

```
t.test(y1, y2, paired=TRUE)
```

其中的 y_1 和 y_2 为两个非独立组的数值向量。结果如下：

```
> library(MASS)
> sapply(UScrime[c("U1", "U2")], function(x) (c(mean=mean(x), sd=sd(x))))
      U1      U2
mean 95.5 33.98
sd   18.0  8.45

> with(UScrime, t.test(U1, U2, paired=TRUE))

      Paired t-test

data:  U1 and U2
t = 32.4066, df = 46, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 57.67003 65.30870
sample estimates:
mean of the differences
      61.48936
```

差异的均值（61.5）足够大，可以保证拒绝年长和年轻男性的平均失业率相同的假设。年轻男性的失业率更高。事实上，若总体均值相等，获取一个差异如此大的样本的概率小于0.000 000 000 000 000 22（即 $2.2e-16$ ）。

7.4.3 多于两组的情况

如果想在多于两个的组之间进行比较，应该怎么做？如果能够假设数据是从正态总体中独立抽样而得的，那么你可以使用方差分析（ANOVA）。ANOVA是一套覆盖了许多实验设计和准实验设计的综合方法。就这一点来说，它的内容值得单列一章。你可以随时离开本节转而阅读第9章。

7.5 组间差异的非参数检验

如果数据无法满足t检验或ANOVA的参数假设，可以转而使用非参数方法。举例来说，若结果变量在本质上就严重偏倚或呈现有序关系，那么你可能会希望使用本节中的方法。

7.5.1 两组的比较

若两组数据独立，可以使用Wilcoxon秩和检验（更广为人知的名字是Mann - Whitney U检验）来评估观测是否是从相同的概率分布中抽得的（即，在一个总体中获得更高得分的概率是否比另一个总体要大）。调用格式为：

```
wilcox.test(y ~ x, data)
```

其中的 y 是数值型变量，而 x 是一个二分变量。调用格式或为：

```
wilcox.test(y1, y2)
```

其中的 y_1 和 y_2 为各组的结果变量。可选参数`data`的取值为一个包含了这些变量的矩阵或数据框。默认进行一个双侧检验。你可以添加参数`exact`来进行精确检验，指定`alternative="less"`或`alternative="greater"`进行有方向的检验。

如果你使用Mann - Whitney U检验回答上一节中关于监禁率的问题，将得到这些结果：

```
> with(UScrime, by(Prob, So, median))

So: 0
[1] 0.0382
-----
So: 1
[1] 0.0556

> wilcox.test(Prob ~ So, data=UScrime)

Wilcoxon rank sum test

data: Prob by So
W = 81, p-value = 8.488e-05
alternative hypothesis: true location shift is not equal to 0
```

你可以再次拒绝南方各州和非南方各州监禁率相同的假设 ($p < 0.001$)。

Wilcoxon符号秩检验是非独立样本t检验的一种非参数替代方法。它适用于两组成对数据和无法保证正态性假设的情境。调用格式与Mann - Whitney U检验完全相同，不过还可以添加参数`paired=TRUE`。让我们用它解答上一节中的失业率问题：

```
> sapply(UScrime[c("U1", "U2")], median)
U1 U2
92 34

> with(UScrime, wilcox.test(U1, U2, paired=TRUE))

Wilcoxon signed rank test with continuity correction

data: U1 and U2
V = 1128, p-value = 2.464e-09
alternative hypothesis: true location shift is not equal to 0
```

你再次得到了与配对t检验相同的结论。

在本例中，含参的t检验和与其作用相同的非参数检验得到了相同的结论。当t检验的假设合理时，参数检验的功效更强（更容易发现存在的差异）。而非参数检验在假设非常不合理时（如对于等级有序数据）更适用。

7.5.2 多于两组的比较

在要比较的组数多于两个时，必须转而寻求其他方法。考虑7.4节中的`state.x77`数据集。它包含了美国各州的人口、收入、文盲率、预期寿命、谋杀率和高中毕业率数据。如果你想比较美国四个地区（东北部、南部、中北部和西部）的文盲率，应该怎么做呢？这称为单向设计（one-way

design)，我们可以使用参数或非参数的方法来解决这个问题。

如果无法满足ANOVA设计的假设，那么可以使用非参数方法来评估组间的差异。如果各组独立，则Kruskal-Wallis检验将是一种实用的方法。如果各组不独立（如重复测量设计或随机区组设计），那么Friedman检验会更合适。

Kruskal - Wallis检验的调用格式为：

```
kruskal.test(y ~ A, data)
```

其中的 y 是一个数值型结果变量， A 是一个拥有两个或更多水平的分组变量（grouping variable）。（若有两个水平，则它与Mann - Whitney U检验等价。）而Friedman检验的调用格式为：

```
friedman.test(y ~ A | B, data)
```

