LIST OF FORMULAS FOR STK1100 AND STK1110

(Version of 11. November 2015)

1. Probability

Let $A, B, A_1, A_2, \ldots, B_1, B_2, \ldots$ be events, that is, subsets of a sample space Ω.

a) Axioms:
A probability function P is a function from subsets of the sample space Ω to real numbers, satisfying

$$
\begin{aligned}
&P(\Omega) = 1 \\
&P(A) \geq 0 \\
&P(A_1 \cup A_2) = P(A_1) + P(A_2) \quad \text{if } A_1 \cap A_2 = \emptyset \\
&P\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} P(A_i) \quad \text{if } A_i \cap A_j = \emptyset \text{ for } i \neq j
\end{aligned}
$$

b) $P(A') = 1 - P(A)$

c) $P(\emptyset) = 0$

d) $A \subset B \Rightarrow P(A) \leq P(B)$

e) The addition law of probability/ the sum rule:

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

f) Conditional probability:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{if } P(B) > 0$$

g) Total probability:

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i) \quad \text{if } \bigcup_{i=1}^{n} B_i = \Omega \text{ and } B_i \cap B_j = \emptyset \text{ for } i \neq j$$

h) Bayes' Rule:

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)} \quad \text{under same conditions as in g)}$$

i) A and B are (statistically) independent events if $P(A \cap B) = P(A)P(B)$

j) $A_1, \ldots, A_n$ are (statistically) independent events if

$$P(A_{i_1} \cap \cdots \cap A_{i_m}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_m})$$

for any subset of indexes $i_1, i_2, \ldots, i_m$

k) <u>The product rule:</u>

$$\begin{aligned} &P(A_1 \cap \cdots \cap A_n) \\ &\quad = \; P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1}) \end{aligned}$$

2. Combinatorics

a) Two operations that can be done in respectively n and m different ways can be combined in $n \cdot m$ ways.

b) The number of ordered subsets of r elements drawn with replacement from a set of n elements is n^r

c) The number of ordered subsets of r elements drawn without replacement from a set of n elements is $n(n-1)\cdots(n-r+1)$

d) Number of permutations of n elements is $n! = 1 \cdot 2 \cdot 3 \cdots (n-1) \cdot n$

e) The number of unordered subsets of r elements drawn from a set of n elements is

$$\binom{n}{r} = \frac{n(n-1)\cdots(n-r+1)}{r!} = \frac{n!}{r!\,(n-r)!}$$

f) Number of ways a set of n elements can be divided into r subsets with n_i elements in the ith subset is

$$\binom{n}{n_1 \; n_2 \; \cdots \; n_r} = \frac{n!}{n_1!\, n_2! \, \cdots \, n_r!}$$

3. Probability distributions

a) For a random variable X (discrete or continuous), $F(x) = P(X \leq x)$ is the cumulative distribution function (cdf).

b) For a discrete random variable X which can take the values $x_1, x_2, x_3, \ldots$ we have

$$p(x_j) = P(X = x_j)$$
$$F(x) = \sum_{x_j \leq x} p(x_j)$$

$p(x_j)$ is a point probability if

$$p(x_j) \geq 0 \quad \text{for all } j$$
$$\sum_j p(x_j) = 1$$

c) For a continuous random variable X we have

$$P(a < X < b) = \int_a^b f(x)dx$$
$$F(x) = \int_{-\infty}^x f(u)du$$
$$f(x) = F'(x)$$

$f(x)$ is a probability density function if

$$f(x) \geq 0$$
$$\int_{-\infty}^{\infty} f(x)dx = 1$$

d) For two random variables X and Y (discrete or continuous) the joint cumulative distribution function is $F(x, y) = P(X \leq x, Y \leq y)$

e) For discrete random variables X and Y which can take the values $x_1, x_2, \ldots$ and $y_1, y_2, \ldots$ respectively, we have

$$p(x_i, y_j) = P(X = x_i, Y = y_j)$$
$$F(x, y) = \sum_{x_i \leq x} \sum_{y_j \leq y} p(x_i, y_j)$$

$p(x_i, y_j)$ is a joint point probability if it fullfills the same conditions as in b)

f) For continuous random variables X and Y we have

$$P\left((X,Y) \in A \right) = \int\int_A f(u,v)dvdu$$

$$F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)dvdu$$

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}$$

$f(x,y)$ is a joint probability density function if it fullfills the same conditions as in c)

g) Marginal point probabilities:

$$p_X(x_i) = \sum_j p(x_i,y_j) \qquad \text{(for } X\text{)}$$

$$p_Y(y_j) = \sum_i p(x_i,y_j) \qquad \text{(for } Y\text{)}$$

h) Marginal probability densities:

