

## 第十四周——同分布检验与列联表检验

### 题目目的

- (一) 掌握 Pearson 拟合优度检验。
- (二) 掌握 Kolmogorov-Smirnov 同分布检验。
- (三) 掌握列联表独立性检验。
- (四) 掌握 Fisher 精确性检验。
- (五) 掌握 McNemar 检验和 Kappa 检验。
- (六) 掌握正态性检验方法。

### 题目

**题目一：拟合优度卡方检验。**打开脚本文件 test1401.R，完成下面任务。

自定义函数 fun1401 计算拟合优度卡方检验的检验 P 值，其中输入参数 ni 为观测频数，pi 为理论概率。

检验统计量为

$$Z = \sum_{i=1}^m \frac{(n_i - np_i)^2}{np_i} \sim \chi^2(m - k - 1)$$

其中  $n_i$  为观测频数， $p_i$  为理论频率。

返回的检验 P 值等于  $1 - P(Z \leq z_0)$ 。

检验  $x$  中的数据是否服从参数为 8 的泊松分布。

```
# 完成自定义函数
fun1401 = function(ni, pi){

}

# 测试 lambda=8?
x = scan(file = 'test1401.txt', sep = ',')
```

**题目二：创建脚本文件 test1402.R，完成下面任务。**

某商店调查了某时间段五种啤酒的销售量（单位：件）分别为 210, 312, 170, 85, 223, 请判断顾客对这五种啤酒的偏好有没有显著性差异，其中  $\alpha = 0.05$ 。

**题目三：Kolmogorov-Smirnov 检验。创建脚本文件 test1403.R，验证自由度较大时的 t 分布近似符合标准正态分布，在本实验中，取自由度为 24。**

方法一：单样本的 ks.test 函数验证。

方法二：双样本的 ks.test 函数验证。

**题目四：正态性检验。创建脚本文件 test1404.R，完成下面任务。**

某实验小组有 15 名学生，测量某物品的长度（单位：cm）分别为 19.9, 20.1, 20.0, 20.1, 20.2, 20.0, 19.9, 20.0, 20.0, 20.0, 20.1, 20.0, 20.0, 19.9, 19.8。用 qqnorm 绘制散点图，观察是否近似服从标准正态分布。请 Shapiro-Wilk 检验法检验该实验测得的长度是否符合正态分布，其中  $\alpha = 0.05$ 。

**题目五：独立性与分布齐性检验。**创建脚本文件 `test1405.R`，完成下面任务。

某种药物治疗不同年龄段患者所获得效果统计如下表所示，请分析药物疗效是否与年龄段有关，其中  $\alpha = 0.05$ 。

效果	儿童	中年人	老年人
显著	61	33	29
一般	31	51	41
较差	24	20	12

**题目六：列联表显著性检验。**创建脚本文件 `test1406.R`，完成下面任务。

甲乙两学徒制作手工品，其中甲学徒制作了 7 件工艺器中合格品有 3 件，不合格品有 4 件，乙学徒制作了 10 件工艺品，其中合格品有 6 件，不合格品有 4 件，请问两学徒制作的合格品中有没有显著差异？其中  $\alpha = 0.05$ 。

**题目七：差异性检验。**创建脚本文件 `test1407.R`，完成下面任务。

某高中进行文理分科，由于报文科学生偏少，学校通过宣传，鼓励学生选择文科，得到如下表所示的统计数据。请问学校的宣传是否对学生选科有显著影响？其中  $\alpha = 0.05$ 。

	宣传前报理科人数	宣传前报文科人数
宣传后报理科人数	598	14
宣传后报文科人数	90	48

**题目八：一致性检验。**创建脚本文件 `test1408.R`，完成下面任务。

两位医生 A 和 B 对一批患有某种疾病的病人进行分级诊断，结果如下表所示，请 Kappa 检验分布两位医生的诊断是否一致？其中  $\alpha = 0.05$ 。

A 诊断结果	B 诊断 1 级	B 诊断 2 级	B 诊断 3 级
1 级	13	13	12
2 级	6	9	8
3 级	5	7	9

## 答案及解析

### 题目一：

```
# 完成自定义函数
fun1401 = function(ni, pi){
  m = length(ni)
  n = sum(ni)
  z0 = sum((ni-n*pi)^2/(n*pi))
  1 - pchisq(z0, m-1)
}
# 测试 lambda=8?
x = scan(file = 'test1401.txt', sep = ',')
ni = table(x)
pi = dpois(as.numeric(names(ni)), 8)
fun1401(ni, pi)
```

```
[1] 0.7810643
```

```
# 结论：服从 lambda = 8 的泊松分布
```

### 题目二：

```
# 做均匀分布检验
x = c(210, 312, 170, 85, 223)
chisq.test(x, p = rep(0.2, time = 5))$p.value
```

```
[1] 1.592068e-28
```

```
# 返回检验 P 值近似为 0，不服从均匀分布，
# 即有明显的偏好
```

### 题目三：

```
# 用 rt 产生随机数序列 x，验证随机数近似服从标准正态分布
set.seed(1)
x = rt(100, 24)
#(1)
ks.test(x, pnorm, mean = 0, sd = 1)$p.value
```

```
[1] 0.8848285
```

```
# 由检验 P 值可知，可认为 x 服从标准正态分布
#(2)
# 用 rnorm 产生随机数序列 y，验证两者来源于同一分布
y = rnorm(200)
ks.test(x,y)$p.value
```

```
[1] 0.5175508
```

```
# 检验 P 值，可认为两者是来源于同一分布，即
# x 近似服从标准正态分布
```

## 题目四：

```
# 用 rt 产生随机数序列 x，验证随机数近似服从标准正态分布
set.seed(1)
x = rt(100, 24)
#(1)
ks.test(x, pnorm, mean = 0, sd = 1)$p.value
```

```
[1] 0.8848285
```

```
# 由检验 P 值可知，可认为 x 服从标准正态分布
#(2)
# 用 rnorm 产生随机数序列 y，验证两者来源于同一分布
y = rnorm(200)
ks.test(x,y)$p.value
```

```
[1] 0.5175508
```

```
# 检验 P 值，可认为两者是来源于同一分布，即
# x 近似服从标准正态分布
```

## 题目五：

```
x = c(61, 33, 29, 31, 51, 41, 24, 20, 12)
x = matrix(x, nrow = 3, byrow = TRUE)
chisq.test(x)$p.value
```

```
[1] 0.002189494
```

```
# 检验 P 值约为 0.0022，在显著水平为 0.05 的条件下，
# 可认为疗效与年龄段有关
```

## 题目六：

```
# 小样本，用 Fisher 精确检验
x = matrix(c(3,6,4,4),nrow = 2)
fisher.test(x)$p.value
```

```
[1] 0.6371863
```

```
# 检验 P 值约为 0.64，可认为两学徒的合格率不存在显著差异
```

## 题目七：

```
# 用 McNemar 检验
x = matrix(c(598,14,90,48), byrow = TRUE, nrow = 2)
mcnemar.test(x)$p.value
```

```
[1] 1.918511e-13
```

```
# 检验 P 值约为 0，表明宣传对选科有明显的影响
```

## 题目八：

```
x = matrix(c(13, 13, 12, 6, 9, 8, 5, 7, 9), nrow = 3,
           byrow = TRUE)
# install.packages('epiR')
epiR::epi.kappa(x, method = 'cohen')$z$p.value
```

```
[1] 0.3255218
```

# 返回 P 值约为 0.33，可认为两医生的诊疗结论是一致的