一元线性回归

李德山

[1] 西南科技大学 经济管理学院

[2]School of Economics and Management
Southwest University of Science and Technology

2021 年 3 月 16 日

主要内容

主要内容

主要内容

主要内容

主要内容

一元线性回归模型

- **回归** (regression)是一种寻找被解释变量与解释变量之间的函数关系的数学方法。
  - 线性回归模型包含两个基本要素：1. 被解释变量、解释变量和干扰项之间的线性函数关系；2. 干扰项和解释变量之间的相关性。
  - 求解得到的线性回归函数并不必然反映变量之间的因果关系。
- 为什么在青少年时期要选择上学？如何解释教育投资的回报率？
  - 除了满足好奇心、个人成长外，一个重要的原因是教育能提高未来的收入水平。Mincer(1958) 指出个体选择多上一年学，则需要推迟一年挣钱（同时还需要交学费）；为弥补其损失，市场均衡条件要求给予受教育多的人更高的未来收入。

一元线性回归模型

- 假设我们研究受教育程度与收入水平（工资）的线性关系：

$$\ln W = \alpha + \beta EDU \tag{1}$$

$\ln W$ 为工资对数，EDU 为受教育年限，α 与β为待估参数。

- α 为截距项，表示当受教育年限为 0 时的工资对数水平。
- β为斜率，表示受教育年限对工资对数的边际效应，即每增加一年教育，将使工资增加百分之几：

$$\beta = \frac{d \ln W}{dEDU} = \frac{dW/W}{dEDU} \approx \frac{\Delta W/W}{\Delta EDU} \tag{2}$$

- 受教育年限只是影响工资或收入的因素之一。将其他无法观测但会影响工资的变量称为**干扰项**ε。那么，方程（1）应该为：

$$\ln W = \alpha + \beta EDU + \varepsilon \tag{3}$$

一元线性回归模型

- 更一般地，假设从总体随机抽取 n 个个体，则一元线性回归模型可以写成：

$$y_i = \alpha + \beta x_i + \varepsilon_i \tag{4}$$

y_i: 被解释变量（dependend variable）
x_i: 解释变量（indepedend variable）
α: 截距项（intercept）或常数项（constant）
β：斜率（slope）
α 和 β：统称回归系数或参数（paramenters）
ε_i: 误差项（error term）或扰动项（disturbance），包括遗漏的其他因素、变量的测量误差、回归函数的设定误差以及人类行为的内在随机性等
下标 i：个体 i，比如第 i 个人，第 i 个企业，第 i 个国家等
n：样本容量（sample size）

一元线性回归模型

- 上式右边 $\alpha + \beta x_i$ 称为总体回归线或总体回归函数。
- 模型 $y_i = \alpha + \beta x_i + \varepsilon_i$ 也称为**数据生成过程**。
- 从数据生成角度来看，随机变量 x_i 与 ε_i 首先从相应的概率分布中抽取观测值，确定 x_i 与 ε_i 的取值后，根据回归模型生成 y_i 的取值。
- 由于 ε_i 通常无法观测，故计量经济学的主要任务之一就是通过数据 $\{x_i, y_i\}_{i=1}^n$ 来获取关于总体参数的信息。

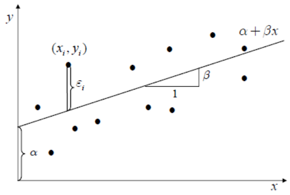


Figure: 数据生成过程

- 在考虑班级规模和学生测试成绩关系的例子中（Stock and Watson,2011）。这里我们根据 1999 年 420 个 K-8 加利福尼亚学区的学生 -教师比和五年级测试成绩的数据画出散点图。可以看出它们之间存在弱相关关系，样本相关系数为 -0.23。