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy \qquad \text{(for } X\text{)}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx \qquad \text{(for } Y\text{)}$$

i) Independence:

The random variables X and Y are independent if

$$p(x_i,y_j) = p_X(x_i)p_Y(y_j) \qquad \text{(discrete)}$$

$$f(x,y) = f_X(x)f_Y(y) \qquad \text{(continuous)}$$

j) Conditional point probabilities:

$$p_{X|Y}(x_i|y_j) = \frac{p(x_i,y_j)}{p_Y(y_j)} \qquad \text{(for } X \text{ given } Y = y_j\text{)}$$

$$p_{Y|X}(y_j|x_i) = \frac{p(x_i,y_j)}{p_X(x_i)} \qquad \text{(for } Y \text{ given } X = x_i\text{)}$$

Assuming $p_Y(y_j) > 0$ and $p_X(x_i) > 0$, respectively. Conditional point probabilities can be treated as regular point probabilities.

k) Conditional probability densities:

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} \qquad \text{(for } X \text{ given } Y = y\text{)}$$
$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} \qquad \text{(for } Y \text{ given } X = x\text{)}$$

Assuming $f_Y(y) > 0$ and $f_X(x) > 0$, respectively. Conditional probability densities can be treated as regular probability densities.

4. Expectation

a) The expected value of a random variable X is defined as

$$E(X) = \sum_j x_j p(x_j) \qquad \text{(discrete)}$$
$$E(X) = \int_{-\infty}^{\infty} x f(x) dx \qquad \text{(continuous)}$$

b) For a real function $g(X)$ of a random variable X, the expectated value is

$$E[g(X)] = \sum_j g(x_j) p(x_j) \qquad \text{(discrete)}$$
$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx \qquad \text{(continuous)}$$

c) $E(a + bX) = a + bE(X)$

d) For a real function $g(X, Y)$ of two random variables X and Y, the expected value is

$$E\big[g(X,Y)\big] = \sum_i \sum_j g(x_i, y_j) p(x_i, y_j) \qquad \text{(discrete)}$$
$$E\big[g(X,Y)\big] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f(x,y) dy dx \qquad \text{(continuous)}$$

e) If X and Y are independent $E\big[g(X)h(Y)\big] = E\big[g(X)\big] \cdot E\big[h(Y)\big]$

f) If X and Y are independent $E(XY) = E(X) \cdot E(Y)$

g) $E\left(a + \sum_{i=1}^{n} b_i X_i\right) = a + \sum_{i=1}^{n} b_i E(X_i)$

h) Conditional expectation:

$$\mathrm{E}(Y|X=x_i) = \sum_j y_j p_{Y|X}(y_j|x_i) \qquad \text{(discrete)}$$

$$\mathrm{E}(Y|X=x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x)dy \qquad \text{(continuous)}$$

5. Variance and standard deviation

a) The variance and standard deviation of a random variable X are defined as

$$\mathrm{V}(X) = \mathrm{E}[(X-\mu)^2]$$
$$\mathrm{sd}(X) = \sqrt{\mathrm{V}(X)}$$

b) $\mathrm{V}(X) = \mathrm{E}(X^2) - \big(\mathrm{E}(X)\big)^2$

c) $\mathrm{V}(a+bX) = b^2\,\mathrm{V}(X)$

d) If $X_1, \ldots, X_n$ are independent we have

$$\mathrm{V}\left(a + \sum_{i=1}^{n} b_i X_i\right) = \sum_{i=1}^{n} {b_i}^2\,\mathrm{V}(X_i)$$

e)

$$\mathrm{V}\left(a + \sum_{i=1}^{n} b_i X_i\right) = \sum_{i=1}^{n} {b_i}^2\,\mathrm{V}(X_i) + \sum_{i=1}^{n}\sum_{j\neq i} b_i b_j \mathrm{Cov}(X_i, X_j)$$

f) Chebyshev's inequality:

