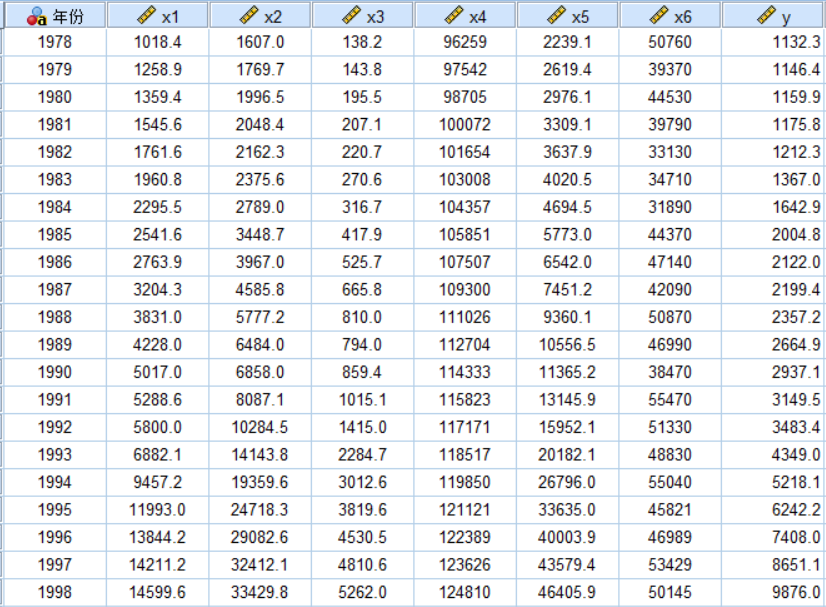
**多元统计分析第一次作业**

**题目**（教材P68第5题）：在研究国家财政收入时，以财政收入y为因变量，考虑以下几个自变量，x1为农业增加值（亿元）；x2为工业增加值（亿元）；x3为建筑业增加值（亿元）；x4为人口数（万人）；x5为社会消费总额（亿元）；x6为受灾面积（万公顷）。据《中国统计年鉴》获得1978-1998年共21个年份的统计数据见下表。试用不同的方法建立回归方程，并对所得模型的优良性做分析。



**解答**：拟选择三种方法建立回归方程。

（1）首先使用全选法进行计算，计算结果如下：

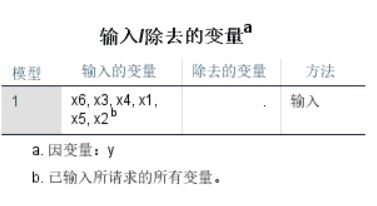


图1

该图1中的内容表示：方法选用了全选法，即自变量包括x1、x2、x3、x4、x5、x6，因变量为y，即国家年财政收入。



图2

图2中显示R方为0.996。R方是样本决定系数（coefficient of determination），其值用于检验模型对样本观测值的拟合程度，越接近于1，模型的拟合优度就越高。本结果的R方接近于1，说明模型能够很好地拟合样本观测值。可是，如果在模型中增加自变量的个数，R方的值也会随之增加，这会给人一种错觉：要想模型拟合效果好，就得尽可能多引进自变量。为了防止这种倾向，人们又定义了引进自由度的修正的R方，即“调整后R方”，一般这个更准确一点。本结果的“调整后R方”为0.995。R方是R的平方值。标准估算的错误的大小反映了建立的模型预测因变量时的精度，在对比多个回归模型的拟合效果时，通常会比较该指标。德宾-沃森检验的目的是为了检验一个回归模型的随机误差项是否存在自相关，即各个样本之间是否存在关联。本结果的德宾-沃森检验值为1.329，教材上讲当该值为2左右时，模型一般不存在一阶自相关，所以该模型存在一定程度的一阶自相关。而且，经验表明，如果模型不存在一阶自相关，一般也不存在高阶序列相关。

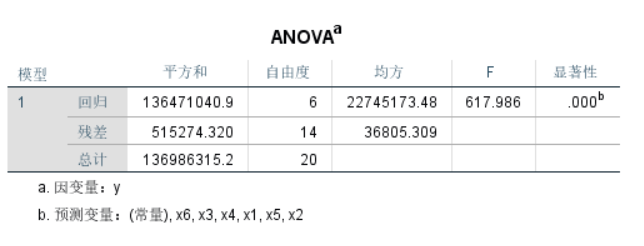
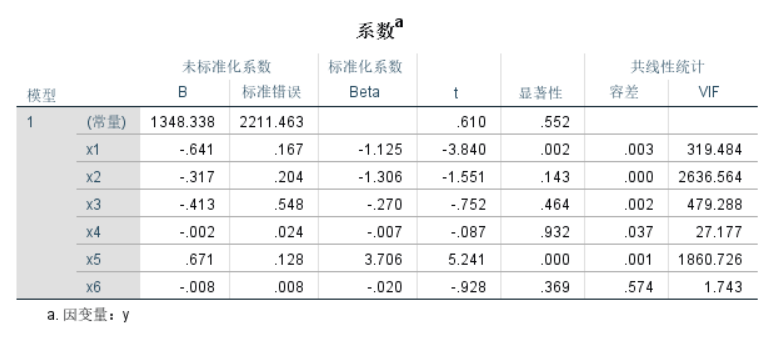


图3

图3表示方差分析结果，主要看F和显著性两个。F统计量的大小代表了线性回归方程是否显著，即因变量和自变量之间是否存在着线性关系，其值越大越好，此结果的F为617.986，表示因变量y和自变量x1-x6之间存在线性关系。显著性可直接与显著性水平α（0.01、0.05）比较得出结果，从表3中我们可以看到，显著性=0.000<0.01，即认为模型1在0.01显著性水平下，由自变量和因变量建立起来的线性关系具有极其显著的统计学意义。

图4

此图重点关注两个部分，一个是t和显著性，另一个是VIF。图3中的F统计量是对模型整体进行检验，这里的t和显著性是对每个变量进行显著性检验。由图4可以看出，以0.01和0.05为检验水平的话，x1、x5显著。VIF，即方差扩大因子法，用来检验模型是否存在多重共线性。由表4可以看出，该模型存在多重共线性。当发现存在多重共线性时，可以通过剔除一些不重要的自变量、增大样本容量等方法来克服多重共线性。

