QUANTITATIVE ANALYSIS OF RESULTS OF A SMARTPHONE SURVEY

1. BACKGROUND

The report aims to summarize the findings of the smart phone survey conducted on behalf of the company. The quantitative survey has been used to obtain insights on the features most preferred by customers as well as the factors that influence them. Features tested for include phone usage statistics, price points and likelihood of purchase. The relevant statistical and inferential tools and best practices have been applied or the same.

2. BIAS AND ERRORS

- While the data indicates that participants from nearly 193 countries have been surveyed, it was found that about 75% of the participants were from only 10 of the countries. This could give rise to **sampling** errors as the data is heavily biased in terms of worldwide representation. Similarly 75% of participants owned an Android device, again a potential marker for sampling errors.
- Further, roughly 88% of participants were between 18-54 years of age. There is not enough representation of older categories, which could give rise to another set of sampling errors.
- Another problematic area was the length of the survey. There were 23 questions in total, which is far more than the recommended number of 10-15. This may have led to respondent fatigue towards the end, and resulted in **non-sampling** errors.
- Grouping of age into categories was another pitfall, as age in inherently continuous. Treating it as individually discrete would have been more appropriate than grouping so. Furthermore, this leads to inaccuracies in causal analysis as we need to encode categories and in the process lose real time differences between values.

3. DATA ELEMENTS CHOSEN FOR ANALYSIS

Demographic

- Age of participant
- Highest educational qualification

Nominal

- Operating system of current phone
- Country of residence

Ordinal

- How often participant uses their phone for given activities
- Current career of participant

Interval

- How influential are the given factors in choosing a new phone
- Star ratings for smart phone activities

4. CENTRAL TENDENCIES AND DISPERSION

4.1 DEMOGRAPHIC DATA

A. Age of participants

AGE	NUMBER OF PARTICIPANTS	%
18 - 24	255	19.36%
25 - 34	390	29.61%
35 - 44	263	19.97%
45 - 54	254	19.29%
55 - 64	41	3.11%
65 - 74	25	1.90%
75 - 84	18	1.37%
85 or older	7	0.53%
Under 18	64	4.86%
TOTAL	**1317**	**100.00%**

B. Educational Qualification

EDUCATIONAL QUALIFICATION	PARTICIPANTS	%
Associate degree in college (2-year)	128	9.72%
Bachelor's degree in college (4-year)	436	33.11%
Doctoral degree	41	3.11%
High school graduate	49	3.72%
Less than high school degree	31	2.35%
Master's degree	426	32.35%
Professional degree (JD, MD)	40	3.04%
Some college but no degree	166	12.60%
GRAND TOTAL	**1317**	**100.00%**

C. Inferences

- Around **88%** of participants were between **18-54 years** of age. There is a poor representation of under 18 year olds and those over 54. This may lead to sampling errors in central tendencies, association tests and causal analysis.
- In terms of educational qualification, the extremes have the poorest representation. Candidates with less than high school education are one extreme and doctoral candidates, the other.

4.2 NOMINAL DATA

A. Operating system of participant's current phone

OS	NO. Of PARTICIPANTS	%
Android	914	75.91%
iOS (Apple)	260	21.59%
Other	16	1.33%
Windows	14	1.16%
GRAND TOTAL	**1204**	**100.00%**

B. Countries

The countries listed below are only those (out of a total of 193) having greater than 10 participants each

COUNTRY NAME	PARTICIPANTS	%
Brazil	45	3.41686
Canada	76	5.77069
China	109	8.27639
India	126	9.5672
Indonesia	36	2.73349
Mexico	66	5.01139
Nigeria	16	1.21488
Spain	43	3.265
United Kingdom of Great Britain and Northern Ireland	41	3.11314
United States of America	437	33.1815
TOTAL	**995/1317**	**75.5505**

C. Inferences

- We can see that nearly **75%** of those surveyed were users of the **android** operating system. This represents a heavy bias in the sample and can give rise results specific to that group.
- The country-wise data indicates that **75%** of participants come from just **10 out of the 193** countries surveyed. Further, about **33%** or roughly one-third of participants were from just the **USA**. This could give rise to sampling errors.

