

计量经济学-作业3-假设检验



P90 习题5

- 5 利用第 5.3 节的方法，对以下双侧假设进行检验。
- a. 对于方程 (5-8)，在 5% 的显著性水平下，检验假设：
- $$H_0: \beta_2 = 160.0$$
- $$H_A: \beta_2 \neq 160.0$$
- b. 对于方程 (5-4)，在 1% 的显著性水平下，检验假设：
- $$H_0: \beta_3 = 0$$
- $$H_A: \beta_3 \neq 0$$
- c. 对于方程 (5-6)，在 5% 的显著性水平下，检验假设：
- $$H_0: \beta_2 = 0$$
- $$H_A: \beta_2 \neq 0$$

题目中涉及的方程如下——

位于回归参数估计量下方括号里的数值为估计量 $\hat{\beta}$ 的标准差

$$\hat{Y} = 300.0 + 10.0X_1 + 200.0X_2 \quad (5-8)$$

(1.0)	(25.0)
$t = 10.0$	8.0
$\bar{R}^2 = 0.90$	$N = 30$

$$\hat{Y}_i = 102.192 - 9.075N_i + 0.3547P_i + 1.288I_i \quad (5-4)$$

(2.053)	(0.0727)	(0.543)
$t = -4.42$	4.48	2.37
$N = 33$	$\bar{R}^2 = 0.579$	

$$CARS_i = 1.30 + 4.91YD_i + 0.00123PRICE_i - 7.14SUV_i \quad (5-6)$$

(2.38)	(0.00044)	(71.38)
$t = 2.1$	2.8	-0.1

回答

由公式 $t_k = \frac{\hat{\beta}_k - \beta_k}{SE(\hat{\beta}_k)}$

- a. 将数据 $\hat{\beta}_2 = 200.0$ 带入上述公式，可知 $t_2 = \frac{200 - 160}{25} = 1.6$ ，样本总数为 30，那么自由度为 $n - k - 1 = 30 - 2 - 1 = 27$ 查表知 $t_c = 2.052$ ，那么 $|t_2| < t_c$ ，因此不能拒绝假设 H_0 ，
- b. 将数据 $\hat{\beta}_3 = 1.288$ 带入上述公式，可知 $t_3 = \frac{1.288 - 0}{0.543} = 2.37$ ，样本总数为 33，那么自由度为 $n - k - 1 = 33 - 3 - 1 = 29$ 查表知 $t_c = 2.756$ ，那么 $|t_3| < t_c$ ，因此不能拒绝假设 H_0
- c. 将数据 $\hat{\beta}_2 = 0.00123$ 带入上述公式，可知 $t_2 = \frac{0.00123 - 0}{0.00044} = 2.8$ ，样本总数为 10，那么自由度为 $n - k - 1 = 10 - 3 - 1 = 6$ 查表知 $t_c = 2.447$ ，那么 $|t_3| > t_c$ ，因此拒绝假设 H_0 。

P91 习题6

- 6 假想现在要估计一个关于房价的模型，来推断面朝海滩的环境对房子价值的影响^⑨。由于一系列理论和数据的可获得性等方面的原因，所以，经过一番研究，决定使用地皮的面积而不是房子本身的面积作为其中一个变量。分析结果如下（括号内的数值为标准差）：

$$\widehat{PRICE}_i = 40 + 35.0 LOT_i - 2.0 AGE_i + 10.0 BED_i - 4.0 FIRE_i + 100 BEACH_i$$

$$(5.0) \quad (1.0) \quad (10.0) \quad (4.0) \quad (10)$$

$$N = 30 \quad \bar{R}^2 = 0.63$$

式中， $PRICE_i$ 代表第*i*幢房子的价格（单位：千万美元）； LOT_i 代表第*i*幢房子所占地皮的面积（单位：千万平方米）； AGE_i 代表第*i*幢房子的已建年限； BED_i 代表第*i*幢房子里的房间数目； $FIRE_i$ 代表第*i*幢房子是否有壁炉，为虚拟变量，如果带有的话则为1，否则为0； $BEACH_i$ 代表第*i*幢房子是不是朝向海滩，为虚拟变量，如果是则为1，否则为0。

