

# 《统计建模与R软件》

## 图书基本信息

书名：《统计建模与R软件》

13位ISBN编号：9787302143666

10位ISBN编号：7302143668

出版时间：2007-5

出版社：清华大学出版社

作者：薛毅

页数：525

版权说明：本站所提供下载的PDF图书仅提供预览和简介以及在线试读，请支持正版图书。

更多资源请访问：[www.tushu111.com](http://www.tushu111.com)

# 《统计建模与R软件》

## 内容概要

统计建模与R软件，ISBN：9787302143666，作者：薛毅

## 书籍目录

- 第1章 概率统计的基本知识
- 第2章 R软件的使用
- 第3章 数据描述性分析
- 第4章 参数估计
- 第5章 假设检验
- 第6章 回归分析
- 第7章 方差分析
- 第8章 应用多元分析 ( )
- 第9章 应用多元分析 ( )
- 第10章 计算机模拟
- 附录 索引
- 参考文献

# 《统计建模与R软件》

## 编辑推荐

《统计建模与R软件》可作为理工、经济、管理、生物等专业学生数理统计课程的辅导教材或教学参考书，也可作为统计计算课程的教材和数学建模竞赛的辅导教材。

## 精彩短评

- 1、差不多是国内R语言方面早起的书籍，比较完整，里面很多的自编函数，目前对于R来说，已经不用那么麻烦了，但是通过此书来学习R挺好
- 2、当年学R就靠它，可惜现在都不用R，就当积累统计基础吧
- 3、氣死了!!!!連個偏最小二乘法的教程都木有!!!!!!!!!!!!枉我還一大早坐了校巴去東風路校區的圖書館借!!!!!!!!!!!!!!!
- 4、很不错，直接用的这个当SPLUS的教材，和R也通用。老师也很好，直接索要课件，第二天就给我发邮箱了。很不错
- 5、整本书的内容还不错。
- 6、:C819/4702
- 7、p67:  $z <- 0:9$ ;  $\text{length}(z)$ 是10不是9....
- 8、没有看R的部分 数学渣的小福音
- 9、R比SAS要好玩的，不过这本好入门
- 10、讲的很细，适合学习。
- 11、R语言入门读物，适合初学者
- 12、感觉很一般，不如直接看R网站上的资源。
- 13、以R为中心，介绍了最重要的一些统计建模的方法，作为一本入门书，统计部分的知识没有强调的很多，对于我这种数学学渣简直是福音，基本上这书你从头看到尾也许有些公式应为没学过会有有的莫名其妙，但是不会说看不懂吧。五星好评。
- 14、很好的哟
- 15、我最近看了国内出版的一系列统计分析软件介绍的书，感觉这本书是最棒的，作者在写书时用了很多心思。优点如下：1.命令介绍详略得当。国内许多软件图书的作者习惯于把软件自带的Help翻译过来就当作命令的介绍。殊不知，在读者第一次接触该命令时不需要那么详细的信息，只需要能够有一个简单的例子就行了。本书作者在这方面做得很好，命令介绍很有克制，详略得当。2.软件操作循序渐进，并且不忽略一些软件实用的必要的小技巧。能够按照作者的介绍对软件Step by Step的操作，逐步深入理解软件。3.统计知识介绍适时，并且自成体系。作者在介绍相关命令的时候，能够介绍必要的统计知识。并且统计知识的介绍比较体系化。可以当做初学者入门的指导。如果对我这种已经学过相关知识的人，当做速查大纲也不错。而国内有些统计软件介绍的图书，如果要明白书中介绍的一个命令经常要参考三四本书才行。4.作者比较热情。昨天向作者索要书中自带的程序，下午发邮件给作者，晚上就有了回音。如果让我横向比较，我可以这么打分：本书 4.5张文彤 SPSS统计分析基础、高级教程 3.5卢文岱 SPSS For Windows 3刘军 UCINET 2.5
- 16、可算没那么难得要死 . .
- 17、看完这本书后，个人总结的学习顺序建议：