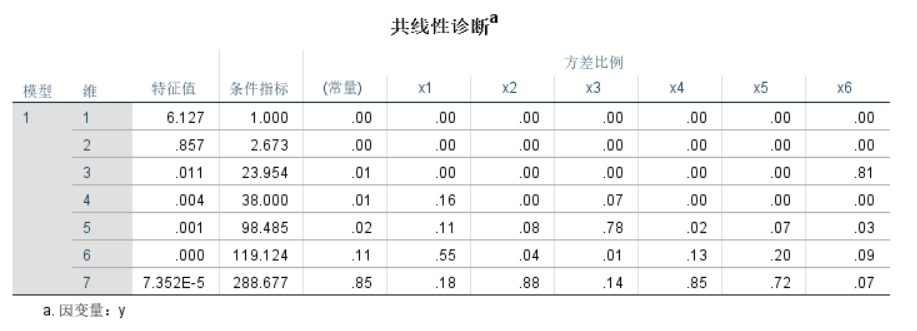


图5

多个维度特征值约为0证明可能存在多重共线性，条件指标大于10时提示有多重共线性。由表5可知，该模型存在多重共线性。

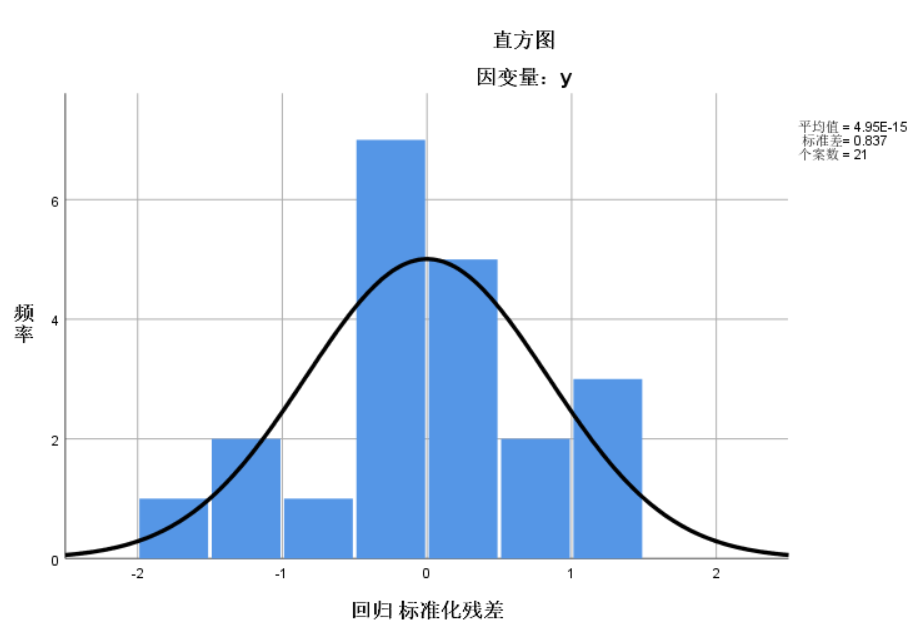


图6

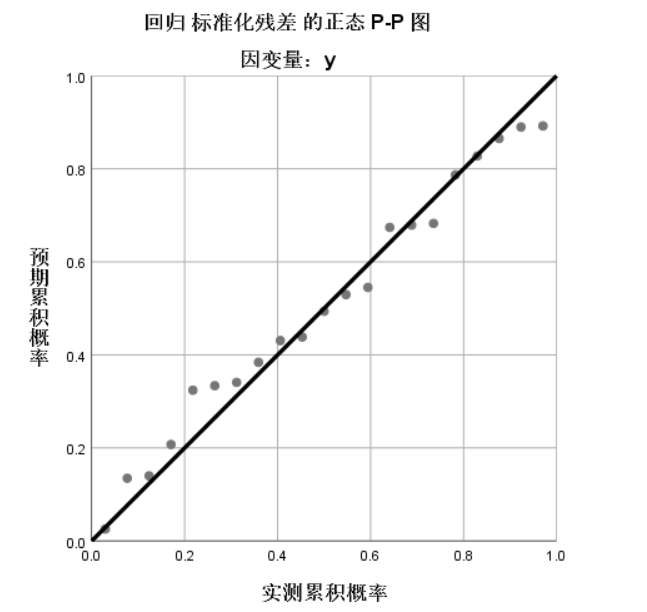


图7

图6、图7的目的是为了检验误差项的正态性。由图可知，本例中误差正态性的假设是合理的。

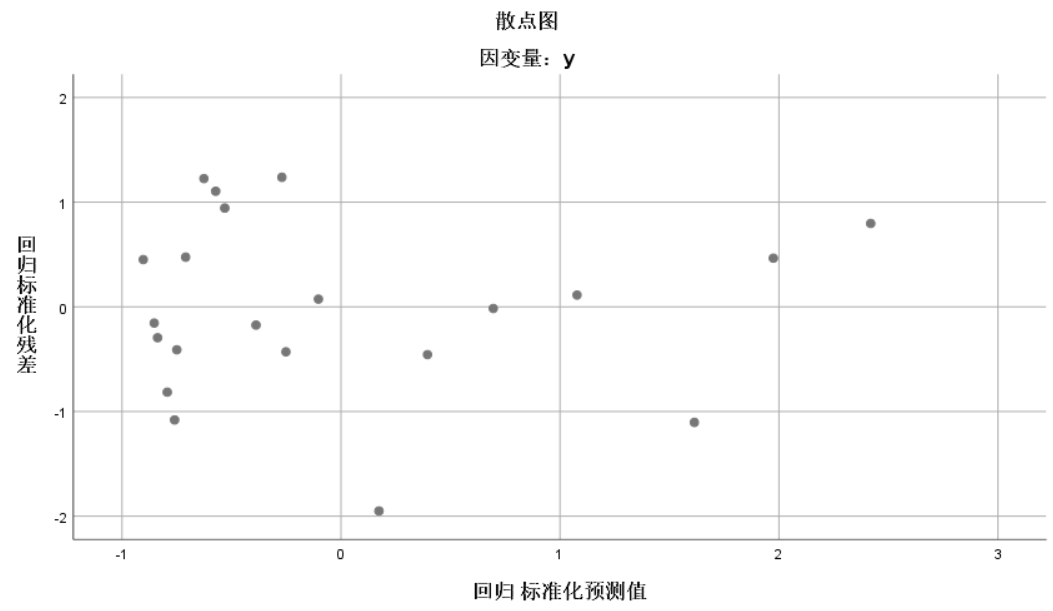


图8

绘制图8散点图的目的是为了对回归函数做线性诊断和检验误差方差齐性。由图8可知，散点大部分集中在区间（-1，1）之间，且分布较为均匀。因此，该例子中的回归函数是自变量的线性函数，且误差方差是齐性的。但由于样本容量较小，只有21个，效果不显著，可适量增加样本。

（2）后退法

首先用m个自变量与y建立一个回归方程，然后在这个方程中剔除一个最不重要的自变量，接着又利用剩下的m-1个自变量与y建立线性回归方程，再剔除一个最不重要的自变量，依次进行下去，直到没有自变量能够剔除为止。



图9

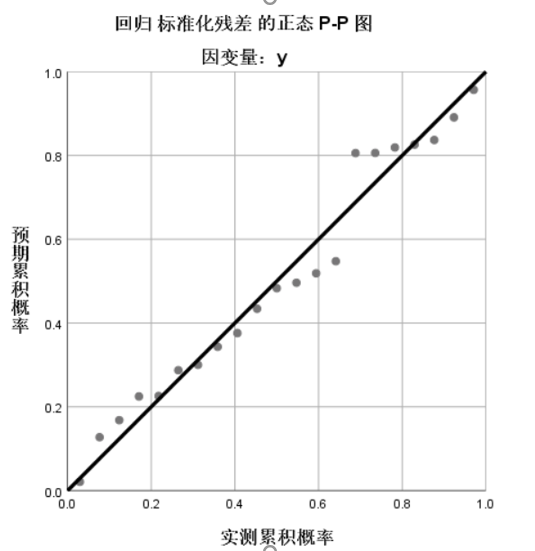
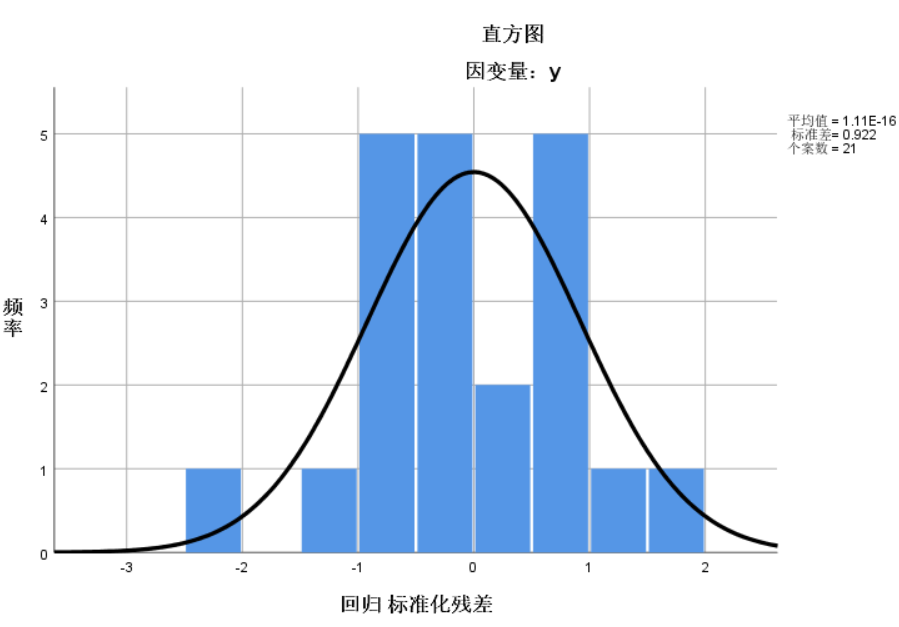


