Interpreting Data Sets

Study Development Factsheet

Interpreting data sets

1. Data cleaning

Look carefully through your data. Ask yourself these questions:

a) Have they missed a very important answer?
If the study really relies on knowing, for example, patient height, and a participant hasn't included their height, you may need to contact them to ask. If the study is anonymous, you'll have to decide whether to remove their answers if this makes them ineligible for the study.

b) Do their answers make sense?
Has the participant said that they are 200 years old? Have they said that they are 20, and have attended university for 30 years? Have they said for one question that they love vanilla cake, and then for another question said they hate all sweet foods?
Flag these answers. You will need to decide whether to contact the participant to check this out or remove their answers from your data.

c) Are there exact duplicates?
Data entry errors can happen and a participant may be included twice. Remove one of them (if you are certain that it's a duplicate and not just two people who are in total agreement).

d) Are there completely blank answers?
These answers can be helpful if you'd like to report on how many participants responded to the questionnaire, but can incorrectly skew your answers when you analyse your data later on. It's worth flagging them so you're aware of them for later.

2. Pose your hypotheses

Try not to let the data influence your ideas. If you collected this data yourself, you should have decided on your hypotheses before you received the data. If you have been asked to

do this analysis by an employer, lecturer, or funder, ask them what they were hoping to learn. If not, look at the variables you have. What interesting relationships might exist between them? If you have been given sales data, you might wonder if, for example, the day of the week affects the final sales price. Choose things that will be measurable based on the data you have. For example, if you'd like to know if age affects injury recovery, but you haven't got any age data, this would not be a hypothesis you could test.

3. **Present your data**

   Create graphs, charts, and diagrams from your data. You can do this in most data collation software (such as SPSS or Excel). If you don't know how to do this, contact Digital Training, your school's technician, or Study Development for help.

   Presenting your data can help you spot patterns and anomalies in your data. The type of graph you should use depends on the type of data and what you'd like to know. If you're looking for a change in pattern over time, something like a scatter graph would show you the general trend of the data. If you'd like to know the distribution of characteristics (e.g. which age group are most of the participants in?) something like a pie chart would be useful. Try out a few different things. You can spot anomalies by finding a pattern in your data and seeing if a data point is very far outside of what you'd expect.

4. **Calculate statistics**

   This step is optional, as the diagrams may be enough to answer your questions. In a typical study, you will have stated which statistical tests you're going to perform before you collected the data, so you'd perform those at this point. If not, refer back to your hypotheses to choose tests that will help to answer your questions.

   You can use things like mean, median, mode and range to compare between groups. For example, you could compare the mean age of people in one department of an office to the mean age of people in another.

   Are you looking for a relationship between two variables? You could then perform a correlation test. Does your hypothesis refer to the difference in two means? This would lead nicely to a t-test or Z-test.

   If you're not confident with statistics, you can book in for a Study Development maths tutorial.

5. Draw conclusions

Explain what your statistical tests and diagrams show, and (if appropriate) recommend follow-up actions. You can also identify any limitations of the data at this point (e.g. “age of participant may affect this, but we did not collect age data, so further study would be needed.”)

Tips

Filter the data into smaller groups

- If one of your hypotheses is about a specific subgroup, filter the data to only include that group. For example, if you think that older female participants will score higher on a test than younger female participants, you can filter out all other genders from your data for now. This will help prevent you from getting overwhelmed and may make it easier to spot patterns.

Use qualitative data

Quantitative data can usually show trends but misses some context. Remember that we cannot usually say whether one variable has **caused** a change in another based on quantitative data, but participants could say whether they feel it has.

Sort the data

If you “sort” a variable from lowest to highest, you can see whether other variables start to show a pattern. This may suggest that there’s a correlation between the variables. You may also notice that a certain characteristic appears more often at the lower end of that variable, and less at the higher end, which again could suggest a relationship between those that variable and characteristic.

Experiment with the data

Whilst you are investigating, try out as many different ideas as you can. Create plenty of diagrams, try out lots of configurations, and propose many questions (even if you think they won’t have a good answer!). Try not to let your own preconceptions affect what you test out, as this can skew your results. Be open to what the data has to show you, rather than trying to fit it to what you’d like to see.

Est. 1841 YORK ST JOHN UNIVERSITY

Library & Learning Services

Support: Study Development offers workshops, short courses, 1 to 1 and small group tutorials.

- Book a tutorial or join a workshop on the Study Development tutorial and workshop webpage or search 'YSJ study development tutorials.'
- Access our Study Success resources on the Study Success webpage or search 'YSJ study success.'

Library and Learning Services

Study Development

Email: studydevelopment@yorksj.ac.uk

Est. 1841 YORK ST JOHN UNIVERSITY