统计方法与机器学习 理论作业2 参考答案

1

（1）对于变换后的数据

$$
\begin{aligned}
\bar{x}' &= \frac{1}{n}\sum_{i=1}^{n}\tilde{x}_i \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{x_i - c_2}{d_2} \\
&= \frac{1}{nd_2}\left(\sum_{i=1}^{n}x_i - nc_2\right) \\
&= \frac{\bar{x} - c_2}{d_2}
\end{aligned} \tag{1}
$$

同理

$$
\bar{y}' = \frac{\bar{y} - c_1}{d_1} \tag{2}
$$

因此

$$
\begin{aligned}
\tilde{l}_{xx} &= \sum_{i=1}^{n}\left(\tilde{x}_i - \bar{x}'\right)^2 \\
&= \sum_{i=1}^{n}\left(\frac{x_i - c_2}{d_2} - \frac{\bar{x} - c_2}{d_2}\right)^2 \\
&= \frac{1}{d_2^2}l_{xx}
\end{aligned} \tag{3}
$$

同理

$$
\begin{aligned}
\tilde{l}_{xy} &= \sum_{i=1}^{n}\left(\frac{x_i - c_2}{d_2} - \frac{\bar{x} - c_2}{d_2}\right)\left(\frac{y_i - c_1}{d_1} - \frac{\bar{y} - c_1}{d_1}\right) \\
&= \sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{d_2}\right)\left(\frac{y_i - \bar{y}}{d_1}\right) \\
&= \frac{1}{d_1 d_2}l_{xy}
\end{aligned} \tag{4}
$$

因此

$$
\begin{aligned}
\hat{\beta}_1' &= \tilde{l}_{xx}^{-1}\tilde{l}_{xy} \\
&= \frac{d_2^2}{l_{xx}}\cdot\frac{l_{xy}}{d_1 d_2} \\
&= \frac{d_2}{d_1}\hat{\beta}_1
\end{aligned} \tag{5}
$$

于是

$$
\begin{aligned}
\hat{\beta}_0' &= \bar{y}' - \hat{\beta}_1'\bar{x}' \\
&= \frac{\bar{y}-c_1}{d_1} - \frac{d_2}{d_1}\hat{\beta}_1 \cdot \frac{\bar{x}-c_2}{d_2} \\
&= \frac{\bar{y}-c_1}{d_1} - \frac{\bar{x}-c_2}{d_1}\hat{\beta}_1 \\
&= \frac{1}{d_1}\left(\hat{\beta}_0 + c_2\hat{\beta}_1 - c_1\right)
\end{aligned} \tag{6}
$$

也即**变换后数据的最小二乘估计** $\hat{\beta}_0', \hat{\beta}_1'$ **和原数据的最小二乘估计** $\hat{\beta}_0, \hat{\beta}_1$ **间**的关系为

$$
\begin{cases}
\hat{\beta}_0' = \dfrac{1}{d_1}\left(\hat{\beta}_0 + c_2\hat{\beta}_1 - c_1\right) \\
\hat{\beta}_1' = \dfrac{d_2}{d_1}\hat{\beta}_1
\end{cases} \tag{7}
$$

与上面的过程类似，我们同样可以快速得到**总偏差平方和**的关系

$$
\begin{aligned}
SS_T' &= \sum_{i=1}^{n}\left(\tilde{y}_i - \bar{y}'\right)^2 \\
&= \sum_{i=1}^{n}\left(\frac{y_i-\bar{y}}{d_1}\right)^2 \\
&= \frac{1}{d_1^2}SS_T
\end{aligned} \tag{8}
$$

而由于

$$
\begin{aligned}
\hat{y}_i' &= \hat{\beta}_0' + \hat{\beta}_1'\tilde{x}_i \\
&= \frac{1}{d_1}\left(\hat{\beta}_0 + c_2\hat{\beta}_1 - c_1\right) + \frac{d_2}{d_1}\hat{\beta}_1 \cdot \frac{x_i-c_2}{d_2} \\
&= \frac{1}{d_1}\left(\hat{\beta}_0 + x_i\hat{\beta}_1 - c_1\right) \\
&= \frac{\hat{y}_i-c_1}{d_1}
\end{aligned} \tag{9}
$$

