Quantitative Analysis In Political Science Midterm Review

Chapter 1- The definition and measurement of concepts

-> Concept: an ideal or mental construct that represents phenomena in the real world (ex: democracy, party affiliation, liberalism, political participation, social status, etc.)
- in empirical political science, concepts refer to facts, not values.
- concepts much vary and they must be concrete

- primary goals of political research are to describe content and to analyze relationships known between them.

-> Conceptual question: a question expressed using ideas. frequently unclear and difficult to answer

-> Concrete question: a question expressed using tangible properties. can be answered empirically.

-> Conceptual definition: clearly describes the concept's measurable properties and specifies the units of analysis(people, nations, states, etc)

-> Operational definition: describes the instrument to be used in measuring the concept and putting a conceptual definition "into operation". (ex: to measure liberalism, set of questions asked to raise, keep the same or lower government spending on things like education, welfare etc. Liberalism could be operationally defined as the number of times a respondent said "increase" because it shows more liberal attitudes)

-> Conceptual definitions: defining a concept by its empirical meaning. settling on a set of properties that best represent the concept and write down the definition of the concept. Must communicate 3 things:
1. the variation within a measurable characteristic or set of characteristics
2. the subjects or groups to which the concept applies
3. how the characteristics is to be measured

-> Example: "The concept of ***(economic liberalism)*** is defined as the extent to which ***(individuals)*** exhibit the characteristic of (**supporting government spending for social programs**)"

-> Conceptual dimension: defined by a set of concrete traits of similar type.

-> Multidimensional concept: has two or more distinct groups of empirical characteristics. Each group contains empirical traits that are similar to each other

-> Unit of analysis: the entity(person, city, country, university, agency, etc) we want to describe and analyze. They can either be individual level or aggregate level.

-> Individual level unit of analysis: when a concept describes a phenomenon at its lowest possible level. (ex: most survey research deals with concepts that apply to individual persons)

-> Aggregate level unit of analysis: a collection of individual entities(ex: neighbourhoods are aggregate level units. ex: a university administrator who wondered if student satisfaction is affected by class size would gather information on each class, an aggregation of individual students.)

-> Ecological fallacy: a classical problem that arises when an aggregate level phenomenon is used to make inferences at the individual level.

-> Operational Definition: describes how the concept is to be measured empirically.

-> Intended characteristic: the property we want to measure

-> Unintended characteristic: the property we do not want our instrument to measure

-> The Hawthorne Effect: ex: test anxiety may make a student score low on their test because of them getting nervous, despite their grasp of the subject.

->Systematic measurement error: introduces consistent, chronic distortion into an empirical measurement. Consistently mis measures the characteristic the researcher is after.

-> Random measurement error: introduces chaotic distortion into the measurement process, producing inconsistent operational readings of the concept.

-> Reliability of a measurement: the extent to which it is a consistent measure of a concept.

-> Validity of a measurement: the extent to which records the true value of the intended characteristic and does not measure any unintended characteristics. If a measurement is reliable, it will have consistent results.

<u>Evaluating Reliability</u>

-> Test-retest method: method to view consistency. the investigator applies the measure once and then applies it again to the same units of analysis. If the measurement is reliable, then the two results should be the same or similar. If a great deal of random measurement error is present, the two results will be very different.

-> Alternative form method: method to view consistency. the investigator administers two different but equivalent versions of the instrument- one form at time point 1 and the equivalent form at time point 2.

-> Split half method: method to view consistency. based on the idea that an operational measurement obtained from half of a scale's item should be the same as the measurement obtained from the other half. the investigatorr divides the scale items in two groups, calculates separate scores, and then compares the measurements. if the items are reliably measuring the same concept, the two sets of scores should be the same.

Evaluating Validity

-> face validity approach: method to evaluate validity. the investigator uses informed judgement to determine whether an operational procedure is measuring what it is supposed to measure.

