MATH 545
DR. MCLOUGHLIN'S CLASS
STATISTICAL FORMULAE
HANDOUT 3

Let D = $\{X_1, X_2, X_3, \ldots, X_n\}$ be a finite data set.

Let $X_1, X_2, X_3, \ldots, X_n$ be a finite random sample for X.

The **arithmetic mean of the sample** is the value $\overline{X}$ where $\overline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j = \frac{X_1 + X_2 + \ldots + X_n}{n}$

The **variance**, s^2, is defined as $s^2 = \frac{\sum_{k=1}^{n}(X_k - \overline{X})^2}{n-1}$

The **standard deviation**, s, is defined as $s = \sqrt{\frac{\sum_{k=1}^{n}(X_k - \overline{X})^2}{(n-1)}}$

The **mean absolute deviation (MAD)**, M, is defined as $M = \frac{\sum_{k=1}^{n}\left|X_k - \overline{X}\right|}{n}$

The **range** is the (highest value – lowest value)

Let $X_1, X_2, X_3, \ldots, X_n$ be a finite random sample for X. The **mode of the sample** is the value of X which occurs most frequently such that there is at least one value that occurs with lesser frequency.

Let $X_1, X_2, X_3, \ldots, X_n$ be a finite random sample for X. The **median of the sample** is the value of X which occurs in the centre of ordered values of the sample if there are an odd number of values and it is the arithmetic mean of the centre two values if there are an even number of values.

The **geometric mean of the sample** is the value G where $G = \sqrt[n]{\prod_{j=1}^{n} X_j} = \sqrt[n]{X_1 \cdot X_2 \cdot \cdots \cdot X_n}$

The **harmonic mean of the sample** is the value H where $H = \frac{n}{\sum_{j=1}^{n}\frac{1}{X_j}} = \frac{n}{\frac{1}{X_1} + \frac{1}{X_2} + \cdots + \frac{1}{X_n}}$

The **coefficient of variation of the sample** is the value C where $C = \frac{s}{\overline{X}} \cdot (100\%)$

Computational examples:

Suppose the X_i are given by $X_1 = 0.1$, $X_2 = 0.1$, $X_3 = 0.1$, and $X_4 = 10$.

$\overline{X} = 2.575$, $S^2 = 24.5025$, $S = \sqrt{24.5025} = 4.95$, $m_o = 0.1$, $m_d = 0.1$, etc.

Suppose the X_i are given by $X_1 = 0.1$, $X_2 = 10$, $X_3 = 10$, and $X_4 = 10$.

$\overline{X} = 7.525$, $S^2 = 57.1725$, $S = \sqrt{57.1725}$, $S \approx 7.561249897$, $m_o = 10$, $m_d = 10$, etc.

Suppose the X_i are given by $X_1 = 1$, $X_2 = 2$, $X_3 = 3$, and $X_4 = 4$.

$\overline{X} = \frac{5}{2}$, $S^2 = \frac{5}{3}$, $S = \sqrt{\frac{5}{3}}$, $S \approx 1.290994449$, m_0 does not exist, $m_d = 2.5$, etc.

To convert from X values to Z values it is the case that: $Z = \frac{x - \mu_x}{\sigma_x}$

$P(X \geq x) = P(Z \geq z)$ when $X \sim N(x, \mu, \sigma)$ when $X \sim N(x, \mu, \sigma)$

$$\mathcal{N}(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x_i - \mu)^2}{2\sigma^2}} \text{ where } x \in \mathbb{R}$$

$$\mathcal{N}(z, \mu = 0, \sigma = 1) = \frac{1}{\sqrt{2\pi}} e^{\frac{-(z)^2}{2}} \text{ where } z \in \mathbb{R}$$

The approximate areas (probabilities) are as follows: