基于非参数bootstrap方法的一元线性回归分析原理与应用

1. 摘要

一元线性回归分析是分析一个因变量与一个自变量之间存在的线性关系。故构建因变量与自变量的关系模型：Y= a+bx+ε，ε~N（0，σ²），即一元线性回归模型。为了研究这种关系，近似地转而去研究因变量的数学期望与自变量之间的线性关系：μ(x）=a+bx，即Y关于x的回归函数。同时，运用非参数bootstrap方法，对回归函数中参数的估计量进行标准误差的估计，以及置信区间的求解。

关键词：一元线性回归 bootstrap方法

1. 引言

一元线性回归模型是最简单的计量经济学模型。在模型中只有一个解释变量, 其参数估计方法也是最简单的。通过最简单模型的参数估计, 可以较清楚地参数估计方法的原理。同时对于理解各类研究中的参数取样也具有及其重要意义。[1] 与此同时，当今计算机技术的高度发展 ,使统计研究 及其应用跃上了一个新台阶。 这不仅提高了计 算的速度 ,而且可以把统计学家从求解数学难 题中解放出来 ,并逐渐形成一种面向应用的、基于大量计算的统计思维—— 模拟抽样统计推断 , Bootstrap法就是其中的一种。[2]

1. 数学模型与原理

假设随机变量Y 与可控普通变量x之间存在相关关系，若直接研究Y与x之间的关系，情况比较复杂，所以寻求与Y近似的近似量来替代。设变量η为Y的近似量，二者的均方误差为E[(Y-η)²]，当η=E（Y）时，均方误差有最小值，即η与Y最为近似，故取Y的期望为近似值。此时，Y的期望为关于x的函数，故记为μ（x），为Y关于x 的回归函数。特殊地，若μ（x）=a+bx，即与x为一元线性关系，该函数为一元线性回归函数，进行一元线性回归分析。

则由Y~N（a+bx，σ²）得

Y=μ+ε=a+bx+ε

即建立了Y的模型，称为一元线性回归模型，其中b为回归系数，ε~N（0，σ²），为非人为可控的随机误差。

（一）a，b的估计

样本：（x1，Y1）…（xn,Yn）,由Y~N（a+bx，σ²），且Y1…Yn相互独立，故可由最大似然估计法估计未知参数a，b。

Y的联合概率密度为：

=

显然，要使L有最大值，只需Q=有最小值即可

故分别对Q求a，b的偏导，并令其为0，可解得a，b的最大似然估计值为

设 ，，

进而，可取为a+bx的估计，即

（二）σ²的估计

估计的均方误差 ]=E(ε²)=D（έ）+[E（έ）²=σ²，所以当σ²越小时，估计值a+bx的偏差越小，Y与x的关系越有效，故需对σ²的值进行估计。记为当x=xi时的估计值，则yi-为xi处的残差。

记为残差平方和，将其展开后可化解为。

又可得，故 ，即有，故可得σ²的无偏估计量：

（三）线性假设的显著性检验

实际问题中，要先对μ与x是否具有一元线性关系进行检验，此后的一元回归分析才有价值。故要对b是否为0进行检验。

：b=0. ：b≠0.

因为，且~，故在σ²未知的情况下，采用t检验法。可得，其中。当为真时，b=0，，故在显著性水平α下，可得的拒绝域为

（四）b的bootstrap置信区间

运用非参数bootstrap法求b的置信区间

设有样本S=【（x1，Y1）…（xn,Yn）】，由原样本可求得b的估计量，独立地从原样本中有放回地抽出B个容量为n的bootstrap样本，即子样本。对每个bootstrap样本分别求出b的估计：…。将得到的bootstrap估计按从小到大排列（不妨假定），取 用对应的的分布作为R(S)的近似分布，可求得R(S\*)的分布的近似分位数与，使得近似地有

