LIST OF FORMULAS STK2120

(Version from 21. November 2014)

1 One-way analysis of variance

Assume $X_{ij} = \mu + \alpha_i + \epsilon_{ij}$; $j = 1, 2, \ldots, J_i$; $i = 1, 2, \ldots, I$; where ϵ_{ij}-s are independent and $N(0, \sigma^2)$ distributed. Then:

(*a*) The total sum of squares $SST = \sum_{i=1}^{I} \sum_{j=1}^{J_i} (X_{ij} - \bar{X}_{..})^2$ can be written as $SST = SSE + SSTr$ where

$SSE = \sum_{i=1}^{I} \sum_{j=1}^{J_i} (X_{ij} - \bar{X}_{i.})^2$ is the sum of squares for error or the sum of squares within groups

$SSTr = \sum_{i=1}^{I} J_i (\bar{X}_{i.} - \bar{X}_{..})^2$ is the sum of squares for treatment or the sum of squares between groups

(*b*) SSE and $SSTr$ are independent

(*c*) $MSE = SSE/[\sum_{i=1}^{I} (J_i - 1)]$ is an unbiased estimator for σ^2.
SSE/σ^2 is chi-squared distributed with $\sum_{i=1}^{I} (J_i - 1)$ degrees of freedom

(*d*) If all α_i-s are equal to zero, $SSTr/\sigma^2$ is chi-squared distributed with $I - 1$ degrees of freedom

(*e*) If $J_i = J$ for $i = 1, \ldots, I$, then

$\max_{i_1, i_2} |(\bar{X}_{i_1 \cdot} - \mu_{i_1}) - (\bar{X}_{i_2 \cdot} - \mu_{i_2})| / \sqrt{MSE/J}$

is distributed as the studentized range with parameters I and $I(J - 1)$.

2 Two-way analysis of variance

Assume $X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$; $k = 1, \ldots, K$; $j = 1, \ldots, J$; $i = 1, \ldots, I$; where ϵ_{ijk}-s are independent and $N(0, \sigma^2)$ distributed. Then:

(*a*) The total sum of squares $SST = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (X_{ijk} - \bar{X}_{...})^2$ can be written as $SST = SSA + SSB + SSAB + SSE$ where

$SSA = JK \sum_{i=1}^{I} (\bar{X}_{i..} - \bar{X}_{...})^2$

$SSB = IK \sum_{j=1}^{J} (\bar{X}_{.j.} - \bar{X}_{...})^2$

$SSAB = K \sum_{i=1}^{I} \sum_{j=1}^{J} (\bar{X}_{ij.} - \bar{X}_{i..} - \bar{X}_{.j.} + \bar{X}_{...})^2$

$SSE = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (X_{ijk} - \bar{X}_{ij.})^2$

(*b*) SSA, SSB, $SSAB$ and SSE are independent

(c) $MSE = SSE/IJ(K-1)$ is an unbiased estimator for σ^2.
SSE/σ^2 is chi-squared distributed with $IJ(K-1)$ degrees of freedom.

(d) If all α_i-s are equal to zero, SSA/σ^2 is chi-squared distributed with $I-1$ degrees of freedom

(e) If all β_j-s are equal to zero, SSB/σ^2 is chi-squared distributed with $J-1$ degrees of freedom

(f) If all γ_{ij}-s are equal to zero, $SSAB/\sigma^2$ is chi-squared distributed with $(I-1)(J-1)$ degrees of freedom

3 Block design (two-way anova without replications)

Assume $X_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$; $j = 1, \ldots, J$; $i = 1, \ldots, I$; where ϵ_{ij}-s are independent and $N(0, \sigma^2)$ distributed. Then:

(a) The total sum of squares $SST = \sum_{i=1}^{I} \sum_{j=1}^{J} (X_{ij} - \bar{X}_{..})^2$ can be written as $SST = SSA + SSB + SSE$ where

$SSA = J \sum_{i=1}^{I} (\bar{X}_{i\cdot} - \bar{X}_{..})^2$

$SSB = I \sum_{j=1}^{J} (\bar{X}_{\cdot j} - \bar{X}_{..})^2$

$SSE = \sum_{i=1}^{I} \sum_{j=1}^{J} (X_{ij} - \bar{X}_{i\cdot} - \bar{X}_{\cdot j} + \bar{X}_{..})^2$

(b) SSA, SSB and SSE are independent

(c) $MSE = SSE/[(I-1)(J-1)]$ is an unbiased estimator for σ^2.
SSE/σ^2 is chi-squared distributed with $(I-1)(J-1)$ degrees of freedom.

