第五节　习题课

例1 考虑回归模型

$$y_i = a + bx_i + \varepsilon_i$$

其中 $y_i = \left(Y_i - \bar{Y}\right), x_i = \left(X_i - \bar{X}\right)$。这时，回归直线必定经过原点。此结论正确或错误？请证明。

证明：由一元线性回归最小二乘估计量的方法

$$\hat{b} = \frac{\sum_{i=1}^{n} x_i y_i - \bar{x}\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i^2 - \bar{x}\sum_{i=1}^{n} x_i} \qquad \hat{a} = \bar{y} - \hat{b}\bar{x}$$

本题中 $\bar{y}=0$，$\bar{x}=0$，所以 $\hat{a} = \bar{y} - \hat{b}\bar{x}=0$

结论正确。

例2 考虑回归模型，其中Y和X都不为0。

$$\frac{1}{Y_i}=a+b\frac{1}{X_i}+\varepsilon_i$$

（1）这是一个线性回归模型吗？

（2）怎样估计这个模型？

（3）随着X趋向无穷大，Y有怎样的行为？

解：

（1）模型关于系数线性，因此是线性回归模型。

（2）令$y_i=\frac{1}{Y_i}$，$x_i=\frac{1}{X_i}$, 作y_i关于x_i的线性回归。

(3) Y_i趋于$\frac{1}{a}$.

例3 考虑下列回归模型，

$$\ln Y_i^* = \alpha_1 + \alpha_2 \ln X_i^* + \varepsilon_i^* \qquad \ln Y_i = \beta_1 + \beta_2 \ln X_i + \varepsilon_i$$

其中$Y_i^* = w_1 Y_i$，$X_i^* = w_2 X_i$，w为常数。

（1）建立这两组回归系数之间的关系？

（2）两模型的R^2有何不同？

证明:(1)第一个回归模型等价于

$$\ln w_1 Y_i = \alpha_1 + \alpha_2 \ln w_2 X_i + \varepsilon_i$$

$$\ln w_1 + \ln Y_i = \alpha_1 + \alpha_2 \ln w_2 + \alpha_2 \ln X_i + \varepsilon_i$$

$$\ln Y_i = (\alpha_1 + \alpha_2 \ln w_2 - \ln w_1) + \alpha_2 \ln X_i + \varepsilon_i$$

因此 $\beta_1 = \alpha_1 + \alpha_2 \ln w_2 - \ln w_1$

$\beta_2 = \alpha_2$

（2） R^2相同.

例4 考虑下列回归模型，

模型A： $Y_t = \alpha_1 + \alpha_2 X_{2t} + \alpha_3 X_{3t} + \varepsilon_{1t}$

模型B： $(Y_t - X_{2t}) = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \varepsilon_{2t}$

（1）α_1和β_1的最小二乘估计会不会是一样的？

（2）α_2和β_2的最小二乘估计会不会是一样的？

（3）α_3和β_3的最小二乘估计会不会是一样的？

（4）你能比较两个模型的R^2吗？为什么？

解:(1) $\alpha_1=\beta_1$ （2）$\alpha_2=1+\beta_2$ （3）$\alpha_3=\beta_3$

（4） 不可以比较R^2.

通过R^2来评比两个模型，要求样本大小和因变量都必须相同，而解释变量则可取任何形式。

例5 假设估计消费函数：$Y_i=\alpha_1+\alpha_2 X_i+\varepsilon_{1i}$

储蓄函数：$Z_i=\beta_1+\beta_2 X_i+\varepsilon_{2i}$，

其中Y为消费，Z为储蓄，X为收入，并且$X=Y+Z$。

（1）模型的回归系数有什么关系？

（2）两个模型的残差平方和会不会是一样的？

（3）你能比较两个模型的R^2吗？为什么？

解:(1)

$$\hat{\alpha}_2=\frac{\sum_{i=1}^{n}X_iY_i-\bar{X}\sum_{i=1}^{n}Y_i}{\sum_{i=1}^{n}X_i^2-\bar{X}\sum_{i=1}^{n}X_i}=\frac{\sum_{i=1}^{n}X_i(X_i-Z_i)-\bar{X}\sum_{i=1}^{n}(X_i-Z_i)}{\sum_{i=1}^{n}X_i^2-\bar{X}\sum_{i=1}^{n}X_i}$$

$$=\frac{\sum_{i=1}^{n}X_i^2-\sum_{i=1}^{n}X_iZ_i-\bar{X}\sum_{i=1}^{n}X_i+\bar{X}\sum_{i=1}^{n}Z_i}{\sum_{i=1}^{n}X_i^2-\bar{X}\sum_{i=1}^{n}X_i}=1-\hat{\beta}_2$$

$$\hat{\alpha}_1 = \overline{Y} - \hat{\alpha}_2\overline{X} = \overline{X} - \overline{Z} - (1-\hat{\beta}_2)\overline{X} = -\overline{Z} + \hat{\beta}_2\overline{X} = -\hat{\beta}_1$$

