Intro to Stats

Probability

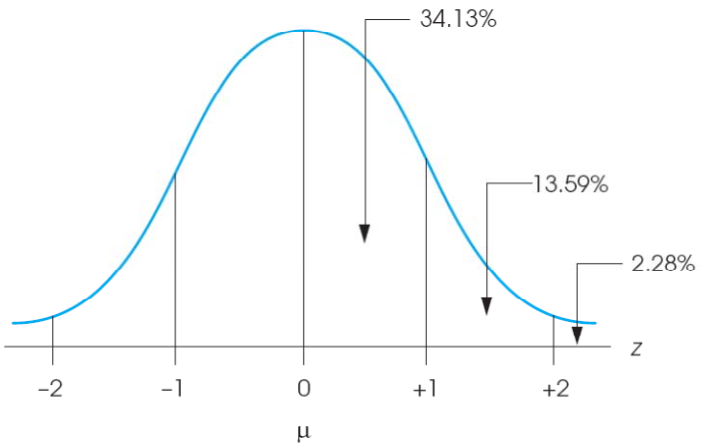
Probability theory

- Provides a basis for thinking about the probability of possible outcomes
- & can be used to determine how confident we can be in an effect
- (not due to chance)

Assumptions

- Large sets of data should approximate a normal curve
  - (it is assumed that the population is normally distributed)
- "I know of scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expressed by the normal curve… Whenever a large sample of chaotic elements are taken in hand and marshaled in the order of their magnitude, an unsuspected and most beautiful form of regularity proves to have been latent all along."
  --Sir Francis Galton, Natural Inheritance, 1889

Normal Distribution



Properties of Normal Distribution

- 1. Mean = median = mode
- 2. Symmetrical
- 3. Asymptotic tails

Standard Scores

- Can therefore compare between samples if standardized with respect to the mean and SD
- Standard scores: comparable because in standard deviation units
- The z score is a common standard score
  - The number of SDs above the mean
  - Are comparable across different distributions

Z Scores

$$z = \frac{(X - \overline{X})}{s}$$

z = z score
X = individual score
$\overline{X}$ = mean of distribution
s = SD of distribution

Parts of the *z*-score

- The *z*-score can be broken into the sign and the magnitude of the value
  - The sign (+/−) tell us the position of the score relative to the mean
    - A positive *z* score means that the value is above the mean
    - A negative *z* score means that the value is below the mean
  - The number (magnitude) itself tells us the distance the score is from the mean in terms of the number of standard deviations

Properties of z

- The shape of the distribution will be the same as the original
- The mean will always be zero
- The standard deviation will always be one (always reliable axes in z-score units)

- Z-scores can be used to compare across different data sets

Jump to Probability

- A certain % of scores fall below or above each z score in a predictable way (as seen earlier)
- This can give us the probability of attaining a particular score
  - Does the sample differ from the population (is there a difference between groups?)
  - We have to specify the risk we're willing to take of being wrong

*Use a table for exact percents

Testing Hypotheses

- If the probability of an event is extreme, it is unlikely that it is due to chance
- Psychology uses 5% ($p < .05$) as the extreme probability
  - z-score > 1.96
- If significant ($p < .05$), the research hypothesis is a more likely explanation than the null