《计量经济学导论》

第3次作业 虚拟变量

说明：1）本次作业作业请于5月30日前交，可提前交，但不接受迟交；

2）本次作业需同时提交电子版与纸质版，电子版通过qq发送给助教；

1. 假定你通过调查收集到了关于工资、教育程度、经验以及性别等个人信息。除此之外，你还询问了有关个人吸食大麻的情况（询问被调查者“在过去的一个月之内吸食了多少次大麻”）。然后，建立了以下的模型：



因此，100 表示（每月）每增加一次大麻的吸食会导致工资的近似百分比变化。

1. **请修正模型使得你可以直接检验吸食大麻对男女生工资会有不同的影响；并说明在新建的模型中如何检验“吸食大麻对男女生工资具有相同的影响”的假设？**

**解答：**

在模型中引入usage与female的交互项usage·female，修正后的模型如下：

要检验“吸食大麻对男女生工资具有相同的影响”的假设，则令=0；≠0；

进行t检验，若显著，则拒绝原假设，支持备择假设；若不显著，则不能拒绝原假设。

1. **如果你认为不应该直接使用“吸食大麻的次数”作为解释变量，而最好将吸食大麻的变量表示成一个四值的分类变量：从不吸食，轻度吸食（每月1-5次）、中度吸食（每月6-10次）、重度吸食（每月10次以上）。那么此时的模型应该做出什么样的调整来使得你可以考察不同程度的大麻吸食对工资的影响？**

**解答：**

由题可得：

转化为四个二值虚拟变量

；；；；

引入，修正后的模型如下：

1. **在b）部分建立的模型中，如何检验大麻吸食对工资没有影响的原假设？**

**解答：**

原假设表述为===0；使用F检验。若F统计量显著，则拒绝原假设，支持备择假设；若F统计量不显著，则不能拒绝原假设。

**d）如果吸食大麻与饮酒之间是正相关的，并且饮酒与工资之间是负向关联（在控制教育、经验以及性别之后）。那么，如果遗漏反映个人饮酒情况的变量会对的估计产生什么样的影响？**

**解答：**

+

——吸食大麻与饮酒的相关性，为正数；——吸食大麻的样本方差；——饮酒的样本方差；——工资关于饮酒的回归系数，为负数；

因此，

2. 使用数据gpa2.dta完成以下的练习。以下是有关该数据的简要介绍：

该数据一共收集了4,137个大学生的信息，主要变量信息包括：

sat: 大学入学成绩

tothrs: total hours through fall semester学习时间

colgpa: grade point average (GPA) after fall semester GPA分数

athlete: =1 if athlete（运动员）

verbmath: verbal/math SAT score口语及学生成绩

hsize: size graduating class, 100s班级规模（单位为100人）

hsrank: rank in graduating class班级排名

hsperc: 100\*(hsrank/hssize) 班级排名

female: =1 if female

white: =1 if white

black: =1 if black

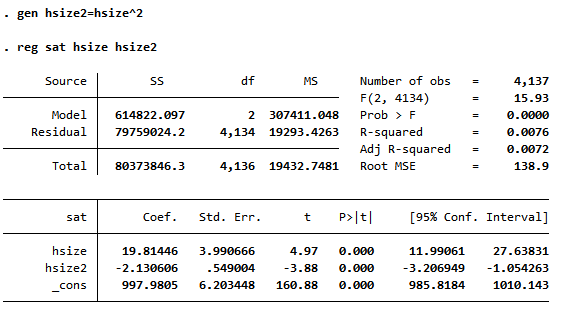
hsizesq: hsizeˆ2

**a）估计以下的模型：**

****

**其中hsize代表高中时的班级人数（单位是百人）。报告回归的基本结果。并检验hsize平方项的系数是否显著？**

**解答：**



SSM=614822.097 ,SSR=79759024.2 ,SST=80373846.3，R-squared=0.0076。hsize的回归系数为19.81446，标准误为3.990666；hsiz的回归系数为-2.130606,标准误为0.549004;截距系数为997.9805。