如果你的计量统计基础超扎实，直接看第二章看起，后面每一章直接做例题，R就学的差不多了。

如果计量统计基础不扎实，先看一遍第二章，然后返回第一章，然后跳三章，顺序看下去，这一遍只看理论讲解部分，看完一遍后，再返回第二章，第三章，顺序看下去，这一遍重点做题，当然是在R里模仿书中例题写一遍。

我感觉光看，不上手，效果还是不行的，必须得上手写程序，这样才扎实，使用的时候才能想起来。

缺点：书中的计量和统计部分还是初级的，很浅，当然这本书的目的就是学习R语言。

在R语言的学习方面，绘图讲的也是一点点。绘图我是结合另一本专门讲lattice的书学习的。

- 18、当时学R时竟然在图书馆看到本R的书，欣喜如狂
- 19、不是统计专业的读的很费劲！好书 值得再读2遍

## 《统计建模与R软件》

- 20、书籍很棒，没有破损，赞一个
- 21、以前自我感觉概率论与数理统计学得不错，回头看其实很多地方知其然不知其所以然，现在遇到问题就在其他地方检索一下，算是恶补了。
- 22、学习R的不二选择。。
- 23、不错，里面讲解很详细
- 24、我学R的入门书籍，推荐
- 25、很好的R教材
- 26、结构清晰，覆盖面广，适合入门。
- 27、统计学讲得比较多，R则涉及的不深。
- 28、中文书里边儿真心不错 第一章很难 因为数理统计不懂 之后的内容很实用 之后能用来当参考书
- 29、1刷，30%
- 30、所见最好的R入门cookbook
- 31、作图可以看看R in action，不过我觉得有基础的直接学下册就可以了，上册偏概率，6章之后才是计量
- 32、挺好的，帮同学买的，满意。
- 33、既回顾了统计学的知识，又学习了R的操作。
- 34、R建模方面很不错的基础读物。
- 35、浅显易懂
- 36、国内的R语言书也不少，这本是很经典的，值得细读
- 37、入门圣经
- 38、Statistical Modelling作业奋斗完毕，终于可以将这本书搁置到一边.....
- 39、断断续续看了一年 今天上午终于看完了
- 40、我觉得这本书是中文学习R语言的优秀教材。
- 41、本书为什么这么晦涩，就是因为作者在某些篇章中点到即至的做法，内容写了一堆，但是假设说明完全都没写。知道的人觉得本书太浅，不知道的人觉得本书太深，有些内容就是不能读。
- 42、书本很新，而且比其他地方便宜呢，很好的书~
- 43、书是免费接的，省了我不少钱。这次要好好练习
- 44、除了R软件详细用法之外，还介绍了各种统计分析的方法，有丰富实例，实在是难得的一本好书。
- 45、全面系统，理论和实践兼顾的初中级资料 案例丰富 适合有一点统计基础
- 46、基础中的基础
- 47、统计建模与R软件较好的结合了统计知识与R软件。由于是06年编写的，现在R软件已经更新到3.0，书中也有少量程序不能被调用，好在给出了原始代码，可以自己输入。综合来说，是一部很好的基础教材。很值得初学者阅读，并运行其中的程序。学习此书必会获益匪浅。
- 48、第一章其实可以不看的，如果学过概率的话。  
请问哪本lattice的书比较好啊，需要用R作图。
- 49、很多都抄袭R的入门手册，手法太低劣了，连变量名都一样
- 50、我一共买了两本R语言的书,这本的前一部份,实际上就是R开发工具上自带的R介绍PDF文件中的一部分翻译.适合初学者,更加偏重于统计学理论专业的人学习.对于销售统计分析方面都相对比较少,可能2006年写的吧,当前比较需要的应用,提及太少.对于了解R语言还是足够.书中应该简单讲些如何在不同MySQL版本中,进行数据抓取,然后分析.这方面非常薄弱.
- 51、给单位采购很不错经常来
- 52、花樣作死大比拼的一學期...
- 53、感慨一下本科时候学习能力真是弱爆了...
- 54、有点旧了。
- 55、R软件及统计相关知识的一本好书
- 56、纸张比较差，看起来像盗版
- 57、自学必备神书
- 58、从图书馆借来当帮助用，后面的PCA和因子分析好复杂，这部分原理可以看《漫画统计学之因子