4.3 ORDINAL DATA

A. How often does the participant use their phone for the given activities

Scale : 1 - Almost every hour , 2 - A few times a day , 3 - A few times a week ,

4 - A few times a month , 5 - Seldom or never

	MEASURES OF CENTRAL TENDENCY						
ACTIVITIES	**AVERAGE**	**COUNT**	**MIN**	**MAX**	**VARIANCE**	**STDEV**	**MODE**
Phone/Video Calls	2.34	1204	1	5	1.13	1.07	2.00
Texting	1.96	1204	1	5	1.19	1.09	1.00
Email/Calendar	1.90	1204	1	5	1.13	1.07	1.00
Gathering Information	2.43	1204	1	5	1.05	1.02	2.00
Social Media	2.99	1204	1	5	1.99	1.41	2.00
Games	3.89	1204	1	5	1.50	1.23	5.00
Taking pictures/videos	2.59	1204	1	5	1.08	1.04	3.00
Maps and navigation	3.28	1204	1	5	1.18	1.09	3.00
Consuming Media	3.00	1204	1	5	1.53	1.24	2.00
Health/Fitness	4.32	1204	1	5	1.48	1.22	5.00
Other	3.64	84	2	5	1.10	1.05	4.00

The below table shows percentage of participants who chose a given rating, for each smart phone feature

	RATING				
FEATURES	**1**	**2**	**3**	**4**	**5**
Phone/Video Calls	22.42%	36.63%	25.89%	11.68%	3.37%
Texting	43.16%	32.32%	15.16%	6.42%	2.95%
Email/Calendar	43.79%	35.47%	10.84%	6.00%	3.89%
Gathering Information	17.68%	43.05%	24.00%	11.58%	3.68%
Social Media	17.89%	24.11%	19.58%	15.79%	22.63%
Games	3.05%	14.84%	16.00%	21.16%	44.95%
Taking pictures/videos	14.53%	33.26%	34.32%	14.11%	3.79%
Maps and navigation	4.42%	19.37%	35.58%	25.58%	15.05%
Consuming Media	12.00%	27.37%	25.47%	20.32%	14.84%
Health/Fitness	5.16%	7.89%	7.79%	7.89%	71.26%

Inferences

- Based on the mode, we can see that maximum candidates seem to **seldom/never** use their phones for **games** and **tracking health and fitness**. The percentage distribution of rating also agrees with this observation as nearly 71% agreed that they seldom use their phone for health apps and around 60% indicated that they use their device for games occasionally to rarely.
- The mode seems to say that most participants seem to use their phones for **texting and emails** almost **every hour**. This agrees well with the distribution heavily pointing towards maximum usage of these features.
- The mode for calling, searching for information and social media indicates that they are used a few times daily. However, the distribution for calling is almost evenly spread across hourly, daily and weekly usage so we cannot conclusively rely on the mode. The same applies for social media as well on account of reasonably even percentage spreads.
- The rating for information gathering is uniform across both measurements and the mode and distribution agree well. Maximum participants seem to search the internet a few times a day.
- We can see that **dispersion** is between **1-1.5** for most criteria, indicating a relatively **high** dispersion of rating values.

B. What best describes your current career ?

Scale

1- **Student**

2- **Entry-level to a few years' experience in my work**

3- **Experienced specialist or professional**

4- **Manager**

5- **Consultant**

6- **Senior management or leadership**

CAREER	NO. OF PARTICIPANTS	%
I'm a consultant	177	13.44%
I'm a manager	227	17.24%
I'm a student	274	20.80%
I'm an experienced specialist or professional	295	22.40%
I'm entry-level to a few years' experience in my work	276	20.96%
I'm in senior management or leadership	68	5.16%
GRAND TOTAL	**1317**	**100.00%**

Inferences

There seems to be a relatively **balanced distribution** of experienced professionals, fresh graduates and students.

4.4 INTERVAL DATA

A. Influence of given factors when participant purchases a new phone

Scale

1- **Not Influential**

2- **Somewhat Influential**

3- **Very Influential**

	MEASURES OF CENTRAL TENDENCY						
FACTORS	**COUNT**	**AVERAGE**	**MAX**	**MIN**	**VARIANCE**	**STDEV**	**MODE**
Price	1283	2.49	3	1	0.47	0.69	3
Advertising	1277	1.57	3	1	0.51	0.71	1
Brand	1277	2.12	3	1	0.57	0.75	2
Seeing it in the store	1277	1.77	3	1	0.44	0.66	2
Family	1277	1.76	3	1	0.44	0.67	2
Expert Reviews	1277	1.56	3	1	0.50	0.71	1
Buyer Reviews	1283	1.76	3	1	0.48	0.69	2
Friends	1277	2.07	3	1	0.50	0.71	2
Others	99	1.89	3	1	0.73	0.86	1

The below table shows the percentage of participants who chose a given rating, for each smart phone feature

	RATINGS		
ATTRIBUTES	**1**	**2**	**3**
Price	11.15%	28.76%	60.09%
Advertising	56.30%	30.54%	13.16%
Brand	23.18%	41.74%	35.08%
Seeing it in the store	36.10%	50.67%	13.23%
Family	37.12%	49.80%	13.08%
Expert Reviews	56.85%	30.38%	12.76%
Buyer Reviews	38.82%	46.61%	14.58%
Friends	21.93%	49.26%	28.82%