(d) If all α_i-s are equal to zero, SSA/σ^2 is chi-squared distributed with $I-1$ degrees of freedom

(e) If all β_j-s are equal to zero, SSB/σ^2 is chi-squared distributed with $J-1$ degrees of freedom

4 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:

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

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

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

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

(c) Let $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_k x_{ik}$, and let $SSE = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$. Then $S^2 = \text{SSE}/[n-(k+1)]$ 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.

(d) 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 $\text{Var}(\hat{\beta}_j)$ (in b). Then $(\hat{\beta}_j - \beta_j)/S_{\hat{\beta}_j} \sim t_{n-(k+1)}$.

5 Two-way tables and chi-square tests

(a) Assume $(N_1, \ldots, N_k)$ is multinomially distributed with probabilities p_i, where $\sum_{i=1}^{k} N_i = n$ and $\sum_{i=1}^{k} p_i = 1$.

If $p_i = \pi_i(\boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$, and $\hat{\boldsymbol{\theta}}$ is the maximum likelihood estimator for $\boldsymbol{\theta}$, then

$$\chi^2 = \sum_{i=1}^{k} \frac{(N_i - E_i)^2}{E_i}$$

is approximately chi-squared distributed with $k - 1 - m$ degrees of freedom when $E_i = n\pi_i(\hat{\boldsymbol{\theta}}) \geq 5$ for (almost) all i

(b) Test for homogeneity: Assume that for $i = 1, \ldots, I$, $(N_{i1}, \ldots, N_{iJ})$ are independent and multinomially distributed with probabilities p_{ij}, where $\sum_{j=1}^{J} p_{ij} = 1$.

If $p_{1j} = \cdots = p_{Ij}$, then

$$\chi^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(N_{ij} - E_{ij})^2}{E_{ij}}$$

is approximately chi-squared distributed with $(I-1)(J-1)$ degrees of freedom when
$E_{ij} = (N_{i\cdot} N_{\cdot j})/N_{\cdot\cdot} \geq 5$ for (almost) all i, j

(c) Test for independence: Assume $(N_{11}, \ldots, N_{1J}, N_{21}, \ldots, N_{2J}, \ldots, N_{I1}, \ldots, N_{IJ})$ is multinomially distributed with probabilities p_{ij}, where $\sum_{i=1}^{I} \sum_{j=1}^{J} p_{ij} = 1$.

If $p_{ij} = p_{i\cdot} p_{\cdot j}$ for all i, j, then

$$\chi^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(N_{ij} - E_{ij})^2}{E_{ij}}$$

is approximately chi-squared distributed with $(I-1)(J-1)$ degrees of freedom when
$E_{ij} = (N_{i\cdot}N_{\cdot j})/N_{\cdot\cdot} \geq 5$ for (almost) all i, j

6 The method of maximum likelihood

Assume $X_1, X_2, \ldots, X_n$ have a joint density function $f(x_1, x_2, \ldots, x_n \,|\, \boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$ is a parameter vector (scalar if $p = 1$). We assume that $f(x_1, x_2, \ldots, x_n \,|\, \boldsymbol{\theta})$ satisfies certain differentiability conditions.

(a) Given observed values $X_i = x_i$; $i = 1, \ldots, n$; the likelihood function is $L(\boldsymbol{\theta}) = f(x_1, x_2, \ldots, x_n \,|\, \boldsymbol{\theta})$ and the log-likelihood function is $l(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta})$.

(b) The maximum likelihood *estimate* is the value of $\boldsymbol{\theta}$ maximizing $L(\boldsymbol{\theta})$ or equivalently maximising $l(\boldsymbol{\theta})$. If we replace the observed x_i-s by the random variables X_i, we get the maximum likelihood *estimator*.

