Class 2

Stats is already based on logic that we have from the real world.
Sampling Frame: What you have access to.
Variables: something that could vary, and something we're trying to assess.
- Attributes OR Quantities
Values: are the numbers or categories that a variable can take
Scores: a value that a variable can take

Variables: Physical or abstraact attributes or quantities that we wish to measure. A variable can take on specific values.
- Score: is a value that an individual has on a particular Variable

Quantitative Variable: They are number values, often reflect how much of some quantity and individual posesses.

Ratio of scales: Have Units of measurment and an Absolute Zero. It is the ratio of two values away from zero. It allows us to measure the relative distance from one point to another. For example if person one is 3 feet tall and person two is 6 feet tall, person two is 2x taller that person 1. CONTINUOS

Quantitative Variables can be Discrete or Continuos.

Discrete: Whole numbers, 1,2,3,4,5,6,7,8. For example: there cannot be 27.8 students in a class, so the # of students in a class is a discrete variable.

Continuous: Real Numbers, can have decimals. For example height. Continuos variables can be any number

Quantitative can be discrete
QUALITATIVE are always discrete
Ordinal Variables are always discrete

Interval scales: When the unit of measurement is the same, but no absolute point of zero.
- Fairenheit or celsius, do not indicate an absence of temperature, meaning there is no null. 12 degrees celsius is not twice as hot as 6 degrees celsius.
(Unit of measurement as in, the distance between 1 degree celsius and 2 degrees is that same as the distance between 5 degrees and 6 degrees)
All ratio scales are equal interval scales but not all equal interval scales are equal ratio scales.

Qualitative Variable: Have variables that are qualities or categories, they are referred to as nominal or categorical Variables.

Categorical Variables can be discrete categories.
Nominal Variables are simply names

So we must use the NOMINAL SCALE: Because unlike Quantitative we cant naturally order the variables.

ORDINAL VARIABLES: THese rest somewhere between Qualitative and Quantitative. For example, University Rankings are Quatitative as they have a natural order. First-Second-Third. However a ranking is not an equal interval scale, because the difference in quality between first and second is not the same as 4th and 5th. Because there is a natural ordering to scores on an ordinal variable, but no unit of measurment, we say that ordinal variables are measured on an ordinal scale.

Steven's Hierarchy of Scales

Ratio Scale
- Natural ordering
- Unit of measurment
- An Absolute Zero

Interval Scale
- Natural ordering
- Unit of measurment

Ordinal Scales
- Natural ordering
- No Unit of measurment

Nominal Scale
- Unordered Categories

You cant chang what would be a ratio scale by putting them into categories to an ordinal 1)10lb-20lb 2) 21lb-30lb etc...

Measurement: REQUIRE MEASUREMENT TOOLS

Psychologial Construct: a hypothetical attribute that is thought to explain some aspect of behavious, but which cannot be measured directly with a measuring utensil.
- Intelligence
- Depression

-Happiness
-Sociability

Conceptualize: provides meaniting to one construct in abstract or theoretical terms
Opterational: Defines what I am going to measure and define how I am going to measure it.
Operational Measures: tools used to measure psychological constructs> more often than not its a questionannaire to capture the score which will give the abstract notion of intelligence.

Conceptual: Making a definition
Operational: defining how we are going to test that definition

Reliability: How many time your measuring tool can measure the same object and gives us the same answer
Nothing is perfectly reliable.

Validity: If it measures what its supposed to measure

Two ways of looking at operationalized variables.
- The properties of measurments themselves
-The relationship between the numbers themselves and the psychological construct they supposedly represent.

Measurement Error: No measuring device is perfect because no measuring device is perfectly reliable and objects often vary about some average value.
- any measurement is an estimate of your theoretical average value.

Population of situation: everything possible score within your group of interest.
Sample: a subset of said group of interest.

Parameter: is a characteristic of a population
- example: average heartrate of canadians

Statistic: characterisitic of a sample:
Example: average heartrate of sample of canadians

Inferential statistics
- when we infer population parameters from sample statistics we are doing inferention statistics
- if I am using a stat from a sample to make a prediction of a population its an inference
- If I use a stat from the sample to make a statement about the sample it is descriptive.

Random Sampling: Means that all members of the population had an equal chance of being in the sample. Ex: the jelly beans had an equal shot of being scooped.

