Antonio Remiro Azócar - Computational Modelling for Biomedical Imaging: Coursework 2.1

Q1

(a) The random seed is set to three for result reproducibility. The mean ($\pm$ st.dev) of Sample 1 is 0.966 ± 0.169 and that of Sample 2 is 1.503 ± 0.200. These values are as expected given that the noise generation is zero-mean.

(b) A two-sample t-test is performed on the generated samples via the MATLAB command `ttest2(group1, group2)`. This command returns `hyp` $= 1$, corresponding to rejecting the null hypothesis (the means of both samples are equal) at the 5% significance level. We also obtain a structure with fields, `stats`, containing the t-statistic value `tstat = -10.26`. These values are appropriate; the large magnitude of `tstat` in comparison to its table value (for its corresponding degrees of freedom) corresponds to the sample means being identified as different.

(ci) The design matrix has two columns (two groups) and 50 rows (data points). The first element of the first 25 rows is one; the second element is set to zero. The second element of the last 25 rows is one; the first element is zero. There are two pivot columns; hence the dimension of the matrix's column space (rank) $C(X)$ is dim$(X) = 2$.

(cii) We have a linear model, $Y = X\beta + e$. For uncorrelated errors and identical weights for each data point, regression parameters are:

$$\hat{\beta} = (X^TX)^{-1}X^TY.$$

This arises since for linear regression, $SSE = Y^TY - 2Y^TX\beta + \beta X^TX\beta$. This implies $\frac{\partial SSE}{\partial \beta} = -2X^TY + 2X^TX\beta$; setting this to zero (to minimise SSE) and rearranging gives the expression above. Since the fitted values are now,

$$\hat{Y} = X\hat{\beta} = X(X^TX)^{-1}X^TY,$$

we can establish a projection operator,

$$P_X = X(X^TX)^{-1}X^T,$$

corresponding to $C(X)$. P_X is a perpendicular projection operator satisfying the properties: **(i) Idempotence:** ($MM = M$ for a given matrix M). For P_X, we have:

$$P_XP_X = X(X^TX)^{-1}X^TX(X^TX)^{-1}X^T.$$

P_X is idempotent since $(X^TX)^{-1}(X^TX) = I$; therefore $P_XP_X = X(X^TX)^{-1}X^T = P_X$. **(ii) Symmetry:** ($M = M^T$ for a given matrix M). For P_X, we have:

$$\begin{aligned} P_X^T = (X(X^TX)^{-1}X^T)^T = (X^T)^T((X^TX)^{-1})^TX^T = \\ X((X^TX)^T)^{-1}X^T = X((X^T(X^T)^T)^{-1}X^T = \\ X((X^TX)^{-1}X^T = P_X. \end{aligned}$$

Therefore $P_X = P_X^T$ and P_X is symmetric. The P_X for $C(X)$ in the question is a 50×50 matrix. The first 25 elements of the first 25 rows are 0.04; the last 25 elements are 0. The last 25 elements of the last 25 rows are 0.04; the first 25 elements are 0. The trace of P_X is 2; therefore for $P_X = X(X^TX)^{-1}X^T$, tr(P_X) =rank(X) (the dimension of the column space of X). This means the trace of the projection matrix is the dimension of the target space.

(ciii) We can use P_X to determine $\hat{Y}$ (the projection of Y into $C(X)$), via:

$$\hat{Y} = P_XY = X(X^TX)^{-1}X^TY.$$

$\hat{Y}$ represents the fitted values for each group (the least squares estimators for our model). For this exercise, $\hat{Y}$ is a column vector of 50 elements whose first 25 elements are 0.996 (Sample 1 mean) and whose last 25 elements are 1.503 (Sample 2 mean).

The assumptions of the linear model imply $E(Y) = X\beta$. β is unknown, thereby $E(Y)$ is found in the column space of X, $C(X)$. Hence, $C(X)$ is known as the estimation space. Least squares estimate $\hat{\beta}$ finds the vector in such space closest to Y. Since $C(X)$ is a subspace of $\mathbb{R}^n$, such vector projects Y orthogonally into $C(X)$.

(civ) $R_X = I - P_X$ is also a perpendicular projection operator. The necessary and sufficient conditions for a square matrix to be a perpendicular projector are idempotence and symmetry. We prove that R_X has both properties. **(i) Idempotence**. For R_X we have:

$$\begin{aligned} R_XR_X = (I - P_X)(I - P_X) = II - 2IP_X + P_XP_X = \\ I - 2P_X + P_X = I - P_X = R_X. \end{aligned}$$

(ii) Symmetry. We have:

$$R_X^T = (I - P_X)^T = I^T - P_X^T = I - P_X,$$

since we proved before P_X is symmetric ($P_X^T = P_X$). Therefore, $R_X^T = I - P_X = R_X$ and R_X is symmetric. R_X is computed and stored in the code as `Rx`. As expected, it is idempotent and symmetric. We also have $\hat{e} = Y - P_XY = (I - P_X)Y = R_XY$, so R_X projects Y orthogonally into the error space.

