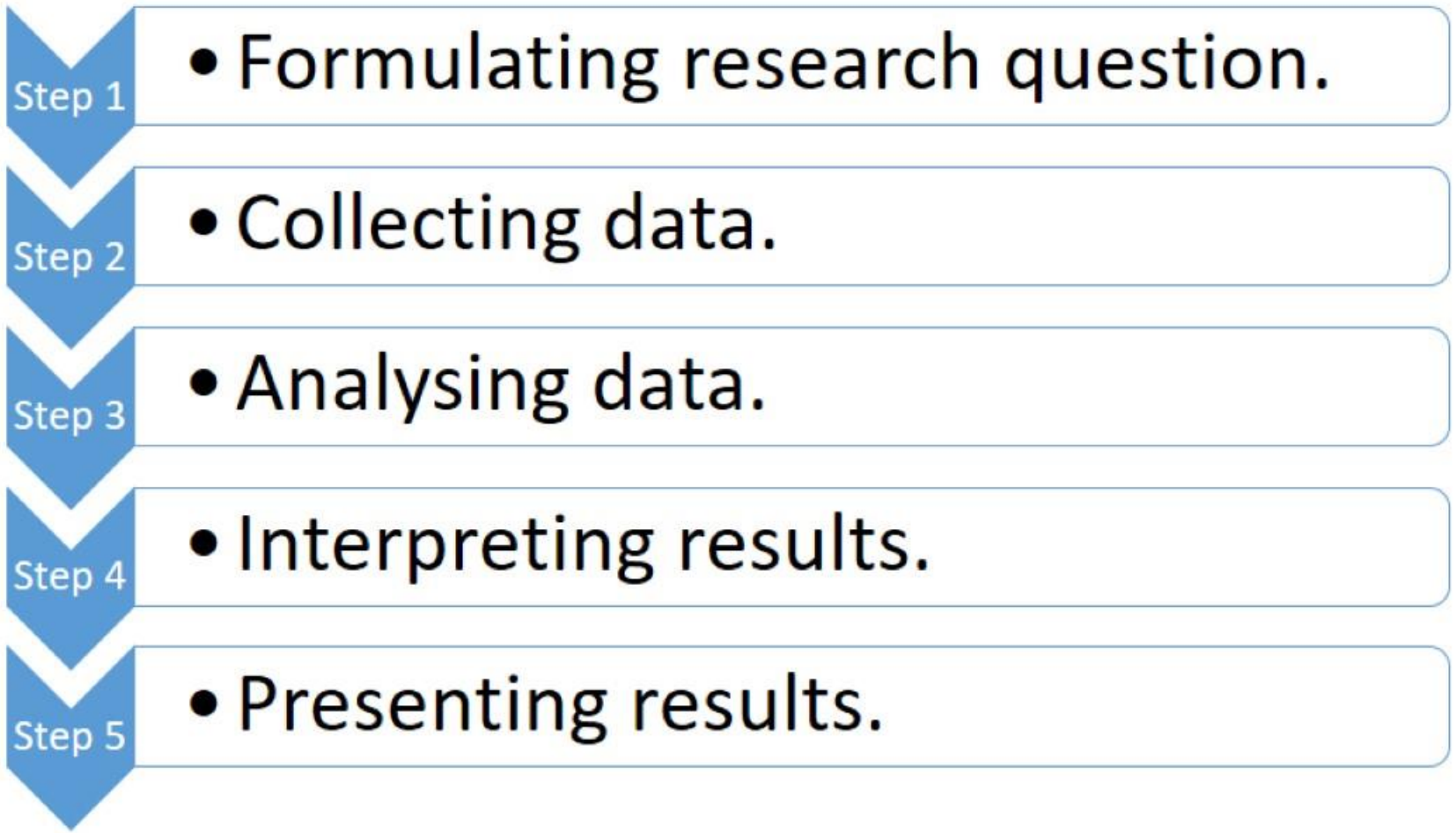
Biz Stats
Statistical Poster Project
AY1920

Introduction

1. This is a group assignment. There should be 3-4 students in each group.
2. Each group is to investigate a real life problem using real data.
3. Consult your lecturer on the feasibility of your problem before collecting data.
4. This assessment contributes to 20% of ICA.
5. This project runs from week 3 to week 17.
6. Peer evaluation will be part of the project marks.
7. The deadline for poster submission/presentation is **Week 17 and Week 18**. Late submission will result in loss of mark.

Statistical Problem-Solving Process



Good Example

- Clear and focused research question that can be studied by collecting data.
- Collect relevant primary data or show citations for secondary data.
- Analyse data using numerical/graphical summaries and statistical concepts (confidence interval/hypothesis testing).
- Show correct Minitab outputs.
- PowerPoint slide is well organised and show only relevant results and concise explanations.

Background Information:

According to NIE, most of the ITE students are from Normal Technical stream(NT) and the Government saw the need to equip these students who are deemed less inclined to academic studies with "the requisite skills and attitudes to enable them to contribute to the national economy".Therefore many people believed that O'level students are smarter than ITE students.

Our aim is to investigate whether this belief is true. We hypothesize that ITE students can averagely do better academically in polytechnics than O'level students.

Research Question:

Do O'level students obtain significantly higher GPA points compared to ITE students?

Data:

We went around Singapore Polytechnic to conduct our survey and sharing our survey via whatsapp to our friends.The survey was done using google forms.

In our survey, from Fig B we ask them who do they think would achieve a better GPA after completing their diploma based on their personal perspective. Surprisingly, almost half of the student from ITE picked student with O Level Cert ,while students with O Level cert picked themselves more than student with ITE Cert

Conclusion:

It seems that O'level students are more biased in their perception of ITE students' academic capability compared to ITE student's perception on O'level academic capability.

