

# 统计推断第一次随机模拟实验报告

## 【实验目的】

针对不同的总体（正态、指数、cauchy、二项、poisson）来通过 R 实验了解样本均值与样本方差之间的关系（是否相关或者独立）。

## 【实验设计】

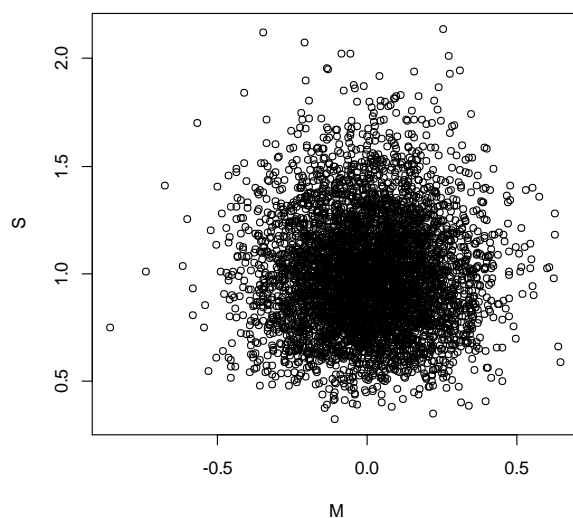
以下均通过 R 语言实现：

- （1）确定实验的总体，对不同的总体的参数（如正态分布的均值和方差，二项分布的概率等），生成样本数（n）分别为 30,100,1000 的随机变量样本集合 X。
- （2）对每一种样本数，利用 R 语言循环语句，重复生成 5000 次样本集合 X。
- （3）对每一种样本数（30,100,1000），计算样本集合 X 的均值（mean）和方差（var）。
- （4）作图，将每种样本数下算出的均值和方差作为二维向量画在图上。
- （5）观察均值与方差的相关性，得出结论。

## 【模拟实验结果】

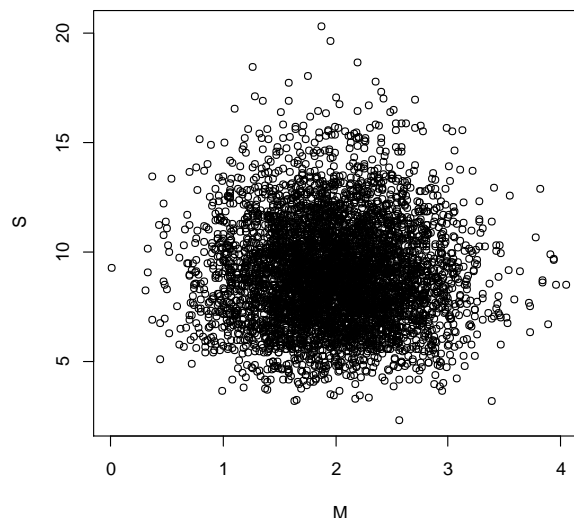
### 1、正态随机变量

标准正态随机变量，样本总量30



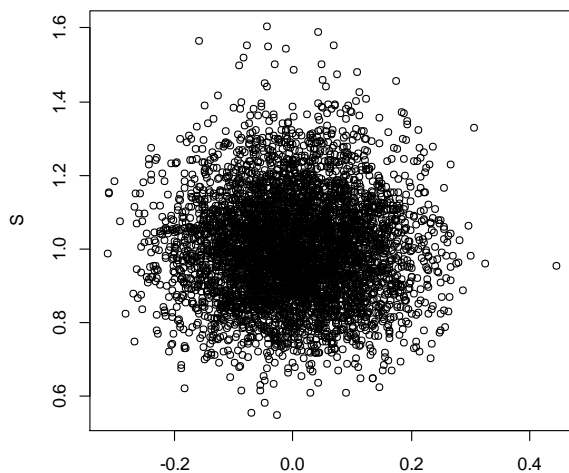
Pearson 相关系数为-0.01  
样本均值标准差为 0.18  
样本方差标准差为 0.26

期望2标准差3的正态随机变量，样本总量30



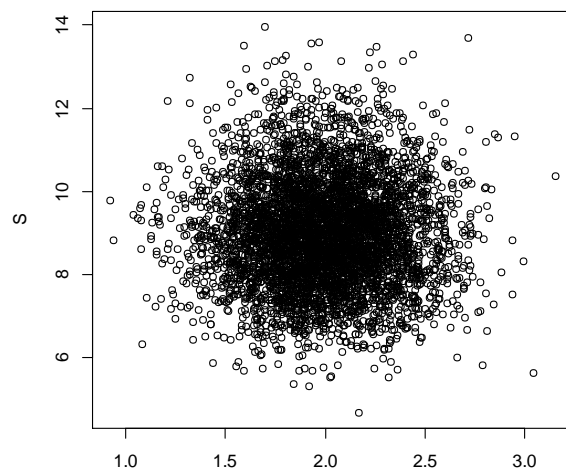
Pearson 相关系数为-0.02  
样本均值标准差为 0.55  
样本方差标准差为 2.34

