SCHOOL OF ENGINEERING
TAYLOR'S UNIVERSITY

ENG 61103

Data Measurement, Analysis and Experimental Design

Assignment Report

Module Coordinator

Mr. Mohd. Hardie Hidayat bin Mohyi

Name	Student ID	Signature
Hafis Murty	0326357	
Wong Yi-San	0326426	
Eugene Gow Jun Yi	0326755	
Samuel Tee Jian Yang	0331293	
Muhammad Zharif Nichol	0331279	

Due Date : 17 November 2019

Date of Submission : 16 November 2019

Content Table

INTRODUCTION

It is common knowledge that Malaysia is host to the highest rate of obesity and overweight civilians in Asia with 64% of male and 65% of female population being either obese or overweight. A consequence of this is that it can lead to health issues such as diabetes, chronic heart conditions and many more. Diabetes amongst adults aged 18 and above has increased from 11.6% to 17.5% over a period of 9 years from 2006 to 2015. **[1]**

Our team, Team Solar Fam, consists of 5 people.

- Leader: Arielle Wong
- Member 1: Hafis Murty
- Member 2: Eugene Gow
- Member 3: Samuel Tee
- Member 4: Muhammad Zharif Nichol

Figure 1: Team Photo

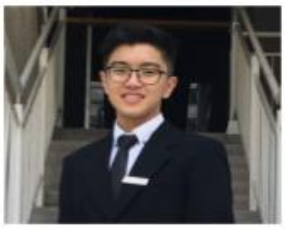
From Left to Right: Samuel, Muhammad Zharif, Hafis, Eugene and Arielle.

ORGANIZATION CHART

LITERATURE REVIEW
ARIELLE WONG

RESULTS & CALCULATIONS
HAFIS MURTY

METHODOLOGY
EUGENE GOW



DISCUSSION & CONCLUSION
SAMUEL TEE

INTRODUCTION
ZHARIF NICHOL

As a team, our main objective is to raise awareness to the frightful state of the Malaysian population and let the people know that it is important to stay active in today's world. We will be providing a statistical analysis by using the ANOVA 1 Way method. We chose to conduct this experiment from 9am to 5pm which would portray a time frame that would follow the average working Malaysian. In addition, we are using an app available from the Google Play Store or the Apple Store called, Sehat Weight Loss Coach, to show how many steps per day the average Malaysian does. This app will calculate how many steps the candidate will take within the time period of a day-today basis by using the accelerometer within the candidate's phone. We chose 4 candidates from 4 certain age groups to show how active these different age groups are on a day-to-day basis. We conducted this experiment over a period of 2 weeks.

We have come up with an initial hypothesis, h0 that the average Malaysian would meet the average daily steps taken by their respective age groups according to studies previously conducted.

LITERATURE REVIEW

Figure 2: Official application icon/logo

Sehat is an application under the category of "Health and Fitness" available in Google Play Store for Android devices. It is a comprehensive lifestyle tracker that aims to help its users to eat healthy and lose weight using their calorie counter and step tracker, among many other functions. It is an alpha version and caters for the user's country, meaning that the data should be more accurate. For step tracking, Sehat utilises Google Fit for Android users and Apple Health for Apple users for the purpose of this experiment, the step tracker function was utilised to obtain data on the number of steps taken in a day by individuals of various age groups. [2]

Figure 2 shows the homepage of the application which summarises all the data that has been collected in that particular day according to different functions. The third row displays the number of steps taken out of the recommended 10,000 steps a day. In the case where the number of steps taken surpasses the recommended 10,000 steps, the application will still continue to track and display the exact number of steps taken.

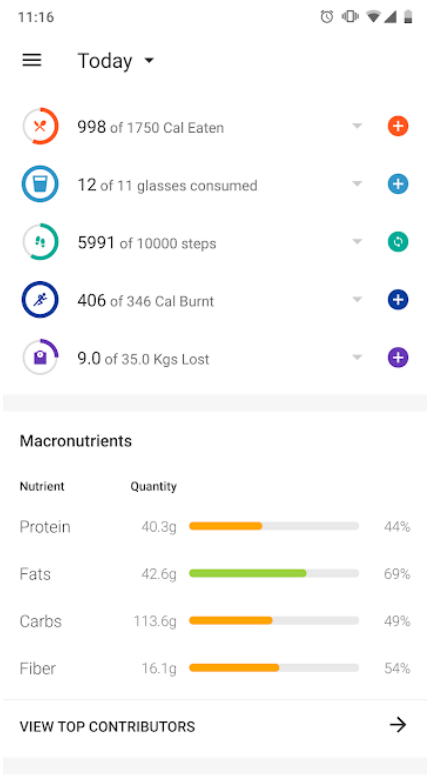


Figure 3: Homepage of Sehat application

In order to track steps more accurately, the device has to be directly or indirectly attached to the user's body during motion, whether it may be in the user's pocket, hands, etc. The direction of the device (straight or upside down) does not matter. This kind of application also works without GPS unless there are other features in the application such as displaying the map of the track you followed. [3]

