Demystifying statistics

(or what do those little letters mean?)

Do these groups differ?

WF	TL
55%	70%

- Different children will perform slightly differently, and there will thus be a distribution of scores

Do these groups differ?

WF	TL
55%	70%
40%	45%
60%	50%
57%	72%
70%	55%

- Statistics provide us with a means of determining whether an apparent difference is consistent enough to be real.

In order to do statistics…

- Outcome measures must be numerical
- Need to be able to define who fits in which group
  - If you wish to compare language-delayed kids with those who are not.
    - What does it MEAN to be language-delayed?
    - What is the cut-off?
  - If you want to compare elderly with normal hearing to elderly with hearing loss.
    - What is NORMAL hearing?
      - Is it having thresholds at 0dB? Or less than 10 dB?
    - What is hearing loss?
      - Is it having a loss at all frequencies? Only at 8000 Hz? Is it a loss of 10 dB? Or 80 dB? Or anything in between?

Definitions

- Conceptual definition
- Operational definition
  - Defined in terms of
    - A procedure (set of operations)
    - How the concept is measured
- Many operational definitions may be possible for the same concept.
- It is important to keep the underling concept separate from the way that they operationally define that concept.

Scales of measurement

- **Nominal**
  - Numbers as labels
  - categories bear no numerical relationship to each other; they are qualitatively different
- **Ordinal**
  - Ranked order
  - Not equal distance between rankings

Chocolate Spinach Asparagus

Scales of measurement, cont.

- Ordinal scale categorizes data into ranks
  - e.g. order in competition (1st, 2nd, 3rd, . . .)
  - attitude scale (strongly agree, agree, neutral, disagree, strongly disagree)
- Intervals are not meaningful. So:
  - 1, 2, 3, 4, 5
  - 1, 7, 33, 95, 783
  - strongly agree, agree, neutral, disagree, strongly disagree

  are identical in ordinal terms

Scales of measurement, cont.

- **Interval**
  - Numbers are in order, and all steps are equidistant
  - The difference between 1 and 2 is the same as that between 3 and 4 or between 4 and 5.
  - The difference between 1 and 3 is twice as much as the difference between 1 and 2.
- Example: temperature

Scales of measurement, cont.

- For interval scales, even though the difference between items can be compared, the numbers themselves cannot.
- Although the difference between 10 and 20 is twice that between 10 and 15, it does not mean that 20 is twice what 10 is.
  - In temperature, 20 degrees is not twice as warm as 10 degrees.

Scales of measurement, cont.

- **Ratio**
  - Like interval scales, but there is an absolute, or true zero
  - Can thus be manipulated arithmetically
  - Can compare the values themselves.
    - 200 ms is twice as long as 100 ms (even though 20 degrees F is not twice as warm as 10 degrees F)

Fundamentals underlying statistics

- You can't test everyone.
- You cannot know how representative your sample is.
- What you really want to know is, given the difference you find in the sample, is it likely that there is really a difference in the entire population.

A population is the entire set of people.
A sample is a subset drawn from that population.

Is the difference "real"?

- The larger the sample, the better an estimate it should be of the population.
- The larger the difference in the sample, the more likely it would be present in the population.
- The more representative the sample, the more generalizable the results.

Normal distribution

- Gaussian curve.
- The largest number of people at the center
- Symmetrical around that (there are as many people below as above).
- There are fewer and fewer people as you get further out towards the extreme values.

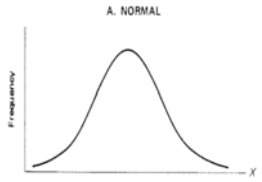
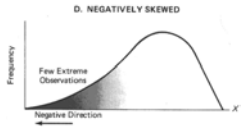


Image: R. S. Witte, *Statistics*, 4th ed.

Skewed distribution

- Not symmetrical.
- There are more extreme scores in one direction.



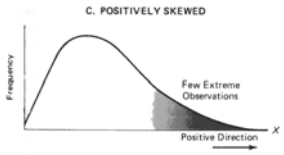


Image: R. S. Witte, *Statistics*, 4th ed.

