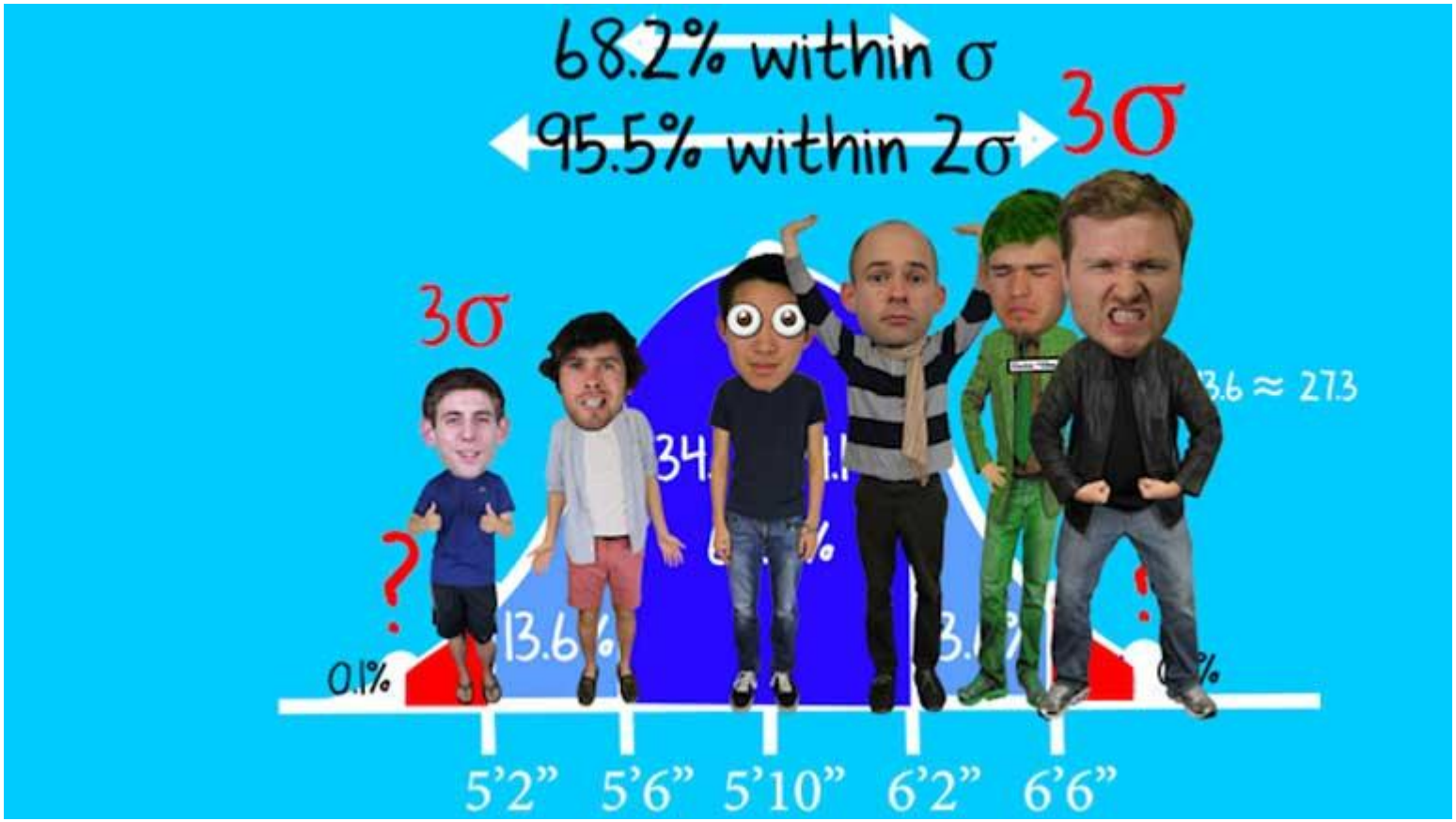


Excel数据分析师突击——从入门到精通到项目实战
第5周

法律声明

【声明】本视频和幻灯片为炼数成金网络课程的教学资料，所有资料只能在课程内使用，不得在课程以外范围散播，违者将可能被追究法律和经济责任。

课程详情访问炼数成金培训网站

http://edu.dataguru.cn

关注炼数成金企业微信

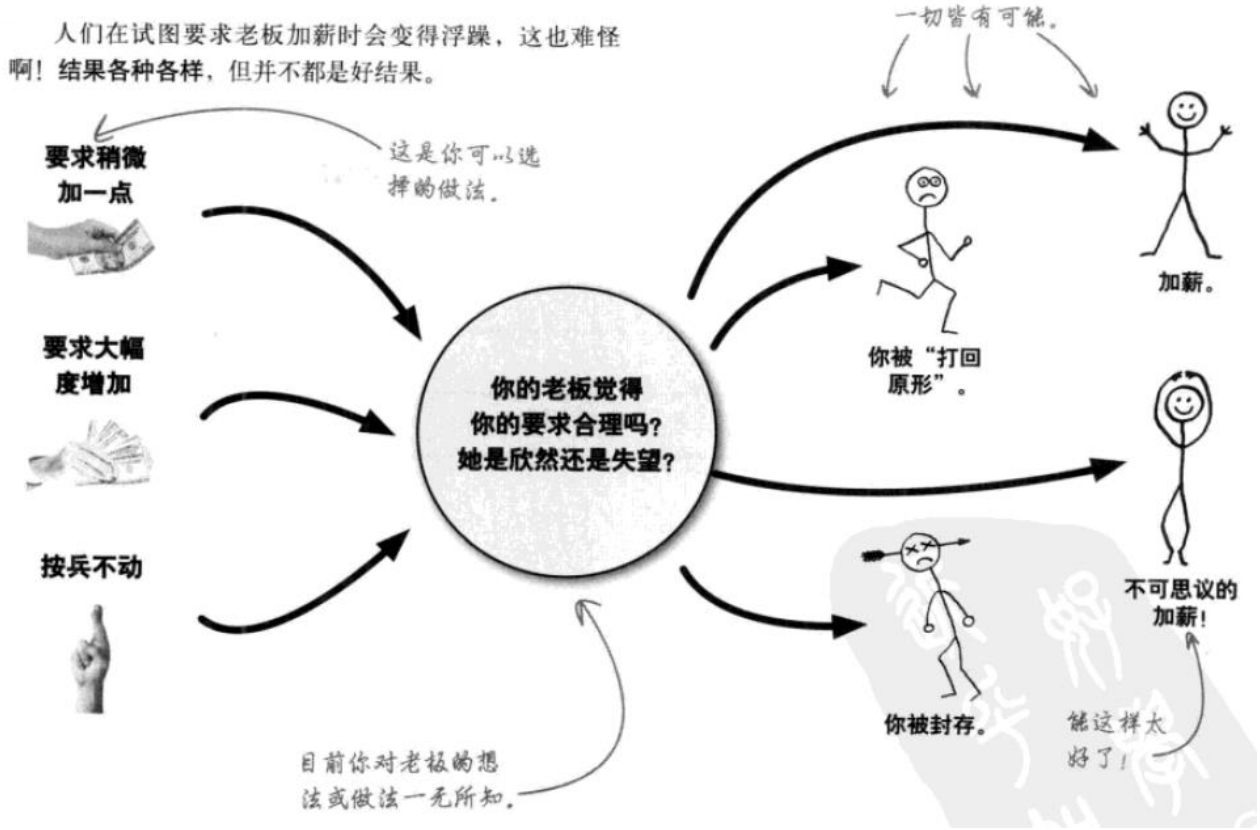
■提供全面的数据价值资讯，涵盖商业智能与数据分析、大数据、企业信息化、数字化技术等，各种高性价比课程信息，赶紧掏出您的手机关注吧！

