

- 4.1 多元线性回归模型的假设与一元线性回归模型有什么异同？
- 4.2 多元线性回归模型的矩条件如何得出，样本矩条件有多少？
- 4.3 自变量无完全共线性假设起到了什么作用？在参数估计的哪个环节用到？
- 4.4 多元线性回归模型的拟合优度 $R^2 = 1$ 表明了什麼？与 R^2 相比，调整 R^2 有什么优点？
- 4.5 公式 (4.16) 给出了同方差和随机抽样假设下多元线性回归模型 OLS 估计的方差计算公式。与第三章结论 6 给出的一元线性回归模型斜率参数 OLS 估计的方差计算公式 $\hat{\sigma}_{\beta_1}^2$ 相比，二者有什么联系？
- 4.6 结论 5、结论 8 和结论 10 有什么区别和联系？
- 4.7 多元线性回归模型的 F 检验如果原假设被拒绝意味着什麼？与 t -检验有什麼关系？
- 4.8 White 异方差检验的原假设是什麼？在采用 EViews 进行异方差的 White 检验时，勾选项 Include White cross terms 的作用是什麼？
- 4.9 什麼是解释变量的内生性？内生解释变量会带来什麼问题？
- 4.10 表 4.3 给出了 1983 年至 2006 年间我国农业生产的有关数据，其中 y 表示粮食产量， x_1 表示农作物成灾面积， x_2 表示农作物播种面积。

表 4.3

年份	y	x_1	x_2
1983	38727.5	16209.3	11404.7
1984	40730.5	15264	11288.4
...
2005	48402.2	19966	10427.8
2006	49747.9	20545.5	10548.9

数据来源：《中国统计年鉴》

从本书网页下载数据，并进行如下操作：

- (1) 采用 OLS 估计模型

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

，并检验 β_0 、 β_1 和 β_2 是否为 0。

- (2) 采用 OLS 方法估计模型

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \delta t + \varepsilon$$

，其中变量 t 表示年份变量，1983 年取 1，此后每增加一年 t 的值增加 1。检验 δ 是否为 0。

- (3) 采用 OLS 方法估计模型

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \delta t + \varepsilon$$

。(i) 计算参数估计 $\hat{\beta}_0$ 、 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 的方差膨胀因子，评价共线性对参数估计方差以及 t 检验的影响。(ii) 检验模型误差项是否存在序列相关，如果存在，采用 Newey-West 调整方法计算参数估计的 HAC 方差以及相应的 t -检验，并与先前的估计和检验结果进行比较。

- 4.11 表 4.4 给出的是一项香烟消费抽样调查数据，共有 807 个抽烟者接受调查。调查数据包括被调查者受教育年限 (edu)、是否为白人 (white) (白人取 1，否则取 0)、年龄 (age)、收入 (inc)、每天抽烟数量 (cigs)、每包香烟价格 (cigp) (美元)、所在州是否允许在餐馆抽烟 (restrn) (禁止为 1，否则取 0)，以及收入自然对数 (linc)、年龄的平方 (agesq)

和价格自然对数 $\ln c_{igp}$ 。

表 4.4

序号	<i>edu</i>	<i>cigp</i>	<i>whi</i>	<i>age</i>	<i>inc</i>	<i>cigs</i>	<i>restrn</i>	<i>lninc</i>	<i>agesq</i>	<i>ln c_{igp}</i>
1	16	60.506	1	46	20000	0	0	9.903487	2116	4.102743
2	16	57.883	1	40	30000	0	0	10.30895	1600	4.058424
...
806	10	59.988	1	18	20000	0	0	9.903487	324	4.094144
807	10	59.652	1	47	30000	20	0	10.30895	2209	4.088528

从本书网页下载，进行如下操作。

- (1) 以每天抽烟数量为被解释变量，其余为解释变量建立香烟需求模型，采用所给数据用 OLS 方法对模型进行估计。
- (2) 用怀特方法检验模型误差项是否存在异方差。如果存在异方差，哪些变量对方差影响较为明显。采用 White 方法计算 OLS 估计量方差和标准误并进行 t 检验，将检验结果与 (1) 中结果进行比较。
- (3) 计算每个参数的 VIF，分析自变量存在的共线性。
- (4) 在估计出的模型基础上，分析各个因素对吸烟量的影响。分析年龄对吸烟量的影响，在什么年龄抽样量最大？禁止在餐馆吸烟对吸烟量有多大的影响？