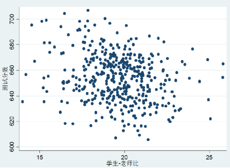


Figure: 学生 -教师比与测试成绩的散点图

- 那么，我们如何在图中画出这条向右下倾斜的线呢？最常用的方法就是**普通最小二乘法**（Ordinary Least Squares,OLS）来拟合这些数据。

OLS 估计量的推导

- 希望在 (x, y) 平面上找到一条直线，使得此直线离所有这些点最近。在此平面上，任意给定一条直线 $y_i = \hat{\alpha} + \hat{\beta} x_i$, 计算出每个点（观测值）到这条线的距离，$e_i = y_i - \hat{\alpha} + \hat{\beta} x_i$，称为**残差**。

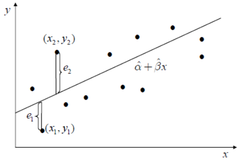


Figure: 残差平方和最小化

- 如果直接把残差加起来，$\sum_{i=1}^{n} e_i$，会出现正负相抵的现象。解决方法之一使用绝对值，$\sum_{i=1}^{n} |e_i|$。但绝对值无法微分，故考虑其平方，$\sum_{i=1}^{n} (e_i)^2 = \sum_{i=1}^{n} (y_i - \hat{\alpha} + \hat{\beta} x_i)^2$，称为残差平方和（SSR）。

OLS 估计量的推导

- OLS **的目标函数** 可以写为：

$$\min_{\hat{\alpha},\hat{\beta}} \sum_{i=1}^{n} {e_i}^2 = \sum_{i=1}^{n} (y_i - \hat{\alpha} + \hat{\beta} x_i)^2 \tag{5}$$

  - 此最小化问题的一阶条件为：

$$\begin{cases} \frac{\partial}{\partial \hat{\alpha}} \sum_{i=1}^{n} {e_i}^2 = -2 \sum_{i=1}^{n} (y_i - \hat{\alpha} + \hat{\beta} x_i) = 0 \\ \frac{\partial}{\partial \hat{\beta}} \sum_{i=1}^{n} {e_i}^2 = -2 \sum_{i=1}^{n} (y_i - \hat{\alpha} + \hat{\beta} x_i) x_i = 0 \end{cases} \tag{6}$$

- 消去方程左边的“-2”可得：

$$\begin{cases} \sum_{i=1}^{n}(y_i-\hat{\alpha}+\hat{\beta}x_i)=0 \\ \sum_{i=1}^{n}(y_i-\hat{\alpha}+\hat{\beta}x_i)x_i=0 \end{cases} \tag{7}$$

- 对上式各项分别求和、移项可得：

$$\begin{cases} n\hat{\alpha}+\hat{\beta}\sum_{i=1}^{n}x_i=\sum_{i=1}^{n}y_i \\ \hat{\alpha}\sum_{i=1}^{n}x_i+\hat{\beta}\sum_{i=1}^{n}x_i^2=\sum_{i=1}^{n}x_iy_i \end{cases} \tag{8}$$

OLS 估计量的推导

- 这是有关估计量 $\hat{\alpha}$，$\hat{\beta}$ 的二元一次线性方程组，称为**正则方程组**。从方程组（8）的第一个方程可得：

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \tag{9}$$

其中，$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$,$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$。将方程（9）代入到方程组（8）的第二个方程中可得：

$$(\bar{y} - \hat{\beta}\bar{x})\sum_{i=1}^{n} x_i + \hat{\beta}\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \tag{10}$$

合并同类项，移项可得：

$$\hat{\beta}(\sum_{i=1}^{n} x_i^2 - \bar{x}\sum_{i=1}^{n} x_i) = \sum_{i=1}^{n} x_i y_i - \bar{y}\sum_{i=1}^{n} x_i \tag{11}$$

根据 $\sum_{i=1}^{n} x_i = n\bar{x}$，求解 $\hat{\beta}$：

OLS 估计量的推导

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} \tag{12}$$

上式可以写成离差形式：

$$\hat{\beta} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{s_{xy}}{s_x^2} \tag{13}$$

$$\sum_{i=1}^{n} (x_i - \bar{x}) = 0$$

$$\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i (y_i - \bar{y}) = \sum_{i=1}^{n} (x_i - \bar{x}) y_i = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}$$