直方图与数据分布

数据的图形表示方法不计其数，直方图就是其中一种用于独立数据分布、差异、集中趋势的图表，那我们到底从直方图看出了什么？数据的分布是否有规律可循？我们怎么总结这些分布规律？

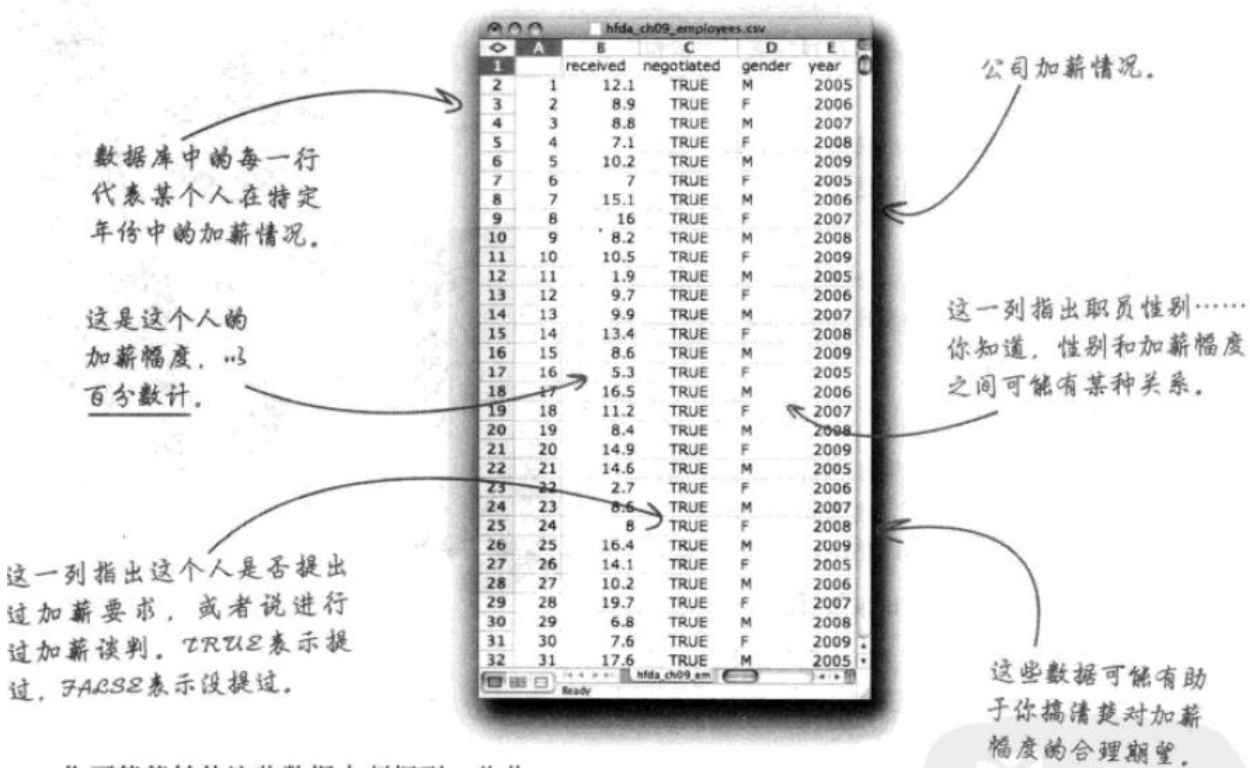
加薪？！

- 员工考核即将到来，你正在犹豫，是否应该向老板提出加薪要求



加薪？！

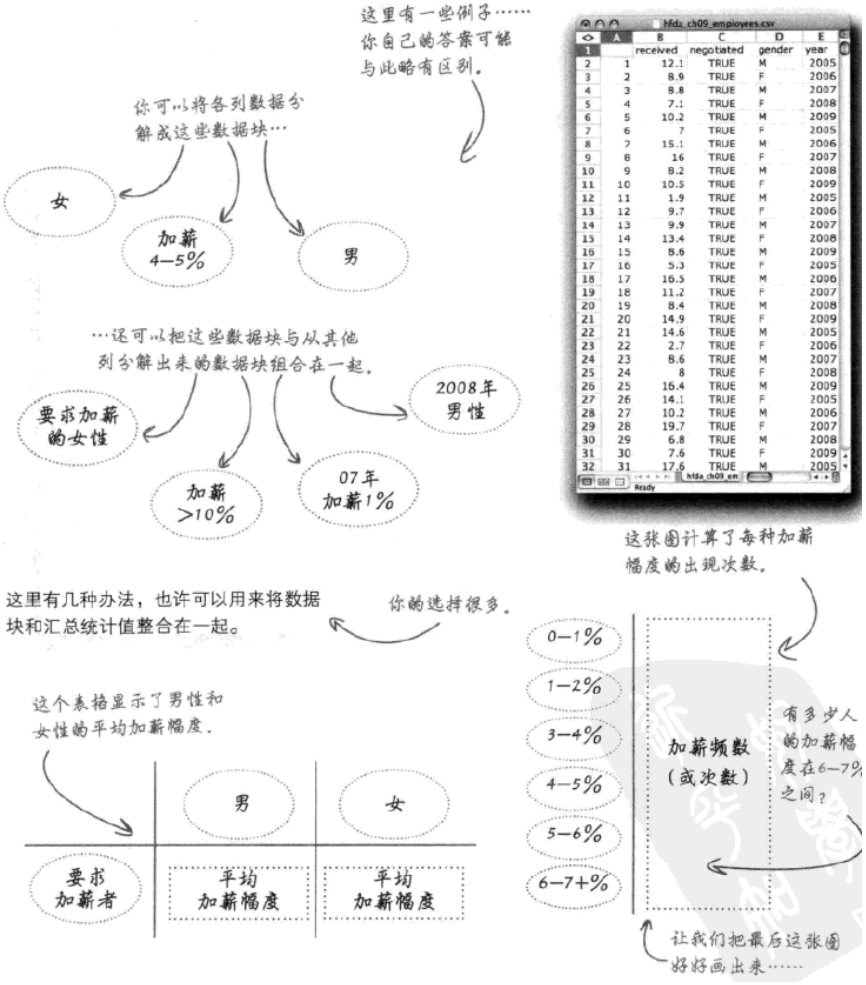
- 借助数据的力量



	A	B	C	D	E
1		received	negotiated	gender	year
2	1	12.1	TRUE	M	2005
3	2	8.9	TRUE	F	2006
4	3	8.8	TRUE	M	2007
5	4	7.1	TRUE	F	2008
6	5	10.2	TRUE	M	2009
7	6	7	TRUE	F	2005
8	7	15.1	TRUE	M	2006
9	8	16	TRUE	F	2007
10	9	8.2	TRUE	M	2008
11	10	10.5	TRUE	F	2009
12	11	1.9	TRUE	M	2005
13	12	9.7	TRUE	F	2006
14	13	9.9	TRUE	M	2007
15	14	13.4	TRUE	F	2008
16	15	8.6	TRUE	M	2009
17	16	5.3	TRUE	F	2005
18	17	16.5	TRUE	M	2006
19	18	11.2	TRUE	F	2007
20	19	8.4	TRUE	M	2008
21	20	14.9	TRUE	F	2009
22	21	14.6	TRUE	M	2005
23	22	2.7	TRUE	F	2006
24	23	8.6	TRUE	M	2007
25	24	8	TRUE	F	2008
26	25	16.4	TRUE	M	2009
27	26	14.1	TRUE	F	2005
28	27	10.2	TRUE	M	2006
29	28	19.7	TRUE	F	2007
30	29	6.8	TRUE	M	2008
31	30	7.6	TRUE	F	2009
32	31	17.6	TRUE	M	2005

