$$E(X'AX)=\mu'A\mu+\mathrm{tr}(A\Sigma)$$

引理 3 多元正态向量的任意线性组合仍然是正态的。

引理 4 假设 $X\sim N_n(0,\ \Sigma)$，Σ 是正定的，则

$$X'\Sigma^{-1}X\sim\chi^2(n)$$

引理 5 假设 $X_1\sim\chi^2(n)$，$X_2\sim\chi^2_m$ 且它们相互独立，则

$$X_1+X_2\sim\chi^2(n+m)$$

引理 6 假设 $X\sim N_n(0,\ I_n)$，A 是一 $n\times n$ 的幂等矩阵，其秩为 r，则

$$X'AX\sim\chi^2(r)$$

引理 7 假设 $X\sim N_n(0,\ I_n)$，A 是一 $n\times n$ 的对称矩阵，B 是一 $m\times n$ 的矩阵，且 $BA=0$，则 BX 与 $X'AX$ 相互独立。

引理 8 假设 $X\sim N_n(0,\ I_n)$，A 和 B 是一 $n\times n$ 的对称矩阵，且 $BA=0$，则 $X'BX$ 与 $X'AX$ 相互独立。

习 题 三

1. 试对二元线性回归模型 $y_i=\beta_0+\beta_1x_{1i}+\beta_2x_{2i}+\varepsilon_i(i=1,\ 2,\ \cdots,\ n)$ 作回归分析。要求：

(1) 求出未知参数 β_0、β_1、β_2 的最小二乘估计量 $\hat{\beta}_0$、$\hat{\beta}_1$、$\hat{\beta}_2$。

(2) 求出随机误差项 u 的方差 σ^2 的无偏估计量。

(3) 对样本回归方程作拟合优度检验。

(4) 对总体回归方程的显著性进行 F 检验。

(5) 对 β_1、β_2 的显著性进行 t 检验。

(6) 当 $x_0=(1,\ x_{10},\ x_{20})'$ 时，写出 $E(y_0\mid x_0)$ 和 y_0 的置信度为 95%的预测区间。

2. 在多元线性回归分析中，称

$$R=\sqrt{S_R/S_T}=\sqrt{1-S_E/S_T}$$

为自变量 x_1，$x_2\cdots$，x_p 对 y 的复相关系数。试说明复相关系数的统计意义，特别地，当 $R=0$ 时，是否排斥 y 与 x_1，$x_2\cdots$，x_p 之间，存在某种很密切的非线性关系？

3. 设有模型 $y=\beta_0+\beta_1x_1+\beta_2x_2+\varepsilon$，试在下列条件下：

(1) $\beta_1+\beta_2=1$；

(2) $\beta_1=\beta_2$。

分别求出β_1和β_2的最小二乘估计量。

4. 根据100对(x_i, y_i)的观察值计算出：

$$\sum \dot{x}_1^2=12, \sum \dot{x}\dot{y}=-9, \sum \dot{y}^2=30$$

要求：

(1) 求出一元模型 $y=\beta_0+\beta_1x_1+u$ 中的β_1的最小二乘估计量及其相应的标准差估计量。

(2) 后来发现y还受x_2的影响，于是，将一元模型改为二元模型 $y=\alpha_0+\alpha_1x_1+\alpha_2x_2+v$，收集$x_2$的相应观察值并计算出：

$$\sum \dot{x}_2^2=6, \sum \dot{x}_2\dot{y}=8, \sum \dot{x}_1\dot{x}_2=2$$

求二元模型中的α_1、α_2的最小二乘估计量及其相应的标准差估计量。

(3) 一元模型中的$\hat{\beta}_1$与二元模型中的$\hat{\alpha}_1$是否相等？为什么？

5. 考虑以下预测的回归方程：

$$\hat{Y}_t=-120+0.10F_t+5.33RS_t, \bar{R}^2=0.50$$

其中，Y_t——第t年的玉米产量(蒲式耳/亩)；

F_t——第t年的施肥强度(磅/亩)；

RS_t——第t年的降雨量(英寸)。

要求：

(1) 从F和RS对Y的影响方面，说出本方程中系数0.10和5.33的含义。

(2) 常数项-120是否意味着玉米的负产量可能存在？

(3) 假定β_F的真实值为0.40，则估计值是否有偏？为什么？

(4) 假定该方程并不满足所有的古典模型假设，即并不是最佳线性无偏估计值，则是否意味着β_{RS}的真实值绝对不等于5.33？为什么？

6. 已知线性回归模型$\boldsymbol{Y}=\boldsymbol{X}\beta+\varepsilon$式中$\varepsilon\sim(0,\sigma^2\boldsymbol{I})$，$n=13$且$k=3$($n$为样本容量，$k$为参数的个数)，由二次型$(\boldsymbol{Y}-\boldsymbol{X}\beta)'(\boldsymbol{Y}-\boldsymbol{X}\beta)$的最小化得到如下线性方程组：

$$\begin{aligned}\hat{\beta}_1+2\hat{\beta}_2+\hat{\beta}_3&=3\\2\hat{\beta}_1+5\hat{\beta}_2+\hat{\beta}_3&=9\\\hat{\beta}_1+\hat{\beta}_2+6\hat{\beta}_3&=-8\end{aligned}$$