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} ({x_i}^2 - 2x_i\bar{x} + \bar{x}^2) = \sum_{i=1}^{n} x_i^2 - 2n(\bar{x})^2 + n(\bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n(\bar{x})^2$$

OLS 估计量的推导

- **方程（13）的分母不能等于 0。**
  - 从而可得：$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$
  - 解释变量 x_i 应有所变动，不能是常数，是对数据的最基本要求。
  - 如果 x_i 没有任何变化，则相同的 x_i 取值将对应于不同的 y_i 的取值，那么就无法估计 x 对 y 的作用。

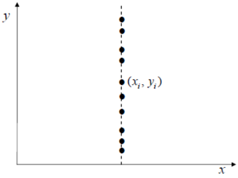


Figure: 解释变量没有变化的情形

OLS 估计量的推导

- 根据方程（9）与（13）可求解 OLS 的估计量 $\hat{\alpha}$,$\hat{\beta}$，得到 $\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i$, 称为**样本回归线**或样本回归函数。
  - 从方程（9）可知，样本回归线一定经过 $(\bar{x}, \bar{y})$。

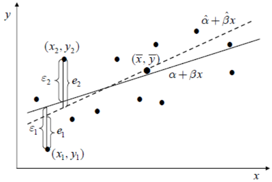


Figure: 总体回归线与样本回归线

OLS 统计量的代数性质与拟合优度

- **拟合值或预测值：** $\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i$
- **残差：** $e_i = y_i - \hat{\alpha} + \hat{\beta} x_i {=} y_i - \hat{y}_i$
  - 根据正则方程组（7）：

$$\begin{cases} \sum_{i=1}^{n} e_i {=} 0 \\ \sum_{i=1}^{n} x_i e_i {=} 0 \end{cases} \tag{14}$$

  - 写成向量内积的形式：

$$\begin{pmatrix} 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix} = 0, \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix} \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix} = 0 \tag{15}$$

  - 即 $\mathbf{1}'\boldsymbol{e} = \mathbf{0}, \boldsymbol{x}'\boldsymbol{e} = \mathbf{0}$。

OLS 统计量的代数性质与拟合优度

- 故常数向量与残差向量正交，而且残差向量也与解释向量正交。（1）：OLS 残差和及其样本均值都为零；（2）：解释变量与 OLS 残差的样本协方差为零。
- 残差向量也与拟合值向量正交；被解释变量的均值恰好等于拟合值的均值。（证略）

• **平方和分解公式** ： 被解释变量可分解为相互正交的两部分：

$$y_i = \hat{y}_i + e_i \tag{16}$$

- 如果回归方程有常数项，则被解释变量的离差平方和 (总平方和) $\sum_{i=1}^{n} (y_i - \bar{y})^2$ (Total Sum of Squares,TSS) 可分解为：

$$\underbrace{\sum_{i=1}^{n} (y_i - \bar{y})^2}_{TSS} = \underbrace{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}_{ESS} + \underbrace{\sum_{i=1}^{n} {e_i}^2}_{RSS} \tag{17}$$

OLS 统计量的代数性质与拟合优度

- 式（17）称为“平方和分解公式”，右边第一项为 $\sum\limits_{i=1}^{n}(\hat{y}_i-\bar{y})^2$，由于 $\bar{y}=\bar{\hat{y}}$，故可写为 $\sum\limits_{i=1}^{n}(\hat{y}_i-\bar{\hat{y}})^2$, 称为**解释平方和**（ESS），也称回归平方和, 可由模型解释的部分。
- 右边第二项为**残差平方和**（RSS），是模型无法解释的部分。
- 平方和分解公式能够成立，正是由于 OLS 的正交性（证略，见 Wooldridge,p35 和附录 A）。
- 如果没有常数项，则无法保证 $\sum\limits_{i=1}^{n}{e_i}^2=0$，故平方和分解公式不成立。