Reference:

- https://repository.nie.edu.sg/bitstream/10497/3360/1/CRP26_05JA_Conf06(ERAS)_Albright.pdf

- https://www.google.com.sg/intl/en-GB/forms/about/

Done by: Leonard and Rui Sheng

Fig A

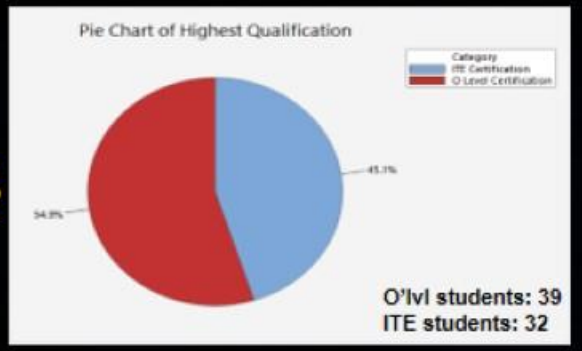


Fig B

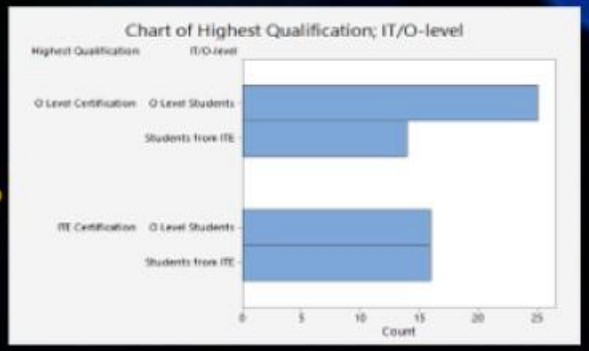
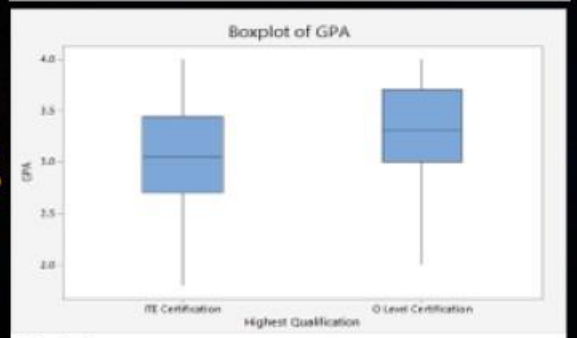


Fig C



Statistics

Variable	Highest Qualification	N	Mean	StDev
GPA	ITE Certification	32	3.07063	0.53012
	O Level Certification	39	3.25569	0.50605

Fig D

Method

μ_1: mean of GPA when Highest Qualification = ITE Certification
μ_2: mean of GPA when Highest Qualification = O Level Certification
Difference: $\mu_1 - \mu_2$

Equal variances are assumed for this analysis.

Test

Null hypothesis	$H_0: \mu_1 - \mu_2 = 0$
Alternative hypothesis	$H_1: \mu_1 - \mu_2 < 0$

T-Value	DF	P-Value
-1.50	69	0.0690

Fig E

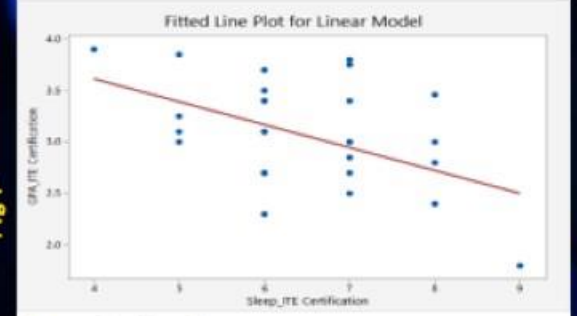
Statistics

Variable	N	Mean	StDev	Median
Sleep_ITE Certification	29	6.5172	1.1533	6.0000
Sleep_O Level Certification	32	5.9688	1.4916	6.0000

Fig F

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.457598	24.59%	21.80%	12.03%



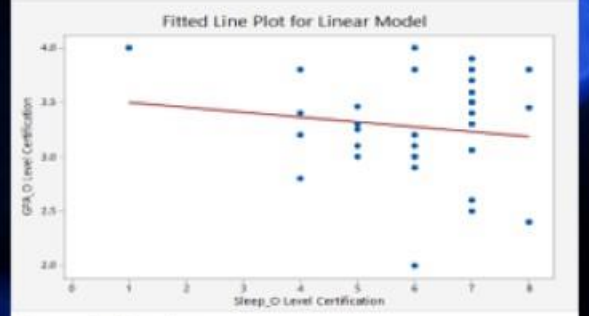
Regression Equation

GPA_ITE Certification = 4.5040 − 0.22252 Sleep_ITE Certification

Fig G

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.484590	1.92%	0.00%	0.00%



Regression Equation

GPA_O Level Certification = 3.5423 − 0.04473 Sleep_O Level Certification

From Fig C, the box plot, we can see from the upper quartile that students with O Levels Cert obtain a slightly higher gpa than students with ITE cert, in fact the lower quartile of the O'level student is almost the same as the interquartile of the student with ITE cert. The upper 75% gpa of O'Level Cert is higher than the lower 50% of GPA by the ITE students.

