

装

订

线



数学模型论文

题目**： 血样的分组检验**

第 组

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 学号 | 姓名 | 班级 | 分工 | 贡献率 |
| 20161101079 | 李策 | 16计科2 | 分析数据求解 | 34% |
| 20161101058 | 赵志淳 | 16计科2 | 总结 | 33% |
| 20161101240 | 张悄 | 16信安2 | 建立模型 | 33% |

网络空间安全与计算机学院

2018年6月10日

# 血样的分组检验

# 摘　　要

# 本文主要为了解决减少血样检验次数这个实际问题 ,为了在人群中 (数量很大, 基本上是健 康人)找出某种病毒的感染者,为减少检验次数(目的是降低费用) ,通常采用筛选的办法：即假设 人群总数为N, 将人群分成M组,将每组的血样混在一起进行化验, 若化验结果呈阳性,则需要对该组的每个人重新进行化验 , 以确定谁是病毒感染者;若化验结果呈阴性, 则表明该组全体成员均为阴性,不需要重新化验。采用概率论中的方法将分组的个数与阳性样本的合数做比较，分情况讨论具体问题，最后运用概率论的知识得出结果，再用较多数据检测结果的正确性。

关键词：古典概型；全概率公式；分组讨论

1　　问题提出

　　在一个很大的人群中通过血样检验普查某种疾病，假定血样为阳性的概率为P，（通常P很小）为减少检验次数，将人群分组，一组人的血样混合在一起化验。当某组的混合血样呈阴性时，即可不经检验就判定该组每个人的血样都是阴性；而当某组的混合血样样本为阳性时，则可判定该组至少有一人血样为阳性，于是需要进一步对这组的每个人都再做检验。

问题：当P固定时（0.01%、0.1%），如何分组，即多少人一组，可使平均检验次数最少，与不分组的情况作比较。

2　　模型假设

2.1　结合本问题的实际，对此模型做出假设如下

1. 为确保人是一个个体，整数部分不同的，结果中的小数部分都舍弃；整数部分相同的，保留相同的小数位后一位再比较大小
2. 每个人的血样在没有检测之前都是相同，分组也不受到外界条件干扰
3. 每个组的人数都相同且均为整数

2.2　模型中的未知变量假设如下

　　N 总检验人数

　R 阳性比率

　P 每组人数

　X 阳性人数

　M分组个数

　Y 假设的组数

　S1 X个人分到Y个组平均组数

　S2 X个人分到Y个组平均组数

3　　问题分析

建设总检验人数N为固定一稍大的数据，由阳性比率R可以确定阳性组人数：X，对所有人进行分组之后，可以确定有多少组M和每组多少人P，我们用概率学的算法算出X个人分到M组中每一组的概率之后，进而求其数学期望，求出平均有多少组分到了人，最后再乘每组人数可算出平均要检测多少个人，再和不分组的情况做比较。

当阳性人数R确定之后，我们根据组数M和R的关系将模型分为两类：M>=R和M<R。我们认为，当M>=R时，这种情况无需考虑分组，因为最坏情况即每组中都有人的情况是由M决定的，而M是一定值，因此分组个数越小越好。我们将用我们问题一的模型验证这一猜测。**①**