将数字转为图表

- 如何将数据信息整合在一张图中？



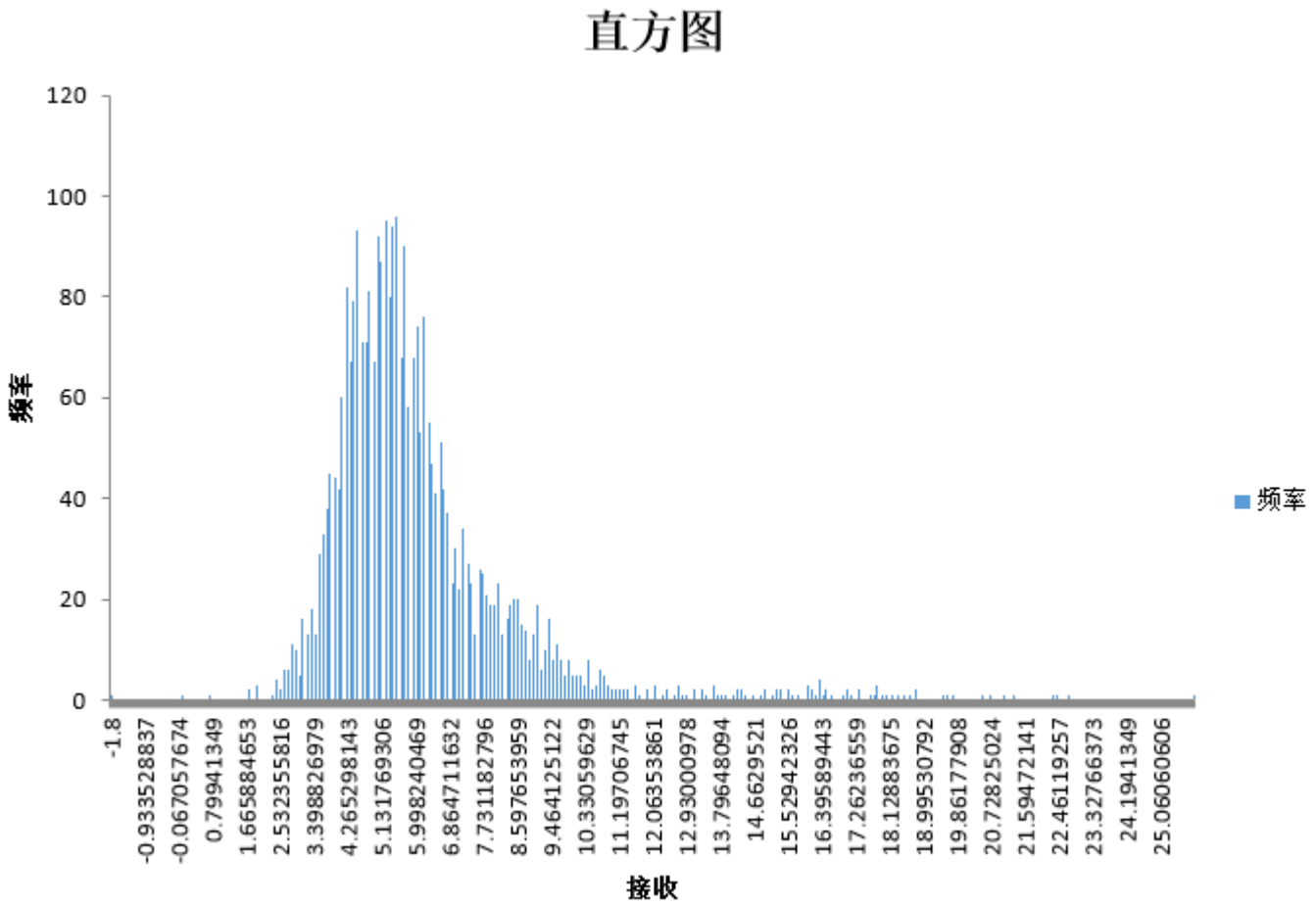
直方图

- 什么是直方图？
  - 频数直方图
  - 频率直方图

- 怎么画直方图？
  - Excel中的直方图本质是柱形图
  - 步骤：1. 设置分组；2. 统计分组数量；3. 作图

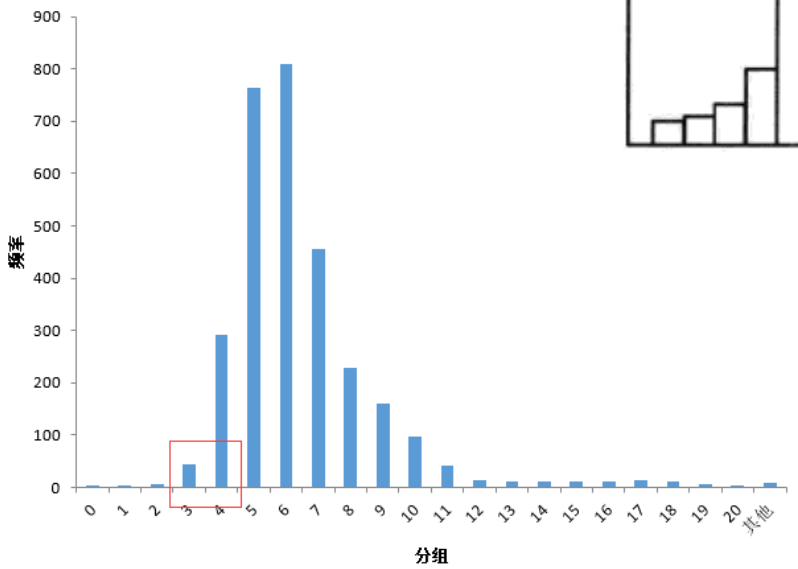
直方图

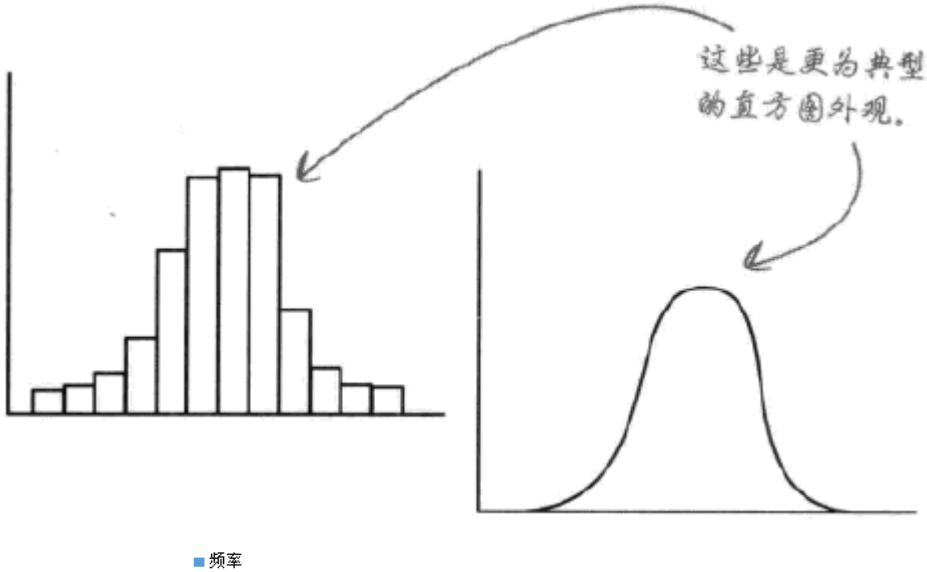
- 展示数据分布情况



直方图

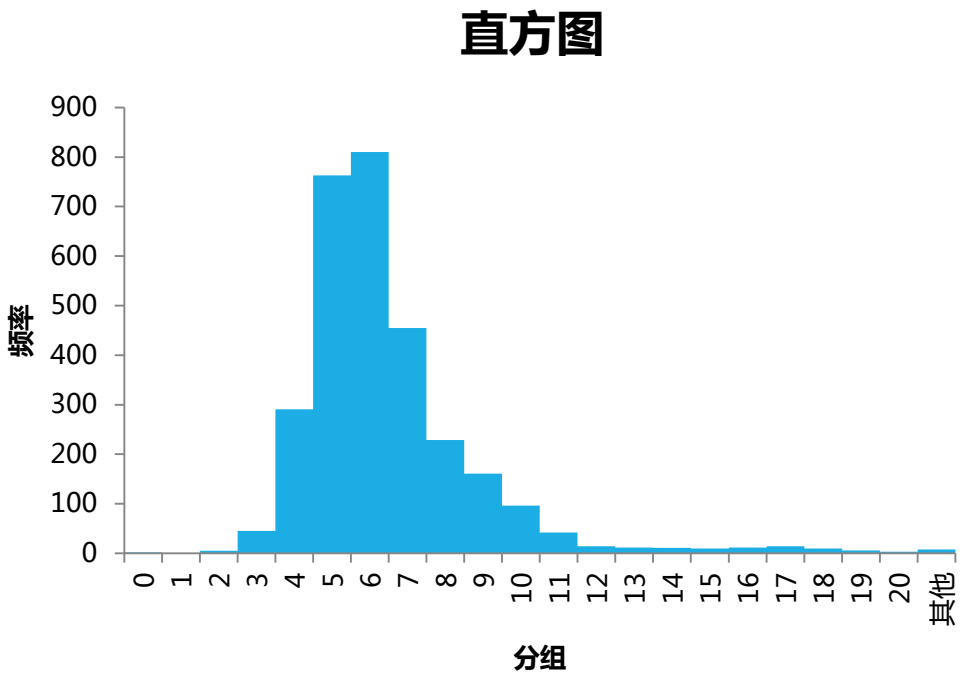
