

---

# COMP0118 - Assignment 2

Student ID: 22197823 - 10th March 2023

## 1 Part 1

*Note: Code for all of Part 1 is in the code pdf pages 2-10*

### Q1.a)

I generate two synthetic groups of data as asked in the question. The generated data can be seen in *Figure 1*. Because the noise is being sampled from  $\mathcal{N}(0, 0.2)$  I would expect the sum of square difference from the mean to be  $N\sigma^2 = 20 \times 0.2^2 = 0.8$ . For group 1 and 2 I have SSD of 0.70 and 0.86, and means of 1.5 and 1.94 respectively. This is as expected.

### Q1.b)

I use the `scipy.stats.ttest_ind()` method in Python to calculate the t-statistic for 2 independent samples, each sample having identical variances. The data meets these requirements. I calculate a t-statistic of  $-7.04$

### Q1.c)

i) I construct the  $Y$  vector of length 40 by concatenating group 1 with group 2 data. I now can construct  $X_i$ , this is the dummy variable column for membership to group  $i$ . The design matrix  $X$  can be seen in *Figure 2*.  $\text{Rank}(X)$  is equal to  $\text{Dim}(X)$ , it can clearly be seen that  $X_1$  and  $X_2$  are linearly independent and therefore  $\text{Dim}(X) = C(X) = 2$

ii) To derive  $P_X$  I first look at the solution space  $C(X)$  which is spanned by the columns of the design matrix. There is a simple diagram showing this in *Figure 3*. One can see  $Y - \hat{Y} = Y - X\beta$  minimises the distance between the solution  $Y$  and the solution space  $C(X)$ . It is therefore orthogonal to  $C(X)$  which implies  $X^T(Y - X\beta) = 0$ . Following this derivation through in *Equation 1* I get  $P_X = X(X^T X)^{-1} X^T$ . Calculating  $P_X$  for our data I get the solution shown in *Figure 4*. From this I can see that  $P_X P_X = P_X$  and  $P_X$  is clearly symmetric, therefore  $P_X$  satisfies the key properties of being a perpendicular projection matrix.

$\text{Trace}(P_X) = 2$ , this is the sum of eigenvalues of  $P_X$ . Because I have  $P_X P_X = P_X$ , I can deduce all the eigenvalues of  $P_X$  must equal to 1 or 0, therefore  $\text{Trace}(P_X)$  is the number of non-zero eigenvalues  $P_X$  has which is equivalent to  $\text{Dim}(P_X)$ . Because  $P_X$  projects onto the solution space  $C(X)$  I therefore have  $\text{Trace}(P_X) = \text{Dim}(C(X))$ .

iii) Using  $P_X$  to get  $\hat{Y}$  can be seen in *Figure 5*.  $\hat{Y}$  is the projection of  $Y$  into the solution space  $C(X)$  with the minimum absolute error size ( $|\hat{e}| = |Y - \hat{Y}|$ ).  $C(X)$  is the estimation space because it is subspace of all possible estimations of  $\hat{Y} = X\beta$ , and is spanned by the column vectors of the design matrix  $X$ .

iv)  $R_X = (I - P_X)$  can be seen in *Figure 6*. In *Equation 2* I show  $R_X$  satisfies the perpendicular projection mapping criteria.

v) To estimate the error I use  $\hat{e} = R_X Y = (I - P_X)Y = Y - \hat{Y}$ . A printout of  $\hat{e}$  can be seen in *Equation 3*.  $R_X$  is a projection mapping, therefore I can use the same logic as Q1.c.iii) to get  $\text{Trace}(R_X) = \text{Dim}(R_X)$ . Also  $R_X$  is projecting into the error space which is  $C(X)^\perp$ , therefore  $\text{Trace}(R_X) = \text{Dim}(C(X)^\perp) = 38$

vi) Using *Equation 4* I calculate the angle between  $\hat{e}$  and  $\hat{Y}$  to be  $\frac{\pi}{2}$ . This shows  $\hat{Y}$  and  $\hat{e}$  are perpendicular which is as expected. This can be seen by referring to the *Figure 3*.  $\hat{Y}$  is a perpendicular projection of  $Y$  onto the solution space, so by definition  $\hat{e} = Y - \hat{Y}$  is perpendicular.

vii) For a GLM:  $Y = X\beta + e$ , where  $e$  is the error, I can estimate the parameter  $\beta$  with  $\hat{\beta}$  using *Equation 5*. This is called the Least Squares estimate because the first line in *Equation 5*,  $\hat{e}^T(X\alpha) = 0$  for all  $\alpha \in \mathbb{R}$  is equivalent to finding a solution that is minimising the sum of the size of the squared error terms. Using this formula I get  $\beta = [1.4978, 1.9380]$ .

viii) Calculating the variance of  $\hat{e}$  I get  $\hat{\sigma}^2 = 0.03906$ . This is called the MSE because the numerator  $\hat{e} \cdot \hat{e} = \sum_i \hat{e}_i^2$  is the sum of the squared error, and the denominator is the degrees of freedom of the error space, hence I get the Mean Squared Error.

ix) My calculation of  $S_\beta$  can be seen in *Equation 6*. The diagonals give the variance of the parameters, therefore I have the standard deviation of both  $\beta_1$  and  $\beta_2$  to be 0.0442

x) Comparing the means I can create the null hypothesis  $H_0$ : The mean of group 1 and group 2 are identical. Therefore we have  $\beta_1 = \beta_2 \implies \beta_1 - \beta_2 = 0$ , which gives the contrast vector of  $\lambda = [1, -1]^T$ . This

---

gives the reduced model  $X_0 = [1, \dots, 1]^T \in \mathbb{R}^{40}$ .