(cv) $\hat{e}$, the projection of Y into the error space $C(X)^{\perp}$, is given by:

$$\hat{e} = Y - \hat{Y} = (I - P_X)Y = R_XY.$$

$\hat{e}$ is computed and stored as `ehat`. `ehat` is a column vector of 50 elements. $C(X)^{\perp}$ has therefore a dimension of one.

(cvi) The angle θ between $\hat{e}$ and $\hat{Y}$ is calculated via:

$$\theta = \arccos(\hat{e} \cdot \hat{Y}).$$

We have $\theta = \frac{\pi}{2}$ radians. Hence the projection of Y into the error space $C(X)^{\perp}$ is perpendicular to its projection into $C(X)$. This result is expected; the residuals are orthogonal to X's column space since $P_X = X$ so that $\hat{e}^TX = ((I - P_X)Y)^TX = Y^T(I - P_X)X = Y^T(X - P_X) = 0$.

(cvii) We have a linear model:

$$Y = X\beta + e.$$

Running the regression is equivalent to minimising the sum of square errors (SSE) of the prediction. We are trying to minimise:

$$SSE = e^Te,$$

where $e = Y - X\beta$. Therefore,

$$\begin{aligned} SSE = (Y - X\beta)^T(Y - X\beta) = (Y^T - X^T\beta^T)(Y - X\beta) \\ = Y^TY - Y^TX\beta - \beta^TX^TY + \beta^TX^TX\beta = \\ Y^TY - 2Y^TX\beta + \beta X^TX\beta, \end{aligned}$$

where we have used $Y^TX\beta = \beta^TX^TY$. We next take the partial derivative of the SSE wrt β and set it to zero:

$$\frac{\partial SSE}{\partial \beta} = -2X^TY + 2X^T\hat{\beta} = 0,$$

where $\hat{\beta}$ is the minimising parameter. The previous step explains why we are performing a 'least-squares' estimate. By setting the partial derivative to zero we are finding an expression for the matrix of estimates which minimises the predictive SSE. We rearrange and simplify the above expression to,

$$X^T X\hat{\beta} = X^T Y.$$

Solving for $\hat{\beta}$ yields the model parameters,

$$\hat{\beta} = (X^T X)^{-1} X^T Y.$$

The derived formula is utilised to determine $\hat{\beta}$ for the exercise. The variable is stored as `betahat`. We obtain $\hat{\beta}_1 = 0.966$ and $\hat{\beta}_2 = 1.503$, the group means.

(cviii) In the case, $n = 50$, $\dim(X) = 2$. We compute $\hat{\sigma}^2 = 0.034$; this value is stored as `var`. We can denote the number of parameters estimated in **1cvii** as $p = 2$. These are β_1, β_2; note p is equivalent to the rank of X, here $\dim(X)$. We observe that the denominator is the sample size reduced by the number of parameters already estimated. From **1cvii**, we have the numerator $\hat{e}^T\hat{e} = SSE$. Therefore, $\hat{\sigma}^2 = \frac{SSE}{n-p}$ is the mean square error over the system's degrees of freedom (we divide over $n-p$ instead of n to remove the bias in the variance of the unobserved errors).

(cix) We compute,

$$S_{\hat{\beta}} = \begin{pmatrix} 0.0014 & 0 \\ 0 & 0.0014 \end{pmatrix}.$$

The covariance matrix has only non-zero diagonal terms. Therefore, the model parameters are independent from each other. The standard deviation of the model parameters is given by the square root of the diagonal entries, which is 0.037.

(cx) A vector is called a contrast vector if its elements sum to zero; it is orthogonal to the vector of ones. We derive the contrast vector $\lambda = [1 - 1]^T$ to compute differences in the means of both groups. The constraint $\lambda^T\beta = \lambda_1\beta_1 + \lambda_2\beta_2 = 0$, gives a new β (lets denote this β_0), with $\beta_1 - \beta_2 = 0 \to \beta_1 = \beta_2 = \beta_0$. Substituting this into our GLM yields $Y = (X_1 + X_2)\beta_0 + e = X_0\beta_0 + e$, with $X_0 = X_1 + X_2$ i.e. X_0 is a column vector of 50 ones.

(cxi) We recalculate $\hat{\beta}_0$ via:

$$\hat{\beta}_0 = (X_0^T X_0)^{-1} X_0^T Y = 1.235.$$

Note this is the mean over all observations. The new error is given by $\hat{e}_0 = Y - X_0\hat{\beta}_0$. This value is subtracted from that computed in **1cv** and the resulting 'additional error' is 0.268. the exercise,

$$\nu_1 = \text{tr}(P_X - P_{X_0}) = 1,$$

$$\nu_2 = \text{tr}(I - P_X) = 48.$$

We compute $SSR(X_0) = \hat{e}_0^T\hat{e}_0$, $SSR(X) = \hat{e}^T\hat{e}$, and obtain $F_{\nu_1,\nu_2} = 105.3$. Note we have a reduced model (X_0) with p_1 parameters nested within a full model (X) with p_2 parameters, $p_2 > p_1$. We have degrees of freedom $v_1 = p_2 - p_1 = 2 - 1 = 1$, $v_2 = n - p_2 = 50 - 2 = 48$ for the F-statistic in question.