标准正态随机变量，样本总量100



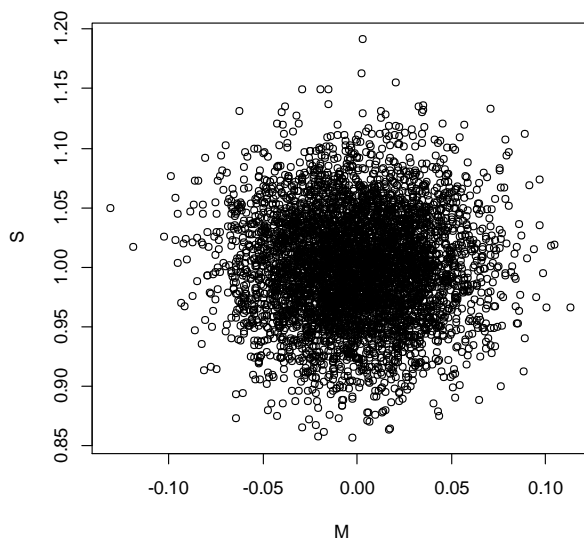
Pearson 相关系数为 0.00076  
样本均值标准差为 0.10  
样本方差标准差为 0.14

期望2标准差3的正态随机变量，样本总量100



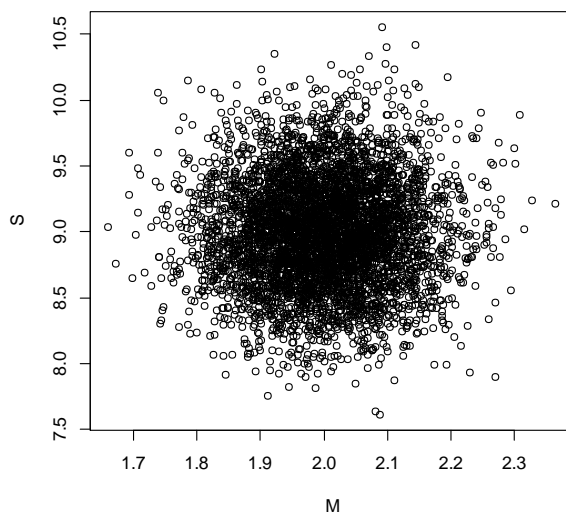
Pearson 相关系数为 -0.0014  
样本均值标准差为 0.30  
样本方差标准差为 1.28

