

计量经济学-作业1

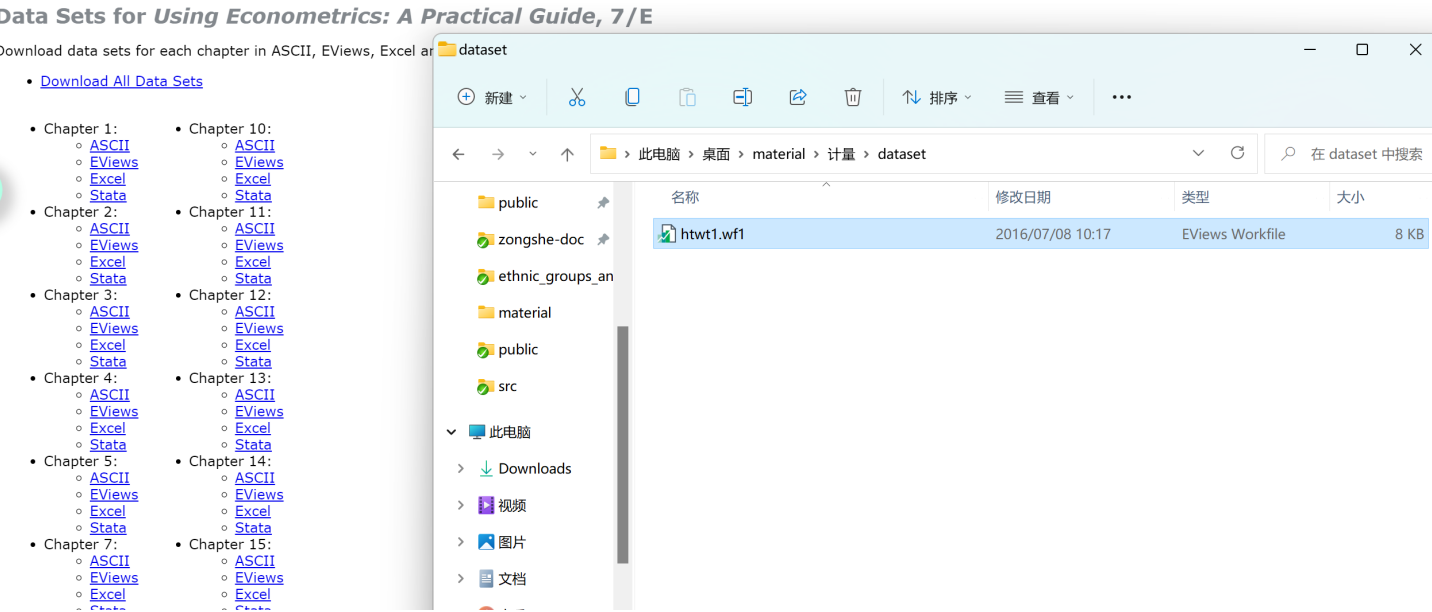


第二题

2 利用计算机软件，以及表 1-2 中体重 (Y) 和身高 (X) 的数据，对方程 (1-19) 重新进行估计。这里有两种录入数据的方法：直接录入数据；从课程网站 www.pearsonhighered.com/studenmund 中下载数据文件 HTWT1 (按 Stata, EViews, Excel 或 ASCII 形式)，下载好数据后，运行 $Y=f(X)$ ，得出的结果应该与方程 (1-19) 一致。运用不同的程序对方程进行回归时，需要不同的指令。关于 Stata 和 EViews 的使用方法，详见附录 A 中的答案或附录 1A。