## 《统计建模与R软件》

分析. (日) 高桥 信》

- 59、建议读下册，上册部分都是基本概统，下册才是真正的计量和建模。越来越喜欢R，体积小，速度快，界面也简洁，不像matlab和sas，看的眼花缭乱的
- 60、满基础的一本书，粗略的看下。
- 61、本书讲的主要是一些统计学方法，以及R语言的实现
- 62、偏统计的内容多了些，不过确实是一本好书来着。 ，
- 63、建议R基础比较好的，可以看看
- 64、全面，易读，基础
- 65、适合初学者吧我觉得但是蛮好懂的纸张的质量没想象中好啊
- 66、书中有大量案例，还有习题，很多的方法也有详细的解释，数学原理推导也有详细的说明，推荐好书。
- 67、数据分析（难道真要接这个offer了吗.....
- 68、屌丝的R入门选读。拿来做入门熟悉，真正使用的还是<Machine learning for hacker>
- 69、适合没有R基础的，最好结合《R in action》，也可以在coursera上面听课，多者结合最好。
- 70、这本书很不错，我用起来感觉很适用，就是有点简单
- 71、我真的觉得自己太水了.....OTLL
- 72、跟着视频，还是有点难.....
- 73、還沒看完呢 但覺得講的很詳細
- 74、看了开头就差不多了
- 75、难得一见的中国本土好教程

1、从统计学角度学R和从R角度学统计之前粗略的看了下《R语言实战》《R经典实例》等书，更多的是介绍R的命令，给人以很零碎的感觉，看完你基本上只能对R有个大概的认识，很难结合实际来使用它。这本书的好处是，统计的推导，详略得当，太深的略过，中低的都讲明白了，难得的是，一遍看一下，你基本都能明白，为什么要做这样那样的检验，他们的公式是怎样的，对应这样的公式，如果我要用R软件实现，应该如何写代码，基本每个例题的代码，都在R脚本文件中给出了，数据也是直接可以复制或者可以转化后使用的。针对前面提出来的关于其他书籍只是零碎的函数和命令的结合的问题，这本书是解决了的，有丰富的实例，也有相应的代码。还要感谢的一点是，以前一直不知道，如何用向量的形式，来写和解答各种数学式子，这本书上的很大代码让我有了认识（虽然我掌握的不好）。期待有更多的好书。

2、看完这本书后，个人总结的学习顺序建议：如果你的计量统计基础超扎实，直接看第二章看起，后面每一章直接做例题，R就学的差不多了。如果计量统计基础不扎实，先看一遍第二章，然后返回第一章，然后跳三章，顺序看下去，这一遍只看理论讲解部分，看完一遍后，再返回第二章，第三章，顺序看下去，这一遍重点做题，当然是在R里模仿书中例题写一遍。我感觉光看，不上手，效果还是不行的，必须得上手写程序，这样才扎实，使用的时候才能想起来。缺点：书中的计量和统计部分还是初级的，很浅，当然这本书的目的就是学习R语言。在R语言的学习方面，绘图讲的也是一点点。绘图我是结合另一本专门讲lattice的书学习的。

