理论题 1

- 令 $\boldsymbol{H}=\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ 是一个帽子矩阵（如定理 1-1），$\boldsymbol{I}$ 为单位阵。证明：$\boldsymbol{I}-\boldsymbol{H}$ 是一个对称且幂等的矩阵。并计算这个矩阵的秩。

证明：$\because$ H 是 n 阶对称矩阵

$\therefore H_{ij}=H_{ji}$

$\because I_{ij}=I_{ji}$

$\therefore I_{ij}-H_{ij}=I_{ji}-H_{ji}$

$\therefore (I-H)_{ij}=(I-H)_{ji}$

$\therefore$ I−H 是一个对称矩阵

$$(I-H)^2=(I-H)(I-H)$$
$$=I^2-IH-HI+H^2$$
$$=I-H-H+H$$
$$=I-H$$

$\therefore$ I−H 是一个幂等矩阵.

$\therefore \operatorname{rank}(I-H)=\operatorname{tr}(I-H)$.
$=\operatorname{tr}I-\operatorname{tr}H$
$=n-p-1$. （p 为自变量数）

理论题 2

- 在一个多元线性回归模型中，响应变量 y_i 的回归值为

$$\hat{y}_i=\hat{\beta}_0+\hat{\beta}_1x_1+\cdots+\hat{\beta}_px_p.$$

$\boldsymbol{X}$ 是一个满秩矩阵，证明：$\sum_{i=1}^{n}(y_i-\hat{y}_i)=0$。

证明：回归模型：$\hat{y}=X\hat{\beta}$.

要证：$\sum_{i=1}^{n}(y_i-\hat{y}_i)=0 \iff 1^T(y-\hat{y})=0$.

由最小二乘估计，得：$\hat{\beta}=(X'X)^{-1}X'y$.

$$\therefore 1^T(y-\hat{y})=1^T(y-X\hat{\beta})$$
$$=1^T(y-X(X'X)^{-1}X'y)$$
$$=1^T(I-X(X'X)^{-1}X')y$$
$$=(1^T-1^TX(X'X)^{-1}X')y$$
$$=(1^T-(X(X'X)^{-1}X'\cdot 1)^T)y.$$

令 $c=(1,0,0,\cdots 0)^T$，则 $1=Xc$.

$\therefore X(X'X)^{-1}X'\cdot 1 = X(X'X)^{-1}X'Xc = Xc = 1$.

所以 $1^T(y-\hat{y}) = (1^T-1^T)y = 0$.

$\therefore \sum_{i=1}^{n}(y_i-\hat{y}_i)=0$.

理论题 3

- 在多元线性回归模型

$$y=\beta_0+\beta_1x_1+\cdots+\beta_px_p+\varepsilon$$

中，我们有数据 $\{(y_i,x_{i1},x_{i2},\cdots,x_{ip})\}_{i=1}^n$。我们可以得到最小二乘估计，记为 $\hat{\boldsymbol{\beta}}=(\hat\beta_0,\hat\beta_1,\cdots,\hat\beta_p)'$。
如果我们对 $y_1,y_2,\cdots,y_n$ 进行中心化，对每一维自变量 $x_{1j},x_{2j},\cdots,x_{nj}$ 均进行了标准化，$j=1,2,\cdots,p$，那么，我们得到的最小二乘估计为 $\tilde{\boldsymbol{\beta}}=(\tilde\beta_0,\tilde\beta_1,\cdots,\tilde\beta_p)'$。
请回答:
  - 这两个估计 $\tilde{\boldsymbol{\beta}}$ 和 $\hat{\boldsymbol{\beta}}$ 之间有什么关系?
  - 求 $\tilde{\boldsymbol{\beta}}$ 的期望和方差。

最小二乘估计为

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

$$= \begin{pmatrix} n & 1_n'X_0 \\ X_0'1_n & X_0'X_0 \end{pmatrix}^{-1} \begin{pmatrix} 1_n' \\ X_0' \end{pmatrix} y$$

$$= \begin{pmatrix} n^{-1}+n^{-2}1_n'X_0A_0X_0'1_n & -n^{-1}1_n'X_0A_0 \\ -n^{-1}A_0X_0'1_n & A_0 \end{pmatrix} \begin{pmatrix} 1_n' \\ X_0' \end{pmatrix} y$$

$$= \begin{pmatrix} n^{-1}1_n'+n^{-2}1_n'X_0A_0X_0'1_n1_n' - n^{-1}1_n'X_0A_0X_0' \\ -n^{-1}A_0X_0'1_n1_n' + A_0X_0' \end{pmatrix} y$$

其中 $A_0 = (X_0'X_0 - n^{-1}X_0'1_n1_n'X_0)^{-1}$

理论题 4

- 已知单因子方差分析模型

$$y_{ij} = \mu_i + \varepsilon_{ij}, \quad i = 1, 2, \cdots, a; j = 1, 2, \cdots, m,$$

其中，ε_{ij} 是独立同分布的随机变量，其分布为 $N(0, \sigma^2)$。我们观测到的数据为 $\{y_{ij}\}$。
证明：单因子方差分析模型可以看作一种多元线性回归模型。提示：
  - 构造一个合适的设计矩阵 $\boldsymbol{X}$；
  - 定义响应变量向量、回归参数向量、设计矩阵、误差向量，并写出“数据版”的多元线性回归模型；
  - 最小二乘法估计回归参数向量，并与 μ_i 进行比较；
  - 利用 F 检验，对所构造对多元线性回归模型进行模型显著性检验，并与方差分析的结果进行比较。

令 $y_{ij} = \sum_{j=k}^{a} \mu_k x_k + \varepsilon_{ij}$，其中 $x_k = \begin{cases} 1, & k = i \\ 0, & k \neq i \end{cases}$.

线性回归模型：

$y = (y_{11}, y_{12}, \ldots, y_{am})^T$.

$\beta = (\mu_1, \mu_2, \ldots, \mu_a)^T$.

$X = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix}_{(a\times m)\times a}$

$e = (\varepsilon_{11}, \varepsilon_{12}, \ldots, \varepsilon_{am})^T$.

$y = X\beta + e$.

最小二乘估计：

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

$$= \begin{pmatrix} 1^T 1 & & \\ & \ddots & \\ & & 1^T 1 \end{pmatrix}^{-1} \begin{pmatrix} 1^T & & \\ & \ddots & \\ & & 1^T \end{pmatrix} \begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{am} \end{pmatrix}$$

$$= \frac{1}{m} \begin{pmatrix} y_{11} + y_{12} + \cdots + y_{1m} \\ y_{21} + y_{22} + \cdots + y_{2m} \\ \vdots \\ y_{a1} + y_{a2} + \cdots + y_{am} \end{pmatrix}$$

$$\hat{\mu}_i = \frac{1}{m} \sum_{j=1}^{m} y_{ij}.$$

显著性检验：

$H_0: \mu_1 = \mu_2 = \cdots = \mu_a = 0$

vs $H_1: \exists_{i \in \{1,2,\ldots,a\}} \ \mu_i \neq 0.$

检验统计量：

$$F_0 = \frac{SS_R/(a-1)}{SS_E/(n-a)}.$$

$\therefore$ one-way ANOVA 可看作多元线性回归模型.