- Excel的直方图跟一般意义上的直方图看起来有点不一样





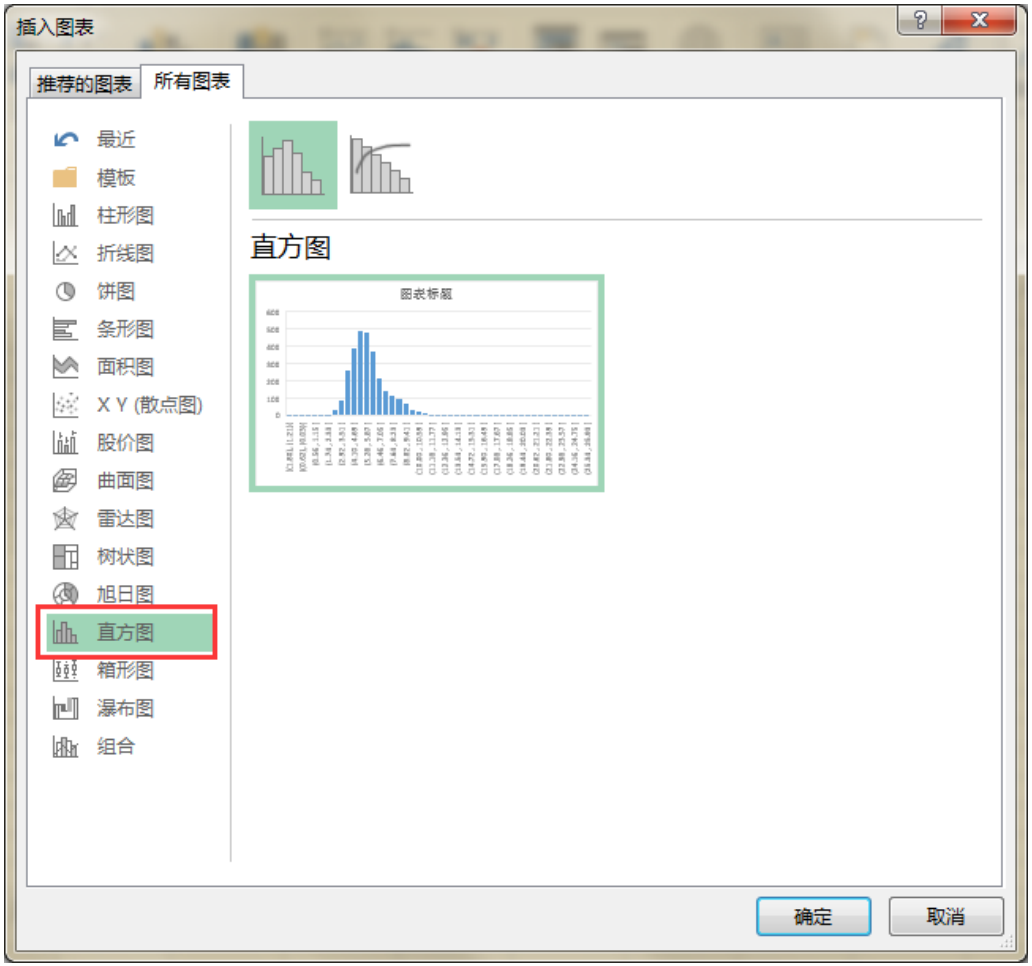
直方图

- 如何在Excel中画一张正确的直方图？

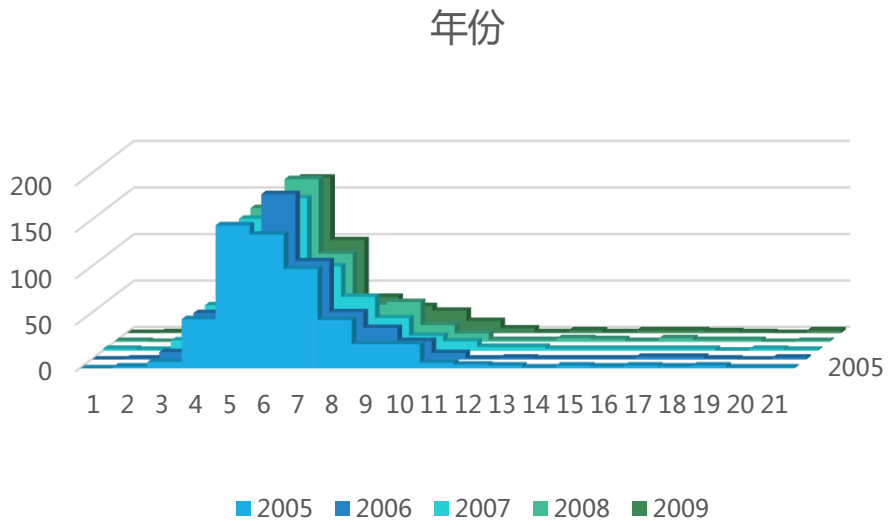


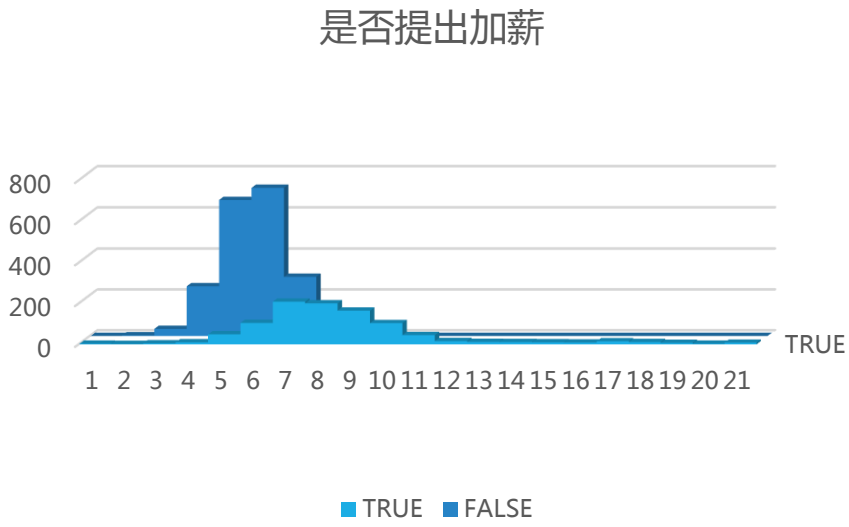
直方图

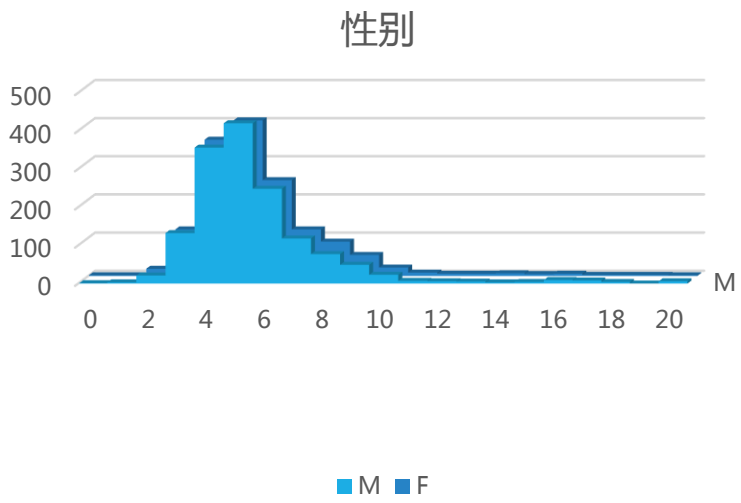
- Excel2016新增的直方图



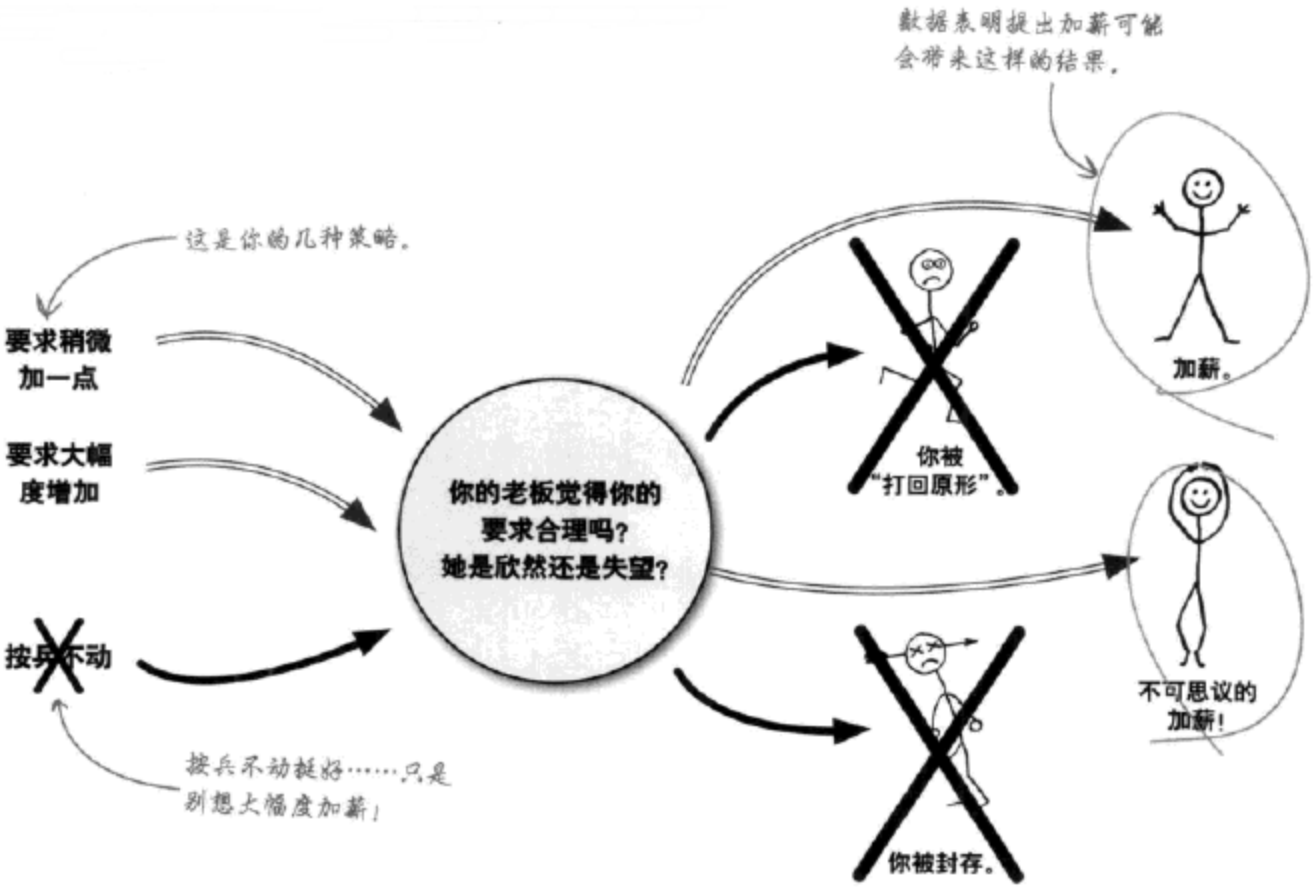
