

CHAPTER 4

# Computational Routines

## A. DESCRIPTIVE STATISTICS

**A1: Sum**

$$\sum X = \quad (1)$$

**A2: Mean**

$$\bar{X} = \frac{\sum X}{n} \quad (2)$$

**A3: Sum of Squares**

$$\sum (X - \bar{X})^2 \quad (3)$$

**A4: Variance**

$$VSS = \frac{\sum (X - \bar{X})^2}{n} \quad (4)$$

**A5: Standard Deviation**

$$SD = \sqrt{V} \quad (5)$$

**A6: Standard Error of the Mean**

$$SEM = \frac{SD}{\sqrt{n}} \quad (6)$$

**A7: Confidence Limits**

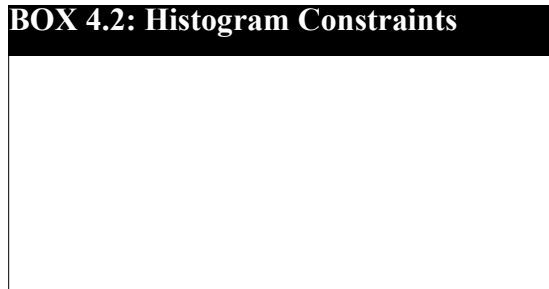
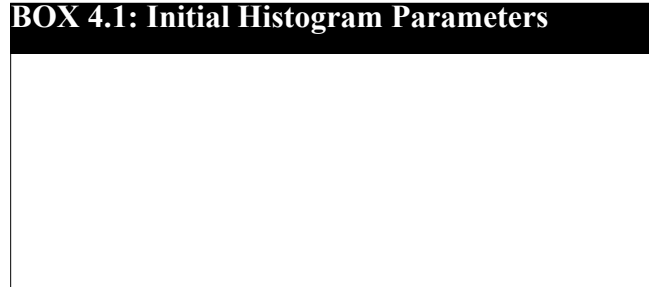
$$\text{Confidence Interval} = \bar{X} \pm t_{\alpha/2} \cdot SEM \quad (7)$$

**A8: Median**

The Median is defined as that observation in the sample that has 50% of the sample's observations being smaller than itself and 50% being larger than itself. In the event that the number of observations in the sample is an even number, then the Median is defined as the mean of the two central observations.

**B. HISTOGRAM PLOT**

STATsimple's Histogram Plot allows you to define plotting parameters such as the number of classes, the lower limit and the upper limit. Initially, STATsimple calculates those parameters as defined in Box 4.1. You may then change them within the constraints listed in Box 4.2.



The class width (CW) and class boundaries are calculated from the defined number of classes (NC), lower limit (LL) and upper limit (UL):

$$CW = \frac{UL - LL}{NC} \tag{8}$$

$$LCB = LL + (i-1)CW \tag{9}$$

$$UCB = LCB + CW \tag{10}$$

where  $LCB_i$  is the Lower Class Boundary of the  $i^{th}$  Class and  $UCB_i$  is the Upper Class Boundary of the  $i^{th}$  Class.

An observation,  $x$ , belongs to the  $i^{th}$  Class if:

$$x \geq LCB \quad \text{and} \quad x < UCB \tag{11}$$

**C. STUDENT'S T-TEST**

The Student's t-Test is used to determine whether two samples were likely taken from the same population. This version of STATsimple uses a two-tailed t-Test for unpaired samples. "V" is the sample variance.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{V_1 + V_2}{n_1 + n_2 - 2}}}$$
(12)

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{V_1 + V_2}{n_1 + n_2 - 2}}}$$
(13)

Degrees of Freedom (DF) =  $n_1 + n_2 - 2$  (14)

**D. LINEAR REGRESSION**

Linear Regression is used to calculate the best fitting straight line through a set of data points. The result is given as a linear function of the form:

$$Y = aX + b \quad [r]$$
(15)

where "r" is the regression coefficient.

$$a = \frac{\sum(XY) - \bar{X}\bar{Y}}{\sum(X^2) - n\bar{X}^2}$$
(16)

$$b = \bar{Y} - a\bar{X}$$
(17)

$$r = \frac{\sum(XY) - \bar{X}\bar{Y}}{\sqrt{(\sum(X^2) - n\bar{X}^2)(\sum(Y^2) - n\bar{Y}^2)}}$$
(18)