chinese constitutes 38%, American constitutes only 8%.

Age group in Population distribution:

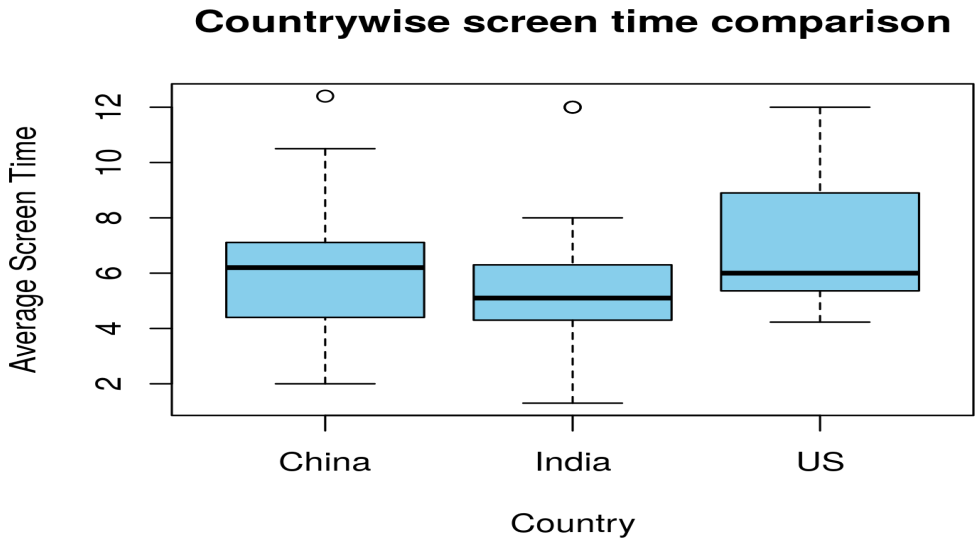
Comparing the age distribution in terms of ethnicities, we found that sample data of Chinese population is more in 18-23 age group than 24-29 whereas Indian population is not at all present

above 42 age group, which shows data above 42 age is negligible and can be ignored. In addition, American data points should also ruled out based on the sample frame this study states.



(ii) Age Group vs Population distribution

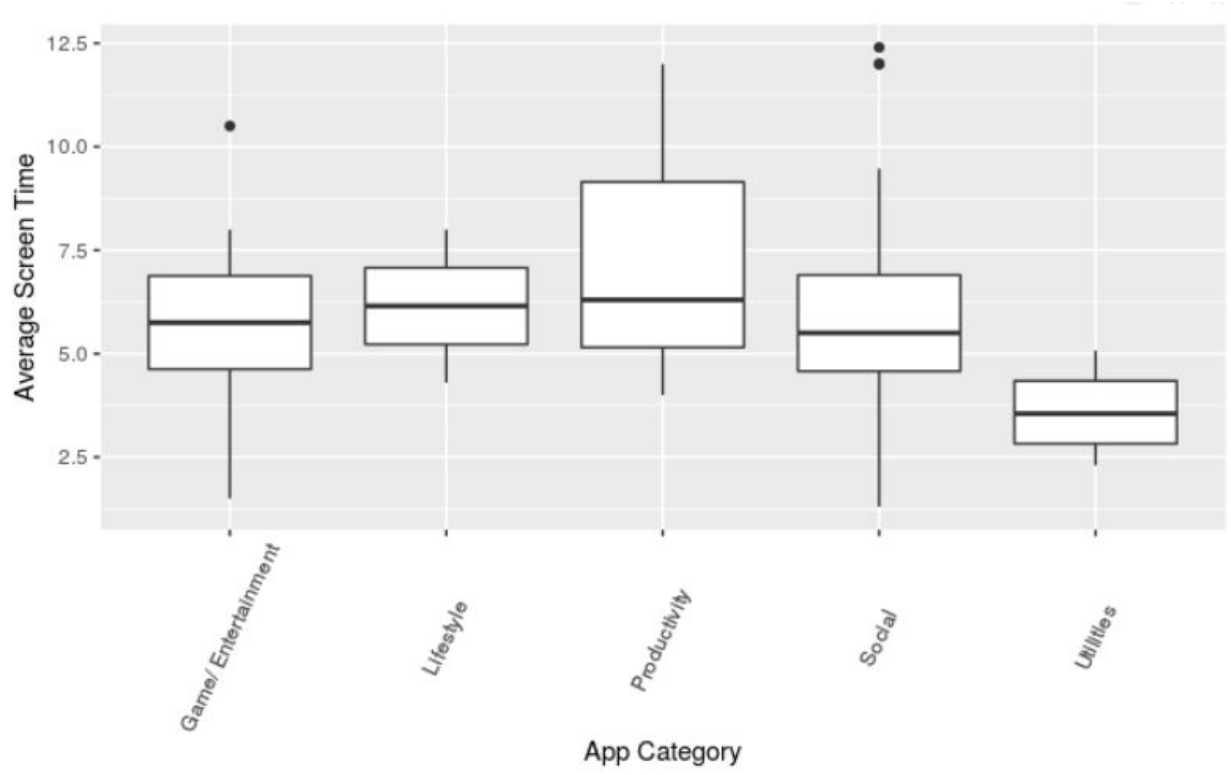
Ethnicities in average screen-time (in hours):



