Formula sheet

Sample statistics

Sample mean:

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

Sample variance:

$$s_x^2 = \frac{1}{n-1}\sum (x_i - \bar{x})^2 = \frac{1}{n-1}\left(\sum x_i^2 - n\bar{x}^2\right)$$

Sample covariance:

$$\begin{aligned} s_{xy} &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \\ &= \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}\right) \end{aligned}$$

Sample correlation:

$$\frac{s_{xy}}{s_x s_y}$$

Percentile position (of ordered data):

$$\frac{p}{100}(n+1)$$

Combinatorics

Permutation:

$$P_r^n = \frac{n!}{(n-r)!}$$

Combination:

$$C_r^n = \frac{n!}{r!(n-r)!}$$

Probability

Addition law:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Conditional probability:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad P(B|A) = \frac{P(A \cap B)}{P(A)}$$

Bayes Theorem:

$$P(A_i|B) = \frac{P(B|A_i)\,P(A_i)}{P(B)} = \frac{P(B|A_i)\,P(A_i)}{\sum_j P(B|A_j)\,P(A_j)}$$

Random Variables

Mean:

$$E(X) = \int_{x \in S} x f(x) dx \text{ or } \sum_{\forall x \in S} x p(x)$$

Variance:

$$V(X) = E[\{X - E(X)\}^2] = E(X^2) - E^2(X)$$

Covariance:

$$\begin{aligned} \mathrm{Cov}(X,Y) &= E[\{X - E(X)\}\{Y - E(Y)\}] \\ &= E(XY) - E(X)E(Y) \end{aligned}$$

Correlation:

$$\frac{\mathrm{Cov}(X,Y)}{\sqrt{V(X)V(Y)}}$$

Population Percentiles:

$$F(x_p) = \int_{-\infty}^{x_p} f(t)dt = \frac{p}{100}$$

Conditional Distribution:

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}$$

Conditional Expectation:

$$E(Y|X) = \int_{\infty}^{\infty} y f_{Y|X}(y|x) dy$$

Chebyshev Inequality:

- $P(|X - \mu_X| \geq k\sigma) \leq \frac{1}{k^2}$
- $P(|X - \mu_X| < k\sigma) \geq 1 - \frac{1}{k^2}$

Cumulative Distribution Function:

$$F(x) = \int_{-\infty}^{x} f(t)dt \text{ or } \sum_{t \leq x} p(t)$$

Linear transformations

Mean:

$$E(aX + b) = aE(X) + b$$

Variance:

$$V(aX + b) = a^2 V(X)$$