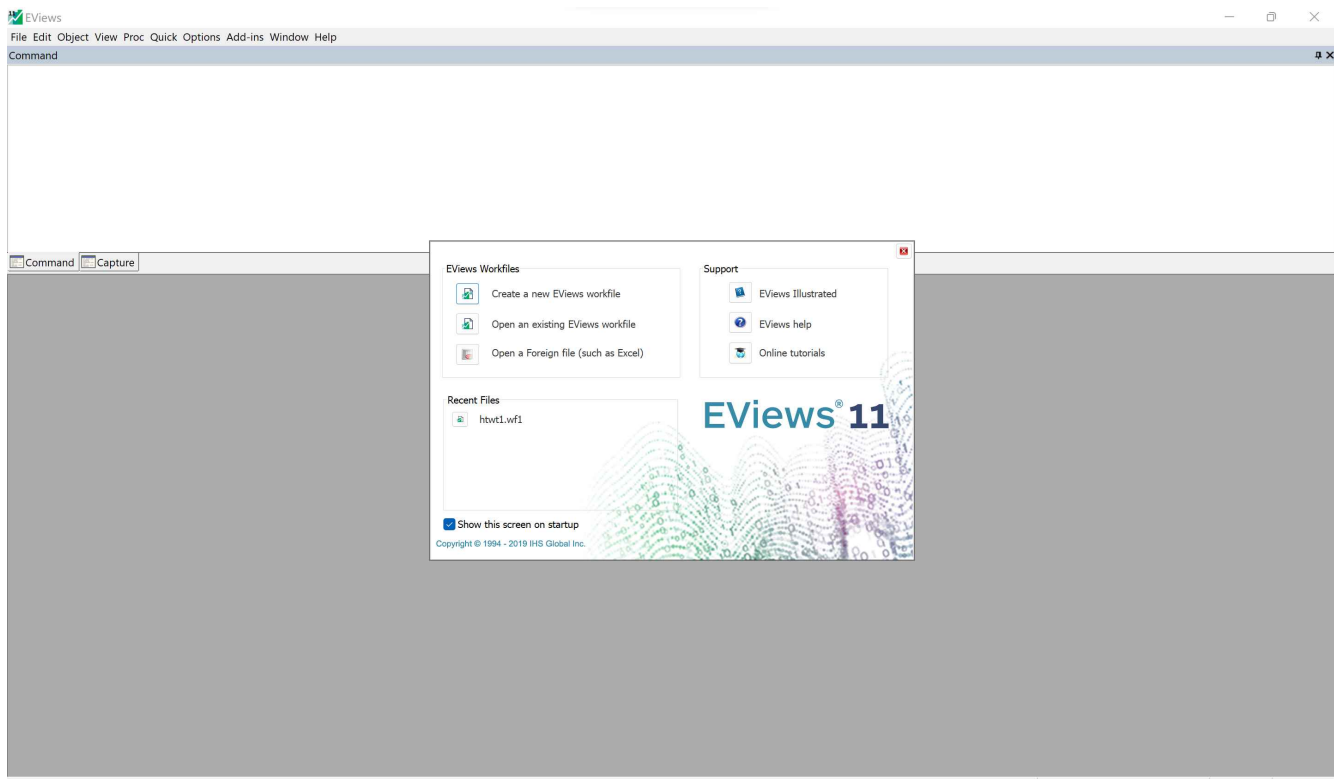
步骤

1. 使用课程网站下载相关数据集



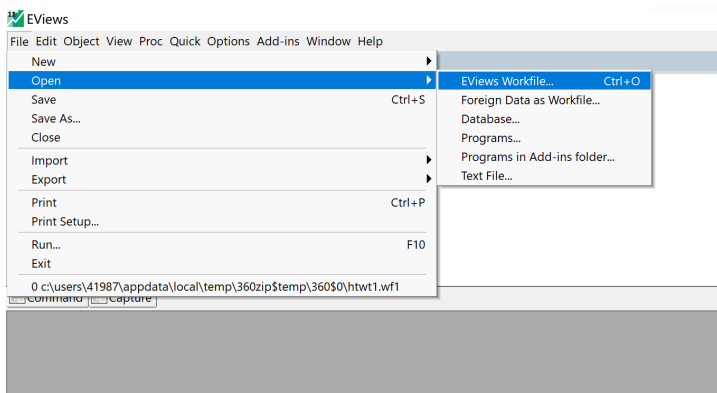
数据集下载

2. 打开review

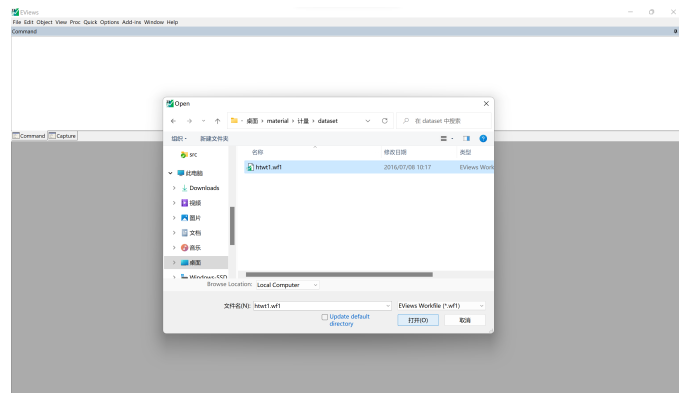


eview启动界面

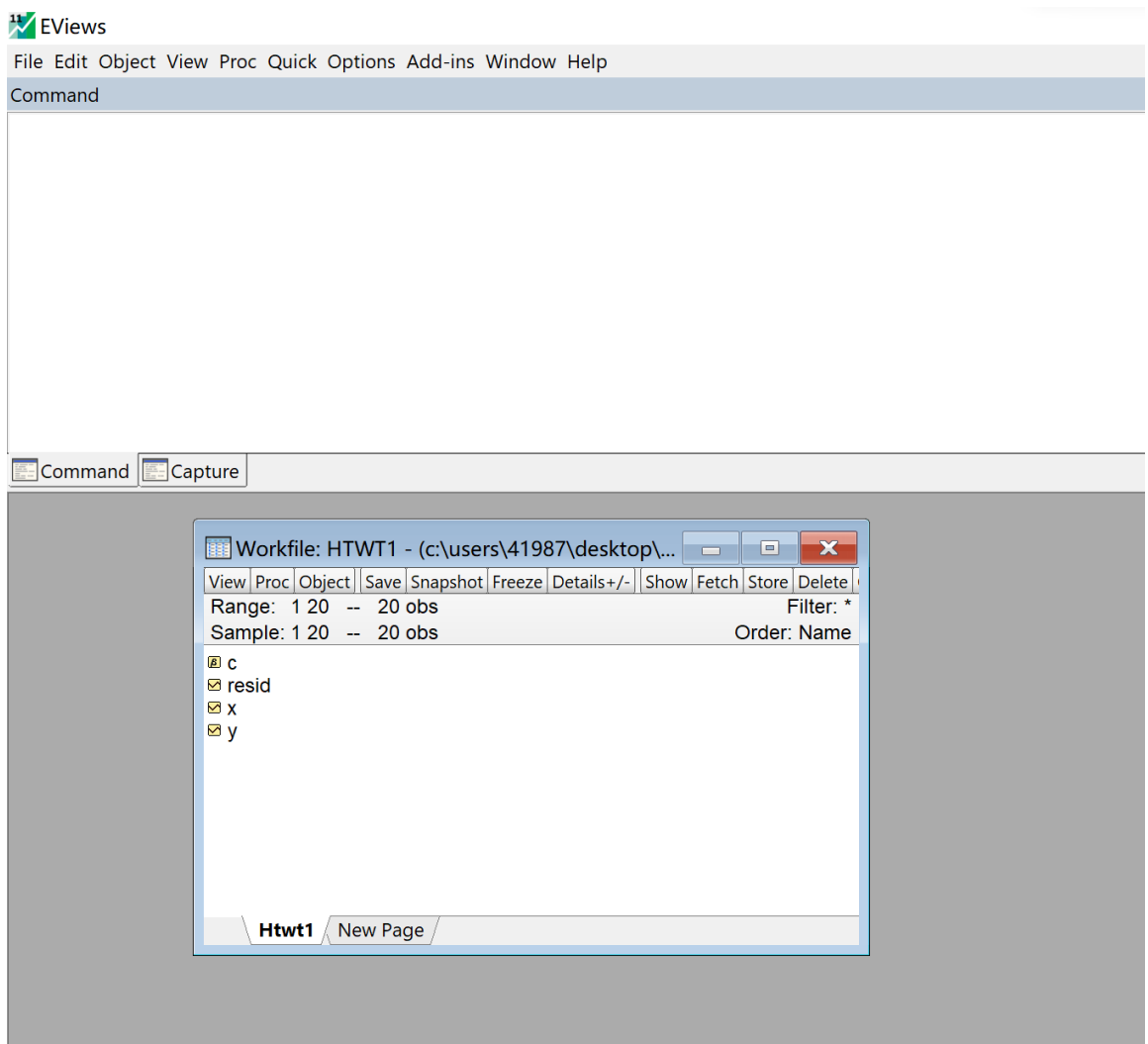
3. 导入数据集



新建workfile