标准正态随机变量，样本总量1000



Pearson 相关系数为 0.0041  
样本均值标准差为 0.032  
样本方差标准差为 0.045

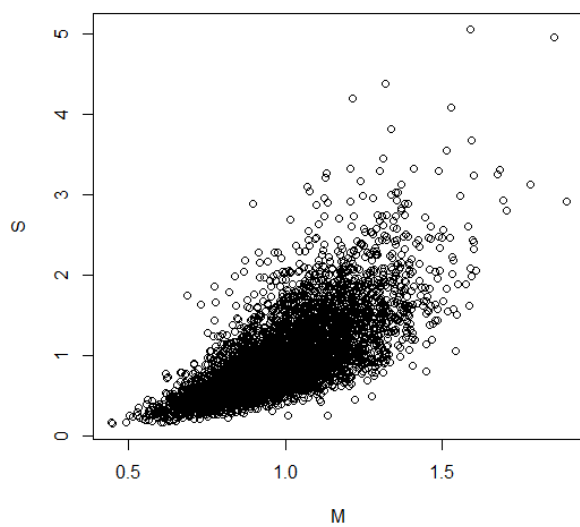
期望2标准差3的正态随机变量，样本总量1000



Pearson 相关系数为 0.013  
样本均值标准差为 0.096  
样本方差标准差为 0.41

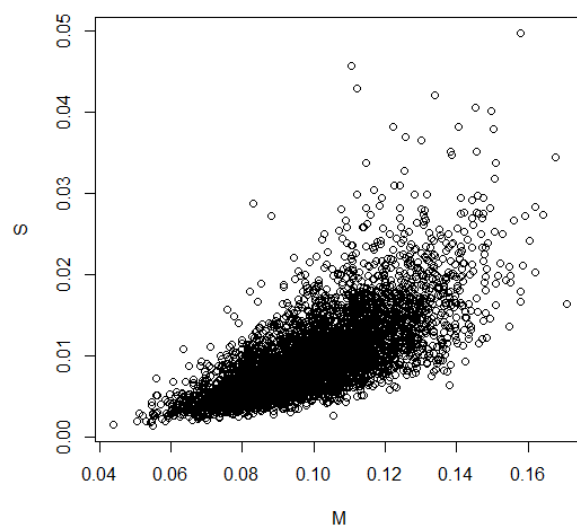
## 2、指数分布随机变量

参数为1的指数分布随机变量，样本总量30



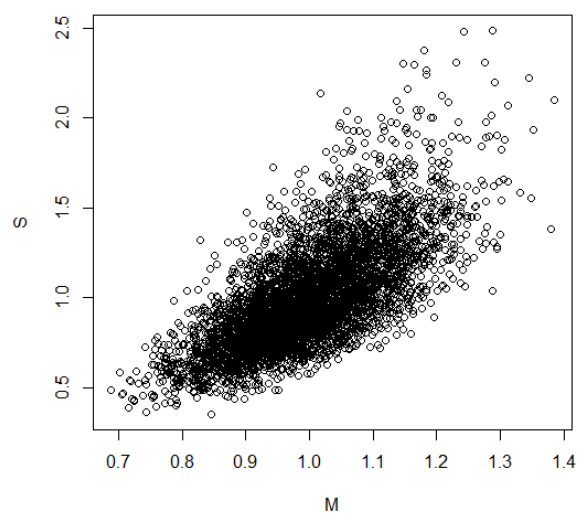
Pearson 相关系数为 0.71  
样本均值标准差为 0.18  
样本方差标准差为 0.51

参数为10的指数分布随机变量，样本总量30



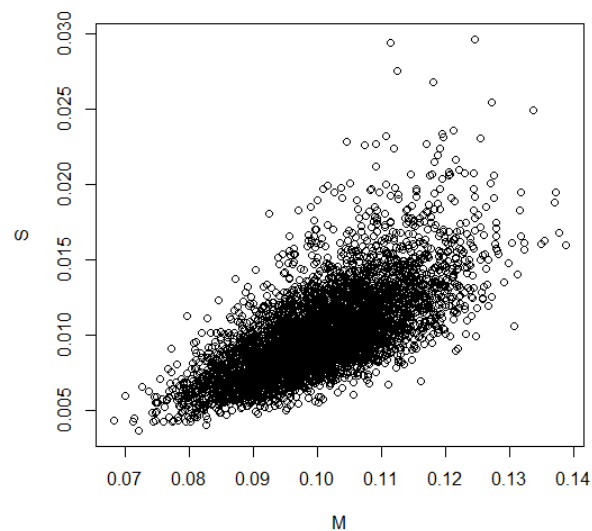
Pearson 相关系数为 0.71  
样本均值标准差为 0.018  
样本方差标准差为 0.0051

参数为1的指数分布随机变量，样本总量100



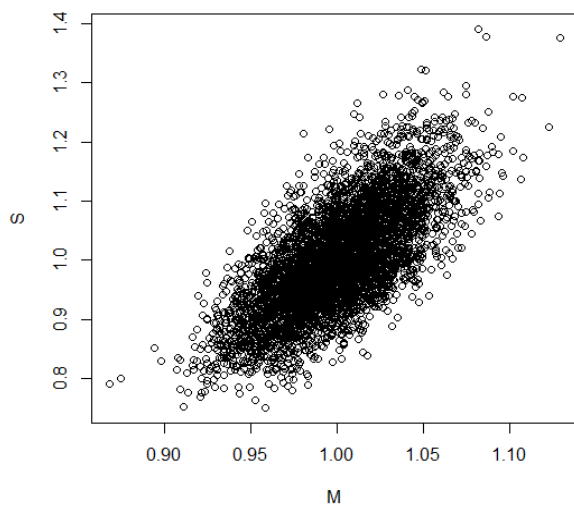
Pearson 相关系数为 0.71  
样本均值标准差为 0.10  
样本方差标准差为 0.28

