

```
• begin
•     using StatsKit    ,StatsPlots    ,Distributions    ,Images    ,FileIO    , ImageIO
• end
```

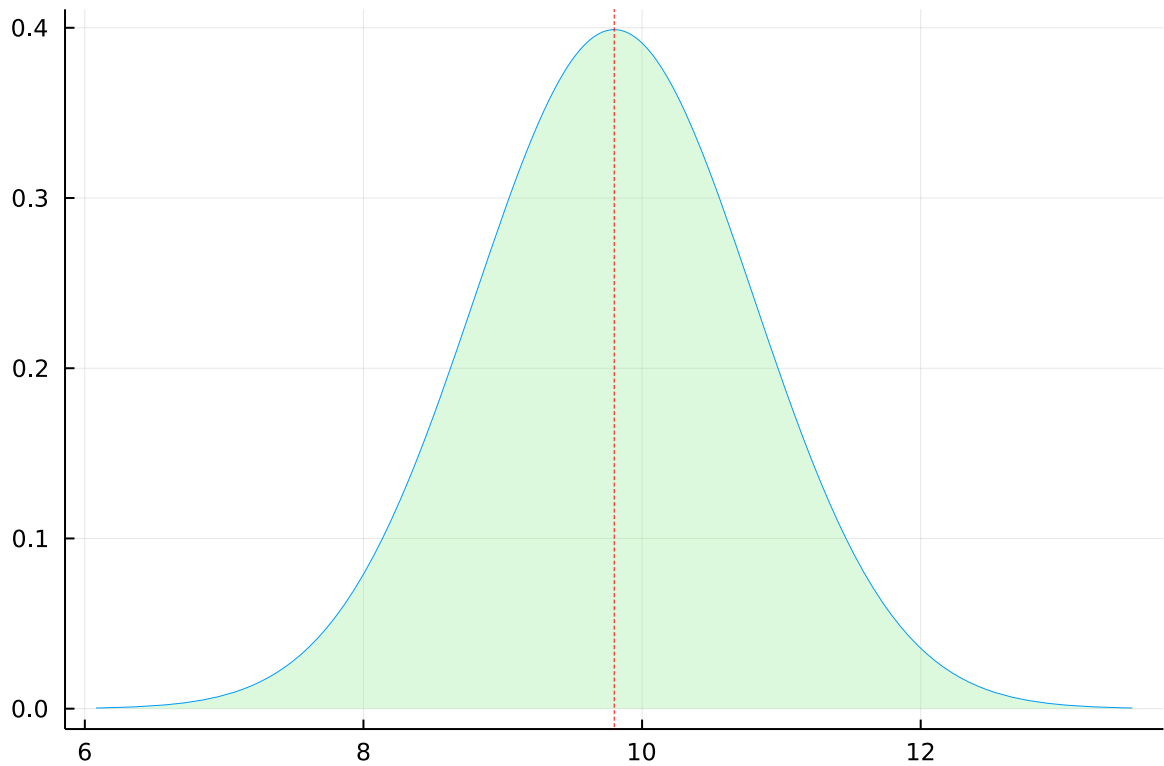
100 米短跑里的数据

一般来说跑道上会有八个人, 如果让着八个人反复的跑, 每次的成绩和排名都是不同. 每次跑步成绩取得均值, 尽管每次成绩都会有一点点差异, 但是基本不会有太大的波动, 这个可以用标准正态分布作为理论模型. 引入理论模型目的不在于被理论框架束缚, 只是提供一个大概的思考方向. 这就是我们使用统计模型的图纸.

假设最优秀的百米选手的成绩为 **9.80** 秒, 标准差为 **1**, 首先可以依据正态分布绘制比赛成绩的理论分布.

比赛成绩与运动员状态, 外界因素都有关, 这些随机因素有时会让成绩高于真实水平, 有时会让运动员成绩变差.