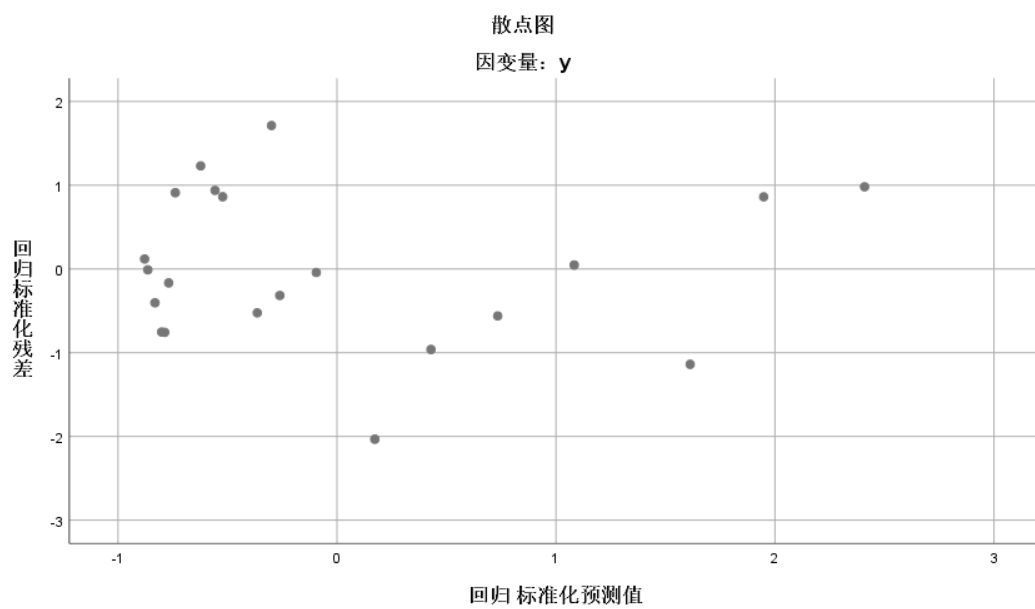
图10



图11

由图9、10、11可知，后退法最后留下的自变量包括x1、x2、x5三个。与全选法的结果进行对比，由图11可以发现，三个自变量的显著性得到了改善，三个自变量对因变量具有显著影响。但VIF还是比较大，表示还是存在多重共线性，必须采取一些策略。





组图12

图12为了检验误差项的正态性，为了对回归函数做线性诊断和检验误差方差齐性。由图可知，本例中误差正态性的假设是合理的，该例子中的回归函数是自变量的线性函数，且误差方差是齐性的。

（3）逐步回归法

基本思想：有进有出，将变量一个一个引入，引入变量的条件是通过了偏F统计量的检验。同时，每引入一个新变量后，对已入选方程的老变量进行检验，将经检验认为不显著的变量剔除，此过程经过若干步，直到既不能引入新变量，又不能剔除老变量为止。