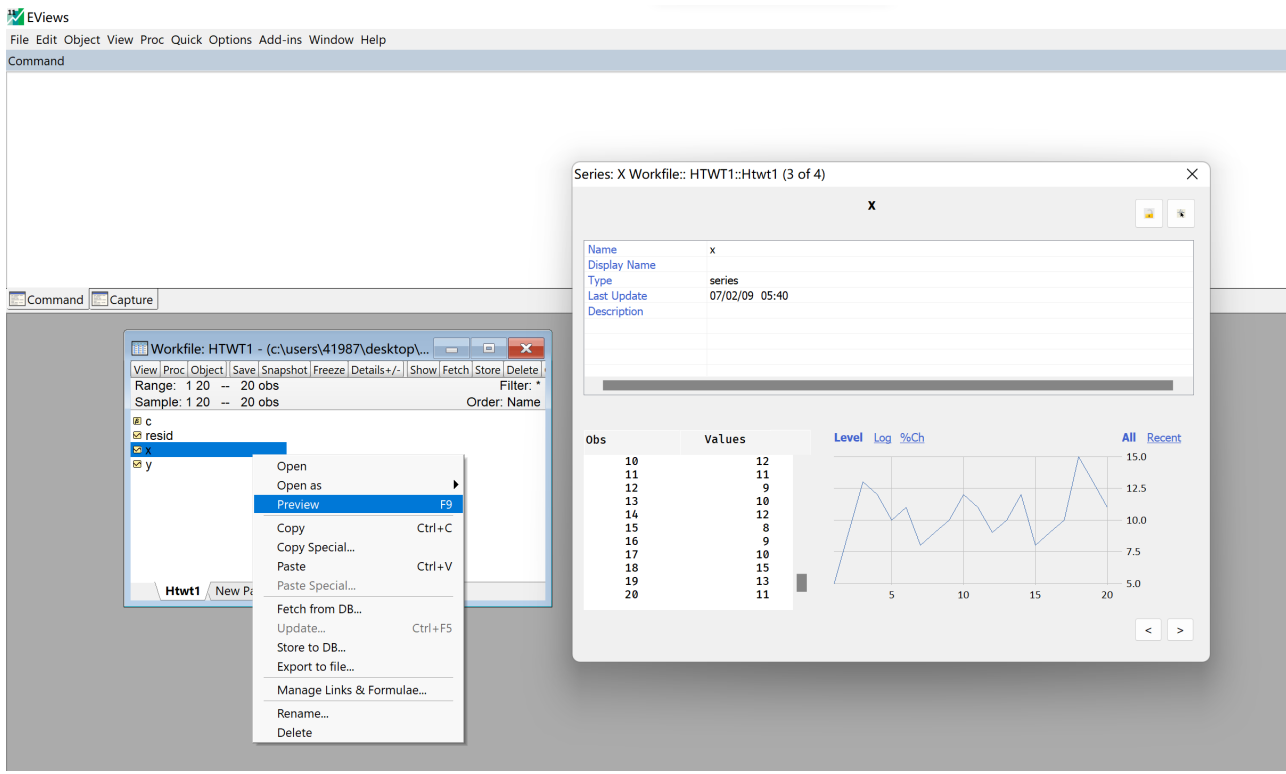


导入下载好的数据集



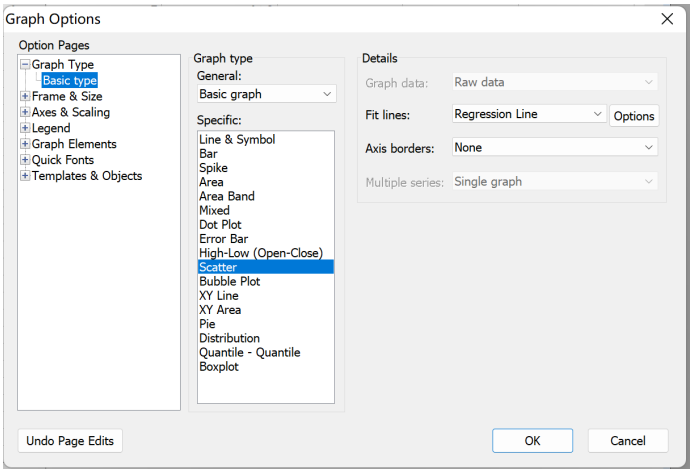
导入成功

4. 对数据进行预览

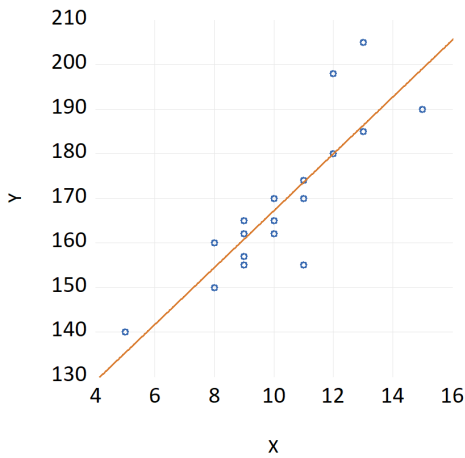


预览数据

a. 绘制散点图查看相关数据的分布



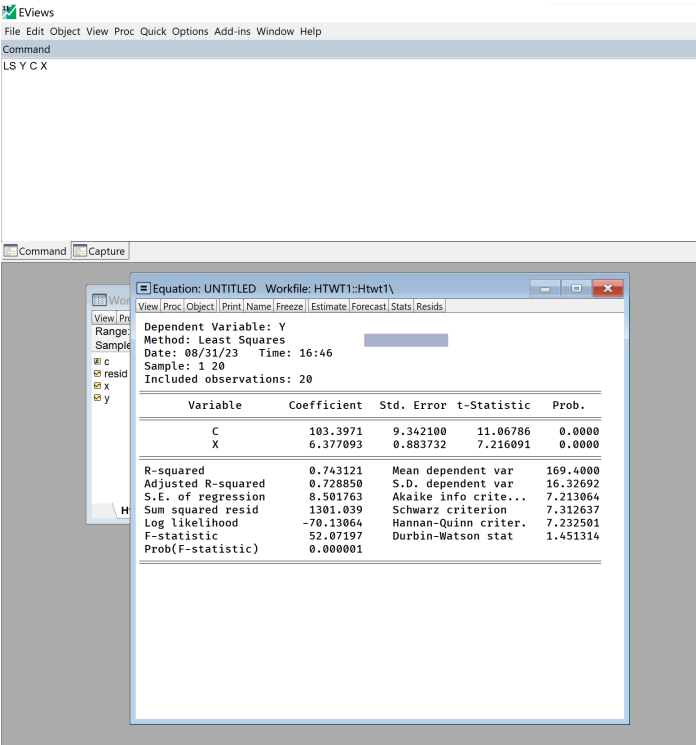
带回归拟合曲线的散点图绘制-步骤a



