

# STAT 608 HW 2

Jack Cunningham (jgavc@tamu.edu)

9/13/24

1)

We can represent our model in matrix form below:

$$y = X\alpha + e_i$$

Where  $\alpha = \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix}$ . And design matrix X is:

$$X = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{bmatrix}$$

We are looking for a geometric reason that  $\sum_{i=1}^5 \hat{e}_i = 0$ . We know that  $\hat{e}$  exists in the null-space of the column space of X,  $C(X)$ . So  $\hat{e}^T v = 0$  where v is  $Xa$ . Then, using matrix notation we can represent the sum as the below, where  $\hat{e}_5 = [\hat{e}_1, \hat{e}_2, \hat{e}_3, \hat{e}_4, \hat{e}_5]^T$  and  $1_5 = [1, 1, 1, 1, 1]^T$ :

$$\sum_{i=1}^5 \hat{e}_i = \hat{e}_5^T \cdot 1_5$$

We can easily see that  $1_5$  is in the column space of X. It is simply X's second column. So we can say  $\sum_{i=1}^5 \hat{e}_i = 0$ .

2)

I left intercept  $\beta_0$  in this model because I didn't want to assume that the weight gain of the rats in absence of sugar and fat in their diet would be 0. So:

$$Y = X\beta + e$$

Where the design matrix X is:

$$X = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

And vector  $\beta$  is:

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

The estimate for  $\beta$  is:

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

3)

a)

The first parameter  $\alpha_1$  should be the expected value of the first group  $E[Y|x_1 = 1]$ , so it should be the mean of the response variables with  $x_1 = 1$ . The second parameter  $\alpha_2$  should be expected value of the second group  $E[Y|x_2 = 1]$ , so it should be the mean of the response variables with  $x_2 = 1$ .

b)

In this case the design matrix X is the below:

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} \\ x_{2,1} & x_{2,2} \\ \dots & \dots \\ x_{n,1} & x_{n,2} \end{bmatrix}$$

And we can say that there are m people in group 1 and n-m people in group 2. So:

$$\sum_{i=1}^n x_{i,1} = m, \sum_{i=1}^n x_{i,2} = n - m$$

We can use these facts to estimate  $\hat{\alpha}$ , the usual formula is below:

$$\hat{\alpha} = (X^T X)^{-1} X^T y$$

Where  $X^T X$  is:

$$X^T X = \begin{bmatrix} \sum_{i=1}^n x_{i,1}^2 & \sum_{i=1}^n x_{i,1} x_{i,2} \\ \sum_{i=1}^n x_{i,1} x_{i,2} & \sum_{i=1}^n x_{i,2}^2 \end{bmatrix} = \begin{bmatrix} m & 0 \\ 0 & n - m \end{bmatrix}$$

Then  $(X^T X)^{-1}$  is:

$$(X^T X)^{-1} = \frac{1}{m(n-m)} \begin{bmatrix} m & 0 \\ 0 & n - m \end{bmatrix} = \begin{bmatrix} \frac{1}{m} & 0 \\ 0 & \frac{1}{n-m} \end{bmatrix}$$

And  $X^T y$  is the below, where  $\bar{y}_1$  and  $\bar{y}_2$  are the mean responses for groups 1 and 2.

$$X^T y = \begin{bmatrix} \sum_{i=1}^n x_{i,1} y_i \\ \sum_{i=1}^n x_{i,2} y_i \end{bmatrix} = \begin{bmatrix} m \bar{y}_1 \\ (n - m) \bar{y}_2 \end{bmatrix}$$

So the final estimate,  $\hat{\alpha}$  is:

$$\hat{\alpha} = \begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \end{bmatrix}$$

This is consistent with what we stated in part a, the coefficients  $\hat{\alpha}_1, \hat{\alpha}_2$  are  $E[Y|x_1 = 1] = \bar{y}_1$  and  $E[Y|x_2 = 1] = \bar{y}_2$ .

4)

a)

The design matrix  $X$  in this case is:

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} \\ x_{2,1} & x_{2,2} \\ \dots & \dots \\ x_{n,1} & x_{n,2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}$$

With  $\beta$ :

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

We use the linear model below:

$$Y = X\beta + e$$

And the estimate  $\hat{\beta}$  is:

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

Starting with  $X^T X$ :

$$X^T X = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$$

So  $(X^T X)^{-1}$  is:

$$(X^T X)^{-1} = \frac{1}{5} \begin{bmatrix} 3 & -2 \\ -2 & 3 \end{bmatrix}$$

And  $X^T y$  is:

$$X^T y = \begin{bmatrix} y_1 + y_3 + y_4 \\ y_2 + y_3 + y_4 \end{bmatrix}$$

So our final estimate  $\hat{\beta}$  is:

$$\hat{\beta} = \frac{1}{5} \begin{bmatrix} 3y_1 - 2y_2 + y_3 + y_4 \\ -2y_1 + 3y_2 + y_3 + y_4 \end{bmatrix}$$

b)

A way I made sense of the estimate's numerator is by just doing a little bit of algebra:

$$\hat{\beta} = \frac{1}{5} \begin{bmatrix} 3y_1 + (y_3 - y_2) + (y_4 - y_2) \\ 3y_2 + (y_3 - y_1) + (y_4 - y_1) \end{bmatrix}$$

Observations 1 and 2 are pretty straightforward, a good estimate of each coin's weight, is obtained by weighing both of them individually.

In Observations 3 and 4 we weigh both coins at the same time. Let's say we are estimating coin 1's weight. It's sensible to take the total weight of both coins and subtract away the weight of coin 2, we would end up with the weight of coin 1 with random error. But since we don't know the true weight of coin 2, we subtract away the estimated weight of coin 2 that we obtained from observation 2. That is what the  $(y_3 - y_2)$  and  $(y_4 - y_2)$  terms are.

When it comes to weighing our two estimates I think there is some intuitiveness to observations 1 and 2 having more weight on the final estimates of coin weight,  $\hat{\beta}_1$  and  $\hat{\beta}_2$ . For simplicity let's continue to focus on coin 1. When we attempt to isolate coin 1's weight from observations 3 and 4 we have the following terms :  $(y_3 - y_2)$  and  $(y_4 - y_2)$ . We know that  $y_i = X_i\beta + e_i$ , each of these terms is subject to the random error from the weighing of both coins together  $(y_3, y_4)$  and the random error in weighing coin 2 individually as in  $y_2$ . This makes them less precise estimates of coin 1's weight when compared to observation 1 which is subject to only the error in  $y_1$ . Our final estimate  $\hat{\beta}_1$  therefore gives more credence to observation 1.

5)

The correct answer is d.

RSS is the variability unexplained by the model, SSreg is the variability explained by the model. They total up to the SST, the total sample variability. So:

$$SST = SSreg + RSS$$

It is clear that model 1 has y values closer to the regression line than model 2, this indicates a smaller residual sum of squares in model 1. So:

$$RSS_1 < RSS_2$$

Then to evaluate SSreg we can say:

$$SSreg_1 = SST - RSS_1$$

and

$$SSreg_2 = SST - RSS_2$$

Since  $RSS_1 < RSS_2$  we can conclude that SSreg for model 1 is greater than SSreg for model 2.

6)

a)

We want to show that  $(y_i - \hat{y}_i) = (y_i - \bar{y}) - \hat{\beta}_1(x_i - \bar{x})$ . Starting with  $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ :

$$(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))$$

We know that  $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$  so:

$$(y_i - (\bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 x_i))$$

This leads us to our desired conclusion:

$$(y_i - \bar{y}) - \hat{\beta}_1(x_i - \bar{x})$$

b)

We want to show that  $(\hat{y}_i - \bar{y}) = \hat{\beta}_1(x_i - \bar{x})$ . Starting with  $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ :

$$(\hat{\beta}_0 + \hat{\beta}_1 x_i - \bar{y})$$

We know that  $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$  so:

$$(\bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 x_i - \bar{y})$$

This leads to:

$$\hat{\beta}_1(x_i - \bar{x})$$

c)

Using our previous results we have:

$$\begin{aligned} \sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) &= \sum_{i=1}^n ((y_i - \bar{y}) - \hat{\beta}_1(x_i - \bar{x}))(\hat{\beta}_1(x_i - \bar{x})) \\ &= \sum_{i=1}^n (\hat{\beta}_1(x_i - \bar{x})(y_i - \bar{y}) - \hat{\beta}_1^2(x_i - \bar{x})^2) \end{aligned}$$

We know that  $\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = SXY$  and  $\sum_{i=1}^n (x_i - \bar{x})^2 = SXX$ . So we have:

$$\hat{\beta}_1 SXY - \hat{\beta}_1^2 SXX$$

Using the fact that  $\hat{\beta}_1 = \frac{SXY}{SXX}$ :

$$\sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = \frac{SXY^2}{SXX} - \frac{SXY^2}{SXX} = 0$$

7)

a)

In order for  $t = (\hat{\beta}_1 - 0)/se(\hat{\beta}_1)$  to have a t distribution we must assume that  $e_i \sim N(0, \sigma^2)$ , that  $e_i$  are i.i.d and that the t statistic of  $\beta_1$  and  $\chi$  statistic of  $\hat{\sigma}$  are independently distributed. We then can say  $\hat{\beta}_1 \sim N(\beta_1, \frac{\sigma^2}{SXX})$ . So:

$$SD(\hat{\beta}_1) = \frac{\sigma}{\sqrt{SXX}}$$

To estimate  $\sigma$  we estimate  $\hat{\sigma}^2$ :

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

With our assumption of  $e \sim N(0, \sigma^2 I_n)$  we can say:

$$\frac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi_{n-2}^2$$

Then:

$$\frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}/\sqrt{SXX}} \sim t_{n-2}$$

b)

As the sample size increases the t statistic approaches a standard normal distribution but we still need these assumptions to be reasonably met.

8)

With  $x$  as a random  $n$ -dimensional vector and  $E(x) = \mu$ :

$$\sum = E[(x - \mu)(x - \mu)^T] = E[x(x - \mu)^T - \mu(x - \mu)^T]$$

$$\sum = E[xx^T - x\mu^T - \mu x^T + \mu\mu^T]$$

$$\sum = E[xx^T] - E[x]\mu^T - \mu E[x^T] + \mu\mu^T$$

Here we use that  $E(x) = \mu$  then we get to the desired result:

$$\sum = E[xx^T] - \mu\mu^T - \mu\mu^T + \mu\mu^T = E[xx^T] - \mu\mu^T$$