因此**回归平方和**

$$
\begin{aligned}
SS_R' &= \sum_{i=1}^{n}\left(\hat{y}_i' - \bar{y}'\right)^2 \\
&= \sum_{i=1}^{n}\left(\frac{\hat{y}_i-c_1}{d_1} - \frac{\bar{y}-c_1}{d_1}\right)^2 \\
&= \frac{1}{d_1^2}\sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2 \\
&= \frac{1}{d_1^2}SS_R
\end{aligned} \tag{10}
$$

同理，**残差平方和**

$$
\begin{aligned}
SS_E' &= \sum_{i=1}^{n}\left(y_i' - \hat{y}_i'\right)^2 \\
&= \sum_{i=1}^{n}\left(\frac{y_i - c_1}{d_1} - \frac{\hat{y}_i - c_1}{d_1}\right)^2 \\
&= \frac{1}{d_1^2}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 \\
&= \frac{1}{d_1^2} SS_E
\end{aligned} \tag{11}
$$

(2) 由（1）的结论易见

$$
F_0' = \frac{SS_R'}{SS_E'/(n-2)} = \frac{\frac{1}{d_1^2}SS_R}{\frac{1}{d_1^2(n-2)}SS_E} = \frac{SS_R}{SS_E/(n-2)} = F_0 \tag{12}
$$

即其 F 统计量保持不变

2

由最小二乘估计可知，y 关于 x 的回归方程为

$$
\hat{y} = a + bx, \begin{cases} a = \bar{y} - b\bar{x} \\ b = \dfrac{l_{xy}}{l_{xx}} \end{cases} \tag{13}
$$

x 关于 y 的回归方程为

$$
\hat{x} = c + dy, \begin{cases} c = \bar{x} - d\bar{y} \\ d = \dfrac{l_{xy}}{l_{yy}} \end{cases} \tag{14}
$$

将下式代入上式，可得其交点方程为

$$
y = a + b(c + dy) \tag{15}
$$

化简得 $(1 - bd)y = \bar{y}(1 - bd)$

当两直线重合时，该方程对一切 y 恒成立，即 $1 - bd = 0$

代入原表达式可知该条件等价于

$$
\frac{l_{xy}^2}{l_{xx}l_{yy}} = 1 \tag{16}
$$

也即相关系数

$$
r^2 = 1 \Rightarrow r = \pm 1 \tag{17}
$$

当两直线不重合时，$r \neq \pm 1$

此时易见必存在交点，且交点处

$$y=\frac{\bar{y}(1-bd)}{1-bd}=\bar{y} \tag{18}$$

代入原式得此时 $x=\bar{x}$。

故交点坐标为 $(\bar{x},\bar{y})$

3

易见

$$\begin{aligned}(\boldsymbol{I}-\boldsymbol{H})^T&=(\boldsymbol{I}-\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T)^T\\&=\boldsymbol{I}^T-\boldsymbol{X}\big((\boldsymbol{X}^T\boldsymbol{X})^{-1}\big)^T\boldsymbol{X}^T\\&=\boldsymbol{I}^T-\boldsymbol{X}\big((\boldsymbol{X}^T\boldsymbol{X})^T\big)^{-1}\boldsymbol{X}^T\\&=\boldsymbol{I}-\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\\&=\boldsymbol{I}-\boldsymbol{H}\end{aligned} \tag{19}$$

且

$$\begin{aligned}(\boldsymbol{I}-\boldsymbol{H})^2&=\boldsymbol{I}^2-\boldsymbol{H}\boldsymbol{I}-\boldsymbol{I}\boldsymbol{H}+\boldsymbol{H}^2\\&=\boldsymbol{I}-2\boldsymbol{H}+\boldsymbol{H}^2\\&=\boldsymbol{I}-2\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T+\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\\&=\boldsymbol{I}-2\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T+\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\\&=\boldsymbol{I}-\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\\&=\boldsymbol{I}-\boldsymbol{H}\end{aligned} \tag{20}$$

