



Course > Unit 6 Linear Regression > Homework 10 Linear regression > 1. Hypothesis Testing

## 1. Hypothesis Testing Setup:

Suppose we have n observations  $(\mathbf{X}_i, Y_i)$ , where  $Y_i \in \mathbb{R}$  is the dependent variable,  $\mathbf{X}_i \in \mathbb{R}^p$  is the **column**  $p \times 1$  vector of deterministic explanatory variables, and the relation between  $Y_i$  and  $\mathbf{X}_i$  is given by

$$Y_i = \mathbf{X}_i^T eta + \epsilon_i, \qquad i = 1, \dots, n.$$

where  $\epsilon_i$  are i.i.d.  $\mathcal{N}\left(0,\sigma^2
ight)$ . As usual, let  $\mathbb X$  denote the n imes p matrix whose rows are  $\mathbf X_i^T$  .

Unless otherwise stated, assume  $\mathbb{X}^T\mathbb{X} = \tau \mathbf{I}$  and that  $\tau$ ,  $\sigma^2$  are known constants.

(a)

2/2 points (graded)

Recall that under reasonable assumptions (which is certainly satisfied in linear regression with Gaussian noise), the Fisher Information of a parameter  $\theta$  given a family of distributions  $\mathbf{P}_{\theta}$  can be computed via the following formula:

$$I\left( heta
ight) = -\sum_{i=1}^{n} H_{ heta} \ln f\left(Y_{i}; heta
ight)$$

where  $H_{\theta}$  denotes the Hessian differentiation operator with respect to  $\theta$ . (Recall the definition in <u>lecture 9</u>).

In terms of  $\mathbb{X}$ ,  $\sigma^2$ , compute the Fisher  $I(\beta)$  information of  $\beta$ .

(Type **X** for  $\mathbb{X}$ , **trans(X)** for the transpose  $\mathbf{X}^T$  of a matrix  $\mathbb{X}$ , and  $\mathbf{X}^{\wedge}$ (-1) for the inverse  $\mathbb{X}^{-1}$  of a matrix  $\mathbb{X}$ .)

$$I(\beta) =$$
 (trans(X)\*X)/sigma^2

Plugging in  $\mathbb{X}^T \mathbb{X} = \tau \mathbf{I}$ , then the Fisher Information simplifies to a scalar multiple of  $\mathbf{I}$ , so that it is a matrix of the form  $\lambda \mathbf{I}$ . Find the multiplicative constant  $\lambda$ , in terms of  $\tau$  and  $\sigma$ .

$$\lambda = \begin{bmatrix} au/sigma^2 \end{bmatrix}$$

**STANDARD NOTATION** 

Submit

You have used 1 of 3 attempts

✓ Correct (2/2 points)

(b)

3/3 points (graded)

**Instructions:** Fill in the blank in terms of  $\sigma$  and  $\tau$ .

Based on the calculation of the Fisher Information (or by other means), we can conclude that the Maximum Likelihood Estimator  $\hat{\beta}$  has entries  $\hat{\beta}_1, \dots, \hat{\beta}_p$  that are independent Gaussians, with variance:

$$\mathsf{Var}\,(\hat{eta}_i) = egin{bmatrix} \mathsf{sigma^2/tau} \end{bmatrix}$$

Suppose we wish to test the hypotheses

$$H_0:eta_1=eta_2, \qquad H_1:eta_1
eqeta_2.$$

Based on the observation made above, a suitable test statistic is  $T_n=rac{\hat{eta}_1-\hat{eta}_2}{\sqrt{{\sf Var}(\hat{eta}_1-\hat{eta}_2)}}.$ 

Find the denominator  $\sqrt{{
m Var}\,(\hat{eta}_1-\hat{eta}_2)}$  (including the square root) in terms of  $\sigma$  and  $\tau$ .

$$\sqrt{\mathsf{Var}(\hat{eta}_1 - \hat{eta}_2)} = \mathsf{sigma*sqrt}(2/\mathsf{tau})$$

What is the appropriate test at significance level lpha=0.01?

(Let  $q_{\alpha}$  denote the standard normal lpha-quantile for each respective choice below. )

$$igcup \psi = \mathbf{1}\left(T_n > q_{0.01}
ight)$$

$$\bigcirc \ \psi = \mathbf{1} \left( T_n > q_{0.005} 
ight)$$

$$\bigcirc \ \psi = \mathbf{1} \left( |T_n| > q_{0.01} 
ight)$$

$$left \psi = \mathbf{1}\left(|T_n| > q_{0.005}
ight)$$



STANDARD NOTATION

Submit

You have used 1 of 3 attempts

✓ Correct (3/3 points)

(c)

2/2 points (graded)

Suppose we instead wish to test the hypotheses  $H_0:(eta_1,eta_2,eta_3)=(0,0,0)$ ,  $H_1:(eta_1,eta_2,eta_3)
eq (0,0,0)$ .

Let  $\gamma$  be some appropriate value corresponding to the significance level, to be determined later. Choose all  $\psi$  that correctly represents the Bonferroni Test of  $H_0$  against  $H_1$ .

$$\psi = \mathbf{1} \left\{ rac{|\hat{eta}_1 + \hat{eta}_2 + \hat{eta}_3|}{3} > q_\gamma 
ight\}$$

$$\psi = \mathbf{1} \left\{ rac{\max(|\hat{eta}_1|,|\hat{eta}_2|,|\hat{eta}_3|)}{\sqrt{\sigma^2/ au}} > q_\gamma 
ight\}$$

$$\psi = \prod_{i=1}^3 \mathbf{1} \left\{ rac{|eta_i|}{\sqrt{\sigma^2/ au}} > q_\gamma 
ight\}$$

$$lacksquare \psi = \mathbf{1} \left\{ |\hat{eta}_1 - \hat{eta}_2 - \hat{eta}_3| > q_\gamma 
ight\}$$



In the Bonferroni test of significance level  $\alpha=0.01$  for testing this particular  $H_0$  against  $H_1$ , what is the numerical value of  $\gamma$ ? Input a fraction or round to the nearest  $10^{-5}$ , if necessary.

0.001666667

Submit

You have used 1 of 3 attempts

✓ Correct (2/2 points)

(d)

5/5 points (graded)

**Instructions:** For an arbitrary significance level  $\alpha \in (0,1)$ , compute an order  $(1-\alpha)$  confidence interval for  $\beta_1=0$  by filling in the blanks. (In other words, find a confidence interval with confidence level  $1-\alpha$ .) Unless otherwise specified, express your answers in terms of  $\sigma^2$ ,  $\tau$ ,  $\alpha$  and the quantile q.

(Type  $q(\alpha)$  to denote  $q_{\alpha}$ , the  $1-\alpha$ -th quantile of the standard Gaussian.)

• The random variable  $\hat{\beta}_1 - \beta_1$  is a Gaussian RV, with a variance that we computed earlier. Find the value of C such that  $\mathbf{P}(-C \leq \hat{\beta}_1 - \beta_1 \leq C) = 1 - \alpha$ .

This gives us the confidence interval  $I=[\hat{eta}_1-C,\hat{eta}_1+C].$ 

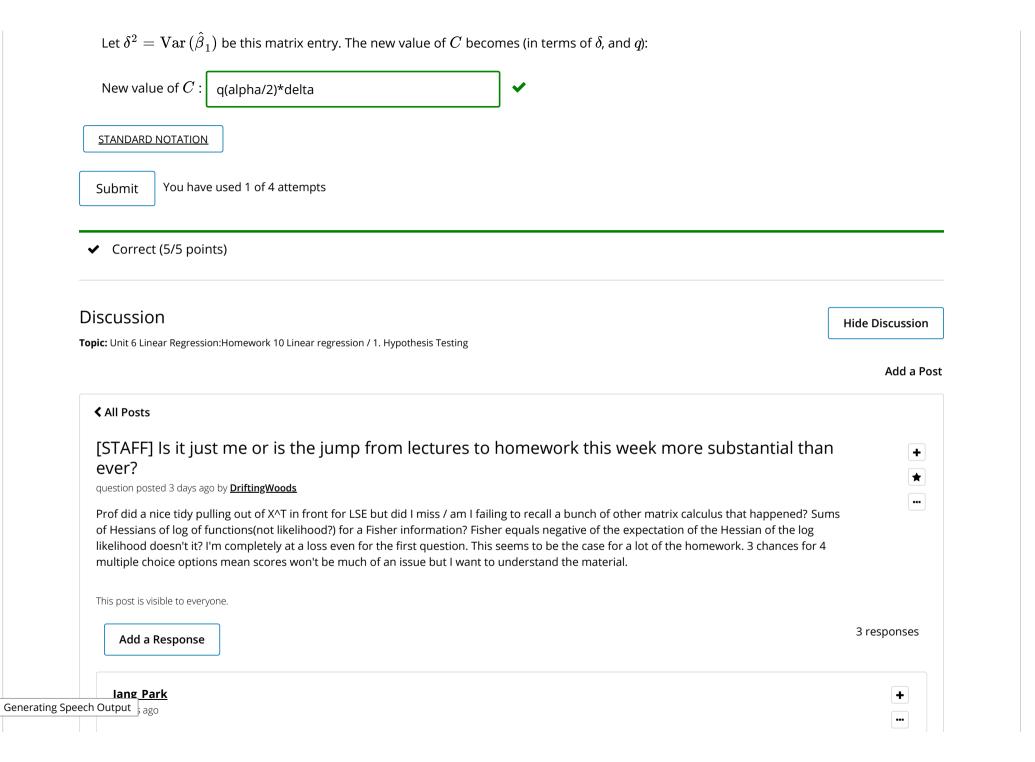
• Revisit part (b). If  $\mathbb{X}^T\mathbb{X}$  were not diagonal, then in terms of  $\sigma$  and  $\mathbb{X}$ , the covariance matrix of  $\hat{\beta}$  is

$$\Sigma_{\hat{eta}} = \operatorname{sigma^2*(trans(X)*X)^(-1)}$$

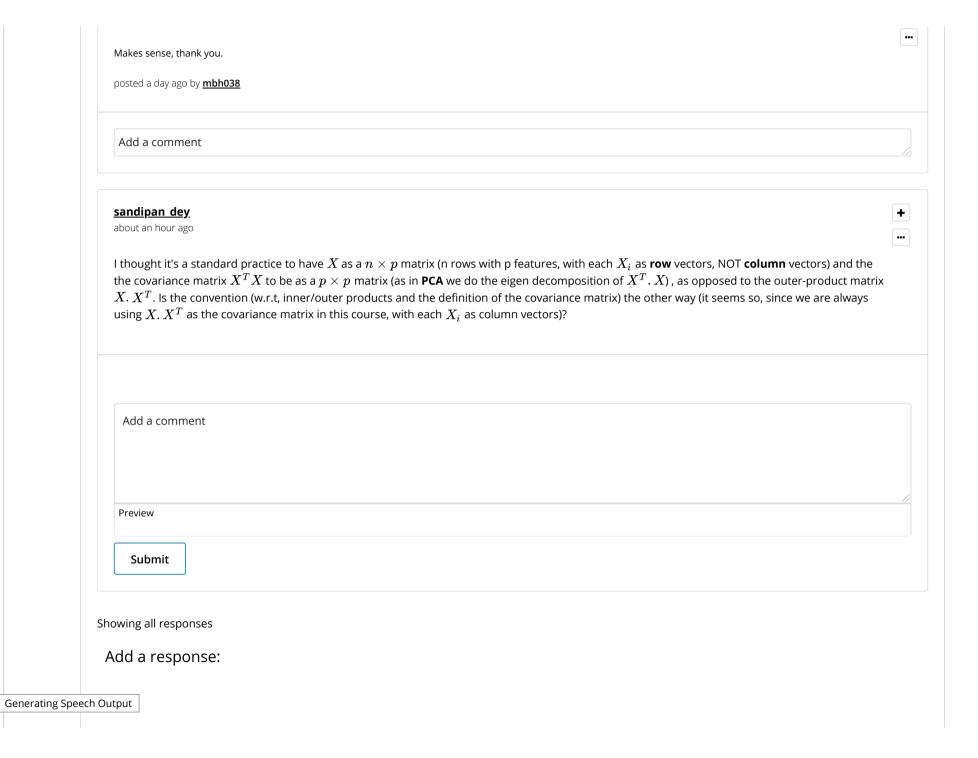
• The variance of  $\hat{\beta}_1$  can be expressed in terms of a particular (i,j) entry of this matrix (the answer to the previous part), where the row-column ordered pair (i,j) is:

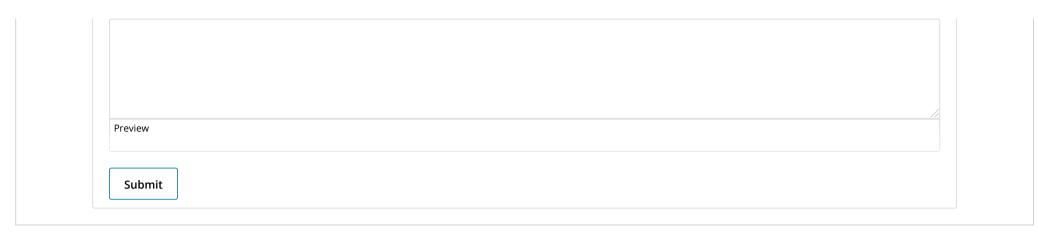
Generating Speech Output

, 🗸 j



Doing the matrix calculus for this problem would be substantial and prone to calculation errors. I would suggest to look for an easier way relating the Fisher Information Matrix to a quantity that we already derived and know. However, taking this approach, it is essential to be able to justify why this can be done. Thank you - I have some better intuition as to what is going on now. posted 2 days ago by **DriftingWoods** Add a comment **Erocha** (Community TA) 2 days ago Note that the Fisher information is just the regular one we saw in class. The difference, I believe, is that you we have a deterministic design, so we have to consider all observations. It is like taking the log-likelihood for all observations (what is the distribution of which one of them, can you tell?). Moreover, note that the Hessian is linear. In addition, you may find useful to consider the sum of rank one matrices. Prof. Gilbert Strang actually has a lecture on that on OCW. ••• I had the same puzzles as you @DriftingWoods. I thought it must be that here we had to use all the observations, whereas before when calculating Fisher Information we just considered one observation. I do not understand, however, why the deterministic design forces us to use all the observations. posted a day ago by mbh038 ••• Note the summation in the definition above. It is including all observations. As you said, previously, the Fisher information had been defined for one observation. So my intuition for understanding the expression above is that before you treated X as fixed. Here,  $\mathbb X$  is what is fixed, and you can only define  $\hat{\beta}$  with this full matrix. a day ago by **Erocha** (Community TA) Generating Speech Output





© All Rights Reserved