UNIVERSITY COLLEGE LONDON

Computational Modelling in Biomedical Imaging - Coursework 2

Author:
RĂZVAN VALENTIN MARINESCU
razvan.marinescu.14@ucl.ac.uk

EPSRC CENTRE FOR DOCTORAL TRAINING IN MEDICAL IMAGING
UNIVERSITY COLLEGE LONDON

March 8, 2015

Part 1 - Q1

(a)

Sample	True mean	Sample mean	True standard deviation	Sample standard deviation
sample 1	1	0.9848	0.25	0.2842
sample 2	1.5	1.4944	0.25	0.2281

The new values are as expected, lying within a 0.04 tolerance level.

(b)

The t-test results in a t-statistic of -6.9923, rejecting the null hypothesis with a p-value of 7.55e-09 which is very low. We are very confident the two samples were generated from distributions with different means, which is indeed the case (means of 1 and 1.5).

(c)

i. $X = \begin{pmatrix} 1 & \dots & 1 & 0 & \dots & 0 \\ 0 & \dots & 0 & 1 & \dots & 1 \end{pmatrix}^T$ and $dim(X) = 2$ because X is made of two column vectors that are linearly independent.

ii. $Y = X\beta \rightarrow X^TY = X^TX\beta \rightarrow (X^TX)^{-1}X^TY = \beta \rightarrow X(X^TX)^{-1}X^TY = X\beta$. Since $MY = X\beta$, we deduce $M = X(X^TX)^{-1}X^T$. In our case $M = 0.04 * \begin{pmatrix} \mathbb{I}_N & 0 \\ 0 & \mathbb{I}_N \end{pmatrix}$ where $\mathbb{I}_N$ is an $N \times N$ matrix full of 1s (not to be confused with I_n).

iii. $\hat{Y} = MY = [0.9848 \dots 0.9848 \; 1.4944 \dots 1.4944]^T$ and $\hat{e} = (I - M)Y$. The cosine between the vectors is almost zero (-4.1352e-16) which means that the vectors are perpendicular. This is what we expected, since the column space and error space are orthogonal.

iv. We know that $Y = X\beta$. Pre-multiplying by X^T gives $X^TY = X^TX\beta$. Now matrix X^TX is a square matrix and can normally be inverted. We therefore further pre-multiply by $(X^TX)^{-1}$ and get $(X^TX)^{-1}X^TY = \beta$, the model parameters. For our problem we get the estimated parameters $\beta = [0.9848, 1.4944]$, which are very close to the ground truth values of $[1, 1.5]$

v.

$$\hat{\sigma}^2 = \frac{\hat{e}'\hat{e}}{n - dim(X)} = \frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{n - dim(X)}$$

This is called the *Mean Squared Error* because it is an unbiased estimate of the error in the prediction. The upper term $(Y - X\hat{\beta})'(Y - X\hat{\beta})$ is the sum of squared differenced between the observations and the predicted values. This is then divided by the number of degrees of freedom, which is $N - 2$ in our case. For our error vector we get $\hat{\sigma}^2 = 0.0664$

vi. $S_{\hat{\beta}} = \begin{pmatrix} 0.0027 & 0 \\ 0 & 0.0027 \end{pmatrix}$. From $S_{\hat{\beta}}$ we can calculate $std(\beta_1) = \sqrt{Sb(1,1)} = 0.0515$. Similarly we get $std(\beta_2) = \sqrt{S_{\hat{\beta}}(2,2)} = 0.0515$

vii. $C = [1, -1]^T \rightarrow C'\beta = \beta_1 - \beta_2 = 0$. By setting $\beta_1 = \beta_2 = \gamma$ we then solve $Y = X_0\gamma = XU\gamma$ so it must be the case that $\beta = U\gamma$. Solving this for our conditions gives $U = [1, 1]^T$. Then the reduced model $X_0 = XU = [1 \dots 1]^T$ where the length of X_0 is 50.

viii. If $\hat{Y}$ and $\hat{Y_c}$ are the predicted values using the full and constrained models respectively, then $\hat{Y}_{\perp c} = \hat{Y} - \hat{Y_c}$ is the additional error. For our data we get $|\hat{Y_{\perp c}}| = 1.8017$. Using the MSE we got from (v), we get

$$F = \frac{|\hat{Y}_{\perp c}|^2/r}{|\hat{e}|^2/(N-p)} = 48.89$$

Since $F > 3.68$, we reject H_0 at $a = 0.05$ confidence level (i.e. the two samples come from different groups).

