

# 第七次上机

生信 2001 张子栋 2020317210101

GitHub 地址: [MarkdownNotes/R at main · Bluuur/MarkdownNotes \(github.com\)](https://github.com/Bluuur/MarkdownNotes)

1.(1)

```
1 myData <- read.table('/home/ubuntu/R_course/hg19_gene_table.txt', header =  
  T)  
2 idx <- grep('_', myData$chrom)  
3 myData <- myData[-idx,]  
4 table <- table(myData$chrom, myData$strand)  
5 table <- as.data.frame.matrix(table)  
6 # (1) 对不同 strand 基因在各染色体上的比例是否相同做统计检验  
7 # 相同, p-value < 0.05  
8 chisq.test(table)  
9  
10 # (2) 对不同 strand 基因在 chr4, chr5 上的比例是否相同做统计检验  
11 # 不相同 p-value > 0.05  
12 chisq.test(table[17:18,])  
13  
14 # (3) 计算基因长度 glen  
15 glen <- myData$txEnd - myData$txStart + 1  
16 # (3) (a)  
17 lg1 <- log(glen)  
18 # (3) (b)  
19 lg1[lg1 < 5] <- NA  
20 pairwise.t.test(lg1, myData$chrom, p.adjust.method = 'fdr')  
21 # (3) (c)  
22 # 不是来自相同分布  
23 wilcox.test(glen[17:18])
```

```
1 warning message in chisq.test(table):  
2 "Chi-squared approximation may be incorrect"
```

```
1      Pearson's Chi-squared test  
2  
3 data:  table  
4 x-squared = 110.35, df = 24, p-value = 4.891e-13
```

```
1      Pearson's Chi-squared test with Yates' continuity correction  
2  
3 data:  table[17:18, ]  
4 x-squared = 0.96437, df = 1, p-value = 0.3261
```

```

1      Pairwise comparisons using t tests with pooled SD
2
3 data:  lg1 and myData$chrom
4
5      chr1      chr10      chr11      chr12      chr13      chr14      chr15      chr16      chr17
6 chr10 < 2e-16 -          -          -          -          -          -          -          -
7 chr11 7.6e-12 < 2e-16 -          -          -          -          -          -          -
8 chr12 4.8e-05 < 2e-16 < 2e-16 -          -          -          -          -
9 chr13 1.0e-15 0.00972 < 2e-16 1.8e-06 -          -          -          -          -
10 chr14 0.00583 < 2e-16 2.1e-14 0.57480 1.2e-06 -          -          -          -
11 chr15 < 2e-16 1.2e-06 < 2e-16 4.4e-05 0.19618 3.0e-05 -          -          -
12 chr16 < 2e-16 < 2e-16 0.01927 < 2e-16 < 2e-16 < 2e-16 < 2e-16 -          -
13 chr17 < 2e-16 < 2e-16 0.00022 < 2e-16 < 2e-16 < 2e-16 < 2e-16 0.29820 -
14 chr18 < 2e-16 0.00610 < 2e-16 < 2e-16 1.1e-05 < 2e-16 7.4e-10 < 2e-16 < 2e-
15 chr19 < 2e-16 < 2e-16 5.4e-12 < 2e-16 < 2e-16 < 2e-16 < 2e-16 0.00014
0.00318
16 chr2  < 2e-16 0.12948 < 2e-16 < 2e-16 0.11045 < 2e-16 0.00015 < 2e-16 < 2e-
17 chr20 0.67794 < 2e-16 7.2e-06 0.00076 1.6e-12 0.01126 1.3e-11 5.4e-10 5.3e-
18 chr21 0.03828 9.7e-12 1.0e-08 0.75448 0.00011 0.92047 0.00193 1.7e-12 1.1e-
19 chr22 0.09513 < 2e-16 0.00480 1.8e-05 1.8e-14 0.00057 1.5e-13 7.1e-06 2.3e-
20 chr3  < 2e-16 0.00744 < 2e-16 2.9e-14 0.51260 8.2e-13 0.01090 < 2e-16 < 2e-
21 chr4  < 2e-16 0.43687 < 2e-16 < 2e-16 0.05984 4.8e-16 0.00012 < 2e-16 < 2e-
22 chr5  < 2e-16 0.00061 < 2e-16 3.8e-10 0.97958 1.4e-09 0.10736 < 2e-16 < 2e-
23 chr6  0.24959 < 2e-16 3.2e-12 0.01124 1.8e-11 0.10599 1.0e-10 < 2e-16 < 2e-
24 chr7  < 2e-16 2.1e-09 < 2e-16 0.00018 0.05984 0.00012 0.54074 < 2e-16 < 2e-
25 chr8  < 2e-16 0.00052 < 2e-16 2.3e-08 0.82241 3.7e-08 0.20102 < 2e-16 < 2e-
26 chr9  0.11632 < 2e-16 5.4e-12 0.06111 1.4e-09 0.26826 1.9e-08 < 2e-16 < 2e-
27 chrM  0.00533 0.00101 0.00982 0.00357 0.00160 0.00385 0.00201 0.01255
0.01407
28 chrX  8.5e-08 < 2e-16 0.49193 3.9e-16 < 2e-16 4.5e-11 < 2e-16 0.00480 4.0e-
05