Since phones usually have accelerometer sensors built in it, the application is able to monitor changes in accelerometer and conclude if a step is taken. The accelerometer supplies information on the angle at which the device is being held, direction, speed at which it is moved and gravity. The accelerometer provides these values whenever they change, along with the time of occurrence. The application stores these values in its database and looks for a pattern. [3]

In more detail, the accelerometer provides three values; x-axis, y-axis and z-axis. When the user takes the first step, the x-axis value increases at a rate depending on the force at which the user starts to move. Meanwhile, the y-axis value, which determines the relative change in the

height of the device along with the force, increases when the step rises from the ground and decreases when the step touches the ground. Based on the pattern, a step is detected. In terms of accuracy, the values are close to the actual count but activities like talking on the phone or playing games may increase the step count. Hence, it is advisable to start and end the application at the appropriate times; right before and immediately after walking. [3]

As mentioned, this experiment aims to measure the number of steps taken in a day for different age groups using an application. As of now, there are a fair number of existing research papers that have similar topic although not specifically based on Malaysians or Southeast Asians. Most of the research papers center around the usage of pedometers, although there is a 2009 study that focuses on accelerometer-determined steps per day in US adults [4]. From Figure 4, the table indicates the mean wear time, uncensored and censored steps per day. Censored steps were obtained from the uncensored steps after excluding steps that were taken at an intensity of less than 500 activity counts per minute. On average per day, US adults took 9676 ± 107 uncensored steps or 6540 ± 106 censored steps. Thus, the research paper concludes that the results indicate that US adults take approximately 10,000 uncensored steps in a day based on an accelerometer. However, this may be improbable due to several limitations such as unstandardized cut-off-point values (intensity of step) and the uncertainty of the number of days for the accelerometer to be worn to achieve a reliable estimate of the habitual activity (walking). Hence, the research paper's use of less than 500 activity count per minute threshold remains valid until a better conversion factor exists for translating accelerometer-determined steps. [4]

On a side note, there is even a 2018 study on the validity of applications to measure steps. Although the research paper focuses on specific applications like iPhone SE, Garmin Vivofit 2, etc., the research concludes that there are applications that can be used as accurate tools for step counting in different age groups and walking conditions (slow, etc.). [5]

TABLE 2. Mean accelerometer wear time, uncensored steps per day, and censored steps per day in 2005–2006 NHANES adults (20 to 80+ yr).

	N	Mean (SE)	95% CI
Wear time (h)[a]	3744	14.0 (0.1)	13.9–14.2
Men	1781	14.3 (0.1)	14.1–14.4
Women	1963	13.8 (0.1)	13.7–14.0
Uncensored steps	3744	9676 (107)	9449–9903
Men	1781	10,578 (134)	10,293–10,863
Women	1963	8882 (124)	8618–9147
Censored steps[b]	3744	6540 (1060)	6315–6766
Men	1781	7431 (129)	7156–7706
Women	1963	5756 (120)	5501–6011

[a] Days with less than 10 h wear time eliminated.
[b] Eliminates any step where accelerometer activity counts per minute are <500.

Figure 4: Table of results extracted from research paper [4]

METHODOLOGY

We decided to divide the general population into 4 main age groups namely Children (7-12), Teens (13-20), Adults (21-64) and Elderly (65 >). The next thing we agreed on is what is the appropriate time to conduct this experiment which is from 9am till 5pm. Using the phone app "Sehat Weight Loss Coach", we made our candidates keep the phone with them all the time from the start of the experiment (9pm) till the end of the experiment (5pm).

Before starting the experiment, we took down the number of steps of that candidate before 9am if the app has already recorded the number of steps prior to the start of the experiment. By 5pm, we noted the number of steps of that candidate for the day and deducted the number of steps taken by candidate before the start of the experiment to get the net number of steps by the candidate from 9am to 5pm.

We managed to get data from 4 different people from each age group (i.e 4 children, 4 teens, 4 adults, 4 elderly) for us to be able to get an average reading to represent the age group. After obtaining all the data, it is compiled and tabulated into a table format as shown below.

