Symbols and Equations

name	*symbol*	***equation***	***use***
i[th] person's *score* on variable X	x_i	x_i	raw data
sample *mean*	$\bar{x}$	$\frac{\sum x_i}{n}$	Most common measure of central tendency for interval or ratio data.
population *mean*	μ	estimated by: x_i	The most often used population *parameter* for central tendency.
i[th] person's deviation score	d_i	$x_i - \bar{x}$	Difference between person *i*'s score and the mean. Used in calculation of standard deviation, *z*-scores, covariance, correlation, and analysis of variance.
i[th] person's *z-score*	z_i	$\frac{x_i - \bar{x}}{s}$ or $\frac{d_i}{s}$	Standard score. Tells where the person is located in the distribution relative to the rest of the sample. Can be compared to *z*-scores for other variables.
sample *standard deviation*	s or sd	$\sqrt{\frac{\sum d_i^2}{n-1}}$	Most common measure of dispersion for interval or ratio data; used in the calculation of *z*-scores and other statistics and parameters
population *standard deviation*	σ	estimated by: $\sqrt{\frac{\sum d_i^2}{n-1}}$	Most used population parameter for dispersion. Note "n-1" in denominator – this gives an unbiased estimate
sample *variance*	s^2	$\frac{\sum d_i^2}{n-1}$	Fundamental measure of dispersion for an interval or ratio scaled variable.
population *variance*	σ^2	estimated by: $\frac{\sum d_i^2}{n-1}$	Fundamental population measure of variability of an interval or ratio scaled variable. Note "n-1" in denominator – this gives an unbiased estimate
standard error of the mean	$\sigma_{\bar{x}}$	estimated by: $\frac{s}{\sqrt{n}}$	Standard deviation of the sampling distribution of means. Used to tell how large an error you must accept, at a specified level of confidence, when using your sample's mean to estimate the population's mean (confidence estimates/intervals).

standard error of the difference between two means	$\sigma_{\bar{x}_1-\bar{x}_2}$	estimated by: $\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$	Standard deviation of the sampling distribution of differences between means. **Used in *z*-test** to determine whether the difference between two sample means could be due to sampling variability.
z-test for a single mean	z	$\frac{\bar{x}-\mu}{\sigma_{\bar{x}}}$	Critical ratio used to test the significance of the difference between a sample mean and a hypothetical population mean. Used to determine whether a sample could have come from a population with a particular mean. Used for samples with more than 30 members.
z-test for significance of difference between sample means	*z*	$\frac{\bar{x}_1-\bar{x}_2}{\sigma_{\bar{x}_1-\bar{x}_2}}$	Critical ratio used to test the significance of the difference between two sample means. Used for samples with more than 30 members.
corrected standard error of the difference between means (takes sample size into account;)	$SE_{\bar{x}_1-\bar{x}_2}$	$\sqrt{\left(\frac{\sum d_1^2+\sum d_2^2}{n_1+n_2-2}\right)\left(\frac{n_1+n_2}{n_1\, n_2}\right)}$	Standard deviation of the sampling distribution of differences between means corrected for sample size. ***Used in t-test*** to determine whether the difference between two sample means could be due to sampling variability.
t-test for significance of difference between sample means	*t*	$\frac{\bar{x}_1-\bar{x}_2}{SE_{\bar{x}_1-\bar{x}_2}}$	Critical ratio used to test the significance of the difference between two sample means. Used for samples with less than 30 members.
t-test for a single mean	*t*	$\frac{\bar{x}_1-\mu}{\sigma_{\bar{x}_1-\bar{x}_2}}$	Critical ratio used to test the significance of the difference between a sample mean and a hypothetical population mean. Used for samples with less than 30 members.
covariance for a pair of continuous variables	cov_{xy}	$\frac{\sum d_{x_i} d_{y_i}}{n}$	Relatively crude measure, based on deviation scores, of the strength of the relationship between a pair of continuous variables.
Pearson product-moment correlation (Pearson's r)	r_{xy}	$\frac{\sum z_{x_i} z_{y_i}}{n}$	Extremely useful measure, based on standard scores, of the strength of the relationship between a pair of continuous variables.