(c) The maximum likelihood estimate $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_p)$ is a solution of the equations $s_j(\boldsymbol{\theta}) = 0$; $j = 1, \ldots, p$; where $s_j(\boldsymbol{\theta}) = (\partial/\partial\theta_j) l(\boldsymbol{\theta})$ are the score functions. The vector of score functions is $\boldsymbol{s}(\boldsymbol{\theta}) = (s_1(\boldsymbol{\theta}), \ldots, s_p(\boldsymbol{\theta}))^T$.

(d) The observed information matrix $\bar{\boldsymbol{J}}(\boldsymbol{\theta})$ is the $p \times p$ matrix with element (i, j) given by $\bar{J}_{ij}(\boldsymbol{\theta}) = -\frac{\partial^2}{\partial\theta_i \partial\theta_j} l(\boldsymbol{\theta})$.

The expected information matrix (or Fisher information matrix) $\bar{\boldsymbol{I}}(\boldsymbol{\theta})$ is the $p \times p$ matrix with element (i, j) given by $\bar{I}_{ij}(\boldsymbol{\theta}) = \mathrm{E}[\bar{J}_{ij}(\boldsymbol{\theta})]$.

For independent and identically distributed observations we have that $\bar{\boldsymbol{I}}(\boldsymbol{\theta}) = n\boldsymbol{I}(\boldsymbol{\theta})$, where $\boldsymbol{I}(\boldsymbol{\theta})$ is the expected information for one observation.

(e) When the equations in (c) do not have an explicit solution, we can find the maximum likelihood estimate by using the Newton-Raphson method:

$$\boldsymbol{\theta}^{(s+1)} = \boldsymbol{\theta}^{(s)} + \bar{\boldsymbol{J}}^{-1}(\boldsymbol{\theta}^{(s)})\boldsymbol{s}(\boldsymbol{\theta}^{(s)}),$$

by using the Fisher scoring algorithm:

$$\boldsymbol{\theta}^{(s+1)} = \boldsymbol{\theta}^{(s)} + \bar{\boldsymbol{I}}^{-1}(\boldsymbol{\theta}^{(s)})\boldsymbol{s}(\boldsymbol{\theta}^{(s)}),$$

or by suitable modifications of these.

(f) When we have "enough" data, $\hat{\theta}_i$ is approximately normally distributed with expectation θ_i and with variance equal to the i-th diagonal element in $\bar{\boldsymbol{I}}^{-1}(\boldsymbol{\theta})$. The covariance between $\hat{\theta}_i$ and $\hat{\theta}_j$ is approximately equal to element (i, j) in $\bar{\boldsymbol{I}}^{-1}(\boldsymbol{\theta})$. We can estimate variances/covariances by plugging in $\hat{\boldsymbol{\theta}}$ instead of $\boldsymbol{\theta}$ in $\bar{\boldsymbol{I}}^{-1}(\boldsymbol{\theta})$ or in $\bar{\boldsymbol{J}}^{-1}(\boldsymbol{\theta})$.

7 Bootstrapping

Assume that the distribution of the data $\boldsymbol{X}$ is described by a distribution function F. Let $\theta = \theta(F)$ be a functional of F estimated by $\hat{\theta} = \hat{\theta}(\boldsymbol{X})$.

(*a*) The bootstrapping idea is to approximate the properties of $\hat{\theta}$ by assuming that an estimate $\widehat{F}$ for F is the true distribution function.

(*b*) Bootstrap estimation of the skewness of $\hat{\theta}$:

$$b_{\hat{\theta}} = \frac{1}{B} \sum_{b=1}^{B} \theta_b^* - \theta(\widehat{F})$$

(*c*) Bootstrap estimation of the standard deviation of $\hat{\theta}$:

$$\sqrt{\mathsf{E}^{\widehat{F}} \left\{ \left(\hat{\theta}(\boldsymbol{X}^*) - \mathsf{E}^{\widehat{F}}[\hat{\theta}(\boldsymbol{X}^*)] \right)^2 \right\}}$$

(*d*) Standard bootstrap confidence interval:

$$(\hat{\theta} - \overline{\delta}, \hat{\theta} - \underline{\delta})$$

where $\underline{\delta}$ and $\overline{\delta}$ are lower and upper $\alpha/2$ quantiles in the bootstrap distribution of $\Delta = \hat{\theta} - \theta$.