(iii) Ethnicities vs Average screen time

On further analysis, this study found that median of screen-time median for Chinese (6.20) sample is higher than the Indian (5.10). Chinese group has an average screen-time 6.09 hours (IQR 4.487 to 7.082) and Indian group has an average screen-time 5.361 hours (IQR 4.30 to 6.30). All these data are interlinked and is important for our study as it provides the basis for statistical analysis. The pattern needs to be further studied and analyzed if there is a difference in average screen-time between these two ethnicities.

Application category and average screen-time (in hours):



(iv) App Category vs Average screen time

Relation between app category and Average screen time is most interesting and is presented in figure no (iv). It shows that out of 90 data points collected, 63 is from Social category which stands for almost 70% of the whole sample. Also 2 outliers spend excessive amount of screen-time in this category. Further, It shows that Lifestyle and Productivity has the highest

median around 6.15 hours whereas utilities has the lowest median which is 3.55 hours. Social being most populous has median of 5.50 hours

Statistical Analysis

Before statistical analysis on two ethnicity groups comparison. Analysts choose the **stratified sampling strategy** as a method to filter out irrelevant data and creates equal-quantity data comparison. Firstly, this strategy filters out the US ethnicity group as it is irrelevant to this part of analysis. Secondly, this strategy divides remaining data into two ethnicity groups which are Chinese and Indian. Then it randomly selects 32 data from Chinese group and 32 data from Indian group for comparison. Finally these two 32-data-point groups will be used for two-sample analysis. Through this stratified sampling strategy, the analysis is more randomized and less selection biased.

Analyst uses two sample t-test for difference in means rather than difference in proportions because this study focuses on comparing average screen-time per day between Indian and Chinese. This study analyzes on one continuous quantitative variable and one category variable. The continuous quantitative variable is the average screen-time for past over 10 days, and the category variable is ethnicity: Chinese or Indian.

Question of Interest:

Is there a difference in the average screen-time per day between Indian and Chinese at Northeastern University, Seattle campus?

Statistical test: Two sample approach T-test.

Explanation: This study focuses on the comparison of Chinese and Indian as two populations in terms of screen-time. The true mean screen-time as well as the true population of each group are unknown. Therefore, the best choice of this study is two sample t-test.

Condition and Requirement

- Is the sample representative of the population?

- ➢ The sample cannot surely be the representative of the population. Because the ratio of Chinese true population versus Indian true population cannot be sure, and true population of these two groups is also unknown.

❖ Question of interest has to do with the difference of means between two populations.

- ➢ The question of interest is correlated to the difference of means between Chinese and Indian group.

❖ 2 independent samples from 2 populations.

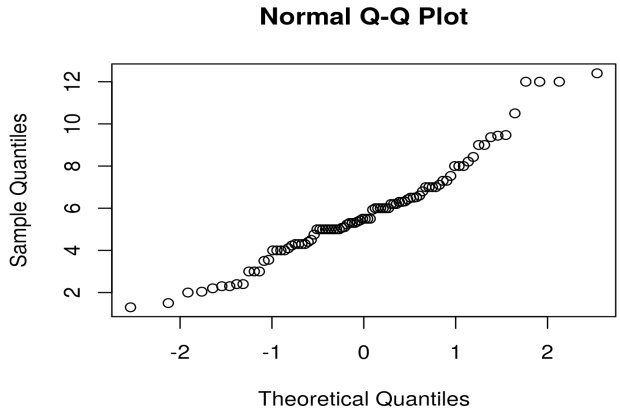
- ➢ Chinese and Indian are two independent samples from two populations for this analysis.