From the Hypothesis Testing shown in Fig D, we can see that although on average the O'Level students get numerically higher GPA than the ITE students, the difference is not statistically significantly higher (P-Value : 0.0690)

Conclusion:

It seems to suggest that on average O'Level students do have higher GPA compared to the ITE students. However, upon deeper analysis as shown in Fig D, the hypothesis test generated a P-value of 0.0690. This shows that there is not enough evidence to show that the mean GPA of O'level students is statistically significantly higher than the mean GPA of the ITE students.

We now want to find if amount of sleep affects students GPA performance. From Fig F & G, it seems that about 25% of the variation in ITE students GPA can be explained by the regression equation with number of hours of sleep as the predictor variable and only about 2% of the variation in O'Level students GPA can be explained by the regression equation with number of hours of sleep as the predictor variable.

Conclusion:

We can see the relationship between the average number of hours a student sleeps could likely affects their GPA, such that for every additional one hour an O'Level student sleeps, it will reduce their GPA by 0.04473. On the other hand, for every additional one hour an ITE students sleeps, it will reduce their GPA by 0.22252. This shows that ITE students should perhaps sleep lesser in order for them to obtain a result better or as good as O'Level students.

From Fig E, we also found that ITE students sleeps an average of about half an hour more than O'Level students. In a week, it would approximately be about three and a half hours more. If the ITE students were to use that time to study, their GPA might increase.

Limitations:

-We only conducted our survey within Singapore Poly
- It is hard to look for students from ITE in the premises because the proportion of O'Level students seems to be higher than ITE students.

Claim:

"Youths only spend 1.5 hours in day with their family members"

Background

According to a survey done by MSF in 2015, youths only spend 1.5 hours with their family in a day. Also, 74% of the youths surveyed in 2013 agreed that maintaining strong family relationships is a very important life goal.

Research Question:

"What is the typical number of hours that youth spend with their family?"

We did a survey titled "Youth and Family Time" Our research participants vary from students of polytechnic, junior colleges and ITE.

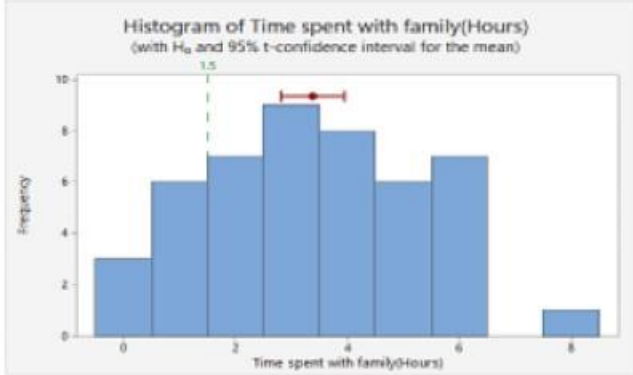
The survey questions are:

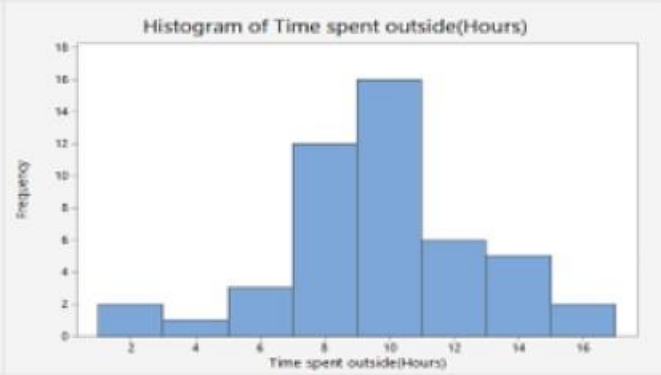
1. What is your age?
2. How many hours do you spend with your family in a day?
3. How many hours do you spend in school / outside of home in a day?
4. How many times do you dine with your family in a week?
5. Do you know about "Eat with your family day"?
6. If yes, have you experienced it?
7. What activities do you wish to do with your family?

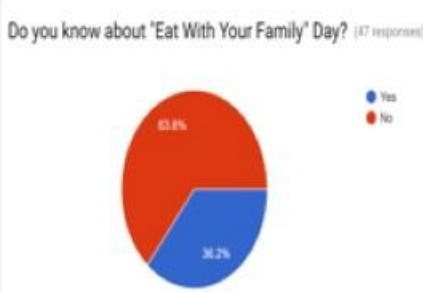
Citation:App.msf.gov.sg. (2016). [online] Available at: https://app.msf.gov.sg/Portals/0/Files/SSPC%20-%20Overview%20of%20Singapore%20Families-%20Final25th.pdf [Accessed 14 Aug. 2016].

Team Members: Huril Jannah (P1541670), Myo Minn Thiha (P1530924), Woravuth Chong (P1531235)

Data







Descriptive Statistics

N	Mean	StDev	SE Mean	95% CI for μ
47	3.3830	1.9287	0.2813	(2.8167, 3.9493)

μ: mean of Time spent with family(Hours)

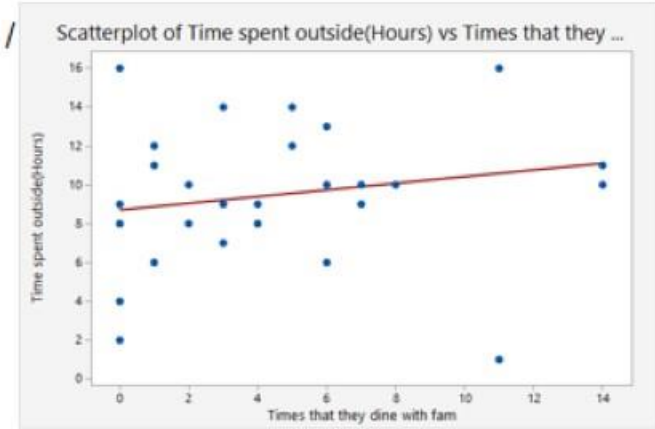
Test

Null hypothesis	H_0: μ = 1.5
Alternative hypothesis	H_1: μ ≠ 1.5

T-Value	P-Value
6.69	<0.0001

Descriptive Statistics

N	Mean	StDev	Minimum	Maximum
47	9.4468	3.0984	1	16



Regression Equation

Time spent outside(Hours) = 8.7066 + 0.1714 Times that they dine with fam

Limitations:

youths who are being home-schooled and a larger group of youths can be surveyed.