- 预期变量 LOT ， BED ， $BEACH$ 的参数为正。建立在 5% 的显著性水平下的恰当假设，并对其进行检验。
- 预期变量 AGE 的参数为负。建立在 10% 的显著性水平下的恰当假设，并对其进行检验。
- 首先，预期变量 $FIRE$ 的参数为正，但你的一个朋友说壁炉很脏，使房间难以保持清洁，使得你不能确定它的参数的符号。运用以 0 为中心的双侧 t 检验，在 5% 的显著性水平下对其进行检验。
- 方程中出现了哪些问题？（提示：是否存在参数估计值的符号与预期不符的情况？是否存在没有显著地异于 0 的参数？）
- 在 d 中列举的问题中，哪个是你最一筹莫展的？请解释原因。
- 对于这个问题，有没有想到解决方案？

回答

- 对于待检验的三个参数我们做出如下假设： $H_0 : \beta \leq 0$ $H_A : \beta > 0$ ，在显著性水平 5% 的条件下分别对三个参数进行假设检验
 - 首先，样本总数为 30，可以求得自由度为 $n - k - 1 = 30 - 5 - 1 = 24$ ，查表得单侧 $t_c = 1.711$
 - $LOT : \hat{\beta}_1 = 35.0$ ，那么 $t_1 = \frac{35 - 0}{5} = 7$ ，可知 $|7| > 1.711$ 即 $|t_1| > t_c$ ，且其预期符号和对立假设相同（均为正数），因此**可以拒绝** H_0
 - $BED : \hat{\beta}_3 = 10$ ，那么 $t_3 = \frac{10 - 0}{10} = 1$ ，可知 $|1| < 1.711$ 即 $|t_3| < t_c$ ，因此**不能拒绝** H_0
 - $BEACH : \hat{\beta}_5 = 100$ ，那么 $t_5 = \frac{100 - 0}{10} = 10$ ，可知 $|10| > 1.711$ 即 $|t_5| > t_c$ ，且其预期符号和对立假设相同（均为正数），因此**可以拒绝** H_0
- 对于待检验的 AGE 参数我们做出如下假设： $H_0 : \beta \geq 0$ $H_A : \beta < 0$ ，在显著性水平 10% 且自由度为 24 的条件下， $t_c = 1.318$ 。 $\hat{\beta}_2 = -2$ ，那么 $t_2 = \frac{-2 - 0}{1} = -2$ ，可知 $|-2| > 1.318$ 即 $|t_2| > t_c$ ，且其预期符号和对立假设相同（均为负数），因此**可以拒绝** H_0

- c. 对于待检验的 **FIRE** 参数我们做出如下假设: $H_0: \beta = 0$ $H_A: \beta \neq 0$, 在显著性水平5%且自由度为24的条件下, $t_c = 2.064$. $\hat{\beta}_2 = -4$, 那么 $t_2 = \frac{-4 - 0}{4} = -1$, 可知 $|-1| < 2.064$ 即 $|t_2| < t_c$, 因此**不能拒绝** H_0
- d. 可能存在的问题:
- **多重共线问题**: BED似乎与LOT有高度相关性(共线性)例如, 房间数目和房屋面积可能存在的较高的相关性。多重共线性可能导致参数估计不准确, 使模型的解释能力下降。
 - **关联度低**: BED和FIRE的系数与零相差不大。这两个变量可能对房价的影响很小。
 - **遗漏变量**: 公式中可能存在未考虑的其他重要影响房价的变量。
- e. 最令人担心的是**BED似乎与LOT有极大的关联性**(多重共线问题), 这会使得方程的经济学意义变得更差, 也可能极大的影响了方程其他系数的估计值
- f. 因为建在较大地块上的海滩屋相较于建在较小地块上的海滩屋更倾向于拥有更多卧室, 导致 **LOT** 和 **BED** 有高度相关性(共线性)。因此如果 **LOT** 在方程中的话, **BED** 就不应该属于方程, 应该将其删除。

P91 习题7

- 7 假设圣迭戈教士棒球队邀请你通过分析去年这支球队在旧体育场馆中每场比赛的观众人数, 来对他们的新体育馆的经济影响进行评估。在对这一课题进行研究后, 你建立了以下模型(括号里的数值为标准差):

$$\widehat{ATT}_i = 25\,000 + 15\,000WIN_i + 4\,000FREE_i - 3\,000DAY_i - 12\,000WEEK_i$$

(15 000) (2 000) (3 000) (3 000)

