

R描述性统计函数

2023年9月18日 23:32

计算	统计检验	统计检验
log(x)	mean(x)	lm(y ~ f+x)
log10(x)	sd(x)	lm(y ~ x1+x2+x3)
exp(x)	var(x)	bartlett.test
sin(x)	median(x)	binom.test
cos(x)	quantile(x,p)	fisher.test
tan(x)	cor(x,y)	chisq.test
asin(x)	t.test()	glm(y ~
acos(x)	lm(y ~ x)	x1+x2+x3,
min(x)	wilcox.test()	binomial)
max(x)	kruskal.test()	friedman.test
range(x)		...
length(x)		

选项和参数

输入控制

file : 接一个文件
data : 一般指要输入一个数据框 ;
x : 表示单独的一个对象 , 一般都是向量 , 也可以是矩阵或者列表 ;
x和y : 函数需要两个输入变量 ;
x , y , z : 函数需要三个输入变量 ;
formula : 公式 ;
na.rm : 删除缺失值 ;

输出控制

调节参数

根据名字判断选项作用

选项接受哪些参数

main : 字符串 , 不能是向量
na.rm : TRUE或者FALSE
axis : side参数只能是1到4
fig : 包含四个元素的向量
.....

1. 数学统计函数

概率分布函数:

d 概率密度函数

p 分布函数

q 分布函数的反函数

r 产生相同分布的随机数

分布	R 中的名称	附加参数
beta	beta	shape1, shape2, ncp
binomial	binom	size, prob
Cauchy	cauchy	location, scale
chi-squared	chisq	df, ncp
exponential	exp	rate
F	f	df1, df2, ncp
gamma	gamma	shape, scale
geometric	geom	prob
hypergeometric	hyper	m, n, k
log-normal	lnorm	meanlog, sdlog
logistic	logis	location, scale
negative binomial	nbinom	size, prob
normal	norm	mean, sd
Poisson	pois	lambda
Student's t	t	df, ncp
uniform	unif	min, max
Weibull	weibull	shape, scale

可用于生成按一定分布的随机数
 例: `rnorm(n=100,mean=15,sd=2)`

2. 描述性统计函数

`mean()` 平均数

`var()` 方差

`sd()` 标准差

`summary()` 给出平均值, 四个分位数的基本信息

`summary(mtcars[myvars])`

分组计算描述性统计量

`aggregate()` 分组进行统计

`aggregate(数据集, by=list(name=列按什么分), 什么统计函数)`

`by(data, INDICES, FUN)`

`data`: 数据框或矩阵

`INDICES` 是一个因子或因子组成的列表, 定义了分组

`FUN` 是任意函数

3. 频数统计函数

`as.factor()` 转换为因子, 再用 `split()` 进行分组切割

或者 `cut(vector, breaks=分割成几段或给个向量来分隔)` 进行切割

函数	描述
<code>table(var1, var2, ..., varN)</code>	使用N个类别型变量(因子)创建一个N维列联表
<code>xtabs(formula, data)</code>	根据一个公式和一个矩阵或数据框创建一个N维列联表
<code>prop.table(table, margins)</code>	依margins定义的边际列表将表中条目表示为分数形式
<code>margin.table(table, margins)</code>	依margins定义的边际列表计算表中条目的和
<code>addmargins(table, margins)</code>	将概述边margins(默认是求和结果)放入表中
<code>ftable(table)</code>	创建一个紧凑的“平铺”式列联表

`table()` 列联表形式统计每个分段有多少数量

`prop.table()` 统计分组概率

`xtabs(~变量, data=数据框数据)` 以变量做成列连表

`margin.table(列连表统计数据, 1)` 统计每列或每行总频数

4. 类别型变量独立性检验

独立性检验是根据频数信息判断两类因子彼此相关或相互独立的假设检验。所谓独立性就是指变量间是独立的，没有关系

原假设：两者独立

卡方检验

table()整理成二维列联表频数图

	None	Some	Marked
Placebo	29	7	7
Treated	13	7	21

chisq.test()

Fisher检验

fisher.test()

Cochran-Mantel-Haenszel检验

mantelhaen.test()

多个变量独立性，和输入变量顺序有关

5. 相关性分析

相关系数可以用来描述定量变量之间的关系。相关性分析指对具备相关性的变量分析，衡量密切程度

cor(x,use=,method=)

参数	描述
x	矩阵或数据框
use	指定缺失数据的处理方式。可选的方式为all.obs(假设不存在缺失数据——遇到缺失数据时将报错)、everything(遇到缺失数据时，相关系数的计算结果将被设为missing)、complete.obs(行删除)以及 pairwise.complete.obs(成对删除，pairwise deletion)
method	指定相关系数的类型。可选类型为pearson、spearman或kendall

偏相关是指在控制一个或多个定量变量时，另外两个定量变量之间的相互关系。

计算偏相关系数用ggm包中的pcor()函数

pcor(u,s)

u为数值向量前两个为计算相关系数的索引；其余为需要排除的条件变量。

s为变量的协方差矩阵

相关性的显著性检验

原假设为变量间独立

cor.test()只能检测两组间的相关显著性

psych包中的corr.test()可检测多组间

结果变量为连续型，比较两个组间结果是否有差异

t检验

独立样本的t检验

原假设两组样本均值相等，结果没有差异

r与多数统计软件不同的时t检验默认假定方差不相等，可以加参数var.equal=TRUE来假定方差相等

t.test(y ~ x, data)

其中的y是一个数值型变量，x是一个二分变量。调用格式或为：

非独立样本的t检验

加入paired=TRUE的参数

6. 非参数检验

总体不满足正态分布

两组的比较(一个为类别变量，另一个为数值变量)

Wilcoxon秩和检验(Mann-Whitney U检验)

wilcox.test(y~x, data)

7. 绘图函数

基础绘图系统，系统自带的graphics包plot(),par()对绘图参数进行设置

lattice包

ggplot2包

grid包

8. 自定义函数

利用function函数来声明

```
myfun <- function ( 选项参数 )  
  {  
    函数体  
  }
```