D-Day: Oct 13 9:30 AM

Q1:

在回归分析中，对数据进行变换

$$\tilde{y}_i = \frac{y_i - c_1}{d_1}, \quad \tilde{x}_i = \frac{x_i - c_2}{d_2}, \quad i = 1, 2, \cdots, n,$$

其中，选取c_1, c_2, d_1, d_2为适当的常数。请回答：

- 试建立由原始数据和变换后数据得到的最小二乘估计、总偏差平方和、回归平方和以及残差平方和之间的关系；
- 证明：由原始数据和变换后数据得到的F统计量的值保持不变。

Q2:

对给定的n组数据$(x_i, y_i), i = 1, 2, \cdots, n$, 若我们关心的是$y$如何依赖$x$的取值而变动，则可以建立回归方程

$$\hat{y} = a + bx.$$

反之，若我们关心的是x如何依赖y的取值而变动,则可以建立另一个回归方程

$$\hat{x} = c + dy.$$

试问这两条直线在直角坐标系中是否重合？为什么？若不重合,它们有无交点？若有,试给出交点的坐标。

Q3:

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

Q4:

在一个多元线性回归模型中，响应变量y_i的回归值为

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_p x_p.$$

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

Q5:

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$\hat{\beta}=(\hat{\beta}_0,\hat{\beta}_1,\cdots,\hat{\beta}_p)'$。

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$\tilde{\beta}=(\tilde{\beta}_0,\tilde{\beta}_1,\cdots,\tilde{\beta}_p)'$。

请回答：

- 这两个估计$\tilde{\beta}$和$\hat{\beta}$之间有什么关系?
- 求$\tilde{\beta}$的期望和方差。

Q6:

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\bm{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。