Interpretation of results

From the data obtained, it is observed that on average youths spend about 3.3 hours with their families in a day. This contradicts the claim of youths spending only 1.5 hours a day with their family members.

A P-value of 0.0001 indicates that it is rare to see youths spend on average 1.5 hours a day with their family and hence, the claim that we got from the web can be rejected. We can say status quo is not true for our sample size.

However, as we can see from the scatter plot, even though they spend more hours with their family, it doesn't mean that they will dine with their family.

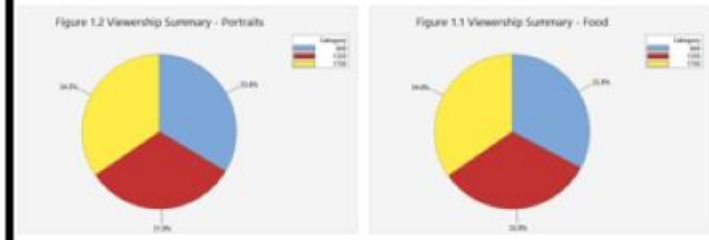
Centre for Fathering has been promoting "eat with your family" day to enhance family bonding annually but it seems that only 36% of the youths know about this annual event. Therefore, if the MSF would want to enhance family bonding through dining, they should work hard to ensure more organizations and schools are on board to promote this annual event and subsequently increase the occurrence.

Background information: A post by The Huffington Post claims that by uploading a photograph onto Instagram at the time of 5pm will result in the most number of viewership. We want to investigate and see if this claim is true.

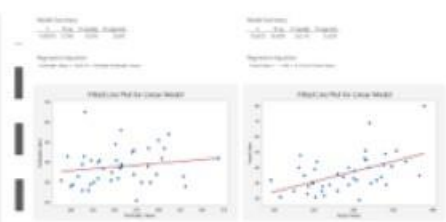
Our Question: 🤔
Does uploading a photograph onto Instagram at 5pm result in more viewership than uploading the photograph at other times of the day?

To set things right:
We would also like to 😎estigate and see if there is a relationship on how different types of photograph to see if the claim is valid. In our investigation, we have decided to choose the genre of Food and Portraits.
Our primary source of data comes from the viewership option that is made available in Instagram. Over the span of 1 month, we have uploaded a total of 84 photographs at 0900H, 1300H and 1700H. We have recorded the total viewership for each timeslot for a duration of 4 hours each.

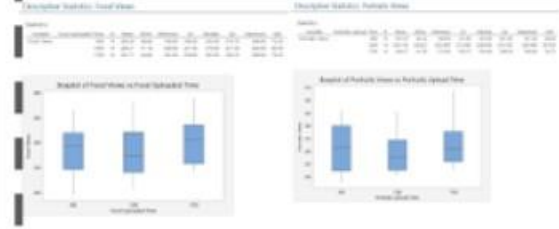
The number of 'likes' have been recorded as well to determine the popularity between Food and Portraits.



A pie chart of the viewership summary for both Food and Portraits are shown in Figure 1.1 and Figure 1.2 respectively. From both the figure we can tell that the total viewership for Food is more than the viewership for Portraits. However, the viewership for both Portraits and Food had the most number of views at the time of 1700H, and the least during 1300H. This could be due to the social media habits of the netizens whereby the usage of mobile phone increased tremendously immediately right after working hours.



These linear regression graph shows us the relationship between the number of likes and the number of viewership. The result of both linear regression graph are not very promising as the R-square results tells us that the data is not very useful in determining the number of likes for every unit of viewerships. It is however, that for the food photography, the number of likes will be 0.13143 times of the viewership obtained. Similarly, for portraiture photography, the number of likes will be 0.03644 times of the viewership obtained.



The use of descriptive statistics(refer to the 2 tables above) helps us with the analysing of important details such as the average viewership that was obtained across the 3 different uploading times. From the statistics we can see that a photograph that was uploaded at the time of 1700H had more mean viewership regardless of the type of photography involved. In addition to that, both Food and Portraits had the most number of viewership when uploaded at 1700H. It is interesting to note that the spread of viewership were moderately consistent for photos that were uploaded at 5pm. From this descriptive statistics it does indeed shows that the viewership were ranked first among both types of photography at a uploading time of 1700H.

Fun Facts!
Do you know that Portraiture photography garner more likes than Food photography? The next time you want to get more likes on your Instagram, upload photos of people!

Instagram viewership?

Our limitations: Viewership can be inconsistent in certain cases as the the viewers may have irregular browsing habits and patterns. For example, an individual may choose to browse through his phone only on weekends. This will result in some minor inconsistency.