❖ The population data must be normally distributed.

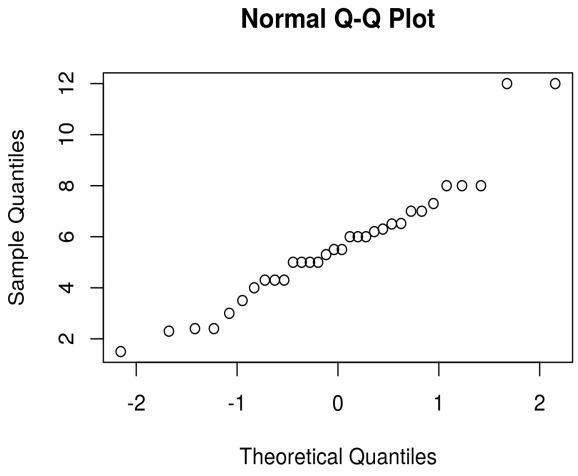
- ➢ QQ plots shows both sample data are nearly normal distributed.
- ➢ Sample size is 32 of each sample which is larger than 30.

QQ - Plot:

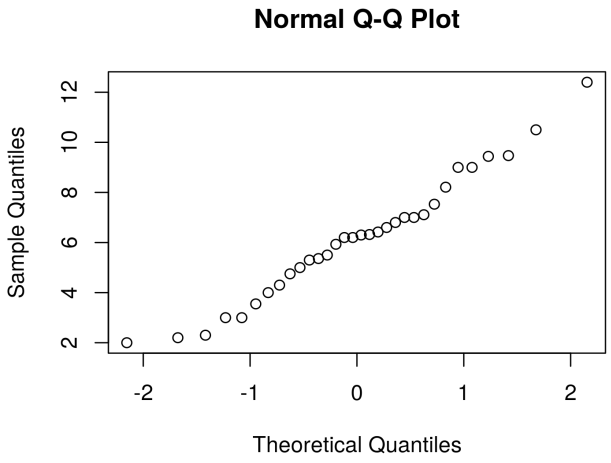
Three Normal QQ plots indicates the data is nearly normal distributed in total sample, Indian group sample and Chinese sample.In this study, the qqnorm() function will be utilized to plot the graph and draw a conclusion based on whether average screen time for both population is distributed normally or not.



(v) Total Sample



(vi) Indian Sample



(vii) Chinese Sample

Based on the Normal Q-Q plots, the study suggests Chinese and Indian group distributions are all nearly normal distributed. Therefore, these data can be used for two sample t-test in the next step.

Parameter:

We are interested in the true population mean difference in screen-time per day between Chinese μ_c and Indian μ_i at Northeastern University, Seattle campus.

Hypothesis:

$$\mu_c - \mu_i = 0$$

Null Hypothesis (H0): The true population mean screen-time for Chinese is equal to the true population mean screen-time for Indian.

$$\mu_c - \mu_i \neq 0$$

Alternate Hypothesis (HA):The true population mean screen-time for Chinese is different from the true population mean screen-time for Indian.

Sample Statistic

Here $\overline{x_i}$ = Average screen-time of Indian Sample whereas $\overline{x_c}$ = Average screen-time of Chinese Sample

$$\overline{x_c} - \overline{x_i}$$

Test Statistic

Here S_i^2 = Variance of Indian Sample whereas S_c^2 = Variance of Chinese Sample.