带回归拟合曲线的散点图绘制-步骤b

5. 输入指令 `LS Y C X` 进行回归

其中LS指令表示用**最小二乘法**进行回归拟合



回归拟合与拟合结果

左图相关英文的解释——

Sample样本情况

Included observation包含的数据量

Coefficient估计的系数值

Std.Error标准误

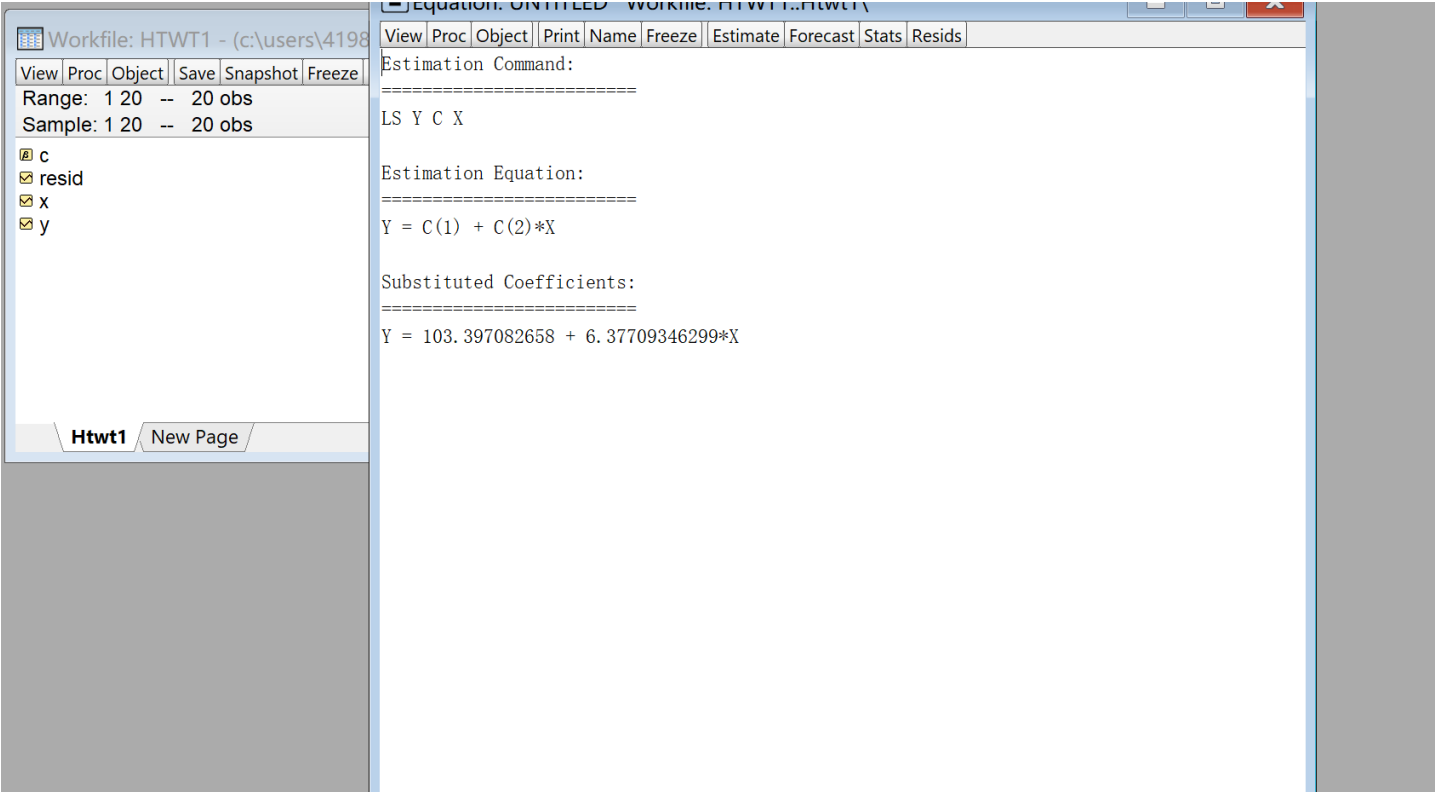
t-Statistic t统计量的值

Prob. 系数的显著性水平（p值）

R-squared R方

Prob（F-statistic）模型的显著性水平（p值）

结果



拟合结果

可以看到X的系数约为6.377，常数C的值约为103.3971，那么可以得出回归方程

$$\hat{Y}_i = 6.377093 \times X_i + 103.3971$$

也即

估计出的体重 = 103.3971 + 6.377 × 身高（单位：英寸，高于5英尺的部分）

与书中1-19得出结果近乎一致。

第四题

4 回到第 1.5 节中的住房价格模型。考察方程：

$$SIZE_i = -290 + 3.62PRICE_i \tag{1-23}$$

式中， $SIZE_i$ 代表第 i 套住房的面积（单位：平方英尺）； $PRICE_i$ 代表第 i 套住房的价格（单位：千美元）。

- a. 认真解释各回归参数估计值的意义。
- b. 假设该方程解释了住房面积变动的大部分原因（超过 80%）。方程是否表明，高价格的住房面积也大？如果不是，方程表明了什么？
- c. 如果价格单位不是千美元，而是美元，方程的参数估计值会有什么变化？请详细阐述。