$N = 35$ $\bar{R}^2 = 0.41$

式中, ATT_i 代表第*i*场比赛的观众人数; WIN_i 代表第*i*场比赛对手赢的概率; $FREE_i$ 代表第*i*场比赛是不是一场“激励”性质的比赛, 即有免费礼物送给到场球迷的比赛, 为虚拟变量, 如果是的话则为1, 否则为0; DAY_i 代表第*i*场比赛是否在白天举行, 为虚拟变量, 如果是白天举行则为1, 如果比赛为夜场或黄昏时举行则为0; $WEEK_i$ 代表第*i*场比赛是否在工作日举行, 为虚拟变量, 如果在工作日进行则为1, 如果在周末进行则为0。

- 预期变量 WIN 和 $FREE$ 的参数为正。建立在5%的显著性水平下的恰当假设, 并对其进行检验。
- 预期变量 $WEEK$ 的参数为负。建立在1%的显著性水平下的恰当假设, 并对其进行检验。
- 将变量 DAY 也加入模型中, 是因为管理者认为它很重要, 但是不能确定日场比赛对观众人数的影响。在5%的显著性水平下, 运用以0为中心的双侧*t*检验, 对其进行检验。
- 方程中会出现什么问题?(提示: 是否存在与预期不符的参数符号? 是否不存在显著地异于0的参数?)
- 在d中列举的问题中哪个是你最一筹莫展的? 请解释原因。
- 对于这个问题, 有没有想到解决方案?(提示: 不是只有运动迷才能回答这个问题。打个比方, 如果你喜欢音乐, 思考一下露天音乐会的到场人数。)

回答

- a. 对于待检验的两个参数我们做出如下假设： $H_0 : \beta \leq 0$ $H_A : \beta > 0$ ，在显著性水平5%的条件下分别对两个参数进行假设检验
- 首先，样本总数为35，可以求得自由度为 $n - k - 1 = 35 - 4 - 1 = 30$ ，查表得单侧 $t_c = 1.697$
 - WIN: $\hat{\beta}_1 = 15000$ ，那么 $t_1 = \frac{15000 - 0}{15000} = 1$ ，可知 $|1| < 1.697$ 即 $|t_1| < t_c$ ，因此**不能拒绝** H_0
 - FREE: $\hat{\beta}_2 = 4000$ ，那么 $t_2 = \frac{4000 - 0}{2000} = 2$ ，可知 $|2| > 1.697$ 即 $|t_2| > t_c$ ，且其预期符号和对立假设相同（均为正数），因此**可以拒绝** H_0
- b. 对于待检验的 WEEK 参数我们做出如下假设： $H_0 : \beta \geq 0$ $H_A : \beta < 0$ ，在显著性水平1%且自由度为30的条件下， $t_c = 2.457$ 。 $\hat{\beta}_4 = -12000$ ，那么 $t_4 = \frac{-12000 - 0}{3000} = -4$ ，可知 $|-4| > 2.457$ 即 $|t_4| > t_c$ ，且其预期符号和对立假设相同（均为负数），因此**可以拒绝** H_0
- c. 对于待检验的 DAY 参数我们做出如下假设： $H_0 : \beta = 0$ $H_A : \beta \neq 0$ ，在显著性水平5%且自由度为30的条件下， $t_c = 2.042$ 。 $\hat{\beta}_3 = -3000$ ，那么 $t_3 = \frac{-3000 - 0}{3000} = -1$ ，可知 $|-1| < 2.042$ 即 $|t_3| < t_c$ ，因此**不能拒绝** H_0
- d. 可能存在以下的问题：
- 关联度低**：DAY和WIN的系数与零相差不大。这两个变量可能对观众人数的影响很小。
 - 遗漏变量**：公式中可能存在未考虑的其他重要影响观众人数的变量。
- e. 潜在的遗漏变量更令我担忧，因为他可能极大的影响了方程其他系数的估计值，也会使得方程的经济学意义受到很大影响
- f. 在我看来可能遗漏了 天气WEATHER 这一变量，在天气好的时候，人们才更愿意来观看比赛。可以在方程中新增这个变量重新进行拟合。

P93 习题9

本页剩余空间无法展示原始问题截图，请接下一页

- 9 弗雷德里克·舒特 (Frederick Schut) 和范·贝赫艾克 (Van Bergeijk)^⑨在发表的文章中, 采用了 32 个国家组成的截面数据来估计药品价格模型, 以弄清医药行业是否存在国际价格歧视。作者认为如果存在价格歧视, 则在恰当设定的价格方程中, 人均收入的参数应显著地为正。之所以这样认为, 是因为他们觉得人均收入是按下列方式来度量价格歧视的: 支付能力越高, 药品需求的价格弹性就越低 (以绝对值形式), 从而价格歧视者索取的价格就越高。而且, 作者认为如果允许存在药品专利, 药品价格会较高, 而对药品进行管制、鼓励竞争或一个国家的药品市场规模相对比较大的话, 药品价格会较低。估计结果如下所示 (括号内的数值为标准差):

$$\hat{P}_i = 38.22 + 1.43GDPN_i - 0.6CVN_i + 7.31PP_i - 15.63DPC_i - 11.38IPC_i \quad (5-10)$$