Table 1: Data table constructed for the experiment

Age Group	No. of Candidates			
	Candidate 1	**Candidate 2**	**Candidate 3**	**Candidate 4**
Children (7-12)				
Teens (13-20)				
Adults (21-64)				
Elderly (65 >)				

We considered 9am to 5pm as an appropriate time to conduct the experiment was because we had found out that most of the candidates of all age groups will perform their daily/regular exercise before 9am as well as after 5pm. Some of the sports activities such as playing badminton, tennis or swimming will not allow the candidates to carry their smartphones with them. Thus, it will be fair only by taking the results from 9am to 5pm. Besides that, we did not take into account ages below 7 years old because it is hard to monitor infants and toddlers as well as getting them to cooperate in getting accurate results for this experiment. Also, we had chosen this application to obtain the results as it is user friendly with its software interface that is well recognised and easy to use. This application also has good reviews on the Google Play store where it had helped many users to record their daily steps as well as to keep track of their calories efficiently. Hence, we considered this application to be a reliable one since it was recommended by many users and even falls on the recommended list on the Google Play Store.

Throughout the whole process of taking down the results, a few **calibration** steps were taken to improve the accuracy and precision of the results.

This includes using the same smartphone as well as the same application to take down the readings for all the age group. The reason being of not using a different smartphone even with a same application is because different smartphones might have different orientation sensors to determine the location of the device. Hence, this might affect the consistency of the readings taken.

Secondly, we had made sure that the smartphone was kept with the candidates at all times. In this case, there might be a problem for candidates aged 13 and below because the size of the smartphone is too large to fit in some of the children's pocket. In order to overcome this issue, we had provided the candidates who were facing this issue with a smartphone neck holder or a lanyard holder for them. This ensured the smartphone was kept with them throughout the whole time even for those who were forgetful of where they had left their smartphone.

Besides, the readings for all age groups were taken from 4 different individuals instead of 1 or 2 individuals. This is because different individuals have different lifestyles which will determine whether the individual is active or less active. Different jobs will be another factor as well to determine whether the job scope requires the individual to move around or to sit in the office for most times. Consequently, it will be better for us to obtain results from more candidates which allows us to obtain an average reading that contributes to a more accurate result.

Furthermore, we also suggested the candidates to take down the number of steps before 9am and after 5pm. At the end of the day, those number of steps taken will then be deducted from the number of steps taken from 9am to 5pm. This will allow us to obtain a net number of steps only during the experiment period.

Another optional precaution step where we had also considered which was to set an alarm or reminder on the smartphone to notify the candidates that the experiment period had just started or ended. This allows the candidates to only track the number of steps during the experiment period.

RESULTS AND DISCUSSION

The data collected from the Google Fit and Sehat application which displays the number of footsteps taken by each Malaysian candidate during the experiment period (9am-5pm) were then tabulated as shown in **Table 2**. Some examples of the data collected based on the different age groups selected are as shown in the figures below. By conducting an ANOVA 1 way (Analysis of Variance) test, we were able to further analyse all the data in more detail. All the calculations are shown below and the Residual and Normal Assumption graphs were also obtained based on the results calculated. The results were then analysed in the discussion section that follows. A conclusion was also drawn from this experiment based on the initial hypothesis set by the team, which were known as $h_o \; and \; h_1$. The hypothesis made were:

h_o = On average, the normal human being (Malaysian) takes about 6000 steps a day regardless of age groups

h_1 = On average, the normal human being (Malaysian) does not take about 6000 steps a day due to age groups

Table 2: Results Obtained

Age Group	**No. of trials**				**Average**	**Total**
	Trial 1	**Trail 2**	**Trial 3**	**Trial 4**		
Children (7-12)	6201	6853	6037	5947	6259.5	25038
Teens (13-20)	7791	7465	7561	7249	7516.5	30066
Adults (21-64)	6000	6512	5787	6243	6135.5	24542
Elderly (65 >)	3427	3507	3222	3108	3316.0	13264