解答

- a. $SIZE_i = -290 + 3.62PRICE_i$ 之中包含了两个参数，3.62 和 -290
- 3.62：是估计的斜率，其标识这栋房子的价格上涨1千美元的房子面积的改变（以平方英尺为单位）。
 - 290：是估计的截距，代表当价格为零时，这栋房子对应的面积。（这里对应面积为-290，为负数，显然不可能，造成这样很有可能因为该估计值可能包含遗漏变量的价值常数）
- b. 不正确，方程并不表明高价格意味着房子面积大或者低价格意味着房子面积小，方程所展现的是存在于房价和其大小之间的统计学联系。
- c. 新的斜率参数由3.62变为0.00362，其余估值完全不变。即方程变为
- $$SIZE_i = -290 + 0.00362PRICE_i$$

第六题

- 6 你的朋友在大学里有一份工作，就是给毕业的校友打电话，寻求他们对学校年度基金的捐献。她想要知道她打去的电话是否能起到作用。为了检验学生的电话对基金捐献的影响，她搜集了50个校友的数据，并估计了如下的方程：

$$\widehat{GIFT}_i = 2.29 + 0.001INCOME_i + 4.62CALLS_i \quad (1-26)$$

式中， $GIFT_i$ 代表第*i*位校友对2016年年度基金的捐献金额（单位：美元）； $INCOME_i$ 代表估计的第*i*位校友在2016年的收入（单位：美元）； $CALLS_i$ 代表为鼓励校友捐献，给第*i*位校友打电话的次数。

- 详细说明每一个参数估计值的意义。估计参数的符号与你预期的相一致吗？
- 为什么方程中左边的变量是 \widehat{GIFT}_i 而不是 $GIFT_i$ ？
- 在估计出的方程中，你的朋友并没有加入随机误差项。这种做法对吗？请说明理由。
- 假如你的朋友决定将变量 $INCOME$ 的单位从“美元”变为“千美元”。方程中的参数估计值会发生什么变化？请详细说明。
- 如果方程中可以再加入一个变量，应该加入什么变量？请说明理由。

解答

- a. $\widehat{GIFT}_i = 2.29 + 0.001INCOME_i + 4.62CALLS_i$ 之中包含了3个参数，2.29、0.001 和 4.62
- 2.29：是估计的常数项，而这个估计值是某校友在**没有收入和没有被打进电话游说的情况下**捐赠的数额。
 - 0.001：是对收入的斜率参数的估计，其反应在**游说电话数保持不变的情况下**，校友收入每增加1美元可能会相应地将捐赠额增加多少。

- 4.62：是对电话斜率参数的估计，其反应在校友收入保持不变的情况下，每多打人一通电话可能会相应地将捐赠额提高多少。
 - 估计的斜率参数的符号符合我们的期望
- b. 方程的右边包含了估计值（如 $INCOME$ ），当我们估计了方程，左边的变量就是被解释变量的估计值，因此需要写为题中形式 \widehat{GIFT}_i ，其表明这是一个估计值
- c. 误差项是不可观测的，也不能引入我们实际上用来算 \widehat{GIFT}_i 的估计方程中。
- d. 方程会变为 $\widehat{GIFT}_i = 2.29 + 1.00INCOME_i + 4.62CALLS_i$ ，即除了收入的斜率参数的估计发生了数量级的改变，其他没有发生改变。
- e. 可以加入**校友年龄**，根据相关调查显示，在年龄为中年的人群往往更愿意为基金会捐款，而青年由于薪资收入不高，往往不愿意捐赠；老年群体因为没有收入也往往不会捐赠；而处于中年的群体往往事业有成，也对学校有着一份怀念之情，往往更容易发生捐赠。因此加入校友年龄可能会提升拟合的效果。