Blekinge Institute of Technology
Institute of mathematics
and natural sciences

Tables and Formulas

Mathematical Statistics

(Intended for use at written examination)
by
Mattias Dahl, Efraim Laksman och Robert Nyqvist

1 Distributions

1.1 Continuous distributions (density functions)

1.1.1 Uniform distribution

$$f_X(x) = \frac{1}{b-a}, a \leq x \leq b, \quad E(X) = \frac{a+b}{2}, \quad V(X) = \frac{(b-a)^2}{12}$$

1.1.2 Exponential distribution

$$f_X(x) = \lambda e^{-\lambda x}, x \geq 0, \quad E(X) = \frac{1}{\lambda}, \quad V(X) = \frac{1}{\lambda^2}$$

1.1.3 Weibull distribution

$$f_X(x) = \frac{c}{b}\left(\frac{x}{b}\right)^{c-1} e^{-(x/b)^c}, x \geq 0, \quad E(X) = b\Gamma((c+1)/c)$$

$$V(X) = b^2(\Gamma((c+2)/c) - \Gamma((c+2)/c)^2), \text{ Gammafunction} : \Gamma(x) = \int_0^\infty \mathrm{e}^{-y} y^{x-1} dy$$

1.1.4 Gaussian distribution (normal distribution)

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right), \quad E(X) = m, \quad V(X) = \sigma^2$$

1.2 Discrete distributions (probability functions)

1.2.1 Binomial distribution

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, x = 0, 1, \ldots, n, \quad E(X) = np, \quad V(X) = np(1-p)$$

1.2.2 Hypergeometric distribution

$$p_X(x) = \frac{\binom{Np}{x}\binom{Nq}{n-x}}{\binom{N}{n}}, \quad E(X) = np, \quad V(X) = npq\frac{N-n}{N-1}$$

1.2.3 Poisson distribution

$$p_X(x) = e^{-m}\frac{m^x}{x!}, x = 0, 1, 2, \ldots, \quad E(X) = V(X) = m$$

1.3 Approximations

1.3.1 Binomial distribution may be approximated with Poisson distribution if $p \leq 0.1, n \geq 10$.

1.3.2 Binomial distribution may be approximated with normal distribution if $npq \geq 10$.

1.3.3 Poisson distribution may be approximated with normal distribution if $m \geq 15$.

1.4 Gauss formulas for approximation:

$$E(g(X)) \approx g(E(X)),$$
$$V(g(X)) \approx (g'(E(X)))^2 V(X)$$

2 Stochastic processes

2.1 Stochastic processes in discrete time

2.1.1 Stationary (and asymptotic) state

$$\pi = \pi P$$

2.2 Stochastic processes in continuous time

2.2.1 Connection between hazard function $\lambda(t)$ and reliability $R(t)$

$$R(t) = e^{-\int_0^t \lambda(u)du}$$

$$\lambda(t) = -\frac{R'(t)}{R(t)}$$

2.2.2 The distribution vector $p(t)$ and intensity matrix A satisfy the relation

$$p'(t) = p(t)A$$

2.2.3 Stationary (and asymptotic) state

$$0 = \pi A$$

2.2.4 Erlang's formula

$$\pi_k = \frac{\rho^k/k!}{\rho^0/0! + \rho^1/1! + \rho^2/2! + \ldots + \rho^n/n!}, k = 0, 1, \ldots, n$$

3 Estimations

3.1 Point estimations

3.1.1 Expected value

$$\hat{m} = \bar{x} = \frac{1}{n}\sum x$$

3.1.2 Standard deviation

$$s = \sqrt{\frac{1}{n-1}(\sum x^2 - \frac{1}{n}(\sum x)^2)}$$

3.1.3 The least square function

$$Q_0(x_1, \ldots, x_n|\theta) = \sum(x - m(\theta))^2$$

3.1.4 The likelihood function (continuous s.v.)

$$L(x_1, \ldots, x_n|\theta) = f_{X_1}(x_1|\theta)\cdots f_{X_n}(x_n|\theta)$$

3.1.5 The likelihood function (discrete s.v.)

$$L(x_1, \ldots, x_n|\theta) = p_{X_1}(x_1|\theta)\cdots p_{X_n}(x_n|\theta)$$

3.2 Confidence intervals

3.2.1 Normal distribution with known standard deviation

$$\bar{x} \pm \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

3.2.2 Normal distribution with unknown standard deviation

$$\bar{x} \pm t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}$$

3.2.3 Prediction interval with linear regression

$$\hat{a} + \hat{b}x_0 \pm t(n-2)s\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$$