3、这本书是将教材课本中介绍的统计方法与计算机语言R进行结合，通过编程实现对统计学理论知识的理解，通过接口封装，进一步增加了其可用性。整本书的使用价值很高，通过理解其中的实例，结合自己的工作，你会有更深的心得体会。这本书就是两条线，一条线是统计学中的各种基础知识和方法，另一条线是R语言实现。然而这两条线如果放在过去，那么我一定会在一段相当长的时间段内不断分别沿着这两条线去琢磨。可以说，这两条线非常复杂，内容也非常多，按照这种读书方法最终收获的是“模糊的记忆”，没有什么系统性的感触，也没有与工作结合的认识。现在看来，我觉得一种比较好的阅读本书的方式是划分成如下几步。第一步，粗看目录，知道整本书的大体框架；第二步，看每一章的序或介绍，通过看序，知道每一章的核心要解决的问题是什么；第三步，看每一章的实例，知道解决实例的方法。【通过目录看书的框架】要了解统计学的基本知识，那就是chapter1；要了解R语言的语法，那就是chapter2；统计学与R的初步结合，基于描述性分析，那就是chapter3，其中用R画图也是本章；个人认为本书最大价值在于5-9章。假设推断 in chapter5；回归分析 in chapter6；方差分析 in chapter7；判别分析与聚类分析 in chapter8；主成分分析、因子分析、相关分析 in chapter9；ok，第一步达到了，本书的框架就是理论到实践的打通，由浅入深。【核心关键词的含义】你知道的分布有哪些？均匀分布、正态分布、泊松分布、卡方分布；统计分析：分为统计描述和统计推断两个部分。统计描述：是通过绘制图表、计算统计量等方法来描述数据的分布特点。是数据分析的基本步骤，也是记性统计推断的基础。统计推断：主要包含参数估计和假设检验两个主要内容。参数估计：通过样本来估计总体分布所包含的未知参数，叫做参数估计，是统计推断的一种重要形式。常见的方法有两种，一种称为点估计，另一种是区间估计。假设检验：通过样本数据对某个事先做出的假设按照某种设计好的方法进行检验，判断该假设是否正确。是统计推断的另一种表现形式。假设检验的参数方法：在统计推断问题中，如果假定或者给定了总体分布的具体形式（如正态分布），只是其中含有若干未知参数，需要基于总体的样本对参数做出估计或者进行某种假设校验，这种推断方法称为参数方法。假设检验的非参数方法：在许多实际问题中，都不知道总体是怎样的一种分布，很难对总体分布做出正确的假定，对于这种不假定总体分布的具体形式，尽量从样本本身来获取所需要的信息的统计方法称为非参数方法。回归分析：变量间的关系有两类，一类是完全确定的关系，可以用函数形式表示；还有一类是变量间有一定的关系，但是由于情况复杂或者存在不可避免的误差无法做到精确确定，为了能够研究这类变量之间的关系，常常通过大量试验或观测获取数据，用统计方法去寻找两者之间的关系，注意，这种关系反映的是变量间的统计规律。这种统计方法便是回归分析。方差分析：在实际工作中，影响一件事情的因素有很多，人们总是希望通过各种试验来观察各种因素对试验结果的影响。方差分析就是研究一种或多种维度的变化对试验结果指标是否有显著影响，从而找出较优的试验条件或者生产条件的一种常用统计方法。基本目标就是通过比较各种原因在总变异中的重要程度，对原因进行排序，从而实现下一步工作方向的确。判别分析：也就是机器学习中的有监督学

习，鸡仔已知有多少类的情况下，并且在有训练样本的前提下，利用训练样本对数据进行分类。是用于判断个体所属群体的一种统计方法。聚类分析：预先不知道有多少类别，很据某种规则将样本进行分类。是一类将数据所对应的研究对象进行分类的统计方法。主成分分析：PCA。是将多指标化为少数几个综合指标的一种统计分析方法。通过降维技术把多个变量化成少数几个主成分的方法。这些主成分能够反映原始变量的大部分信息，通常表现为原始变量的线性组合。因子分析：是PCA的推广和扩展，也是降维的一种方法，是一种用来分析隐藏在表面现象背后的因子作用的一类统计模型。相关分析：分析两组随机变量之间相关性程度的一种统计方法。

4、中文写R的书里，算是最好的了，因为本来就没几本。作者避免了过多的数理统计证明，更偏向应用。给出了很多的实例。内容比较简单，生物、经济等学科的学过简单统计学就基本能看了，不过这些读者用SPSS更多吧。当然有数理统计背景阅读起来更轻松。

5、在豆瓣读书里检索本书时，自己大吃一惊——本书25页？想想都不可能！买到手之后才知道是525页，请修改该书的简介~！之前想当然的觉得R会像SPSS，其实它更像SAS，但是比SAS更开放，而且是免费嵌入电脑系统来使用，这是它能渐渐风靡起来的原因吧！