分组表示


年份
200
150
100
50
0
2005
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
2005 2006 2007 2008 2009


是否提出加薪
800
600
400
200
0
TRUE
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
TRUE FALSE

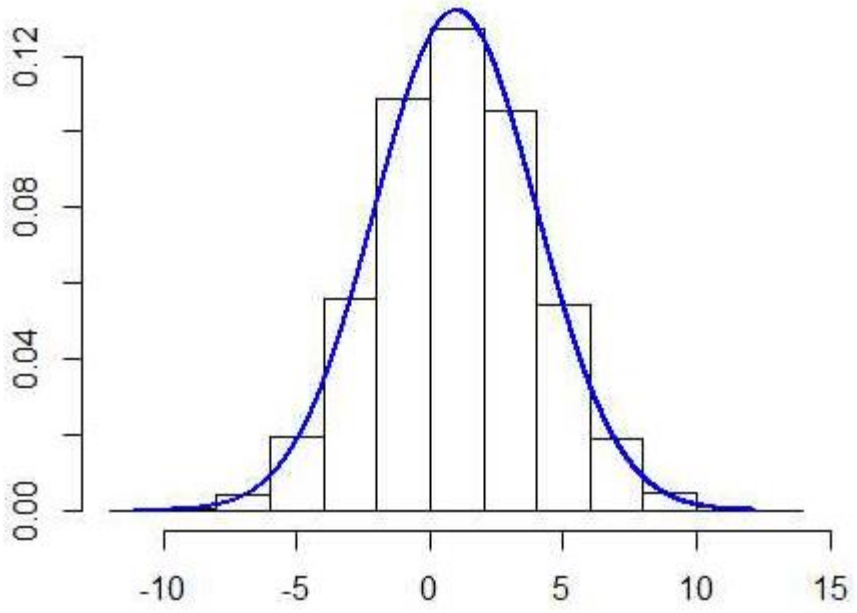

性别
500
400
300
200
100
0
M
0 2 4 6 8 10 12 14 16 18 20