Inferences

- It can be observed that **price** appears to be the **most influential parameter** while considering a new phone, according to most candidates, as indicated by the mode. The rating distribution indicates **60%** of people believe it is very influential, in agreement with the mode.
- Majority of the people seem to feel that **Advertising** and **expert reviews** do not influence their purchase decisions. The mode for both is 1 i.e. **Not Influential** and this matches the percentage ratings indicating that 56% gave a rating of 1 for each of the two.
- The mode rating for **brand** is 2 i.e. Somewhat influential, but the distribution is spread out with nearly **42%** saying it's **somewhat influential** and **35%** saying it's **very influential**. This seems to indicate that brand is an **important factor** while making purchase decisions.
- Seeing it in the store, buyer reviews, family and friends all appear to be somewhat influential, as felt by a majority of people. This observation is consistent across both tables. While opinions of friends seems to lean towards very influential, as witnessed by the percentages, family, buyer reviews and seeing it in the store lean majorly towards being uninfluential.
- Standard deviation is **below 1** and is indicative of relatively **low dispersion** with respect to the mean rating for influence.

B. Star ratings for importance of different phone activities

Scale

1 - **Unimportant**

2 - **Minimally important**

3 - **Neutral**

4 - **Somewhat important**

5 - **Most Important**

	MEASURES OF CENTRAL TENDENCY						
FEATURES	**COUNT**	**AVERAGE**	**MAX**	**MIN**	**VARIANCE**	**STDEV**	**MODE**
Calls	1204	3.73	5	1	2.12	1.46	5
Texting	1204	3.97	5	1	1.76	1.33	5
Email/C	1198	3.98	5	1	1.70	1.30	5
Information	1204	3.49	5	1	2.19	1.48	5
SM	1198	2.96	5	1	2.47	1.57	1
Games	1204	2.33	5	1	2.09	1.45	1
Pictures	1204	3.50	5	1	2.15	1.47	5
GPS Nav	1204	2.89	5	1	2.42	1.56	1
Media	1204	2.98	5	1	2.36	1.54	1
Health/Fitness	1198	2.26	5	1	1.81	1.35	1
Others	114	2.84	5	1	2.36	1.54	1

The below table shows the percentage of participants who chose a given rating, for each smart phone feature

	RATINGS				
FEATURES	**1**	**2**	**3**	**4**	**5**
Calls	14.45%	7.64%	13.29%	19.35%	45.27%
Texting	10.38%	5.40%	10.80%	23.50%	49.92%
Email/C	9.10%	6.01%	13.27%	21.12%	50.50%
Information	16.20%	12.29%	14.37%	21.10%	36.05%
SM	28.63%	14.11%	15.53%	16.19%	25.54%
Games	42.94%	18.60%	13.12%	12.79%	12.54%
Pictures	16.03%	10.96%	15.61%	21.43%	35.96%
GPS Nav	29.07%	16.03%	15.78%	15.28%	23.84%
Media	26.25%	15.61%	16.94%	16.61%	24.58%
Health/Fitness	42.40%	19.03%	17.36%	12.60%	8.60%

Inferences

- **Calls, texting and emails** have a mode of 5, indicating that maximum participants felt them to be the **most important**. This can also be observed in the second table where **more than 50%** of participants gave them a rating between **4-5**.
- While the mode for **information gathering and taking pictures** is 5, the distribution is spread out between 4-5 for both. This tells us that candidates appear to attach a **significant** amount of importance to both.
- The mode for **social media** use is 1. But the second table tells us that only 28% feel it is unimportant. Another 25% actually rated it as most important, showcasing highly varied opinions. **No clear conclusion** can be drawn from this observation, except that the variability in term is high. A similar distribution is observed for **consumption of media** and **GPS** navigation as well.
- Majority of participants seem to feel **games** and **health apps** are **unimportant or minimally important**. The mode rating for both is 1, in agreement with the 42% of people rating them as unimportant.
- Standard deviation is between **1.3-1.5** for all values, which is quite high. It indicates of **high** levels of **dispersion** about the mean.

5. ASSOCIATION ANALYSIS – CHI SQUARE TEST

- For this test, the **variables** considered are **Age of participants** and **how often a participant uses their phone for accessing social media**.
- **Null hypothesis** : The two variables are independent i.e. there is not statistical correlation between them.
- **Alternative hypothesis** : The two variables are associated and a dependence exists.