参数为10的指数分布随机变量，样本总量100



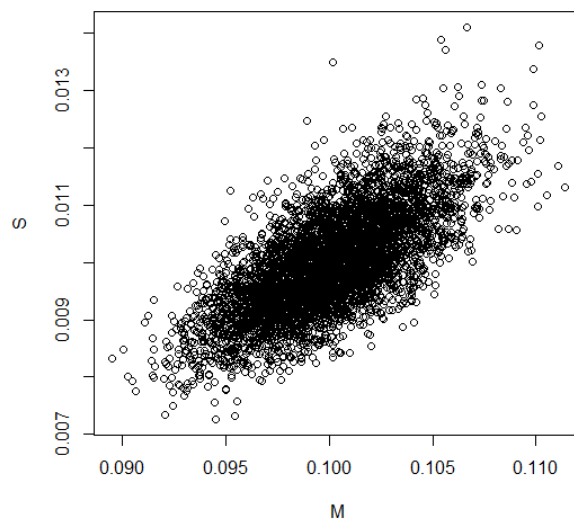
Pearson 相关系数为 0.70  
样本均值标准差为 0.010  
样本方差标准差为 0.0029

参数为1的指数分布随机变量，样本总量1000



Pearson 相关系数为 0.71  
样本均值标准差为 0.031  
样本方差标准差为 0.088

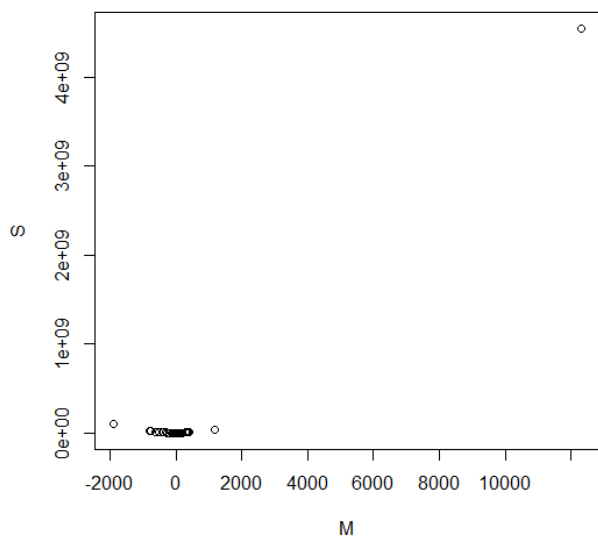
参数为10的指数分布随机变量，样本总量1000



Pearson 相关系数为 0.71  
样本均值标准差为 0.0031  
样本方差标准差为 0.00089

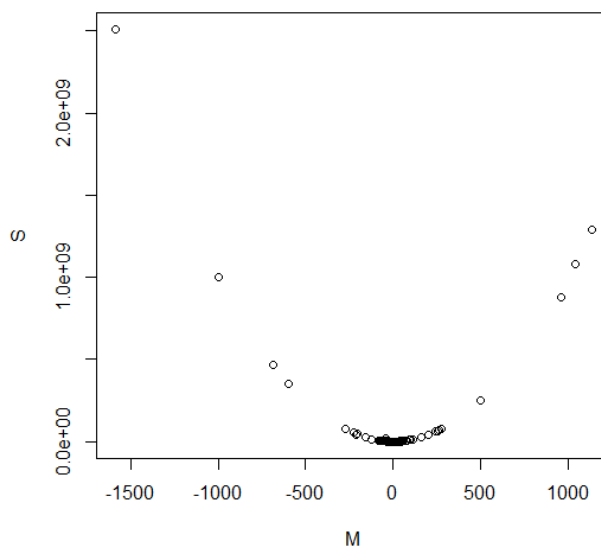
### 3、Cauchy 分布随机变量：

标准cauchy分布随机变量，样本总量30



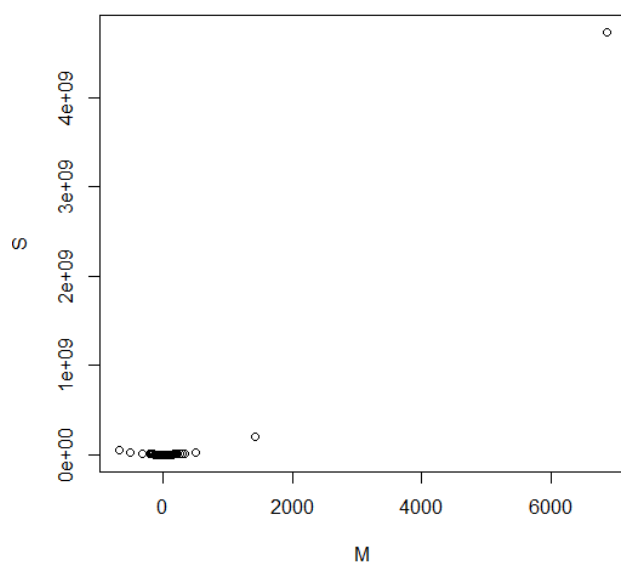
Pearson 相关系数为 0.97  
样本均值标准差为 179.04  
样本方差标准差为 64310558