OLS 统计量的代数性质与拟合优度

- OLS 的样本回归线为离所有样本点最近的直线。但究竟离这些样本点有多近？如果模型可以解释的部分所占比重越大，则样本回归线的拟合程度越好。

- **定义** 拟合优度（goodness of fit），也称为可决系数，R^2 为:

$$R^2 = \frac{\sum\limits_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum\limits_{i=1}^{n}(y_i - \bar{y})^2} = 1 - \frac{\sum\limits_{i=1}^{n}{e_i}^2}{\sum\limits_{i=1}^{n}(y_i - \bar{y})^2} \tag{18}$$

- 在有常数项的情况下，拟合优度等于被解释变量与拟合值之间的相关系数的平方，即 $R^2 = [Corr(y_i, \hat{y}_i)]^2$。
- 显然，$R^2 = ESS/TSS = 1 - RSS/TSS$。$0 \leq R^2 \leq 1$。$R^2$ 越高，则样本回归线对数据的拟合程度越好。

OLS 统计量的代数性质与拟合优度

- 如果 $R^2 = 1$，则解释变量 x 可以完全解释 y 的变动。此时，残差平方和为零，所有残差均为零，故所有样本点都在样本回归线上。
- 如果 $R^2 = 0$，则解释变量 x 对于解释 y 没有任何帮助。此时，回归平方和为零，$\hat{y}_i \equiv \bar{y}$，故样本回归线为一条水平线，与 x 轴平行。这也意味着 $\hat{\beta} = 0$。

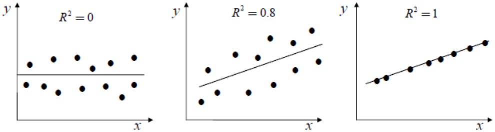


Figure: 拟合优度

- R^2 只是反映拟合程度的好坏，除此之外并无太多意义。评估回归方程是否显著，应使用 F 检验。
- 思考题：无常数项的回归，OLS 的正交性还成立吗？

OLS 统计量的代数性质与拟合优度

- 变量的度量单位的改变会影响 R^2 大小吗？不会。
  - 变量使用原始形式或对数形式可以得到四个函数形式组合。

表 2-3 含对数的函数形式总览

模型	因变量	自变量	对 β_1 的解释
水平值—水平值	y	x	$\Delta y=\beta_1\Delta x$
水平值—对数	y	$\log(x)$	$\Delta y=(\beta_1/100)\%\Delta x$
对数—水平值	$\log(y)$	x	$\%\Delta y=(100\beta_1)\Delta x$
对数—对数	$\log(y)$	$\log(x)$	$\%\Delta y=\beta_1\%\Delta x$

  - 在对数 -水平值模型中，$100\beta_1$ 也被称为 y 对 x 的半弹性。在对数 -对数模型中，β_1 也被称为 y 对 x 的弹性。
  - “线性”回归的含义：方程中的参数 β_1 和 α 是线性的，至于为 y 和 x 与我们所关注的被解释变量和解释变量有何联系，并没有限制。（线性意味着 x 的 1 单位的变化，将使 y 的期望值改变 β 之多）

- **回归标准误**（Standard error of the regression,SER）：是回归误差 e_i 的标准差估计量。
  - 因为回归误差 e_i 的单位同 Y_i 一样，所以 SER 是用因变量单位度量的观测值在回归线附近的离散程度。
  - 由于回归误差不可观测，因此利用样本中相应的 OLS 残差 $\hat{e}_i$ 计算 SER。

$$SER = s_{\hat{e}} = \sqrt{{s_{\hat{e}}}^2} = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}{e_i}^2} = \sqrt{\frac{RSS}{n-2}} \tag{19}$$

    其中 ${s_{\hat{e}}}^2$ 的公式用到了 OLS 残差的样本均值为零的结论。
  - 这里使用除数 $n-2$，而不是 n 的理由：即它修正了由于估计两个回归系数引入的微小向下偏差。这是因为两个回归系数是估计的，损失了数据的 2 个自由度。但是当 n 很大时，就差别不大了。
  - R^2 小 SER 大：R^2 小说明存在影响被解释变量的其他重要因素。SER 大说明只用该解释变量预测被解释变量通常会有较大的误差，散点图在回归线附近较为分散。