要求：

(1) 把问题写成矩阵向量的形式；用求逆矩阵的方法求解之。

(2) 如果$\boldsymbol{Y}'\boldsymbol{Y}=53$，求$\hat{\sigma}^2$。

(3) 求出$\hat{\beta}$的方差—协方差矩阵。

7. 经研究发现，学生用于购买书籍及课外读物的支出与本人受教育年限和其家庭收入水平有关，对18名学生进行调查的统计资料如下表所示：

学生序号	购买书籍及课外读物的支出 y(元/年)	受教育年限 x_1(年)	家庭月可支配收入 x_2(元/月)
1	450.5	4	171.2
2	507.7	4	174.2
3	613.9	5	204.3
4	563.4	4	218.7
5	501.5	4	219.4
6	781.5	7	240.4
7	541.8	4	273.5
8	611.1	5	294.8
9	1 222.1	10	330.2
10	793.2	7	333.1
11	660.8	5	366.0
12	792.7	6	350.9
13	580.8	4	357.9
14	612.7	5	359.0
15	890.8	7	371.9
16	1 121.0	9	435.3
17	1 094.2	8	523.9
18	1 253.0	10	604.1

要求：

(1) 试求出学生购买书籍及课外读物的支出 y 与受教育年限 x_1 和家庭收入水平 x_2 的估计的回归方程 $\hat{y}=\hat{\beta}_0+\hat{\beta}_1x_1+\hat{\beta}_2x_2$。

(2) 对 β_1、β_2 的显著性进行 t 检验，计算 R^2。

(3) 假设有一学生的受教育年限 $x_1=10$ 年，家庭收入水平 $x_2=480$ 元/月，试预测该学生全年购买书籍及课外读物的支出，并求出相应的预测区间 ($\alpha=0.05$)。

8. 下表给出三变量模型的回归结果：

方差来源	平方和	自由度	均方
回归	65 965	—	—
剩余	—	—	—
总和	66 042	14	

要求：

(1) 样本容量是多少？

(2) 求 S_E。

(3) S_R 和 S_E 的自由度各是多少？

(4) 求 R^2。

(5) 检验假设：x_2 和 x_3 对 y 无影响。你用什么假设检验？为什么？

(6) 根据以上信息，你能否确定 x_2 和 x_3 各自对 y 的贡献？

9. 研究同一地区土壤内所含植物可给态磷的情况，得到 18 组数据如下，其中，

x_1：土壤内所含无机磷浓度；

x_2：土壤内溶于 K_2CO_3 溶液并受溴化物水解的有机磷；

x_3：土壤内溶于 K_2CO_3 溶液但不溶于溴化物的有机磷；

y：栽在20℃土壤内的玉米中的可给态磷。

单位为百万分之一，已知 y 对 x_1、x_2、x_3 存在线性相关，求它们的回归方程，并进行检验。

土壤样本	x_1	x_2	x_3	y
1	0.4	53	158	64
2	0.4	23	163	60
3	3.1	19	37	71
4	0.6	34	157	61
5	4.7	24	59	54
6	1.7	65	123	77
7	9.4	44	46	81
8	10.1	31	117	93
9	11.6	29	173	93
10	12.6	58	112	51
11	10.9	37	111	76

(续表)

土壤样本	x_1	x_2	x_3	y
12	23.1	46	114	96
13	23.1	50	134	77
14	21.6	44	73	93
15	23.1	56	168	95
16	1.9	36	143	54
17	26.8	58	202	168
18	29.9	51	124	99

10. 下面给出依据 15 个观察值计算得到的数据：

$$\bar{y} = 367.693,\ \bar{x}_2 = 402.760,\ \bar{x}_3 = 8.0,\ \sum y_i^2 = 66\,042.269$$

$$\sum x_{2i}^2 = 84\,855.096,\ \sum x_{3i}^2 = 280.0,\ \sum y_i x_{2i} = 74\,778.346$$

$$\sum y_i x_{3i} = 4\,250.9,\ \sum x_{2i} x_{3i} = 4\,796.0$$

要求：

(1) 估计三个多元回归系数。

(2) 估计它们的标准差；并求出 R^2。

(3) 估计 β_2、β_3 置信度为 95%的置信区间。

(4) 在 $\alpha = 0.05$ 下，检验估计的每个回归系数的统计显著性(双边检验)。

(5) 在 $\alpha = 0.05$ 下，并给出方差分析表。

11. 考虑以下方程(括号内为估计标准差)：

$$\begin{array}{l} \hat{W}_t = 8.562 + \underset{(0.080)}{0.364P_t} + \underset{(0.072)}{0.004P_{t-1}} - \underset{(0.658)}{2.560U_t} \\ \qquad n = 19 \qquad R^2 = 0.873 \end{array}$$