4　　模型的建立与求解

4.1　问题模型建立

因为R是一个很小的数字，我们假设总人数N=100000设计模型。R的数值我们取两个：0.01%、0.1%考虑。

则阳性人数为

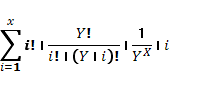


分组的个数M根绝具体情况设定，如果趋势不明显，则进一步放大或 缩小。

基于问题二的猜想，我们把X个人分到Y个组中的概率分成两种情况：

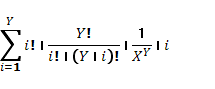
X<=M

S1=



X>M

S2=



4.2　问题求解

根据S1、S2的算式，最终的求得的平均检测人数为

X<=M

Min{}

将N=100000，R=0.01%带入得M=10，取M的范围为[10，50]，使用工具软件Matlab输入上述数据和公式计算得下表：

表1　　R=0.01%时的数据表



使用Matlab自带的绘图软件把上述结果绘成散点图：

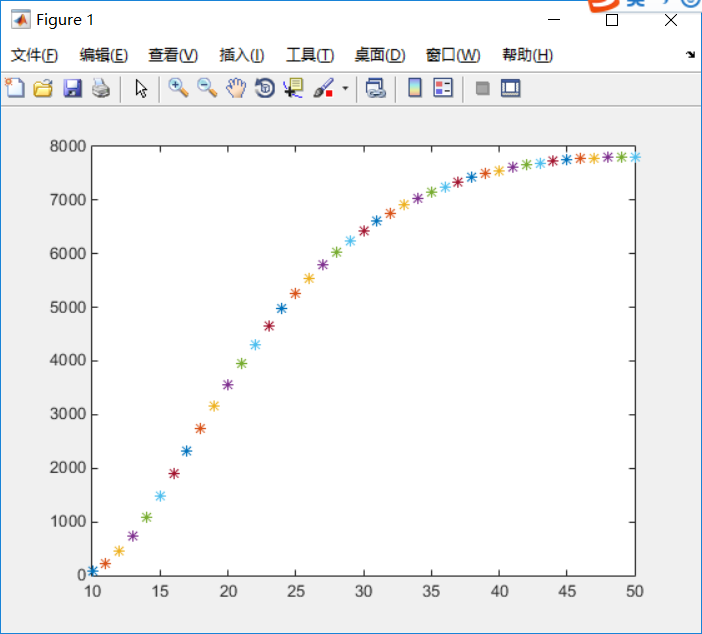


图1　　R=0.01%时的散点图

取N=20000 R=0.1%带入得M=20，取M的范围为[20,50]，使用工具软件Matlab输入上述数据和公式计算得下表：

表2　　R=0.1%时的数据表



　　使用Matlab自带的绘图软件把上述结果绘成散点图：

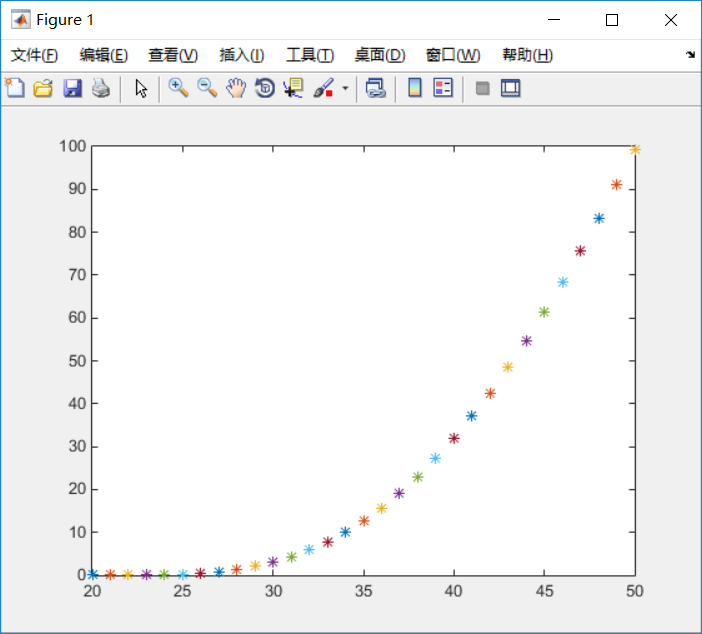


图2　　R=0.1%时的散点图

根据上两种条件，我们得出总结：在此种情况下，在我们的模型中检验个数是递增的，因此当X=M时，取得结果最小。

X>=M

Min{}

将N=100000，R=0.01%带入得M=10，取M的范围为[2，9]，使用工具软件Matlab输入上述数据和公式计算得下表：

表3　　R=0.01%时的数据表2



使用Matlab自带的绘图软件把上述结果绘成散点图：

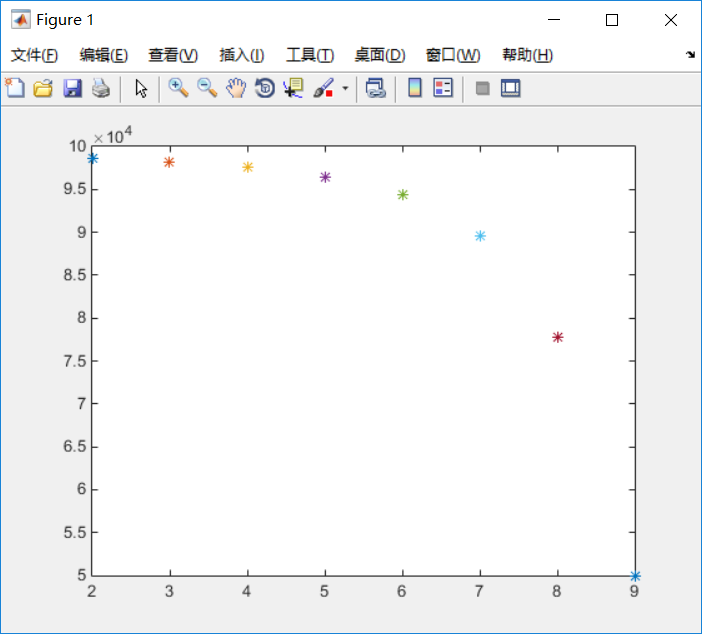


图3　　R=0.01%时的散点图2

取N=20000 R=0.1%带入得X=20，取M的范围为[2，20]，使用工具软件Matlab输入上述数据和公式计算得下表：

表4　　R=0.1%时的数据表2



　　使用Matlab自带的绘图软件把上述结果绘成散点图：

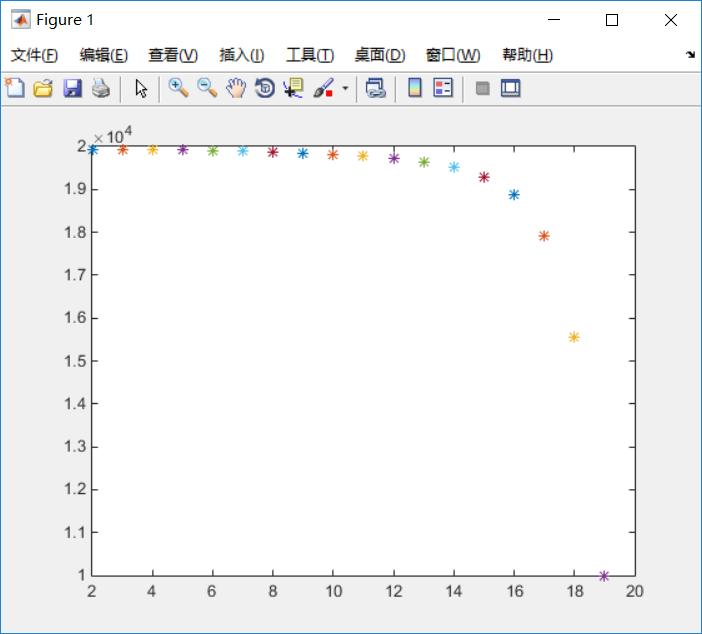


图4　　R=0.1%时的散点图2

在不分组的情况下，两种情况的检验个数均为样本个数：N

5　结果分析与总结

在我们的模型中，我们把问题的求解分成了两部分情况：

在第一种情况下，即X<=M时，需要检验的人数同分组的个数成正比，根据概率学知识可知，当分组个数多于人数时，不是每个分组内都可能有阳性血样样本，因此阳性血样样本分布的广泛性是造成检验次数多的原因。而在我们的数据中，当阳性血样样本个数等于分数个数时，平均需要检测的个数最少，但是我们后来经过大量计算，发现，此种情况发生的概率十分小，而我们计算的过程中因为太过理想化的假设而放大了此概率，导致了计算结果部分失真。