(0.21)	(0.22)	(6.12)	(6.93)	(7.16)
$t = 6.69$	-2.66	1.19	-2.25	-1.59
$N = 32$		$\bar{R}^2 = 0.775$		

式中, P_i 代表第 i 个国家与美国相比的相对药品价格水平; $GDPN_i$ 代表第 i 个国家相对于美国的人均国内生产总值; CVN_i 代表第 i 个国家相对于美国的人均药品消费量; PP_i 代表医药产品的专利在第 i 个国家是否得到认证, 为虚拟变量, 如果得到认证则为 1, 否则为 0; DPC_i 代

表第 i 个国家是否实施严格的价格管制, 为虚拟变量, 如果施严格管制则为 1, 否则为 0; IPC_i 代表第 i 个国家鼓励价格竞争, 为虚拟变量, 如果鼓励竞争则为 1, 否则为 0。

- a. 建立关于方程中参数的合理假设, 并在 5% 的显著性水平下, 对其进行 t 检验。
- b. 在置信水平为 90% 的情况下, 求出每个斜率参数的置信区间。
- c. 你认同弗雷德里克·舒特和范·贝赫艾克关于存在价格歧视的观点吗? 为什么?
- d. 如果作者没有将每个国家的药品价格、人均收入、人均药品消费与美国的相比, 估计结果会有何不同? 请解释你的观点。
- e. 使用 EViews 软件或计算机中的其他电脑软件以及表 5-2 中的数据重现回归结果。

回答

- a. 在显著性水平 5% 的条件下分别对 5 个参数进行假设检验。其中, 样本总数为 32, 可以求得自由度为 $n - k - 1 = 32 - 5 - 1 = 26$, 查表得单侧 $t_c = 1.706$
 - GDPN: 做出假设 $H_0: \beta \leq 0$ $H_A: \beta > 0$, 且预测参数为正。 $t_1 = 6.69$, 可知 $|6.69| > 1.706$ 即 $|t_1| > t_c$, 且其预期符号和对立假设相同 (均为正数), 因此**可以拒绝** H_0
 - CVN: 做出假设 $H_0: \beta \geq 0$ $H_A: \beta < 0$, 且预测参数为负 (因为根据题意一个国家的药品市场规模相对比较大的话, 药品价格会较低)。 $t_2 = -2.66$, 可知 $|-2.66| > 1.706$ 即 $|t_2| > t_c$, 且其预期符号和对立假设相同 (均为负数), 因此**可以拒绝** H_0
 - PP: 做出假设 $H_0: \beta \leq 0$ $H_A: \beta > 0$, 且预测参数为正。 (因为根据题意如果允许存在药品专利, 药品价格会较高)。 $t_3 = 1.19$, 可知 $|1.19| < 1.706$ 即 $|t_3| < t_c$, 因此**不能拒绝** H_0

- **DPC** :做出假设 $H_0 : \beta \geq 0$ $H_A : \beta < 0$,且预测参数为负 (因为根据题意一个国家如果对药品进行管制, 药品价格会较低) 。 $t_4 = -2.25$, 可知 $|-2.25| > 1.706$ 即 $|t_4| > t_c$, 且其预期符号和对立假设相同 (均为负数) , 因此**可以拒绝** H_0
- **IPC** :做出假设 $H_0 : \beta \geq 0$ $H_A : \beta < 0$,且预测参数为负 (因为根据题意如果允许药品价格竞争, 药品价格会较低) 。 $t_5 = 1.59$, 可知 $|1.59| < 1.706$ 即 $|t_5| < t_c$, 因此**不能拒绝** H_0

b. 置信区间可以表示为 $(\hat{\beta} - t_c \cdot SE(\hat{\beta}), \hat{\beta} + t_c \cdot SE(\hat{\beta}))$, 在显著性水平10% (置信水平90%) 且自由度为26的条件下, $t_c = 1.706$.