（2） 模型1的残差平方和为

$$Q_1 = \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 = \sum_{i=1}^{n}\left(Y_i - \hat{\alpha}_1 - \hat{\alpha}_2 X_i\right)^2 = \sum_{i=1}^{n}\left(X_i - Z_i + \hat{\beta}_1 - X_i + \hat{\beta}_2 X_i\right)^2$$

$$= \sum_{i=1}^{n}\left(Z_i - \hat{\beta}_1 - \hat{\beta}_2 X_i\right)^2 = Q_2$$

（3） 不可以比较R^2.

通过R^2来评比两个模型，要求样本大小和因变量都必须相同，而解释变量则可取任何形式。

例6 考虑以下回归模型：$Y_i = \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{\varepsilon}_i$ 。

（1）模型的回归系数怎样估计？ （2） $\sum_{i=1}^{n} \hat{\varepsilon}_i = 0$？

（3） $\sum_{i=1}^{n} \hat{\varepsilon}_i X_{2i} = \sum_{i=1}^{n} \hat{\varepsilon}_i X_{2i} = 0$？

解：(1) 正规方程为

$$\begin{cases} \hat{\beta}_2 \sum_{i=1}^{n} X_{2i}^2 + \hat{\beta}_3 \sum_{i=1}^{n} X_{2i} X_{3i} = \sum_{i=1}^{n} X_{2i} Y_i \\ \hat{\beta}_2 \sum_{i=1}^{n} X_{2i} X_{3i} + \hat{\beta}_3 \sum_{i=1}^{n} X_{3i}^2 = \sum_{i=1}^{n} X_{3i} Y_i \end{cases}$$

(2)错 (3)对

对于有截距项的模型来说，$\sum_{i=1}^{n} \hat{\varepsilon}_i = 0$总是成立的。

对于无截距项的模型来说，$\sum_{i=1}^{n} \hat{\varepsilon}_i = 0$不一定成立。

例7　假使回归模型：

$$\ln\left(Y_i / X_{2i}\right) = \alpha_1 + \alpha_2 \ln X_{2i} + \alpha_3 \ln X_{3i} + \mu_i$$

中的回归系数及其标准误差均已知，那么怎样估计以下回归模型的参数及其标准误差？

$$\ln Y_i = \beta_1 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + \varepsilon_i \quad .$$

解：原回归模型等价于

$$\ln Y_i - \ln X_{2i} = \alpha_1 + \alpha_2 \ln X_{2i} + \alpha_3 \ln X_{3i} + \mu_i$$

即：　$\ln Y_i = \alpha_1 + (\alpha_2 + 1)\ln X_{2i} + \alpha_3 \ln X_{3i} + \mu_i$

因此：　$\hat{\alpha}_1 = \hat{\beta}_1$，$\hat{\alpha}_2 + 1 = \hat{\beta}_2$，$\hat{\alpha}_3 = \hat{\beta}_3$，

且　$\sigma(\hat{\alpha}_i) = \sigma(\hat{\beta}_i)$

例8　给定如下回归结果:

$$\hat{Y}_i = 16899 \quad -2978.5X_{2i}$$

$$t = (8.5152) \quad (-4.7280) \qquad R^2 = 0.6149$$

以及

$$\hat{Y}_i = 9734.2 \quad -3782.2X_{2i} \quad +2815X_{3i}$$

$$t = (3.3705) \quad (-6.6070) \quad (2.9712) \qquad R^2 = 0.7706$$

你能推测出这些结果所依据的样本大小吗?

解：考虑第一个回归模型，因为

$$F_\alpha(1, n-2) = \left[t_\alpha(n-2)\right]^2 = \left[-4.7280\right]^2 = 22.3540$$