2-Sample t: Food Uploaded Time 1700, Food Uploaded Time 0900

2-Sample t: Portraits Uploaded Time 1700, Portraits Uploaded Time 0900

2-Sample t: Food Uploaded Time 1700, Food Uploaded Time 1300

2-Sample t: Portraits Uploaded Time 1700, Portraits Uploaded Time 1300

We have conducted Hypothesis testing on both Portraits and Food where our Null Hypothesis is the viewership at 1700H and our alternative hypothesis is the viewership at either 1300 or 0900H (refer to the attached table above). From the results, we can see that the p-value is significantly large and is greater than 5%. It is not surprising to see a result like this and there is insufficient evidence to reject the claim(refer to "background information"). We thought that it may only be applicable to Food photography only but the results are consistent throughout even with Portraiture.

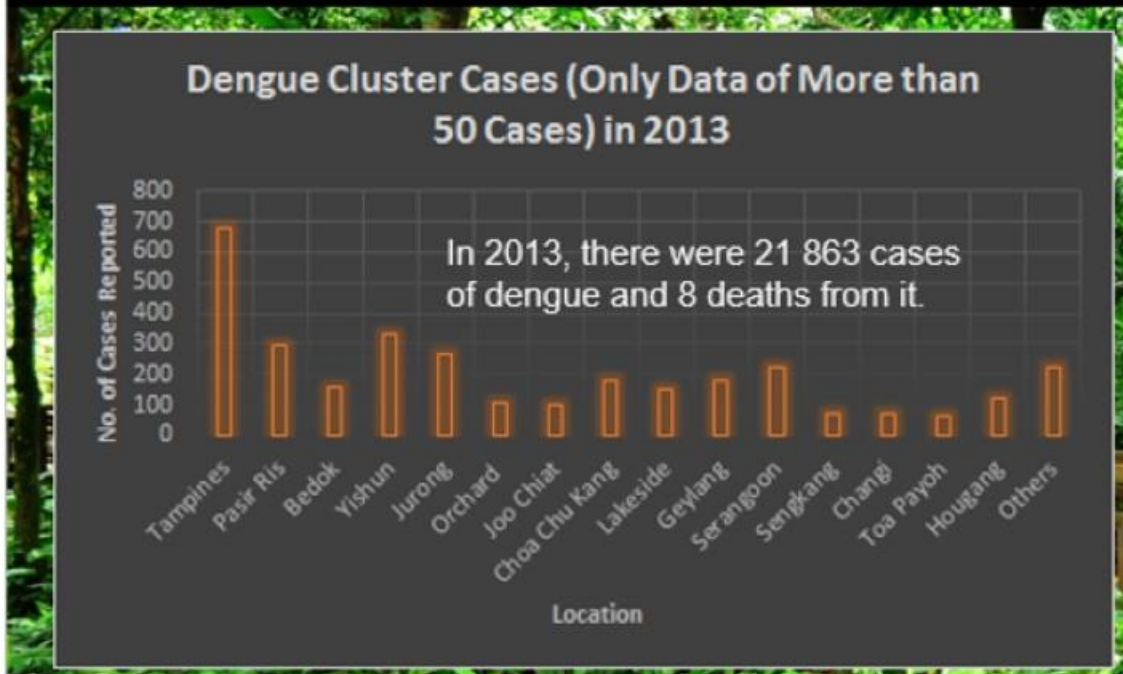
In conclusion, we our research have shown that uploading a photograph regardless of the genre at 5pm will result in the highest number of viewership, in which correspond to the article's claim. So the next time you want to get massive viewership for your product advertisement, upload it at 5pm! Statistics have shown that there will be a heightened viewership when you upload your pictures at this 'golden hour'!

Done by:
Sri Mayang (P1630738)
Teo Shao Zun (P1631106)

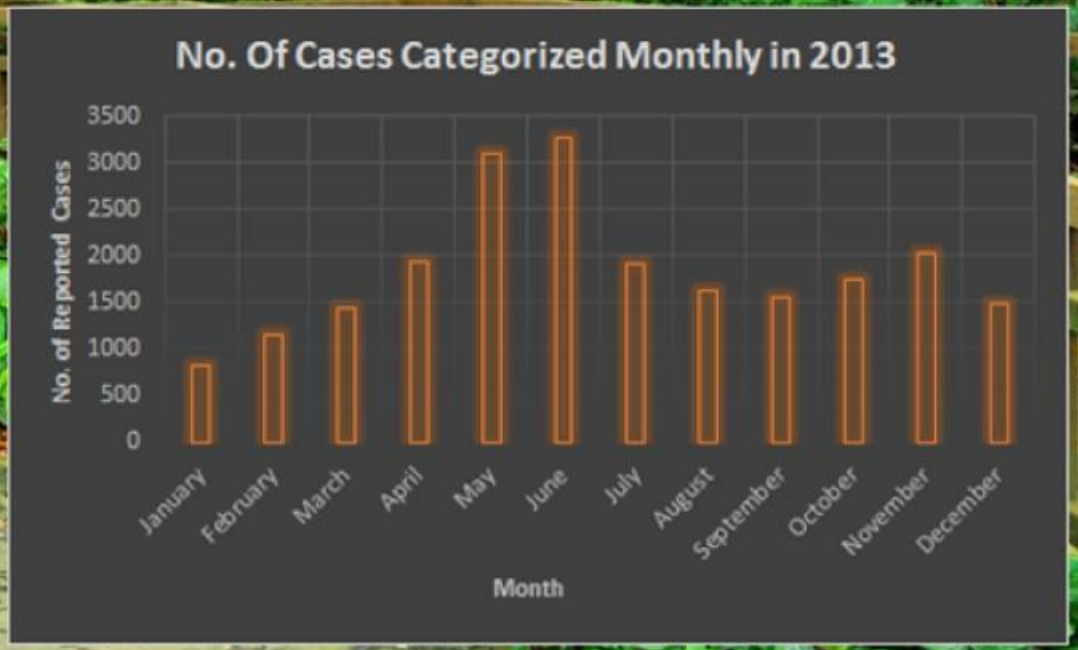
Below Average Example

- Vague research question.
- Data collection are not done properly.
- Data analysis are erroneous.
- Graphs are not clearly labelled. Wrong choice of graphs.
- Explanations are not relevant to task.
- PowerPoint slide is cramped with unnecessary information.