```



```

62 chr16 - - - - -
63 chr17 - - - - -
64 chr18 - - - - -
65 chr19 - - - - -
66 chr2 - - - - -
67 chr20 - - - - -
68 chr21 - - - - -
69 chr22 - - - - -
70 chr3 - - - - -
71 chr4 - - - - -
72 chr5 - - - - -
73 chr6 - - - - -
74 chr7 2.4e-10 - - - -
75 chr8 1.2e-15 0.04376 - - -
76 chr9 0.64586 6.9e-08 1.9e-12 - -
77 chrM 0.00478 0.00220 0.00165 0.00452 - -
78 chrX 1.0e-08 < 2e-16 < 2e-16 7.7e-09 0.00918 -
79 chrY 0.39231 4.2e-05 2.4e-07 0.27649 0.00583 0.03541
80
81 P value adjustment method: fdr

```

```

1 wilcoxon signed rank exact test
2
3 data: glen[17:18]
4 v = 3, p-value = 0.5
5 alternative hypothesis: true location is not equal to 0

```

2

```

1 myData <- read.table('/home/ubuntu/R_course/UCB_admission.txt', header = T)
2 # (1)
3 table <- table(myData$sex, myData$accepted)
4 prop.test(table)
5 # p-value < 0.05 录取率相同
6
7 # (2)
8 acceptRate <- prop.table(table(myData$department, myData$accepted), 1)
9 acceptRate
10 sexRate <- prop.table(table(myData$department, myData$sex), 1)
11 sexRate
12 cor(acceptRate[, 2], sexRate[, 1])
13 cor.test(acceptRate[, 2], sexRate[, 1])
14
15
16 # (3)
17 plot(x = acceptRate[, 2], sexRate[, 1])
18 print('-----')
19 print('分析性别录取率')
20 prop.table(table(myData$accepted, myData$sex, myData$department), c(2, 3))
21 b <- as.array(table(myData$accepted, myData$sex, myData$department))
22 print('-----')
23 print('差异性检验')
24 apply(b, 3, chisq.test)

```

```

1      2-sample test for equality of proportions with continuity correction
2
3  data:  table
4  X-squared = 91.61, df = 1, p-value < 2.2e-16
5  alternative hypothesis: two.sided
6  95 percent confidence interval:
7    0.1129887 0.1703022
8  sample estimates:
9    prop 1    prop 2
10  0.6964578 0.5548123

```

```

1      N      Y
2  A 0.35584137 0.64415863
3  B 0.36752137 0.63247863
4  C 0.64923747 0.35076253
5  D 0.66035354 0.33964646
6  E 0.74828767 0.25171233
7  F 0.93557423 0.06442577

```

```

1      female      male
2  A 0.11575563 0.88424437
3  B 0.04273504 0.95726496
4  C 0.64596950 0.35403050
5  D 0.47348485 0.52651515
6  E 0.67294521 0.32705479
7  F 0.47759104 0.52240896

```

相关系数为 -0.785793618734187

```

1      Pearson's product-moment correlation
2
3  data:  acceptRate[, 2] and sexRate[, 1]
4  t = -2.541, df = 4, p-value = 0.06391
5  alternative hypothesis: true correlation is not equal to 0
6  95 percent confidence interval:
7    -0.9753531 0.0711263
8  sample estimates:
9    cor
10  -0.7857936

```

```

1  [1] "-----"
2  [1] "分析性别录取率"
3
4  , , = A
5
6      female      male
7  N 0.17592593 0.37939394
8  Y 0.82407407 0.62060606

```

```

9
10 , , = B
11     female      male
12 N 0.32000000 0.36964286
13 Y 0.68000000 0.63035714
14
15 , , = C
16     female      male
17 N 0.65935919 0.63076923
18 Y 0.34064081 0.36923077
19
20 , , = D
21     female      male
22 N 0.65066667 0.66906475
23 Y 0.34933333 0.33093525
24
25 , , = E
26     female      male
27 N 0.76081425 0.72251309
28 Y 0.23918575 0.27748691
29
30 , , = F
31     female      male
32 N 0.92961877 0.94101877
33 Y 0.07038123 0.05898123

```

```

1 [1] "-----"
2 [1] "差异性检验"

```

```

1 $A
2
3     Pearson's Chi-squared test with Yates' continuity correction
4
5 data:  array(newX[, i], d.call, dn.call)
6 x-squared = 16.372, df = 1, p-value = 5.205e-05

```

```

1 $B
2
3     Pearson's Chi-squared test with Yates' continuity correction
4
5 data:  array(newX[, i], d.call, dn.call)
6 x-squared = 0.085098, df = 1, p-value = 0.7705

```

```

1 $C
2
3     Pearson's Chi-squared test with Yates' continuity correction
4
5 data:  array(newX[, i], d.call, dn.call)
6 x-squared = 0.63322, df = 1, p-value = 0.4262

```

```
1 $D
2
3     Pearson's Chi-squared test with Yates' continuity correction
4
5 data:  array(newX[, i], d.call, dn.call)
6 x-squared = 0.22159, df = 1, p-value = 0.6378
```

```
1 $E
2
3     Pearson's Chi-squared test with Yates' continuity correction
4
5 data:  array(newX[, i], d.call, dn.call)
6 x-squared = 0.80805, df = 1, p-value = 0.3687
```

```
1 $F
2
3     Pearson's Chi-squared test with Yates' continuity correction
4
5 data:  array(newX[, i], d.call, dn.call)
6 x-squared = 0.21824, df = 1, p-value = 0.6404
```