**xi)** Comparing the full model to the reduced model I get an F-Statistic of 38, with (1, 38) degrees of freedom (using *Equation 7*. This gives a p-value of  $3.388 \times 10^{-7}$ , giving strong evidence that the means of both groups are not the same.

**xii)** Calculating the t-statistic I get  $t = -7.04$ . This is the same value I got in Q1.b) which is expected.

**xiii)**  $\hat{\beta}_1$  and  $\hat{\beta}_2$  are the means of the sample populations for group 1 and group 2 respectively. This is because  $\hat{\beta}$  is estimated by minimising the squared error between  $Y$  and  $\hat{Y}$ , and this is given when  $\hat{\beta}_1$  and  $\hat{\beta}_2$  are the group means. The ground truth values for  $\hat{\beta}$  are the ground truth means of the sample populations which is 1.5 and 2.0 for group 1 and 2 respectively.

**xiv)** Ground truth deviation error  $e = Y_{gt} - Y$  where  $Y_{gt}$  is the ground truth of the sample population. To project  $e$  into  $C(X)$  I use the  $P_X$  projection mapping. The solution of this can be seen in *Figure 7*. The first 20 elements are equal to  $\beta_1 - \hat{\beta}_1$ , and the last 20 elements are equal to  $\beta_2 - \hat{\beta}_2$ .

**xv)** To project  $e$  into  $C(X)^\perp$  we use  $R_X$ . The solution can be seen in *Equation 8*. This is identical to  $\hat{e}$ .  
**Q1.d)**

**i)** The design matrix for the GLM with an intercept is shown in *Figure 8*.  $DIM(X) = 2$  because  $X_0 = X_1 + X_2$  so the first column is not linearly independent of the remaining 2 columns.

**ii)** Calculating  $P_X$  for this GLM the same way as before we get the exact same projection mapping as we did in Q1.c.ii, seen in *Figure 4*. This is because the solution space  $C(X)$  is still spanned by column vectors  $X_1$  and  $X_2$ , therefore the projection mapping is projecting into the same space Q1.c.ii)

**iii)** Having a Null Hypothesis  $H_0$  of both groups having the same mean, will give  $\beta_1 = \beta_2$ , therefore we have contrast vector  $\lambda = [0, 1, -1]^T$ . The related reduced model can be seen in *Figure 9*

**iv)** The t-statistic for this model is -7.04, which is the same as Q1.c and Q1.b

**v)** For the GLM we have  $\hat{\beta}_0 + \hat{\beta}_1$  equal to the mean of sample group 1, and  $\hat{\beta}_0 + \hat{\beta}_2$  equal to the mean of sample group 2.

**Q1.e)**

**i)** For the new GLM the design matrix is shown in *Figure 10*.  $Dim(X) = 2$

**ii)** A Null hypothesis of the groups having the same mean I have the contrast vector  $\lambda = [0, 1]^T$ .

**iii)** Using the same method as before I calculate a t-statistic of -7.04, the same as before.

**iv)** The GLM doesn't have a membership column vector for group 2, therefore  $\hat{\beta}_0$  is equal to the mean of sample group 2. This implies that  $\hat{\beta}_0 + \hat{\beta}_1$  must be equal to the sample mean of group 1.

**Q1.f)**

We cannot complete the same hypothesis test using this model. We need a solution space of at least 2 dimensions when completing the test, because when creating the reduced model it by definition has to have less dimensions, and this is not possible with the proposed GLM which only has 1 dimension in its solution space.

**Q2.a)**

In Python to calculate the 1 sample t-test I used the function `scipy.stats.ttest_rel()`. I calculate a t-statistic of -6.133. This is different to what was calculated before because this test is assuming the two groups are no longer independent (and still have the same mean and variance).

**Q2.b)**

**i)** The design matrix for this model is shown in *Figure 12*. The rank of the design matrix is 21 (the first column is equal to the sum of the last 20 columns).

**ii)** In *Equation 9* I derive that  $\lambda = [0, 1, 0, \dots, 0]^T$

**iii)** I calculate the same t-statistic -6.133 as I did in Q2.a). This confirms the GLM is setup assuming the two group samples are no-longer independent.