这就在随机实验中引入正态分布的意义, 与这个情况类似的事件都可以正态分布作为模型, 根据数据做一下修剪.



```
• begin
•   μ,σ=9.8,1
•   span=8:0.01:12
•   d1=Normal(μ,σ)
•   plot(d1,label=false,fill =(0,:lightgreen),fa=0.3,lw=0.3)
•   vline!([μ],ls=:dash, lw=0.5,lc=:red,label=false)
• end
```

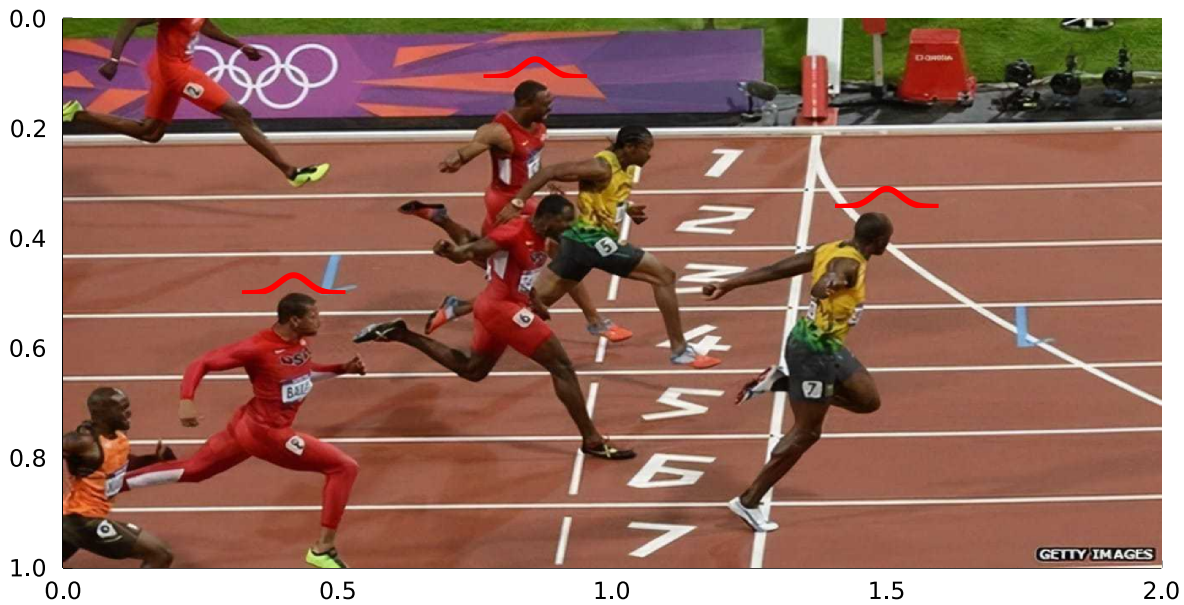
成绩的多元性

实际对于一场比赛的成绩向量,或者均值是一个多元统计结果.

每个参赛的选手自身的成绩都是从一个正态分布中抽样出来的结果.

体育比赛结果的不确定性就是由每个选手的成绩的抽样概率决定的.

参见下面这张图. 参考了游戏里的参考值, 每个运动员的成绩实际就是一个正态分布



```

• begin
•   url="https://tva1.sinaimg.cn/large/e6c9d24egy1h5z83kbnlhj21040kataf.jpg"
•   img_file = download(url)
•   img=load(img_file)
•   x = range(0, 2, length=size(img,1))
•   y = range(0, 1, length=size(img,2))
•   plot(x,y,img)
•   n1=Normal(1,2)
•   BB1 = bbox(0.7,0.23, 0.1,0.15)
•   BB2 = bbox(0.16,0.35, 0.1,0.15)
•   BB3 = bbox(0.38,0.05, 0.1,0.15)
•   plot!(n1,inset = (1, BB1), subplot = 2,axis=false,label=false,ylims=
(0,1.2),color=:red,frame=:none,background=false,lw=3)
•   plot!(n1,inset = (1, BB2), subplot = 3,axis=false,label=false,ylims=
(0,1.2),color=:red,frame=:none,background=false,lw=3)
•   plot!(n1,inset = (1, BB3), subplot = 4,axis=false,label=false,ylims=
(0,1.2),color=:red,frame=:none,background=false,lw=3)
• end

```