标准cauchy分布随机变量，样本总量1000



Pearson 相关系数为-0.19  
样本均值标准差为 41.97  
样本方差标准差为 47615710

标准cauchy分布随机变量，样本总量100



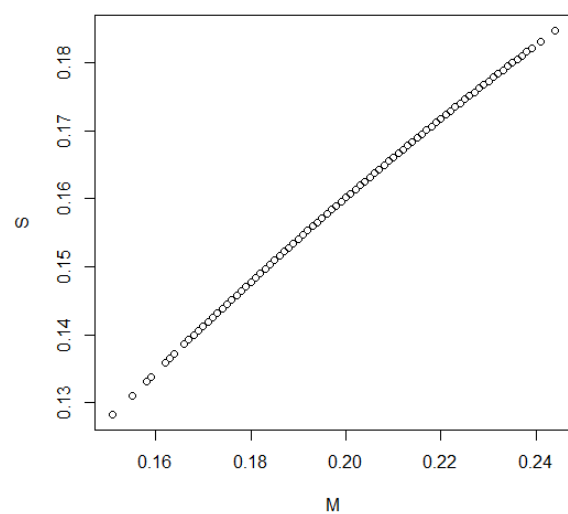
Pearson 相关系数为 0.97

样本均值标准差为 101.43

样本方差标准差为 66947742

4、二项分布随机变量（这里可变参数很多，只取样本总量 1000 的情况，其他情况结果类似）

$p=0.2$ 伯努利分布随机变量，样本总量1000

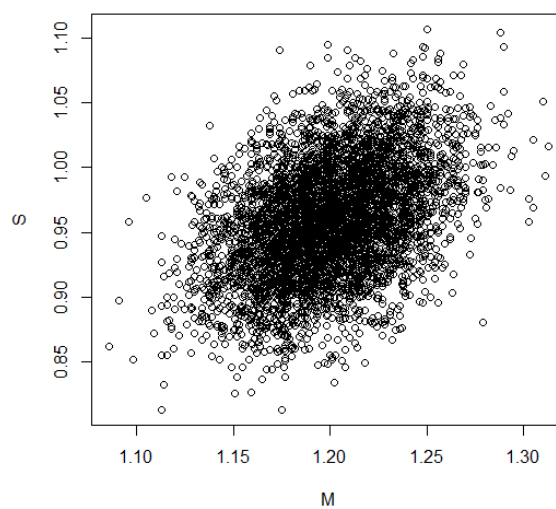


Pearson 相关系数为 1.0

样本均值标准差为 0.013

样本方差标准差为 0.0076

$p=0.2, n=6$ 的二项分布随机变量，样本总量1000