## 2 Part 2

*Note: Code for Part 2 Q1 is in the **code pdf pages 11-17***

**Q1.a)** Generated data can be seen graphically in *Figure 13*. The calculated 2 sample t-statistic for this data is -4.86 with a p-value of 0.00039. This gives strong evidence that the means of the two samples are not

---

equal.

**Q1.b)**

i), ii), iii), iv) In Python I construct an array  $D$  which holds all the generated data. I then construct an array of all possible unordered permutations of this data such that group 1 and group 2 keep their original sample sizes. There are 3003 unique permutations, so the set of all unique permutations of the data is  $3003 \times 14$ . Calculating the t-statistic for each unique permutation I get the histogram and empirical distribution shown in *Figure 14*. The original t-statistic was  $-4.86$ , using the empirical distribution there are only 4 permutations (including the original) that have an equally or more extreme value of the t-statistic in a 2-tailed test. This gives a p-value of  $4/3003 = 0.00133$ .

Compared to Q1.a) the p-values are different. We wouldn't expect them to be exactly the same, because the empirical distribution p-value only depends on the different relative orderings of the permutations, whereas the Q1.a) p-value is dependent on the relative values of the groups. For example if I reduce the smallest value in group 1 by  $m \in \mathbb{R}_+$  then the p-value in Q1.a) would change, but the p-value I calculated in Q1.b) would remain the same because group 1 and group 2 would still be the 4th most extreme possible t-statistic out of all the permutations, giving the same p-value.

**Q1.c)**

Repeating the above but with the mean-difference statistic =  $\text{mean}(\text{group1}) - \text{mean}(\text{group2})$ , I get the mean-diff statistic of  $-0.48$  for the original data. A histogram and empirical distribution can be seen in *Figure 15*. Similar to above I complete a 2-tailed test to calculate the p-value, so there are 4 equally or more extreme values of the mean-diff statistic, giving a p-value of  $0.00133$  which is the same as Q1.b). For the same reason this is different to Q1.a).

**Q1.d)**

i), ii) I repeat Q1.b) but only using sub-sample of all permutations. I create the sub-sample by sampling from all permutations without replacement 1000 times and I ensure the first sample is the original non-permuted dataset. This way all the permutations in my sub-sample are unique. The histogram and empirical distribution can be seen in *Figure 16*. With my 1000 sample there are 4 equally or more extreme values of t-statistic (again doing a 2-tailed test). This gives a p-value of  $4/1000 = 0.004$ . This is much higher than Q1.c) and Q1.b). This method is inherently random due to the sampling, and if I changed the random seed in my Python script I might get a different answer. It is very susceptible to the number of permutations that are equally or more extreme than the original data.

iii) Due to my sampling method there are only unique permutations. The expected value of the p-value won't change due to sampling with or without replacement, but sampling without replacement will reduce the variance of the p-value.

**Q2**

*Note: Code for Part 2 Q2 is in the **code pdf pages 18-22***

a) I will be using the GLM:  $Y = X_1\beta_1 + X_2\beta_2 + e$ , where  $e$  is sampled from  $\mathcal{N}(0, \sigma^2)$ . This gives the design matrix detailed in *Equation 17* with a  $\text{Dim}(X) = 2$ . Using the same method detailed in Part 1 Q1.c) I calculate the t-statistic for all relevant voxels. The maximum t-statistic calculated is  $6.53$ .

b) I repeat the previous method for every permutation of the voxels in each group. There are 12870 unique unordered permutations for two groups of 8, therefore we calculate 12870 max t-statistics. The histogram and empirical distribution of this can be seen on *Figure 18*.

c) Using the empirical distribution the original grouping t-statistic of  $6.53$  has a p-value of  $0.0918$ . Therefore our data doesn't hit the standard 5% significance level, and we don't have reason to reject the Null hypothesis that the original voxel groupings have different means.

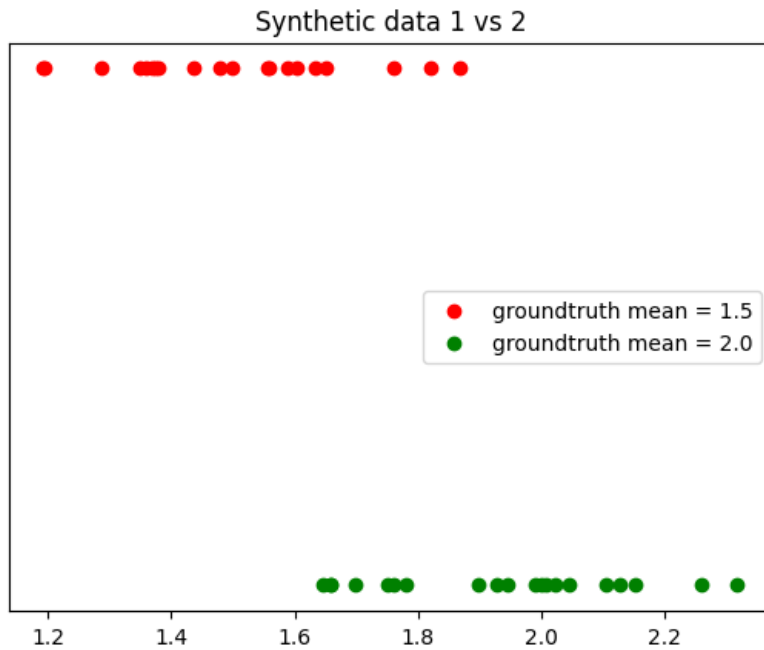
d) Using the empirical distribution to find the 95th percentile max t-statistic. This gives the maximum t-statistic threshold for a p-value of 5% to be  $6.93826$ .

