# Statistical Equations

You are permitted to use the information on these pages during your exams. These pages are not guaranteed to contain all the information you will need. If you find information which you believe should be included, you may request that the information be added. All changes to these pages must be made prior to the exam.

**Probabilities**

Marginal probability: Probability of a single event occurring

$\Pr(A)$

Complement probability: Probability of a single event not occurring

$\Pr(A') = 1 - \Pr(A)$

Joint probability: Probability of two or more events occurring together ($\cap$ = "and")

$\Pr(A \cap B \cap ... N) = \Pr(A)\Pr(B)...\Pr(N)$ when $A, B, ..., N$ are independent

$\Pr(A \cap B \cap ... N) = \Pr(A \mid B)\Pr(B) + ... + \Pr(A \mid N)\Pr(N)$ when $A, B, ..., N$ are dependent

Disjoint probability: Probability of either (or both) of two events occurring ($\cup$ = "or")

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

Conditional probability: Probability of one event occurring given the occurrence of a related event ($\mid$ = "given")

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

Events $A$ and $B$ are unrelated if $\Pr(A \mid B) = \Pr(A)$

Bayes' Theorem: Reversed conditional probability

$$\Pr(A \mid B) = \frac{\Pr(B \mid A)\Pr(A)}{\Pr(B)}$$

$\Pr(B) = \Pr(B \mid A_1)\Pr(A_1) + ... + \Pr(B \mid A_n)\Pr(A_n)$ events $A_1, ..., A_n$ are mutually exclusive and jointly exhaustive

Mutually Exclusive

Events are mutually exclusive when they can never occur together.

Jointly Exhaustive

Events are jointly exhaustive when at least one of them must occur.

Chebyshev's Inequality

Regardless of the distribution from which the observations are drawn,

$$\Pr\left(m - ks < x < m + ks\right) \geq 1 - \frac{1}{k^2} \quad \text{for } k > 1$$

The probability of an observation falling within $k$ standard deviations of the mean is $1 - 1/k^2$.

**Combinatorics**

Combinations

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

Excel function: =COMBIN(n, x)

Permutations

$${}_nP_x = \frac{n!}{(n-x)!}$$

Excel function: =PERMUT(n, x)

**Sample and Population Statistics**

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

Population Mean: $m$

Sample Variance: $s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

Population Variance: $s_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - m_x)^2$

Variance of a Proportion: $s_p^2 = \frac{p(1-p)}{n}$

Sample Covariance: $s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i y_i - n\overline{xy}\right)$

Population Covariance: $s_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - m_x)(y_i - m_y) = \frac{1}{n}\left(\sum_{i=1}^{n} x_i y_i - n m_x m_y\right)$

Sample Correlation: $r_{xy} = \frac{s_{xy}}{s_x s_y}$

Population Correlation: $r_{xy} = \frac{s_{xy}}{s_x s_y}$

Standard Deviation $= \sqrt{\text{Variance}}$

Excel functions:

| | |
|---|---|
| =AVERAGE(*range*) | mean |
| =STDEV(*range*) | sample standard deviation |
| =STDEVP(*range*) | population standard deviation |
| =VAR(*range*) | sample variance |
| =VARP(*range*) | population variance |
| =COVAR(*range*)**n*/(*n*-1) | sample covariance |
| =COVAR(*range*) | population covariance |

**Identifying Distributions**

| Distribution | Known Information |
|---|---|
| Binomial | Probability of a single success<br>Number of trials<br>Number of successes |
| Negative Binomial | Probability of a single success<br>Number of trials<br>Number of successes<br>Trials stop when last success is achieved |
| Hypergeometric | Number of possible trials<br>Number of actual trials<br>Number of possible successes<br>Number of actual successes |
| Poisson | Average successes per unit time<br>Number of successes per unit time |
| Exponential | Average successes per unit time<br>Maximum time to the next success |

**Discrete Distributions**

Uniform Distribution

$$\Pr(X = x) = \begin{cases} \dfrac{1}{b-a} & \text{for } a < x < b \\ 0 & \text{for } x < a,\ x > b \end{cases}$$

$$\text{Mean of Uniform Distribution} = \frac{1}{2}(a+b)$$

$$\text{Standard Deviation of Uniform Distribution} = \sqrt{\frac{1}{12}(b-a)(b+a)}$$

Binomial Distribution

$$\Pr(x \text{ successs in a sample of } n) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$\text{Mean of Binomial Distribution} = np$$

$$\text{Standard Deviation of Binomial Distribution} = \sqrt{np(1-p)}$$

Excel function: =BINOMDIST(x, n, p, TRUE/FALSE)