最小二乘假设

假设 1：关于参数是线性的

- 在回归模型中，被解释变量与解释变量和误差项之间的关系如下：

$$y_i = \alpha + \beta x_i + u_i$$

- x,y,u 都被视为表述总体模型时的随机变量。通过适当选择 x,y，我们能够得到有趣的非线性关系。

假设 2：(X_i, Y_i) 独立同分布

- 我们是通过随机抽样的方式从总体中获得样本。如果从总体中抽取了 n 个样本，那么它们具有相同的分布，又因为是随机抽取的，故它们也是独立的。即它们是 $i.i.d.$ 的。
- 无论是截面数据还是时间序列时间，并不是所有的抽样方案都能得到关于 (X_i, Y_i) 的 $i.i.d.$ 观测值。

最小二乘假设

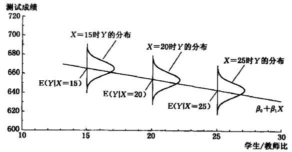
假设 3：解释变量的样本要有波动

- 也就是说在样本中，解释变量 x 不是一个不变常数。
- 在做回归分析之前检查一下 x 的描述性统计，如果 x 的样本标准差为零，则这个假设不成立。

最小二乘假设

假设 4：零条件均值

- **给定 x_i 的条件下，u_i（有的教材用 e_i）的条件分布均值为零** 。这个假设是说，“丢弃”在残差项里的其他因素与 x_i 无关。意味着 $Corr(x_i, u_i) = 0$。
- 将假设 $E(u_i) = 0$ 结合 $E(u_i) = E(u_i|x_i)$(称为 u 的均值独立与 x) 便得到**零条件均值假定**:$E(u_i|x_i) = 0$。



- 零条件均值假定给定了 β 的另一种非常有用的解释。以 x 为条件将方程（4）取期望值，并利用 $E(u|x) = 0$，便得到：$E(y|x) = \alpha + \beta x$。对任何给定的 x，y 的分布都以 $E(y|x)$ 为中心（见图）。

最小二乘假设与 OLS 估计量的抽样分布

假设 5：同方差假设，$Var(u|x) = \sigma^2$

- 给定解释变量的任意值，误差都具有相同的方差。
- 根据 $E(y|x) = \alpha + \beta x; Var(u|x) = \sigma^2$K 可以知道，给定 x,y 的条件期望线性于 x，但给定定 x,y 的方差却是常数。（见图）
- 当 $Var(u|x) = \sigma^2$ 取决于 x 时，则称误差项表现出**异方差性**（heteroskedasticity）。由于 $Var(u|x) = Var(y|x)$，所以只要，$Var(y|x)$ 是 x 的函数，便出现了异方差性。

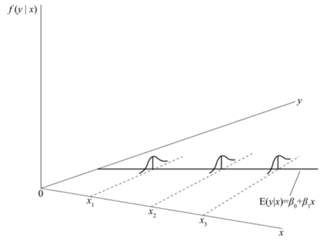
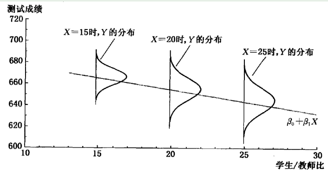


图 2－8　同方差下的简单回归分析



最小二乘假设

假设 6：X_i Y_i 不太可能出现较大异常值

- 较大的奇异值会使得 OLS 结果产生误差。这个假设使得 X_i Y_i 具有非零有限四阶矩：$0 < E(X_i^4) < \infty$,$0 < E(Y_i^4) < \infty$。另一种表述是 X_i Y_i 具有有限峰度。
- 出现大的异常值可能是数据录入错误或是单位错误。可以通过画图来检查。