因此 $\boldsymbol{I}-\boldsymbol{H}$ 是一个对称且幂等的矩阵。

由于 $\boldsymbol{I}-\boldsymbol{H}$ 为幂等矩阵，故有 $\operatorname{rank}(\boldsymbol{I}-\boldsymbol{H})=\operatorname{tr}(\boldsymbol{I}-\boldsymbol{H})$

因此

$$\begin{aligned}\operatorname{rank}(\boldsymbol{I}-\boldsymbol{H})&=\operatorname{tr}(\boldsymbol{I}-\boldsymbol{H})\\&=\operatorname{tr}\boldsymbol{I}-\operatorname{tr}\boldsymbol{H}\\&=n-p-1\end{aligned} \tag{21}$$

其中 p 为自变量个数（或 $\boldsymbol{X}$ 的行维数减一）

4

该回归模型可写为

$$\hat{\boldsymbol{y}}=\boldsymbol{X}\hat{\boldsymbol{\beta}} \tag{22}$$

故待证结论

$$\sum_{i=1}^{n}(y_i-\hat{y}_i)=0\Leftrightarrow\mathbf{1}^T\left(\boldsymbol{y}-\hat{\boldsymbol{y}}\right)=0 \tag{23}$$

（其中 $\mathbf{1}$ 为元素全为 1 的列向量）

由回归系数的最小二乘估计解 $\hat{\boldsymbol{\beta}}=(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$ 可知

$$
\begin{aligned}
\mathbf{1}^T\left(\boldsymbol{y}-\hat{\boldsymbol{y}}\right) &= \mathbf{1}^T\left(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}\right)\\
&= \mathbf{1}^T\left(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}\right)\\
&= \mathbf{1}^T\left(\boldsymbol{y}-\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}\right)\\
&= \mathbf{1}^T\left(\boldsymbol{I}-\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\right)\boldsymbol{y}\\
&= \left(\mathbf{1}^T-\mathbf{1}^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\right)\boldsymbol{y}\\
&= \left(\mathbf{1}^T-\left(\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\cdot\mathbf{1}\right)^T\right)\boldsymbol{y}
\end{aligned}
\tag{24}
$$

注意到 $\mathbf{1}$ 即为 $\boldsymbol{X}$ 的第一列，因此若令 $\boldsymbol{c}=(1,0,\cdots,0)^T$，则 $\mathbf{1}=\boldsymbol{X}\boldsymbol{c}$

于是

$$
\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\cdot\mathbf{1}=\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{c}=\boldsymbol{X}\boldsymbol{c}=\mathbf{1}
\tag{25}
$$

因此

$$
\mathbf{1}^T\left(\boldsymbol{y}-\hat{\boldsymbol{y}}\right)=\left(\mathbf{1}^T-\mathbf{1}^T\right)\boldsymbol{y}=0
\tag{26}
$$

也即

$$
\sum_{i=1}^{n}(y_i-\hat{y}_i)=0
\tag{27}
$$

5

(1) 记中心化后的因变量向量为 $\boldsymbol{y}^*$，标准化后的自变量矩阵为 $\boldsymbol{X}^{**}=(\mathbf{0}\quad\boldsymbol{X}_s)$

令 $\boldsymbol{A}_s=\left(\boldsymbol{X}_s^T(\boldsymbol{I}_n-\boldsymbol{H}_{1_n})\boldsymbol{X}_s\right)^{-1}$

由题1结论可知，$\boldsymbol{I}_n-\boldsymbol{H}_{1_n}$ 为对称幂等矩阵

故

$$
\begin{aligned}
\boldsymbol{A}_s &= \left(\boldsymbol{X}_s^T(\boldsymbol{I}_n-\boldsymbol{H}_{1_n})\boldsymbol{X}_s\right)^{-1}\\
&= \left(\boldsymbol{X}_s^T(\boldsymbol{I}_n-\boldsymbol{H}_{1_n})(\boldsymbol{I}_n-\boldsymbol{H}_{1_n})\boldsymbol{X}_0\boldsymbol{L}\right)^{-1}\\
&= \left(\boldsymbol{X}_s^T(\boldsymbol{I}_n-\boldsymbol{H}_{1_n})\boldsymbol{X}_0\boldsymbol{L}\right)^{-1}\\
&= \left(\boldsymbol{X}_s^T\boldsymbol{X}_s\right)^{-1}
\end{aligned}
\tag{28}
$$

