线性方程组系数矩阵的秩与统计自由度的关系探讨

曹 昭

【摘 要】关于统计变量自由度的确定是统计研究的重要内容之一，以线性方程组系数矩阵的秩和其自由未知量的个数之间的关系为依据，对统计变量的自由度进行探讨，对我们把握自由度的实质和内涵有着重要的启迪意义。

【关键词】系数矩阵；秩；自由未知量；自由度

在统计学中，关于变量的自由度问题一直是一个让初学者非常迷惑的问题。而且，在一般的教科书或统计资料中，很少对统计变量的自由度问题给予充分的分析和讨论。这就使得很多人对自由度的含义缺乏正确的认识和理解。因此，对自由度的含义及其蕴含的数学思想，作出清楚的说明，就有着非常重要的理论和现实意义。本文试图通过分析线性方程组系数矩阵的秩与其自由未知量之间的关系，进一步探讨统计变量自由度的内涵。然后，对常用统计方法中变量自由度的确定作出理论上的阐释。

线性方程组系数矩阵的秩与自由未知量

我们知道，对于线性方程组来讲，其是否有解，取决于其系数矩阵或增广矩阵秩的取值情况。同时，其解的形式，也就是其解集中包含的自由未知量的个数也取决于线性方程组的系数矩阵或增广矩阵的秩的取值情况。接下来，我们分别就齐次线性方程组和非齐次线性方程组解的结构，与其系数矩阵和增广矩阵秩的关系进行简单分析和说明。

一般的齐次线性方程，可以用矩阵表示为如下等式：

$$AX=0$$

其中

$A=(a_{ij})m\times n$，$X=(x_1,x_2,\cdots x_n)^T$，

$0=(0,0,\cdots,0)^T$

根据线性代数的有关知识，我们可以得出以下几个方面的结论：

1. 如果线性方程组 $AX=0$ 的系数矩阵的秩 $r(A)=r$，且 r 大于 0 小于 n，则方程组会有无数个解，它的任一个基础解系都含有 $n-r$ 个线性无关的解向量。在这种情况下，如果我们对其系数矩阵进行初等变换，将其简化为阶梯型矩阵，我们会发现齐次线性方程组自由未知量的个数为 $n-r$ 个。

2. 如果线性方程组 $AX=0$ 的系数矩阵的秩 $r(A)=0$，则任意 n 个线性无关的向量都是次方程组的基础解系。如果我们对其系数矩阵进行初等变换，将其简化为阶梯型矩阵，我们会发现齐次线性方程组自由未知量的个数为 n 个。

3. 如果线性方程组 $AX=0$ 的系数矩阵的秩 $r(A)=n$，那么，次线性方程组具有唯一解，也就是零解。它不存在基础解系。如果我们对其系数矩阵进行初等变换，将其简化为阶梯型矩阵，我们可以看出齐次线性方程组自由未知量的个数为 0。

综合以上三个方面的内容，我们得知：为了求出线性方程组 $AX=0$ 的

基础解系或全部解，我们可以对其系数矩阵施以初等变换，化为简化的阶梯型矩阵。如果 $r(A)=r$，无论 r 的取值状况如何，则在其解集中，可选定的自由未知量的个数都为 $n-r$。

对于非齐次线性方程组

$$AX = b$$

其中

$A=(a_{ij})m\times n$，$X=(x_1, x_2,\cdots x_n)^T$，

$b=(b_1, b_2, \cdots, b_n)^T$

其对应的齐次线性方程组 $AX = 0$ 称为它的导出组。非齐次线性方程组有无解及其解的结构情况，取决于其系数矩阵和增广矩阵的秩是否相等，以及它们秩的取值情况如下：

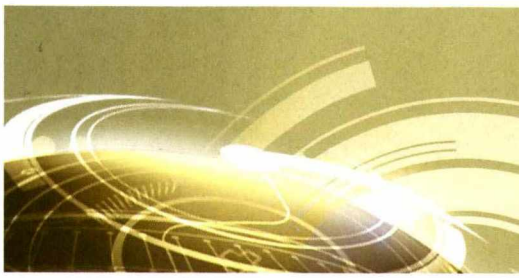
1. 当系数矩阵的秩与增广矩阵的秩相等，且它们的取值 r 都小于 n 时，非齐次线性方程有无穷解，它的每一个基础解系包含 $n-r$ 个线性无关的向量。如果把增广矩阵进行初等变换，简化为阶梯型矩阵，可以发现方程组自由未知量的个数为 $n-r$。2. 当系数矩阵的秩与增广矩阵的秩相等，且它们的取值 r 都等于 n 时，非齐次线性方程有唯一解，它没有基础解系。如果把增广矩阵进行初等变换，简化为阶梯型矩阵，可以发现方程组自由未知量的个数为 0。3. 当系数矩阵与增广矩阵的秩不相等时，方程组没有解。通过以上对齐次线性方程组和非齐次线性方程组解的结构情况进行分析，我们可以发现，无论是齐次线性方程组还是非齐次线性方程组，当它们有解时，每一个基础解系都包含 $n-r$ 个线性无关的向量（r 是它们系数矩阵或增广矩阵的秩），如果对它们的系数矩阵或增广矩阵进行初等变换，我们发现，线性方程组自由未知量的个数都是 $n-r$。