(cxii) We obtain $t = -10.26$. This value is stored as `tstat` and is identical as that in **1cii**.

(cxiii) Parameters $\hat{\beta}_1$, $\hat{\beta}_2$ represent the means of group 1 and group 2, respectively. Their ground truth (g.t.) values should be $\hat{\beta}_1^{(GT)} = 1$, $\hat{\beta}_2^{(GT)} = 1.5$.

(cxiv) The projection of the g.t. deviation e into $C(X)$ is computed via, $e = y - X\hat{\beta}^{(GT)}$. e is a 50×1 vector. It represents the difference between $\hat{\beta}^{(GT)}$ and $\hat{\beta}$. This is because $P_x e = \hat{y} - X\hat{\beta}^{(GT)} = X\hat{\beta} - X\hat{\beta}^{(GL)} = X(\hat{\beta} - \hat{\beta}^{(GL)})$.

(cxv) The projection of e into $C(X)^\perp$ is computed via $\hat{y} - X\hat{\beta}^{(GT)} = P_X e$,(since $P_X y = \hat{y}$ and $P_x X\hat{\beta}^{(GT)} = X\hat{\beta}^{(GT)}$). $\hat{e}$ represents $\hat{e} = R_X e = (I - P_X)e$, the orthogonal projection of the ground truth deviation into error space.

(di) The design matrix has three columns and 50 rows. All entries are one in the first column. The second and third columns are identical to the design matrix in **1ci**. We still have two pivot columns; since the first is a linear combination of the the last two (therefore linearly dependent). Hence the dimension of the matrix's column space (rank) $C(X)$ is dim(X) $= 2$.

(dii) We utilise the MATLAB operator `pinv` (instead of `inv`) to compute the pseudoinverse of $X^T X$. The P_X for $C(X)$ in the question is the same as that in **1cii**. The estimation space (col.space of X) is the same as that in **c**.

(diii) The appropriate contrast vector is $\lambda = [0 \ \ 1 \ \ -1]^T$. We have $\lambda^T\beta = \lambda_0\beta_0 + \lambda_1\beta_1 + \lambda_2\beta_2 = 0$, with $\beta_1 - \beta_2 = 0 \to \beta_1 = \beta_2$. Substituting this into our GLM yields $X_0\beta_0 + (X_1 + X_2)\beta_1 + e = X_0\beta_0 + X_3\beta_1 + e$, with $X_3 = X_1 + X_2$ i.e. X_3 is a column vector of 50 ones. X_0 is also a column vector of 50 ones. Hence we have $X_0 = X_3 \to Y = X_0(\beta_0 + \beta_1) + e = X_0\beta^* + e$ for a given parameter $\beta^* = \beta_0 + \beta_1$. We have obtained the same reduced model as for **1cx**.

(div) We obtain $t = -10.26$, the same t as in **b**, **cxii**.

(dv) Parameter $\hat{\beta}_0$ represents a bias term applied to all values. $\hat{\beta}_1$, $\hat{\beta}_2$ are the differences between the means of the first and second group and $\hat{\beta}_0$.

(ei) The design matrix has two columns and 50 rows. All entries are one in the first column. The first 25 elements of the second column are ones; the rest are zero. We have two pivot columns. The dimension of the matrix's column space (rank) $C(X)$ is dim(X) $= 2$.

(eii) The appropriate contrast vector is $\lambda = [0 \ \ 1]^T$. We have $\lambda^T\beta = \lambda_0\beta_0 + \lambda_1\beta_1 = 0$. Substituting in $\lambda_0 = 0$, $\lambda_1 = 1$ yields $\beta_1 = 0$. Hence we have $Y = X_0\beta_0 + e$; the same reduced model as for **1cx**, **1diii**.

(eiii) We have $t = -10.26$, the same t as in **b**, **cxii**, **div**.

(eiv) Parameter $\hat{\beta}_0$ represents the mean of the second sample set. $\hat{\beta}_1$ is the difference between the means of both sample sets.

(f) We cannot test the hypothesis with such model because with only X_0 in our design matrix we have no information about any sample mean.

Q2

(a) The MATLAB command `ttest(group1, group2)` returns `hyp` $= 1$, corresponding to rejecting the null hypothesis (the means of both samples are equal) at the 5% significance level. We obtain `tstat = -10.29`, a very similar value for the result in **Q1**.

(bi) The design matrix has 50 rows and 27 columns. All entries are one in the first column. The first 25 elements of the second column are one; the rest are zero. Columns 3 to 27 have ones in the entries of their corresponding subjects (for example column 3 has ones in its first and 26th elements). We have 26 pivot columns; since the first is a linear combination of the the last 25 (therefore linearly dependent). Hence the dimension of the matrix's column space (rank) $C(X)$ is $\dim(X) = 26$.

(bii) The appropriate contrast vector is $\lambda = [0\ \ 1\ \ 0\ \ 0\ \ 0\ \ ...]^T$ with the last 25 entries as zero.
(biii) We obtain $t - 10.29$, an identical t as in **Q1**. Note this time we divide by ($n - \dim(X) = 24$) to obtain $\hat{\sigma}^2$.