- **GDPN** :
 $(\hat{\beta}_1 - t_c \cdot SE(\hat{\beta}_1), \hat{\beta}_1 + t_c \cdot SE(\hat{\beta}_1)) = (1.43 - 1.706 \times 0.21, 1.43 + 1.706 \times 0.21)$
 \Rightarrow 置信区间为 $1.07 < \hat{\beta} < 1.79$
- **CVN** :
 $(\hat{\beta}_2 - t_c \cdot SE(\hat{\beta}_2), \hat{\beta}_2 + t_c \cdot SE(\hat{\beta}_2)) = (-0.6 - 1.706 \times 0.22, -0.6 + 1.706 \times 0.22)$
 \Rightarrow 置信区间为 $-0.98 < \hat{\beta} < -0.22$
- **PP** :
 $(\hat{\beta}_3 - t_c \cdot SE(\hat{\beta}_3), \hat{\beta}_3 + t_c \cdot SE(\hat{\beta}_3)) = (7.31 - 1.706 \times 6.12, 7.31 + 1.706 \times 6.12)$
 \Rightarrow 置信区间为 $-3.13 < \hat{\beta} < 17.75$
- **DPC** :
 $(\hat{\beta}_4 - t_c \cdot SE(\hat{\beta}_4), \hat{\beta}_4 + t_c \cdot SE(\hat{\beta}_4)) = (-15.63 - 1.706 \times 6.93, -15.63 + 1.706 \times 6.93)$
 \Rightarrow 置信区间为 $-27.45 < \hat{\beta} < -3.81$
- **IPC** :
 $(\hat{\beta}_5 - t_c \cdot SE(\hat{\beta}_5), \hat{\beta}_5 + t_c \cdot SE(\hat{\beta}_5)) = (-11.38 - 1.706 \times 7.16, -11.38 + 1.706 \times 7.16)$
 \Rightarrow 置信区间为 $-23.59 < \hat{\beta} < 0.83$

c. 我**同意**其关于价格歧视的观点, 因为这样的估计经过检验,对参数的符号等重要内容符合预期, 具有统计学意义, 总体拟合良好。

d. 相关参数的系数的大小会改变, 但不会改变其符号。

e. 相关过程见下面的部分

回归重现 (e题)