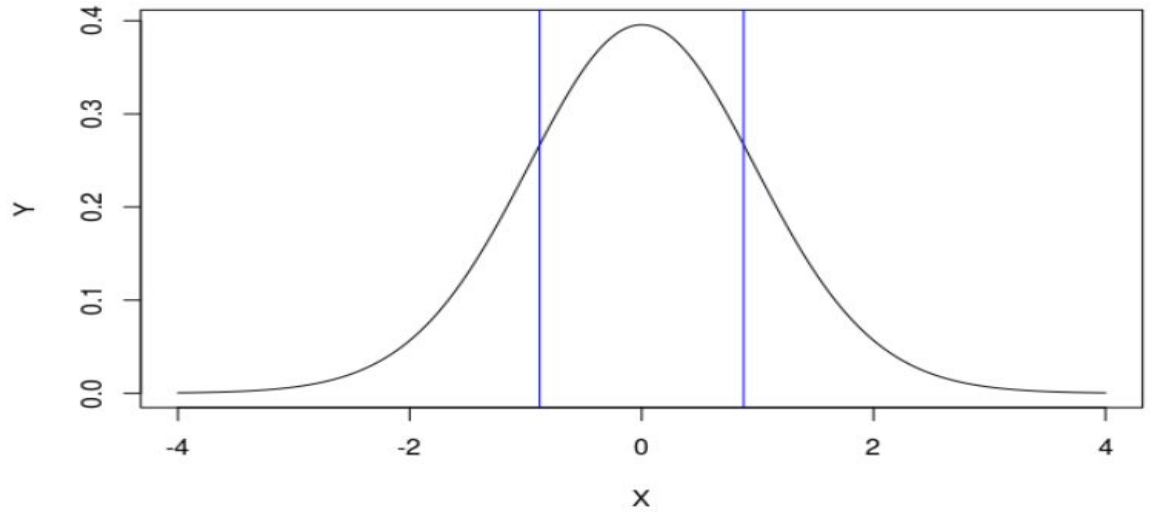
Also n_i = Length of Indian sample size and n_c = length of Chinese sample size.

$$t_{min(n_c-1,n_i-1)} = \frac{(\overline{x_c}-\overline{x_i})-(\mu_c-\mu_i)}{\sqrt{\frac{s_i^2}{n_i}+\frac{s_c^2}{n_c}}}$$

Distribution of the test statistic

$$t_{min(n_c-1,n_i-1)} = \frac{(\overline{x_c}-\overline{x_i})-(\mu_c-\mu_i)}{\sqrt{\frac{s_i^2}{n_i}+\frac{s_c^2}{n_c}}} \sim t_{min(n_c-1,n_i-1)}$$

T- test distribution graph



P- Value

Analysis involves comparing the P-Value with the significance level $\alpha = 0.05$. If P-Value is $< \alpha$ then our null hypothesis is rejected whereas if P-value is $> \alpha$ there exist a weak evidence and we will not be able to reject the null hypothesis

By using the t-test statistical method in R, the p value was calculated and returned 0.3856217 The lower bound of the confidence interval is -0.7069666 and the upper bound of the confidence interval is 1.780092.

Comparison with R built in Welch's T test

As Per the Calculation, P-Value for both the test is almost same

Parameters	T-test statistical method	R Built in Welch's T-Test
P-Value	0.3856	0.3823
Lower Bound	-0.7069	-1.755
Upper bound	1.7800	0.6822

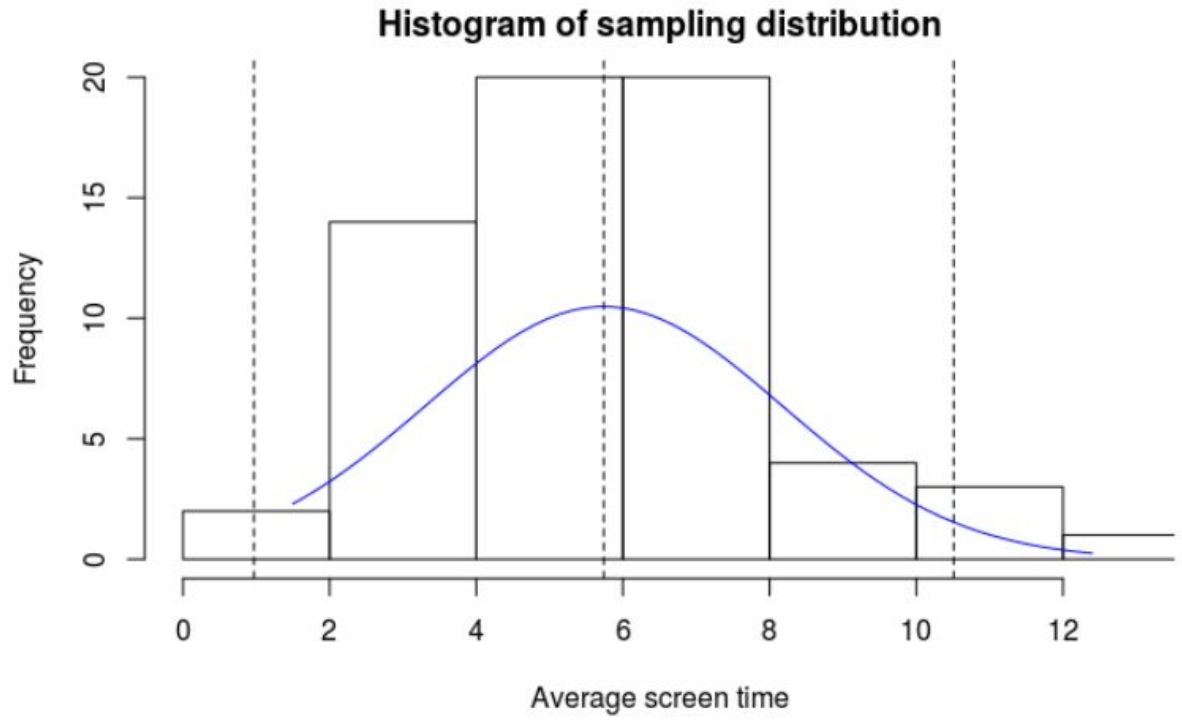
Interpretation:

From the computation, analysts found that there is no evidence (p-value=0.3856217) to suggest that the true population mean screen-time for Chinese is different than Indian. We fail to reject the null hypothesis that there is no difference true mean screen-time between Chinese and Indian at the level $\alpha = 0.05$. With 95% confidence, the true difference in mean screen-time between Chinese and Indian is between -0.7069666 and 1.780092. The null hypothesized difference between the mean screen-time is zero and zero is in the 95% confidence interval which agree with our failure to reject the null hypothesis.

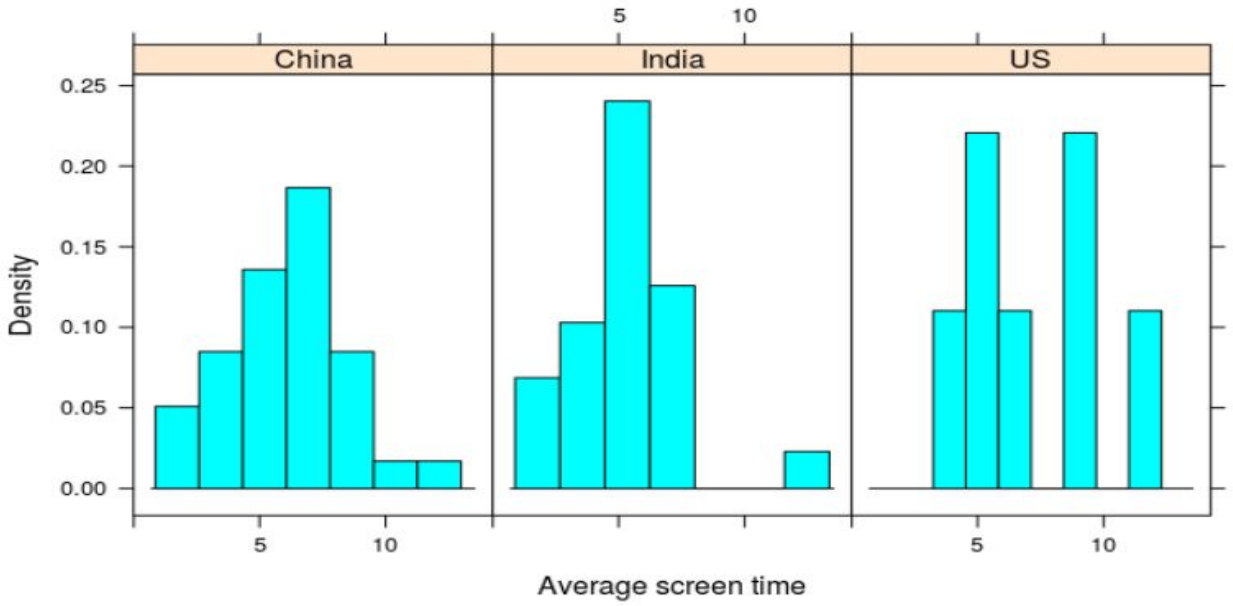
Sample Distribution:

Of 90 Data points, we randomly selected 64 data on which we have plotted normal distribution curve. Here inference is for population which represents northeastern students at Seattle campus. Sampling distribution is a theoretical distribution of a sample statistic. Here in this case , we have taken the sampling distribution so that analyst can hypothesis about the distribution of average screen time

Analyst have plotted mean, upper bound and lower bound line also so as to analyse how close our distribution is to the mean



(viii)Histogram of Sampling Distribution



(ix)Histogram of Survey data

Discussion

Summary and Implications:

In this observational study, analyst compared Chinese and Indian as two sample group to find if there is a difference in the true mean of screen-time between Chinese and Indian population at Northeastern University, Seattle. By collecting data-points on screen-time, most-used app categories from online-survey, analysts were able to characterize smartphone usage by measuring average screen-time for different age groups and different ethnicity groups in the exploratory analysis. The study finds the distribution of sample data collected is skewed to age groups from 18-35. The mean and median of Chinese and Indian groups are fairly close. Through the computation at statistical analysis, analysts find there is no evidence to suggest that the true population mean screen-time for Indian is different than Chinese which cannot reject our null hypothesis.

