Stats Refresher

Measures of central tendency
- Mean
- Median
- Mode

Educational Statistics Translator
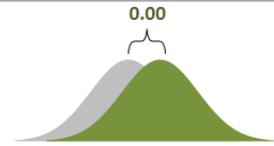
1. Click on the **shaded box** on the right, then click the **down arrow** and **select a measure** from the dropdown menu.
2. Add a **value** to the box to the right of the shaded box.

Effect Size

Effect Size

Effect Size (ES) measures of the magnitude of impact an input (such as a change in instruction) on a population.

Effect Size = **0.00**

0.00

Impact

The improvement index is the change in an average student's **percentile rank** you can expect if the input is given to the student.

This means that 50% of students report increased scores OR a student at the average percentile (50%) could expect to move to the 50th percentile after the input.

Improvement Index = **0%**

☐ = Average percentile
■ = Increase as a result of the input

0   50   100

Standard Deviation

Based on the normal distribution, the change in standard deviation would be 0.0.

3 2 1 0 1 2 3
■ = Change in standard deviation due to input

Correlation

A correlation is a measure of the relationship between the input and the change in the measured outcome. It is represented by the variable r.
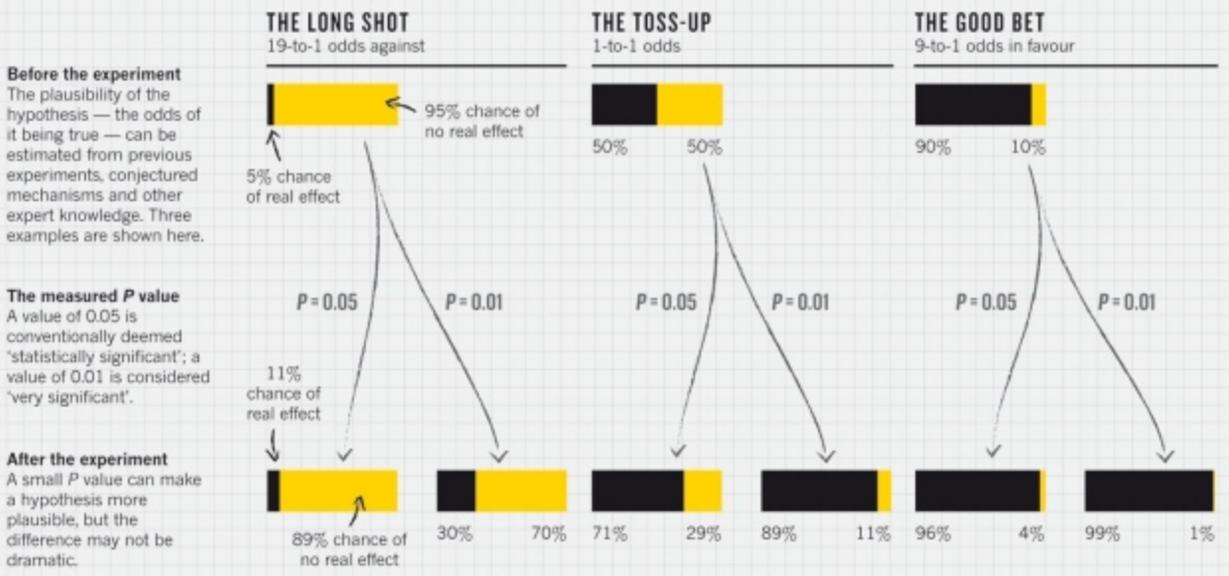
$r = 0$

