Help for the Macintosh DataGen Statistics Window

The Statistics window presents all the basic statistics it can, given the structure of the data in the DataGen Spreadsheet. The sections below describe the five different states of the Statistics window.

- If the DataGen Spreadsheet has no values, or has values in only one variable, then the Statistics window is empty and looks like this:
• If the DataGen Spreadsheet has values in two variables and does not have an equal number of observations in those variables (or if there is a value in one column next to a blank in the other), then the label on the Statistics window changes to Two variables and the Statistics window automatically presents statistics for independent variables:

![DataGen Statistics: Two variables window](image)

Pooled variance

\[ S^2_{\text{pooled}} = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)} \]

Standard error of the difference between two means

\[ S_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s^2_{\text{pooled}}}{n_1} + \frac{s^2_{\text{pooled}}}{n_2}} \]

Observed value of t [H₀: \( \mu_1 = \mu_2 \)]

\[ t_{\text{obs}} = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}} \]

Degrees of freedom

\[ df = (n_1 - 1) + (n_2 - 1) \]

Critical value of t [\( \alpha = .05 \)]

\[ t_{cv} : \text{Student's } t \text{ with } df = (n_1 - 1) + (n_2 - 1) \]

The other statistics \( (n_1, n_2, \bar{X}_1, \bar{X}_2, s_1^2, s_2^2) \) required to compute these statistics are provided in the DataGen Descriptive Statistics window.
If the DataGen Spreadsheet has values in two variables, and has an equal number of observations in those variables (and there are no blanks in one column next to a value in the other), then the label on the Statistics window changes to Two variables (possibly paired) and the Statistics window automatically presents statistics for independent variables (as above) and the statistics for paired variables:

![DataGen Statistics: Two variables (possibly paired)](image)

- **Difference score**
  \[ D = X_1 - X_2 \]

- **Sum of difference scores**
  \[ \sum D \]

- **Mean of the difference scores**
  \[ \bar{D} = \frac{\sum D}{n} \]

- **Standard deviation of the differences**
  \[ s_D = \sqrt{\frac{\sum (D - \bar{D})^2}{n-1}} \]

- **Standard error [of the means of the differences]**
  \[ s_{\bar{D}} = \frac{s_D}{\sqrt{n}} \]

- **Observed value of \( t \)**  
  \[ t = \frac{\bar{D}}{s_{\bar{D}}} \]  
  \[ \text{[H}_0: \mu_1 = \mu_2] \]

- **Degrees of freedom**
  \[ n - 1 \]  
  [where \( n \) is number of pairs]

- **Critical value of \( t \)**
  \[ t_{cv} : \text{Student's} t \text{ with } df = n - 1 \]

- **Sum of the cross products**
  \[ \sum X_1 X_2 \]

- **Pearson correlation coefficient**
  \[ r = \frac{\sum z_1 z_2}{n-1} \]  
  where \( z_1 \) is z-score for \( X_1 \), etc.

- **Regression equation**
  \[ \bar{Y}_2 = b \bar{X}_1 + a \]
  \[ b = r \frac{s_2}{s_1} ; a = \bar{Y}_2 - b \bar{X}_1 \]
• If the DataGen Spreadsheet has values in three or more variables, and does not have an equal number of observations in all those variables, then the label on the Statistics window changes to **Three or more variables** and the Statistics window automatically presents statistics for independent samples analysis of variance:

![DataGen Statistics: 3 or more variables](image)

- **Pooled variance**
  \[ s^2_{\text{pooled}} = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2 + \cdots + (n_k-1)s_k^2}{n_G - k} \]

- **Total number of observations**
  \[ n_G = n_1 + n_2 + \cdots + n_k \]

- **Sum of all the observations**
  \[ \sum X_{ij} \]

- **Sum of all the squared observations**
  \[ \sum X_{ij}^2 \]

- **Variance of the means (unweighted)**
  \[ \frac{\sum(\bar{x}_j - \bar{x}_G)^2}{k-1} \]

- **Sum of squares between groups**
  \[ SS_B = \sum(\bar{x}_j - \bar{x}_G)^2 \]

- **Sum of squares within groups**
  \[ SS_W = \sum(X_{ij} - \bar{x}_j)^2 \]

- **Sum of squares total**
  \[ SS_T = \sum(X_{ij} - \bar{x}_G)^2 \]

- **Degrees of freedom between groups**
  \[ df_B = k - 1 \]

- **Degrees of freedom within groups**
  \[ df_W = n_G - k \]

- **Degrees of freedom total**
  \[ df_T = n_G - 1 \]

- **Mean square between groups**
  \[ MS_B = \frac{SS_B}{df_B} \]

- **Mean square within groups**
  \[ MS_W = \frac{SS_W}{df_W} \]

- **F ratio**
  \[ F_{\text{obs}} = \frac{MS_B}{MS_W} \]
If the DataGen Spreadsheet has values in three or more variables, and does have an equal number of observations (and parallel structure) in all those variables, then the label on the Statistics window changes to **Three or more variables (possibly related)** and the Statistics window automatically presents statistics for both the independent samples and the repeated measures analysis of variance:

![DataGen Statistics: 3 or more variables (possibly related)](image)

- **Number of subjects in each group:** $n$
- **Sum of squares between subjects:** $SS_{between\ subjects} = \sum (\bar{X}_i - \bar{X}_G)^2$
- **Sum of squares between occasions:** $SS_{between\ occasions} = \sum (\bar{X}_j - \bar{X}_G)^2$
- **Sum of squares residual:** $SS_{residual} = \sum (X_{ij} - \bar{X}_j)^2$
- **Sum of squares total:** $SS_{total} = \sum (X_{ij} - \bar{X}_i - \bar{X}_j + \bar{X}_G)^2$
- **Degrees of freedom between subjects:** $df_{between\ subjects} = n - 1$
- **Degrees of freedom between occasions:** $df_{between\ occasions} = k - 1$
- **Degrees of freedom residual:** $df_{residual} = (n - 1)(k - 1)$
- **Degrees of freedom total:** $df_{total} = nk - 1$
- **Mean square between subjects:** $MS_{between\ subjects} = \frac{SS_{between\ subjects}}{df_{between\ subjects}}$
- **Mean square between occasions:** $MS_{between\ occasions} = \frac{SS_{between\ occasions}}{df_{between\ occasions}}$
- **Mean square residual:** $MS_{residual} = \frac{SS_{residual}}{df_{residual}}$
- **$F$ ratio:** $F_{obs} = \frac{MS_{between\ occasions}}{MS_{residual}}$