



上海交通大学  
SHANGHAI JIAO TONG UNIVERSITY



# t-test using R

资料搜索：谭润滋

ppt制作：黄宇翔

ppt讲解：余晨阳





# 目录

- `t.test()`
- `pt()`
- `dt()`
- `qt()`
- `rt()`





## t.test()

- R环境下可调用t.test()函数对数据作t检验。
- 函数一般形式为：`t.test(x = mydata1, y = mydata2, alternative = "two.sided", mu = mean, paired = FALSE, var.equal = FALSE, conf.level = 0.95 )`





#t-test:

```
>t.test(x = mydata1, y = mydata2, alternative = "two.sided",  
mu = mean, paired = FALSE, var.equal = FALSE, conf.level =  
0.95 )
```

- **x, y**为待检验的数据集，其中**x**不可为空，**y**默认为**NULL**。若仅输入**x**，则函数进行**单样本t检验**。当同时输入**x, y**时，函数进行**双样本t检验**。





#t-test:

```
>t.test(x = mydata1, y = mydata2, mu = mean, alternative =  
"two.sided", paired = FALSE, var.equal = FALSE, conf.level =  
0.95 )
```

- **mu**为假设均值，表示原假设中事先判断的均值，默认值为0。
- **alternative**用于指定所求置信区间的类型，默认为“two.sided”，表示求双尾的置信区间；若为“less”则求置信上限；若为“greater”则求置信下限。



#t-test:

```
>t.test(x = mydata1, y = mydata2, alternative = "two.sided",  
mu = mean, paired = FALSE, var.equal = FALSE, conf.level =  
0.95 )
```

- **paired**是逻辑值，表示是否进行配对样本t检验，默认FALSE。
- **var.equal**是逻辑值，表示双样本检验时两个总体的方差是否相等，默认FALSE。
- **conf.level**用来表示区间的置信水平，默认值为**0.95**。



- `t.test()`只适用于知道原始数据的样本。若原始数据未知，则需要采用其他办法。例如

已知某水样中含碳酸钙的真值为20.7mg/L，现用某法重复测定该水样12次，碳酸钙的含量分别为

**20.99,20.41,20.10,20.00,20.91,22.60,20.99,20.42,20.90,22.99,23.12,20.89。**

该法测定碳酸钙含量所得的均值与诊治有无显著差异？





#代码如下

```
> x <- c(20.99,20.41,20.10,20.00,20.91,22.60,20.99,20.42,20.90,22.99,23.12,20.89)  
> t.test(x, alternative = "greater", mu = 20.7 )
```

### One Sample t-test

data: x

t = 1.5665, df = 11, p-value = 0.07276

alternative hypothesis: true mean is greater than 20.7

95 percent confidence interval:

20.62778    Inf

sample estimates:

mean of x

21.19333







# pt()

- R下的分布函数，其形式为**pt(q, df, ncp, lower.tail = TRUE, log.p = FALSE)**，输出结果为 $P(X < t)$ 。
- **q**为t值，可以是包含多个值的向量。
- **df**为自由度。
- **lower.tail**默认为False，输出 $P(X < t)$ ；为TRUE时，输出结果为 $P(X > t)$ 。
- **Log.p**=TRUE时，结果取关于自然底数e的对数。





dt()

- R下的概率密度函数，其形式为**dt(x, df, ncp, log = FALSE)**
- 参数x为t，返回值是概率密度，即t分布分布图上某一点的高度值。





# qt()

- R下的分位数函数，其形式为**qt(p, df, ncp, lower.tail = TRUE, log.p = FALSE)**
- 参数

是概率值，该函数相当于pt()的逆运算。





$rt()$

- R下的t分布随机生成函数，其形式为 $rt(n, df, ncp)$
- 用于生产 $n$ 个一定自由度下符合标准t分布的随机数。





# 例题

未知原始数据的参数检验方法如下：

## 1、单样本假设检验

健康成年男子脉搏均数为72次/分。某医生在某山区随机抽查健康成年男子25人，其脉搏均数为74.2次/分，标准差为6.5次/分。根据这个资料能否认为某山区健康成年男子脉搏数与一般健康成年男子的不同？





> #根据公式算出t值

```
> x <- 74.2
```

```
> mu <- 72
```

```
> s <- 6.5
```

```
> n <- 25
```

```
> t <- (x-mu) / (s/sqrt(n))
```

```
> t
```

```
[1] 1.692308
```

> #用pt()函数，输入t值和自由度df ( n-1 ) ，得到p值

```
> p <- pt(t,df=24)
```

```
> p
```

```
[1] 0.9482341
```





- 2、双样本配对检验
- 慢性支气管炎病人血中胆碱酯酶活性常常偏高。某校药理教研室将同性别同年龄的病人与健康人配成8对，测量该值加以比较，配对两组人差值的均值为0.625，标准差为0.78，问可否通过这一资料得出较明确的结论？





> #依次输入配对样本的差值d、标准差s、配对数n

> d <- 0.625;s <- 0.78;n <- 8

> #算t值

> t <- d / (s/sqrt(n))

> #输入自由度n-1, pt()函数得到p值

> df <- n-1

> p <- pt(t,df)

> t

[1] 2.266368

> p

[1] 0.9711069







- 3、双样本独立检验
- 测量某两个地区水中碳酸钙的含量，分别从两个地区随机抽取20份样品进行碳酸钙检测，分别得到两个地区碳酸钙含量的均数和标准差，结果见下图。试判断两个地区水中碳酸钙的含量是否有差异？





```
> #输入对照组实验组均值x1, x2; 组数n1, n2; 方差s1, s2
> x1<-20.95; x2<-21.79; n1<-20; n2<-20; s1<-5.89; s2<-3.43
> #计算两独立样本共同的标准差
> sc <- sqrt((1/n1+1/n2)*((n1-1)*s1**2+(n2-1)*s2**2)/(n1+n2-2))
> #t值, 自由度df, p值
> t <- (x2-x1)/sc
> df <- n1+n2-2
> p <- pt(t,df)
> t
[1] 0.5511486
> p
[1] 0.7076209
```





# 练习

- 1、用仪器间接测量某生物实验室的连续灭菌设备，重复5次，结果分别为 $125^{\circ}\text{C}$ ， $126.5^{\circ}\text{C}$ ， $124.5^{\circ}\text{C}$ ， $126.0^{\circ}\text{C}$ ， $127.5^{\circ}\text{C}$ 。若测量值服从 $N(\mu, \sigma^2)$ ，置信度为0.05，可否认为该仪器温度低于 $127.7^{\circ}\text{C}$ （真实值）？
- 2、测量温度对某微生物存活时间的影响，甲、乙两组分别在70和80摄氏度条件下独立测量8次，两组分别需要的灭菌时间均值为20.9和20.1，标准差为0.94和0.91。
  - (1) 若知道两种温度下方差的真值相等， $\alpha=0.05$ ，可否认为温度对存活时间无影响？
  - (2) 若不知道两种温度方差是否相同， $\alpha=0.05$ ，可否认为温度对存活时间无影响？



上海交通大学  
SHANGHAI JIAO TONG UNIVERSITY



谢谢