-> construct validity approach: method to evaluate validity. the researcher examines the empirical relationships between a measurement and other concepts to which it should be related.

Atanasoff- Berry Computer

- invented and built in 1937-1942
- to speed up solutions of mathematical and statistical problems
- no patent
- US district Court in 1973

-> recognized Juhn Atanasoff as the inventor of the electronic digital computer

- Anulled ENIAC computer patents

Variables in SPSS (Statistical Package for the Social Sciences)

- *Data Editor:* mode for viewing, entering and editing data

-> data view (for data entry)
-> variable view (for entering information about each variable)
-> variable name (usually one word)
-> variable label (usually more than one word)

- *Output:* where statistical results and graphics are retrieved

-> can be exported/saved as a Microsoft Word file

Basic operations in SPSS (Statistical Package for the Social Sciences)

- *Value label:* the label attached to a code (ex: 1=male, 2=female)
- *Missing Values:* typically negative values (-5, -1, etc) or every large values (9999, 9998, 999, etc)

-> make sure that you check and define as missing such values before doing any analysis

Chapter 2- Measuring and describing variables

-> Variables: empirical measurement of a concept. Contains different values. (ex: political party affiliation in Canada variable has the following values- conservative, liberal, NDP, etc.). Always have at least one name and 2 values. (ex: marital status is the name of the variable and widowed/divorced are the values)

-> World Dataset: used to identify variables and values.

Types of Measurement Levels/Variables

- **Nominal:** communicates differences between units of analysis on the characteristics being measured. (ex: numbers 1-5 are associated with each value of marital status. 1 means "married" and 5 means "divorced". 1 is different from 5. Marital status is an example of nominal level variable. Numbers simply show they are different categories)

- **Ordinal:** communicates relative differences between units of analysis where a level of measurement describing a variable with values can have a rank-order along some dimension but cannot find the average value The ranking is reflected in the variable's numeric code. . (ex: scale for support for same sex marriage which can have four values- 1=strongly oppose, 2=oppose, 3=support, 4=strongly support)

- **Interval/ratio:** communicates exact differences between units of analysis. values are rank-ordered and have equal distances between adjacent values. (ex: age measured in years is an interval level variable: 18 years, 24 years, 77 years, etc. How much of a difference is there between two units of analysis? Income-$ is another example. Variables values do not need to be represented by set of numeric codes.)

-> Additive Index(ordinal style measurement/like the Likert Scale): provides a more precise and reliable measurement of a characteristic (ex: ten government programs and add up the number of times they say "increase spending". Scores on the index would range from 0 to 10).

Describing Variables

-> Central Tendency: a variables central tendency is its typical average value. Measured in three ways:

- Mode: most common value of the variable, the value that contains the largest number of units of analysis. Used to describe the central tendency of any variable. For nominal level variables, it is the only measure that may be used.

- Median: used to describe higher levels of measurement for ordinal and interval variables. it is the middle value. Comes closest to the average. Looks at the highest and lowest values, and goes to the middle value.

- Mean: used to describe interval level variables. Comes closest to everyday average. Sum up all the cases' individual values and divide it by the number of cases to come up with the mean value.

-> Dispersion: variable can also be described by this. The variation or spread of cases across its values.

-> Frequency Distribution: tables or charts that summarize the distribution of variables and variables' values. Commonly used in data presentations of all kinds- survey research, marketing studies, journalistic polls, etc.).
- First column of each frequency distribution lists the variable's values
- The second column reports the number of individuals at each of the variables values, totalled at the bottom column. (The total frequency)

Frequency Graphs/Charts
- graphs/charts can be used to illustrate frequency distributions

-> Bar chart: most appropriate for nominal and ordinal variables
-> Pie chart: most appropriate for nominal and ordinal variables with small number of values (categories)
-> Histogram: most appropriate for interval-ratio variables with large number of values

-> Cumulative percentage: records the percentage of cases at or below any given value of the variable

-> Percentile: reports the percentage of cases in a distribution that lie below it.