在第二种情况下，即X>M时，需要检验的人数同分组个数成反比，我们推测此种情况主要是因为大量的人数造成了每个分组当中有阳性血样样本的概率大大增加，且因为分组个数过少，每个分组中含有大量的样本，造成了平均需要检测的样本个数偏多。随着分组个数的增多，每个分组中是否含有阳性血样样本的概率和分组样本数量均在下降，因此，总的平均检验个数均在下降。

综合两种情况和我们采用的R的大小来看，对于任意的R，随着分组个数的逐渐增加，平均需要检测的样本数目先降低、后增加，在R\*N的数值附近时将达到最小，且两组当中最坏的结果均要好于M=1，即不分组的情况。

但是，我们计算是全部基于概率学讨论，而且在第一种情况下计算存在失真②，因此计算的结果只能作为参考，不能大范围的推广，还需加以验证和修改。

参考文献

[1]　　盛骤.概率论与数理统计.北京.高等教育出版社，2008：P16~19

[2]　　姜启源.数学模型.北京.高等教育出版社，2011：P300~302

# 附　　录

所有实验过程用的Matlab程序源码：

function a=getequal(n,m)

syms q;

a=symsum((factorial(n)/factorial(n-q))\*q/n.^m,1,m);

function a=test(n,m)

syms q;

a=symsum((factorial(q)\*(q-1))/(factorial(m)),1,m);

>> for i=20:50

a=getequal(i,20)\*20000/i;

plot(i,a,'\*');

hold on;

end

>> for i=20:50

a=getequal(i,20)\*20000/i;

vpa(a,4)

end

>> for i=10:50

a=getequal(i,10)\*20000/i;

plot(i,a,'\*');

hold on;

end

>> for i=10:50

a=getequal(i,10)\*20000/i;

vpa(a,4)

end

>> for i=2:19

a=test(20,20-i+1)\*20000/(20-i+1);

plot(i,a,'\*');

hold on;

end

>> for i=2:19

a=test(20,20-i+1)\*20000/(20-i+1);

vpa(a,4)

end

>>

for i=2:9

a=test(10,10-i+1)\*100000/(10-i+1);

plot(i,a,'\*');

hold on;

end

>>

for i=2:9

a=test(10,10-i+1)\*100000/(10-i+1);

vpa(a,4)

end