中国科学技术大学数学科学学院

2021～2022 学年第 2 学期期中考试试卷

■A 卷　　□B 卷

课程名称　回归分析　　　　课程编号 00136301

考试时间　2022/5/6　　　　考试形式　闭卷

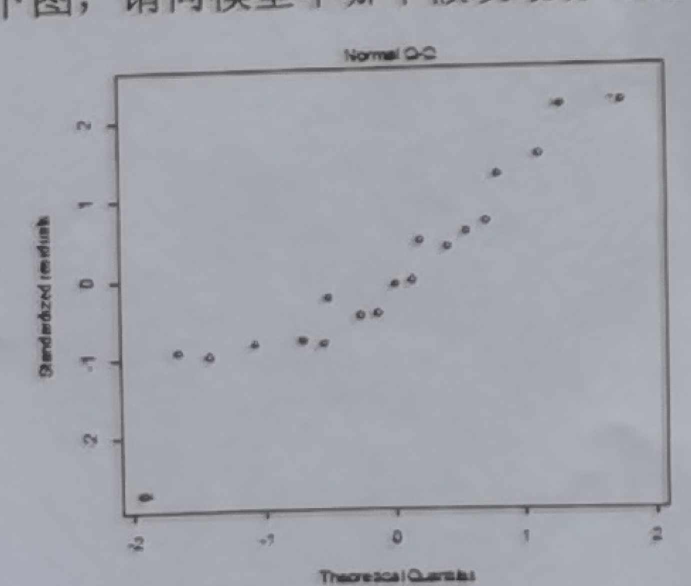
姓　名________　学　号________　学　院________

题 号	一	1	2	3	4	5	6	总 分
得 分								

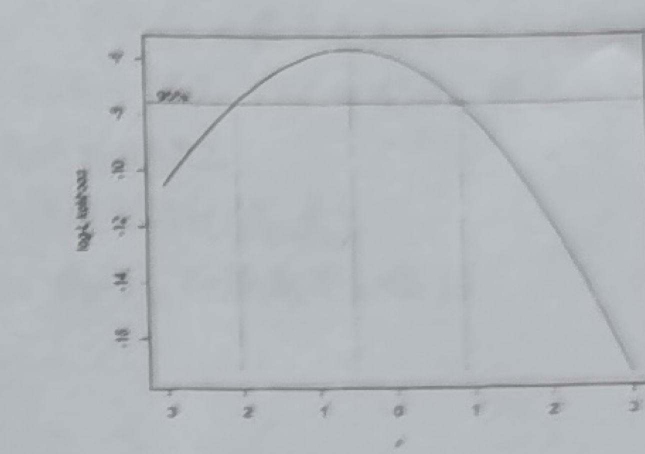
一　简答题（共 30 分）

I　（5 分）考虑简单线性回归模型 $Y=\beta_0+\beta_1X+e$，$e\sim N(\mu,\sigma^2I)$，OLS 估计$\hat\beta_0$ 是否是β_0的最佳线性无偏估计？说明理由。

II　（5 分）考虑简单线性回归模型 $Y=\beta_0+\beta_1X+e$，随机误差项$\mathbf{e}$服从正态分布，$E\mathbf{e}=\mathbf{0}$，$\mathrm{Cov}(\mathbf{e})=\sigma^2\mathbf{I}$。观察下图，请问模型中哪个假设最有可能不成立？



III　（5 分）观察如下由响应变量 Y 和预测变量X的观测值经 Box-Cox 变换后其 log 似然关于参数λ的函数图，请对 Y 给出合适的 Box-Cox 变换，进而给出变换后的线性回归模型（包括模型假设）。



IV　（5 分）考虑无截距简单线性回归模型 $Y=\beta X+e$，$e_1,\cdots,e_n$ 独立且$Ee_i=0$，$\mathrm{Var}(e_i)=c\sigma_i^2$，$i=1,\cdots,n$. 其中$\sigma_i^2$已知，$c>0$未知。请给出$\beta$的唯一最小方差无偏估计。

V （10 分）考虑线性回归模型$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{e}$，$\mathbf{e} \sim N(\mu, \sigma^2 I_n)$。记$h_{ii}, \hat{e}_i$ 分别为第i个案例的位势和 OLS 拟合后第i个案例的残差，$\hat{\sigma}^2_{(i)}$是剔除第i个案例后，剩余$n-1$个案例 OLS 拟合得到的残差均方。证明：

$$r_i^* = \frac{\hat{e}_i}{\hat{\sigma}_{(i)}\sqrt{1-h_{ii}}} \sim t_{n-p-1}.$$

提示：记$\boldsymbol{x}_i^T$为$\boldsymbol{X}$的第i行，$\widehat{\boldsymbol{\beta}}$和$\widehat{\boldsymbol{\beta}}_{(i)}$分别为用$n$个案例和剔除第$i$个案例后剩余$n-1$个案例拟合得到的$\boldsymbol{\beta}_{p\times 1}$的 OLS 估计，满足 $\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(i)} = \frac{\hat{e}_i}{1-h_{ii}}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i$.

