Collect data

**Categorical name** can be present values

**Numeric values** show how much of something

**Discrete variable**: derives from accounting process

**Continuous variable**: derives from measure process

Measurement scale:

*Interval scale*: categorizes, rank order items

*Ratio scale:* has equal intervals, has true zero point

*Nominal scale*: categorizes items; like gender, languages

*Ordinal scale*: categorizes ranking items

Collect date from:

Primary source

----statistical studies: observational

Secondary source

----business databases

----public business activities

----survey responses

Original variable; like major, income level

Recoded variable; like social class

Judgment sample: collect data from pre-selected experts in the subject matter, can’t be generalized

Probability Sampling methods

Simple random sample: every item from a frame has the same chance of selection as every other item.

Systematic sampling: choose the first item to be selected at random from the first K items in the frame.

Stratified random sampling: elements in the population are divided into groups called strata such that each element in the pop belongs to only one stratum.

Cluster sampling: elements in the population are first divided into groups called clusters.

Ethical issues

Researcher bias – coverage and nonresponse error manipulation

Sampling errors not reported

Measurement errors

---researchers using biased questions

---participants lying

CH 2

A statistician is someone who insists on being certain about uncertainty

Summary table: tallies frequencies/% of items in a set of distinguishable categories

Contingency table: cross-tabulates the value of 2 or more categorical variables to see pattens

Ordered array: the value of a numerical variable in order from smallest to largest value.

Bar chart: represent the frequency or % for each category

Pie chart

Pareto chart: vertical bars in descending order

Side-by-side bar chart

Histogram: represent class interval from a frequency or percentage distribution

Percentage polygon: midpoint of each class interval to represent data of each class and then plots the midpoints

Ogive: cumulative percentage destructions to plot cumulative percentages along the Y axis

Scatter plot: graphical representation between the relationship of 2 variables

CH 3

Mode: value occurs often

Range = largest value --- smallest value

Z score = (x—AVG(x))/S

Skewness is a numerical measure of the shape of a distribution

Kurtosis: measures the peakedness of the curve of a distribution

Quartiles: splits data values into four equal parts.

Q1 = first quartner = (n+1)/4

Q3 = third quarter = 3(n+1)/4

Interquartile range = Q3 – Q1

Boxplot 5 factors: Min, Q1, Median, Q3, Max

Sxy = Sum(x-AVG(x))(y-AVG(y))/(n-1)

=CORREL(A#:A#,B#:B#)

Probability: is the numerical value to present chance, likelihood, or possibility that a particular event will occur.

Probabilistic predictions reduce uncertainty.

Increasing likelihood of occurrence varies from 0-1

3 types:

--priori: an occurrence based on prior knowledge of the process involved

Simple pro = X/T

--empirical: probability based on observed data from a survey or an experiment

--subjective: pro that differs from person to person

Events, sample space

Requirements:

Outcome must be 0-1

The sum of the pro for all outcomes must be = 1

Joint event: an event has two or more characteristics.

Sample space = all possible outcome

Decision tree: an alternative to the contingency table represents all possible decisions.