This study in another way, provides interesting findings may be useful for further study in this phone usage topic. It can be used as a reference for measurement on phone usage. Though this study is not able to analyze most-used app categories further, analyst can work on in the future to find the most-used app category with further data collection. From these findings and further development on this study, analysts will be able to better understand how people spend their screen-time and what may be the influential factors that lead to high screen-time. It ultimately helps app developers to better develop attractive products , scientists to find correlation between productivity and phone usage, and phone users to optimize their time management.

Limitation:

However, this study has its limitations. These findings on smartphone screen-time, app categories between Chinese and Indian are highly based on survey taken by participants themselves. Screen-time, most-used app categories may differ across age and ethnicity in a larger and less-biased environment setting, but this study cannot fully support it due to its narrow scoped setting and potential biased environment. These findings cannot support conclusions on causation since there are so many confounding factors exist. Biases can be avoided with further steps:

Firstly, have a larger scope on the sample to ensure the coverage of sample to population. Secondly, create a well-developed sampling strategy to avoid sampling errors and provide better sampling selection. Thirdly, rule out potential confounding factors for the study. Lastly, Our sample size is small, so it is difficult to infer relationships between Indian And chinese population average screen time.

Next Step:

There can be multiple steps which if taken can make this study more robust

1. We should take large sample which represents our population so that relations can be drawn out of them.
2. Biases should be avoided like there exist data where user might have filled wrong information which can result in misleading.
3. App Categories defined in this study can be different for other groups so there should be standardised way to define category.

References

Aust, F., & Barth, M. (2018). papaja: Create APA manuscripts with R Markdown. Retrieved from https://github.com/crsh/papaja

Dolliver, M. (2019, May 30). US Time Spent with Media 2019.Retrieved from https://www.emarketer.com/content/us-time-spent-with-media-2019.

R Core Team. (2019). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Appendix