其中，W —— t 年的每位雇员的工资和薪水；

P —— t 年的物价水平；

U —— t 年的失业率。

要求：

(1) 对个人收入估计的回归系数进行假设检验。

(2) 讨论 P_{t-1} 在理论上的正确性，对本模型的正确性进行讨论；P_{t-1} 是否应从方程中删除？为什么？

12. 下表是某种商品的需求量、价格和消费者收入10年的时间序列资料：

年　份	1	2	3	4	5	6	7	8	9	10
需求量(吨) y	59 190	65 450	62 360	64 700	67 400	64 440	68 000	72 400	75 710	70 680
价　格(元) x_1	23.56	24.44	32.07	32.46	31.15	34.14	35.30	38.70	39.63	46.68
收　入(元) x_2	76 200	91 200	106 700	111 600	119 000	129 200	143 400	159 600	180 000	193 000

要求：

(1) 已知商品需求量 y 是其价格 x_1 和消费者收入 x_2 的函数，试求 y 对 x_1 和 x_2 的回归方程 $\hat{y}=\hat{\beta}_0+\hat{\beta}_1x_1+\hat{\beta}_2x_2$。

(2) 求 y 的总变差中未被 x_1 和 x_2 解释的部分，并对回归方程进行显著性检验。

(3) 对回归参数 $\hat{\beta}_1$、$\hat{\beta}_2$ 进行显著性 t 检验。

13. 逐步回归的基本思想是什么？试用 Hald 数据来说明逐步回归法。这里给出 Hald 水泥问题中各变量的观测数据，列出了所有可能的回归，其中，

x_1：水泥中 $3CaO\cdot Al_2O_3$ 的含量(%)；

x_2：水泥中 $3CaO\cdot SiO_2$ 的含量(%)；

x_3：水泥中 $4CaO\cdot Al_2O_3\cdot Fe_2O_3$ 的含量(%)；

x_4：水泥中 $2CaO\cdot SiO_2$ 的含量(%)；

x_5：y = 每克水泥释放的热量(卡路里)。

原始数据和(或)变换后的数据

序号	x_1	x_2	x_3	x_4	x_5
1	7.000 000 00	26.000 000 00	6.000 000 00	60.000 000 00	78.500 000 00
2	1.000 000 00	29.000 000 00	15.000 000 00	52.000 000 00	74.300 000 00
3	11.000 000 09	56.000 000 00	8.000 000 00	20.000 000 00	104.300 000 00
4	11.000 000 00	31.000 000 00	8.000 000 00	47.000 000 00	87.600 000 00
5	7.000 000 00	52.000 000 00	6.000 000 00	33.000 000 00	95.900 000 00
6	11.000 000 00	55.000 000 00	9.000 000 00	22.000 000 00	109.200 000 00
7	3.000 000 00	71.000 000 00	17.000 000 00	6.000 000 00	102.700 000 00
8	1.000 000 00	31.000 000 00	22.000 000 00	44.000 000 00	72.500 000 00

（续表）

序号	x_1	x_2	x_3	x_4	x_5
9	2.000 000 00	54.000 000 00	18.000 000 00	22.000 000 00	93.100 000 00
10	21.000 000 00	47.000 000 00	4.000 000 00	26.000 000 00	115.900 000 00
11	1.000 000 00	40.000 000 00	23.000 000 00	34.000 000 00	83.800 000 00
12	11.000 000 00	66.000 000 00	9.000 000 00	12.000 000 00	113.300 000 00
13	10.000 000 00	68.000 000 00	8.000 000 00	12.000 000 00	109.400 000 00

14. 为了提高煤的净化效率，收集了一些有关的数据，请你提出一个提高净化率的方案。数据如下：

y：净化后煤溶液中所含杂质的重量(是衡量净化效率的指标)；

x_1：输入净化过程的溶液所含的煤及杂质的比；

x_2：溶液的 Ph 值；

x_3：溶液的流量。

净化煤溶液数据

序　号	x_1	x_2	x_3	y
1	1.50	6.00	1 315	243
2	1.50	6.00	1 315	261
3	1.50	9.00	1 890	244
4	1.50	9.00	1 890	285
5	2.00	7.50	1 575	202
6	2.00	7.50	1 575	180
7	2.00	7.50	1 575	183
8	2.00	7.50	1 575	207
9	2.50	9.00	1 315	216
10	2.50	9.00	1 315	160
11	2.50	6.00	1 890	104
12	2.50	6.00	1 890	110

15. 证明第六节关于 $\hat{y}_0$ 的六个性质。