

# 统计方法的几个应用

- 普查 (Census)
- 调查 (Survey)
  - 样本
  - 群体

# 样本选取

- H1 : 平均体重为60Kg
- 检验



统计的基本功能是基于**样本**的结果，给出**群体**的结论

# 基本假设

- 大数定理
- 小数定理
- 概率与直觉

# 二项分布

$$b(x, n, p) = C_n^x p^x q^{n-x}$$

其中b表示二项分布的概率，n表示试验次数，x表示出现某个结果的次数。

生产商通常都会跟零件供应商约定供货合格率，并对每批供货进行抽检。设约定的合格品率为97%，如果每批随机抽10件，那么抽出1件不合格时，整批的零件的合格率是不是达不到97%？

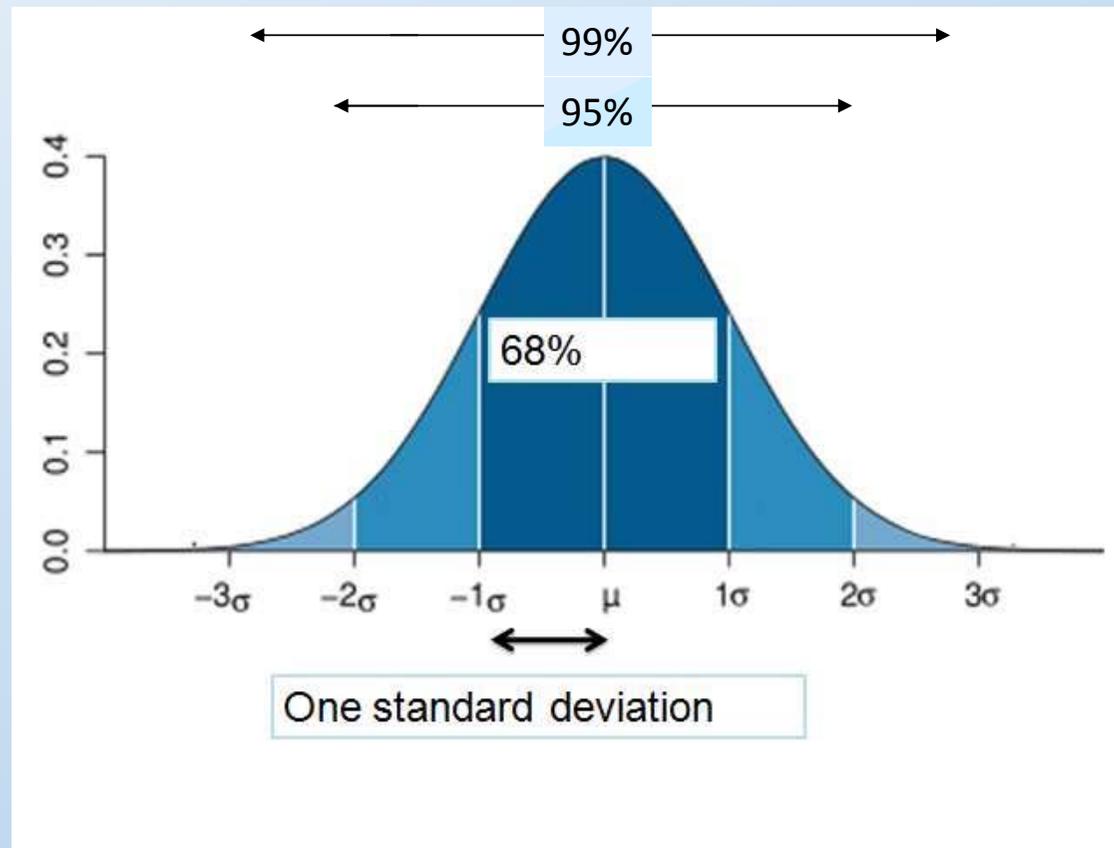
$$b(9, 10, 0.97) = C_{10}^9 \times 0.97^9 \times 0.03^1 = 0.228$$