## 章节试读

### 1、《统计建模与R软件》的笔记-第42页

下载地址：<http://cran.r-project.org/bin/windows/base/>

### 2、《统计建模与R软件》的笔记-第130页

老师把high-level,low-level翻译成高水平作图和低水平作图未免太差强人意，译成全景作图函数和要素作图函数怎么样？

\*事实上，R的所有作图函数分为两类，一类是高层函数（high-level），用以生成新的图形，另一类就是低层函数（low-level），这一类指的正是绘制图形元素的这些基础函数。

——谢益辉《现代统计图形》

### 3、《统计建模与R软件》的笔记-第63页

```
> labs<-paste("X",1:6,sep="");labs
[1] "X1" "X2" "X3" "X4" "X5" "X6"
> labs1<-paste("X",1:6,sep=" ");labs1
[1] "X 1" "X 2" "X 3" "X 4" "X 5" "X 6"
> labs1<-paste("X",1:6);labs1
[1] "X 1" "X 2" "X 3" "X 4" "X 5" "X 6"
```

不制定sep的数值，默认为空格。

### 4、《统计建模与R软件》的笔记-第203页

test of hypothesis

概念：

零假设、原假设（null hypothesis）

备择假设（alternative hypothesis）

单边检验VS双边检验

一尾VS二尾

假设检验的依据是样本。样本的某些取值可能对原假设 $H_0$ 有利，而另一些取值可能对 $H_0$ 不利，因此可以根据某种公认的合理准则将样本空间分成两部分。一部分是拒绝域（critical region），当样本落入拒绝域时，便拒绝 $H_0$ 另一部分可称为接受域（acceptance region）

重要的参数检验

Z, t

正态分布为例，给定z值之后，需要考虑 $X > |z|$ 和 $X < -|z|$ 的概率P值

```
P_value<-function(cdf, x, paramet=numeric(0), side=0){
n<-length(paramet)
P<-switch(n+1,
cdf(x),
cdf(x, paramet),
cdf(x, paramet[1], paramet[2]),
cdf(x, paramet[1], paramet[2], paramet[3]))
```

```
)  
if (side<0) P  
else if (side>0) 1-P  
else  
  if (P<1/2) 2*P  
  else 2*(1-P)  
}
```

cdf是分布函数，正态分布是pnorm，x是计算P值的给定值。paramet是对应的分布参数，如正太分布参数是paramet=c(mu, sigma)。side是计算单侧P值或者双侧P值的指标参数，side=-1，计算左侧的P值，输入side=1，计算右侧的P值，输入side=0或默认，计算两侧P值。函数输出值是相应的P值。在得到P值后，其检验标准改为：当P值小于指定的显著水平a时，则拒绝原假设，否则不拒绝。

## 5、《统计建模与R软件》的笔记-第69页

可以用attr(object,name)的形式存取对象object的名为name的属性，例如，

```
> attr(x,"names")  
[1] "apple" "orange"  
> attr(x,"name")  
[1] "apple" "orange"  
> attr(x,"nam")  
[1] "apple" "orange"  
> attr(x,"n")  
[1] "apple" "orange"
```

原因见笔者对另一本书《R语言及Bioconductor在基因组分析中的应用》所做的笔记：  
<http://book.douban.com/annotation/24010007/>

```
attr {base} R Documentation  
Object Attributes  
Description  
Get or set specific attributes of an object.  
Usage  
attr(x, which, exact = FALSE)  
attr(x, which) &lt;- value  
Arguments  
x  
an object whose attributes are to be accessed.  
which  
a non-empty character string specifying which attribute is to be accessed.  
exact  
logical: should which be matched exactly?  
value  
an object, the new value of the attribute, or NULL to remove the attribute.  
这点在R文档中得到了验证。
```

```
> attr(x,"abc")&lt;-"test";x  
apple grapes  
 2.5 2.1  
attr(,"type")
```

```
[1] "fruit"
attr(,"abc")
[1] "test"
> attr(x,"abc")
[1] "test"
> attr(x,"a")
[1] "test"
> attr(x,"abc",exact=TRUE)
[1] "test"
> attr(x,"a",exact=TRUE)
NULL
```

