

定性数据统计分析作业 (2)

钟瑜 222018314210044

2020 年 10 月 23 日

习题二

1. 消费者对糖果颜色的偏好的分布如下表所示：

橙色	黄色	红色	棕色	绿色	蓝色
30 %	20 %	20 %	10 %	10 %	10 %

随机取一袋糖果, 其中的 500 块糖果中, 各种颜色的个数如下所示:

橙色	黄色	红色	棕色	绿色	蓝色
172	124	85	41	36	42

这些数据是否与消费者对糖果颜色的偏好分布相符? (取水平 $\alpha = 0.05$)

解. R 代码如下:

```
1 > x<-c(172,124,85,41,36,42)
2 > p<-c(0.3,0.2,0.2,0.1,0.1,0.1)
3 > n<-sum(x)
4 > chi<-sum((x-n*p)*(x-n*p)/(n*p))
5
6 > p_value=1-pchisq(chi,5);
7 > p_value
8 [1] 0.002876218
9
10 # 或者
11 > chisq.test(x,y=NULL,correct =TRUE,p=p)
12
13 Chi-squared test for given probabilities
14
15 data:  x
16 X-squared = 18.057, df = 5, p-value = 0.002876
```

p 值小于 $\alpha=0.05$, 故这些数据确实与消费者对糖果颜色的偏好不相符.

2. 某菜场为了解顾客对三种肉食(猪肉、牛肉、羊肉)的喜好程度,调查了 200 人。分别有 85 人、41 人、74 人最喜欢猪肉、牛肉、羊肉。取水平 $\alpha=0.05$, 检验“顾客对这三种肉食的喜好程度相同”的假设。

解. *R* 代码如下:

```
1 > x<-c(85,41,74)
2 > chisq.test(x,y=NULL,correct = TRUE,p=rep(1/length(x),length(x)))
3
4 Chi-squared test for given probabilities
5
6 data:  x
7 X-squared = 15.73, df = 2, p-value = 0.0003839
```

p 值小于 $\alpha=0.05$, 故顾客对这三种肉食的喜好程度不相同.

附录

A 关于卡方分布的函数说明

The (non-central) Chi-Squared Distribution

Description

Density, distribution function, quantile function and random generation for the chi-squared (χ^2) distribution with df degrees of freedom and optional non-centrality parameter ncp.

Usage

`dchisq(x, df, ncp = 0, log = FALSE)`

`pchisq(q, df, ncp = 0, lower.tail = TRUE, log.p = FALSE)` 卡方随机变量的累积分布函数

`qchisq(p, df, ncp = 0, lower.tail = TRUE, log.p = FALSE)`

`rchisq(n, df, ncp = 0)`

`dchisq` gives the density,

`pchisq` gives the distribution function,

`qchisq` gives the quantile function,

`rchisq` generates random deviates.

Arguments

x, q: vector of quantiles.

p: vector of probabilities.

n: number of observations. If `length(n) > 1`, the length is taken to be the number required.

df: degrees of freedom (non-negative, but can be non-integer).

ncp: non-centrality parameter (non-negative).

log, log.p: logical; if TRUE, probabilities p are given as $\log(p)$.

lower.tail: logical; if TRUE (default), probabilities are $P[X \leq x]$, otherwise, $P[X > x]$.

B 关于卡方检验的函数说明

Pearson's Chi-squared Test for Count Data

Description

`chisq.test` performs chi-squared contingency table tests and goodness-of-fit tests.

Usage

```
chisq.test(x, y = NULL, correct = TRUE, p = rep(1/length(x), length(x)), rescale.p = FALSE,  
simulate.p.value = FALSE, B = 2000)
```

Arguments

`x`: a numeric vector or matrix. `x` and `y` can also both be factors.

`y`: a numeric vector; ignored if `x` is a matrix. If `x` is a factor, `y` should be a factor of the same length.

`correct`: a logical indicating whether to apply continuity correction when computing the test statistic for 2 by 2 tables: one half is subtracted from all $|O - E|$ differences; however, the correction will not be bigger than the differences themselves. No correction is done if `simulate.p.value = TRUE`.

`p`: a vector of probabilities of the same length of `x`. An error is given if any entry of `p` is negative.
`rescale.p`: a logical scalar; if `TRUE` then `p` is rescaled (if necessary) to sum to 1. If `rescale.p` is `FALSE`, and `p` does not sum to 1, an error is given.

`simulate.p.value`: a logical indicating whether to compute p-values by Monte Carlo simulation.

`B`: an integer specifying the number of replicates used in the Monte Carlo test.