Qualitative and Quantitative Research -

TABLE 1.1 Assumed characteristics of research

Qualitative research	Quantitative research
Uses words	Uses numbers
Concerned with meanings	Concerned with behaviour
Induces hypotheses from data	Begins with hypotheses
Case studies	Generalisations

Source: adapted from Hammersley, 1992

Quantitative

↳ Numerical analysis / data collection

↳ data set(s) that have a numerical aspect or you convert to a numerical format.

↳ allow for numerical analysis and/or statistical tests

Qualitative

↳ Open-ended / Why questions

↳ Response info / data

↳ Gather some document that you will dive into & sort through as you do your analysis

↳ "how" and "why" questions that do not demand numerical data or precise results.

TIPS FOR FINDING DATA

1) Availability - obtain data in timely manner
2) Data Cleanliness - avoid studies with missing data or values that seem unreasonable
3) Good Documentation - what each variable in data set means; how data is collected to assess for bias and create appropriate analysis
    a) Qualitative research - understand story behind docs (where they came from)
        i) Think about potential pieces you or authors may have overlooked (fill in gaps and format for accessibility)

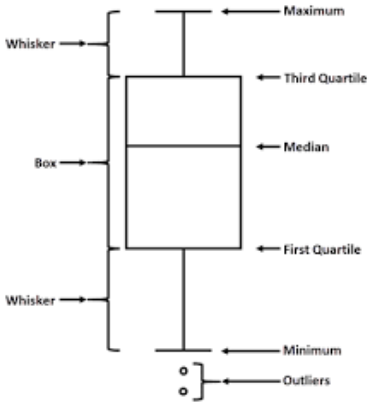
EXPLORATORY DATA ANALYSIS

↳ looking for patterns or trends in your data that will help you refine your analysis plan and figure out which statistical or other test are most appropriate

Once you have your research questions…

1) Visualize / transform your data and use the new info and patterns to refine your analysis and perhaps your research questions
    a) No formal way of doing exploratory data analysis, just what seems reasonable

Numerical Data



↳ Scatterplot (predictor variable on x axis & predicted variable on y axis)

- May discover that your expectation may not have been correct or exist

↳ Boxplot

- Used to compare two groups or predator variables with some outcome variable.
- Often will be categorical predictor variables

↳ Bar charts / Histogram

- Probability of each values occurring by grouping them into bins and then displaying the bar height as the number of values in that group or bin
- Useful when trying to figure out the distribution of your data before a statistical test

BASICS OF PROBABILITY DISTRIBUTIONS

↳ Distribution: way of measuring the partners of occurrence in some data set

Bernoulli :

Single trial or sample (flipping a coin) *only two outcomes*

Each of the two possible outcomes has some probability of occuring

Binomial :

Several bernoulli samples strung together

Some probability of success and some of failure

Geometric :

Number of failures you'd need before the success when your stringing together bernoulli trials or samples

Normal (gaussian) distribution :

Most common ; bell curve → Peak = mean

95% of samples will have some value within one standard deviation

Used with continuous random variables and large enough sample size

Poisson Distribution :
Used when interested in number of events over time period or space
*in general, larger data sets are better

HYPOTHESIS TESTING

↳ Create a hypothesis then use a statistical test to see if or which hypothesis holds true

↳ **Null Hypothesis** = Baseline assumption that say there is no statistically association or difference between two means; "default hypothesis"

*The 2 means are not statistically significantly different.

↳ **Alternative Hypothesis** = What you're hoping is true or suspect may be true. Ex: two groups differ

*There is significant difference between the means

**frame question so it can be answered through some statistical test*

1) Come up with research question
2) Turn your research question into a hypothesis
3) Collect samples or data set
4) Exploratory data analysis (EDA)
5) Significance level
    a) cutoff / tolerance for random error
6) Finalize your statistical test
7) Run your statistical test
8) Examine your output of the test / Ensure that it seems reasonable (subjective???)
9) Report your key statistics or results

STATISTICAL TEST

If you know the population standard deviation then use the Z test
If you don't know the population standard deviation use the T test
Also use the T test if the sample size is greater than 30

↳ **One Sample Z/T tes**t = comparing a sample(s) mean to a predetermined mean (hypothesized mean)

↳ **Paired Z/T test** = comparing the means of two samples from the SAME population

↳ **Two Sample Z/T test** = comparing means from two independent samples from two different populations

T TEST (more common)
Assumptions:

One Sample T Test =

1) Data is continuous not categorical binary
2) Data follow a normal distribution
3) Random sampling within the population

Paired T Test

1) Continuous variables
2) Difference between the matched sample pairs is normally distributed
3) Random sampling

Two Sample T Test

1) Continuous data
2) normally distributed
3) Samples are INDEPENDENT
4) Variances of the two samples are the same or very similar
5) Random sampling from both populations

Z TEST (Assumptions)

One Sample Z Test

1) Continuous Data
2) Data is normal distribution
3) Samples taken randomly
4) Population standard deviation is known

Paired Z Test

1) Continuous data
2) Difference between the pairs of data points follows a normal distribution
3) Random and independent samples
4) Population standard deviation of the difference is known

Two Sample Z Test

1) Continuous variables
2) Random sampling in both populations
3) Independence of samples
4) Variance of the distributions of both populations is known

*one tailed test = determine whether your mean is greater than or less than the sample mean
*two tailed tests = is there a difference between the two means?

Typical Outputs

1) T/Z stat
2) Degrees of freedom (n -1)
3) P value or confidence interval

ANOVA TEST

T / Z test = two samples or one sample and a mean

ANOVA = three or more groups or samples or populations that you want to compare the means across

One- Way ANOVA

↳ one variable differs across the populations

One Way Assumptions

1) Normality assumption
    a) Samples are taken from some approximately normal distribution in the population
2) Independent samples
3) Variance across the groups is the same
4) Dependent variable is continuous

*Null Hypothesis = no difference between the groups and implies that the means are equal or close to equal

*Alternative Hypothesis = means across the groups differ; at least one mean is different from the other two

Two-Way ANOVA

↳ how two categorical factors differ across the populations AND how these categorical factors affect each other

Two Way Assumptions

1) Outcome variable (dependent) is continuous
2) Groups / independent variables should be categorical and independent
3) Samples are independent
4) Variance across groups is at least approximately the same
5) Normality
    a) Samples are taken from some approximately normal distribution in the population

P-VALUES & CONFIDENCE INTERVAL

Significance level = how much tolerance we have for error due to random chance in our results

α = 0.05

P - Value = probability that your result from your statistical significance test is due to random chance

$P < 0.05$ (5%) = result is likely not due to random chance

reject the null hypothesis if it's greater than or equal to .05 than you fail to reject the null

*you can determine the cutoff

Confidence Interval =

- ↳ CI = (1- **α**)
- ↳ provide a range of means that the true population mean is within
- " 95% of the confidence intervals will cover the true population mean"

* population mean = original population that you're sampling from to try to make some inference about