Dengue Cases In Singapore, By Month and Area (2013)

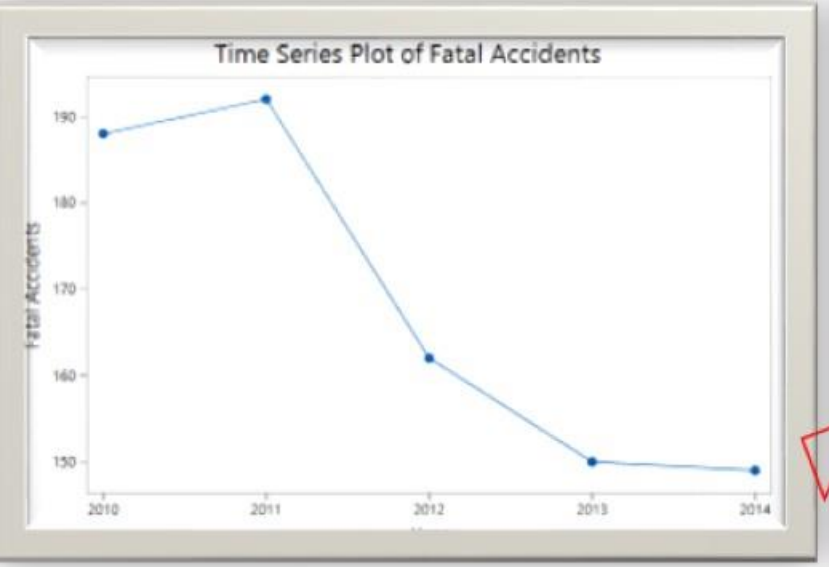


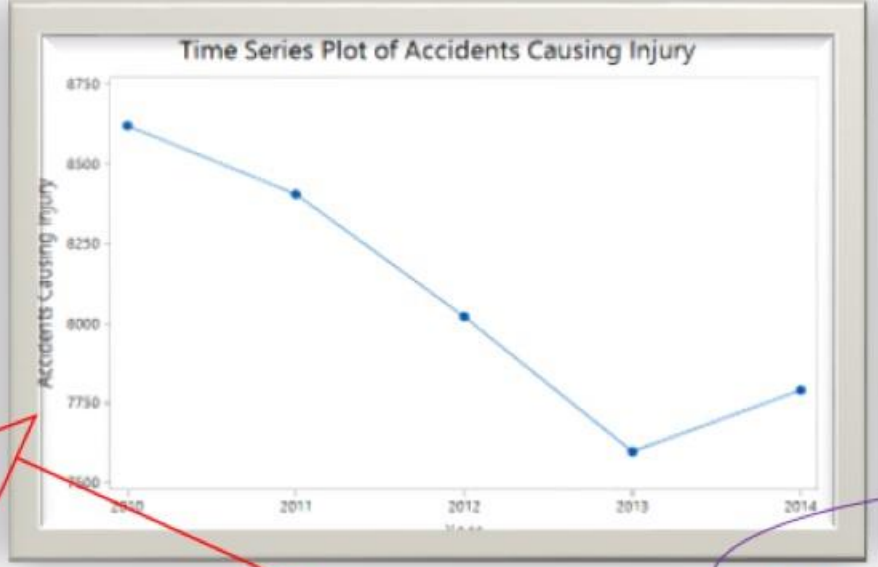
The most number of dengue cases reported in 2013 is during June, whereby approximately 3259 cases were reported and in Tampines with approximately 679 cases with an average of 200 cases in each area.

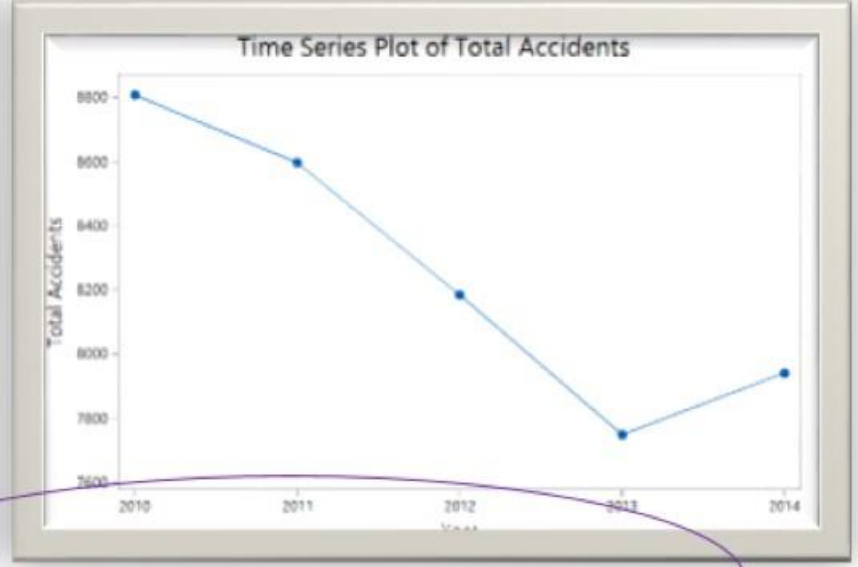


We can conclude that the month of May and June has the most number of dengue cases due to the dengue breeding season