---

## Appendix A: Figures Equations

### 3 Part 1 - Supporting Material

#### 3.1 Q1.a)



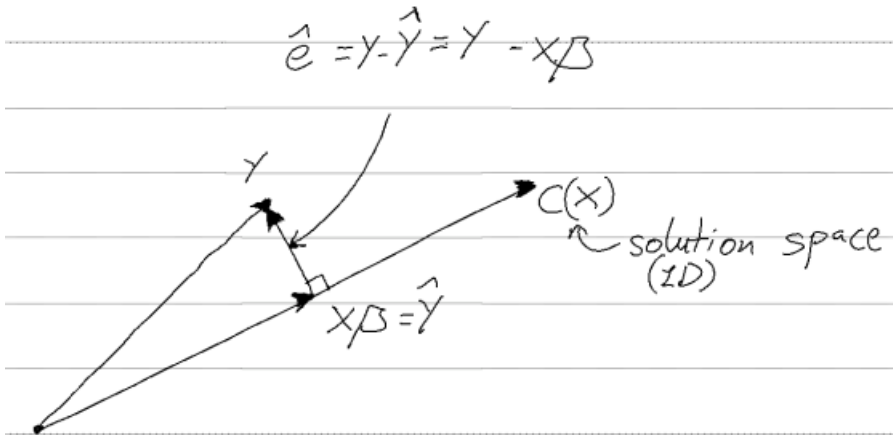
**Figure 1** Synthetic Group 1 (red) and Group 2 (green) generated data, with standard deviation noise of 0.2

#### 3.2 Q1.c.i)

$$X = \begin{bmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix} \begin{matrix} \downarrow 20 \\ \downarrow 20 \end{matrix} \in \mathbb{R}^{40 \times 2}$$

**Figure 2** Design matrix for Group 1 and Group 2 synthetic data for GLM  $Y = X_1\beta_1 + X_2\beta_2 + e$

#### 3.3 Q1.c.ii)

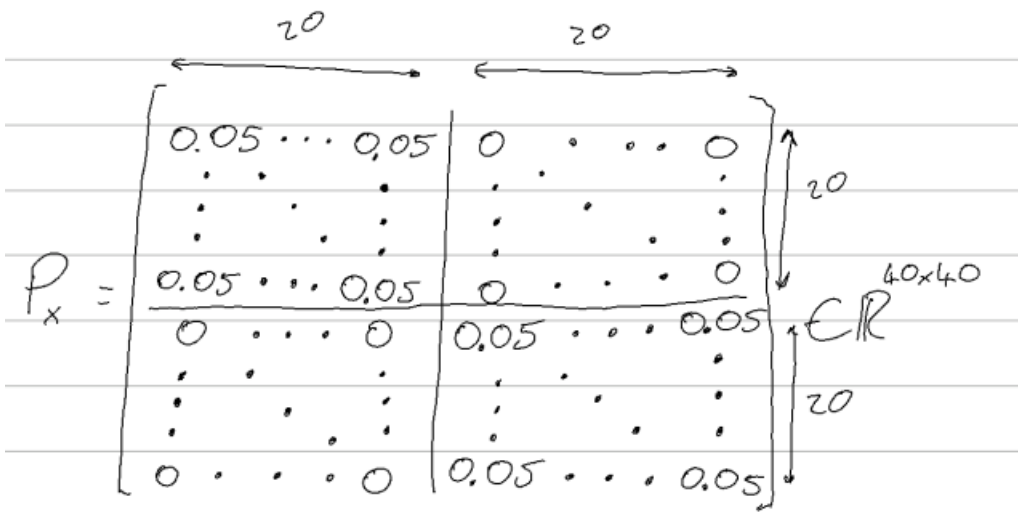


**Figure 3** Diagram of 1D solution space  $C(X)$ .  $\hat{Y}$  is in the solution space and is minimising the distance between the solution space  $C(X)$  and the target solution  $Y$

Derivation of Projection Mapping  $P_X$

$$\begin{aligned}
 X^T(Y - X\beta) &= 0 \\
 X^TY - X^TX\beta &= 0 \\
 X^TX\beta &= X^TY \\
 \beta &= (X^TX)^{-1}X^TY
 \end{aligned} \tag{1}$$

So we have  $\hat{Y} = X\beta = X(X^TX)^{-1}X^TY$ . From the diagram we can see this is projecting  $Y$  onto the solution space. We therefore have the projection mapping  $P_X = X(X^TX)^{-1}X^T$