M F

谈判加薪吧


数据表明提出加薪可能会带来这样的结果。
这是你的几种策略。
要求稍微加一点
要求大幅度增加
按兵不动
你的老板觉得你的要求合理吗？她是欣然还是失望？
你被"打回原形"。
加薪。
不可思议的加薪！
你被封存。
按兵不动挺好……只是别想大幅度加薪！

从直方图到概率密度图

- 直方图————概率密度图



常见分布

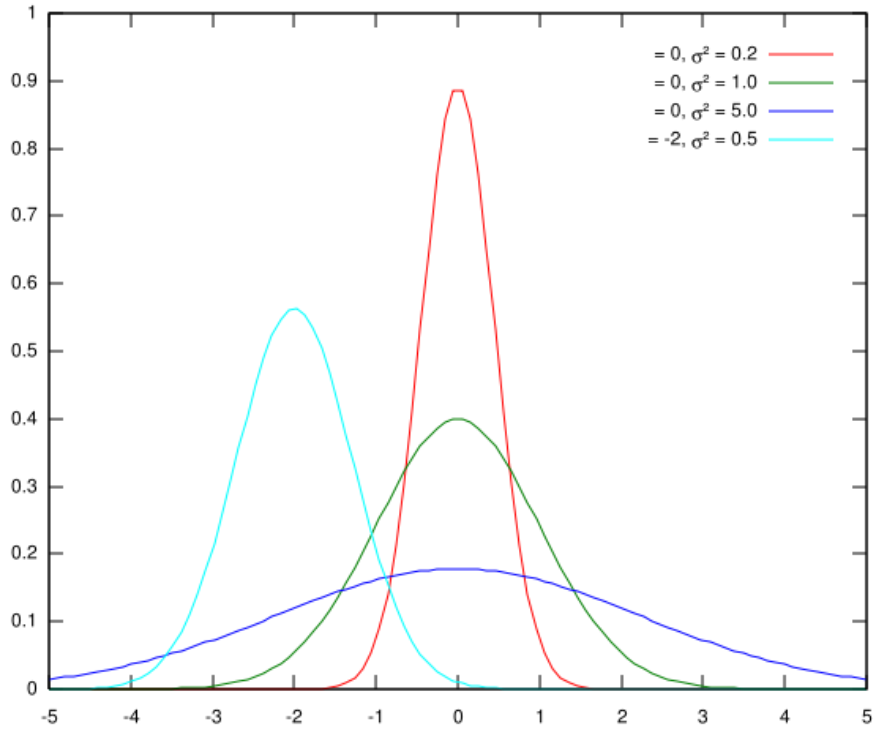
- 连续分布
  - 正态分布
  - 卡方分布
  - T分布
  - F分布
  - 均匀分布
- 离散分布
  - 0-1分布
  - 二项分布
  - 泊松分布