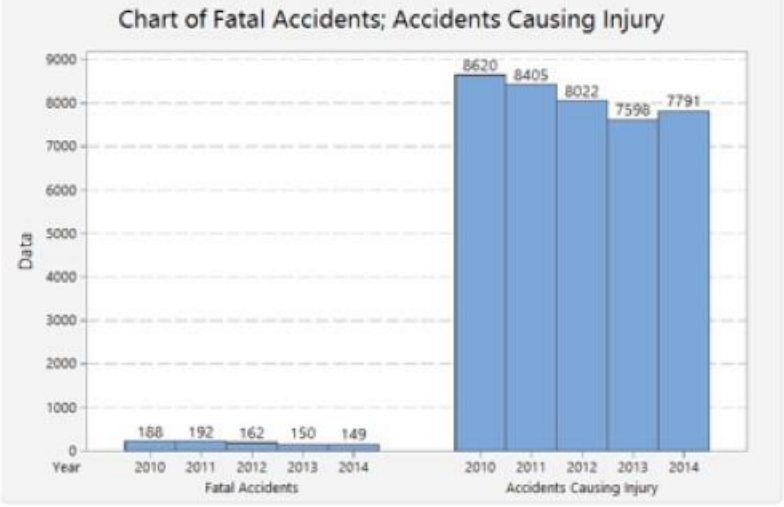
https://www.moh.gov.sg/content/moh_web/home/statistics/infectiousDiseasesStatistics/weekly_infectiousdiseasesbulletin.html?year=2013

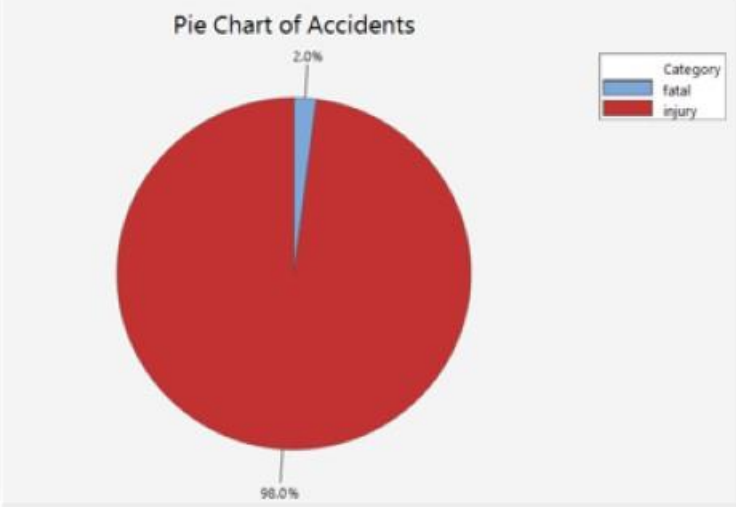






YEAR	FATAL ACCIDENTS	ACCIDENTS CAUSING INJURY	TOTAL ACCIDENTS
2010	188	8620	8808
2011	192	[illegible]	8597
2012	162	[illegible]	8184
2013	150	7598	7748
2014	149	7791	7940





QUESTION- How many accidents causing injuries and fatalities were there from 2010-2014?

-For fatal accidents, as you can see from the graph, there was an increase in the first year from 2010-2011. However, it started to decline drastically from 2011-2012, and declined at a slow rate over the rest of the years from 2012-2014.

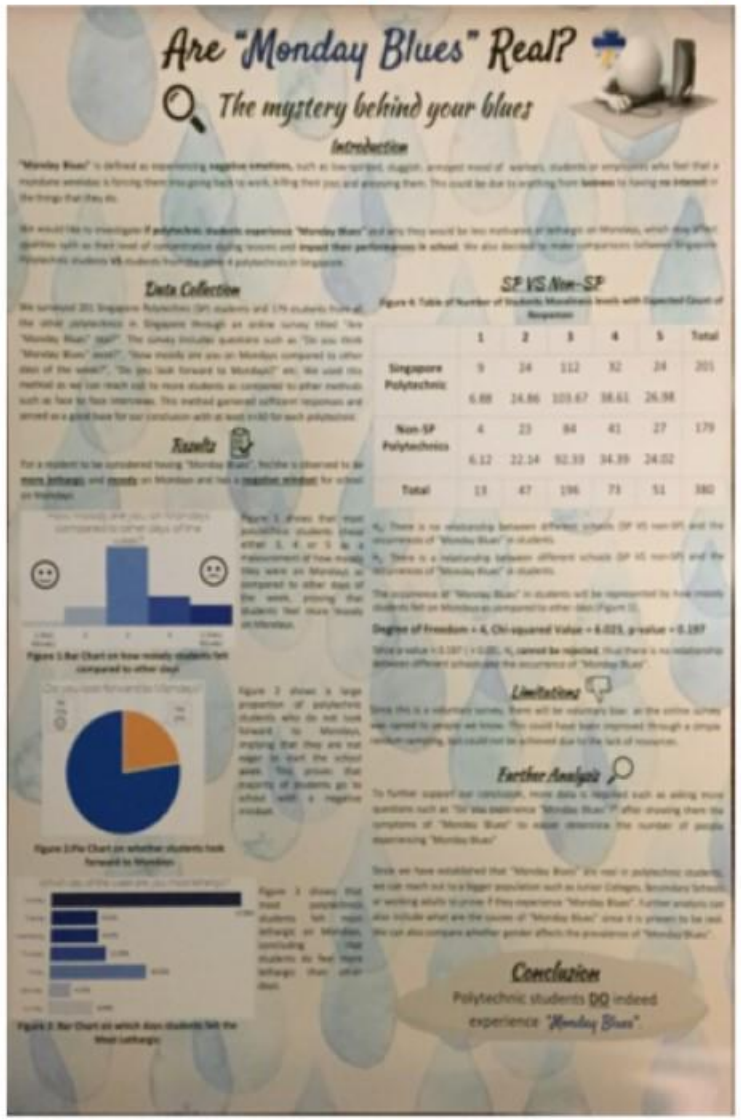
-For accidents causing injury, there was steady decline from 2010-2013, but a sudden increase from 2013-2014.