Let X be a random variable with $\mu = \mathrm{E}(X)$ and $\sigma^2 = \mathrm{V}(X)$.
For all $t > 0$ we have

$$P(|X-\mu| > t) \leq \frac{\sigma^2}{t^2}$$

6. Covariance and correlation

a) Let X and Y be random variables with $\mu_X = \mathrm{E}(X)$, $\sigma_X^2 = \mathrm{V}(X)$, $\mu_Y = \mathrm{E}(Y)$ and $\sigma_Y^2 = \mathrm{V}(Y)$. The covariance and correlation of X and Y is then defined as

$$\mathrm{Cov}(X,Y) = \mathrm{E}\big[(X-\mu_X)(Y-\mu_Y)\big]$$
$$\rho = \mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y}$$

b) $\text{Cov}(X, X) = \text{V}(X)$

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

d) X, Y independent $\Rightarrow \text{Cov}(X, Y) = 0$

e)

$$\text{Cov}\left(a + \sum_{i=1}^{n} b_i X_i\,,\; c + \sum_{j=1}^{m} d_j Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} b_i d_j \text{Cov}(X_i, Y_j)$$

f) $-1 \leq \text{Corr}(X, Y) \leq 1$ and $\text{Corr}(X, Y) = \pm 1$ if and only if there exists two numbers a, b such that $Y = a + bX$ (except, eventually, on areas of zero probability)

7. Moment generating functions

a) For a random variable X (discrete or continuous) the moment generating function is $M_X(t) = \text{E}(e^{tX})$

b) If the moment generating function $M_X(t)$ exists for t in an open interval containing 0, then it uniquely determines the distribution of X.

c) If the moment generating function $M_X(t)$ exists for t in an open interval containing 0, then all moments of X exist, and we can find the rth moment by $\text{E}(X^r) = M_X^{(r)}(0)$

d) $M_{a+bX}(t) = e^{at} M_X(bt)$

e) If X and Y are independent: $M_{X+Y}(t) = M_X(t) M_Y(t)$

8. Some discrete probability distributions

a) Binomial distribution:

Point probability: $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad k = 0, 1, \ldots, n$

Moment generating function: $M_X(t) = (1 - p + pe^t)^n$

Expectation: $\text{E}(X) = np$

Variance : $\text{V}(X) = np(1-p)$

Approximation 1: $Z = \dfrac{X - np}{\sqrt{np(1-p)}}$ is approximately normally distributed when np and $n(1-p)$ both are sufficiently big (at least 10)

Approximation 2: X is approximately Poisson distributed with parameter $\lambda = np$ when n is big and p is small

Sum rule: $X \sim$ binomial (n, p), $Y \sim$ binomial (m, p) and X, Y independent $\Rightarrow X + Y \sim$ binomial $(n + m, p)$

b) <u>Geometric distribution</u>:

Point probability: $P(X = k) = (1 - p)^{k-1} p \qquad k = 1, 2, \ldots$

Moment generating function: $M_X(t) = e^t p / [1 - (1 - p)e^t]$

Expectation: $\mathrm{E}(X) = 1/p$

Variance: $\mathrm{V}(x) = (1 - p)/p^2$

Sum rule: If X is geometrically distributed with probability p then $X - 1$ is negative binomial $(1, p)$. Then if X and Y are geometrically distributed with same p and independent then $X + Y - 2$ is negative binomial $(2, p)$

c) <u>Negative binomial distribution</u>:

Point probability: $P(X = k) = \binom{k+r-1}{r-1} p^r (1 - p)^k \qquad k = 0, 1, 2, \ldots$

Moment generating function: $M_X(t) = \{p/[1 - (1 - p)e^t]\}^r$

Expectation: $\mathrm{E}(X) = r(1 - p)/p$

Variance: $\mathrm{V}(X) = r(1 - p)/p^2$

Sum rule: $X \sim$ negative binomial (r_1, p),
$Y \sim$ negative binomial (r_2, p)
and X, Y independent
$\Rightarrow X + Y \sim$ negative binomial $(r_1 + r_2, p)$

d) <u>Hypergeometric distribution</u>:

Point probability: $P(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$

Expectation: $\mathrm{E}(X) = n \cdot \frac{M}{N}$

Variance: $\mathrm{V}(X) = n\frac{M}{N}(1 - \frac{M}{N})\frac{N-n}{N-1}$

Approximation: X is approximately binomial $(n, \frac{M}{N})$ when n is much smaller than N

e) <u>Poisson distribution</u>:

Point probability: $P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \qquad k = 0, 1, \ldots$