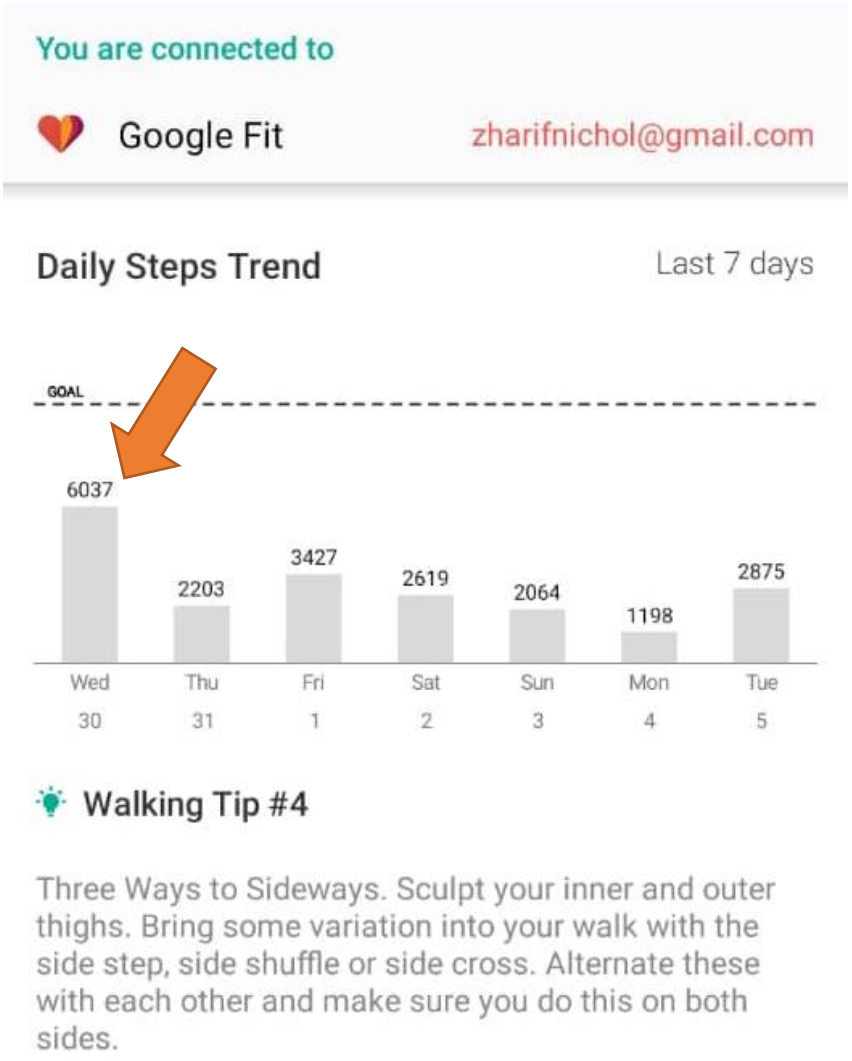


Figure 5: Example Data for Number of Steps of Children (Arrow)

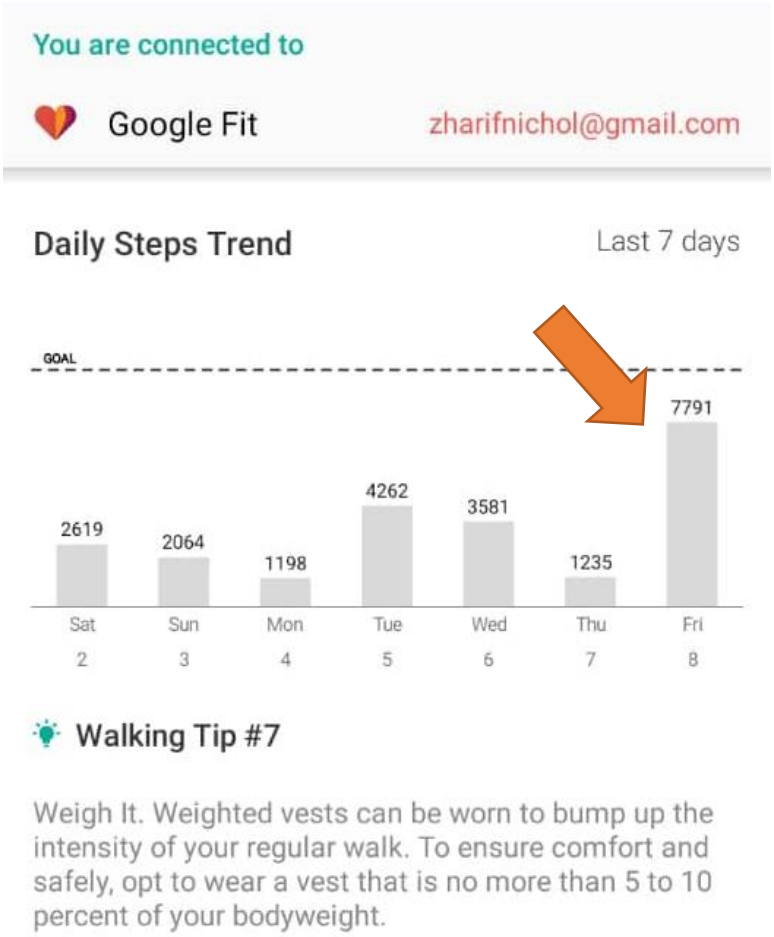


Figure 6: Example Data for Number of Steps of Teens (Arrow)