**Figure 4** Calculated  $P_X$  for answer Part 1 Q1.c.ii)

### 3.4 Q1.c.iii)

$$\hat{Y} = P_X Y = \begin{bmatrix} 1.49782 \\ \vdots \\ 1.49782 \\ 1.93798 \\ \vdots \\ 1.93798 \end{bmatrix} \begin{matrix} 20 \\ \downarrow \\ 20 \end{matrix} \in \mathbb{R}^{20}$$

**Figure 5** Calculated  $\hat{Y}$  for answer Part 1 Q1.c.iii)

### 3.5 Q1.c.iv)

$$R_X = I - P_X = \begin{bmatrix} \xrightarrow{20} & \xrightarrow{20} \\ \begin{matrix} 0.95 & -0.05 & \dots & -0.05 \\ -0.05 & 0.95 & & \vdots \\ \vdots & & & -0.05 \\ -0.05 & \dots & -0.05 & 0.95 \end{matrix} & \begin{matrix} 0 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & 0 \end{matrix} \\ \hline \begin{matrix} 0 & \dots & 0 \\ \vdots & & \vdots \\ \vdots & & \vdots \\ 0 & \dots & 0 \end{matrix} & \begin{matrix} 0.95 & -0.05 & \dots & -0.05 \\ -0.05 & \dots & & \vdots \\ \vdots & & & -0.05 \\ -0.05 & \dots & -0.05 & 0.95 \end{matrix} \end{bmatrix} \begin{matrix} 20 \\ \downarrow \\ 20 \end{matrix}$$

**Figure 6** Calculation of  $R_X$  for q1.c.iv)

Derivation that  $R_X$  is a perpendicular projection mapping.

$$\begin{aligned} R_X R_X &= (I - P_X)(I - P_X) = I - P_X P_X = I - P_X = R_X \\ R_X^T &= (I - P_X)^T = (I - P_X) = R_X \end{aligned} \quad (2)$$

### 3.6 Q1.c.v)

Estimated error

$$\hat{e} = \quad (3)$$

[0.090, 0.263, 0.060, 0.001, -0.120, -0.129, -0.305, -0.140, -0.124, 0.106, 0.370, 0.058, 0.154, -0.019, -0.061, 0.324, -0.149, -0.210, -0.303, 0.135, 0.108, 0.169, -0.177, 0.323, 0.216, 0.007, 0.381, -0.279, 0.052, -0.009, -0.039, -0.240, 0.070, -0.158, -0.280, -0.188, 0.086, 0.189, 0.062, -0.292]

---

### 3.7 Q1.c.vi)

Calculating angle  $\theta$  between vectors  $\hat{e}$  and  $\hat{Y}$

$$\theta = \cos^{-1} \left( \frac{\hat{e} \cdot \hat{Y}}{|\hat{e}| |\hat{Y}|} \right) \quad (4)$$

### 3.8 Q1.c.vii)

Derivation of Projection Mapping  $P_X$

$$\begin{aligned} \forall \alpha \in \mathbb{R} \\ \hat{e} \in C(X)^\perp \implies \hat{e}^T(X\alpha) = 0 &= (Y - X\hat{\beta})^T(X\alpha) \\ &= (Y^T X - \hat{\beta}^T X^T X)\alpha \\ &= Y^T X - \hat{\beta}^T X^T X \\ &= X^T Y - X^T X \hat{\beta} \\ \beta &= (X^T X)^{-1} X^T Y \end{aligned} \quad (5)$$

### 3.9 Q1.c.ix)

Calculated variance matrix of our parameter estimates:

$$S_{\hat{\beta}} = \begin{bmatrix} 0.00195 & 0 \\ 0 & 0.00195 \end{bmatrix} \quad (6)$$

### 3.10 Q1.c.xi)

Calculated variance matrix of our parameter estimates:

$$F_{v_1, v_2} = \frac{(SSR(X_0) - SSR(X))/v_1}{SSR(X)/v_2} \quad (7)$$

where,  $v_1 = \text{tr}(P_X - P_{X_0})$   
 $v_2 = \text{tr}(I - P_X)$

### 3.11 Q1.c.xiv)

$$y_{gt} = \begin{bmatrix} 1.5 \\ \vdots \\ 1.5 \\ 2.0 \\ \vdots \\ 2.0 \end{bmatrix} \begin{matrix} | \\ 20 \\ \downarrow \\ 20 \end{matrix} \quad e = y_{gt} - y$$

$$P_X e = \begin{bmatrix} 0.00218 \\ \vdots \\ 0.00218 \\ 0.06202 \\ \vdots \\ 0.06202 \end{bmatrix} \begin{matrix} | \\ 20 \\ \downarrow \\ 20 \end{matrix}$$

