UNIT V: Analysis of Non-numerical and Numerical Data

SWK 3300
Kimberly Baker-Abrams

In qualitative research:

analysis is on going (occurs as data is gathered)
must be careful not to draw conclusions
before all data is gathered

Qualitative data analysis is less standardized than quantitative analysis because:

- researchers rarely know specifics at beginning
- analysis does not draw on established body of knowledge
- narrative is more imprecise than numbers

"Grounded Theory"

inductive process of discovering theory from data
theory is created from and grounded
in observations and measurements

The basis of qualitative analysis is often a description - this includes:

all collected information
in-depth overview of all aspects of study's constraints

The analysis process occurs by:

- assembling data
- organizing, classifying, and editing
- chronological or thematic narrative
- categories and codes for content
- use of code-book
- use of indigenous categories

"Folk Terms" vs "Cover Terms"

- Folk terms- words or categories used by the subjects
- Cover terms - words or categories used by the researcher

Domain Analysis is:

- Process of putting data into categories
- Looking for: similarities/differences, flow of data, themes and theory that connects data
- Often shown in visual display of data (text p. 226)

In qualitative analysis the emphasis is on finding patterns, understanding events and using models to present what is found.

The alternative hypothesis is:

a test hypothesis (to be ruled out)
the qualitative equivalent for null hypothesis

Triangulation is important because:

the more connected the ideas are to multiple sources
the more likely you have validity

Looking for missing information is as important in qualitative research as is connecting existing data. The missing information is often referred to as "Negative Evidence".

Forms of negative evidence include:

- events that do not occur
- a population not being aware of events or overlooking common events
- a population wanting to hide certain events
- effects of preconceived notions
- unconscious and conscious non-reporting

Addition issues to consider for non-numerical data analysis:

- What points of view are not included?
- What do the events look like from the perspective of broader society?
- Has care been given to be sensitive to vulnerable populations and other social divisions?

Results of analysis should:

be understandable and meaningful to the population that was studied (and broader society as well).

Descriptive statistics:

a means of summarizing the characteristics of a sample or the relationship among variables

Considerations for a frequency distribution:

- accurately describing the number of times a value occurs in a sample
- incomplete data is important to consider
- use of visual representation (text ch. 12)

Measures of central tendency

- Mode (most common)
- Median (middle point)
- Mean (average - most commonly used)

If the values form a normal distribution it will present as a "bell-shaped" curve. In a normal distribution, the three measures of central tendency are equal.

If the values are unevenly distributed the result is a skewed distribution (pulled toward outliers)

("a distribution in which most of the scores are concentrated at one end of the distribution rather than in the middle" text definition).

The measures of central tendency summarize the characteristics of the middle of the distribution, other characteristics of a distribution include the spread, dispersion and deviation. These characteristics are referred to as the measures of variability or measures of dispersion.

Range

- the distance between the highest and lowest values
- used at interval and ratio level
- to calculate subtract the lowest value from the highest value = range

Percentile

a number that divides the range of a data set so that a given percentage lies below this number

Standard deviation

- most comprehensive and widely used measure of variation
- used at ratio or interval levels
- a measure of variability that averages the distance of each value from the mean
- only calculated by hand when there are few cases (see text 246-247)

Analysis that involves two or more variables is called bivariate or multivariate.

Cross-tabulation (contingency table)

- examination of the dependent variable using the independent variable
- useful at any measurement level
- distribution of cases into categories to show which are "contingent" upon other variables

Correlation

- plotted on a chart called a scattergram
- correlation is shown by how closely the values approximate a straight line (+/- or curved)

In using descriptive statistics (whether univariate or multivariate) the most effective way to describe them is to display the results visually.