## 6、《统计建模与R软件》的笔记-第284页

## 7、《统计建模与R软件》的笔记-第63页

```
> paste(c('a','b'),collapse='.')
[1] "a.b"
```

`paste {base}` R Documentation  
Concatenate Strings  
Description  
Concatenate vectors after converting to character.  
Usage  
`paste (... , sep = " ", collapse = NULL)`  
`paste0(... , collapse = NULL)`  
Arguments  
...  
one or more R objects, to be converted to character vectors.  
`sep`  
a character string to separate the terms. Not `NA_character_`.  
`collapse`  
an optional character string to separate the results. Not `NA_character_`.  
Details  
`paste` converts its arguments (via `as.character`) to character strings, and concatenates them (separating them by the string given by `sep`). If the arguments are vectors, they are concatenated term-by-term to give a character vector result. Vector arguments are recycled as needed, with zero-length arguments being recycled to `""`.  
Note that `paste()` coerces `NA_character_`, the character missing value, to `"NA"` which may seem undesirable, e.g., when pasting two character vectors, or very desirable, e.g. in `paste("the value of p is ", p)`.  
`paste0(... , collapse)` is equivalent to `paste(... , sep = "", collapse)`, slightly more efficiently.  
If a value is specified for `collapse`, the values in the result are then concatenated into a single string, with the elements being separated by the value of `collapse`.

```
> paste(1:10) #same as as.character(1:10)
[1] "1" "2" "3" "4" "5" "6" "7" "8" "9" "10"> paste(1:10,collapse='s')
[1] "1s2s3s4s5s6s7s8s9s10"
> paste(1:10,collapse="s")
```

```
[1] "1s2s3s4s5s6s7s8s9s10"
```

注意不同。书中的用法，1-10分别作为一个字符串，一共10个。而加上collapse，则连成了一个长字符串。

以下的代码和图片或许解释更清楚（通过RStudio）：

```
> x<-paste(1:10);x
```

```
[1] "1" "2" "3" "4" "5" "6" "7" "8" "9" "10"
```

```
> y<-paste(1:10,collapse='s');y
```

```
[1] "1s2s3s4s5s6s7s8s9s10"书中的用法，1-10分别作为一个字符串，一共10个。而加上collapse，则连成了一个长字符串。
```

根据R文档，paste(1:10)被转换为字符；而paste(1:10,collapse='s')则把结果里所有字符串连接起来成一个大字符串，用s来间隔。而以下结果为之所以没有被s连接，因为本身"1 2 3 4 5"就是唯一的结果字符串：

```
> x1<-paste(1,2,3,4,5,collapse='s');x1
```

```
[1] "1 2 3 4 5"
```

```
> x2<-paste("1","2","3","4","5",collapse='s');x2
```

```
[1] "1 2 3 4 5"
```

## 8、《统计建模与R软件》的笔记-第五章

mark一下，终于看完第五章了。

数理统计的底子太弱，看各种检验真是累死人了

## 9、《统计建模与R软件》的笔记-第90页

```
> rt<-read.table("R_YiXue/Chapter02/houses.data")
```

```
> rt
```

```
Price Floor Area Rooms Age Cent.heat
```

```
01 52.00 111 830 5 6.2 no
```

```
02 54.75 128 710 5 7.5 no
```

```
03 57.50 101 1000 5 4.2 no
```

```
04 57.50 131 690 6 8.8 no
```

```
05 59.75 93 900 5 1.9 yes
```

```
> is.data.frame(rt)
```

```
[1] TRUE
```

```
> rt2<-read.table("R_YiXue/Chapter02/houses2.data")
```

```
> rt2
```

```
V1 V2 V3 V4 V5 V6
```

```
1 Price Floor Area Rooms Age Cent.heat
```

```
2 52.00 111.0 830 5 6.2 no
```

```
3 54.75 128.0 710 5 7.5 no
```

```
4 57.50 101.0 1000 5 4.2 no
```

```
5 57.50 131.0 690 6 8.8 no
```

```
6 59.75 93.0 900 5 1.9 yes
```

```
> rt2<-read.table("R_YiXue/Chapter02/houses2.data",header=TRUE)
```

```
> rt2
```

```
Price Floor Area Rooms Age Cent.heat
```

```
1 52.00 111 830 5 6.2 no
```

```
2 54.75 128 710 5 7.5 no
```

```
3 57.50 101 1000 5 4.2 no
```

```
4 57.50 131 690 6 8.8 no
5 59.75 93 900 5 1.9 yes
> read.table("R_YiXue/Chapter02/houses2.data",skip=1)
Warning message:
In read.table("R_YiXue/Chapter02/houses2.data", skip = 1) :
  incomplete final line found by readTableHeader on 'R_YiXue/Chapter02/houses2.data'
> read.table("R_YiXue/Chapter02/houses2.data",header=TRUE,skip=1)
Warning message:
In read.table("R_YiXue/Chapter02/houses2.data", header = TRUE, skip = 1) :
  incomplete final line found by readTableHeader on 'R_YiXue/Chapter02/houses2.data'
> rt3
  V1 V2 V3 V4 V5 V6
1 52.00 111 830 5 6.2 no
2 54.75 128 710 5 7.5 no
3 57.50 101 1000 5 4.2 no
4 57.50 131 690 6 8.8 no
5 59.75 93 900 5 1.9 yes
> rt4
 X52.00 X111.0 X830 X5 X6.2 no
1 54.75 128 710 5 7.5 no
2 57.50 101 1000 5 4.2 no
3 57.50 131 690 6 8.8 no
4 59.75 93 900 5 1.9 yes
```