线性方程的自由未知量与统计变量的自由度

通过分析齐次线性方程组和非齐次线性方程组解的结构与它们的系数矩阵或增广矩阵的秩之间的关系，我们发现对线性方程组而言，在它们的解集中包含的自由未知量的个数与它们的系数矩阵或增广矩阵的秩有着密切关系，自由未知量的个数始终为 $n-r$。实际上，在统计研究过程中，统计变量的自由度的值，非常类似线性方程组的系数矩阵或增广矩阵转化为阶梯型矩阵后自由未知量的个数。通过分析线性方程组系数矩阵或增广矩阵的秩与确定其基础解系的时候自由未知量的取值个数之间的关系，我们能够理解不同统计方法是如何确定统计变量的自由度的。

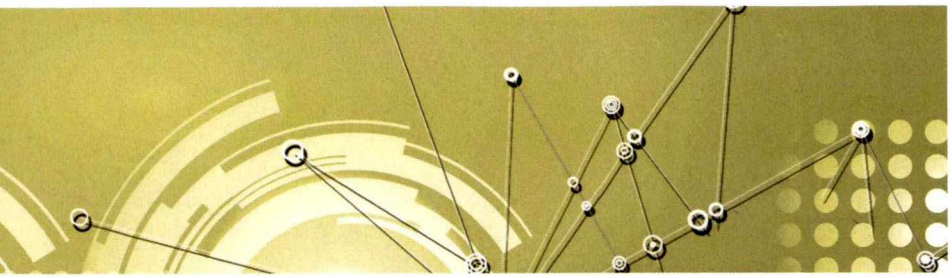
接下来，我们先对线性方程组系数矩阵或增广矩阵秩的确定方法及其数学含义作出说明。然后，在此基础上说明统计变量的自由度与线性方程组基础解系的确定过程中自由未知量的个数之间的关系。

一般而言，我们通过对线性方程组的系数矩阵或增广矩阵施以初等变换，把它们化为阶梯型矩阵，就可以求出其系数矩阵或增广矩阵的秩。对线性方程组的系数矩阵或增广矩阵进行初等变换，实际上相当于对线性方程组反复进行以下三种变换：一是交换两个方程的位置；二是用一个非零的数乘某一方程；三是把一个方程的若干倍加到另一个方程上去。这三种变换不会改变方程组的解，因此，最终的阶梯形矩阵对应的方程组与原方程组同解。当方程组有无穷多组解时，如果 $r(A)=r(Ab)=r$，则方程组中应



有 $n-r$ 个变量作为自由未知量，而其他变量可用这些自由未知量表示，从而求得方程组的一般解。另外，对线性方程组的系数矩阵或增广矩阵施以初等变换后我们可以立刻求出这两种矩阵的秩。对于这两种矩阵的秩，有着不同的理解方法。最通常的理解方法有两种：一种是把线性方程组的系数矩阵或增广矩阵看作一个由 n 个向量组成的向量组，矩阵的秩表示在这个向量组中，其最大线性无关组中包含的向量的个数；另一种理解方法是，如果系数矩阵或增广矩阵的秩为 r，那么，线性方程组系数矩阵（为方阵且阶数为 n）所有元素构成的行列式的不为零的子式的最高阶数为 r。

为了更好地理解统计变量自由度的含义，我们可以对线性方程组增广矩阵的秩作出另一种阐释：非齐次线性方程组增广矩阵的秩，实际上可以看作是方程组中 n 个未知变量之间的不同数量关系。也就是说，如果一个非齐次线性方程组增广矩阵的秩为 r，那么这个方程组中的 n 个未知变量之间必须满足 r 个不同的数量约束关系。同时，根据前面的分析我们得知，在这 r 个约束条件下，自由未知量的个数必然是 $n-r$。以此道理，我们可以说，在统计研究中，统计变量的自由度不仅取决于这些变量的取值个数 n，还取决于这些变量必须满足的约束条件的个数 r，统计变量的自由度是两者之差，