列出回归方程如下：

hsize平方项的t统计量为-3.88，显著。

**b）基于以上的估计结果，计算最优的班级规模。**

**解答：**

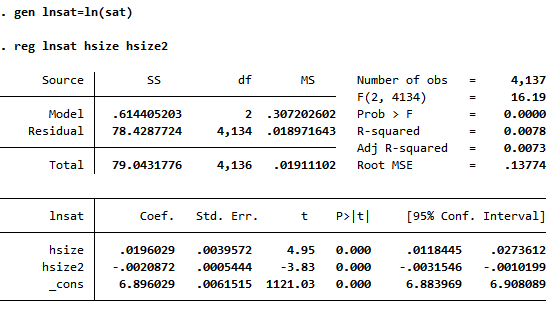
对回归方程求导：

令=0，求解得出hsize\*=4.65

即最优班级规模为465人。

**c)如果使用log(sat)作为被解释变量，再次估计模型，并计算最优的班级规模，请问与b）中的最优规模存在很大的差异吗？**

**解答：**



回归方程：

ln(

求导：

-0.0041744\*hsize

令=0，得hsize\*=4.70；

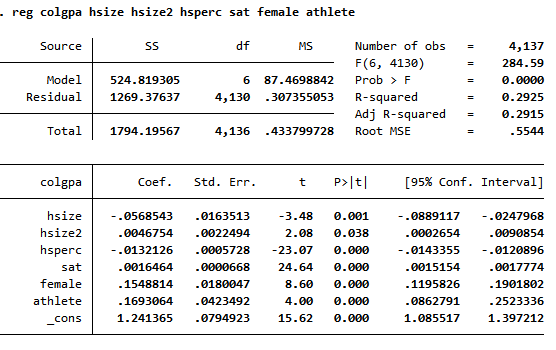
即最优班级规模为470人，与b)相差不大。

**d）估计以下模型：**

****

**请问估计值的含义。基于估计的结果，该估计值统计上是否显著？**

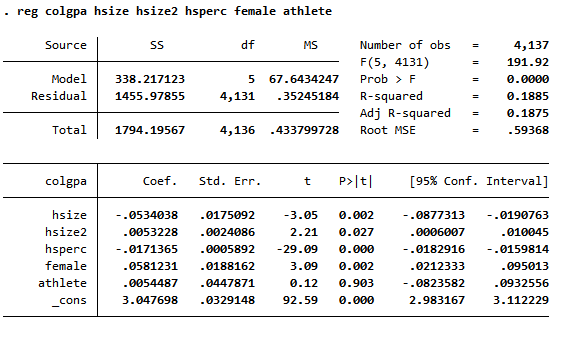
**解答：**



衡量的是运动员与非运动员的平均分数差异；t=4.00,在统计上显著。

**e)将sat从d）中的模型中删除，比较此时估计值是否发生了变化？为什么？**

**解答：**



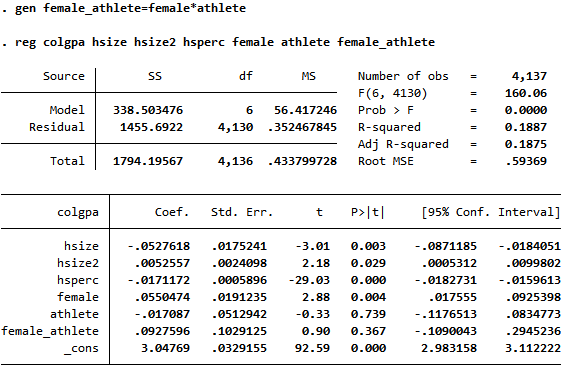
=0.0054487，与d)的估计值发生较大变化。由于sat与athlete具有一定的相关性，排除sat将导致偏误，使估计结果误差变大。

**f）如果在问题d）的模型中允许“是不是运动员”对成绩产生的影响在男女生中有不同，建立对应的模型，并运用估计的结果检验“女性运动员与女性非运动员的成绩没有差异”的原假设。**

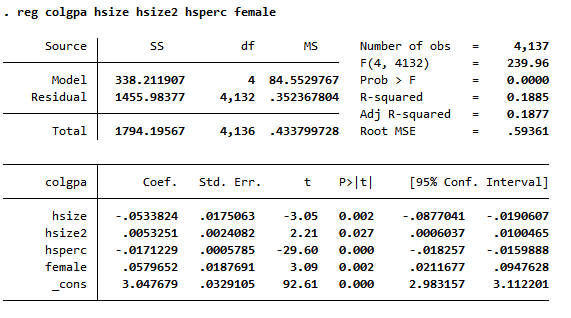
**解答：**

引入交互项athlete×female，修正模型如下：

不受约束模型：



受约束模型：



女性运动员与女性非运动员的成绩差异为，原假设表述为

构造F统计量：=0.4136,计算得p值=0.661292791，不显著，不能拒绝原假设。可认为女性运动员与女性非运动员的成绩没有差异。