图13

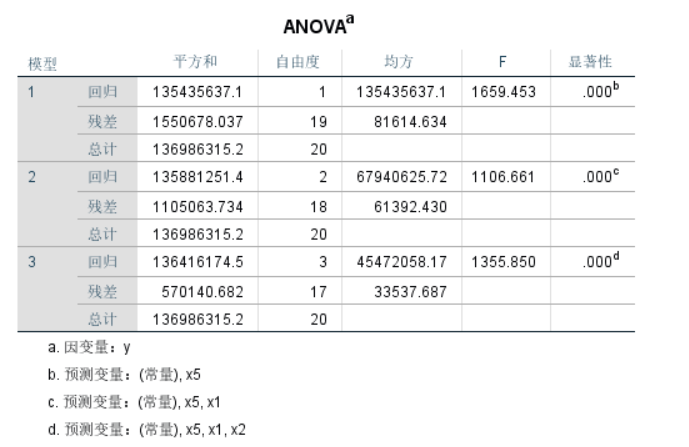


图14

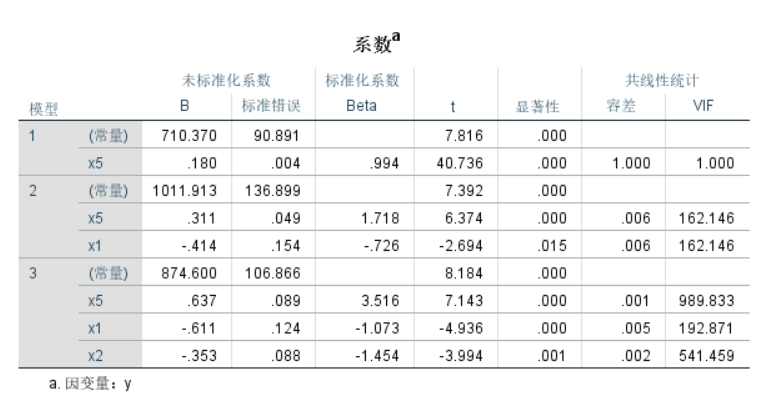
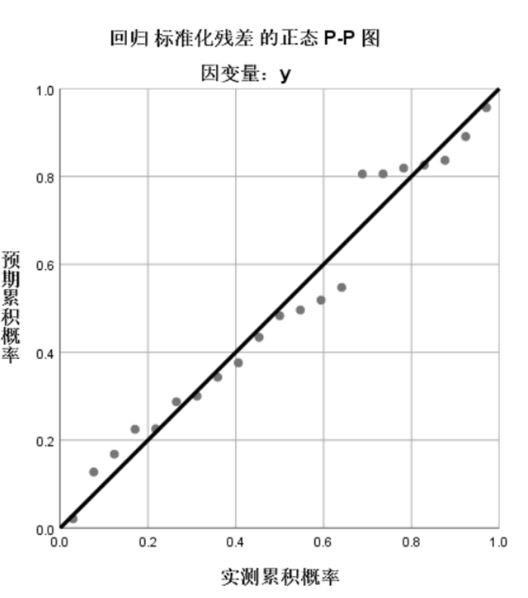
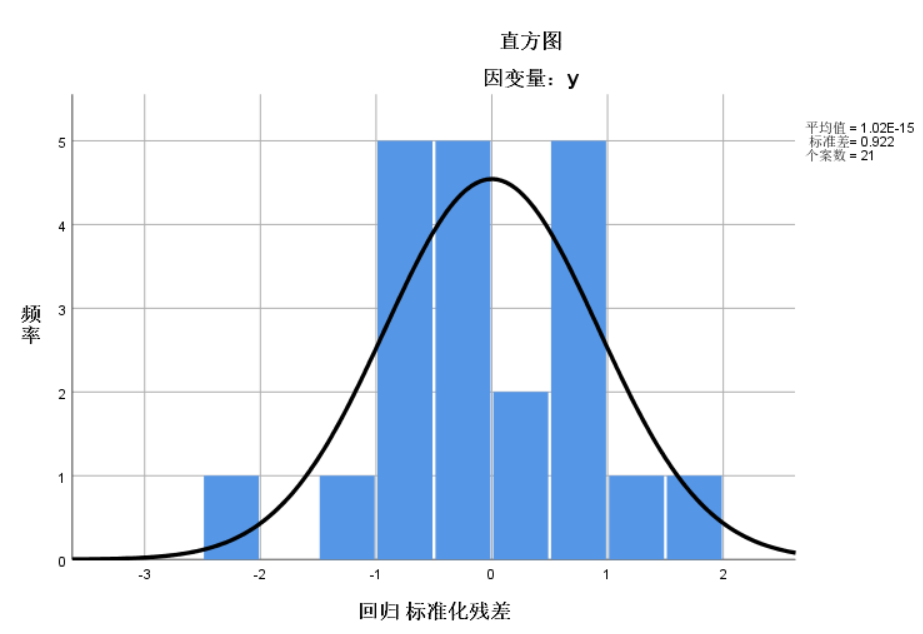