Spearman rank correlation (Spearman's rho)	r_s	$1-\frac{6\sum d_i^2}{n^3-n}$	Useful measure of the strength of the relationship between a pair of ordinal variables. Based on squared differences in ranks of values in each pair (d_i^2). May also be used for continuous data.
Fisher's r to Z	Z	$1/2\left[\ln(1+r)-\ln(1-r)\right]$	Used to transform Pearson's *r* or Spearman's rho into Z which has a normally-distributed sampling distribution so it can be tested for significance.
standard error of the difference between two Zs *(for Z from Fisher's r to Z)*	$\sigma_{Z_1-Z_2}$	estimated by: $\sqrt{\frac{1}{n_1-3}+\frac{1}{n_2-3}}$	Standard deviation of the sampling distribution of differences between two *Z*s from Fisher's *r* to *Z*. Used to determine whether the difference between a pair of sample correlations could be due to sampling variability. Used for Pearson's *r* and Spearman's rho.
z-test for significance of difference between correlations	z	$\frac{Z_1-Z_2}{\sigma_{Z_1-Z_2}}$	Critical ratio used to test the significance of the difference between a pair of Pearson's *r* or Spearman's rho correlations.
standard error of Z (for Z from Fisher's r to Z)	σ_Z	estimated by: $\frac{1}{\sqrt{n-3}}$	Standard deviation of the sampling distribution of Z from Fisher's *r* to *Z*. Can be used for estimating confidence intervals of Pearson's *r* or Spearman's rho.
t- test for significance of correlation	t	$r\sqrt{(n-2)/(1-r^2)}$	Critical ratio used to test the significance of Pearson's *r* or Spearman's rho.
between-groups Sum of Squares	SS_b	$n_1\left(\bar{x}_1-\bar{X}_T\right)^2+n_2\left(\bar{x}_2-\bar{X}_T\right)^2$	Measure of variability between groups. Used as the numerator in the *F*-ratio for ANOVA.
within-groups Sum of Squares	SS_w	$\sum\left(x_{i,1}-\bar{x}_1\right)^2+\sum\left(x_{i,2}-\bar{x}_2\right)^2+\ldots$	Measure of variability within groups. Used as the denominator in the *F*-ratio for ANOVA.
F-ratio	F	$\frac{MS_b}{MS_w}=\frac{SS_b/df_b}{SS_w/df_w}$	Critical ratio used to test the significance of the difference between sample means in a one-way Analysis of variance (ANOVA).

There are two major classes of standard deviations: those that apply to samples and those that apply to populations. You directly *calculate* the ones that apply to *samples* and you *estimate* the ones that apply to *populations*.

Statistical tests: their variables and their uses

	variables	level of scaling	sample size	purpose	example
χ^2 **chi-squared**	one or more categorical variables	nominal	no more than 20% of cells have expected frequency less than 5	are row and column variables independent?	compare men and women in terms of passing or failing a course. IV: male/female; DV: pass/fail
t-test	1 discrete IV; 1 continuous DV	IV -- nominal DV -- int/ratio	less than 30 cases	could the difference between two sample means be due to sampling variability?	compare ages of samples of men and women. IV: male/female; DV: age
z-test of difference between means	1 discrete IV; 1 continuous DV	IV -- nominal DV -- int/ratio	30 or more cases	could the difference between two sample means be due to sampling variability?	compare ages of samples of men and women IV: male/female; DV: age
z-test of a single mean	1 continuous variable	sample mean; population mean	30 or more cases	could the difference between $\bar{x}$ and μ be due to sampling variability?	could this sample have come from a population with $\mu = 95.3$?
ANOVA	discrete IV(s); 1 continuous DV	IV(s) -- nominal DV -- int/ratio	30 or more cases	could the difference between sample means be due to sampling variability?	compare ages of samples of students majoring in CMNS, PSYC, MATH, ECON, HIST. IV: major; DV: age
regression	continuous IV(s); 1 continuous DV	all interval or ratio	30 or more cases	describe the linear contributions of a set of independent variables on one dependent variable.	examine factors (age, GPA, undergrad major, family income) that contribute to success in graduate school. IVs: age, CGPA, major, income; DV: graduate performance
***t*-test for significance of correlation**	2 continuous or ordinal variables	both at least ordinal	20 or more cases	could sample correlation be due to sampling variability?	relation between age and amount of time spent sleeping at night. V1: age; V2: time sleeping
***z*-test for correlation differences**	2 pairs of continuous or ordinal variables	all at least ordinal	30 or more cases	could difference between sample correlations be due to sampling variability?	compare *r* (age - GPA) to *r* (age - coffee consumption). V1: age; V2: GPA; V3: coffee consumption
z-test for proportion differences	1 discrete IV; 1"continuous" DV	IV -- nominal DV -- ratio only	30 or more cases	could the difference between a pair of proportions be due to sampling variability?	compare proportion of men and women who study math. IV: male/female; DV: % who study math