# Chapter 13

# Theory for Linear Models

Theory for linear models is used to show that linear models have good statistical properties. This chapter needs a lot more work.

Suppose the linear model is $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{e}$ where $\boldsymbol{X}$ is an $n \times p$ matrix, $\boldsymbol{\beta}$ is a $p \times 1$ vector and $\boldsymbol{e}$ and $\boldsymbol{Y}$ are $n \times l$ vectors.

Assume that the $e_i$ are iid with zero mean and variance $\mathrm{V}(e_i) = \sigma^2$.

Linear model theory previously proved in the text includes Propositions 2.1, 2.2, 2.3, 2.10, 3.1, 3.2, 3.3, and 4.1. Some matrix manipulations are illustrated in Example 4.1.

Unproved results include Propositions 2.4, 2.5, 2.9, 2.11, Theorems 2.6, 2.7, and 2.8. Also see Equation (2.23).

Also assume that the model includes all possible terms so may overfit but does not underfit. Then $\hat{\boldsymbol{Y}} = \boldsymbol{HY}$ and $\mathrm{Cov}(\hat{\boldsymbol{Y}}) = \sigma^2 \boldsymbol{HIH}^T = \sigma^2 \boldsymbol{H}$. Thus

$$\frac{1}{n} \sum_{i=1} V(\hat{Y}_i) = \frac{1}{n} tr(\sigma^2 \boldsymbol{H}) = \frac{\sigma^2}{n} tr((\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{X} = \frac{\sigma^2 p}{n}$$

where $tr(\boldsymbol{A})$ is the trace operation. Hence if only $k$ parameters are needed and $p >> k$, then serious overfitting occurs and increases $\frac{1}{n} \sum_{i=1} V(\hat{Y}_i)$. This result implies Equation (3.7).

## 13.1 Complements

Texts on the theory of linear models include Christensen (2002), Freedman (2005), Graybill (2000), Guttman (1982), Hocking (2003), Porat (1993), Rao

(1973), Ravishanker and Dey (2002), Rencher and Schaalje (2008), Scheffé (1959), Searle (1971) and Seber and Lee (2003).

## 13.2 Problems

**Problems with an asterisk * are especially important.**

**13.1.** Suppose $Y_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + e_i$ where the errors are iid double exponential $(0, \sigma)$ where $\sigma > 0$. Then the likelihood function is

$$L(\boldsymbol{\beta}, \sigma) = \frac{1}{2^n}\frac{1}{\sigma^n}\exp\left(\frac{-1}{\sigma}\sum_{i=1}^n |Y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}|\right).$$

Suppose that $\tilde{\boldsymbol{\beta}}$ is a minimizer of $\sum_{i=1}^n |Y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}|$.

a) By direct maximization, show that $\tilde{\boldsymbol{\beta}}$ is an MLE of $\boldsymbol{\beta}$ regardless of the value of $\sigma$.

b) Find an MLE of $\sigma$ by maximizing

$$L(\sigma) \equiv L(\tilde{\boldsymbol{\beta}}, \sigma) = \frac{1}{2^n}\frac{1}{\sigma^n}\exp\left(\frac{-1}{\sigma}\sum_{i=1}^n |Y_i - \boldsymbol{x}_i^T\tilde{\boldsymbol{\beta}}|\right).$$

**13.2.** Consider the model $Y_i = \beta_0 + \beta_1 X_{i,1} + \cdots + \beta_{p-1}X_{i,p-1} + \epsilon_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + \epsilon_i$. The least squares estimator $\hat{\boldsymbol{\beta}}$ minimizes

$$Q_{OLS}(\boldsymbol{\eta}) = \sum_{i=1}^n (Y_i - \boldsymbol{x}_i^T\boldsymbol{\eta})^2$$

and the weighted least squares estimator minimizes

$$Q_{WLS}(\boldsymbol{\eta}) = \sum_{i=1}^n w_i(Y_i - \boldsymbol{x}_i^T\boldsymbol{\eta})^2$$

where the $w_i, Y_i$ and $\boldsymbol{x}_i$ are known quantities. Show that

$$\sum_{i=1}^n w_i(Y_i - \boldsymbol{x}_i^T\boldsymbol{\eta})^2 = \sum_{i=1}^n (\tilde{Y}_i - \tilde{\boldsymbol{x}}_i^T\boldsymbol{\eta})^2$$

by identifying $\tilde{Y}_i$ and $\tilde{\boldsymbol{x}}_i$. (Hence the WLS estimator is obtained from the least squares regression of $\tilde{Y}_i$ on $\tilde{\boldsymbol{x}}_i$ without an intercept.)

**13.3.** Find the vector $\boldsymbol{b}$ such that $\boldsymbol{b}^T\boldsymbol{Y}$ is an unbiased estimator for $E(Y_i)$ if the usual linear model holds.

**13.4.** Write the following quantities as $\boldsymbol{b}^T\boldsymbol{Y}$ or $\boldsymbol{Y}^T\boldsymbol{A}\boldsymbol{Y}$ or $\boldsymbol{A}\boldsymbol{Y}$.

a) $\overline{Y}$

b) $\sum_i (Y_i - \hat{Y}_i)^2$

c) $\sum_i (\hat{Y}_i)^2$

d) $\hat{\boldsymbol{\beta}}$

e) $\hat{\boldsymbol{Y}}$

**13.5.** Show that $\boldsymbol{I} - \boldsymbol{H} = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$ is idempotent, that is, show that $(\boldsymbol{I} - \boldsymbol{H})(\boldsymbol{I} - \boldsymbol{H}) = (\boldsymbol{I} - \boldsymbol{H})^2 = \boldsymbol{I} - \boldsymbol{H}$.

**13.6.** Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be matrices with the same number of rows. If $\boldsymbol{C}$ is another matrix such that $\boldsymbol{A} = \boldsymbol{B}\boldsymbol{C}$, is it true that $rank(\boldsymbol{A}) = rank(\boldsymbol{B})$? Prove or give a counterexample.

**13.7.** Let $\boldsymbol{x}$ be an $n \times 1$ vector and let $\boldsymbol{B}$ be an $n \times n$ matrix. Show that $\boldsymbol{x}^T\boldsymbol{B}\boldsymbol{x} = \boldsymbol{x}^T\boldsymbol{B}^T\boldsymbol{x}$.

(The point of this problem is that if $\boldsymbol{B}$ is not a symmetric $n \times n$ matrix, then $\boldsymbol{x}^T\boldsymbol{B}\boldsymbol{x} = \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$ where $\boldsymbol{A} = \dfrac{\boldsymbol{B} + \boldsymbol{B}^T}{2}$ is a symmetric $n \times n$ matrix.)