1. Code for Test

```
# Installing Packages
install.packages("ggplot2")
install.packages("mosaic")
install.packages("Sleuth3")
library(ggplot2)        # plotting & data
library(dplyr)          # data manipulation
library(mosaic)
library(Sleuth3)
# Loading Survey data in RStudio
Survey_Data <- read.csv("/cloud/project/Survey_Data.csv")
summary(Survey_Data)
# Exploratory Analysis
# We Explored data for App category and its usage based on country to get the
distribution. We got the visibility of most popular Appilcation category
among different countries
table(Survey_Data$App.Category, Survey_Data$Country)
# Here we explored data from different prispective to get the distribution of
3 ethicities among different age group
table(Survey_Data$Age , Survey_Data$Country)
# Average Screen time distribution
ggplot(Survey_Data, aes(x=Survey_Data$Average.Screen.time.last.10.days,
color=Survey_Data$Country)) +
  geom_histogram(fill="white", position="dodge", bins = 10)+
  theme(legend.position="bottom") + labs(x = "Average Screen Time", color =
"Legend: Country wise average screen time")
# Age Group Vs Average Screen time
ggplot(Survey_Data, aes(x = Survey_Data$Age, y =
Survey_Data$Average.Screen.time.last.10.days, color =
factor(Survey_Data$Country))) +
  geom_point() +
```

```
  labs(x = "Age-Group", y = "Average Screen Time", color = "Legend: Country
wise average screen time")+ theme(legend.position = "bottom")
# Sample distribution among Age-group
ggplot(Survey_Data, aes(x = Survey_Data$Age, fill = Survey_Data$Country)) +
  geom_bar() +   labs(x = "Age-Group", y = "No of people",
                      color = "Legend: Country wise count")+
theme(legend.position = "bottom")
# Ethnicities wise Screen time comparison
boxplot( Survey_Data$Average.Screen.time.last.10.days~Survey_Data$Country,
ylab = 'Average Screen Time', xlab = 'Country', main = 'Ethnicitieswise
screen time comparison', col = 'skyblue')
# App category Vs Average screen time : Screen Time here shows its usage
across app categories along with outliers to get more visibility
ggplot(Survey_Data, aes(x = Survey_Data$App.Category, y =
Survey_Data$Average.Screen.time.last.10.days)) +
  geom_boxplot() + guides(fill = TRUE) +
  theme(axis.text.x = element_text(angle = 65, vjust=0.5, hjust=0.5))+ labs(x
= "App Category", y = "Average Screen Time")
pairs(Survey_Data$Average.Screen.time.last.24.hour~Survey_Data$Age, col =
Survey_Data$Country)
# Statistical Analysis
# Sampling data for 32 Indians and Chinese for calculating average screen
time
library(dplyr)
set.seed(0)
India <- filter(Survey_Data, Survey_Data$Country == "India")
India_2 <- India[sample(nrow(India),32),]

China <- filter(Survey_Data, Survey_Data$Country == "China")
China_2 <- China[sample(nrow(China),32),]

sampledata <- rbind(China_2,India_2)
```

```
sampledata
# Summary for Mean & median for sample population
summary(sampledata)
# QQ Plot
qqnorm(Survey_Data$Average.Screen.time.last.10.days)
qqnorm(India_2$Average.Screen.time.last.10.days)
qqnorm(China_2$Average.Screen.time.last.10.days)
# Sample Mean
X_bar_i <- mean(India_2$Average.Screen.time.last.10.days)
X_bar_c <- mean(China_2$Average.Screen.time.last.10.days)
X_bar_i
X_bar_c
# Sample Variance
s_i <- sd(India_2$Average.Screen.time.last.10.days)**2
s_c <- sd(China_2$Average.Screen.time.last.10.days)**2
s_i
s_c
# Sample Size
n_i <- length(India_2$Average.Screen.time.last.10.days)
n_c <- length(China_2$Average.Screen.time.last.10.days)
n_i
n_c
#Null Hypothises
mu <- 0
# T Test
t <- (X_bar_c - X_bar_i - mu)/sqrt((s_i/n_i) + (s_c/n_c))
t
# p-value for two sided upper
two_sided_diff <- pt(q=t, df = min(n_i, n_c)-1, lower.tail = FALSE) * 2
two_sided_diff
Alpha <- 0.05
Confidence_Interval <- 0.95
```

```
# Lower Bound
L_bound <- (X_bar_c - X_bar_i) + (qt(0.025, min(n_i, n_c)-1)* sqrt((s_i/n_i)
+(s_c/n_c)))
L_bound
# Upper Bound
U_bound <- (X_bar_c - X_bar_i) + (qt(0.975, min(n_i, n_c)-1)* sqrt((s_i/n_i)
+(s_c/n_c)))
U_bound
# R built in t-test function
t.test(India_2$Average.Screen.time.last.10.days,
China_2$Average.Screen.time.last.10.days)
#histograms
histogram(~ Survey_Data$Average.Screen.time.last.10.days |
Survey_Data$Country)
histogram (India_2$Average.Screen.time.last.10.days)
histogram(China_2$Average.Screen.time.last.10.days)
# Histogram of the sampling distribution
mu <- mean(Survey_Data$Average.Screen.time.last.10.days)
sd <- sd(Survey_Data$Average.Screen.time.last.10.days)
h <- hist(Survey_Data$Average.Screen.time.last.10.days, xlim = c(0,13))
lb <- mu - 1.96*sd
ub <- mu + 1.96*sd
abline(v = c(mu, lb, ub), lty = 2)
x_axis <-
seq(min(Survey_Data$Average.Screen.time.last.10.days),max(Survey_Data$Average
.Screen.time.last.10.days),length=80)
y_axis <- dnorm(x_axis, mu, sd)*length(x_axis)
lines(x_axis, y_axis, col = "blue")
# T- Test distribution graph
n <- min(n_c, n_i)
X <- seq(-4, 4, .01)
Y <- dt(X, n-1)
```

```
plot(X, Y, type = 'l')
abline(v = c(t, -t),  col = "blue")
# Confidence Interval graph
plot(X, Y, type = 'l')
abline(v = qnorm(0.975), col = "Green")
abline(v = qnorm(0.025), col = "Green")
abline(v = 0, col = "black")
# Difference between normal distribution & T distribution
plot(X,Y,type = 'l')
lines(X,dnorm(X), col = 'yellow')
```

