

## Stat705 PS 3. Due in class September 29nd

NOTE 1: It is fine to work with other students on problem sets, and that is encouraged. Each person's write up must be done separately though, and identical answers (to relatively complicated problems) from different students will not be graded.

NOTE 2: If you use a source other than the textbook to do these problems, you must list the source. Additionally, it is never OK to copy from another source verbatim! If we notice that, the problem won't be graded.

Please read chapter 2 in Plane Answers...

1. Suppose  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ ,  $\mathbf{y}$  is length  $n$ ,  $\mathbf{X}$  is  $n$  by  $p$  with  $n > p$  and rank  $p$ , and  $E(\mathbf{e}) = \mathbf{0}$ ,  $\text{var}(\mathbf{e}) = \sigma^2\mathbf{I}$ . Also assume that the first column of  $\mathbf{X}$  is all 1s. Let  $\mathbf{P}$  be the perpendicular projection matrix that projects onto a linear combination of the columns of  $\mathbf{X}$ . Let  $\hat{\mathbf{y}} = \mathbf{P}\mathbf{y}$  and  $\mathbf{r} = (\mathbf{I} - \mathbf{P})\mathbf{y}$ .

Use facts about perpendicular projections to show the following.

- (a)  $\mathbf{1}^T\mathbf{r} = 0$ .
  - (b)  $\hat{\mathbf{y}}^T\mathbf{r} = 0$ .
  - (c)  $\mathbf{1}^T\hat{\mathbf{y}}/n = \bar{y}$ .
  - (d)  $\mathbf{r}^T\mathbf{X} = \mathbf{0}$ .
  - (e)  $\hat{\beta}_0 + \hat{\beta}_1\bar{x}_2 + \dots + \hat{\beta}_{p-1}\bar{x}_p = \bar{y}$  where  $\bar{x}_i$  is the sample mean of the  $i$ th column of  $\mathbf{X}$ .
2. Let  $(\mathbf{y}|\boldsymbol{\beta}) \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$  and  $(\boldsymbol{\beta}) \sim N(\boldsymbol{\mu}, \Sigma)$ .

- (a) Verify each step in the following:

$$\log \{(\mathbf{y}|\boldsymbol{\beta})(\boldsymbol{\beta})\} = C_1 - \frac{1}{2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T \left( \frac{\mathbf{X}^T\mathbf{X}}{\sigma^2} \right) (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) - \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu})$$

$$\begin{aligned}
&= C_2 - \frac{1}{2} \left\{ \boldsymbol{\beta}^T \left( \frac{\mathbf{X}^T \mathbf{X}}{\sigma^2} + \Sigma^{-1} \right) \boldsymbol{\beta} - 2 \boldsymbol{\beta}^T \left( \frac{\mathbf{X}^T \mathbf{X}}{\sigma^2} \hat{\boldsymbol{\beta}} + \Sigma^{-1} \boldsymbol{\mu} \right) \right\} \\
&= C_3 - \frac{1}{2} \left[ \{ \boldsymbol{\beta} - m(\boldsymbol{\beta}) \}^T V(\boldsymbol{\beta})^{-1} \{ \boldsymbol{\beta} - m(\boldsymbol{\beta}) \} \right]
\end{aligned}$$

where  $m(\boldsymbol{\beta})$  and  $V(\boldsymbol{\beta})$  were given in class.

(b) What is the distribution of  $(\boldsymbol{\beta}|\mathbf{y})$  and why?

3. Suppose  $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$  where  $\mathbf{y}$  is length  $n$  and  $\mathbf{X}$  is  $n$  by  $p$  with  $n > p$  and rank  $p$ ,

(a) Show that the MLE of  $\sigma^2$  is  $\hat{\sigma}^2 = \mathbf{y}^T(\mathbf{I} - \mathbf{P})\mathbf{y}/n$ .

(b) Find the bias of the MLE. If it is biased, propose an unbiased estimator.

(c) REML (restricted or residual) maximum likelihood is a method to estimate a variance that is often unbiased (or at least less biased) than maximum likelihood.

It considers maximum likelihood using the density of  $\mathbf{K}\mathbf{y}$  where  $\mathbf{K}$  is chosen to be orthogonal to  $E(\mathbf{y})$ .

i. Suppose  $\mathbf{X} = \mathbf{1}_n$  (a vector of all 1s). Let  $\mathbf{K}$  be the first  $n - 1$  rows of  $\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n$ .

Find the log density of  $\mathbf{K}\mathbf{y}$ . (Remember:  $\det(cA) = c^K \det(A)$  where  $A$  is a  $K$  by  $K$  matrix.)

ii. Show that  $\hat{\sigma}_{REML}^2 = \frac{\mathbf{y}^T \mathbf{K}^T (\mathbf{K}\mathbf{K}^T)^{-1} \mathbf{K}\mathbf{y}}{n-1}$  maximizes the likelihood of  $\sigma^2$  given the data  $\mathbf{K}\mathbf{y}$ .

iii. Show that  $\hat{\sigma}_{REML}^2 = S^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$ . Hint: Use SVD to show that  $\mathbf{K}^T(\mathbf{K}\mathbf{K}^T)^{-1}\mathbf{K} = \mathbf{I} - \mathbf{P}$ .

iv. Show that the estimator you proposed in 3b is REML with  $\mathbf{K}$  as the first  $n - p$  rows of  $\mathbf{I} - \mathbf{P}$ .