其中的 y 是数值型结果变量， A 是一个分组变量，而 B 是一个用以认定匹配观测的区组变量（blocking variable）。在以上两例中， $data$ 皆为可选参数，它指定了包含这些变量的矩阵或数据框。

让我们利用Kruskal - Wallis检验回答文盲率的问题。首先，你必须将地区的名称添加到数据集中。这些信息包含在随R基础安装分发的state.region数据集中。

```
states <- as.data.frame(cbind(state.region, state.x77))
```

现在就可以进行检验了：

```
> kruskal.test(Illiteracy ~ state.region, data=states)
```

```
Kruskal-Wallis rank sum test
```

```
data: states$Illiteracy by states$state.region
```

```
Kruskal-Wallis chi-squared = 22.7, df = 3, p-value = 4.726e-05
```

显著性检验的结果意味着美国四个地区的文盲率各不相同（ $p < 0.001$ ）。

虽然你可以拒绝不存在差异的原假设，但这个检验并没有告诉你哪些地区显著地与其他地区不同。要回答这个问题，你可以使用Mann - Whitney U检验每次比较两组数据。一种更为优雅的方法是在控制犯第一类错误的概率（发现一个事实上并不存在的差异的概率）的前提下，执行可以同步进行的多组比较，这样可以直接完成所有组之间的成对比较。npmc包提供了所需要的非参数多组比较程序。

说实话，我将本章标题中基本的定义拓展了不止一点点，但由于在这里讲非常合适，所以希望你能够容忍我的做法。第一步，请先安装npmc包。此包中的npmc()函数接受的输入为一个两列的数据框，其中一列名为var（因变量），另一列名为class（分组变量）。代码清单7-20中包含了可以用来完成计算的代码。

代码清单7-20 非参数多组比较

```
> class <- state.region
> var <- state.x77[,c("Illiteracy")]
> mydata <- as.data.frame(cbind(class, var))
> rm(class, var)
> library(npmc)
> summary(npmc(mydata), type="BF")
```

```
$'Data-structure'
      group.index  class.level nobs
Northeast          1    Northeast   9
South              2      South    16
North Central      3 North Central  12
West               4        West   13

$'Results of the multiple Behrens-Fisher-Test'
      cmp effect lower.cl upper.cl p.value.1s p.value.2s
1 1-2 0.8750 0.66149 1.0885 0.000665 0.00135
2 1-3 0.1898 -0.13797 0.5176 0.999999 0.06547
3 1-4 0.3974 -0.00554 0.8004 0.998030 0.92004
4 2-3 0.0104 -0.02060 0.0414 1.000000 0.00000
5 2-4 0.1875 -0.07923 0.4542 1.000000 0.02113
6 3-4 0.5641 0.18740 0.9408 0.797198 0.98430

> aggregate(mydata, by=list(mydata$class), median)
      Group.1 class  var
1          1     1 1.10
2          2     2 1.75
3          3     3 0.70
4          4     4 0.60
```

① 成对组间比较结果

② 各类别的文盲率中间值

调用了npmc的语句生成了六对统计比较结果（东北部对南部、东北部对中北部、东北部对西部、南部对中北部、南部对西部，以及中北部对西部）**①**。可以从双侧的p值（p.value.2s）看出南部与其他三个地区显著不同，而其他三个地区之间并没有什么不同。在**②**处可以看到南部的文盲率中间值更高。注意，npmc在计算积分时使用了随机数，所以每次计算的结果会有轻微的不同。

7.6 组间差异的可视化

在7.4节和7.5节中，我们关注了进行组间比较的统计方法。使用视觉直观地检查组间差异，同样是全面的数据分析策略中的一个重要组成部分。它允许你评估差异的量级、甄别出任何会影响结果的分布特征（如偏倚、双峰或离群点）并衡量检验假定的合理程度。R中提供了许多比较组间数据的图形方法，其中包括6.5节中讲解的箱线图（简单箱线图、含凹槽的箱线图、小提琴图）、6.4.1节中叠加的核密度图，以及在第9章中讨论的评估检验假定的图形方法。

7.7 小结

在本章中，我们评述了R中用于生成统计概要和进行假设检验的函数。我们关注了样本统计量和频数表、独立性检验和类别型变量的相关性度量、定量变量的相关系数（和连带的显著性检验）以及两组或更多组定量结果变量的比较。

下一章中，我们将探索一元回归和多元回归，讨论的焦点在于如何理解一个预测变量（一元回归）或多个预测变量（多元回归）与某个被预测变量或效标变量（criterion variable）之间的关系。图形将有助于诊断潜在的问题、评估和提高模型的拟合精度，并发现数据中意料之外的信息瑰宝。