STATISTICS

Scales of measurement

Scale	Description	Examples
Nominal	Same vs. different. Qualitative. Categorical (each category has own unique identity).	Sex (male vs. female). University. Country.
Ordinal	Same vs. different. Greater than vs. less than. Qualitative.	Order of finishing in a race, doesn't matter how long gap is between them (still have 1st, 2nd, 3rd etc.) Friendship (not intervally spaced). Attractiveness (more attractive, about the same, less attractive).
Interval	Same vs. different. Greater than vs. less than. Equal intervals. Quantitative. Equally spaced intervals but no absolute zero.	Temperature—can measure, difference between degrees is always the same no matter what level of temp, no absolute zero temperature (absence of temperature, 0°C is arbitrary because different in Fahrenheit).
Ratio	Same vs. different. Greater than vs. less than. Equal intervals. Absolute zero. Quantitative.	Distance or length—absolute 0 so can make ratios (e.g. 180cm person is twice as tall as 90cm person). Weight. Level of income.

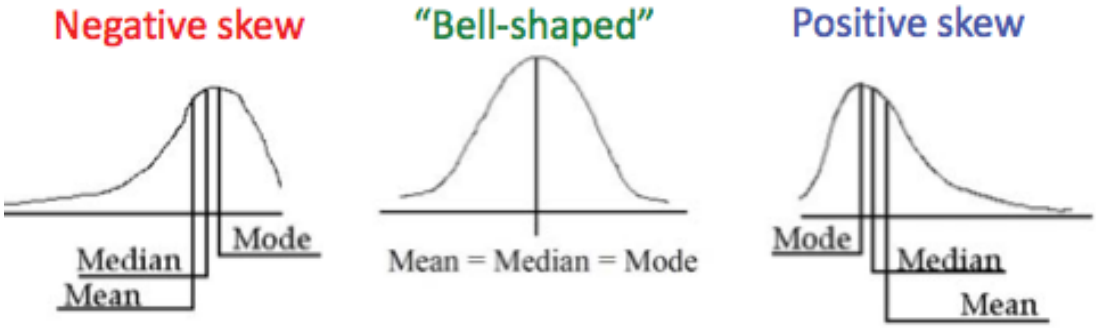
Descriptive statistics

Used to succinctly describe data in general terms so you can make statements about the data.

Measure	Description	Pros and cons
Mode	Score (x value) with highest frequency, most common	May not be unique (e.g. bimodal). May not exist (e.g. rectangular).
Median	Score that divides the distribution into 2 (roughly) equal parts. (n + 1)/2 in set of ordered scores.	Not affected by outliers (e.g. 7, 8, 100), can get rid of them so they're not biasing measure of central tendency. But, not sensitive to all scores (may go from 2 extremes with nothing in the middle, e.g. 7, 7, 8, 98, 100).
Mean	Balance point of the distribution. Population mean (mu) = sum of all scores (sigma X) / number of observations.	Sensitive to all scores in the distribution. Knowing the mean of a distribution helps you interpret a score—e.g. with grades can see if they are better than average, which may differ greatly between subjects (so the exact number may not be the most meaningful thing).

Distribution and central tendency

Relationship between mode, median and mean changes depending on the distribution.



Very different distributions can have the same mean, which limits its usefulness.
E.g. bell curve and bimodal distributions are both symmetrical and can have the same mean, so it doesn't actually tell you much about the distribution.

Bell curve: scores more tightly clustered around the mean so less variance in the scores.
For bimodal and flatter bell curve, scores are more widely dispersed so more variance.

Deviation scores

Deviation score = how far away a score is from the mean (X – u).
Negative answer means below the mean, positive means above it.

The mean is the value with the smallest average deviation score. The sum of all deviation scores in a set of observations is 0.

Summation notation

Data: X: 7, 7, 8, 10, 11 (rewrite: $X_1=7$ $X_2=7$ $X_3=8$ $X_4=10$ $X_5=13$)

Finish here

$$\boxed{\sum_{i=1}^{N} X_i} = X_1 + X_2 + X_3 + \ldots\ldots + X_N$$

Start here

Count these

Shorthand

Longhand

$$\sum_{i=2}^{3} X_i = X_2 + X_3$$

$$\sum_{i=1}^{2} X_i^2 = (X_1)^2 + (X_2)^2$$

$$\left(\sum_{i=1}^{N} X_i\right)^2 = (X_1 + X_2 + X_3 + X_4 + X_5)^2$$

i = 2 means start from the 2nd score.