2. Raw Data Set

Country	Age	Average Screen time last 10 days	Top App	App Category
India	24-29	5.3	Game	Game/ Entertainment
India	30-35	3	WhatsApp	Social
China	24-29	5.36	Wechat	Social
China	18-23	3	Messages	Utilities
China	24-29	7	WeChat	Social
US	24-29	5.42	social	Social
China	18-23	7.53	WeChat	Social
China	24-29	9	wechat	Social
China	18-23	9.47	Wechat	Social
China	24-29	6.32	Wechat	Social
China	24-29	4.3	wechat	Social
China	24-29	3.55	Wechat	Social

China	18-23	5.5	WeChat	Social
China	18-23	2.3	Maps	Utilities
US	18-23	8.43	Facebook	Social
India	18-23	1.5	YouTube	Game/ Entertainment
US	30-35	4.23	LinkedIn	Social
India	18-23	3.5	Twitter	Social
India	18-23	12	Instagram	Social
India	24-29	2.04	WhatsApp	Social
India	24-29	7.3	Whatsapp	Social
US	18-23	6	Instagram	Social
India	24-29	5	WhatsApp	Social
India	30-35	6.52	Youtube	Game/ Entertainment
India	18-23	6	YouTube	Game/ Entertainment
India	24-29	1.3	Whatsapp	Social
India	24-29	6.5	WhatsApp	Social
India	24-29	5.23	WhatsApp	Social
US	24-29	12	Gmail	Productivity
India	18-23	5	Instagram	Social
India	18-23	8	Uber	Lifestyle
India	24-29	7.3	Instagram	Social
India	18-23	5	WhatsApp	Social
US	18-23	5.3	WhatsApp	Social

India	18-23	4.3	WhatsApp	Social
India	24-29	5	Whatsapp	Social
India	24-29	4	Twitter	Social
India	18-23	12	What's up	Social
China	24-29	4.75	QQ	Social
India	18-23	4.5	PUBG	Game/ Entertainment
India	18-23	5.5	Netflix	Game/ Entertainment
India	24-29	6	PUBG	Game/ Entertainment
India	24-29	7	You tube	Game/ Entertainment
India	24-29	5	Whatsapp	Social
India	24-29	6.3	Instagram	Social
India	24-29	2.4	LinkedIn	Social
US	24-29	9.37	Whatsapp	Social
India	18-23	5	YouTube	Game/ Entertainment
India	24-29	5.5	TWITTER	Social
India	18-23	2.4	Youtube	Game/ Entertainment
India	18-23	4	Quora	Productivity
India	18-23	4.1	Chrome	Utilities
India	18-23	2.3	What's app	Social
India	18-23	8	Netflix	Game/ Entertainment

India	18-23	8	YouTube	Game/ Entertainment
India	18-23	6.2	Instagram	Social
India	18-23	5	YouTube	Game/ Entertainment
India	18-23	4.3	Pubg	Game/ Entertainment
India	18-23	6	Youtube	Game/ Entertainment
India	18-23	5.1	LinkedIn	Social
India	18-23	6	Whatsapp	Social
India	24-29	4	Instagram	Social
India	24-29	5.5	Facebook	Social
India	24-29	6.5	Instagram	Social
India	24-29	7	WhatsApp	Social
India	30-35	6	Whatsapp	Social
India	30-35	5	LinkedIn	Social
India	30-35	4.3	Swiggy	Lifestyle
China	30-35	4.4	Weibo	Social
China	24-29	6.2	Wechat	Social
China	18-23	5	QQ	Social
China	18-23	6.42	Sina Weibo	Social
China	18-23	5.93	Wechat	Social
China	30-35	5.3	Wechat	Social
China	18-23	9.44	Wechat	Social
China	24-29	2	Wechat	Social

China	47 and above	3	Wechat	Social
China	18-23	6.8	Wechat	Social
China	18-23	6.6	Wechat	Social
China	18-23	8.21	Wechat	Social
China	18-23	6.2	Wechat	Social
China	18-23	4	Wechat	Social
China	18-23	7.11	Wechat	Social
China	42-47	10.5	Iqiyi	Game/ Entertainment
China	18-23	5.07	google chrome	Utilities
China	24-29	12.4	Weichat	Social
China	24-29	9	Wechat	Social
China	24-29	7	Bilibili	Game/ Entertainment
China	18-23	2.2	Honor of Kings	Game/ Entertainment
China	30-35	6.3	Dingding	Productivity