Statistics

Measurement

Measurement is defined as a set of rules for assigning numbers to represent objects, traits, attributes, or behaviors

A *variable* is something that varies (eye color), a *constant* does not (pi)

Variables can be *discrete* (finite range – sex, race) or *continuous* (infinite range – time, distance)

Scales of Measurement

Nominal Scales are qualitative system for categorizing objects or people

- Gender: Female =1, Male = 2; Eye Color: Brown =1, Blue =2, Green = 3.

Ordinal Scales allow you to rank people or objects according to the quantity of a characteristic

- Class Rank: 1 = Valedictorian, 2 = Salutatorian, 3 = 3rd Rank, etc.

Scales of Measurement

Interval Scales allow ranking on a scale with equal units

- IQs, GRE scores

Ratio Scales have the properties of interval scales with a true zero point

- Height in inches, weight in pounds

Why "Scale" matters

There is a hierarchy among the scales

Nominal scales are the least sophisticated (provide the least information) and ratio scales are the most sophisticated (provide the most information)

Interval and ratio level data allow the use of the more powerful parametric statistical procedures

Types of Statistics

Statistics is the branch of mathematics dedicated to organizing, depicting, summarizing, analyzing, and dealing with numerical data

Can be descriptive or inferential

Distributions

Frequency distributions offer a great way to visually inspect data before running inferential statistics

For example:

Table 2.2: Distribution of Scores for 20 Students

Student	Homework Scores
Cindy	7
Tommy	8
Paula	9
Steven	6
Angela	5
Robert	6
Kim	10
Charles	9
Julie	9
Shawn	9
Karen	8
Paul	4
Teresa	5
Freddie	6
Tammy	7
Shelly	8
Carol	8
Johnny	7
Kevin	8
Randy	5

Mean = 7.3
Median = 7.5
Mode = 8

Table 2.3: Ungrouped Frequency Distribution

Score	Frequency
10	1
9	4
8	5
7	4
6	3
5	2
4	1

Note: This reflects the same distribution of scores depicted in Table 2.2.

Ungrouped FDs give information on all scores in a set of data

Table 2.4: Grouped Frequency Distribution

Class Interval	Frequency
125 – 129	6
120 – 124	14
115 – 119	17
110 – 114	23
105 – 109	27
100 – 104	42
95 – 99	39
90 – 94	25
85 – 89	22
80 – 84	17
75 – 79	13
70 – 74	5

Note: This presents a grouped frequency distribution of 250 hypothetical scores that are grouped into class intervals that incorporate 5 score values.

Grouped FDs give information on all score ranges in a set of data

Graphs can be a visually interesting and meaningful way to convey information about a set of scores

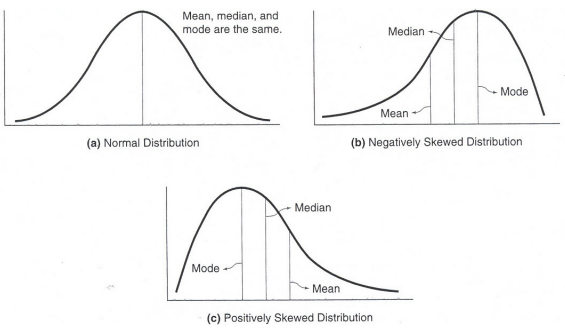
Let's see it in action!

Measures of Central Tendency

Mean
- The arithmetic average

Median
- Divides a distribution arranged in an order of magnitude in half

Mode
- Most frequently occurring value in a distribution



Measures of Variability

Range
- Distance between extreme points in a distribution

Variance
- Sum of the deviations between each value in a distribution and the mean of the distribution, or $\Sigma(X - M)^2$, divided by N

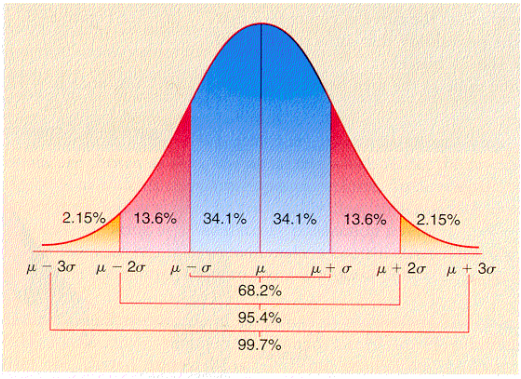
Standard Deviation
- Square root of the variance, a gauge of variability in a set of scores