又由于

$$
\mathbf{1}_n^T\boldsymbol{X}_s=\mathbf{1}_n^T(\boldsymbol{I}_n-\boldsymbol{H}_{1_n})\boldsymbol{X}_0\boldsymbol{L}=0
\tag{29}
$$

因此

$$
\begin{aligned}
\tilde{\boldsymbol{\beta}} &= \left((\boldsymbol{X}^{**})^T(\boldsymbol{X}^{**})\right)^{-1}(\boldsymbol{X}^{**})^T\boldsymbol{y}^* \\
&= \begin{pmatrix} n^{-1}\mathbf{1}_n^T + n^{-2}\mathbf{1}_n^T\boldsymbol{X}_s\boldsymbol{A}_s\boldsymbol{X}_s^T\mathbf{1}_n\mathbf{1}_n^T - n^{-1}\mathbf{1}_n^T\boldsymbol{X}_s\boldsymbol{A}_s\boldsymbol{X}_s^T \\ -n^{-1}\boldsymbol{A}_s\boldsymbol{X}_s^T\mathbf{1}_n\mathbf{1}_n^T + \boldsymbol{A}_s\boldsymbol{X}_s^T \end{pmatrix}\boldsymbol{y}^* \\
&= \begin{pmatrix} n^{-1}\mathbf{1}_n^T \\ -n^{-1}\boldsymbol{A}_s\boldsymbol{X}_s^T\mathbf{1}_n\mathbf{1}_n^T + \boldsymbol{A}_s\boldsymbol{X}_s^T \end{pmatrix}(\boldsymbol{I}_n - \boldsymbol{H}_{1_n})\boldsymbol{y} \\
&= \begin{pmatrix} n^{-1}\mathbf{1}_n^T(\boldsymbol{I}_n - \boldsymbol{H}_{1_n})\boldsymbol{y} \\ \left(-n^{-1}\boldsymbol{A}_s\boldsymbol{X}_s^T\mathbf{1}_n\mathbf{1}_n^T + \boldsymbol{A}_s\boldsymbol{X}_s^T\right)(\boldsymbol{I}_n - \boldsymbol{H}_{1_n})\boldsymbol{y} \end{pmatrix} \\
&= \begin{pmatrix} 0 \\ \boldsymbol{A}_s\boldsymbol{X}_s^T(\boldsymbol{I}_n - \boldsymbol{H}_{1_n})\boldsymbol{y} \end{pmatrix} \\
&= \begin{pmatrix} 0 \\ \left(\boldsymbol{X}_s^T\boldsymbol{X}_s\right)^{-1}\boldsymbol{X}_s^T\boldsymbol{y}^* \end{pmatrix} \\
&= \begin{pmatrix} 0 \\ \sqrt{L_{yy}}\left(\boldsymbol{X}_s^T\boldsymbol{X}_s\right)^{-1}\boldsymbol{X}_s^T\boldsymbol{y}^{**} \end{pmatrix} \\
&= \begin{pmatrix} 0 \\ \sqrt{L_{yy}}\hat{\boldsymbol{\beta}}_{s,slope} \end{pmatrix} \\
&= \begin{pmatrix} 0 \\ \boldsymbol{L}^{-1}\hat{\boldsymbol{\beta}}_{slope} \end{pmatrix}
\end{aligned} \tag{30}
$$

由于

$$
\hat{\boldsymbol{\beta}} = \begin{pmatrix} \beta_0 \\ \hat{\boldsymbol{\beta}}_{slope} \end{pmatrix} \tag{31}
$$

