

# ST758, Homework 2

Due Oct 15, 2013

1. (a) Read in the file ‘longley.dat’ on course webpage, which has the response  $\mathbf{y}$  in the first column and six explanatory variables in the other columns.
- (b) Compute the  $6 \times 6$  sample covariance matrix and call it  $\mathbf{V}$ .
- (c) Compute the  $6 \times 6$  correlation coefficient matrix  $\mathbf{C}$  from  $\mathbf{V}$ . What do you observe in  $\mathbf{C}$ ?
- (d) Partition the matrix  $\mathbf{V}$  as

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix},$$

where the blocks  $\mathbf{V}_{ij}$  have size  $3 \times 3$ . Compute  $\mathbf{V}_{22} - \mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{V}_{12}$ , using Cholesky decomposition.

- (e) Compute  $\mathbf{V}_{22} - \mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{V}_{12}$  again, using sweeping.
  - (f) Assume linear model  $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$ . Include an intercept in your model and compute the regression coefficients  $\hat{\boldsymbol{\beta}}$ , their standard errors, variance estimate  $\hat{\sigma}^2$ , fitted values  $\hat{\mathbf{y}}$ , and residuals  $\hat{\mathbf{e}}$  using three methods – Cholesky, QR, and sweeping. Please compute them directly; you can use other “black-box” function, e.g. `lm()`, only to check.
  - (g) Let  $\mathbf{X}_i \in \mathbb{R}^{n \times i}$  contain the first  $i$  columns of the design matrix. That is  $\mathbf{X}_1 = \mathbf{1}_n$ ,  $\mathbf{X}_2 = (\mathbf{1}_n, \mathbf{x}_1)$ ,  $\mathbf{X}_3 = (\mathbf{1}_n, \mathbf{x}_1, \mathbf{x}_2)$ , and so on. Compute  $\mathbf{y}^T \mathbf{P}_{\mathbf{X}_1} \mathbf{y}$ ,  $\mathbf{y}^T (\mathbf{P}_{\mathbf{X}_i} - \mathbf{P}_{\mathbf{X}_{i-1}}) \mathbf{y}$ ,  $i = 2, \dots, 7$ , and  $\mathbf{y}^T (\mathbf{I}_n - \mathbf{P}_{\mathbf{X}}) \mathbf{y}$ , where  $\mathbf{P}_{\mathbf{A}}$  denotes the orthogonal projection onto the column space of a matrix  $\mathbf{A}$ . (Hint: don’t ever think about forming these projection matrices.)
2. Let  $\mathbf{X} \in \mathbb{R}^{n \times p}$ ,  $n \geq p$ , be the design matrix in linear regression. The least squares solution is given by

$$\hat{\boldsymbol{\beta}} = \min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2.$$

- (a) Show that  $\hat{\boldsymbol{\beta}}$  is a least squares solution if and only if it satisfies the normal equation  $\mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{X}^T \mathbf{y}$ .
- (b) If  $\mathbf{X}$  has rank  $r < p$ , the least squares solution is not unique. Show that any vector  $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^- \mathbf{X}^T \mathbf{y}$ , where  $(\mathbf{X}^T \mathbf{X})^-$  is any generalized inverse, is a least squares solution, and the residual vector  $\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}$  is invariant to the choice of the generalized inverse.
- (c) Assume  $\mathbf{X}$  has rank  $r < p$  and the *QR decomposition with (column) pivoting* yields

$$\mathbf{X} = \mathbf{Q} \begin{pmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{0}_{(n-r) \times r} & \mathbf{0}_{(n-r) \times (p-r)} \end{pmatrix} \mathbf{\Pi}^T, \quad (1)$$

where  $\mathbf{Q} \in \mathbb{R}^{n \times n}$  is orthogonal,  $\mathbf{R}_{11} \in \mathbb{R}^{r \times r}$  is upper triangular with positive diagonal entries, and  $\mathbf{\Pi} \in \mathbb{R}^{p \times p}$  is a permutation matrix. Show that

$$\mathbf{X}^- = \mathbf{\Pi} \begin{pmatrix} \mathbf{R}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{Q}^T$$

is a generalized inverse of  $\mathbf{X}$ . Is this generalized inverse the Moore-Penrose inverse of  $\mathbf{X}$ ? The `qr()` function in R performs QR decomposition with (column) pivoting. Study the documentation of `qr()` and do the following on the Longley data in Q1.

- (d) Add an extra column (sum of the intercept and the first predictor) to the original design matrix such that  $\mathbf{X}_{\text{new}} = (\mathbf{1}_n, \mathbf{x}_1, \mathbf{1}_n + \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_6)$ . Compute a least squares solution using the generalized inverse in (c) obtained from QR. Compare the least squares solution, fitted values, and residuals to those obtained in Q1(f).
- (e) Given the singular value decomposition (SVD)

$$\mathbf{X} = \mathbf{U} \begin{pmatrix} \Sigma_{11} & \mathbf{0}_{r \times (p-r)} \\ \mathbf{0}_{(n-r) \times r} & \mathbf{0}_{(n-r) \times (p-r)} \end{pmatrix} \mathbf{V}^\top,$$

where  $\mathbf{U} \in \mathbb{R}^{n \times n}$  and  $\mathbf{V} \in \mathbb{R}^{p \times p}$  are orthogonal, and  $\Sigma_{11} \in \mathbb{R}^{r \times r}$  is diagonal with positive diagonal entries. Show that

$$\mathbf{X}^+ = \mathbf{V} \begin{pmatrix} \Sigma_{11}^{-1} & \mathbf{0}_{r \times (n-r)} \\ \mathbf{0}_{(p-r) \times r} & \mathbf{0}_{(p-r) \times (n-r)} \end{pmatrix} \mathbf{U}^\top$$

is the Moore-Penrose inverse of  $\mathbf{X}$ .