又因为 $F = (n-2)\dfrac{\hat{r}^2}{1-\hat{r}^2} = (n-2)\dfrac{0.6149}{1-0.6149}$

解得 $n = 16$

例9　在对一个含有30个厂商的随机样本做的平均薪金（W）对职工人数（N）的回归中，得到如下的回归结果：

$$\hat{W} = 7.5 \quad +0.009N$$

$$t = \qquad (16.10) \qquad R^2 = 0.90$$

以及　$$\hat{W}/N = 0.008 \quad +7.8\left(1/N\right)$$

$$t = (14.43) \quad (76.58) \quad R^2 = 0.99$$

(1)从模型1到模型2作者做了什么假定？他是否担心过异方差性？

（2）怎样把两个模型的截距和斜率联系起来？

解：

（1）考虑了异方差性，假设方差的形式和N^2有关。

（2）　回归系数近似相等

例**10**　在模型 $Y_i=\beta X_i+\varepsilon_i$ （注意没有截距项）

其中$E\left(\varepsilon_i\right)=0, Cov\left(\varepsilon_i,\varepsilon_j\right)=0,\ Var\left(\varepsilon_i\right)=\sigma^2 X_i^2$

证明 $Var\left(\hat{\beta}\right)=\dfrac{\sigma^2\sum X_i^4}{\left(\sum X_i^2\right)^2}$

例11　在模型 $Y_i=\beta X_i+\varepsilon_i$ （$i=1,2$）中

已知ε_1~N(0,σ^2),ε_2~N(0,2σ^2),而且他们相互独立。

（1）若$X_1=1, X_2=-1$，计算β的加权最小二乘估计值及其方差。

（2）若此时不正确地假定了误差方差相同（比方说都等于σ^2）

那么最小二乘估计量是什么，其方差又是多少？

(提示：参考例10)

解：

（1）$\hat{\beta}=\dfrac{2Y_1-Y_2}{3},\ D(\hat{\beta})=\dfrac{2}{3}\sigma^2$

(2) $\hat{\beta}=\dfrac{Y_1-Y_2}{2},\ D(\hat{\beta})=\dfrac{1}{2}\sigma^2$

例12　给定一个含有50个观测样本和4个解释变量的模型，如果

（1）d=1.05；　（2）d=1.40；　（3）d=2.5；　（4）d=3.97

请对相关性进行判断（α取0.05）。

解：（1）正相关；　（2）无结论；

（3）无结论；　（4）负相关

例13 假设模型: $Y_i = \alpha_1 + \alpha_2 X_i + \varepsilon_i$ 中ε_i确实序列无关,且ε_i~N(0,σ^2).
那么，如果我们假定了 $\varepsilon_i=\rho\varepsilon_{i-1}+\mu_i$，并使用了差分回归：

$$Y_i - \rho Y_{i-1} = \alpha_1(1-\rho) + \alpha_2 X_i - \rho\alpha_2 X_{i-1} + \mu_i$$

讨论μ_i的性质。

解：$E\mu_i=0$

$$D\mu_i=(1+\rho^2)\sigma^2$$

$$Cov(\mu_i,\mu_{i-1})=-\rho\sigma^2$$

例**14**　考虑如下模型：

$$Y_i = \alpha_1 + \alpha_2 D_i + \beta X_i + \varepsilon_i$$

其中 ***Y***=一位大学教授的年薪，***X***=从教年限，***D***=性别。

考虑虚变量***D***采用下列三种取值方式：

（**1**）当为男性时***D***取**1**，对女性***D***取值**0**；

（**2**）当为男性时***D***取**2**，对女性***D***取值**1**；

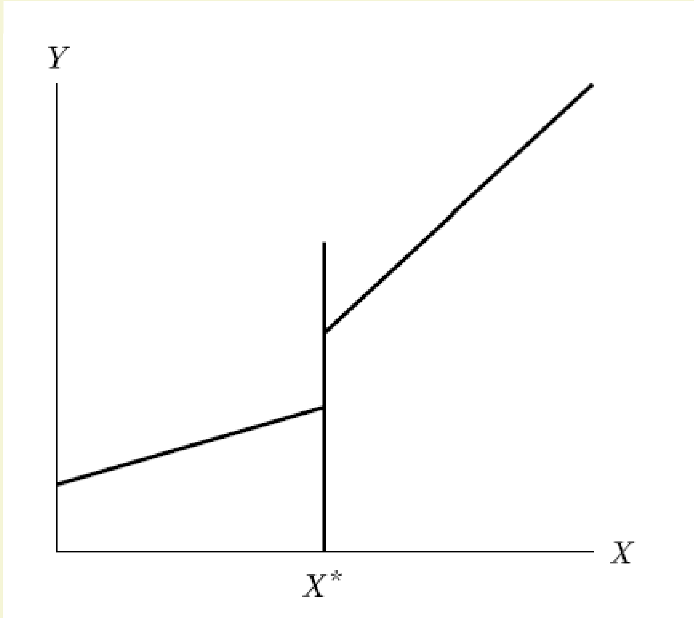
（**3**）当为男性时***D***取**1**，对女性***D***取值**-1**；