0   Strength of Correlation   1

Its square (r^2) tells you how much of the variance in the outcome is explained by the program.
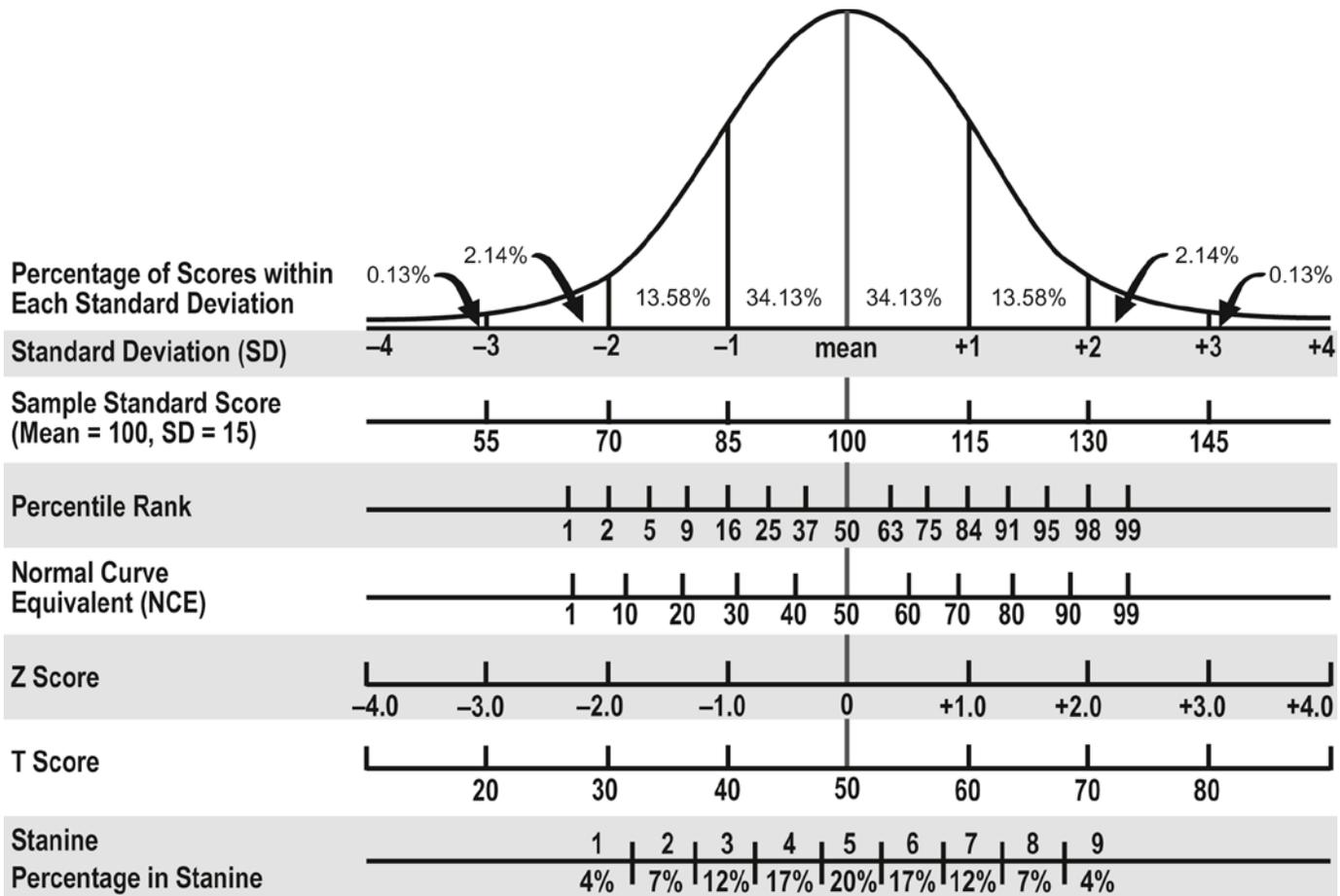
$r^2 = 0\%$

PROBABLE CAUSE

A P value measures whether an observed result can be attributed to chance. But it cannot answer a researcher's real question: what are the odds that a hypothesis is correct? Those odds depend on how strong the result was and, most importantly, on how plausibile the hypothesis is in the first place.

■ Chance of real effect
☐ Chance of no real effect

	THE LONG SHOT 19-to-1 odds against	THE TOSS-UP 1-to-1 odds	THE GOOD BET 9-to-1 odds in favour

Before the experiment
The plausibility of the hypothesis — the odds of it being true — can be estimated from previous experiments, conjectured mechanisms and other expert knowledge. Three examples are shown here.

95% chance of no real effect

5% chance of real effect

50%   50%

90%   10%

The measured P value
A value of 0.05 is conventionally deemed "statistically significant"; a value of 0.01 is considered "very significant".

P = 0.05   P = 0.01   P = 0.05   P = 0.01   P = 0.05   P = 0.01

11% chance of real effect

After the experiment
A small P value can make a hypothesis more plausible, but the difference may not be dramatic.

89% chance of no real effect

30% 70%   71% 29%   89% 11%   96% 4%   99% 1%

Source: http://www.nature.com/news/scientific-method-statistical-errors-1.14700

The normal curve, percentiles, and selected standard scores

Percentage of Scores within Each Standard Deviation	0.13%	2.14%	13.58%	34.13%	34.13%	13.58%	2.14%	0.13%	
Standard Deviation (SD)	−4	−3	−2	−1	mean	+1	+2	+3	+4
Sample Standard Score (Mean = 100, SD = 15)		55	70	85	100	115	130	145	
Percentile Rank			1 2 5 9	16 25 37	50	63 75 84	91 95 98 99		
Normal Curve Equivalent (NCE)			1 10	20 30 40	50	60 70	80 90 99		
Z Score	−4.0	−3.0	−2.0	−1.0	0	+1.0	+2.0	+3.0	+4.0
T Score		20	30	40	50	60	70	80	
Stanine			1 2 3	4 5	6	7 8	9		
Percentage in Stanine			4% 7% 12%	17% 20%	17%	12% 7%	4%		

Normal Distribution

For many kinds of tests, developers often make the assumption that individual test scores have an approximately bell-shaped, or normal, distribution. The figure above illustrates, for a normal distribution, the relationship between standard scores and other derived scores: percentile ranks, stanines, and normal curve equivalents.