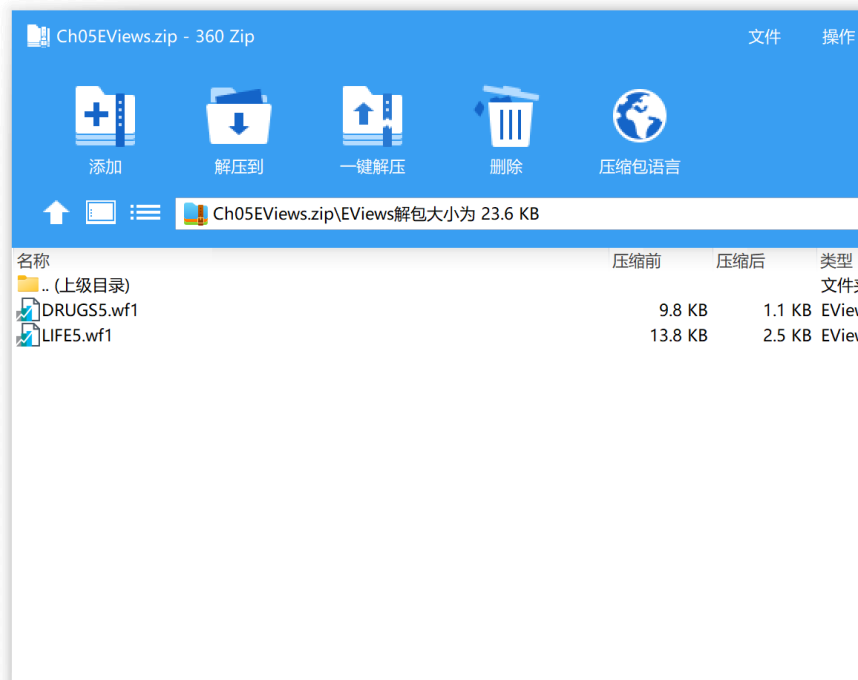
1. 下载相关数据集

本页剩余内容无法展示截图, 请看下一页。

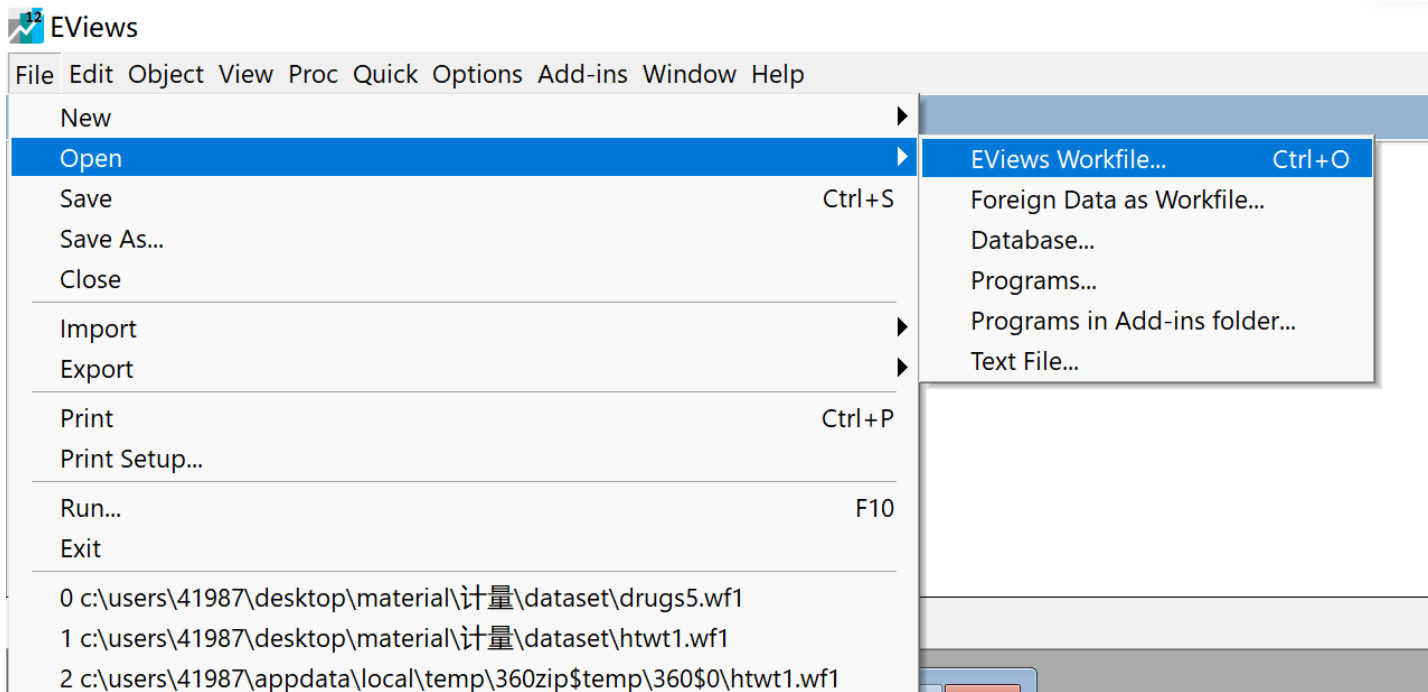
Data Sets for Using Econometrics: A Practical Guide, 7/E

Download data sets for each chapter in ASCII, EViews, Excel and Stata formats in .zip archive format:

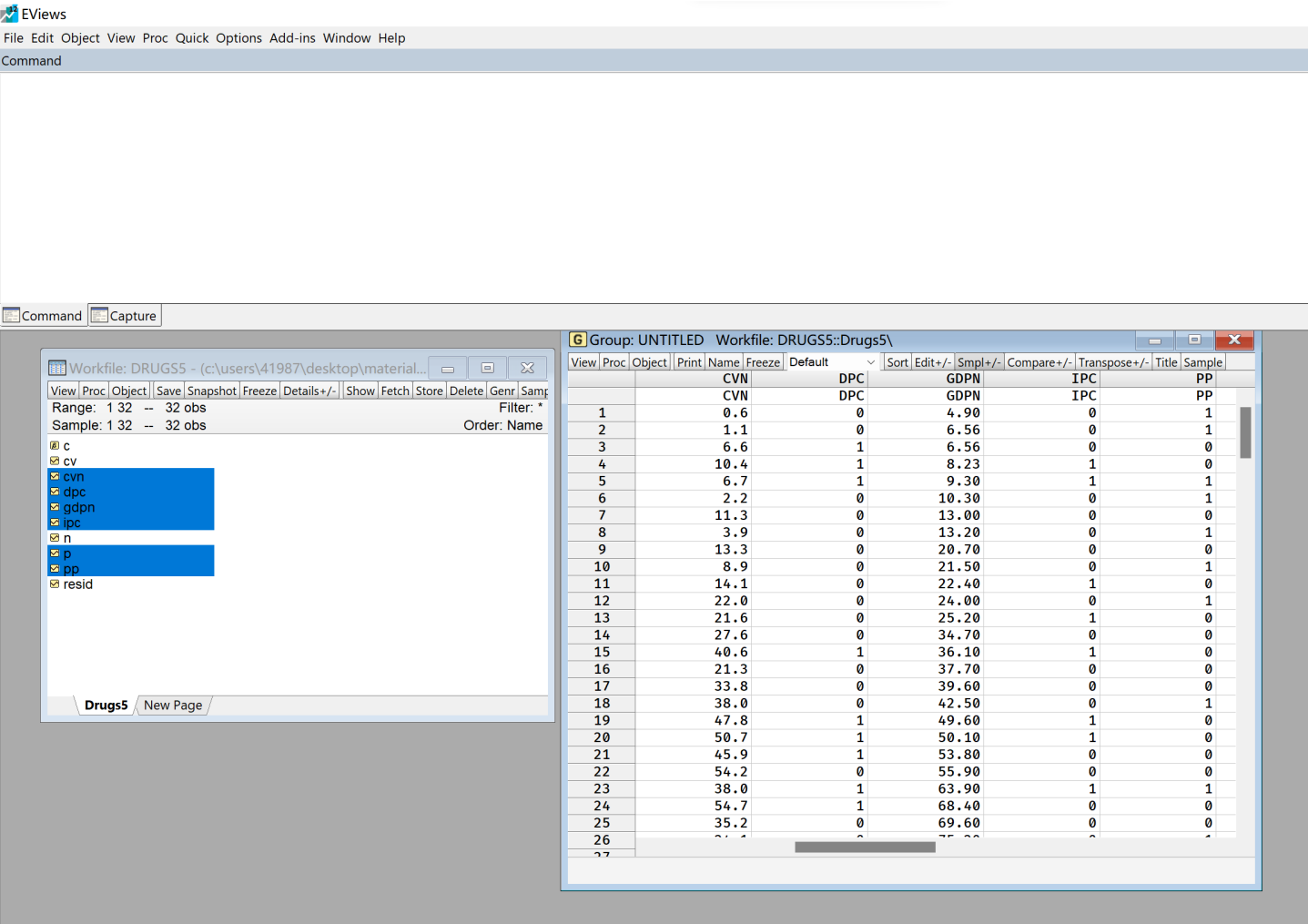
- [Download All Data Sets](#)
- Chapter 1:
 - [ASCII](#)
 - [EViews](#)
 - [Excel](#)
 - [Stata](#)
- Chapter 2:
 - [ASCII](#)
 - [EViews](#)
 - [Excel](#)
 - [Stata](#)
- Chapter 3:
 - [ASCII](#)
 - [EViews](#)
 - [Excel](#)
 - [Stata](#)
- Chapter 4:
 - [ASCII](#)
 - [EViews](#)
 - [Excel](#)
 - [Stata](#)
- Chapter 5:
 - [ASCII](#)
 - [EViews](#)
 - [Excel](#)
 - [Stata](#)
- Chapter 7:
 - [ASCII](#)
 - [EViews](#)
 - [Excel](#)
 - [Stata](#)
- Chapter 8:
 - [ASCII](#)
 - [EViews](#)
 - [Excel](#)
 - [Stata](#)
- Chapter 10:
 - [ASCII](#)
 - [EViews](#)
 - [Excel](#)
 - [Stata](#)
- Chapter 11:
 - [ASCII](#)
 - [EViews](#)
 - [Excel](#)
 - [Stata](#)
- Chapter 12:
 - [ASCII](#)
 - [EViews](#)
 - [Excel](#)
 - [Stata](#)
- Chapter 13:
 - [ASCII](#)
 - [EViews](#)
 - [Excel](#)
 - [Stata](#)
- Chapter 14:
 - [ASCII](#)
 - [EViews](#)
 - [Excel](#)
 - [Stata](#)
- Chapter 15:
 - [ASCII](#)
 - [EViews](#)
 - [Excel](#)
 - [Stata](#)