Sampling Bias: Means that not all members of the population had an equal chance of being selected in the sample.

Sampling error: is the difference between the sample statistic and the population parameter.

Class 3

Chapter 2

Distribution: Conveys the relative frequency with which values of a variable occur in a sample or population. Also conveys the relative standing of scores within a sample or population.

n= Sample/population size

Frequency Table (Qualitative): Conveys the number or proportion of scores in a sample or population having each value of a variable.

The variable here is qualitative - Area Preference.

60 participants chose their favorite brand. f = frequency. Also known as a raw frequency count.

The number of people that chose a specific brand is known as raw frequency counts.

The proportion of participants preferring each brand is marked by the symbol p Otherwise known as a relative frequency

You get this by dividing the frequency of a chosen brand by the total number of participants.

THIS INFORMATION CAN BE CONVEYED GRAPHICALLY USING A BAR GRAPH

X-axis: different brand names = area preference

Y-axis showing the proportion of each student's preferred brand.

Frequency Tables in Discrete Quantitative Variables

Conveys the number or proportion of scores in a sample or population having specific values of the variable. The frequency table may also convey the number or proportion of scores at or below a given value of the variable.

Difference between Quantitative frequency tables and Qualitative:

Quantitative measures Cumulative Frequency = Cumulative f: The number of scores at or below a given value of a variable.

Cumulative Population also known as P: the proportion of scores at or below a given value of a variable.

Percentile rank: Cumulative proportion (P) multiplied by 100.

HISTOGRAMS OF A DISCRETE QUANTITATIVE VARIABLE

A graphical depiction of the number or proportion of scores in a set having a specific value of a variable.

Differences between a bar graph and a Histogram

Unlike in a bar graph, there are no spaces between the bars in a histogram. When the bars tough it shows the continuity.

There is no natural ordering to a qualitative variable so you cant make a histogram.

Grouped frequency Tables
- Intervals: the range of possible values on a quantitative variable
-Score limit (ex: 80-84)
Real Limit: 79.5-84.5

INTERVAL MIDPOINT = max-min /2

PROBABILITY

Proportion of succeses is the number of successes devided by the number of trials
an event is one or more of the possible outcomes of the sampling expirament.

Probability distributions: Discrete Variables

- The proability distributions for a discrete variable simply expresses the probability of each value of the variable

Distribution of scores:

- Function: There is a single value of y for every possible x value
    - ex y=4x
- Density: Number of things per interval

CLASS 5:
CENTRAL TENDENCY

mean: Average

The fancy E means sum
learn rules for when to uses which* found on the slides.***
OVERALL MEAN is known as a GRAND MEAN.
(n1*m1) + (n2*m2) / n1+n2

MEDIAN: haveway point of all marks

1) sort all scores
2) Count all scores
3) find the score in the center.

- THIS IS BEST WHEN THE DISTRIBUTION IS SKEWED.

MODE: most frequent in a data set.
1112222333445**66666**778889999

- looking for the most typical case (Nominal data)

Bimodal has multiple mode. there is not one particular popular trend.

VARIABILITY or Dispersion: how spread out the scores are in the distribution.

RANGE: Difference between highest point and lowest point.
SUM OF SQUARES IS TOTAL VARIABILITY
VARIANCE: TOTAL VARIABILITY AVERAGED.
2 3 5 7 8

1) Deviation
    a) 2-5 = -3
    b) 3-5= -2

c) 5-5=0
d) 7-5=2
e) 8-5=3

2) SQUARE
   a) 9
   b) 4
   c) 0
   d) 4
   e) 9
3) Sum of all squared deviation / n
   a) 26/5= 5.2
   b) n-1 when it is a sample
   c) N as is when its
4) Standard deviation
   a) Square root your average squared deviation from the mean.

Adding a constant (c) to all scores in a mean has no effect on variance or standard deviation.

Standard deviation is a more intuitive quantity to compare the variability.

Given that I have this probability of occurence how likely is not null hypothesis.

1) -1.3
2) 1.67
3) 0.67
4) -.66

Critical Z: How far one mean is from the other

.4938

Critical Z of 5% is always 1.645
Critical Z of 2.5% is always 1.96
Mean= 80
alpha .05
Std= 16

another sample
n = 64
Std =16

standard error = 2