因此

$$
\tilde{\boldsymbol{\beta}} = \begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{L}^{-1} \end{pmatrix}\hat{\boldsymbol{\beta}} \tag{32}
$$

(2) 由（1）的结论易得

$$
\begin{aligned}
E\left(\tilde{\boldsymbol{\beta}}\right) &= E\left(\begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{L}^{-1} \end{pmatrix}\hat{\boldsymbol{\beta}}\right) \\
&= \begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{L}^{-1} \end{pmatrix}E\left(\hat{\boldsymbol{\beta}}\right) \\
&= \begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{L}^{-1} \end{pmatrix}\boldsymbol{\beta}
\end{aligned} \tag{33}
$$

且

$$
\begin{aligned}
Var\left(\tilde{\boldsymbol{\beta}}\right) &= Var\left(\begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{L}^{-1} \end{pmatrix}\hat{\boldsymbol{\beta}}\right) \\
&= \begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{L}^{-1} \end{pmatrix}Var\left(\hat{\boldsymbol{\beta}}\right)\begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{L}^{-T} \end{pmatrix} \\
&= \sigma^2\begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{L}^{-1} \end{pmatrix}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{L}^{-T} \end{pmatrix}
\end{aligned} \tag{34}
$$

6

受到矩阵形式下回归模型的启发，我们引入虚拟变量 $x_1, x_2, \cdots . x_n$，使得

$$y_{ij} = \sum_{j=k}^{a} \mu_k x_k + \varepsilon_{ij} \tag{35}$$

其中

$$x_k = \begin{cases} 1, k = i \\ 0, k \neq i \end{cases} \tag{36}$$

于是我们就可以写出对应的线性回归模型：

响应变量 $\boldsymbol{y} = (y_{11}, y_{12}, \cdots, y_{am})^T$

参数向量 $\boldsymbol{\beta} = (\mu_1, \mu_2, \cdots, \mu_a)^T$

自变量矩阵

$$\boldsymbol{X} = \begin{pmatrix} \mathbf{1} & & & \\ & \mathbf{1} & & \\ & & \ddots & \\ & & & \mathbf{1} \end{pmatrix}_{(a \times m) \times a} \tag{37}$$

误差矩阵 $\boldsymbol{e} = (\varepsilon_{11}, \varepsilon_{12}, \cdots, \varepsilon_{am})^T$

回归模型为 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$

故而由线性回归的最小二乘估计可知

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \left(\boldsymbol{X}^T \boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \boldsymbol{y} \\ &= \begin{pmatrix} \mathbf{1}^T\mathbf{1} & & & \\ & \mathbf{1}^T\mathbf{1} & & \\ & & \ddots & \\ & & & \mathbf{1}^T\mathbf{1} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{1}^T & & & \\ & \mathbf{1}^T & & \\ & & \ddots & \\ & & & \mathbf{1}^T \end{pmatrix} \begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{am} \end{pmatrix} \\ &= \frac{1}{m} \begin{pmatrix} y_{11} + y_{12} + \cdots + y_{1m} \\ y_{21} + y_{22} + \cdots + y_{2m} \\ \vdots \\ y_{a1} + y_{a2} + \cdots + y_{am} \end{pmatrix} \end{aligned} \tag{38}$$

而由于 $\boldsymbol{\beta} = (\mu_1, \mu_2, \cdots, \mu_a)^T$

因此

$$\hat{\mu}_i = \frac{1}{m} \sum_{j=1}^{m} y_{ij} \tag{39}$$

对该回归模型进行显著性检验，则假设检验问题为

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_a = 0 \ \mathbf{v.s} \ H_1 : \exists i \in \{1, 2, \cdots, a\} \ \mathbf{s.t} \ \mu_i \neq 0 \tag{40}$$

于是其检验统计量即为

$$F_0 = \frac{SS_R/(a-1)}{SS_E/(n-a)} \tag{41}$$

此即为单因素方差分析的检验统计量。

由此可见，单因子方差分析模型可以看作一种带有哑元（Dummy Variable）的多元线性回归模型。