-> Bimodal distribution: when a frequency distribution has two different values that are heavily populated with cases.

-> Interquartile range: defined as the range of a variable's values that defines the "middle half" of a distribution- the range between the upper half and the lower half of a bar graph

Chapter 6: Foundations of Statistical inference

-> Descriptive statistics: describes and summarizes data

-> Inferential statistics: inferences about data, hypothesis testing. set of procedures for deciding how closely a relationship we observe in a sample corresponds to the unobserved relationship in the population from which the sample was drawn. (ex: 5 point feeling difference thermometer between men and women represents real gender difference in the population in preference of vote)

-> Population: the universe of cases the researcher wants to describe(attitudes, beliefs, behaviours, of large groups)

-> Population parameter: figuring out a population's characteristics, its parameters, etc.

-> Census: allows researcher to obtain measurements for all members of a population.

-> Sample: a number of cases or observations drawn from a population. Researchers use this to examine a population. Sampling is used in survey research. It is absolute size of sample that is most important. As sample size increases, sample errors decrease

-> Population: all cases about which one seeks knowledge to know of

->Coding: transforming a measure into numbers (ex: those who say they are single coded as "5")

Designing Survey

- Research question and/or hypothesis
- Selecting sample
- Questionnaire

-> language/languages

- Type of survey
- Pilot testing

Type of Samples

- *Representative sample:* a sample containing the essential characteristics of the population

- *Probability sample:* a sample selected using a random process such that each element in the population has a known probability of being selected

- *Weight:* makes a sample representative when there are small deviations

-> example: many samples in World Values Surveys

Non-probability Sampling

- *Convenience sample:* participants / respondents are selected based on their convenient availability (ex: students in one class)

- *Snowball sample:* the researcher makes contact with some individuals, who in turn provide contacts for other participants (ex: sometimes students who participate in survey studies are asked to come up with the names of some non students who may be willing to participate)

Non probability sampling(generally not representative of entire population)

- *Purposive sample:* non random selection of participants / respondents based on a specific purpose of a project (ex: sample of Facebook users)

- *Volunteer sample :* self selected participants

- *Quota sample:* set quotas for specific groups

-> Sampling error: errors of estimation that occur as a result of differences between the characteristics of the sample and those of the population

-> Non response: occurs if people selected for the sample do not supply the required data

-> The response rate: the percentage of the sample that participates in the study

Standardized/ Structured interviews and questionnaires

- *standardized/structured interviews and questionnaires:* used in survey research. standardizes asking and recording of answers. The question should be asked exactly how stated in interviews. Closed and open ended questions.

- *questionnaires:* contains a list of survey questions and answers. interview schedule. helpful in analyzing survey data.

Types of Surveys

- Telephone surveys (cheaper and quicker to administer, cant exceed 30mins)
- Online surveys (quick and cheap, but some don't have internet access)
- Face to face Surveys (time consuming and costly)
- Mail surveys (questionnaires, no interview needed, expensive, low response rate
- Omnibus surveys (surveys comprised of small surveys, cheap)
- Census (surveys entire population, very expensive)

Respondent problems in survey research

- *acquiescence:* the respondent is just trying to please the researcher

-> ex: the respondent may agree with all the items, or disagree with all terms, just to be "cooperative"

- *social desirability:* repsondents may be reluctant to provide a sincere answer if doing so will make them appear to be unlikeable or bad people

- *laziness or boredom:* may prevent a respondent from giving a genuine response

Chapter 3- Proposing explanations, framing hypothesis, and making comparisons

-> Hypothesis: a testable statement about the empirical relationship between cause and effect.

-> Dependent Variable: a variable assumed to depend on the independent variable, a variable we are interested in explaining

-> Independent Variable: a variable which presumed to affect the dependent variable, a variable which we use to explain the dependent variable.

