Measurement presumes that individuals differ in terms of some characteristic – a characteristic we are interested in measuring. We may measure this characteristic because we want to know how much of it a person has or has obtained; we may want to know where a person falls in the population in terms of how much of the characteristic she or he has.

The basis for measurement is that the answers to these questions are unknown, but we assume that individuals differ. We measure primarily to understand these differences (for any purpose).

Educational Statistics

We hope to manipulate measurements in a way to summarize some set of measurements taken on our population of interest. We use indicators of central tendency, indicators of variability, and indicators of relationship to accomplish this.

Moments about the distribution: mean, variance, skewness, kurtosis. Standard normal distribution has a mean = 0, variance = 1, skewness = 0, kurtosis = 0.

Variability

Variability is an indicator to describe the degree of differences among the measures we have acquired from individuals. Our primary interest is in these differences and variability indicates the extent to which individuals differ.

Deviations $(X - \overline{X})$

Problem arises when summing deviations, sum = 0

Squared Deviations $(X - \overline{X})^2$

Avoid the sum of deviations problem; other nice properties

The metric is the original measure squared (not easy to interpret)

Sum of Squares = $SS = \sum (X - \overline{X})^2$

Indicator of total variation (deviations) in the sample (population)

Average Sum of Squares = S^2 = Variance = $\dfrac{\sum (X - \overline{X})^2}{N}$

Standard Deviation (S) is the square root of the variance (average sum of squares) the metric is the same as the metric of the original measure

Use of S implies a normal distribution

Sometimes we want to compare scores measured on different scales and a convenient indicator is the number of standard deviations each score is away from the mean.

z-score: $\frac{(X - \overline{X})}{S}$

Summary indicator of the deviation – the number of standard deviations a given score is from the mean. It is a standardized score (standardized by the S).
A z score of 0 is always the mean; a z score of +1.0 means that the score is one standard deviation above the mean. This transformation maintains the rank order of scores.

Relationships

Cross Product: $\sum[(X - \overline{X})(Y - \overline{Y})]$

Covariation: $\frac{\sum[(X - \overline{X})(Y - \overline{Y})]}{N}$

This indicates the degree to which two variables co-vary: that is, do both measures indicate deviation simultaneously – does an individual differ on both measures, in one direction or the other? Covariance is computed by taking the average of the products of deviations on two measures.

Correlation: $\frac{\sum[(X - \overline{X})(Y - \overline{Y})]}{N S_X S_Y} = r_{XY} = \frac{Cov_{XY}}{S_X S_Y}$

This is a standardized indicator of covariance. Since some variables may be measured on a scale with a wide range, the covariance will also be large. To standardize these effects, the covariance can be standardized by dividing it by the standard deviations of both measures.

This is similar to standardizing a score (z-scores). In fact, if scores are in z-score metric, you can take the average products of the z-scores to compute the correlation.

$$r = \frac{\sum z_X z_Y}{N}$$

The Normal Curve

$\boldsymbol{N}(\mu, \sigma^2)$

Mode=mean=median

+/- 1 sd occurs at the points of inflection

The Standard Normal Curve $\boldsymbol{N}(0, 1)$

We know the Area contained under the curve

Standard Scores (z scores)

Same shape as the raw score distribution
Mean of z-scores is zero
Standard deviation is one

$$z = \frac{X - \mu}{\sigma}$$