Moment generating function: $M_X(t) = e^{\lambda(e^t - 1)}$

Expectation: $\mathrm{E}(X) = \lambda$

Variance: $\quad \mathrm{V}(X) = \lambda$

Approximation: $\quad Z = \dfrac{X - \lambda}{\sqrt{\lambda}}$ is approximately normally distributed
when λ is sufficiently big (at least 10)

Sum rule: $\quad X \sim$Poisson (λ_1), $\quad Y \sim$Poisson (λ_2)
and X, Y independent $\Rightarrow X + Y \sim$ Poisson $(\lambda_1 + \lambda_2)$

e) Multinomial distribution:

Point probability: $P(N_1 = n_1, \ldots, N_r = n_r) = \frac{n!}{n_1! \cdots n_r!} p_1^{n_1} \cdots p_r^{n_r}$
Here $\sum_{i=1}^{r} p_i = 1$ and $\sum_{i=1}^{r} n_i = n$

Marginal distribution: $N_i \sim$ binomial(n, p_i)

9. Some continuous probability distributions

a) Normal distribution:

Density: $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} \qquad -\infty < x < \infty$

Moment generating function: $\quad M_X(t) = e^{\mu t} e^{\sigma^2 t^2/2}$

Expectation: $\mathrm{E}(X) = \mu$

Variance: $\mathrm{V}(X) = \sigma^2$

Transformation: $X \sim N(\mu, \sigma^2) \Rightarrow a + bX \sim N(a + b\mu, b^2\sigma^2)$
$X \sim N(\mu, \sigma^2) \Rightarrow Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$

Sum rule: $X \sim N(\mu_X, \sigma_X^2)$, $\quad Y \sim N(\mu_Y, \sigma_Y^2)$,
X, Y independent
$\Rightarrow X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$

b) Exponential distribution:

Density: $f(x) = \lambda e^{-\lambda x} \qquad x > 0$

Moment generating function: $\quad M_X(t) = \lambda/(\lambda - t)$ for $t < \lambda$

Expectation: $\mathrm{E}(X) = 1/\lambda$

Variance: $\mathrm{V}(X) = 1/\lambda^2$

Sum rule: $X \sim \exp(\lambda)$, $\quad Y \sim \exp(\lambda)$, $\quad X$ and Y independent
$\Rightarrow X + Y \sim$ gamma$(2, 1/\lambda)$

c) Gamma distribution:

Density: $f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} \qquad x > 0$

Gamma function: $\Gamma(\alpha) = \int_0^\infty u^{\alpha-1}e^{-u}du$
$\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$
$\Gamma(n) = (n-1)!$ when n is an integer
$\Gamma(1/2) = \sqrt{\pi}, \quad \Gamma(1) = 1$

Moment generating function: $M_X(t) = [1/(1-\beta t)]^\alpha$

Expectation: $\mathrm{E}(X) = \alpha\beta$

Variance: $\mathrm{V}(X) = \alpha\beta^2$

Sum rule: $X \sim \text{gamma}(\alpha,\beta), \quad Y \sim \text{gamma}(\delta,\beta)$,
X and Y independent $\Rightarrow X+Y \sim \text{gamma}(\alpha+\delta,\beta)$

d) Chi-squared distribution:

Density: $f(v) = \frac{1}{2^{n/2}\Gamma(n/2)} v^{(n/2)-1}e^{-v/2} \qquad v > 0$
n degrees of freedom

Expectation: $\mathrm{E}(V) = n$

Variance: $\mathrm{V}(V) = 2n$

Sum rule: $U \sim \chi^2_n$, $V \sim \chi^2_m$, U and V independent
$\Rightarrow U+V \sim \chi^2_{n+m}$

Result: $Z \sim N(0,1) \Rightarrow Z^2 \sim \chi^2_1$

e) Student's t-distribution:

Density: $f(t) = \frac{\Gamma[(n+1)/2]}{\sqrt{n\pi}\Gamma(n/2)}(1+\frac{t^2}{n})^{-(n+1)/2} \qquad -\infty < t < \infty$
n degrees of freedom

Expectation: $\mathrm{E}(T) = 0 \qquad (n \geq 2)$

Variance: $\mathrm{V}(T) = n/(n-2) \qquad (n \geq 3)$

Result: $Z \sim N(0,1), \quad U \sim \chi^2_n, \quad Z,U$ independent $\Rightarrow Z/\sqrt{U/n} \sim t_n$

f) Binormal distribution:

Density:

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right]\right\}$$

Marginal distribution: $X \sim N(\mu_X,\sigma_X^2), \quad Y \sim N(\mu_Y,\sigma_Y^2)$

Correlation: $\mathrm{Corr}(X,Y) = \rho$

Conditional distribution: Given $X=x$, Y is normally distributed with expectation $\mathrm{E}(Y|X=x) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)$ and variance $\mathrm{V}(Y|X=x) = \sigma_Y^2(1-\rho^2)$

10. One normally distributed sample

If $X_1, X_2, \ldots, X_n$ are independent and $N(\mu, \sigma^2)$ distributed then we have that:

a) $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ are independent

b) $\overline{X} \sim N(\mu, \sigma^2/n)$

c) $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$

d) $\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$

11. Two normally distributed samples

Let $X_1, X_2, \ldots, X_n$ be independent and $N(\mu_X, \sigma^2)$ distributed, and $Y_1, Y_2, \ldots, Y_m$ independent and $N(\mu_Y, \sigma^2)$ distributed. The two samples are independent of each other. Let $\overline{X}$, $\overline{Y}$, S_X^2 and S_Y^2 be defined as in 10a). Then we have that:

a) $S_p^2 = [(n-1)S_X^2 + (m-1)S_Y^2]/(m+n-2)$ is a weighted estimator for σ^2

b) $\overline{X} - \overline{Y} \sim N\left(\mu_X - \mu_Y, \sigma^2(\frac{1}{n} + \frac{1}{m})\right)$

c) $(n+m-2)S_p^2/\sigma^2 \sim \chi^2_{m+n-2}$

d) $\frac{\overline{X}-\overline{Y}-(\mu_X-\mu_Y)}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}} \sim t_{m+n-2}$

12. Regression analysis

Assume $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i\,;\ i = 1, 2, \ldots, n\,;$ where x_{ij}-s are given numbers and ϵ_i-s are independent and $N(0, \sigma^2)$ distributed. Then we have that:

a) The least squares estimators for β_0 and β_1 are

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1\overline{x} \quad \text{and} \quad \hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

b) The estimators in a) are normally distributed and unbiased, and

$$\text{Var}(\hat{\beta}_0) = \frac{\sigma^2 \sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n}(x_i - \overline{x})^2} \quad \text{and} \quad \text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

c) Let SSE$= \sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$. Then $S^2 = \text{SSE}/(n-2)$ is an unbiased estimator for σ^2, and $(n-2)S^2/\sigma^2 \sim \chi^2_{n-2}$

13. Multiple linear regression

Assume $Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \epsilon_i$; $i = 1, 2, \ldots, n$; where x_{ij}-s are given numbers and ϵ_i-s are independent and $N(0, \sigma^2)$ distributed. The model can be written in matrix form as $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta}$, where $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$ and $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_k)^T$ are n- and $(k+1)$-dimentional vectors, and $\mathbf{X} = \{x_{ij}\}$ (with $x_{i0} = 1$) is a $n \times (k+1)$-dimentional matrix. Then:

1. The least squares estimator for $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$.

2. Let $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \ldots, \hat{\beta}_k)^T$. Then $\hat{\beta}_j$-s are normally distributed and unbiased, and

$$\mathrm{Var}(\hat{\beta}_j) = \sigma^2 c_{jj} \quad \text{og} \quad \mathrm{Cov}(\hat{\beta}_j, \hat{\beta}_l) = \sigma^2 c_{jl}$$

   where c_{jl} is element (j, l) in the $(k+1) \times (k+1)$ matrix $\mathbf{C} = (\mathbf{X}^T\mathbf{X})^{-1}$.

3. Let $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_k x_{ik}$, og let $SSE = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$. Then $S^2 = \mathrm{SSE}/[n-(k+1)]$ is an unbiased estimator for σ^2, and $[n-(k+1)]S^2/\sigma^2 \sim \chi^2_{n-(k+1)}$. Also, S^2 and $\hat{\boldsymbol{\beta}}$ are independent.

4. Let $S^2_{\hat{\beta}_j}$ be the variance estimator for $\hat{\beta}_j$ we get by replacing σ^2 with S^2 in the formula for $\mathrm{Var}(\hat{\beta}_j)$ (in b). Then $(\hat{\beta}_j - \beta_j)/S_{\hat{\beta}_j} \sim t_{n-(k+1)}$.