-> Intervening variable: a variable that acts as a mediator between an independent variable and a dependent variable. (describes how the independent variable is linked to the dependent variable)

-> Cross tabulation: a table that shows the distribution of cases that have different values on an independent variable.

Hypothesis Testing/ Cross-Tabulation Analysis

- *Bivariate Analysis:* determines whether there is a relationship between two variables. (Bivariate Table of percentages, where dependent variable is in rows and independent variable is in columns. percentage totals are column totals)

- *Tests of statistical significance:* formal and exact way to test hypothesis. Is a relationship between dependent and independent variables statistically significant?

- *Research Hypothesis (H1):* a statement about relationship between independent and dependent variables that we want to prove or disprove(ex: membership in political parties in Canada affected by level of education?)

- *Null Hypothesis (H0):* a statement of “no difference” between independent and dependent variables(ex: membership in political parties in Canada not affected by level of education)

-> Two types of errors: Type I=rejecting a true null hypothesis, Type II=not rejecting a false null hypothesis.

Graphing relationships and Patterns

- *Direct relationship:* a relationship that runs in a positive direction(increase-increase)

- *Inverse relationship:*a relationship that runs in a negative direction(increase-decrease)

- *Linear relationship:* an increase in independent variable=consistent increase/decrease in dependent variable

- *Negative relationship:* lower value of independent variable is associated with higher value of dependent variable

- *Positive relationship:* lower value of independent variable is associated with the lower value of the dependent variable

- *Curvilinear relationship:* the relationship between the independent variable and dependent variable depends on which interval or range of the independent variable is being examined. The relationship may change directions from positive to negative, etc.

Chapter 4 - Measure of Association(Making controlled comparisons)

-> Test group: composed of subjects who receive a treatment that the researcher believes is causally linked to the dependent variable. (ex: the patients who received the surgery)

-> Control group: composed of subjects who do not receive the treatment that the researcher believes is causally linked to the dependent variable(ex: the patients who didn't receive the surgery)

-> Research design: overall set of procedures for evaluating the effect of an independent variable on a dependent variable

-> Experimental design: ensures that the test group and the control group are the same in every way, except one- the independent variable.

-> Controlled comparison design: allows the researcher to observe the effect of the independent variable on the dependent variable

Types of experimental designs

- *Laboratory experiment:* the control group and the test group are studied in an environment created wholly by the investigator

- *Field Experiment:* the control and test groups are studied in their normal surroundings, living their lives as they naturally do.

- *Random assignment:* occurs when every prospective participant has an equal chance of ending up in the control group or the test group.

Chapter 7- Measure of Association and Chi Square Test (x2)

-> Chi Square Test of Significance: helps you decide whether an observed relationship between an independent variable and a dependent variable really exists in the population or whether it could have happened by chance when the sample was drawn. Most appropriate for nominal and ordinal variables.

-> Measure of association: tells the researcher how well the independent variable works in explaining the dependent variable. Provides information about the strength of relationship between variables.

- Cramer's V: appropriate for nominal variables
- Gamma: appropriate for ordinal variables.

Criteria of Strength of Association

- Cramer's V: varies between 0 (no association) and 1 (perfect association)
- Gamma : Varies between -1 (perfect negative association) and 1(perfect positive association)
- 0: no association
- 0-0.1: weak association
- 0.1- 0.3: moderate association
- 0.3 - 1.0: strong association

Directions of Association

- *positive association:* relationship where the variables vary in the same direction

- *example:* positive association between income and education level

- *negative association:* relationship where the variables vary in opposite directions

Independent samples t-test

- used to test research hypothesis

-> example: men and women in Canada have different levels of confidence in political parties

- determines if differences between means of two independent groups/sub samples are significant

- *independent groups/subsamples:* if one member/case in a group does not influence place of another member/case in that group or another group

-> example: male and female respondents are independent groups/subsamples