习　　题

1. 在模型(2.6)中用配方的方法(不求助于求偏导数),以决定最小二乘估计(2.12)和(2.13),并由此得出残差平方和的表达式(2.23).

2. 在模型(2.6)中,假定(2.5)成立,仍记残差为 $\delta_1,\cdots,\delta_n$. 证明以下各点:

(a) $E(\delta_i)=0, i=1,\cdots,n$.

(b) $\delta_1,\cdots,\delta_n$ 不相互独立.

(c) $\mathrm{Var}(\delta_i)=\left(1-\frac{1}{n}-(X_i-\overline{X})^2/S^2\right)\sigma^2$

$(S^2=\sum_{i=1}^n(X_i-\overline{X})^2)$

(d) $\mathrm{Cov}(\delta_i,\delta_j)=-\left(\frac{1}{n}+(X_i-\overline{X})(X_j-\overline{X})/S^2\right)\sigma^2, i\neq j$.

3. 设样本 $X_1,\cdots,X_n\sim N(a,\sigma^2)$, $Y_1,\cdots,Y_m\sim N(b,\sigma^2)$, a,b,σ^2 都未知,为要估计 $b-a$ 或检验假设 $H_0:b-a=c$(c 已知),可利用线性回归的理论去做.指出具体怎样做的办法.

4. 考虑过原点的线性回归模型

$$Y_i=bX_i+e_i, i=1,\cdots,n$$

误差 $e_1,\cdots,e_n$ 仍假定满足条件(2.5).

(a) 给出 b 的最小二乘估计 $\hat{b}$.

(b) 给出残差平方和 $R=\sum_{i=1}^n\delta_i^2$ 的表达式,并证明:$R/(n-1)$是误差方差 σ^2 的无偏估计.这与不一定过原点的模型有何不同?为何有这个不同?

(c) 用附录 A 中的方法,证明当误差服从正态分布时,有 $R/\sigma^2\sim\chi_{n-1}^2$.

(d) 给出回归系数 b 的区间估计.

5. 考虑回归模型(2.4),而 c_1,c_2 为已知常数.假定(2.5)且设误差服从正态分布,求 c_1a+c_2b 的区间估计.

6. 从一元线性回归的讨论中出现几个有趣而初等的数学问题.现列举如下请读者考虑:

(a) 由第 2 题的(c),根据方差非负,可知:对任意 n 个实数 $X_1,\cdots,X_n$,有

$$(X_i-\overline{X})^2\Big/\sum_{j=1}^n(X_j-\overline{X})^2\leqslant 1-\frac{1}{n}, i=1,\cdots,n$$

等号在何时达到?

(b) 在 6.2 节 6.2.3 段末尾处提到的断言:若 $X_1,X_2,\cdots$,是一串实数,记 $\overline{X}_n=(X_1+\cdots+X_n)/n$, $S_n^2=\sum_{i=1}^n(X_i-\overline{X})^2$,则对任何固定的实数 a,有 $(a-\overline{X}_n)^2/S_n^2\to 0$ 当 $n\to\infty$.这个事实的统计意义已在 6.2 节(6.2.3)段中说明过了.

(c) 我们已证明：在模型(2.6)及假定(2.5)之下，$\mathrm{Var}(\hat{\beta}_1)=\sigma^2/S^2$，$S^2=\sum_{i=1}^{n}(X_i-\overline{X})^2$．当然，方差愈小愈好．故如限制试验点 X_i 只能取在某有限区间$[A,B]$内，就有一个如何配置这些点，以使 S^2 达到最大的问题．证明这问题的解是：若 n 为偶数 $2m$，则取 $X_1,\cdots,X_n$ 中有 m 个 A 及 m 个 B；若 n 为奇数 $2m+1$，则取 $X_1,\cdots,X_n$ 中有 m 个 A(或 B)，$m+1$ 个 B(或 A)．不过，在实用上，这个设计并不一定被采用，除非我们对回归函数为线性函数这一点绝无疑义．因为，这个设计只采用两个自变量值，它无法借助于观察数据去发现真实的回归函数与线性函数的可能的偏差．

7. 证明 6.3 节 6.3.1 段末尾处多元回归系数最小二乘估计的三个性质．

8. 设$(X_1,Y_1),\cdots,(X_n,Y_n)$是从二维正态总体 $N(a,b,\sigma_1^2,\sigma_2^2,\rho)$中抽出的样本，以 r 记样本相关系数．用以下的思路证明当 $\rho=0$ 时，$\sqrt{n-2}\,r/\sqrt{1-r^2}\sim t_{n-2}$：固定 $X_1,\cdots,X_n$，考虑 $Y_1,\cdots,Y_n$ 的条件分布，因为 $\rho=0$ 表示 X_i,Y_i 独立，故 $Y_1,\cdots,Y_n$ 的条件分布即是其无条件分布，即 $Y_1,\cdots,Y_n$ 独立，有公共分布 $N(b,\sigma_2^2)$．这可写为回归模型．

$$Y_i=b+\beta_1X_i+e_i,i=1,\cdots,n \tag{1}$$

回归系数 $\beta_1=0,e_i\sim N(0,\sigma^2),\sigma^2=\sigma_2^2$．然后在这个模型中使用(2.26)(记住 $\beta_1=0$)．证明(2.26)式左边正好就是$\sqrt{n-2}\,r/\sqrt{1-r^2}$．这样就证明了在给定 $X_1,\cdots,X_n$ 的条件下，$\sqrt{n-2}\,r/\sqrt{1-r^2}$的条件分布总是 t_{n-2}与 $X_1,\cdots,X_n$ 无关．因此$\sqrt{n-2}\,r/\sqrt{1-r^2}$的无条件分布就是 t_{n-2}．其所以要在给定 $X_1,\cdots,X_n$ 的条件下来考虑，是因为线性模型(1)有关的理论，特别是(2.26)式，都是在 X_i 为常数的情况下给出的．

9. 考虑下面的统计模型：样本 $X_1,\cdots,X_n,Y_1,\cdots,Y_n$ 独立，$X_i\sim N(d_i+a,\sigma^2)$，$Y_i\sim N(d_i+b,\sigma^2)$，$i=1,\cdots,n$．这里 $d_1,\cdots,d_n$ 和 a,b,σ^2 都未知，要检验假设 $H_0:a=b$．

(a) 试通过使用 $Z_i=Y_i-X_i,i=1,\cdots,n$，用 t 检验来处理这个问题．

(b) 说明：这个模型事实上是一个随机区组试验模型，共有 n 个区组，区组大小为 2．写出化到这样一种模型的过程．

(c) 用随机区组模型的 F 检验来处理 H_0 的检验问题，证明它与用(a)中的方法得到的一致．

10. 验证一下,下面的表是正交表:

行号 \ 列号	1	2	3	4	5	6	7
1	1	1	1	1	1	1	1
2	1	2	2	1	1	2	2
3	1	1	1	2	2	2	2
4	1	2	2	2	2	1	1
5	2	1	2	1	2	1	2
6	2	2	1	1	2	2	1
7	2	1	2	2	1	2	1
8	2	2	1	2	1	1	2

按正交表命名法,这个表的名称应是什么?它在用来安排试验时受到哪些限制?现如有三个两水平因子 A,B,C,共做 8 次试验,并分两个区组做,这试验如何用这张表安排?写出其方差分析表.