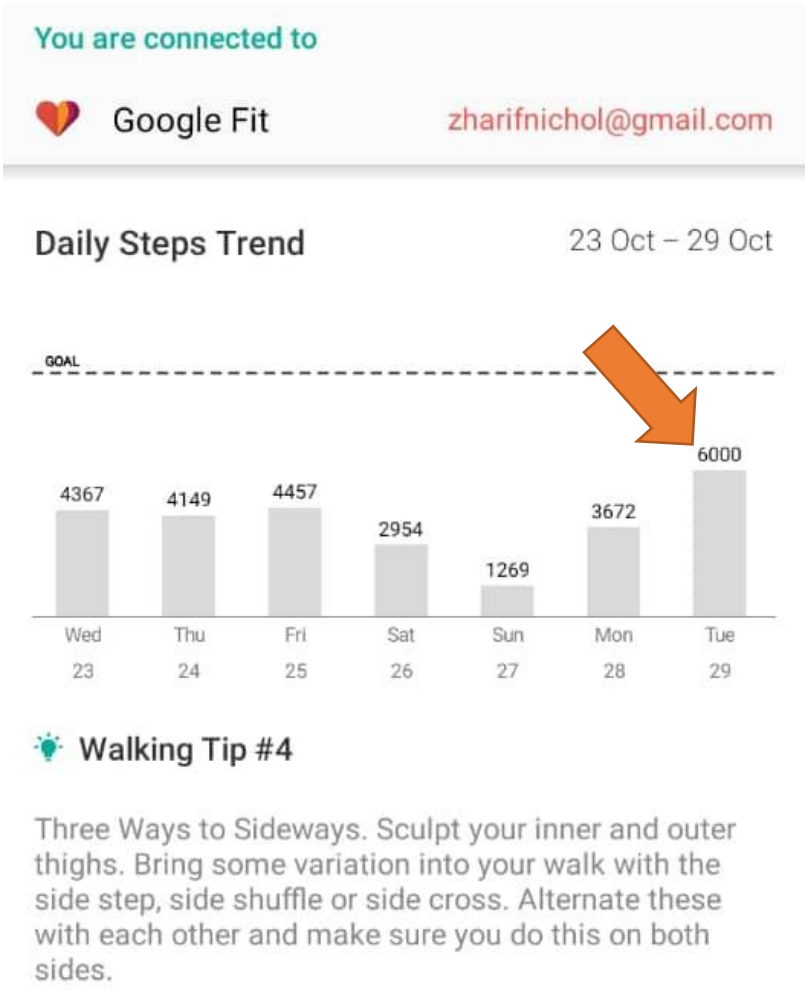


Figure 7: Example Data for Number of Steps of Adults (Arrow)

Table Calculations

y_i, Total number of steps of children = 6201 + 6853 + 6037 + 5947

= 25038

Average number of steps for children = $\frac{6201 + 6853 + 6037 + 5947}{4}$

= 6259.5

y_i, Total number of steps of teens = 7791 + 7465 + 7561 + 7249

= 30066

Average number of steps for teens = $\frac{7791 + 7465 + 7561 + 7249}{4}$

= 7516.5

y_i, Total number of steps of adults = 6000 + 6512 + 5787 + 6243

= 24550

Average number of steps for adults = $\frac{6000 + 6512 + 5787 + 6243}{4}$

= 6135.5

y_i, Total number of steps of elderly = 3427 + 3507 + 3222 + 3108

= 13264

Average number of steps for elderly = $\frac{3427 + 3507 + 3222 + 3108}{4}$

= 3316.0

ANOVA 1 Way Calculations

Table 3: ANOVA 1 Way Formulas

Analysis of Variance Table (One Way, Fixed Model)

Table 3-3 The Analysis of Variance Table for the Single-Factor, Fixed Effects Model

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Between treatments	$SS_{\text{Treatments}} = \frac{1}{n}\sum_{i=1}^{a} y_{i.}^2 - \frac{y_{..}^2}{N}$	$a-1$	$MS_{\text{Treatments}} = \frac{SS_{\text{Treatments}}}{a-1}$	$F_0 = \frac{MS_{\text{Treatments}}}{MS_E}$
Error (within treatments)	$SS_E = SS_T - SS_{\text{Treatments}}$	$N-a$	$MS_E = \frac{SS_E}{a(n-1)}$	
Total	$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}^2 - \frac{y_{..}^2}{N}$	$N-1$		