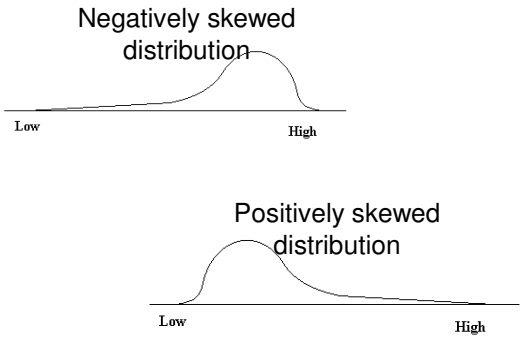
Shapes of Distributions

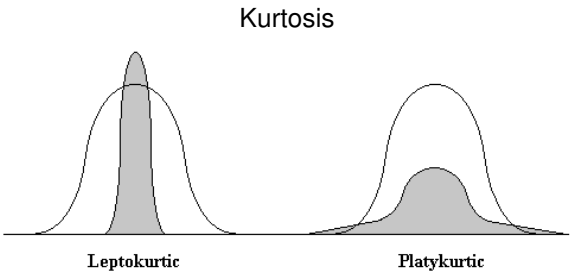
Normal Distribution (bell curve)
- Special symmetric distribution that is unimodal with mode = median = mean

Skewed Distributions

Kurtosis
- Leptokurtic (less dispersion)
- Platykurtic (greater dispersion)







Correlation Coefficients

A correlation coefficient is a mathematical measure of the relationship between two variables

The correlation coefficient was developed by Karl Pearson and is designated by the letter *r*

Remember that variables tend to *regress to the mean*

Correlation (*r*)

Correlations range from -1.0 to +1.0

Correlations differ on two parameters:

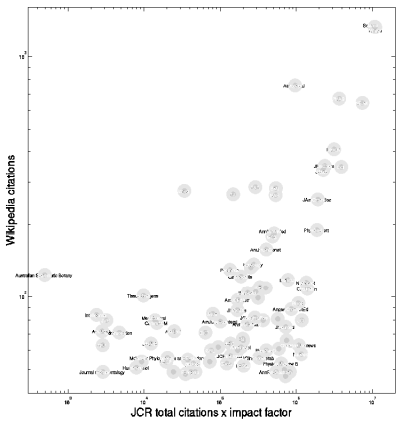
Sign - can be positive or negative. Indicates the pattern of the relationship

Size - a correlation of 0.0 indicates the absence of a relationship; the closer the correlation gets to 1.0, the stronger the relationship; a 1.0 indicates a perfect relationship

Scatterplots

Graph depicting the relationship between two variables (X & Y)

Each mark in the scatterplot actually represents two scores, an individual's scores on the X and the Y variable



Types of Correlations

Pearson Product-Moment Correlation
- Both variables continuous and on an Interval or Ratio scale

Spearman Rank-Difference Correlation
- Both variables on an Ordinal scale

Point-Biserial Correlation
- One variable continuous and on Interval/Ratio scale, the other a genuine dichotomy (e.g., true/false)

Biserial Correlation
- Both variables continuous and on Interval/Ratio scale, but one is reduced to two categories (i.e., dichotomized)

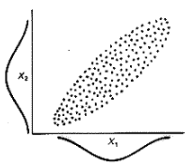
Effecting Factors

Most correlations assume a linear relationship; if another type of relationship exists, traditional correlations may underestimate the correlation

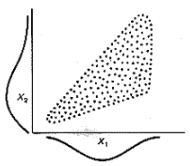
If there is a restriction of range in either variable, the magnitude of the correlation will be reduced

Deviations from Linearity

Homoscedasticity indicates normally distributed variables, heteroscedasticity indicates skewness on one or both



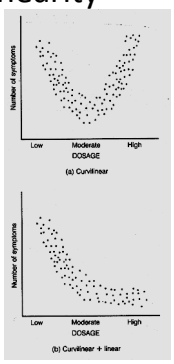
Homoscedasticity with both variables normally distributed



Heteroscedasticity with skewness on one variable

Deviations from Linearity

The relationship between variables can also not be the same throughout their distribution



Interpretation of Correlations

General Guidelines:

< 0.30 Weak

0.30 - 0.70 Moderate

> 0.70 Strong

These, however, are not universally accepted and you might see other guidelines

Statistical Significance

Statistical significance is determined both by the size of the correlation coefficient and the size of the sample

Usually expressed as a *p* value, which tells the probability that the found results are due to chance

Quantitative Interpretation

Coefficient of Determination (r^2)

- The proportion of variance on one variable that is determined or predictable from the other variable

Coefficient of Nondetermination ($1-r^2$)

- The proportion of variance in one variable that is not determined or predictable from the other variable

