

Statistical Poster Project

Instructions:

1. This is a group assignment. There should be 3-4 students per 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 20% of your module score.
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.

Guidelines for Statistical Problem-Solving Process:

Step 1: Formulate Questions

Identify a context that you want to find out more or a claim to be challenged, then formulate the research questions around this context or claim.

The research question should satisfy three conditions: (i) clearly defined, (ii) focused and (iii) can be studied by collecting data.

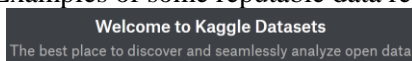
Formulate a strategy to answer the research question.

Step 2: Collect Data

There are three ways to collect data:

➤ Data repository

- Pros: Data are readily available.
Cons: Data may be massive and messy, and may not be from credible sources
- Examples of some reputable data repository sites:



<https://www.kaggle.com/datasets>



<https://data.gov.sg/>



<http://mathforum.org/workshops/sum96/data.collections/datalibrary/data.set6.html>

➤ Surveys

- Pros: Collect only data that you need.
Cons: More resource intensive, and may not be random.
- You have to design a survey form and get participants to respond to the survey.
(Click [here](#) for guide on creating good surveys)
- One of the best ways to implement a survey is to use Google Form or Survey Monkey.
(Click [here](#) for tutorial on how to use Google Form, and [here](#) for Survey Monkey)

- Ensure that you collect some numerical and some categorical data.

Example of questions to get numerical data	Example of questions to get categorical data
1. What is your height in feet and inches? For example, if you are 5 feet and 4 inches, write 5'4". <input type="text"/>	1. How many apps do you currently have on your Facebook account? <input type="radio"/> 0-4 <input type="radio"/> 5-9 <input type="radio"/> 10-19 <input type="radio"/> 20 or more
3. In a typical week, about how much money, in U.S. dollars, do you spend on songs or other music files you download from the internet? <input type="text"/>	2. Do you take nutritional supplements? <input type="radio"/> Yes <input type="radio"/> No
4. During the last week, how many times, if any, has someone threatened you with harm? <input type="text"/>	1. How often do you watch television shows? <input type="radio"/> Extremely often <input type="radio"/> Very often <input type="radio"/> Moderately often <input type="radio"/> Slightly often <input type="radio"/> Not at all often

- Try to avoid the following questions for this assignment

(These are not bad survey questions but it may be beyond the scope of this module to teach you how to prepare and analyse such data)

	Reasons
1. In a typical day, which of the following forms of transportation do you use? (Check all that apply) <input type="checkbox"/> Bus <input type="checkbox"/> Subway <input type="checkbox"/> Train <input type="checkbox"/> Ferry <input type="checkbox"/> Cab <input type="checkbox"/> Automobile <input type="checkbox"/> Airplane	The data collected with multiple answer questions (“Check all that apply”) require a lot of effort to clean and transform the data. So, unless you are willing to spend time to cleaning data, it would be best to avoid multiple answer questions.
5. What does your organization do to raise money? <input type="text"/>	Such question generates textual data, which is beyond the scope of this module to teach you how to analyse such data.

➤ Experiments

- Pros: Collect only data that you need.
Cons: More resource intensive, may only collect small data set.
- Identify dependent variable and independent variables. Make a careful plan to collect all necessary measurements on the variables.
- Examples
 - Recording the number of views when you post a photo on Instagram in the morning, evening and midnight.
 - Recording the fares of Uber and Grab Car of similar distances at peak and non-peak hours.

Step 3: Analyse Data

Analyse the data using the appropriate choice of the following statistical concepts:

- | | |
|---------------------------|---------------------------|
| (i) numerical summaries, | (ii) graphical summaries, |
| (iii) confidence interval | (iv) hypothesis testing. |

Decide which types of graphs are most suitable to represent the data and support the analysis.

Step 4: Interpret results

Based on the data analysis in Step 3, interpret the results and draw valid conclusion in view of the original research question in Step 1.

Step 5: Presenting results

The final report should be a one PowerPoint slide that tells stories about a set of data.


The poster templates are available to download via blackboard. (Poster Project)


The poster should:

- ✓ be simple and visually attractive,
- ✓ include numerical and graphical summaries,
- ✓ contain interpretation of the results,
- ✓ have a logical flow of statistical thinking process,
- ✓ include citations of sources for data.

Resources:

Here are some links to get a general idea about making “statistical posters”:

 Singapore Statistics Poster Competition:
<http://www.science.nus.edu.sg/newshub/2247-2017-singapore-statistics-poster-competition>

 International Statistical Literacy Poster Competition:
http://iase-web.org/islp/documents/Poster_Competition_2014-2015/ISLP%20poster%20session%20presentation%202015.pdf

 American Statistical Association:
<http://washstat.org/poster/Brochure%20front%202016%20-%202012%201%2014%20RD.jpg>

General Scoring Rubric:

Marks	Descriptions
85 – 100 (Excellent)	<ul style="list-style-type: none"> Valid research question satisfying all three conditions (refer to Step 1). Research topic is interesting. Data collected are relevant to answer the statistical question. All variables are identified and defined for analysis. Apply at least three statistical concepts (refer to Step 3) correctly. Show at least two Minitab Express outputs correctly. Concise explanations detailing how justification is arrived supported by statistical evidence. Poster shows creativity and neatly organised with good use of font, colour and space, satisfying all criteria (refer to Step 5). All group members participated equally. Able to answer all questions clearly.
70 – 84 (Good)	<ul style="list-style-type: none"> Valid research question satisfying at least two conditions (refer to Step 1). Data collected are relevant to answer the statistical question. Most variables are identified for analysis. Apply at least two statistical concepts (refer to Step 3) correctly. Show at least two Minitab Express outputs, but with minor errors. Explanations are relevant to task and supported by analysis. Poster satisfies at least three criteria (refer to Step 5). All group members participated. Able to answer most questions.
50 – 69 (Average)	<ul style="list-style-type: none"> Valid research question satisfying at least one condition (refer to Step 1). Data collected are relevant to answer the statistical question. Some variables are not identified or defined for analysis. Apply at least two statistical concepts (refer to Step 3), but with minor errors. Show at least one Minitab Express output. Explanations are relevant to task but not supported by sufficient analysis. Poster satisfies at least two criteria (refer to Step 5). Some group members participated. Able to answer some questions.
< 50 (Poor)	<ul style="list-style-type: none"> Invalid research question. Data collected are not relevant to task. Variables are not defined. Apply incorrect statistical concepts. Minitab Express outputs are incorrect. Explanations are not relevant to task or not supported by analysis. Poster satisfies at least one criteria (refer to Step 5). Late or incomplete submission. Plagiarism. Only 1 or 2 group members participate. Have difficulties answering questions.