Table of Contents

1. Note on statistical symbols
2. Descriptive statistics
3. Standard error and t-tests
4. Post hoc tests
5. ANOVA
6. Regression

1. A note on statistical symbols

The statistical formulas in this document are based on a standard symbol system used in many textbooks and professional publications (see Research Design and Statistical Analysis, 3rd ed., by Myers, Well, and Lorch), rather than the idiosyncratic symbol system used in Statistics for the Behavioral Sciences, 10th, by Gravetter and Wallnau.

These differences might appear superficial to students at first, but there are many advantages to learning a standard symbol system. Learning the standard symbol system will:

1. Better prepare you for future research methods and statistics courses
2. Make it easier to understand online statistics tutorials and videos
3. Make it easier to work with data in a statistical software program, like R
4. Make talking about data easier, for example, using a variation of one symbol to label raw scores and the mean makes it easier to keep track of multiple variables; e.g., $Y\ amd\ \bar{Y}$ (pronounced Y-bar) rather than $X\ and\ M$

$$\bar{Y} = \frac{\sum Y_i}{n} \qquad \text{Arithmetic Mean} \qquad M = \frac{\sum X}{n}$$

The formula on the left, which we will use for the arithmetic mean reads "the mean of Y is equal to the sum of all individual Y scores (Y_i) divided by the sample size (n)." As you progress through the formulas below, additional identifying information is added to some symbols. For example, sums of squares (SS_j) for the t-test is written as "the sum of all individual Y scores in each group (Y_{ij}) minus the group mean ($\bar{Y}_{.j}$) squared."

$$SS_j = \sum(Y_{ij} - \bar{Y}_{.j})^2$$

Finally, it is important to point out that the formulas below use a simplified version of a more formal symbol system so it is easier to focus on the most important elements of the statistic. For example, the arithmetic mean above is a simplified version of this formula:

$$\bar{Y} = \sum_{i=1}^{n} Y_i\,/n$$

2. Descriptive statistics

Summation of Raw Scores	$\sum Y_i$
Sample Mean	$\bar{Y} = \frac{\sum Y_i}{n}$
Squared Deviation Score	$y^2 = (Y_i - \bar{Y})^2$
Sum of Squares	$\sum y^2 = SS = \sum(Y_i - \bar{Y})^2$
Variance	$s^2 = \frac{SS}{n-1}$
Standard Deviation	$s = \sqrt{s^2}$
Estimated Standard Error of the Mean	$s_{\bar{Y}} = \sqrt{\frac{s^2}{n}} = \frac{s}{\sqrt{n}}$

3. Standard error and *t*-tests

Sum of Squares	$SS_j = \sum(Y_{ij} - \bar{Y}_{.j})^2$
Pooled Variance	$s_p^2 = \dfrac{SS_1 + SS_2}{(n_1 - 1) + (n_2 - 1)}$
Estimated Standard Error of the Mean Difference	$s_{\bar{Y}1-\bar{Y}2} = \sqrt{\dfrac{s_p^2}{n_1} + \dfrac{s_p^2}{n_2}}$
Independent-Samples *t*-test	$t = \dfrac{\bar{Y}_1 - \bar{Y}_2}{s_{\bar{Y}1-\bar{Y}2}}$
Independent-Samples *t*-test critical value	$t_{critical} = t_{(n_1-1)+(n_2-1)}$
Paired-samples *t*-test	$t = \dfrac{\bar{Y}_D}{s_{\bar{Y}_D}}$
Paired-samples *t*-test critical value	$t_{critical} = t_{(n_D-1)}$
Cohen's d	$d = \dfrac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{s_p^2}} \quad or \quad d = \dfrac{\bar{Y}_1 - \bar{Y}_2}{s}$

4. Post-hoc tests

Tukey's HSD (all pairwise comparisons)	$HSD = q\sqrt{\dfrac{MS_{Error}}{n}}$
Bonferroni correction (planned comparisons)	$p = \dfrac{\alpha}{K}$