正态分布

- 正态分布 N(μ， σ)
  - 形状特点：对称，呈钟型分布
  - 参数：均值μ，标准差σ
  - 概率密度函数

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

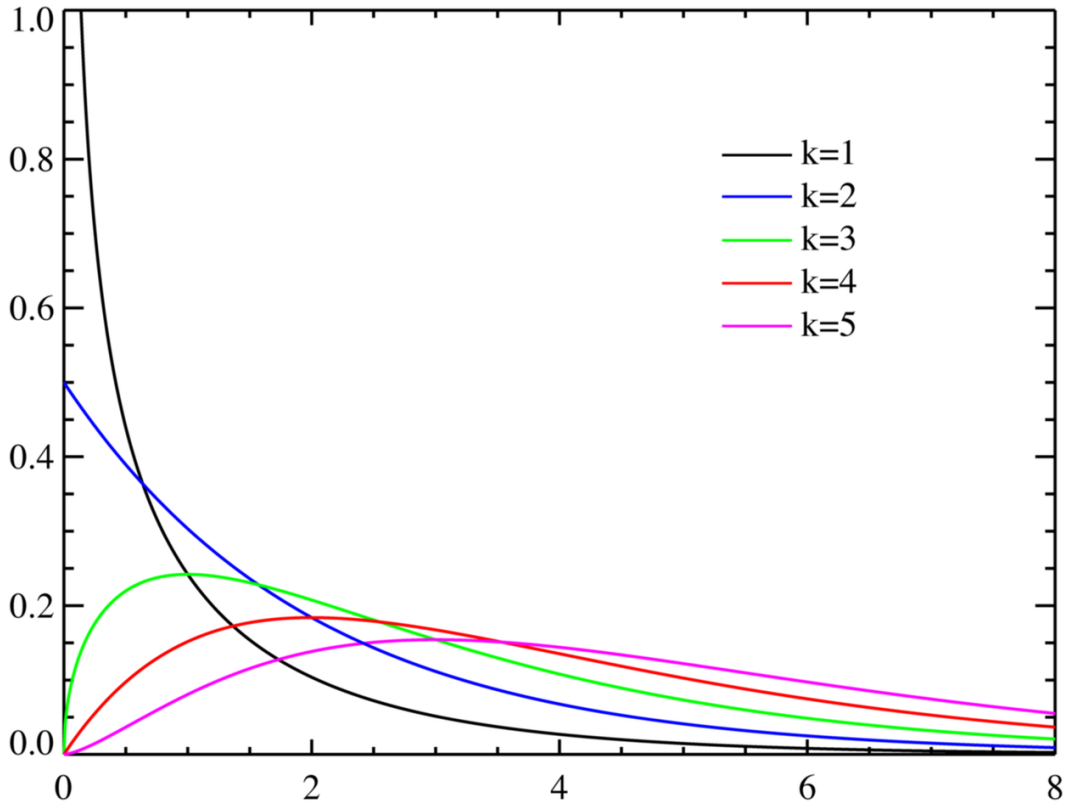
- 标准正态分布（图中绿色线）
  - 均值为0，标准差为1的正态分布



卡方分布

- 卡方分布——k个标准正态分布的平方相加得到的分布
  - 形状特征：取值都大于等于0；不同自由度形状差异大
  - 参数：自由度
  - 概率密度函数

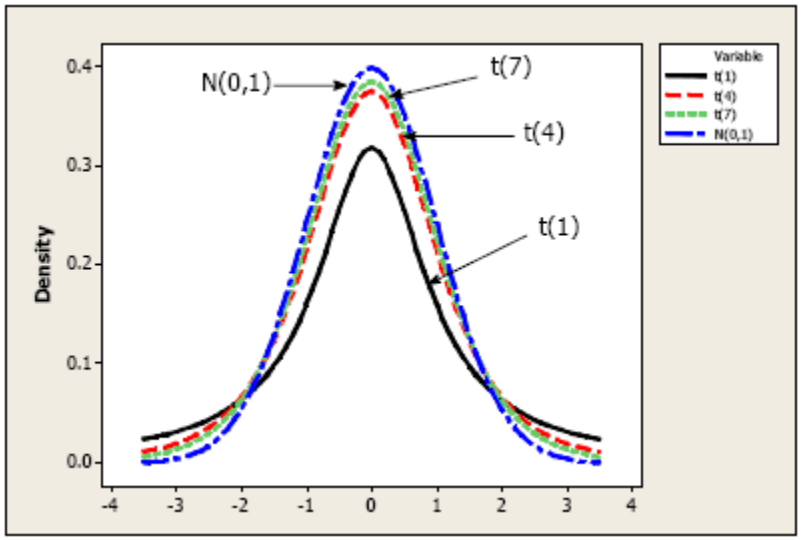
$$f_k(x) = \frac{(1/2)^{k/2}}{\Gamma(k/2)} x^{k/2-1} e^{-x/2}$$



T分布

- T分布——标准正态分布/卡方分布
  - 形状特点：与标准正态分布类似，但比标准正态分布要矮胖
  - 参数：自由度
  - 概率密度函数

$$f(t) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi}\,\Gamma(\nu/2)}(1+t^2/\nu)^{-(\nu+1)/2}$$



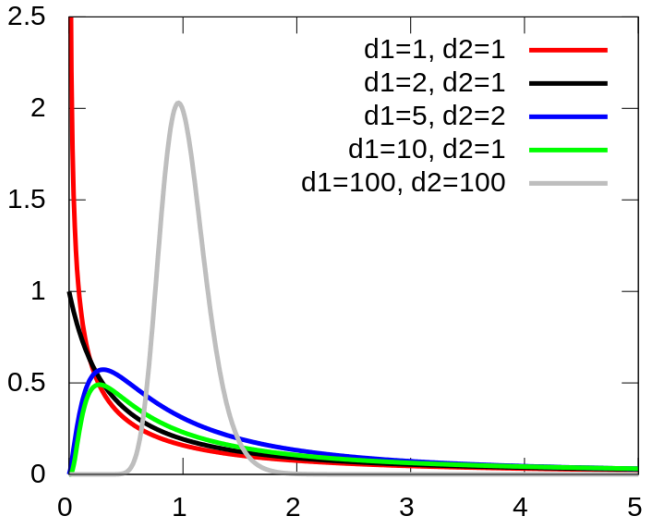
F分布

- F分布——两个卡方分布变量的比率

$$\frac{U_1/d_1}{U_2/d_2} = \frac{U_1/U_2}{d_1/d_2}$$

其中：

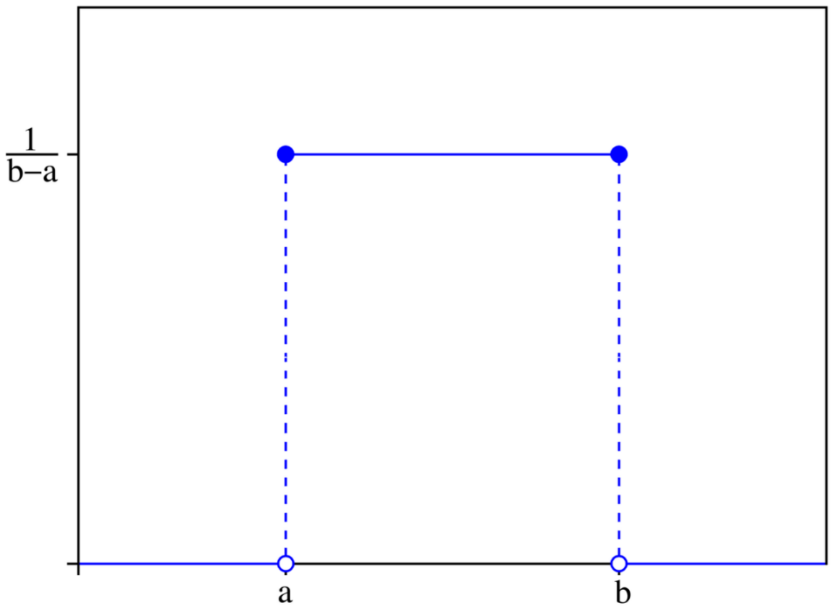
- U_1和U_2呈卡方分布，它们的自由度（degree of freedom）分别是d_1和d_2
- U_1和U_2是相互独立的。