2. 打开review，并导入数据集



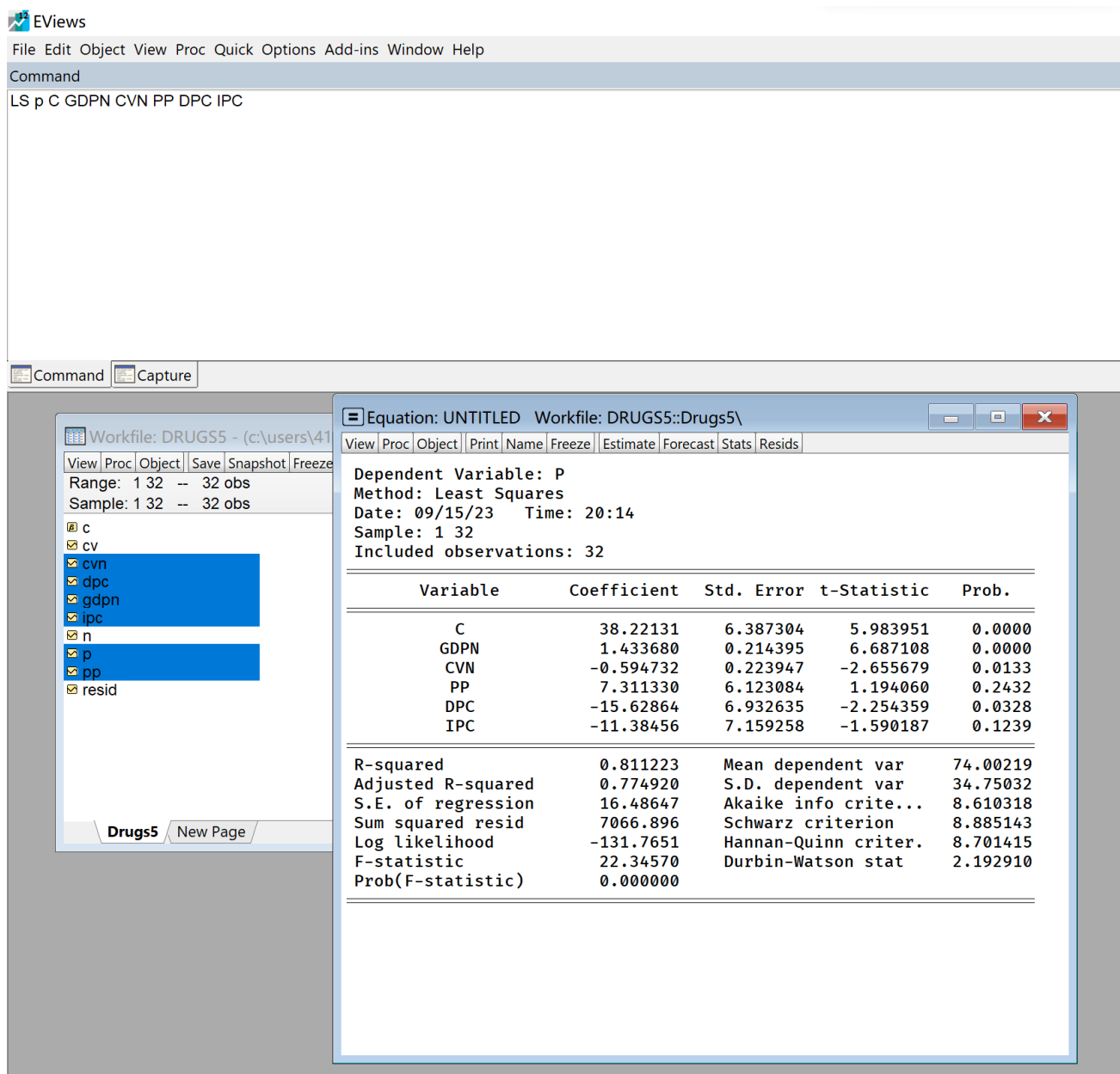
3. 导入成功



4. 输入指令 `LS P C GDPN CVN PP DPC IPC`

指令的意思是用最小二乘法进行拟合，被解释变量为 P ， $GDPN$ 、 CVN 、 PP 、 DPC 、 IPC 均为解释变量， c 为常数项

本页剩余空间无法展示截图，相关截图请看下一页



可以看到其系数（coefficient列）和题目一致，拟合的曲线为

$$\hat{P}_i = 38.22 + 1.43GDPN_i - 0.6CVN_i + 7.31PP_i - 15.63DPC_i - 11.38IPC_i$$

与题目中一致。

且第二列（Std.Error）列代表计算出的标准差；第三列代表与0相比的t估计量，均与题目一致。因此复现成功。