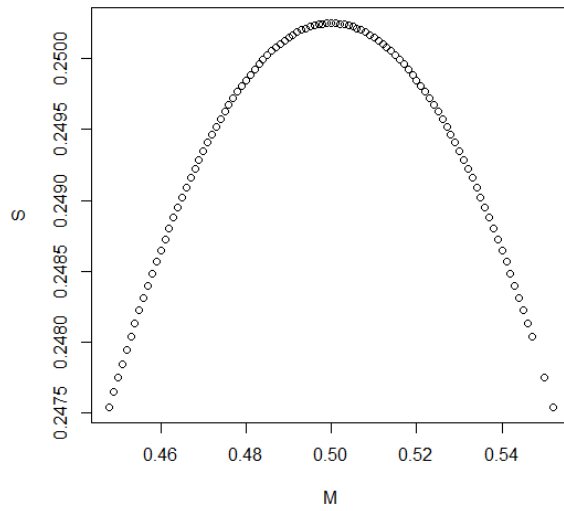


Pearson 相关系数为 0.42

样本均值标准差为 0.031

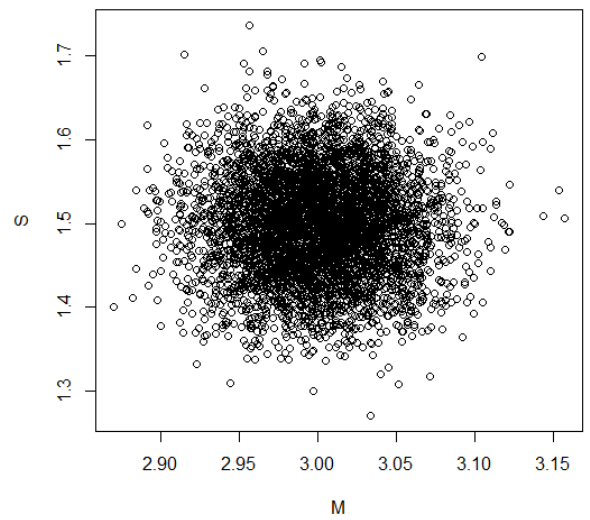
样本方差标准差为 0.043

**p=0.5伯努利分布随机变量，样本总量1000**



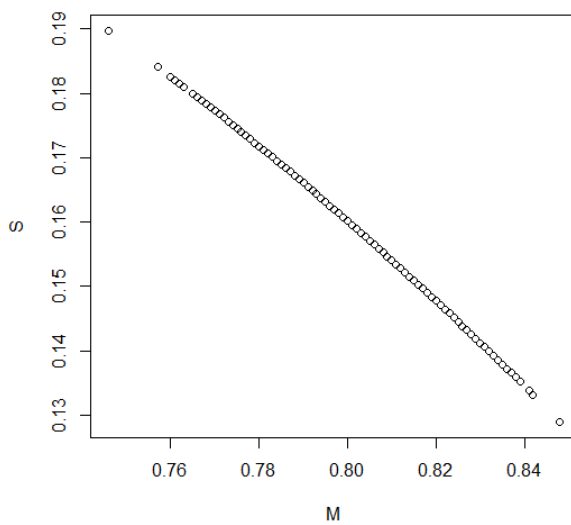
Pearson 相关系数为 0.063  
样本均值标准差为 0.016  
样本方差标准差为 0.00036

**p=0.5,n=6的二项分布随机变量，样本总量1000**



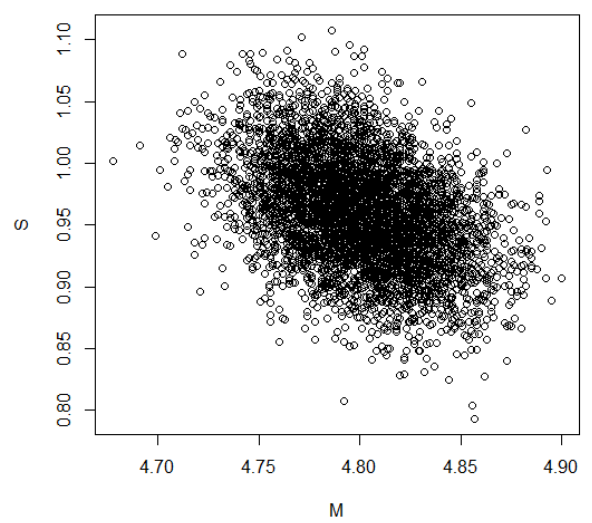
Pearson 相关系数为-0.0043  
样本均值标准差为 0.039  
样本方差标准差为 0.062