二 计算题 （共 70 分，写出必要的公式和计算过程）

1. 研究某种材料单位体积的质量指标Y与成分含量X之间的关系，现考虑如下线性回归模型，

$$Y = \beta_0 + \beta_1 X + e,\ \mathbf{e} \sim N(0, \sigma^2 I_n).$$

已知样本容量$n = 25$，$SXX = 150$，$SXY = 450$，样本均值$\bar{x} = 3$，$\bar{y} = 8$ ，利用 OLS 法拟合得到残差平方和 $RSS = 147.2$.

(a) （10 分）求(β_0, β_1)的 OLS 估计$(\hat{\beta}_0, \hat{\beta}_1)$.

(b) （10 分）请检验H_0：$\beta_0=1$，检验水平$\alpha=0.10$.

2. （10 分）考虑某种材料单位体积的质量指标Y与成分含量X之间的关系，已知样本容量$n = 50$，$SXX = 50$，$SXY = -30$，$SYY = 100$，样本均值$\bar{x} = 3$，$\bar{y} = 2$。现利用 OLS 法拟合如下线性回归模型

$$Y = \beta_0 + \beta_1 X + e.$$

（a） 请给出测定系数R^2；

（b） 假设$\beta_0 = 0$，请给出此时模型的 OLS 拟合回归方程并给出合适的测定系数R^2。

3. （10 分）考虑某种材料单位体积的质量指标Y与成分含量X之间的关系，已知样本容量$n = 20$，$SXX = 80$，样本均值$\bar{x} = 2$。现利用 OLS 法拟合得到如下线性拟合回归方程

$$\hat{Y} = 10 - 2X\ .$$

已知残差均方$\hat{\sigma}^2 = 9$， 第 3 个案例数据$x_3 = 6$，$y_3 = 10$，请检验第 3 个案例是否是强影响案例？检验水平$\alpha = 0.10$.

4. （10 分）调查一个家庭每年的生活支出（Y）关于本地人口密度（X_1），人口总数（X_2）以及人均收入（X_3）的线性回归关系：$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + e$。现有一组样本，利用 R 语言中 anova() 函数对数据进行方差分析得到下表：

来源	自由度（d.f.）	平方和(SS)	均方(MS)	F
关于X_1的回归	1	3060	δ_3	
经X_1调整后关于X_2的回归	1	300	δ_4	
经X_1,X_2调整后关于X_3的回归	1	δ_1	δ_5	
残差	δ_2	2400	δ_6	
总的	19	6080		

部分数据被遮挡。请对如下两个模型比较问题进行 F 检验，并分别回答应选择哪一个模型进行拟合。检验水平$\alpha = 0.10$.

(a) **模型1**：$Y = \beta_0 + e$

模型2：$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e$

(b) **模型3**：$Y = \beta_0 + \beta_2 X_2 + \beta_3 X_3 + e$

模型4：$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + e$

5. （10 分）一公司为了研究产品的营销策略，对产品的销售情况进行了调查。设Y表示某市该产品的人均购买量，X_1表示人均收入，X_2表示人口密度，共 n=25 个城市。预测变量的样本均值和校正叉积和如下：

$$m_1 = 9,\quad m_2 = 8,\quad SX_1X_1 = 100,\quad SX_1X_2 = 50,\quad SX_2X_2 = 100 .$$

OLS 拟合回归方程

$$\hat{Y} = 13 + X_1 - 2X_2$$

残差平方和 $RSS = 660$。现利用上述回归方程对第 26 个城市的人均购买量y_{21}进行预测，已知$x_{26,1} = 11$, $x_{26,2} = 6$。假设这 26 个城市样本的随机误差项$\vec{e} \sim N_{26}(\vec{0}, \sigma^2 I_{26})$，求这第 26 个城市的人均购买量置信度为 0.9 的预测置信区间。

6. （10 分）调查一个家庭每年的生活支出（Y）关于本地人口密度（X_1），人口总数（X_2）以及人均收入（X_3）的线性回归关系，共 n=22 个样本，预测变量和响应变量样本均值

$$m_1 = 8,\quad m_2 = 3,\quad m_3 = 2,\quad \bar{y} = 220$$

X_1, X_2, X_3的校正叉积矩阵及其逆矩阵

$$X_c^T X_c = \begin{bmatrix} 2140 & 77 & 354 \\ 77 & 25 & 17 \\ 354 & 17 & 112 \end{bmatrix},\ (X_c^T X_c)^{-1} = 10^{-3} \times \begin{bmatrix} 1 & -1 & -3 \\ -1 & 46 & -4 \\ -3 & -4 & 19 \end{bmatrix}$$

模型$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + e$的 OLS 拟合回归方程为

$$\hat{Y} = 170 + X_1 + 2X_2 + X_3$$

残差均方$\hat{\sigma}^2 = 4$。假设模型随机误差项$\vec{e} \sim N_n(\vec{0}, \sigma^2 I_n)$，请检验 H_0：$\beta_1 = \beta_2 = \beta_3$，检验水平$\alpha = 0.10$.