Standard Scores

Standard scores are called deviation scores because they indicate the distance between an individual's score and the mean of the norm group in terms of standard deviation units.

Many psychological and educational instruments use a standard score scale with a mean of 100 and a standard deviation of 15. This common scale allows both basic score comparisons and more complex statistical analyses to be made.

Percentile Ranks

A percentile rank indicates the percentage of the norm group that scored below a specific raw score. Percentile ranks are easily explained and widely used. However, unlike standard scores, percentile ranks are not on an equal-interval scale. Percentile ranks are clustered close together at the center of the normal distribution, and are spread more widely at the extremes. Because of this, there is a danger in attempting to interpret differences between percentile ranks. For example, the 10-point difference between the percentile ranks of 45 and 55 represents a smaller difference in the trait being measured than the 10-point difference between the percentile ranks of 85 and 95.

Stanines

The stanine ("*standard nine*") scale ranges from 1 to 9, with a mean of 5 and a standard deviation of about 2. Because it contains only 9 possible score points, the stanine scale does not provide a precise description of score levels. Stanines are commonly used in situations where fine discriminations among scores would not be appropriate.

Normal Curve Equivalents

The normal curve equivalent (NCE) scale divides the normal score distribution into 99 equal intervals. NCE scores range from 1 to 99, with a mean of 50 and a standard deviation of 21.06.

AGS®

Terms

Action Research: A type of research in which educators examine their own practice and evaluate strategies to improve practice and education outcomes.

Bivariate Correlation: A statistical correlation between two variables.

Central Tendency: A score in a set of scores or a frequency distribution that is typical or representative of all the scores. Measures of central tendency are the mean, median and mode.

Coefficient of determination: For bivariate correlations, the coefficient of determination is defined as r^2, which is interpreted as the proportion of variation in the scores that is explained by the relationship between the variables. Note: Correlations indicate statistical, not causal, relationships.

Confidence interval: A range of values that indicates the confidence or probability of observing a particular score or value in a population, usually expressed as standard deviation units above and below the mean.

Correlation coefficient: A number indicating strength & direction of the statistical association between 2 or more variables.

Descriptive statistics: Statistics used to describe, organize and summarize data.

Effect size: The degree to which a practice, program or policy has an effect based on research results, measured in standard deviation units. (Effect size is also referred to as practical significance.) A statistic commonly used to measure effect size is Cohen's d, which social scientists interpret as the following: d = .2, small; d = .5 to .8, medium; and d = .8 and higher, large.

Frequency distribution: The frequency of occurrence of scores in a set. Frequency distributions can be represented in graphs or tables.

Inferential statistics: Statistics used to make inferences about a population based on the scores obtained from a sample. Inferential statistics are based on the mathematics of probability theory, ex., t, F and Chi Square.

Intervention: A procedure, technique or strategy that is designed to modify an ongoing process. In research studies, the intervention also is referred to as a treatment. Most interventions in education are designed to modify directly or indirectly the student-learning process.

Mean: In general, the average score in a set of scores or frequency distribution, calculated as the sum of the scores divided by the number of scores.

Terms

Median: The middle score in a set of scores or frequency distribution such that 50% of the scores are at or below the median score.

Mode: The most frequent score in a set of scores or a frequency distribution.

N (n): The number of scores in a population (N) or a sample (n) of scores.

Normal curve: The bell-shaped curve that results from the graph of a normal frequency distribution.

Normal distribution: A symmetrical frequency distribution in which the scores form a bell-shaped curve, and the mean, median and mode have the same value.

Percent: The proportion of participants who obtain a particular score in a frequency distribution.

Percentile: The percent of participants who score at or below a particular score in a frequency distribution (also referred to as percentile rank).

Population: All individuals or entities belonging to the group that is being studied.

Practical significance: The degree to which a practice, program or policy has enough of an effect to justify its adoption. Practical significance usually is measured with statistics that calculate effect sizes.

Range: The difference between the highest and lowest score in a set of scores or frequency distribution.

Raw score: An original score on a test or other measuring instrument prior to any score transformations.

Sample: A subset of individuals or entities from a population.

Standard deviation: A measure of the variability of the scores in a set of scores or a frequency distribution, equivalent to the average distance of the scores from the mean.

Standard score: A score that transforms an original or raw score into standard deviation units in order to locate the score's position within a frequency distribution. Standard scores also are known as z-scores and are calculated as: $z = $ Raw Score − Mean /Standard Deviation. Sign of the score (plus or minus) indicates whether it is above or below the mean.

Statistically significant: A result that has a low probability (e.g., 5 %) of occurring by chance. Because it is unlikely that a statistically significant result has occurred by chance, the result is said to reflect non-chance factors in the study, such as the effects of a treatment.

Statistics: Methods and rules for organizing and interpreting quantitative observations.