5.1 Condensed version of given data

AGE	FREQUENCY OF SOCIAL MEDIA USAGE
25 - 34	1
55 - 64	5
45 - 54	3
35 - 44	2
… (1317 values)	… (1317 values)
25 – 34	3

5.2 Summarization of Observed Data

	NO. OF PARTICIPANTS PER RATING					
AGE CATEGORIES	**1**	**2**	**3**	**4**	**5**	**TOTAL**
18 - 24	32	40	49	51	43	215
25 - 34	82	93	68	72	70	385
35 - 44	43	78	47	32	57	257
45 - 54	47	42	64	39	50	242
55 - 64	4	6	5	4	19	38
65 - 74	9	3	6	1	3	22
75 - 84	3	3	5	-	5	16
85 or older	1	2	2	-	1	6
Under 18	3	6	3	3	8	23
TOTAL	**224**	**273**	**249**	**202**	**256**	**1204**

5.3 Expected values

	RATINGS				
AGE CATEGORIES	**1**	**2**	**3**	**4**	**5**
18 - 24	40	48.75	44.464286	36.07142857	45.71428571
25 - 34	71.62790698	87.29651	79.622093	64.59302326	81.86046512
35 - 44	47.81395349	58.27326	53.150332	43.1179402	54.64451827
45 - 54	45.02325581	54.87209	50.048173	40.6013289	51.4551495
55 - 64	7.069767442	8.616279	7.858804	6.375415282	8.079734219
65 - 74	4.093023256	4.988372	4.5498339	3.6910299	4.677740864
75 - 84	2.976744186	3.627907	3.3089701	2.684385382	3.401993355
85 or older	1.11627907	1.360465	1.2408638	1.006644518	1.275747508
Under 18	4.279069767	5.215116	4.7566445	3.858803987	4.890365449

5.4 Results

Chi square value = $\sum \frac{(O_i - E_i)^2}{E_i}$

= **72.1362**

Degrees of freedom = (R-1)(C-1) = (9-1)(5-1) = 8*4 = **32**

Critical value (from chi square test table) at significance level of 0.05 = **20.72**

p-Value = 0.000063 ≈ **0.00** (from online p-value calculator)

Inferences

- From the above calculations we can see that the **p-value** is almost zero and is **less than** the significance level of **0.05**
- Further, our **Chi square** value = **72.13** is **greater** than the critical value = **20.7**

Hence we can conclude that our null hypothesis is false and can be rejected. This indicates that there exists an association or statistical correlation between age and how often a participant uses social media on their phone

6. REGRESSION

- The **variables** chosen for causal analysis are the **operating system of current phone of participant** and **how influential the price is when choosing a new phone**.
- Association analysis using chi square test gave a **p-value = 0.000016 ≈ 0.00** at 0.05 level of significance, indicating that the variables are **associated**.
- The idea is to test whether and how much a participant's rating for influence of price depends on the phone they currently own.
- **Independent variable (X) :** OS of the current phone owned by the participant
- **Dependant variable (Y) :** Influence of price while purchasing a phone

The regression equation can be formed as follows :

Influence of Price = Current Phone * coefficient + Intercept

- **Null hypothesis :** The slope for the intercept is zero, meaning that there is no significant linear relationship between the variables
- **Alternative hypothesis :** The slope for the intercept is not zero, meaning they are significantly related

****The values for the responses for current phones were re-encoded for simplicity as follows : iOS -1 , Android – 2, Windows – 3, Others – 4.**

Regression results

REGRESSION STATISTICS	
Multiple R	0.257154
R Square	0.066128
Adjusted R Square	0.06272
Standard Error	0.620554

	COEFFICIENTS	STANDARD ERROR	T STAT	P-VALUE
INTERCEPT	1.859842095	0.151649839	12.26406	7.43E-28
CURRENT PHONE	0.373034618	0.084688582	4.404781	1.52E-05

Interpretation of results

- Firstly the value of **adjusted R-squared** indicates that only **6.6% variation** in dependent variable (influence of price) is explained by the independent variable (OS of phone).
- This may seem at odds with the chi square test, which indicated an association between the variables. A low R-Squared value and a low p-value indicate that the model doesn't explain much of variation of the data but it is still very much significant
- This tells us that the OS of the current phone of the participant, even though significant, is not accounting for much of the mean of the influence of price.
- The **t-value is 4.404**, which is **greater than 1.96**

Hence at a confidence level of 95% we can reject the null hypothesis. This means that current phone OS of the participant does affect how influential they think price is

- However, our **t-value** is quite **small**, which indicates that while the alternative hypothesis is true, the **confidence** in our predictor variable(X) is quite **low**. This is corroborated by the low values of R and Adjusted R-squared values, as only a small proportion of variance in Y is explained by X.

From the coefficient and intercept values, our regression equation now becomes :

Influence of Price = Current phone * 0.373 + 1.879

The basic meaning of this regression is that as we move from iOS to Android to other operating systems, the influence of price increases.

This is a reasonable assumption to make since iPhones are generally costlier, and the people owning them may be affluent enough to not consider price as an influential factor, while buying a new phone. As we move to cheaper phones, the influence of price increases.

****It is to be noted that linear regression is best done with continuous data since categorical and ordinal variables often tend to have a unit difference between criteria and this does not give reliable results in most cases. It would be better to go for ordinal logistic regression in such cases.**