Equations!

X = Categories of IV

f = frequency of scores

∑ (sigma) = sum (to add something up)

Relative Frequency (*rf*) = *f*÷*N*

N = total number of scores

Cumulative frequency (*cf*) = start at bottom *f* and add up

Cumulative relative frequency (*crf*) = *cf*÷*N*

Range = Max # - Min #

Population mean = μ

Sample mean = M or $\bar{x}$

Deviation = x - μ or x-$\bar{x}$

Variance = $\Sigma(x-\bar{x})^2 \div N$

Standard Deviation (SD) = √Variance OR $\sqrt{SD^2}$

Pearson's coefficient of skew = $3(\bar{x}-Mdn) \div SD$

Types of scales of measurement!

1.) **Nominal** ("categories of"):
- No quantitative distinction between observations
- Categories are equivalent and discriminable: one is not better than or higher than the other(s) and can be distinguished from each other
- how many items/people are in one category/group
- do not need/include *crf* or *cf*
- Cant create stem and leaf display

2.) **Ordinal** ("more of"):
- the data can be categorized and ranked
- Cant create stem and leaf display

Types of scales of measurement! (cont)

3.) **Interval** ("how much of"):
- the data can be categorized and ranked, *and evenly spaced* (e.g., temp)
- Arbitrary zero, therefore, cannot speak meaningfully about ratios
- could have negative numbers

4.) **Ratio** ("Proportion of"):
- Equal intervals between objects represent equal differences (Eg., money)
- Has a meaningful zero

How we describe data

"Bell–shaped" curve	Kurtosis
- Normal distribution, Gaussian distribution	- degree to which data values are distributed in the tails of the distribution
	platykurtic distribution = low degree of peakedness (<0)
	normal distribution = mesokurtic distribution (0)
	leptokurtic distribution = high degree of peakedness (>0)

Definitions!

Descriptive statistics: Organizes, summarizes, and communicates a group of numerical observations

Inferential statistics: Allows tests of hypotheses using systematic, objective procedures

Discrete numbers: separate, indivisible categories (eg., 4 or 5 children, not 4.34 children)

Continuous numbers: infinite number of values fall between any two observed values (eg., Age, height, weight, time)

Independent variable (IV): Feature(s) of a study that is/are used to explain or explore the participants behaviour

Dependent Variable (DV): Behaviour of the participants that we are observing, measuring, or recording

Cumulative relative frequency (crf):proportion of scores at or below a particular score

Cumulative frequency (cf): frequency of scores at or below a particular score

Relative frequency (rf): fraction of the total group associated with each scores

Modality:the number of peaks in a frequency distribution of data

positive skew: a lot of data on the lower end of the distribution

negative skew: a lot of data point on the higher end of the distribution

Semi-interquartile Range (SIQR): the distance of a typical value from the median

Definitions! (cont)

Median Absolute deviation (MAD): Absolute measure of how many physical units values deviate from the median

Sum of squared deviations

1.) Compute $\bar{x} = \sum x \; X \div N$
2.) Compute the squared deviation for each score: $(x-\bar{x})2$
3.) Compute the sum of squared deviations (SS)
4.) Divide SS by N for the mean of squared deviations

Graphic Figures!

If you have *nominal or ordinal* data: use **BAR GRAPH**

If you have *Interval or Ratio* data: use **HISTOGRAM**, **LINE GRAPH**, or **POLYGON**

Measures of Central Tendency!

1.) **Mode** (*Mod* or *Mo*)
- most frequent category/score in a distribution
- **ALWAYS** a value that is observed in the dataset
- No inferential statistics
- May not be representative

2.) **Median** (*mdn, md or* $\bar{x}$)
- Physical middle of an ordered set of data (aka, 50th percentile rank)
- less biased when interval/ratio data are severely skewed
- not affected by outliers or extreme scores

Measures of Central Tendency! (cont)

- No inferential statistics

3.) **Mean**

- Average of all numbers
- Most common value used for descriptive/inferential analyses
- Applied only to interval/ratio data
- Is biased if the scores are strongly skewed

Data and Central Tendency!

Nominal: Mode
Ordinal: Mode, Median
Interval/Ratio: Mode, Median, Mean

Measurement and Variance!

Nominal: none
Ordinal: range, SIQR, MAD
Interval/Ratio: Range, SIQR, MAD, variance, SD

Interpretation of skew value

Range of Values	Skew	Data
Between 0 and 0.5	Normal distribution	Use Mean and SD
Between .5 and 1.0	Mild to moderate skew	Use Mean and SD
Between 1.o and 2.0	moderate to strong skew	Use Mean and SD if closer to 1.0 than 2.0
Greater than 2.0	Severe skew	Use Median and MAD

Measures of Variability!

1.) **Range**

- Distance covered by scores in a distribution from the smallest score (min) and largest score (max)
- unreliable: sensitive to extreme values
- least preferred option of measures of variability

2.) **Semi-Interquartile Range (SIGR)**

- Half the range of the middle 50% of observations
- Can be used with ordinal, interval, and ratio scales
- Not affected by outliers or extreme scores
- Some values in the distribution are excluded

3.) **Median Absolute Deviation (MAD)**

- *How to calculate it*

→ Find the median of the data set
→Compute the absolute deviation of each value in the data set from the median
→Subtract the median from the value
remove +/- (if they apply)
→Order the absolute deviation values from low to high:
→Find the median of the ordered deviation values: Mad

Measures of Variability! (cont)

- less sensitive (than standard deviation) to extreme scores or skews in data
- not useful in advanced statistical procedures

4.) **Variance**

- average squared distance from the mean

– for computing descriptive statistics only

5.) **Standard Deviation (SD)**

- measure of the standard/-average distance from the mean (how dispersed the scores are around the mean)
- sensitive to extreme scores or outliers and is therefore biased with skewed distributions

Symmetrical vs. Skewed !

Symmetrical	+ Skewed	- skewed
Mean and median are always the same (in the middle)	mean the closest to the tail end	
mode varies	mode is where the peak is	
	median is in between	
	Tail pointed towards high #	Tail pointed towards low #

Use Median and median absolute deviations for extremely skewed data