Correlation & Prediction

When variables are strongly correlated, knowledge about performance on one variable provides information that can help predict performance on the other variable

Linear regression is a statistical technique for predicting scores on one variable (criterion or Y) given a score on another (predictor or X)

- Predicts criterion scores based on a perfect linear relationship

Essential Facts

Degree of the relationship is indicated by the *r* number, while the direction is indicated by the sign

Correlation does not equal causality

High correlations allow for predictive ability

Test Score Interpretation

Scores and their Interpretation

Raw scores are of limited utility

Norm-referenced are based on a comparison between a test taker's performance and that of other people

Criterion-referenced are when the test taker's performance is compared to a specified level or standard of performance (i.e., criterion)

Score Interpretation

Norm-referenced
- Are relative to the performance of other test takers
- Can be applied to both maximum performance tests and typical response tests

Criterion-referenced
- Are compared to an absolute standard
- Typically only applied to maximum performance tests

Score Interpretations

While people often refer to norm-referenced and criterion-referenced *tests*, this is not technically accurate

The terms norm-referenced and criterion-referenced actually refer to the *interpretation of scores or test performance*, not the test itself

Norm-Referenced Interpretations

The *most important factor* when making norm-referenced interpretations involves the relevance of the group of individuals to which the examinee's performance is compared

Ask yourself, "Are these norms appropriate for this individual?"

- Is the standardization sample representative?
- Is the sample current?
- Is the sample of adequate size?

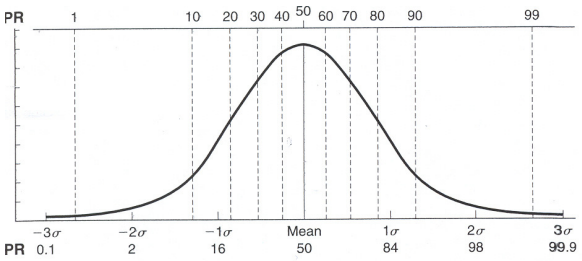
The Normal Distribution

The normal distribution is also referred to as the Gaussian or bell-shaped curve

Characterizes many variables that occur in nature

It is unimodal and symmetrical

Predictable proportions of scores occur at specific points in the distribution



Normal Distribution

The mean equals the median, so the mean score exceeds 50% of scores

Approximately 34% of the scores fall between the mean and 1 SD above the mean, so a score one SD above the mean exceeds about 84% of the scores (i.e., 50% + 34%)

Approximately 14% of the scores fall between the first and second standard distributions, so a score two SDs above the mean exceeds about 98% of the scores (i.e., 50% + 34% + 14%)

Standard Scores

Are linear transformations of raw scores to a scale with a predetermined mean and standard deviation

Use standard deviation units to indicate where a subject's score is located relative to the mean of the distribution

Retain a direct relationship with the raw scores and the distribution retains its original shape

Reflect interval level measurement

Standard Scores Examples

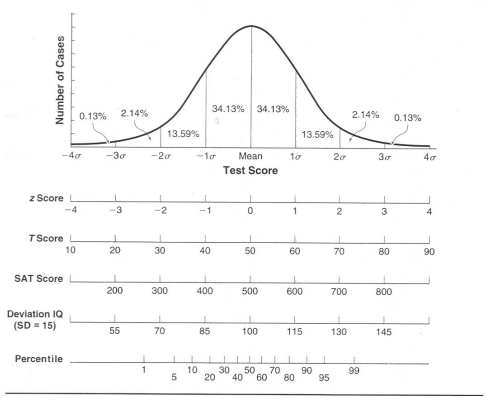
z-scores: mean of 0, SD of 1

T-scores: mean of 50, SD of 10

IQ Scores: mean of 100, SD of 15

CEEB Scores (SAT/GRE): mean of 500, SD of 100

Many writers use the term “standard score” generically



Z Scores

$$z = \frac{X - \overline{X}}{SD_x}$$

X = Raw score

X = Reference group mean

SD_x = Standard deviation of reference group

To transform z scores into other Standard Scores:

New standard score = (z score) (New SS) + (new Mean)

