



专注于商业智能BI和大数据的垂直社区平台

# 数据的分布

Allen

[www.hellobi.com](http://www.hellobi.com)

## 课程目录

- 正态分布
- Poission分布
- 指数分布
- 小结

## R语言中的分布函数

- 分布律： $P\{X = x_k\} = p_k, k = 1, 2, \dots$
- 概率密度函数  $f(x)$  和分布函数  $F(x)$  的关系  $F(x) = P\{X \leq x\} = \int_{-\infty}^x f(t)dt, -\infty < x < \infty$

# 正态分布

- 正态分布：

- 若正态分布的均值是 $\mu$ ，方差是 $\sigma^2$ ，对任意变量 $x$ ,其分布函数为

$$F(x) = \frac{1}{\sqrt{2\pi} \sigma} \int_{-\infty}^x \exp\left\{-\frac{(t-\mu)^2}{2\sigma^2}\right\} dt$$

$pnorm(x, \mu, \sigma)$

- 若正态分布的均值是 $\mu$ ，方差是 $\sigma^2$ ，对任意变量 $x$ ,其概率密度函数为

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

$dnorm(x, \mu, \sigma)$

# 正态分布

- R语言中正态分布所用函数：

函数	描述	参数
$dnorm(x, \mu, \sigma)$	概率密度函数	参数分别是数值向量、均值、标准差
$pnorm(x, \mu, \sigma)$	概率分布函数	参数分别是数值向量、均值、标准差
$qnorm(p, \mu, \sigma)$	计算分位点函数	参数分别是概率构成数值向量、均值、标准差
$rnorm(n, \mu, \sigma)$	生成随机数函数	参数分别是产生随机数的个数、均值、标准差

# Possion分布

- Possion分布：
- Possion分布式离散分布，当x是整数k时，概率为  $P\{X = k\} = \frac{\lambda e^{-\lambda}}{k!}$

*dpois(k, λ)*

# Possion分布

- R语言中Possion分布所用函数：

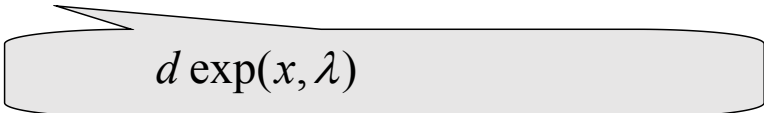
函数	描述	参数
$dpois(k, \lambda)$	求概率	参数分别是数值向量、分布参数
$ppois(k, \lambda)$	概率分布	参数分别是数值向量、分布参数
$qpois(p, \lambda)$	计算分位点函数	参数分别是概率构成数值向量、分布参数
$rpois(n, \lambda)$	生成随机数函数	参数分别是产生随机数的个数、分布参数

# 指数分布

- 指数分布：

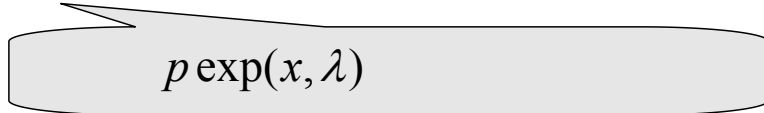
- 若  $\lambda > 0$  是指数分布的一个参数，指数分布的概率密度函数为

$$f(x, \lambda) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$


$$d \exp(x, \lambda)$$

- 若  $\lambda > 0$  是指数分布的一个参数，指数分布的概率分布函数为

$$F(x, \lambda) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$


$$p \exp(x, \lambda)$$



## 指数分布

- R语言中指数分布所用函数：

函数	描述	参数
$d \exp(x, \lambda)$	概率密度函数	参数分别是数值向量、分布参数
$p \exp(x, \lambda)$	概率分布函数	参数分别是数值向量、分布参数
$q \exp(p, \lambda)$	计算分位点函数	参数分别是概率构成数值向量、分布参数
$r \exp(n, \lambda)$	生成随机数函数	参数分别是产生随机数的个数、分布参数

## 其他分布

- R语言中其他分布所用函数：

分布	函数	前缀
F分布	f	d概率密度函数 p概率分布函数 q求分位数 r产生随机数
t分布	t	
卡方分布	chisq	

## 小结

- 正态分布
- Poission分布
- 指数分布
- 小结