Values from data:

a (level) = 4

n (repetition) = 4

N (all samples) = 4 × 4 = 16

y (grand total number of steps) = 25038 + 30066 + 24542 + 13264

= 92910

$SS_{Treatments}$ Calculations

$\frac{1}{n}\sum_{i=1}^{a} y_i^2 - \frac{y^2}{N}$

$= \frac{1}{4}(25038^2 + 30066^2 + 24542^2 + 13264^2) - \frac{92910^2}{16}$

= 37760558.75

SS_T Calculations

$\sum_{i=1}^{a}\sum_{j=1}^{n} y_i^2 - \frac{y^2}{N}$

$= (6201^2 + 6853^2 + 6037^2 + 5947^2 + 7791^2 + 7465^2 + 7561^2 + 7249^2 + 6000^2 +$

$6512^2 + 5787^2 + 6243^2 + 3427^2 + 3507^2 + 3222^2 + 3108^2) - \frac{92910^2}{16}$

= 38808947.75

SS_E Calculations

SS_T - $SS_{Treatments}$

= 38808947.75 - 37760558.75

= 1048389

Degree of Freedom (D.O.F) Calculations

$SS_{Treatments}$ = a – 1

= 4 – 1

= 3

SS_E = N - a

= 16 – 4

= 12

SS_T = N - 1

= 16 – 1

= 15

Mean Square Calculations

$MS_{Treatment} = \frac{SS_{Treatments}}{a-1}$

$= \frac{37760558.75}{4-1}$

= 12586852.92

$MS_E = \frac{SS_E}{a(n-1)}$

$= \frac{1048389}{4(4-1)}$

= 87365.75

F_o Calculations

$F_o = \frac{MS_{Treatments}}{MS_E}$

$= \frac{12586852.92}{87365.75}$

= 144.07

Table 4: Residual values calculation

Age Group	Trial 1	Trial 2	Trial 3	Trial 4
Children (7-12)	-58.5	593.5	-222.5	-312.5
Teens (13-20)	274.5	-51.5	44.5	-267.5
Adults (21-64)	-135.5	376.5	-348.5	107.5
Elderly (65 >)	111	191	-94	-208

Discussion and Analysis

By using the ANOVA 1 Way analysis method, we were able to determine if there is any significant difference when comparing 2 or more groups which were unrelated [1]. In this experiment, we wanted to find out if a Malaysian citizen on average, walks about 6000 steps per day. This leads us to our hypothesis, h_o and on the other hand, h_1 is the hypothesis where a Malaysian citizen on average, does not walk about 6000 steps a day. Using the ANOVA 1 Way method for calculations, we managed to obtain an F_o value. This is as shown below:

$$\text{Calculated } F_o = 144.07$$

The calculated F_o value then had to be compared with theoretical F_o values using an F-distribution table. From the F-distribution table, using significance levels of 5 %, 1 % and 0.1 % respectively, the theoretical F_o values obtained were:

$$\text{At 5 \%} = F_5, 3, 12 = 3.49$$

$$\text{At 1 \%} = F_1, 3, 12 = 5.95$$

$$\text{At 0.1 \%} = F_{0.1}, 3, 12 = 10.8$$

When comparing the F_o values obtained, it can be clearly seen that the calculated value was much more larger than the theoretical F_o values. Therefore, a significance level of 5 % can be taken as all theoretical F_o values are smaller than the calculated value. A 5 % significance level also improves the accuracy of our experiment as a smaller variance gap is used. Hence, this proves that our team's hypothesis, h_1 is accepted as all theoretical F_o values are smaller than the calculated one. In other words, the h_o hypothesis is rejected. This is because all of the theoretical F_o values lie in the rejection region due to the values being smaller than the calculated F_o value. This is shown in the Normal Assumption and Residual graphs as shown below.