**Figure 7** Projection of the ground truth deviation error  $e$  into  $C(X)$

### 3.12 Q1.c.xv)

projection of the ground truth deviation  $e$  into  $C(X)^\perp$

$$R_X e = \quad (8)$$

[-0.090, -0.263, -0.060, -0.001, 0.120, 0.129, 0.305, 0.140, 0.124, -0.106,  
0.370, -0.058, -0.154, 0.019, 0.061, -0.324, 0.149, 0.210, 0.303, -0.135,  
-0.108, -0.169, 0.177, -0.323, -0.216, -0.007, -0.381, 0.279, -0.052, 0.009,  
0.039, 0.240, -0.070, 0.158, 0.280, 0.188, -0.086, -0.189, -0.062, -0.292]

### 3.13 Q1.d.i)

$$X = \begin{bmatrix} 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \end{bmatrix} \begin{matrix} | \\ 20 \\ \downarrow \\ 20 \end{matrix} \in \mathbb{R}^{40 \times 3}$$

**Figure 8** Design matrix for chosen GLM with intercept in Q1.d.i)

### 3.14 Q1.d.iii)



$$X_0 = \begin{bmatrix} 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \end{bmatrix} \begin{matrix} \downarrow \\ 40 \\ \downarrow \end{matrix} \in \mathbb{R}^{40 \times 2}$$

**Figure 9** Reduced Model  $X_0$  for GLM with intercept in Q1.d.iii)

**3.15 Q1.e.i)**

$$X = \begin{bmatrix} 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \\ 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \end{bmatrix} \begin{matrix} \downarrow \\ 20 \\ \downarrow \\ 20 \\ \downarrow \end{matrix} \in \mathbb{R}^{40 \times 2}$$

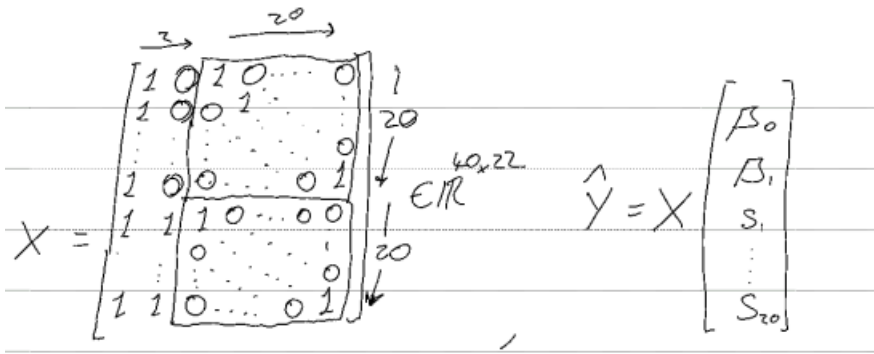
**Figure 10** Design matrix for GLM in Q1.e.i)

**3.16 Q1.e.ii)**

$$X_0 = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \in \mathbb{R}^{40}$$

**Figure 11** Reduced model for GLM in Q1.e.ii)

**3.17 Q2.b.i)**



**Figure 12** Design matrix for GLM in Q2.b.i)

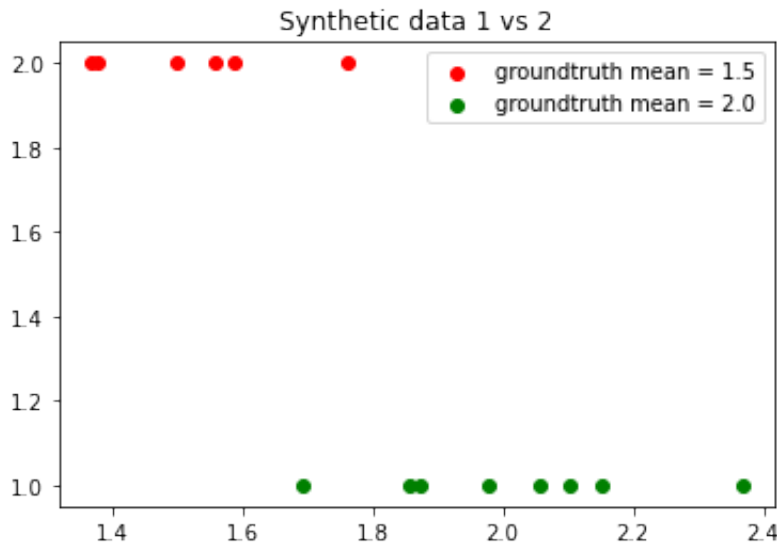
### 3.18 Q2.b.ii)

Contrast vector derivation for sample means being equal

$$\begin{aligned}
 \text{sample 1 mean} &= \text{sample 2 mean} \\
 20\hat{\beta}_0 + \sum_{i=1}^{20} s_i &= 20\hat{\beta}_0 + 20\hat{\beta}_1 + \sum_{i=1}^{20} s_i \\
 20\hat{\beta}_1 &= 0 \\
 \hat{\beta}_1 = 0 &\implies \lambda = [0, 1, 0, \dots, 0]^T
 \end{aligned} \tag{9}$$