Bimodal distributions

- Scores fall generally into two regions

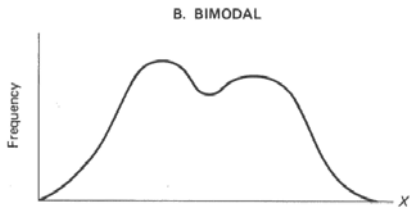
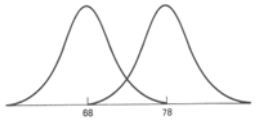


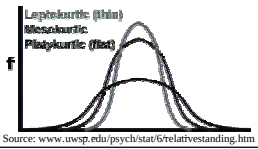
Image: R. S. Witte, *Statistics*, 4th ed.

Other distributions

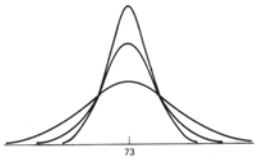
- Distributions can have different centers
- Or different variabilities around the center
- Or different kurtosis



Different Means, Same Standard Deviation



Source: www.uwsp.edu/psych/stat/6/relativestanding.htm



Same Mean, Different Standard Deviations

Image: R. S. Witte, *Statistics*, 4th ed.

Descriptive statistics

- Indicate what the group of people you measured were like as a group – what was their average, or how much did they vary.

Measures of central tendency

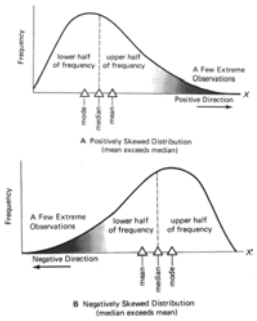
- Mode
  - the single most-frequently occurring score.
- Median
  - the middlemost score
  - the score which has as many scores below it as above it.
- Median
  - the arithmetic average of the set.

- "You know how dumb the average guy is? Well, by definition, half of them are even dumber than that."
  - J.R. "Bob" Dobbs
  - Taken from CHANCE News 13.03, Mar - April. 2004, Copyright 2004 Laurie Snell

How are these different?

- Imagine an exam.
  - 4 of the 10 people in the class aced it (100s)
  - Then one got a 92, another a 90, another an 88, another an 85, one got a 70, and one got a 52.
- The MODE would be 100.
- The MEDIAN would be 91.
- The MEAN would be 88.
- All are accurate, but they tell you different things.

Differences, cont.



- A mean will be most affected by skewness in the distribution.
- Median and mode are not affected by skewness.
- The more skewed a distribution, the more the mean and the median will differ.
- A mode can be used for nominal data, while a median requires ordinal.

Image: R. S. Witte, *Statistics*, 4th ed.

So, what is wrong with this quote?

- "In 1961, the average number of children born to an Australian woman was at its highest level since reliable records began in the 1920s, at 3.55. It was also the only occasion documented this century in which the 3.5 child average was exceeded - meaning women were statistically more likely to have four children than any other number."
  - Sydney Morning Herald
  - 8 February 2004

Taken from CHANCE News 13.03, Mar - April. 2004, Copyright 2004 Laurie Snell

Variability or dispersion

- Range
  - Difference between the top and bottom score (maximum & minimum)
  - Affected by outliers
  - Range tends to get larger as the number of scores gets larger
    - Not a "pure" measure of variability

Variability or dispersion, cont.

- Interquartile range
  - Range of the middle 50% of the population
  - Distance from the 25% mark to the 75% mark.
  - Ignores the real outliers, but gives a measure of the distance between the others.

Variance

- Refers to how tightly or loosely scores cluster around the mean.
- How much scatter there is, or how representative the mean is of what the group of people are doing.
  - Ex: 89 & 91 vs. 80 & 100
- Since average distance from the mean is sometimes positive & sometimes negative, we take squared differences from the mean

Standard deviation

- Square root of the variance
- A rough measure of the average amount by which observations deviate on either side of the mean
- For most distributions, about 2/3 of the people will be within 1 s.d. of the mean.
- Only as little as 5% will be more than 2 s.d. from the mean.
- So s.d. tells you what is typical variability vs. what is more unusual variability

Why do we care about variability?

- Research questions are always really about variability.
  - Variability over time
    - People change, learn, improve
  - Variability across situations
    - People do better with one HA than another
  - Variability across individuals
    - These people perform differently than other people
- Statistics determines whether variability in one thing is related to variability in another thing.

Standardising scores

- These are a means of comparing raw scores on different tests.
- Raw scores are changed to a standardised value, which takes into account both average performance and spread of performance
- Based on how many standard deviations a particular score is from the mean