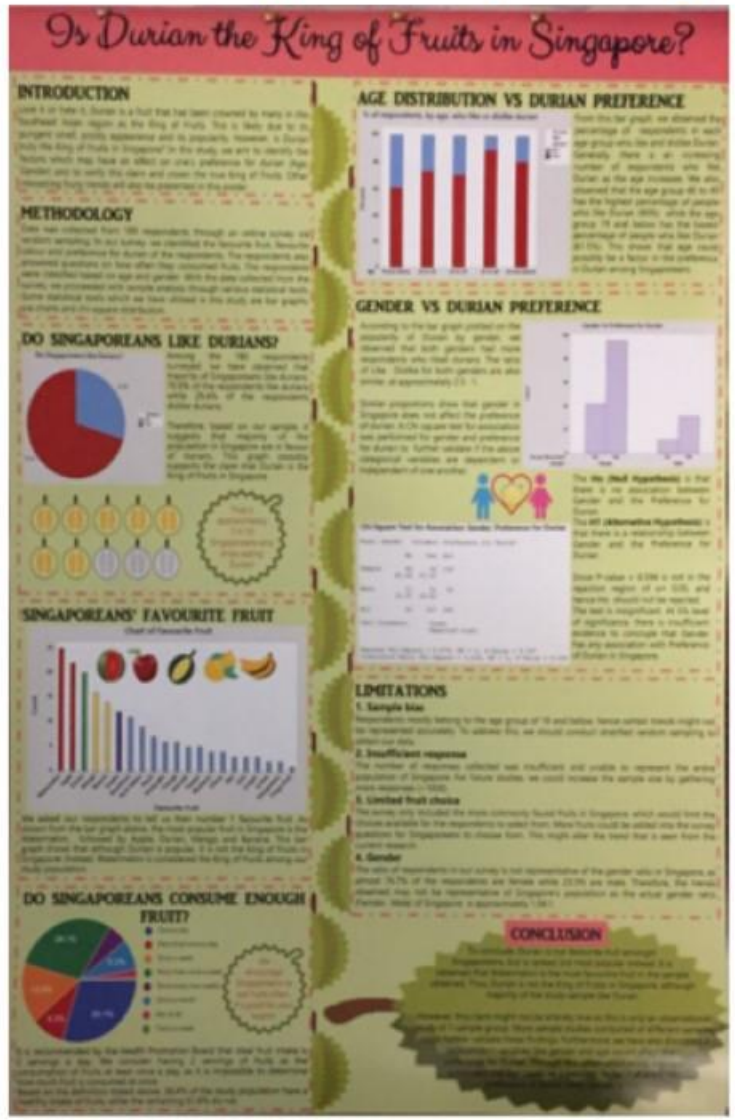
-Overall, for total accidents, there was a steady decrease from 2010-2013, however an increase from 2013-2014 mainly due to the increase in accidents causing injury from 2013-2014.

-As you can see from the pie chart, injury prone accidents make up most, to be exact 98%, of total injuries over the 5 years, while fatal accidents only make up a very small portion of 2%.

CITATION : http://driving-in-singapore.spf.gov.sg/services/driving_in_singapore/documents/Annual_Road_Traffic_Stats.pdf

Examples of past topics (for your reference)





Examples of past topics (for your reference)

Examples of past topics (for your reference)

- Is Uber cheaper in its estimated rides compared to Grab Car?
- Is there such a thing called “Home Team Advantage”?
- What are the factors affecting the winning position in F1 races?
- Are Singaporean teens living an unhealthy lifestyle?
- Is the percentage of teens speaking mother tongue at home lower than adults?
- Does education affect the rate of marriage in Singapore?

(refer also to the “Guide and rubric” document for further info)

		Statistical Poster Project	Students To do:
Week 1			
Week 2			
Week 3		Formulate Questions	•Form groups (3-4 persons per group) •Find a claim about a context / hypothesis to be challenged •Decide on one or two research questions based on the context
Week 4			
Week 5			•Get approval from Tutor regarding the research area/research question
Week 6			
Week 7			
Week 8	MST Week		
Week 9 Week 10 Week 11 Week 12	Vacation	Collect Data	•Design survey forms & data collection plan, or, search and select a data set in archives •May decide to change research questions (subject to Tutor's approval)
Week 13 Week 14		Analyze Data & Interpret Results (Descriptive Statistics)	•Import data into Minitab •Generate numerical and graphical summaries & interpret the results •Report results/interpretation into poster
Week 15 Week 16 Week 17		Analyze Data & Interpret Results (Statistical Tests)	•Generate Statistical tests such as Confidence Interval, Hypothesis testings etc & interpret the results •Report results/conclusions into poster
Week 18		Submission/Presentation	•Poster Ready for presentation
Week 19	EST Week		

Before you start collecting data, make sure your lecturer has approved your research question and data collection plan.

What to do now?

- Form groups, 3-4 persons per group.
- Go to Blackboard now to take a look at the poster templates.
- Find a claim about a context / hypothesis to be challenged
- Decide on one or two research questions based on the context