一个不合格

$$b(2, 10, 0.03) = C_{10}^2 \times 0.03^2 \times 0.97^8 = 0.032$$

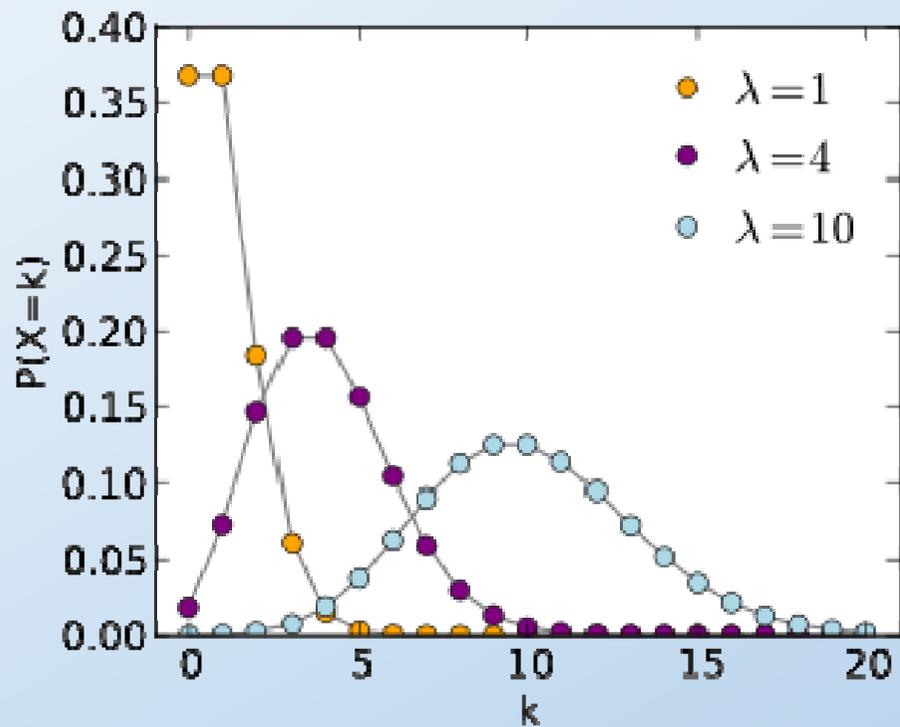
两个不合格

# 高斯分布( $\sigma$ )



$$\frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$$

# Poisson distribution , 泊松分布



$$p(x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

$X=k$	碰上百年一遇
0	0.368
1	0.368
2	0.184
3	0.061
4	0.015
5	0.003
6	0.0005

# German Tank problem



二战期间，德国坦克很有威胁；

如何估算德国坦克的产量？

序列号

- 可以使用的方法：
- 均值乘一个系数，通常是2
- 标准差乘一个系数，通常是4或者6
- 最大值加最小值
- 最大值加标准差
- 中位值乘一个系数

# minimum-variance mean-unbiased estimator

Month	统计估计	情报部门估计	真实值
June 1940	169	1,000	122
June 1941	244	1,550	271
August 1942	327	1,550	342

# 赌博问题

- 囚徒困境 ( tit for tat )
- 赔率 - 马尔可夫过程
- 马丁格尔 ( Martingale ) 策略
- 反马丁格尔 ( Martingale ) 策略

# 作业

- 你有一个硬币；
- 用它产生一系列均值为零，标准差为1的高斯分布

# 内容提要

- SPSS 简介
- SPSS 的基本结构
- 数据准备
- 数据描述
- 图表
- 数据转换
- 其它

# SPSS 简介



- SPSS: Statistical Package for the Social Sciences
- SPSS 在类似的软件中如 S-Plus, R, SAS, 属于比较易学和易操作的.
- 但是同样也可以编程