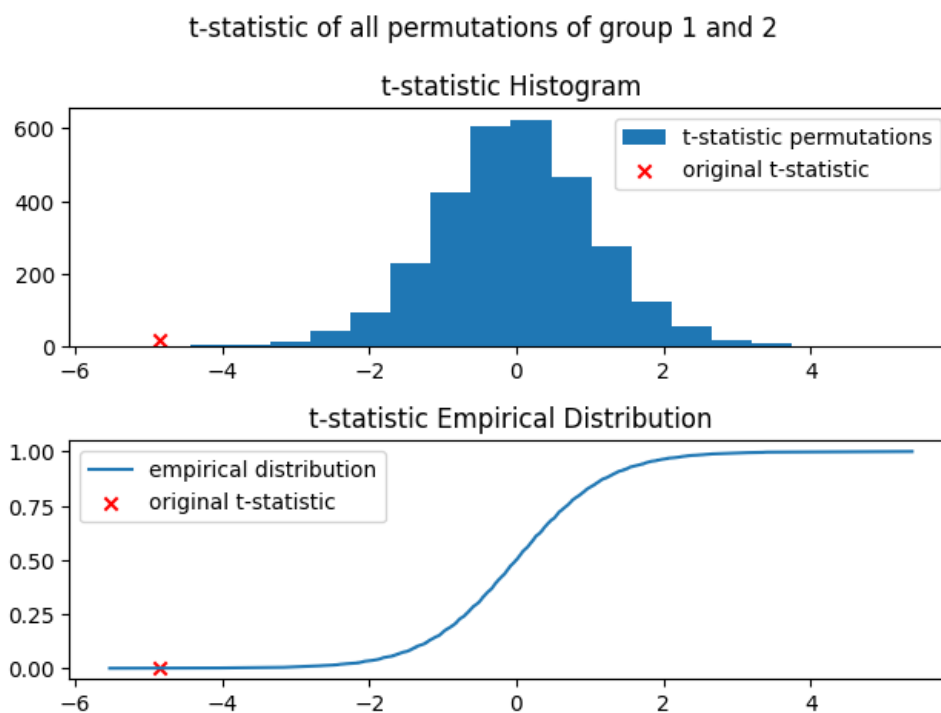
## 4 Part 2 - Supporting Material

### 4.1 Q1.a)



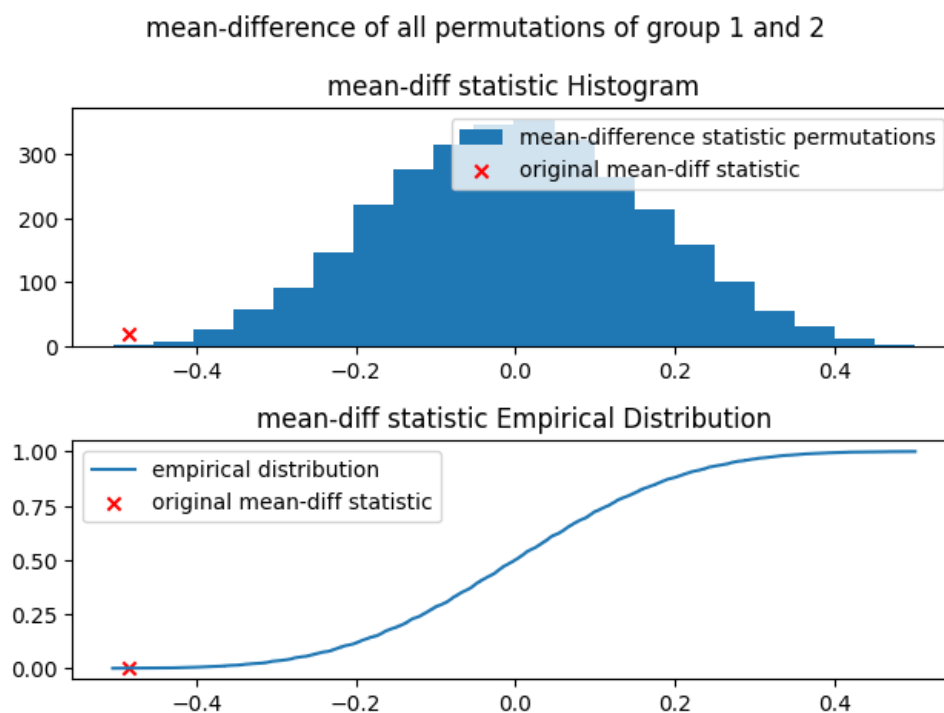
**Figure 13** Generated data for group 1 and 2 with means 1.5 and 2.0 respectively. Both groups have noise with standard deviation of 0.2