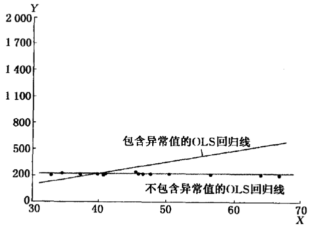


Figure: OLS 对较大异常值的敏感性

- 无偏性的证明。参见伍德里奇（第五版），p44。

证明思路

- 由于 $y_i = \alpha + \beta_1 x_i + e_i$，$y_i - \bar{y} = \beta_1(x_i - \bar{x}) + e_i - \bar{e}$。因此方程（13）中的分子可以化为：

$$
\begin{aligned}
&\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) = \sum_{i=1}^{n}(x_i-\bar{x})[\beta_1(x_i-\bar{x})+(e_i-\bar{e})] \\
=&\beta_1\sum_{i=1}^{n}(x_i-\bar{x})^2+\sum_{i=1}^{n}(x_i-\bar{x})(e_i-\bar{e}) \\
=&\beta_1\sum_{i=1}^{n}(x_i-\bar{x})^2+\sum_{i=1}^{n}(x_i-\bar{x})e_i
\end{aligned}
$$

- 这里用到如下推导：

$$
\sum_{i=1}^{n}(x_i-\bar{x})(e_i-\bar{e}) = \sum_{i=1}^{n}(x_i-\bar{x})e_i - \sum_{i=1}^{n}(x_i-\bar{x})\bar{e} = \sum_{i=1}^{n}(x_i-\bar{x})e_i
$$

最小二乘假设

证明思路

- 上面用到了 $\bar{x}$ 的定义，意味着 $\sum_{i=1}^{n}(x_i-\bar{x})\bar{e}=[\sum_{i=1}^{n}x_i-n\bar{x}]\bar{e}=0$
- 将上式推导代入到方程（13）中，得：

$$\hat{\beta}_1=\beta_1+\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})e_i}{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2} \tag{20}$$

- 对上式两边取期望可得 $\hat{\beta}_1$ 的期望。于是

$$\begin{aligned}E(\hat{\beta}_1)&=\beta_1+E[\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})e_i}{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}]\\&=\beta_1+E[\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})E(e_i|x_1,x_2,\cdots,x_n)}{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}]=\beta_1\end{aligned} \tag{21}$$

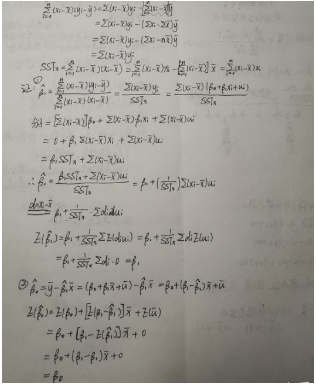
最小二乘假设

证明思路

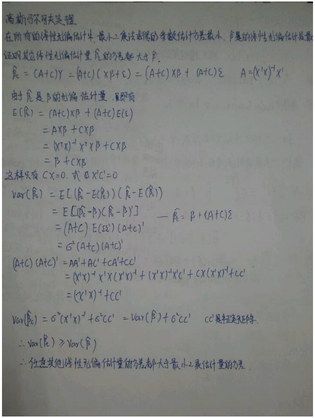
- 上式的第二个等式由期望的迭代原则可以得到。由最小二乘假设可知，除了 x 的第 i 个观测值以外，e_i 与其他所有的 x 观测值都独立，故 $E(e_i|x_1, x_2, \cdots, x_n) = E(e_i|x_i)$。同时由最小二乘的假设可知，$E(e_i|x_i) = 0$。由此，式（21）的第二行中括号的条件期望等于零。因此，$E(\hat{\beta}_1) = \beta_1$, 故 $\hat{\beta}_1$ 是无偏的。

最小二乘假设

- 有效性的证明，参见陈强，高级计量经济学及 Stata 应用（第二版），p19。
- 线性条件无偏估计量和 $Gauss-Markov$ 定理，参见斯托克，计量经济学导论（第五版），p112；p140。（下一章还会详细介绍）


$$\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})=\sum(x_i-\bar{x})y_i-\left[\sum_{i=1}^{n}(x_i-\bar{x})\right]\bar{y}$$
$$=\sum(x_i-\bar{x})y_i-(\sum x_i-\sum\bar{x})\bar{y}$$
$$=\sum(x_i-\bar{x})y_i-(\sum x_i-n\bar{x})\bar{y}$$
$$=\sum(x_i-\bar{x})y_i$$
$$SST_x=\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})=\sum_{i=1}^{n}(x_i-\bar{x})x_i-\left[\sum_{i=1}^{n}(x_i-\bar{x})\right]\bar{x}=\sum_{i=1}^{n}(x_i-\bar{x})x_i$$
证：① $\hat{\beta}_1=\dfrac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})}=\dfrac{\sum(x_i-\bar{x})y_i}{SST_x}=\dfrac{\sum(x_i-\bar{x})(\beta_0+\beta_1x_i+u_i)}{SST_x}$