## 10、《统计建模与R软件》的笔记-第五章 假设检验

### 最后一道题5.18

```
wilcox.test(x,y,paired=TRUE,exact=FALSE) (Wilcox符号秩检验)
p-value = 0.001586 < 0.05 , 应该拒绝H0
```

```
wilcox.test(x,y,paired=FALSE,exact=FALSE) (Wilcox秩和检验)
p-value = 0.05509 > 0.05. 应该保留H0
```

差别怎么这么大？

## 11、《统计建模与R软件》的笔记-第84页

```
5.数组的广义转置> A=array(1:24,dim=c(2,3,4))
> B=aperm(A,c(2,3,1))
B[i,j,k]=A[j,k,i]
我觉得不对。应该是A[j,k,i]=B[k,i,j]
```

## 12、《统计建模与R软件》的笔记-第61页

```
> seq(2,10)
[1] 2 3 4 5 6 7 8 9 10
> seq(2.10)
[1] 1 2
```

即若seq()函数中只有一个参数，默认为是to的值。得到从1到此参数数值的间隔为1的序列。

## 13、《统计建模与R软件》的笔记-第192页

矩法 (method of moments) 是由英国统计学家K Pearson在20世纪初提出来的，它的中心思想就是用样本矩去估计总体矩。要是不知道moment在力学里是力矩的意思，估计该觉得这词又文艺了一把...

## 14、《统计建模与R软件》的笔记-第80页

### 2.1R软件使用

#是说明语句字符，后面的语句是说明语句，增加程序可读性；  
&lt;-表示赋值，c()表示数组，x&lt;-c(...)表示将一组数据赋给变量X，  
平均数：mean(x)，标准差sd(x)；散点图plot(x)；直方图hist(x)；

阅读文件：

```
rt<-read.table("exam0203.txt", head=TRUE);rt  
head=TRUE表示第一行是表头，否则(FALSE)文件第一行作为数据处理；
```

载入文件：

单击“载入”；或者<code>load("myworkspace.RData")

### 2.2数字、字符与向量

1.赋值：<code>x<-c(.....)</code>；  
<code>assign("x",c(.....))</code>；  
<code>y<-c(x,0,x)</code> 定义变量y有11个分量，其中两边是变量x,中间是0

### 2.运算

加+，减-，乘\*，除/，乘方^  
%%表示整数除法，例5%%3=1；%%表示求余数，例5%%3=2  
基本初等函数：log, exp, cos, tan, sqrt  
函数：min(), max(), range()  
which.min(), which.max()表示第几个是最小\大值  
求和：sum()，乘积：prod()  
向量个数：length()  
中位数：median()；均值：mean()；方差var() 标准差 sd()  
顺序统计量（递增）：sort()