以上三种取值方式对预测结果是否会有影响？

是否其中某个取值方式比另外一个方法更好？

解：不会；没有。

例15 考虑以下回归问题。假设在X^*处不仅斜率系数发生了变化，而且回归直线还发生了跳跃。如何引入带有虚变量的回归模型，以考虑回归线在X^*处的跳跃。



解：$Y_i = \alpha_1 + \alpha_2 D + \beta_1 X_i + \beta_2 \left(X_i - X^* \right) D + \varepsilon_i$

其中虚变量的取值为

$$D = \begin{cases} 1 & 当 X_i > X^* \\ 0 & 当 X_i < X^* \end{cases}$$

例16　考虑以下回归问题：

$$Y_i = \beta_1 + \beta_2 D_i + \varepsilon_i .$$

其中，D_i对前20个观测取值0，而对后30个观测取值1, $D\varepsilon_i = 300$。

（1）前20组数据和后30组数据的均值分别是多少。

（2）如已知$\mathrm{Cov}\left(\hat{\beta}_1, \hat{\beta}_2\right) = -15$, 则$D\left(\hat{\beta}_1 + \hat{\beta}_2\right)$?

解：

（1）前20组数据的均值为β_1,

后30组数据的均值为$\beta_1 + \beta_2$。

（2）$D\left(\hat{\beta}_1 + \hat{\beta}_2\right)$=10

例17　考虑下表中的数据，如果你要用以下模型拟合数据

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i,$$

（1）你能估计这3个回归系数吗？

（2）如果不能，你将进行怎样的回归，解决问题？

Y	X_2	X_3
−10	1	1
−8	2	3
−6	3	5
−4	4	7
−2	5	9
0	6	11
2	7	13
4	8	15
6	9	17
8	10	19
10	11	21

解：

（1）$X_{3i}=2X_{2i}-1$，存在完全多重共线，因此不能估计回归系数。

（2）将回归模型整理，得

$$
\begin{aligned}
Y_i &= \beta_1+\beta_2 X_{2i}+\beta_3\left(2X_{2i}-1\right)+\varepsilon_i \\
&=\beta_1-\beta_3+\left(\beta_2+2\beta_3\right)X_{2i}+\varepsilon_i \\
&=\alpha_1+\alpha_2 X_{2i}+\varepsilon_i
\end{aligned}
$$

2.19. 设回归模型指定为

$$Y_i=\beta X_i+u_i$$

这里 u_i 满足所有的基本假设。现提出了 β 的三个估计量：

$$\hat{\beta}_1=\overline{Y}/\overline{X}$$

$$\hat{\beta}_2=\Sigma X_iY_i/\Sigma X_i^2$$

$$\hat{\beta}_3=\Sigma(X_i-\overline{X})(Y_i-\overline{Y})/\Sigma(X_i-\overline{X})^2$$

请回答以下问题：

(1)证明三个估计量都是 β 的无偏估计量；

(2)推导各个估计量的方差，并确定哪个是最小的(如果有的话)？

3.15. 考虑下述两个模拟过程

（1）建立回归模型

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$$

（2）建立回归模型

$$x_{2i} = \alpha_1 + \alpha_2 x_{3i} + u_i'$$

计算回归残差 $\hat{u}_i'$，最后建立回归模型

$$y_i = \beta_1' + \beta_2' \hat{u}_i' + \beta_3' x_{3i} + u_i^*$$

试问：等式 $\hat{\beta}_2 = \hat{\beta}'_2$ 是否成立？请证明之。

3.17. 假设真实模型为

$$y_i = \alpha_1 x_{1i} + \alpha_2 x_{2i} + u_i$$

其中 x_1，x_2 是确定性变量，随机变量 u_i 满足基本假定。现在把模型建立为

$$y_i = \beta_1 x_{1i} + v_i$$

请回答以下问题：

(1) 求出 β_1 的最小二乘估计量 $\hat{\beta}_1$；

(2) 在什么条件下，$E(\hat{\beta}_1) = \alpha_1$；

(3) 证明 $\hat{\beta}_1$ 的方差小于或等于 $\hat{\alpha}_1$ 的方差。

8. 5. 考虑以下模型：$Y_i=\beta_1+\beta_2X_i+\beta_3X_i^2+\beta_4X_i^3+\mu_i$。由于 X^2 和 X^3 是 X 的函数，所以它们之间存在多重共线性，你同意这种说法吗？为什么？

8. 5. 答：不同意。x^2 和 x^3 是 x 的非线性函数，把它们包括在回归模型中并不违反经典线性回归模型的基本假设。