分子 $=\left[\sum(x_i-\bar{x})\right]\beta_0+\sum(x_i-\bar{x})\beta_1x_i+\sum(x_i-\bar{x})u_i$
$$=0+\beta_1\sum(x_i-\bar{x})x_i+\sum(x_i-\bar{x})u_i$$
$$=\beta_1SST_x+\sum(x_i-\bar{x})u_i$$
$$\therefore\hat{\beta}_1=\frac{\beta_1SST_x+\sum(x_i-\bar{x})u_i}{SST_x}=\beta_1+\left(\frac{1}{SST_x}\right)\sum(x_i-\bar{x})u_i$$
$$\overset{d=x_i-\bar{x}}{=\!=}\beta_1+\frac{1}{SST_x}\cdot\sum d_iu_i$$
$$E(\hat{\beta}_1)=\beta_1+\frac{1}{SST_x}\sum E(d_iu_i)=\beta_1+\frac{1}{SST_x}\sum d_iE(u_i)$$
$$=\beta_1+\frac{1}{SST_x}\sum d_i\cdot0=\beta_1$$
② $\hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x}=(\beta_0+\beta_1\bar{x}+\bar{u})-\hat{\beta}_1\bar{x}=\beta_0+(\beta_1-\hat{\beta}_1)\bar{x}+\bar{u}$
$$E(\hat{\beta}_0)=E(\beta_0)+\left[E(\beta_1-\hat{\beta}_1)\right]\bar{x}+E(\bar{u})$$
$$=\beta_0+\left[\beta_1-E(\hat{\beta}_1)\right]\bar{x}+0$$
$$=\beta_0+(\beta_1-\beta_1)\bar{x}+0$$
$$=\beta_0$$


高斯马尔可夫定理

在所有的线性无偏估计中，最小二乘法求得的参数估计方差最小，$\hat{\beta}$ 是最优的线性无偏估计量

证明其它线性无偏估计量 $\hat{\beta}_c$ 的方差都大于 $\hat{\beta}$.
$$\hat{\beta}_c=(A+C)Y=(A+C)(X\beta+\varepsilon)=(A+C)X\beta+(A+C)\varepsilon\qquad A=(X'X)^{-1}X'$$
由于 $\hat{\beta}_c$ 是 β 的无偏估计量，即有
$$E(\hat{\beta}_c)=(A+C)X\beta+(A+C)E(\varepsilon)$$
$$=AX\beta+CX\beta$$
$$=(X^TX)^{-1}X^TX\beta+CX\beta$$
$$=\beta+CX\beta$$
这样只有 $CX=0$，或 $X'C'=0$
$$var(\hat{\beta}_c)=E[(\hat{\beta}_c-E(\hat{\beta}_c))(\hat{\beta}_c-E(\hat{\beta}_c))]$$
$$=E[(\hat{\beta}_c-\beta)(\hat{\beta}_c-\beta)']\quad—\ \hat{\beta}_c=\beta+(A+C)\varepsilon$$
$$=(A+C)E(\varepsilon\varepsilon')(A+C)'$$
$$=\sigma^2(A+C)(A+C)'$$
$$(A+C)(A+C)'=AA'+AC'+CA'+CC'$$
$$=(X'X)^{-1}X'X(X'X)^{-1}+(X'X)^{-1}X'C'+CX(X'X)^{-1}+CC'$$
$$=(X'X)^{-1}+CC'$$
$var(\hat{\beta}_c)=\sigma^2(X'X)^{-1}+\sigma^2CC'=Var(\hat{\beta})+\sigma^2CC'$　CC' 是半正定矩阵.

