一、为研究某种新药对抗凝血酶活力的影响，随机安排新药组病人12例，对照组病人10例，分别测定器抗凝血酶活力（单位：），其结果如下：

新药组：126 125 138 128 123 138 142 116 110 108 113 140

对照组：160 175 177 170 175 153 168 159 160 162

试分析新药组和对照组病人的抗凝血酶活力有无差别（α = 0.05）

（1）检验两组样本方差是否相同。 (15’)

（2）选择最合适的检验方法检验新药组和对照组病人的抗凝血酶活力有无差别。(15’)

Answer:

1. ：对照组和实验组两组样本方差相同

：对照组和实验组两组样本方差不相同

>a<-c(126,125,138,128,123,138,142,116,110,108,113,140)

>b<-c(160,175,177,170,175,153,168,159,160,162)

>var.test(a,b)

F test to compare two variances

data: a and b

F = 2.1512, num df = 11, denom df = 9, p-value = 0.26

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.5498769 7.7181417

sample estimates:

ratio of variances

2.151159

1. value=0.26>0.05,所以对照组和实验组方差相同。
2. ：对照组和实验组两组样本均值相同

：对照组和实验组两组样本均值不相同

使用双样本t.test

> t.test(a,b,var.equal = TRUE)

Two Sample t-test

data: a and b

t = -8.9578, df = 20, p-value = 1.947e-08

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-49.70504 -30.92829

sample estimates:

mean of x mean of y

125.5833 165.9000

P-value<0.05,对照组和新药组病人的抗凝血酶活力有差别

二、对7位健康成年人的血液测量其中的尿酸浓度，分别用手工（X）和仪器 （Y）两种方法测量，结果如下表所示，请用wilcoxon signed-rank test来检测两种测量方法的精度是否存在差异? （α = 0.05）(20’)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 手工（X） | 4.5 | 6.5 | 6 | 9.2 | 10 | 12 | 8.3 |
| 仪器（Y） | 4 | 7.2 | 8 | 14 | 8.8 | 10 | 11.5 |

答案：

R代码如下：

> X <- c(4.5,6.5,6,9.2,10,12,8.3)

> Y <- c(4,7.2,8,14,8.8,10,11.5)

> wilcox.test(X,Y,paired = TRUE,exact = FALSE)

Wilcoxon signed rank test with continuity correction

data: X and Y

V = 8.5, p-value = 0.3972

alternative hypothesis: true location shift is not equal to 0

因为p>0.05,所以两种测量方法不存在差异。

三、在某保险种类中，一次关于2018年的索赔数额（单位：元）的随机抽样为（按升幂排列）：

4152，4579，5053，5112，5745，6250，7081，9048，

12095，14430，17220，20610，22836，48950，67200

已知2017年的索赔数额的中位数为7520元。问2018年索赔的中位数与前一年是否有所变化？（α = 0.05）(15’)

Hint: You can use wilcox.test

答案：

> insure <- c(4152,4579,5053,5112,5745,6250,7081,9048,12095,14430,17220,20610,22836,48950,67200)

> wilcox.test(insure,mu = 7520,conf.int = TRUE)

Wilcoxon signed rank test

data: insure

V = 87, p-value = 0.1354

alternative hypothesis: true location is not equal to 7520

95 percent confidence interval:

6413 27031

sample estimates:

(pseudo)median

12265.75

因为p值=0.1354 >0.05, 故拒绝原假设, 认为2018年索赔的中位数与前一年没有变化.

四、Type 1 diabetes is a multigenic disease caused by T-cell mediated destruction of the insulin producing β-cells. Although conventional (targeted) approaches of identifying causative genes have advanced our knowledge of this disease, many questions remain unanswered.

Here we have a gene data from NOD mouse after(case) and before(control) treatment. The data can be found in "Data.txt”.Use the information mentioned above to answer the following questions:

1. use paired t-test to find genes which have significant expression (p<0.05) between case and control sample. Give the number of differential expressed genes and give the names of top 10 significantly differential expression genes. hint: “apply(data,1,function(x){…})” can apply function to every row in data more quickly than “for{}”, “names()” or “rownames()” can be used to extract names of differentially expressed genes
2. Adjust the p-values in question a) with bonferroni and FDR method to find differentially expressed genes in stringent way( list the differentially expressed gene names and the adjusted p-value).

Hint: you can do the adjustment according to the fomular, or use “p.adjust()” instead.

Answer:

评分细则：(1) read.table 步骤4分；apply(t.test) 步骤10分 ；提取差异基因名字步骤 6分

1. p.adjust(‘bonferroni’) 5分；p.adjust(‘bonferroni’) 5分；第三步骤找基因 5分

(1)利用两样本成对t.test

> Data<-read.table("第四次作业Data.txt")

> Data<-read.table("第四次作业Data.txt", header = TRUE)

> p.value<-apply(Data,1,function(x)t.test(x[1:10],x[11:20],paired = TRUE)$p.value)

> DEG.pair<-rownames(Data)[p.value<0.05]

> sum(p.value<0.05)

[1] 2296

> names(p.value) = rownames(Data)

> names(sort(p.value)[1:10])

1. "28801" "27868" "27438" "21642" "24019" "12323" "12962" "28939" "2387" "18712"

一共有2296个差异基因

(2)

> p.bon<-p.adjust(p.value,'bonferroni')

> p.fdr<-p.adjust(p.value,'fdr')

> rownames(Data)[p.bon<0.05]

character(0)

> rownames(Data)[p.fdr<0.05]

character(0)

由上可知，没有差异基因