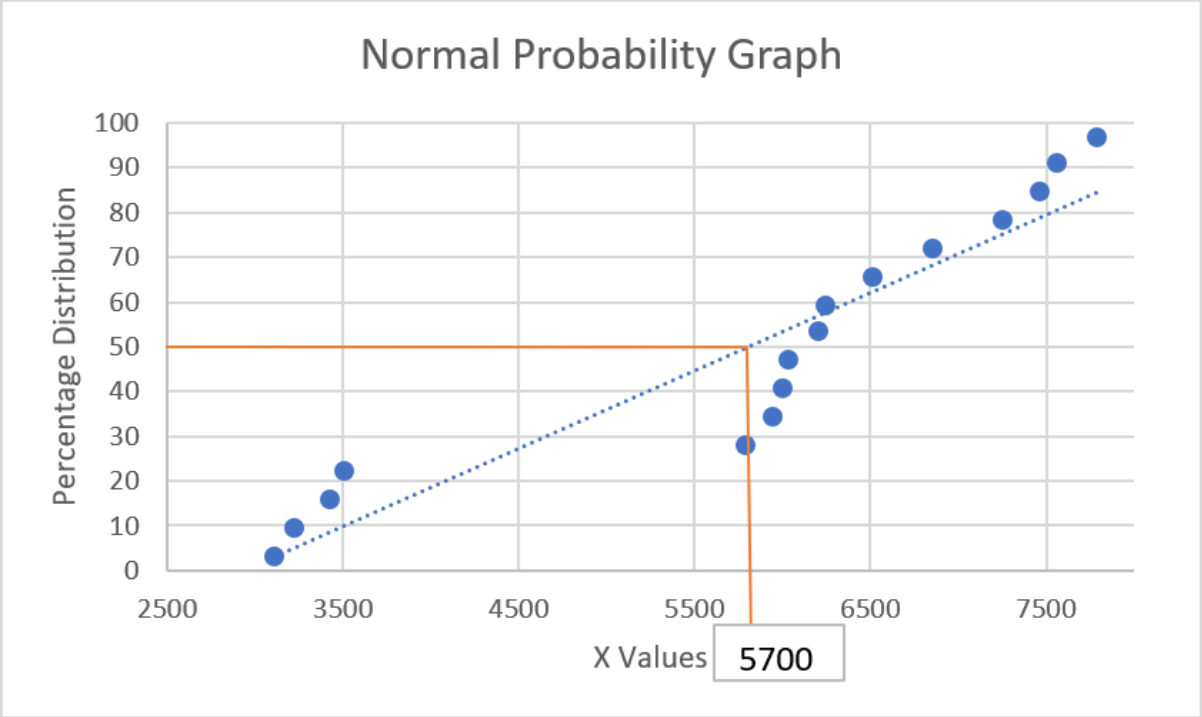


Figure 8: Normal Assumption Graph

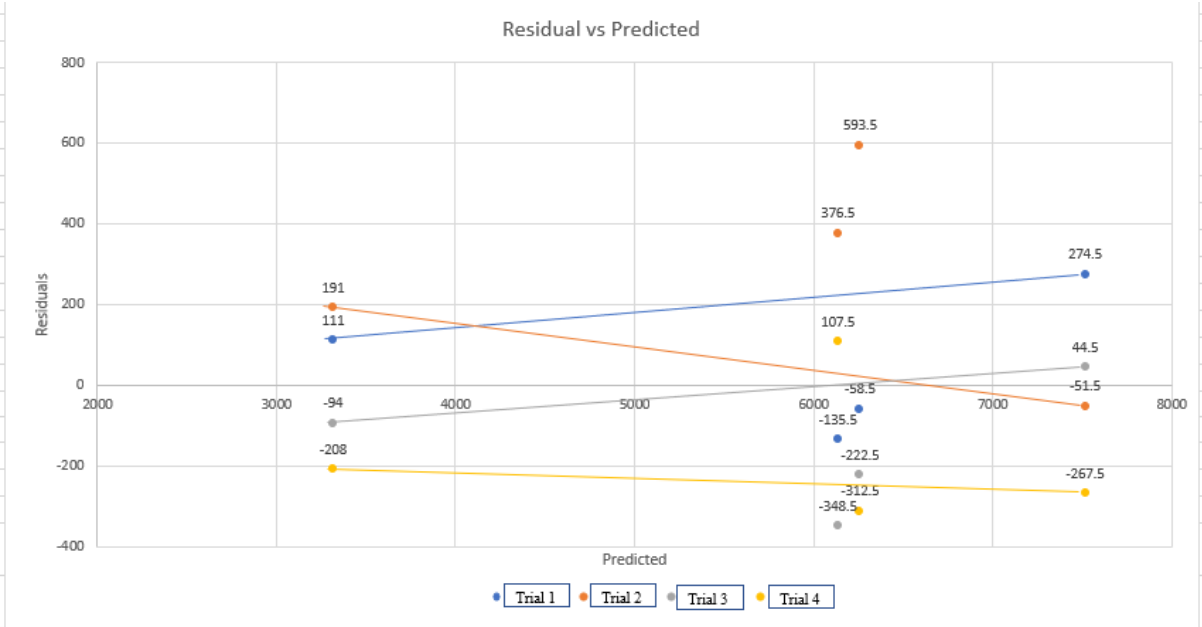


Figure 9: Residual Graph

Based on the normal assumption graph and residual graph plotted, there is further justification as to why h_1 is accepted. As seen in the normal assumption graph, the number of plots above and below the line are not the same. These inequalities indicate h_0 is not valid in terms of this experiment. Hence h_1 is accepted. Besides that, other analysis can be drawn from these graphs. Firstly, it can be seen that the plotted data is more saturated towards the upper portion of the graph. This indicates that the average Malaysian walks a lot depending on their age group (levels). It can also be seen that the data is more compactly grouped in the bottom portion of the graph as compared to the top portion of the group. This shows that the data of number of steps below 5700 is precise but inaccurate. However, the data of the number of steps above 5700 have a higher accuracy but is not precise. Furthermore, higher accuracy is a result of a smaller variance.

By comparing the number of outliers in the residual plots (4) and the normal assumption plot (2), the type of error in this experiment can be determined. The inequalities in the number of outliers explain that the errors in this experiment might be larger or smaller depending on the variance.

Due to h_1 being accepted, a grand average number of steps is then calculated and deducted from the average number of steps from each age group. This was to allow for further analysis to be conducted. The table is as shown below.

Table 5: Grand Average Table

Age Group	Average Number of Steps of Each Age Group	Grand Average Number of Steps	Difference in Average, (Grand Average – Average of Each Group)
Children	6259.50	5806.88	452.62
Teens	7516.50	5806.88	1709.62
Adults	6135.50	5806.88	328.62
Elderly	3316.00	5806.88	- 2490.88

GRAND AVERAGE CALCULATIONS

$$\frac{6259.5 + 7516.5 + 6135.5 + 3316.0}{4} = 5806.875$$

Difference in Average Calculations

For Children = 6259.5 - 5806.88

= 452.62

For Teens = 7516.5 - 5806.88

= 1709.62

For Adults = 6135.5 - 5806.88

= 328.62

For Children = 3316.0 - 5806.88

= - 2490.88

Conclusion