均匀分布

- 均匀分布——每一个取值都具有相同的可能性
  - 参数：上界a；下界b
  - 密度函数

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \le x \le b \\ 0 & \text{elsewhere} \end{cases}$$



离散变量与离散分布

- 连续与离散？
- 什么是随机变量？
  - 本质是一个函数/映射，从随机试验结果到实数上的映射
  - 设X是抛10次硬币得到的正面向上的次数，求P(X=2)
- 离散随机变量与连续随机变量
  - 离散：取值是有限个数值
  - 连续：取值是一个区间范围
- 离散分布与连续分布

0-1分布

- 贝努利试验——只有两种可能结果的单次随机试验
  - 明天是否下雨
  - 1分钟内要等的公交是否会到站
  - 买一次彩票是否会中奖
  - 抛一次硬币是否会得到正面向上
  - ……

- 概率密度函数

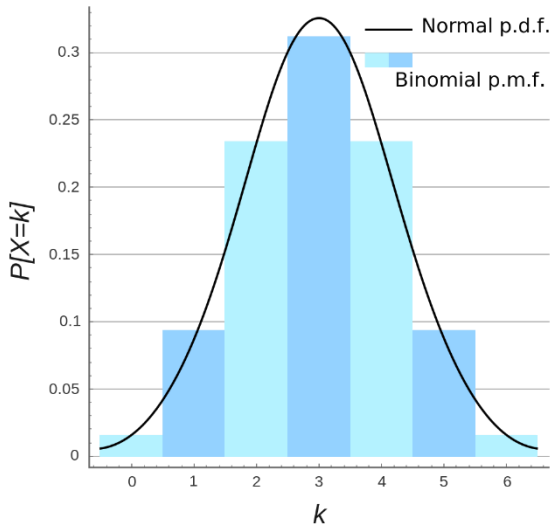
$$f_X(x) = p^x(1-p)^{1-x} = \begin{cases} p & \text{if } x = 1, \\ q \equiv 1-p & \text{if } x = 0, \\ 0 & \text{otherwise.} \end{cases}$$

二项分布

- 二项分布B(n,p)——n个0-1分布相加
  - 抛10次硬币，得到正面向上的次数
  - 可以看做是离散的正态分布

- 概率密度函数

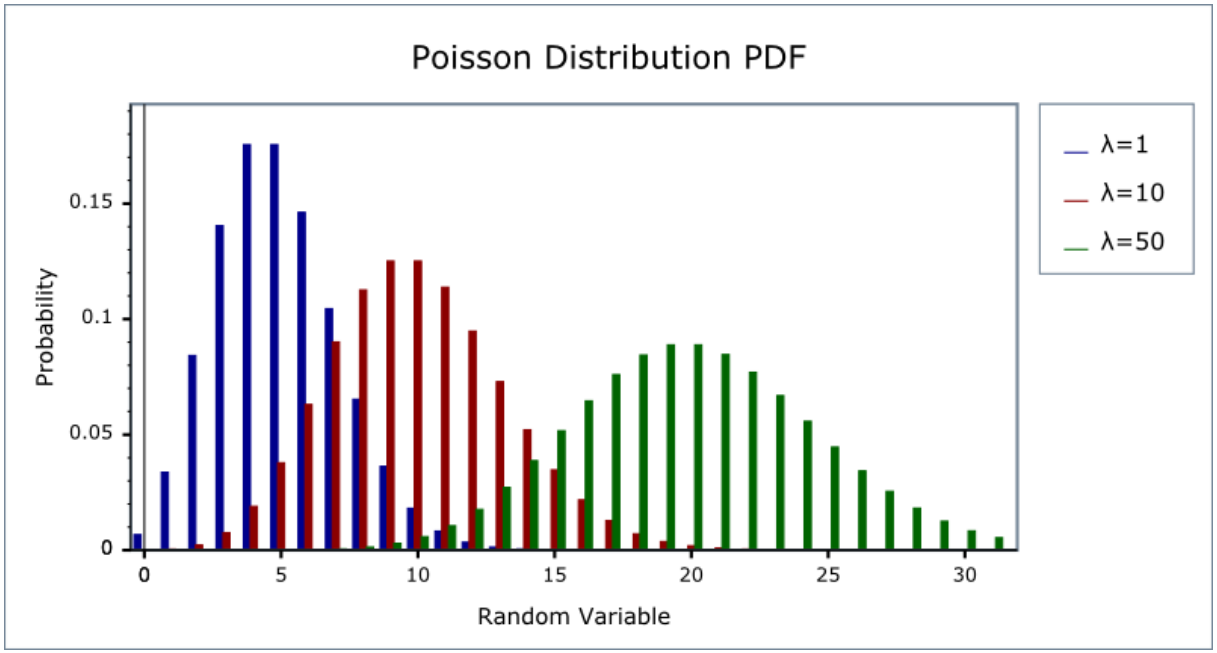
$$f(k;n,p) = \Pr(K = k) = \binom{n}{k} p^k (1-p)^{n-k}$$



泊松分布

◆ 泊松分布

- 适合于描述单位时间内随机事件发生的次数的概率分布
- 10分钟内，某服务台接到电话的次数
- 10年内，地震发生的次数
- 10分钟内，某一公交车站进站公交的数量
- ……



炼数成金逆向收费式网络课程

- Dataguru（炼数成金）是专业数据分析网站，提供教育，媒体，内容，社区，出版，数据分析业务等服务。我们的课程采用新兴的互联网教育形式，独创地发展了逆向收费式网络培训课程模式。既继承传统教育重学习氛围，重竞争压力的特点，同时又发挥互联网的威力打破时空限制，把天南地北志同道合的朋友组织在一起交流学习，使到原先孤立的学习个体组合成有组织的探索力量。并且把原先动辄成千上万的学习成本，直线下降至百元范围，造福大众。我们的目标是：低成本传播高价值知识，构架中国第一的网上知识流转阵地。
- 关于逆向收费式网络的详情，请看我们的培训网站 http://edu.dataguru.cn

Thanks

FAQ时间