3.2.4 Comparison of expected values in two groups with normal distribution with known standard deviation

$$\bar{x} - \bar{y} \pm \lambda_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}$$

3.2.5 Comparison of expected value in two groups with normal distribution with unknown standard deviation

$$\bar{x} - \bar{y} \pm t_{\alpha/2}(n_x - 1 + n_y - 1)s\sqrt{\frac{1}{n_x} + \frac{1}{n_y}},$$

where

$$s = \sqrt{\frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x - 1 + n_y - 1}}$$

3.2.6 Normal approximation of binomial distribution

$$\hat{p} \pm \lambda_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

3.2.7 Normal approximation of Poisson distribution

$$\bar{x} \pm \lambda_{\alpha/2}\sqrt{\frac{\bar{x}}{n}}$$

3.2.8 The standard deviation in a normal distribution

$$\left(\sqrt{\frac{n-1}{\chi^2_{\alpha/2}(n-1)}}, \sqrt{\frac{n-1}{\chi^2_{1-\alpha/2}(n-1)}}\right)$$

4 Hypothesis testing

4.1 Test variables

4.1.1 Normal distribution with known standard deviation

$$u = \frac{\bar{x} - m_0}{\sigma/\sqrt{n}}$$

4.1.2 Normal distribution with unknown standard deviation

$$u = \frac{\bar{x} - m_0}{s/\sqrt{n}}$$

4.1.3 Comparison of expected value in two groups of normal distribution with known standard deviation

$$u = \frac{\bar{x} - \bar{y} - C}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}$$

4.1.4 Comparison of expected value in two groups of normal distribution with unknown but similar standard deviation

$$u = \frac{\bar{x} - \bar{y} - C}{s\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}},$$

where

$$s = \sqrt{\frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x - 1 + n_y - 1}}$$

4.1.5 Normal approximation of binomial distribution

$$u = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

4.1.6 Normal approximation of Poisson distribution

$$u = \frac{\bar{x} - m_0}{\sqrt{m_0/n}}$$

4.1.7 Comparison with some distribution

$$\chi^2 = \sum \frac{(O - E)^2}{E},$$

with $k - a - 1$ degrees of freedom.

4.1.8 Comparison with other observations

$$\chi^2 = \sum \frac{(O - E)^2}{E},$$

with $(r - 1)(k - 1)$ degrees of freedom.

4.2 Simple linear regression

4.2.1 Model

$$y = a + bx + \epsilon$$

4.2.2 Sums of squares

$$S_{xx} = \sum x^2 - \frac{1}{n}(\sum x)^2 = \sum (x - \bar{x})^2$$
$$S_{xy} = \sum xy - \frac{1}{n}(\sum x)(\sum y) = \sum (x - \bar{x})(y - \bar{y})$$
$$S_{yy} = \sum y^2 - \frac{1}{n}(\sum y)^2 = \sum (y - \bar{y})^2$$

4.2.3 Estimators of parameters

$$\hat{b} = \frac{S_{xy}}{S_{xx}}, \quad \hat{a} = \bar{y} - \hat{b}\bar{x}$$

4.2.4 Sum of squares errors

$$Q_0 = S_{yy} - \frac{S_{xy}^2}{S_{xx}}$$

4.2.5 Estimator of standard deviation of residuals

$$s = \sqrt{\frac{Q_0}{n - 2}}$$

4.2.6 Estimators of standard deviation of estimators of parameters

$$s_a = s\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}, \quad s_b = \frac{s}{\sqrt{S_{xx}}}$$

4.2.7 Hypothesis test for parameters

$$u = \frac{a^* - a_0}{s_a}, \quad u = \frac{b^* - b_0}{s_b}$$

4.2.8 Confidence interval for simple linear regression

$$\hat{a} + \hat{b}x_0 \pm t(n - 2)s\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$$

4.2.9 Correlations

$$r_{Pearson} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

$$r_{Spearman} = 1 - \frac{6\sum d^2}{n^3 - n}$$

4.2.10 Hypothesis test for correlation

$$u = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

4.2.11 Prediction interval for simple linear regression

$$\hat{a} + \hat{b}x_0 \pm t(n-2)s\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$$