# EMGMT formula sheet statistics

## Measures

Mean:

$$\mu_x = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Standard score:

$$\frac{x_i - \mu_x}{\sigma_x}$$

Variance of population:

$$\begin{aligned}\sigma_x^2 &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_x)^2 \\ &= E((X - \mu_x)^2) \\ &= E(X^2) - (E(X))^2\end{aligned}$$

Standard deviation:

$$\sqrt{\sigma_x^2}$$

Unbiased estimator for mean:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Unbiased estimator for variance:

$$\hat{\sigma}_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{X})^2$$

Unbiased estimator for standard deviation:

$$\sqrt{\hat{\sigma}_x^2}$$

## Variations and combinations

Variation (no sample replacement, with ordering):

$$P(n,k) = \frac{n!}{(n-k)!}$$

Repeating variation (with sample replacement, with ordering):

$$n^k$$

Combination (no sample replacement, no ordering):

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Repeating combination (with sample replacement, no ordering):

$$\binom{k+n-1}{k} = \frac{(k+n-1)!}{k!(n-1)!}$$

## Probability

Basic rules logic and probability:

$$\neg(A \vee B) = \neg A \wedge \neg B$$

$$\neg(A \wedge B) = \neg A \vee \neg B$$

$$P(\neg A) = 1 - P(A)$$

$$P(A \vee B) = P(A) + P(B) - P(A \wedge B)$$

Bayes:

$$P(A|B) = \frac{P(A \wedge B)}{P(B)}$$

Expected value:

$$\begin{aligned} E(X) &= \int x f(x) dx \quad \text{(continuous case)} \\ & \sum x_i P(x_i) \quad \text{(discrete case)} \end{aligned}$$

Normal (Gaussian) probability density function:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

## Covariance and Correlation

Covariance:

$$\sigma_{XY} = Cov(X, Y) = E(XY) - E(X)E(Y)$$

Unbiased estimator for covariance:

$$\hat{\sigma}_{XY} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{X})(y_i - \bar{Y})$$

Correlation:

$$\rho_{XY} = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2 \sigma_Y^2}}$$

## The Student t-test

One-sample t-test:

$$H_0 : \mu = \mu_0$$

$$T = \frac{\bar{X} - \mu_0}{\hat{\sigma}} \sqrt{n}$$

$$df = n - 1$$

Two-sample unpaired t-test:

$$H_0 : \mu_0 = \mu_1$$

$$T = \frac{\bar{X} - \bar{Y}}{S_{XY} \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

where $S_{XY}$ is the weighted unbiased standard deviation defined as follows:

$$S_{XY} = \sqrt{\frac{(n-1)\sigma_X^2 + (m-1)\sigma_Y^2}{n+m-2}}$$

$$df = n + m - 2$$

Two sample paired t-test:

$$Z = X - Y$$

$$H_0 : \mu_Z = 0$$

$$T = \frac{\bar{Z}}{\hat{\sigma_Z}} \sqrt{n}$$

$$df = n - 1$$