- (f) Show that  $\hat{\beta} = \mathbf{X}^+ \mathbf{y}$  is a least squares solution and has the minimum  $\ell_2$  norm among all least squares solutions.
- (g) Redo part (d) but using the Moore-Penrose inverse obtained from SVD. Compare the least squares solution, fitted values, and residuals to those obtained in Q2(d).

3. Let  $\mathbf{X} \in \mathbb{R}^n$  be a random vector with i.i.d. standard normal entries.

- (a) Show that  $\mathbf{X}$  has density

$$f_{\mathbf{X}}(\mathbf{x}) = \left( \frac{1}{2\pi} \right)^{n/2} e^{-\mathbf{x}^\top \mathbf{x} / 2}.$$

- (b) Any affine transformation  $\mathbf{Y} = \mathbf{A}\mathbf{X} + \boldsymbol{\mu}$  of  $\mathbf{X}$ , where  $\mathbf{A} \in \mathbb{R}^{m \times n}$  and  $\boldsymbol{\mu} \in \mathbb{R}^m$ , is called a multivariate normal random vector. Assume  $\text{rank}(\mathbf{A}) = m$ . Show that

$$\mathbf{E}(\mathbf{Y}) = \boldsymbol{\mu}, \quad \text{Var}(\mathbf{Y}) = \mathbf{A}\mathbf{A}^\top = \boldsymbol{\Omega},$$

and the density of  $\mathbf{Y}$  is

$$f_{\mathbf{Y}}(\mathbf{y}) = \left( \frac{1}{2\pi} \right)^{m/2} |\det(\boldsymbol{\Omega})|^{-1/2} e^{-(\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Omega}^{-1} (\mathbf{y} - \boldsymbol{\mu}) / 2}.$$

- (c) Suppose  $\mathbf{Y}$  is partitioned as two sub-vectors  $\mathbf{Y} = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top$  and  $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \boldsymbol{\mu}_2^\top)^\top$  and  $\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{pmatrix}$  are partitioned compatibly. Show that the conditional distribution of  $\mathbf{Y}_2$  given  $\mathbf{Y}_1$  is normal with mean and variance

$$\begin{aligned} \mathbf{E}(\mathbf{Y}_2 \mid \mathbf{Y}_1) &= \boldsymbol{\Omega}_{21} \boldsymbol{\Omega}_{11}^{-1} (\mathbf{Y}_1 - \boldsymbol{\mu}_1) + \boldsymbol{\mu}_2 \\ \text{Var}(\mathbf{Y}_2 \mid \mathbf{Y}_1) &= \boldsymbol{\Omega}_{22} - \boldsymbol{\Omega}_{21} \boldsymbol{\Omega}_{11}^{-1} \boldsymbol{\Omega}_{12}. \end{aligned}$$

Let  $\mathbf{Y} \sim N(\mathbf{0}_m, \boldsymbol{\Omega}_m)$ , where  $\boldsymbol{\Omega}_m = (r^{|i-j|})_{i,j}$ . Use  $m = 10$  and  $r = 0.9$  in below.

- (d) Generate a  $\mathbf{y}$  from this model, using only the `rnorm()` function.
- (e) Partition  $\mathbf{y}$  as  $(\mathbf{y}_1^T, \mathbf{y}_2^T)^T$ , where  $\mathbf{y}_1 \in \mathbb{R}^6$  and  $\mathbf{y}_2 \in \mathbb{R}^4$ . Compute the following quantities at the generated  $\mathbf{y}$ , using either Cholesky or sweeping.
- Marginal density  $f_{\mathbf{Y}_1}(\mathbf{y}_1)$
  - Conditional mean  $\mathbf{E}(\mathbf{Y}_2 \mid \mathbf{Y}_1)$  and conditional variance  $\mathbf{Var}(\mathbf{Y}_2 \mid \mathbf{Y}_1)$
  - Joint density  $f_{\mathbf{Y}}(\mathbf{y})$
  - Conditional density  $f_{\mathbf{Y}_2 \mid \mathbf{Y}_1}(\mathbf{y}_2 \mid \mathbf{y}_1)$
  - Let  $\mathbf{Z} = \mathbf{Y}_2 - \mathbf{E}(\mathbf{Y}_2 \mid \mathbf{Y}_1)$ . Evaluate density  $f_{\mathbf{Z}}(\mathbf{z})$  at  $\mathbf{z} = \mathbf{y}_2 - \boldsymbol{\Omega}_{21} \boldsymbol{\Omega}_{11}^{-1} \mathbf{y}_1$
- (f) Now you have gained some experience using sweeping or Cholesky to handle multivariate normal computations. Re-think part (e). Are there more efficient ways to compute the multivariate normal density for this covariance structure? If you find one, re-evaluate the joint density  $f_{\mathbf{Y}}(\mathbf{y})$  using your method.