也就是 $n-r$。在下面的内容中，我们会运用这一结论，对常见统计方法中的自由度问题进行分析和说明。

| 常用统计方法中自由度的确定及其理论阐释

在具体的统计方法中，我们经常遇到自由度的问题，例如，在总体方差未知的情况下，小样本均值的 t 检验方法中，自由度的取值为 $n-1$，而在一元线性回归中，确定残差平方和（又称剩余平方和）的自由度时，自由度的取值却为 $n-2$，为什么这两种方法对自由度的确定不同呢？接下来，我们分别对这两种方法自由度确定的理论依据作出说明。

假设初婚年龄服从正态分布，我们根据 9 个人的抽样调查发现，样本均值为 23.5 岁，样本标差为 3 岁，问是否可以认为该地区的初婚年龄超过 20 岁（显著水平为 0.05）？这是一个单边检验问题。为了回答这一问题，我们必须计算统计量 t，计算出统计量后，还必须确定自由度，就此例而言自由度 $k=9-1$，然后根据自由度、相应的显著水平，查表得出临界值，再将临界值和统计量进行比较，最后做出统计判断。一般而言，在进行 t 检验时，统计变量自由度的一般都是 $n-1$。这是因为，当我们进行 t 检验时，我们假定 X_1，$X_2\cdots X_n$ 的均值 X 是既定的。既然这 n 个数的均值是既定的，所以它们的和也是既定的。也就是说

$$X_1+X_2+\cdots X_n=nX \quad (nX\text{为常数})$$

对于 n 个未知数，我们只给出了一个约束条件，相当于一个关于 X 的非齐次线性方程组满足 $r(A)=r(Ab)=1$，其自由未知量必然为 $n-1$，自由度也必然为 $n-1$。

与上述 t 检验方法不同，在利用一元线性回归分析的时候，我们把残差平方和的自由度的确定为 $n-2$，而不是 $n-1$。在一元线性回归分析中，我们把总平方和分解为回归平方和与残差平方和两部分，也即 TSS=RSS+RSSR。在对其进行统计检验的时候，我们需计算统计量 F，其具体的表达式为：$F=RSSR/[RSS/n-2]$。此时，回归平方和的自由度为 1，残差平方和的自由度为 $n-2$。

残差平方和 RSS 的自由度为什么是 $n-2$ 呢？这要从一元线性回归方程的建立说起。我们知道，一元线性回归模型的建立所依据的数学思想是最小二乘法。如果一条直线满足各观测点到它的距离的平方和最小，那么这条直线就是所要求的最佳预测直线。设这条直线为 $y=a+bx$，那么，任一观测点 (x_i, y_i) 到直线的铅直距离为：$y_i-(a+bx_i)$，那么所有各点到该直线的铅直距离的平方和为

$$Q(a,b)=\sum[y_i-(a+bx_i)]^2 \quad (i=1,2,\cdots n)$$

显然 Q 是 a,b 的函数，我们要做的就是求出 a,b 的值，使 Q 达到最小值。对 Q 的表达式分别就 a,b 两个量求导，并令它们的偏导数分别为零，可得下面的方程组：

$$\sum[y_i-(a+bx_i)]=0$$
$$\sum[y_i-(a+bx_i)]x_i=0$$

其中 $i=1,2,\cdots n$。

对上面的方程组继续整理，可得

$$\sum y_i=na+b\sum x_i$$
$$\sum x_iy_i=a\sum x_i+b\sum x_i^2$$

此时，我们可以把 a,b,x_i 都看作常数，这就相当于一个关于 y 的非其次线性方程组满足 $r(A)=r(Ab)=2$，其自由未知量的个数必然为 $n-2$。也相当于我们对 n 个变量 y_1，y_2，$\cdots y_n$ 只给定了 2 个不同的约束条件，变量的自由度必然为 $n-2$。

作者单位：安阳师范学院

参考文献

[1] 徐晓岭．统计分布间的关系探讨[J]．统计与决策，2008，(17)：20-22．

[2] (美) 罗斯．概率论基础教程[M]．郑忠国，詹从赞译．北京：人民邮电出版社，2010．

[3] (美) Freedman，D等．统计学[M]．魏宗舒等译．北京：中国统计出版社，1997．

[4] (美) 费勒．概率论及其应用[M]．胡迪鹤译．北京：人民邮电出版社，2006．