Hypergeometric Distribution

$$\Pr(x \text{ successes in a sample of } n \text{ taken from a population of } N \text{ with } X \text{ successes}) = \frac{\binom{X}{x}\binom{N-X}{n-x}}{\binom{N}{n}}$$

Excel function: =HYPGEOMDIST(x, n, X, N)

Negative Binomial Distribution

$$\text{Pr(achieving } x^{\text{th}} \text{ success on the } n^{\text{th}} \text{ draw)} = \binom{n-1}{x-1} p^{x}(1-p)^{n-x}$$

Excel function: =NEGBINOMDIST(n-x, x, p)

Poisson Distribution

The Poisson Distribution approximates the binomial distribution for $n$ = 30 and $np$ or $n(1 - p) < 5$.

$$\text{Pr(}x\text{ successes per unit time given an average of } l \text{ successes per unit time)} = \frac{e^{-l} l^{x}}{x!}$$

Excel function: =POISSON(x, λ, TRUE/FALSE)

**Continuous Distributions**

Exponential Distribution

$$\text{Pr(success occurring in } x \text{ or fewer time intervals given } l \text{ successes per time interval)} = 1 - e^{-lx}$$

$$\text{E}\left(\text{time intervals to a single success}\right) = \frac{1}{l}$$

$$\text{Var}\left(\text{time intervals to a single success}\right) = \frac{1}{l^{2}}$$

Excel function: =EXPONDIST(x, λ, TRUE)

Gamma Distribution

$$Gamma(x, a, b) = \begin{cases} \dfrac{x^{a} e^{\frac{-x}{b}}}{b^{a}\Gamma(a)} & \forall\, x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

mean = $ab$

variance = $ab^{2}$

Converting a Normal Distribution to a Standard Normal Distribution

$$Z = \frac{X - m}{s}$$

Excel function: =STANDARDIZE(X, μ, σ)

Converting a t-Distribution to a Standard t-Distribution

$$t = \frac{X - \overline{X}}{s}$$

Excel function: =STANDARDIZE(X, X bar, s)

Confidence Interval

Confidence interval for $X$: $X$ ±(Critical Value)(Standard deviation of $X$)

Excel functions:

| | |
|---|---|
| =NORMSINV(p) | Finds critical value, C, such that Pr(Z > C) = p where Z is distributed standard normal |
| =NORMS(C) | Finds Pr(Z > C) where Z is distributed standard normal |
| =TINV(p, df) | Finds critical value, C, such that Pr(-C < T < C) = p where T is distributed $t_{df}$ |
| =TDIST(C, df, D) | Finds Pr(-C < T < C) when D = 2, and finds Pr(T > C) when D = 1 where T is distributed $t_{df}$ |

**Statistical Tests**

Finite Population Correction Factor

$$\text{correction factor} = \frac{N-n}{N-1}$$ Multiply variances of sample means and proportions by this factor when $n / N \geq 0.05$

Test for Difference in Means (large sample)

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (m_1 - m_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Test for Difference in Means (small sample)

$$t = \frac{\bar{X}_1 - \bar{X}_2 - (m_1 - m_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \qquad df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

Test for Difference in Variances

$$F = \frac{s_1^2}{s_2^2}$$ where $s_1 > s_2$ df numerator $= n_1 - 1$ df denominator $= n_2 - 1$

Test for Difference in Multiple Means (note: by definition, this is a one-tailed test)

There are c cases.
Let $X_{ij}$ be the $i^{th}$ observation of the $j^{th}$ case.
Let $n_j$ be the number of observations in the $j^{th}$ case.
Let $\bar{X}_j$ be the mean for the $j^{th}$ case.
Let $\bar{\bar{X}}$ be the grand mean.

Mean squared between deviation (MSB) = $\frac{1}{c-1}\sum_{j=1}^{c} n_j \left(\bar{X}_j - \bar{\bar{X}}\right)^2$

Mean squared within deviation (MSW) = $\frac{1}{n-c}\sum_{j=1}^{c}\sum_{i=1}^{n_j}\left(X_{ij} - \bar{X}_j\right)^2$

$$\text{F-stat} = \frac{\text{MSB}}{\text{MSW}}$$ df numerator $= c - 1$ df denominator $= n - c$

Test for Difference in Paired Means (note: by definition, this is a one-tailed test)

Two sample means are statistically different if $\left|\overline{X}_1 - \overline{X}_2\right|$ > Critical Value

$$\text{Critical Value} = Q\sqrt{\frac{MSW}{2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} \qquad \text{df numerator} = c \qquad \text{df denominator} = n - c$$

Test for Differences in Frequencies

$$c^2 = \sum_{allcells} \frac{(f_o - f_e)^2}{f_e} \qquad \text{df} = (\text{rows-1})(\text{columns-1})$$

$f_o$ = observed frequency
$f_e$ = expected frequency