M7 Aufgabenblatt: Tests

Zehui Bai

Stand: 2022-01-19 23:28 Uhr

Table of Contents

# Aufgabe 1

Angenommen Sie wollen eine Alternative zu einem t-Test für den folgenden Datensatz, um auf Gleichheit der Mittelwerte zwischen den Gruppen A und B zu testen. Wie könnten Sie einen Permutationstest als Alternative konstruieren? Wieviele mögliche Permutationen gibt es hier? (5 Punkte)

Permutation test 置换检验是一种基于大量计算利用样本数据的全排列，进行统计推断的方法，

其对总体分布自由，应用较为广泛，特别适用于总体分布未知的小样本资料，以及某些难以用常规方法分析资料的假设检验问题。在具体使用上它和Bootstrap Methods类似，通过对样本进行顺序上的置换，重新计算统计检验量，构造经验分布，然后在此基础上求出P-value进行推断。

Permutation test步骤：

* A组（共4个数据）：4 8 2 6
* B组（共6个数据）：8 13 7 8 12 15

1. 计算给定数据的均值差 mean(Xa)-mean(Xb) = (4+8+2+6)/4 - (8+13+7+8+12+15)/6 = -5.5
2. 将A、B两组数据合并到一个集合中，从中挑选出4个作为A组的数据（X’a），剩下的作为B组的数据（X’b）。因此 Wieviele mögliche Permutationen gibt es hier? （可以理解为从10个数据里面选4个的可能性 C(10,4)=210）
3. 计算并记录第一步中A组同B组的均值之差。mean(X’a)-mean(X’b)
4. 对前两步重复n次（重复次数越多越稳定），比如n为1000
5. 最后我们得到1000个均值差，然后我们就可以查看观察到的-5.5在这1000个数据中的位置, 在permutation得到的抽样总体中大于-5.5的数值有x个，所以估计的P-value是x/n

更详细的参看课件31页

# Aufgabe 2

Was sind die Vor- und Nachteile eines (exakten) Permutationstests? Wodurch unterscheidet sich ein Random Permutation Test? (4 Punkte)

**Wodurch unterscheidet sich ein Random Permutation Test?** 参见课件38页

当数据的可能排序太多而无法以方便的方式进行完整枚举时，可以创建渐近等效置换测试。这是通过蒙特卡洛抽样生成参考分布来完成的，该抽样对可能的重复样本进行少量（相对于排列总数）随机样本。

Vorteile

1. 任何检验统计量都存在置换检验，无论其分布是否已知。
2. 置换检验是一种非参数估计，是分析不满足传统参数检验（例如 t 检验、方差分析）的统计假设的量化数据的理想选择。
3. 排序置换法计算的 p值都是精确不需要近似的

Nachteile

1. 可置换原则必须得到满足；
2. Assumption required, cause tests of difference in location (like a permutation t-test) require equal variance under the normality assumption.
3. Computation burden 当数据量大时，计算过程将会很缓慢

# Aufgabe 3

Testen Sie für den folgenden Datensatz Exposition und Erkrankung auf Unabhängigkeit

|  |  |  |  |
| --- | --- | --- | --- |
|  | krank | gesund | gesamt |
| Exponiert | 3 | 1 | 4 |
| Nicht exponiert | 1 | 4 | 5 |
| gesamt | 4 | 5 | 9 |

## Chi-Quadrat Test

### Generate datasets  
Testdata <- data.frame(Krank=c( rep(c("krank"),4), rep(c("gesund"),5) ) ,  
 Exponiert = c("Exponiert","Exponiert","Exponiert",  
 "Nicht Exponiert", "Exponiert", "Nicht Exponiert",   
 "Nicht Exponiert", "Nicht Exponiert", "Nicht Exponiert"))  
xtabs(~ Exponiert + Krank, data=Testdata)

## Krank  
## Exponiert gesund krank  
## Exponiert 1 3  
## Nicht Exponiert 4 1

chisq.test(Testdata$Exponiert, Testdata$Krank)

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: Testdata$Exponiert and Testdata$Krank  
## X-squared = 0.95063, df = 1, p-value = 0.3296

## fisher exact test

由于样本量太小，卡方检验渐近无法满足，因此可以执行fisher exact test

fisher.test(Testdata$Exponiert, Testdata$Krank)

##   
## Fisher's Exact Test for Count Data  
##   
## data: Testdata$Exponiert and Testdata$Krank  
## p-value = 0.2063  
## alternative hypothesis: true odds ratio is not equal to 1  
## 95 percent confidence interval:  
## 0.001288082 3.255946155  
## sample estimates:  
## odds ratio   
## 0.1196876

## Random Permutation Test

library(coin)  
## Random permutation test  
chisq\_test(table(Testdata), distribution = approximate(nresample=1000))

##   
## Approximative Pearson Chi-Squared Test  
##   
## data: Exponiert by Krank (gesund, krank)  
## chi-squared = 2.7225, p-value = 0.203

卡方检验的参数假设不满足，但是小样本量时候fisher exact test 也可以提供robust的结果，此外 Random Permutation Test不依赖参数假设，结果和fisher exact test接近，也是很好的选择