记k1=[B], k2=[B]，在原样本S中分别以与为，的估计，故可得b的置信水平为1-α的bootstrap置信区间为

（）

（五）b的标准误差的bootstrap估计

在给出估计量的同时，还需知道这一估计的精度。故使用估计量的标准差来体现估计量的精度，故称为标准误差。

如（四）中所述产生boostrap样本，并独立地计算出…，

，其中

即为的bootstrap估计

1. 案例应用与讨论

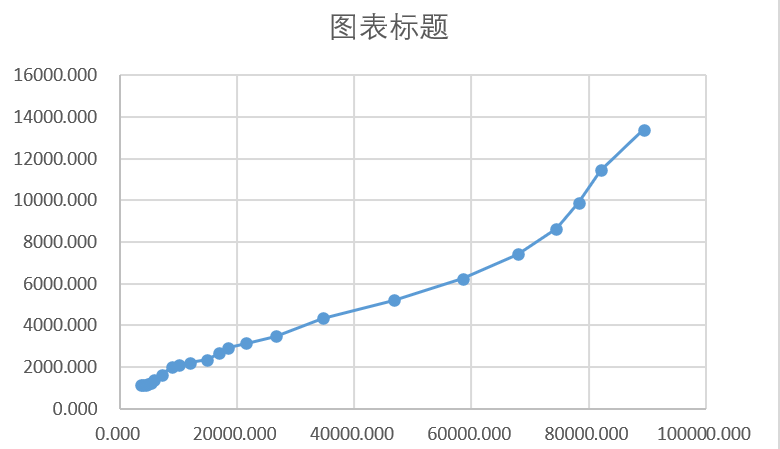
案例：（该案例数据来着焊工师王素霞）



案例分析：

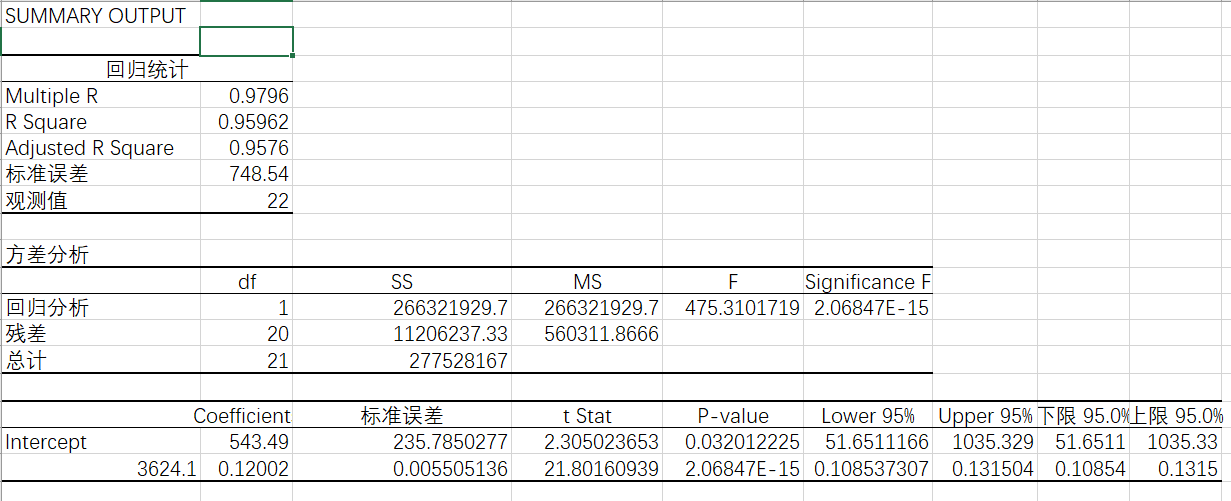
设GDP为x，财政收入为Y，研究x与Y的回归关系。

画出散点图：



由散点图上点的分布规律，可猜测Y与x存在一元线性回归关系y=a+bx。

故求回归方程（α=0.05）。



由excel计算结果可得

故一元线性回归方程为

进行线性假设的显著性检验：

：b=0. ：b≠0.

可得 p值=2.6847E-15<α=0.5，故拒绝，认为回归效果是显著的。

残差平方和，n=23

故=533630.349

同时，b的置信水平为0.95的置信区间为（0.10854，0.1315）

1. 结论

以上建立回归模型，借助一元线性回归函数，对两变量之间的线性关系进行研究，并对参数进行估计，同时采用非参数bootstrap方法，对参数的估计量进行标准误差的估计与置信区间的求解。回归模型不仅可以用来分析解释变量间的关系，还可用于实际的预测与控制，进而广泛应用于自然科学，经济管理，工业工程等领域。Bootstrap方法将原样本看作整体，将反复抽样得到的子样本看作样本，从而进行统计推断。Bootstrap是一种常用的统计推断方法, 它只依赖于给定的观测信息, 不需要其他假设和增加新的观测量, 可以比较方便地应用于实际的数据处理之中.在科学研究中, 它可以大大增强常用的估计、推断等方法。[3]

1. 参考文献

[1]任建英.一元线性回归分析及其应用[J].才智,2012(22):116-117.

[2]陈峰,陆守曾,杨珉.Bootstrap估计及其应用[J].中国卫生统计,1997(05):7-9.

[3]张萍.基于Bootstrap方法的统计分析[J].宜宾学院学报,2011,11(12):31-33.

[4]概率论与数理统计[M].浙江大学，盛骤，谢式千，潘承毅编2018年1月底16次印刷.

七、致谢

衷心感谢这一学期以来王琳老师在《概率论与数理统计》这门课上对我的教导，让我充分领略到了这门数学分支的神奇与奥秘，同时他全新的观点与对学习方法的认知也让我受益匪浅。同时，也要感谢助教老师，林思曼学姐为我们能够进行顺利的线上学习所作出的付出，以及平时耐心细致的解惑与作业批改。