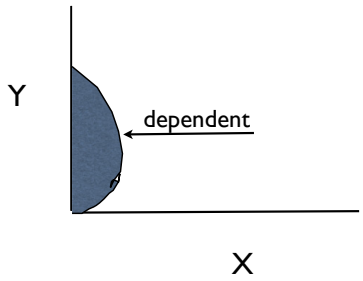
Interpreting graphs

- Level (change = discontinuity)
- Stability
- Trends (pattern within phase = slope, pattern across phases = drift)

In graphing data

X axis = independent variable (horizontal)

Y axis = dependent variable (vertical)



Types of graphs

Graphs and statistics can be deceiving. Be aware of any graph that does not have an absolute zero point for "Y" or that has a discontinuous scale.

Going beyond simply describing the data with descriptive statistics, you will use inferential statistics to test a hypothesis.

Inferential statistics:

- precise method for deciding confidence in results from a sample
- method for deciding if a relationship between variable exists
- relies on probability theory

There are two forms of hypothesis

- Two-tailed (non-directional): states there is an association between variables but gives no prediction about type of relationship
- One-tailed (directional): states that there is an association between variables and gives a prediction for type of relationship

From the hypothesis, a null hypothesis is created. The null hypothesis:

the test hypothesis
asserts that any relationship between variables
is due to chance

A finding is considered to be statistically significant when the null hypothesis can be rejected and the probability that the result is due to chance falls at or below the study's given significance level.

The power of a statistical test is its ability to reject the null hypothesis. The power to reject the null hypothesis increases with the sample size.

Errors in judgment concerning the acceptance or rejection of the null hypothesis are referred to as Type I and Type II errors.

Type I error

rejection of null hypothesis
false conclusion that a relationship exists
when in fact it does not

Type II error

acceptance of null hypothesis
failure to recognize that a relationship does in
fact exist between variables

When a research study has a normal distribution and the dependent variable is measured at least at the interval level and the independent variable has been measured at the nominal level - the differences between the groups of data can be found using a T-Test.

A T-Test

a bivariate statistical test used to determine whether two groups are significantly different based on a particular characteristic

degrees of freedom (df)

- the number of values in the final calculation of a statistic that are free to vary
- you know that value of the mean, therefore it is not free to vary, to calculate df take the total number of cases (-) one (n-1)
- T-test involve two means, therefore to calculate a df take the total number of cases (-) two (n-2)

In social work, it is common for research to involve three or more samples of data to compare. This is when an analysis of variance is used (ANOVA).

ANOVA

used to test differences between the means of three or more groups (one-way anova)

Another way to analyze bivariate data is to examine the strength of the relationships between variables using correlation coefficients.

Pearson's r:

correlation coefficient
used to asses the strength of a relationship
ranges from -1 to +1

The regression line in a scatterplot graph is used to show how much change has occurred in the dependent variable are due to changes in the independent variable (prediction of change).

Multiple regression analysis produces a coefficient that allows each particular outcome (interaction between variables) to be evaluated in direction and the amount of change.

Chi-Square analysis (X^2):

- one of the most widely used statistical tests in social work research.

  As a descriptive statistic: strength of association between variables

  As an inferential statistic: probability that association between variables is due to chance

To use chi-square:

- you must first know what to expect from data (did results differ from expected)
- if a difference exists, there may be an association between variables

  $X^2 = 0$ then independence

  X^2 = more, then there is an association

There are three types of significance in analysis of group data and single-subject studies.

1. Practical or clinical significance
2. Visual significance
3. Statistical significance

auto-correlation

"the relationship between the outcome or dependent variable scores in single-system studies"
(the sample is not statistically independent)

A method to lessen the chance of auto-correlation is using a celeration line:

- connecting the midpoints of two values of the baseline and projecting the line into the intervention period
- if a given proportion of data are on the desired side of the celeration line, then an estimate of significance can be made

In looking at the same results on a curve distribution, if the intervention mean is more than two standard deviations from the baseline mean, then there is a statistically significant change.

In using inferential statistics, make sure to choose the most appropriate test for your data.... and be sure to present the findings in as neutral a manner as possible.