### 3.序列

等差数列：a:b表示从a开始，逐项加1（或减1），直到b  
<code>x<-1:30</code>表示向量（1,2,。。。30）  
等差运算优于乘法运算，1:n-1表示向量1:n-1  
等间隔函数：seq(X1, X2, by=X3), 从X1到X2，间隔X3.不特别注明by，间隔为1  
or: s<-seq(length=x1, from=x2, by=x3)  
重复函数：s<-rep(x, times=3)

### 4.逻辑向量

运算符：&lt;, &lt;=, &gt;, &gt;=, ==, !=（不等于）  
且"&"; 或"|", 非"!"

```
判断 : >all(c(1,2,3,4,5)>3)
[1] FALSE
```

5. 缺失数据NA。。。。待补全

15、《统计建模与R软件》的笔记-第62页

```
> x<-c(0/1,0/0,1/0,NA);x
[1] 0 NaN Inf NA
> is.na(x)
[1] FALSE TRUE FALSE TRUE
> is.nan(x)
[1] FALSE TRUE FALSE FALSE
> is.finite(x)
[1] TRUE FALSE FALSE FALSE
> is.infinite(x)
[1] FALSE FALSE TRUE FALSE
```

数据总体来说分为两类，一类是确定数据，一类是不确定数据。后者标记为NaN。

确定数据可以分为两类，一类是缺失数据，一类是非缺失数据。前者标记为NA。不确定数据也可以认为是缺失数据。

非缺失数据又可氛围两类，一类是有限数据，记为finite；一类是无穷数据，记为infinite。

16、《统计建模与R软件》的笔记-第284页

异常样本的存在往往会给回归模型带来不稳定，为此，人们必须提出所谓回归诊断的问题（regression diagnostic），其主要内容有：  
误差项是否满足独立性、等方差性、正态性  
选择线性模型是否合适  
是否存在异常样本  
回归分析的结果是否对某些样本的依赖过重，即回归模型是否剧本稳定性  
自变量之间是否存在高度相关，即是否有多重共线性问题存在

17、《统计建模与R软件》的笔记-第66页

```
> v[-(1:5)]
[1] 15 16 17 18 19 20
表示扣除相应的元素
> v
[1] 10 11 12 13 14 15 16 17 18 19 20
> v[c(-1,-2,-3,-4,-5)]
[1] 15 16 17 18 19 20
> v[-1]
[1] 11 12 13 14 15 16 17 18 19 20
> v[-2]
[1] 10 12 13 14 15 16 17 18 19 20
```

> v[-5]

```
[1] 10 11 12 13 15 16 17 18 19 20
```

-n为扣除第n个元素

## 18、《统计建模与R软件》的笔记-第90页

### 2.5多维数组和矩阵

定义数组：>z<-1:12

```
>dim(z)<-c(3,4)
```

```
>z
```

```
  [,1] [,2] [,3] [,4]
```

```
[1,]  1  4  7 10
```

```
[2,]  2  5  8 11
```

```
[3,]  3  6  9 12
```

dim 是各维的长度

多维数组：>x<-array(1:20, dim=c(4,5)) 4 × 5的二维数组（矩阵）

```
>x
```

构造矩阵：>A<-matrix(1:15, nrow=3, ncol=5, byrow=TRUE)

```
>A
```

3 × 5的矩阵,byrow=TRUE去掉，按列置放

```
>A<-matrix(c(1,2,2,3,3,4),nrow=2)
```

访问某个元素：>a[1,2:3,2:3]表示第一下标为1，第二、三下标为2或3的元素

数组进行四则运算，形状相同；

形状不同：短向量的数据循环使用

转置运算：>t(A)

求行列式：det()

向量内积：x%\*%y

crossprod(x,y) 内积运算函数，表交叉乘积

向量外积：>x%o%y; outer(x,y)

乘法：A\*B，对应元素乘积（相同维度）

A%\*%B 两个矩阵乘积（A的列数等于B的行数）

对角：diag(v),v为向量，表示以v为对角线元素的对角阵

diag(M),M为矩阵，表示对角线上的元素

方程组：Ax=b

```
x<-solve(A);x
```

```
逆矩阵 B<-solve(A);B
```

# 《统计建模与R软件》

## 版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:[www.tushu111.com](http://www.tushu111.com)