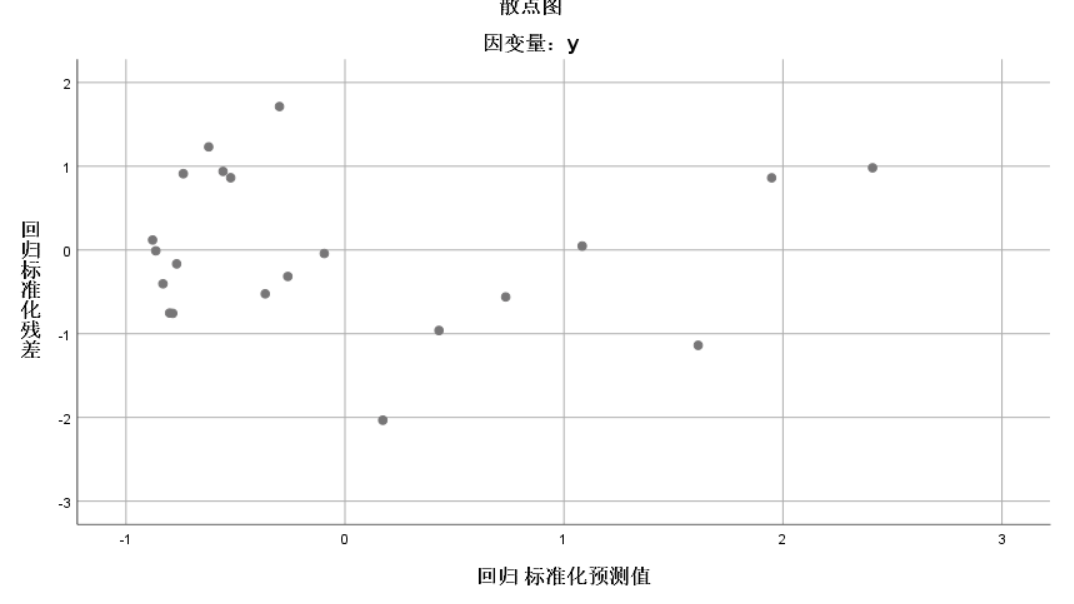


图15

由图15可知，逐步回归法和后退法得到的最终回归方程是一样的。





组图16

图16为了检验误差项的正态性，为了对回归函数做线性诊断和检验误差方差齐性。由图可知，本例中误差正态性的假设是合理的，该例子中的回归函数是自变量的线性函数，且误差方差是齐性的。

**总结**：通过本次作业，熟悉了SPSS软件的操作，学习了几种求解方法。