5. Analysis of Variance (ANOVA) – One Factor, Independent Samples

A. **ANOVA Structural Model**: An expression that describes each raw score in a data set:

Individual Score (Y_i) = Grand Mean + Treatment Effect + Error

or

$$Y_{ij} = \mu + \alpha_j + \varepsilon_{ij}$$

B. **Sample Estimates**: To test sample data, we find an estimate of each population parameter in the model:

Parameter	Estimate
μ	$\bar{Y}..$ = Grand Mean or "The mean of all scores across all conditions."
μ_j	$\bar{Y}._j$ = Condition Mean or "The mean of all scores in one condition."
α_j	$\bar{Y}._j - \bar{Y}..$ = Effect of one condition or "A condition mean minus the grand mean."
ε_{ij}	$Y_{ij} - \bar{Y}._j$ = Error of one score in a condition or "A raw score minus its condition mean."

C. **Partitioned Variability**: The total variability in the outcome (Y) scores $SS_{Total} = \sum(Y_{ij} - \bar{Y}..)^2$ can partitioned into two elements: $SS_A = \sum(\bar{Y}._j - \bar{Y}..)^2$ and $SS_E = \sum(Y_{ij} - \bar{Y}._j)^2$

$$SS_{Total} = SS_A + SS_E$$

or

$$\sum(Y_{ij} - \bar{Y}..)^2 = \sum(\bar{Y}._j - \bar{Y}..)^2 + \sum(Y_{ij} - \bar{Y}._j)^2$$

D. **Source Table and Formulas**

Source	SS	df	MS	F
Factor A	$SS_A = \sum(\bar{Y}._j - \bar{Y}..)^2$	$df_A = a - 1$	$MS_A = \frac{SS_A}{df_A}$	$F = \frac{MS_A}{MS_E}$
Error	$SS_E = \sum(Y_{ij} - \bar{Y}._j)^2$	$df_E = N - a$	$MS_E = \frac{SS_E}{df_E}$	
Total	$SS_T = \sum(Y_{ij} - \bar{Y}..)^2$	$df_T = N - 1$		

6. Linear Regression – One predictor variable

A. **Linear Regression Model**: An expression that describes each raw score in a data set:

Individual Score (Y_i) = Intercept + (Slope * X_i) + Error

or

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

B. **Predicted Regression Model**: The predicted model (regression equation) can be found with sample estimates of the population intercept and slope.

Parameter	Estimate	Interpretation
Y_i	$\hat{Y} = b_0 + b_1 X_i$	The predicted model, or best fit model, gives the mean (predicted) Y score at any X value
β_0	$b_0 = \bar{Y} - b_1 \bar{X}$	The **Y Intercept** is the predicted value of Y at X = 0
β_1	$b_1 = r \frac{s_Y}{s_X}$	The **Regression Coefficient** is the slope of the relationship, that is, the change in Y with each one unit change in X

C. **Partitioned Variability**: The total variability in the outcome (Y) scores $SS_Y = \sum(Y_i - \bar{Y})^2$ can partitioned into two elements: $SS_{Regression} = \sum(\hat{Y} - \bar{Y})^2$ and $SS_{Residual} = \sum(Y_i - \hat{Y})^2$

$$SS_Y = SS_{Regression} + SS_{Residual}$$

or

$$\sum(Y_i - \bar{Y})^2 = \sum(\hat{Y} - \bar{Y})^2 + \sum(Y_i - \hat{Y})^2$$

D. **Source Table and Formulas**

Source	***SS***	***df***	***MS***	***F***
Regression	$SS_{Reg} = \sum(\hat{Y} - \bar{Y})^2$	$df_{reg} = 1$	$MS_{Reg} = \frac{SS_{Reg}}{df_{Reg}}$	$F = \frac{MS_A}{MS_E}$
Residual	$SS_{Res} = \sum(Y_i - \hat{Y})^2$	$df_{Res} = N - 2$	$MS_{Res} = \frac{SS_{Res}}{df_{Res}}$	
Total	$SS_Y = \sum(Y_i - \bar{Y})^2$	$df_Y = N - 1$		

E. **Standard Error of Estimate**

$$s_e = \sqrt{\frac{SS_{Residual}}{df_{Residual}}}$$