**p=0.8伯努利分布随机变量，样本总量1000**



Pearson 相关系数为-1.0  
样本均值标准差为 0.013  
样本方差标准差为 0.0076

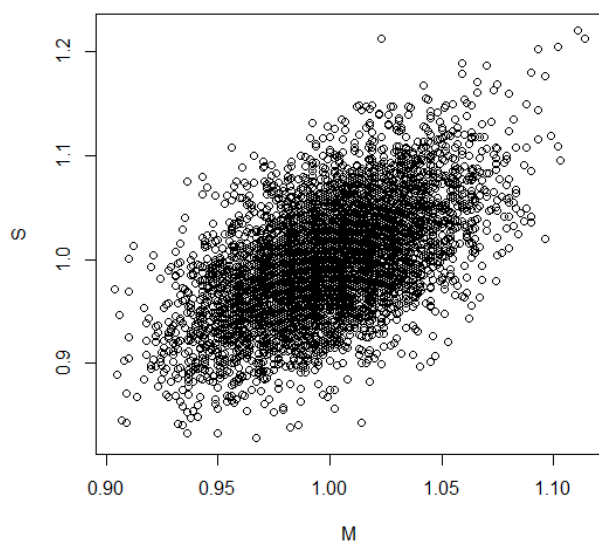
**p=0.8,n=6的二项分布随机变量，样本总量1000**



Pearson 相关系数为-0.42  
样本均值标准差为 0.031  
样本方差标准差为 0.043

## 5、poisson 随机变量

参数为1的poisson分布随机变量，样本总量1000

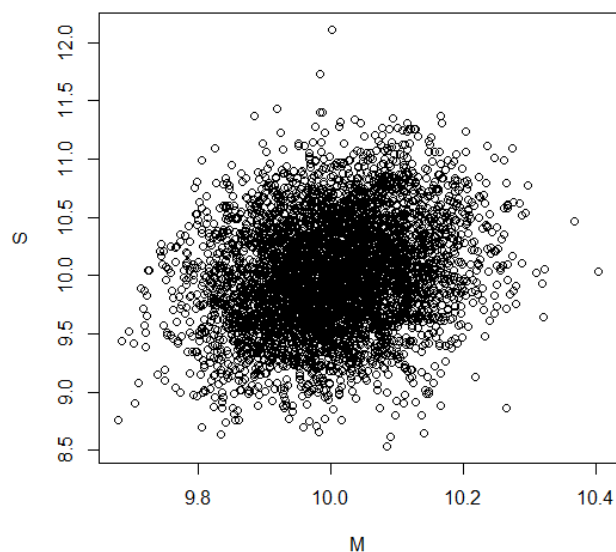


Pearson 相关系数为 0.60

样本均值标准差为 0.032

样本方差标准差为 0.055

参数为10的poisson分布随机变量，样本总量1000

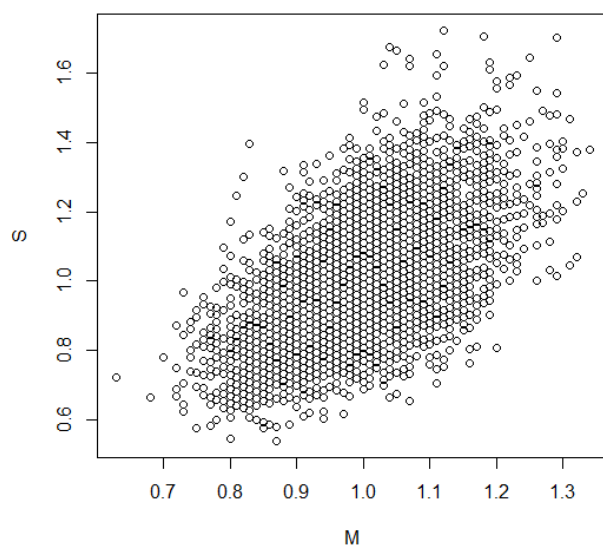


Pearson 相关系数为 0.22

样本均值标准差为 0.10

样本方差标准差为 0.46

参数为1的poisson分布随机变量，样本总量100

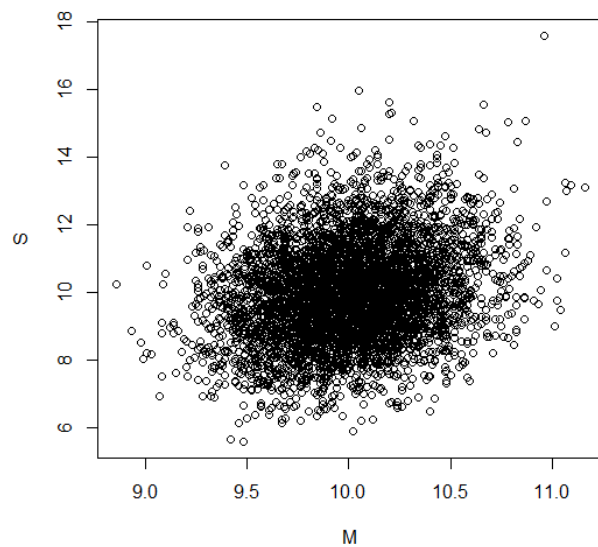


Pearson 相关系数为 0.57

样本均值标准差为 0.10

样本方差标准差为 0.17

参数为10的poisson分布随机变量，样本总量100

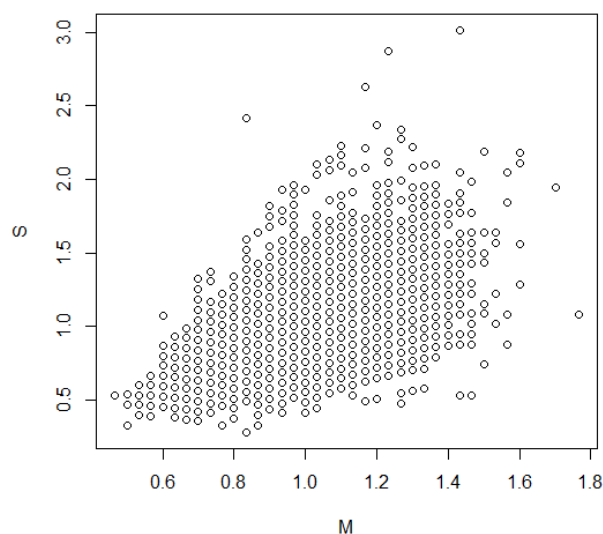


Pearson 相关系数为 0.24

样本均值标准差为 0.32

样本方差标准差为 1.46

参数为1的poisson分布随机变量，样本总量30

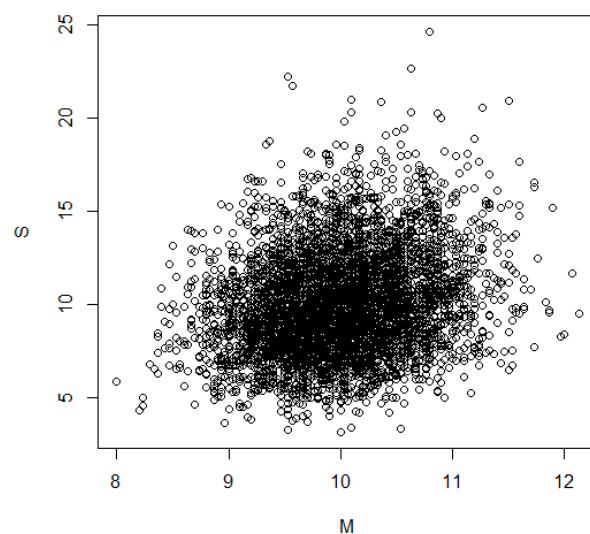


Pearson 相关系数为 0.57

样本均值标准差为 0.18

样本方差标准差为 0.32

参数为10的poisson分布随机变量，样本总量30



Pearson 相关系数为 0.22

样本均值标准差为 0.58

样本方差标准差为 2.71

### 【结果分析与结论】

- (1) 正态分布：对每种样本总量和两种不同的参数，由 pearson 相关系数来看，绝对值普遍十分接近 0，可以推断正态分布随机变量样本均值与方差不存在线性相关性。由图像可以看出，样本均值与方差是独立且不相关的。