◆ 参考答案

1. 多元线性回归模型对模型误差项的假设，是以所有解释变量为条件的。例如假设 1 给出的零条件均值假设中，误差项的条件期望是以所有解释变量为条件的。同样区别也出现在同方差假设、随机抽样假设、正态分布假设等。除了对模型误差项进行必要的假设之外，多元线性回归模型还要假设模型自变量之间不存在完全共线性。一元线性回归模型只有一个自变量，不存在共线性问题。
2. 多元线性回归模型的矩条件从假设 1 得出，根据条件数学期望的性质，从零条件均值假设可以得出每个解释变量和误差项不相关，即与误差项乘积的数学期望为 0。从零条件均值中可以得出 $k+1$ 个矩条件，由类比原则得出的样本矩条件有 $k+1$ 个。
3. 自变量的无完全共线性假设保证了回归模型中的参数是唯一确定的（即可以被识别的）。只有被唯一确定的参数才可以用 OLS 估计。在从样本矩条件形成的方程组中 (4.9) 求出参数的 OLS 估计时，无完全共线性假设能够推出方程组的系数矩阵行列式不等于 0，保证了方程组有唯一解。
4. $R^2 = 1$ 称为完全拟合 (perfect fitting)，表明残差平方和等于 0，由此得出所有残差都等于 0，模型拟合值与实际值相等。根据 R^2 的定义可以得出，只要添加新的自变量到模型中， R^2 的值就会增加，会导致一些对因变量没有关系的变量进入模型。调整 R^2 对模型的自变量个数施加了惩罚，只有引入的变量对因变量有充分解释能力时才能使 \bar{R}^2 增加，以 \bar{R}^2 为标准在嵌套模型中选择的模型更合理。
5. 从公式 (4.16) 看出，OLS 估计的方差为两部分的乘积，第一部分是变量 X_j 与其余自变量无任何共线关系时 OLS 估计的方差，第二部分是源于 X_j 与其余自变量共线而导致的方差膨胀因子 VIF。如果 X_j 与其余自变量无共线关系，则 (4.16) 与第三章结论 6 给出的方差计算公式 $\hat{\sigma}_{\beta_j}^2$ 相同。

6. 三个结论都是关于多元线性回归模型参数 OLS 估计的分布。结论 5 要求的条件最少，仅要求假设 1 和假设 2 成立，最为适用。得出的分布是近似正态分布，并且正态分布的计算要分情况进行：如果存在异方差但没有序列相关，则采用 White 计算方法，如果有异方差也有自相关，则采用 Newey-West 方法计算。结论 8 要求的条件多，需要假设 1~假设 4 都成立，但方差计算公式简单。结论 10 要求的条件最多，需要假设 1~假设 5 成立，但结果很好，得出的分布是精确的正态分布，由于条件苛刻，实际中很少用到。
7. F 检验的原假设是多元线性回归模型中所有的斜率系数都等于 0，拒绝原假设表明，至少有一个回归系数不为 0，即至少有一个自变量对因变量有解释能力。到底哪些自变量对因变量有解释能力，则需要用单个参数的 t -检验来确定。
8. White 异方差检验的原假设是误差项为同方差（即不存在异方差），即辅助回归 (4.27) 中的斜率系数都等于 0。在采用 EViews 进行异方差的 White 检验时，勾选项 Include White cross terms 表明要在辅助回归 (4.27) 的自变量中包含原回归模型自变量的交叉相乘项。
9. 解释变量的内生性是指解释变量和模型误差项相关。解释变量有内生性时，模型参数的 OLS 估计不再具有一致性。
10. (1)、(2)、(3) 采用 EViews 计算。注意，在 EViews 的工作表中产生年份变量 t 时，需用 gener 按钮，并在弹出的对话框中输入 $t=@trend+1$ 。
11. (1)、(2)、(3) 用 EViews 计算。(4) 将估计出的模型看成 age 的二次函数，求函数的极大值点。