Relationship of Different Standard Score Formats

z-scores *X* = 0 *SD* = 1	T-scores *X* = 50 *SD* = 10	Wechsler IQ *X* = 100 *SD* = 15	CEEB scores *X* = 500 *SD* = 100	Percentile rank
2.6	76	139	760	>99
2.4	74	136	740	99
2.2	72	133	720	99
2.0	70	130	700	98
1.8	68	127	680	96
1.6	66	124	660	95
1.4	64	121	640	92
1.2	62	118	620	88
1.0	60	115	600	84
0.8	58	112	580	79
0.6	56	109	560	73
0.4	54	106	540	66
0.2	52	103	520	58
0.0	50	100	500	50
-0.2	48	97	480	42
-0.4	46	94	460	34
-0.6	44	91	440	27
-0.8	42	88	420	21
-1.0	40	85	400	16
-1.2	38	82	380	12
-1.4	36	79	360	8
-1.6	34	76	340	5
-1.8	32	73	320	4
-2.0	30	70	300	2
-2.2	28	67	280	1
-2.4	26	64	260	1
-2.6	24	61	240	1

Note: Adopted from Reynolds (1998). *X* = Mean, *SD* = Standard Deviation

Normalized Standard Scores

Are standard scores based on underlying distributions that were not originally normal, but were transformed into normal distributions

Often involve nonlinear transformations and may not retain a direct relationship with the original raw scores

Are typically interpreted in a manner similar to other standard scores

Percentile Rank

One of the most popular and easily understood ways to interpret and report test performance

Interpreted as reflecting the percentage of individuals scoring below a given point in a distribution

A percentile rank of 80 indicates that 80% of the individuals in the reference group scored below this score

Percentile Rank

While easy to interpret, percentile ranks do not represent interval level measurement

They are compressed near the middle of the distribution where there are large numbers of scores, and spread out near the tails where there are relatively few scores

When interpreting be sure they are not confused with “percent correct”

Quartile Scores

Based on percentile ranks

The lower 25% receive quartile score of 1
26% - 50% a quartile score of 2
51% - 75% a quartile score of 3
The upper 25% a quartile of 4

Decile Scores

Divides the distribution of percentile ranks into ten equal parts

The lowest decile score is 1 and corresponds to scores with a percentile ranks between 0 and 10%

The highest decile score is 10 and corresponds to scores with percentile ranks between 90 and 100%

Grade Equivalents

Norm-referenced score interpretation that identifies the academic "grade level" achieved by the student

Popular in school settings and *appear* easy to interpret, but they need to be interpreted with *extreme caution*

Limitations of GE

Based on assumptions that are not accurate in many situations (e.g., academic skills achieved at a constant rate with no gain or loss during the summer vacation)

There is not a predictable relationship between grade equivalents and percentile ranks

Limitations of GE

A common misperception is that students should receive instruction at the level suggested by their grade equivalents

Experts have numerous concerns about the use of grade equivalents and it is best to avoid their use in most situations

Age Equivalents share many of the limitations of GE and as a general rule they should also be avoided

Criterion-Referenced Scores

The examinee's performance is not compared to that of other people, but to a specified level of performance

Emphasize what the examinee knows or what they can do, not their standing relative to other test takers

The most important consideration is how clearly the knowledge or skill domain is specified or defined

Criterion-Referenced Types

Percent Correct: the student correctly answered 85% of the questions

Mastery Testing: a "cut score" is established and all scores equal to or above this score are reported as "pass"

Standards Based Interpretations: Not Proficient, Partially Proficient, Proficient, & Advanced Performance; Letter Grades = A, B, C, D, & F

Characteristics of Norm-Referenced and Criterion-Referenced Scores

Norm-Referenced Interpretations	Criterion-Referenced Interpretations
Compare performance to a specific reference group – a relative interpretation.	Compare performance to a specific level of performance – an absolute interpretation.
Useful interpretations require a relevant reference group.	Useful interpretations require a carefully defined knowledge or skills domain.
Usually assess a fairly broad range of knowledge or skills.	Usually assess a limited or narrow domain of knowledge or skills.
Typically only has a limited number of items to measure each objective or skill	Typically will have several items to measure each test objective or skill
Items are selected that are of medium difficulty and maximize variance, very difficult and very easy items are usually deleted	Items are selected that provide good coverage of content domain; the difficulty of the items matches the difficulty of content domain
Example: Percentile rank – a percentile rank of 80 indicates that the examinee scored better than 80% of the subjects in the reference group.	Example: Percentage correct – a percentage correct score of 80 indicates that the examinee successfully answered 80% of the test items.

Which One?

It is possible for a test to produce both norm-referenced and criterion-referenced interpretations (e.g., WIAT-II)

While the development of a test producing both norm-referenced and criterion-referenced interpretations may require some compromises, the increased interpretative versatility may justify the compromises

Inter-Test Comparisons

Test scores cannot be meaningfully compared if

- The tests/test versions are different
- The reference groups are different
- The score scales differ

unless the tests/groups/scales have been purposefully equated

Still, the context and background of test takers must be taken into account when comparing