$\therefore var(\hat{\beta}_c)\geqslant Var(\hat{\beta})$

$\therefore$ 任意其他线性无偏估计量的方差都大于最小二乘估计量的方差.

假设检验与置信区间

- 回到班级规模与成绩的例子中。有人会说“班级规模并不会对分数产生影响”，也就是说，$\beta_1 = 0$。下面我们就来检验斜率是否为 0（原假设）。然后判断是否接受（拒绝）原假设。
  - 原假设：$H_0 : E(Y) = \mu_{Y,0}$。备择假设：$H_0 : E(Y) \neq \mu_{Y,0}$。
  - 假设检验步骤：
    1. 计算 $\bar{Y}$ 的标准误 $SE(\bar{Y})$
    2. 计算 t 统计量，即 $t = \frac{\bar{Y}-\mu_{Y,0}}{SE(\bar{Y})}$
    3. 计算 p 值，它表示的是能够拒绝原假设的最小显著水平。由于原假设下，t 统计量在大样本下服从标准正态分布，因此双边假设的 p 值为 $2\Phi(-|t^{act}|)$, 其中 t^{act} 是实际计算得到的 t 统计量值，Φ 为累积标准正态分布。

假设检验与置信区间

- 根据上述检验步骤，我们假设，$H_0 : \beta_1 = 0, H_1 : \beta_1 \neq 0$。
  - 第一步，计算 $\hat{\beta}_1$ 的标准误 $SE(\hat{\beta}_1)$。即：$SE(\hat{\beta}_1) = \sqrt{\hat{\sigma}^2_{\hat{\beta}_1}}$，其中，

$$\hat{\sigma}^2_{\hat{\beta}_1} = \frac{1}{n} \times \frac{\frac{1}{n-2}\sum_{i=1}^{n}(x_i-\bar{x})^2\mu_i^2}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^2}$$

  - 第二步，计算 t 统计量。$t = \frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)}$。
  - 第三步，计算 p 值。

$$\begin{aligned} p_{-value} &= \Pr_{H_0}[|\hat{\beta}_1 - 0| > |\hat{\beta}_1^{act} - 0|] \\ &= \Pr_{H_0}[|\frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)}| > |\frac{\hat{\beta}_1^{act} - 0}{SE(\hat{\beta}_1)}|] \\ &= \Pr_{H_0}(|t| > |t^{act}|) \end{aligned}$$

  - 由于 t 统计量近似标准正态分布，因此，
    $p_{-value} = \Pr(|Z| > |t^{act}|) = 2\Phi(-|t^{act}|)$。如果 p 值小于 5%, 即是说在 5% 的显著性水平下拒绝原假设。或者，若 $|t^{act}| > 1.96$, 则在 5% 显著性水平下拒绝原假设。

- 从样本数据并不能得到系数的真值，但是我们可以根据 OLS 的估计量和标准误构造 β_1,α 的置信区间。
  - 系数 β_1 的**置信区间**有两种等价含义。其一是在 5% 显著水平下利用双边假设检验不能拒绝的取值集合。其二是以 95% 的概率包含 β_1 真值的区间。（称之为 95% 的置信水平）
  - β_1 的 95% 的置信区间为：$[\hat{\beta}_1 - 1.96SE(\hat{\beta}_1), \hat{\beta}_1 + 1.96SE(\hat{\beta}_1)]$。

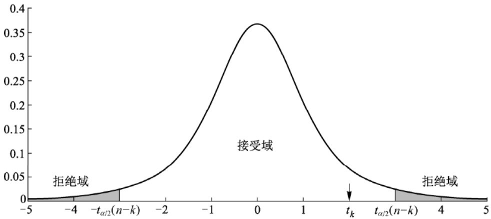


Figure: 双边 t 检验的临界值与拒绝域

参考文献：

Wooldridge. Introductory econometrics: A modern approach. Nelson Education，2015.

陈强. 计量经济学及 Stata 应用 [M]. 高等教育出版社，2015.