Based on the Grand Average table, it can be concluded that the age groups of children, teens and adults have an average daily number of steps of more than 6000. This is due to their positive average difference in values. However, the elderly age group have an average daily number of steps of less than 6000 because of their negative average difference in values. It can also be analysed that the teens age group have the highest average number of steps daily as they have the largest difference in average values. This may be due to the fact that teens are the age group

with the most active daily lifestyle as compared to the others. The elderly age group is the only one with a negative average difference. This means that they do not even meet the average number of steps daily which is 6000. This may be caused by their health conditions and inactive lifestyle which causes this age group to tire easily. Therefore, a final conclusion for this experiment that can be made is that Malaysians below the age of 65 years old walk more than the average amount of 6000 steps per day while those above 65 years of age walk less than the average amount of steps. This also further justifies the acceptance of our h_1 hypothesis whereby the average daily number of steps of a Malaysian is affected by their respective age groups.

References

[1] Who.int. (2019). *Malaysia and WHO call for more investment in primary health care the 21st century*. [online] Available at: https://www.who.int/malaysia/news/detail/08-04-2019-malaysia-and-who-call-for-more-investment-in-primary-health-care-the-21st-century [Accessed 14 Nov. 2019].

[2] "Sehat Weight Loss Coach - Apps on Google Play", Play.google.com, 2019. [Online]. Available: https://play.google.com/store/apps/details?id=com.sehatapp.basic&hl=en. [Accessed: 09- Nov- 2019].

[3] M. Aslam, "Pedometer and Step Counter Mobile Apps : How does it work?", Mohd Aslam - Technical and Non-Technical Blogs, 2013. .

[4] C. TUDOR-LOCKE, W. JOHNSON and P. KATZMARZYK, "Accelerometer-Determined Steps per Day inUS Adults", Medicine & Science in Sports & Exercise, vol. 41, no. 7, pp. 1384-1391, 2009. Available: 10.1249/mss.0b013e318199885c.

[5] C. Höchsmann, R. Knaier, J. Eymann, J. Hintermann, D. Infanger and A. Schmidt-Trucksäss, "Validity of activity trackers, smartphones, and phone applications to measure steps in various walking conditions", Scandinavian Journal of Medicine & Science in Sports, vol. 28, no. 7, pp. 1818-1827, 2018. Available: 10.1111/sms.13074.