- (2) 指数分布：对每种样本总量和两种不同的参数，由 pearson 相关系数来看，均在 0.71 附近，可以推断指数分布随机变量样本均值和方差具有较明显的正相关性。由图像可以看出，样本均值与方差不独立且为比较明显正相关。
- (3) Cauchy 分布：样本量为 30 和 100 的模拟实验得到的 pearson 系数接近 1，样本均值和样本方差明显的正相关。但样本量为 1000 的实验得到的 pearson 系数接近 0，样本均值和方差明显不相关。从图上看，坐标的分布点实则近似一个抛物线，方差极大，所以这样得到的相关性是不可取的。事实上，cauchy 分布随机变量不存在期望，因此样本均值在  $n$  很大时意义不大。  
由于单个随机变量无期望，所以样本均值没有意义，没有讨论。
- (4) 二项分布：根据 pearson 相关系数，样本均值与方差对  $n=1$  的二项分布来说， $p=0.2$  有完全正相关性， $p=0.5$  几乎不相关（这里体现了 pearson 系数的局限性）， $p=0.8$  有完全的负相关性。对  $n=6$  来说， $p=0.2$  有微弱正相关性， $p=0.5$  几乎无相关性， $p=0.8$  有微弱负相关性，但是我们从图上来看， $n=1$ ， $p=0.5$  时，样本均值和方差有明显的抛物线关系，而此时 pearson 相关系数却很小，这表明 pearson 相关系数只能用来检验线性相关，也是改系数的局限性。
- (5) Poisson 分布：样本均值与样本方差对参数在 1 时，根据 pearson 系数，有极其微弱的正相关性（0.6 左右），但参数为 10 时几乎就不存在相关性了。但由于 pearson 系数仅在固定正常数（0.22）附近徘徊，可以断定样本均值和方差并不独立，也几乎不相关。

### 【展望】

这一次模拟实验给我影响最深的就是 pearson 相关系数的局限性，它只能反映线性相关性，具体体现在伯努利分布  $p=0.5$  时的情况。我认为下一步在我水平范围内能研究的问题可以是：



了解其他相关系数以及其分析方式，对类似伯努利分布这种不适用 **pearson** 系数的问题进行解释。