### 4.2 Q1.b)



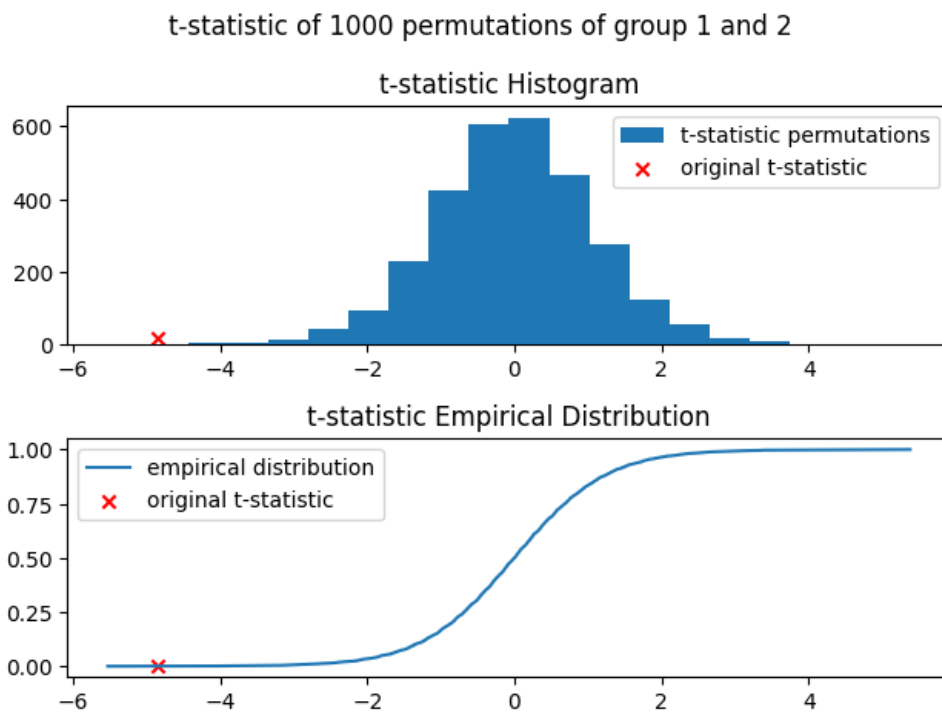
**Figure 14** Empirical distribution of t-statistic of group 1 and group 2 data, calculated by finding all permutations of group 1 and 2 data

#### 4.3 Q1.c)



**Figure 15** Empirical distribution of mean-statistic of group 1 and group 2 data, calculated by finding all permutations of group 1 and 2 data

#### 4.4 Q1.d)



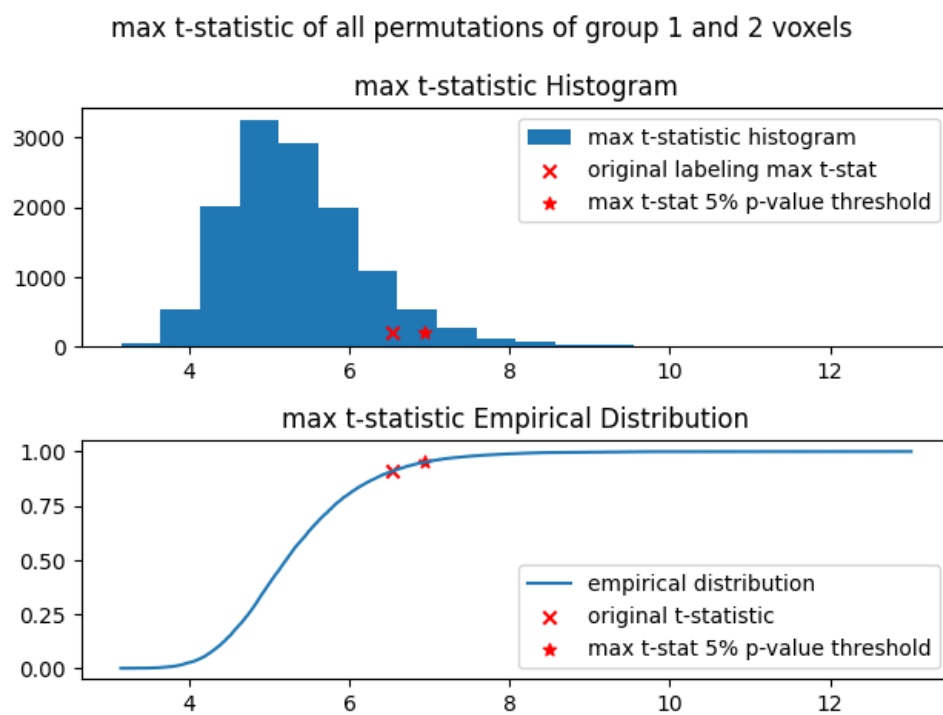
**Figure 16** Empirical distribution of t-statistic of group 1 and group 2 data using a 1000 sub-sample of all permutations

#### 4.5 Q2.b)

$$X = \begin{bmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix} \begin{matrix} 1 \\ 8 \\ \downarrow \\ 1 \\ 8 \\ \downarrow \end{matrix} \in \mathbb{R}^{16 \times 2} \quad \hat{Y} = X\hat{\beta}$$

**Figure 17** Design matrix of GLM chosen in Q2.b)

#### 4.6 Q2.c)



**Figure 18** Histogram and Empirical distribution of the max t-statistic for all voxels. Calculated using all permutations of group 1 and group 2