ix. The t-statistic is -6.9923 and is exactly the same as the one calculated in point b).

x. The model parameters represent the *estimated values* for the means of the distributions that we used to sample the data from. In our case we get $\hat{\beta} = [0.9848, 1.4944]$ which are close to the true values of $\beta = [1, 1.5]$.

xi. $e_c = Me = [-0.0152 \cdots -0.0152 \quad -0.0056 \cdots -0.0056]$ where elements from each part of the array are exactly the differece between $\hat{\beta} - \beta = [-0.0152 \quad -0.0056]$. e_c represents the error component that doesn't come from noise, but from the wrong estimates of $\hat{\beta}$ that are not equal to the true β.

xii. $e_e = (I-M)e = (I-M)(Y - X\beta) = (I-M)Y - (I-M)X\beta = \hat{e} - 0 = \hat{e}$ because $(I-M)$ and $X\beta$ are orthogonal. In our example we indeed get that $e_e = \hat{e}$.

(d)

i. $X = \begin{pmatrix} 1 & \dots & 1 & 1 & \dots & 1 \\ 1 & \dots & 1 & 0 & \dots & 0 \\ 0 & \dots & 0 & 1 & \dots & 1 \end{pmatrix}^T$ and $dim(X) = 2$ because vectors in X are linearly dependent $X_1 = X_2 + X_3$, so at most two vectors are linearly independent.

ii. $C = [0, 1, -1]^T \rightarrow \beta_1 = \beta_2 \rightarrow U = null(C^T) = \begin{pmatrix} -0.707 & 0.707 \\ 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix} \rightarrow X_0 = XU = \begin{pmatrix} -0.207 & 1.207 \\ \vdots & \vdots \\ -0.207 & 1.207 \end{pmatrix}$

iii. $t = -6.9923$

iv. β_0 models the constant factor for both groups, β_1 models the first group and β_2 models the second group. However, by removing one of the parameters we still get the same fit, so one of them is redundant ($dim(X) = 2$).

(e)

i. $X = \begin{pmatrix} 1 & \dots & 1 & 1 & \dots & 1 \\ 1 & \dots & 1 & 0 & \dots & 0 \end{pmatrix}^T$ and $dim(X) = 2$ because X has two vectors that are linearly independent.

ii. $C = [01]^T \rightarrow \beta_0 = 0 \rightarrow U = null(C^T) = [-1, 0]^T \rightarrow X_0 = XU = \begin{pmatrix} -1 \\ \vdots \\ -1 \end{pmatrix}$

iii. $t = -6.9923$

iv. β_0 models a constant term for both groups while β_1 models only the first group. We should have that $\beta_0 + \beta_1 \approx \mu_0 = 1$ and $\beta_0 \approx \mu_1 = 1.5$, which is the case in our solution $\hat{\beta} = [1.49 - 0.5]$.

(f)

We cannot test that the samples have different means with the model $Y = X_0\beta_0$ because this model already assumes that all samples have the same mean. In order to test that the two groups have different means one needs a more complex model.

Part 1 - Q2

(a)

i. $t = -6.9918$, which is different from the t-statistic obtained in question 1. Although the difference is quite small, for a different random number seed, a bigger Δt can be observed (see `p12testT.m`).

(b)

i. $X = \begin{pmatrix} 1 & \dots & 1 & 1 & \dots & 1 \\ 1 & \dots & 1 & 0 & \dots & 0 \\ & I_n & & & I_n & \end{pmatrix}^T$ where I_n is the identity matrix of dimensions $n \times n$.

ii. $C = [0, 1, 0, \dots, 0]^T \rightarrow$ the reduced model $X_0 = \begin{pmatrix} 1 & \dots & 1 & 1 & \dots & 1 \\ & I_n & & & I_n & \end{pmatrix}^T$

iii. $t = -6.9918$, which is the same answer as in part (a).

Part 2 - Q1

(a)

$t = 5.2295$ and $p = 2.114e - 04$

(b)

$p = 3.33e - 04 = 1/3003$ which is close to the original p-value of $2.11e - 04$. However, a better resolution cannot be obtained because the number of permutations is 3003 and with our data only one t-statistic was bigger than the original t-value.

(c)

$p = 3.33e - 04 = 1/3003$ which is the same value as in part (b).

(d)

i.

ii.

iii.

Part 2 - Q2

(a)

(b)

(c)

(d)