test way

XiaoYi

2021/3/4

Table of Contents

# 建立动态链接文档

用 servr::rmdv2("dir") 时，目录中文件名不能有空格

library("tidyverse")  
library("gt")  
library("printr")  
library("readxl")

# 计量资料的统计描述

对象E02\_01:

E02\_01 <- read\_xls("../医学统计学第五版/02章/09 R数据文件/E02\_01.xls")  
E02\_01[[1]]

## [1] 3.96 4.23 4.42 3.59 5.12 4.02 4.32 3.72 4.76 4.16 4.61 4.26 3.77 4.20 4.36  
## [16] 3.07 4.89 3.97 4.28 3.64 4.66 4.04 4.55 4.25 4.63 3.91 4.41 3.52 5.03 4.01  
## [31] 4.30 4.19 4.75 4.14 4.57 4.26 4.56 3.79 3.89 4.21 4.95 3.98 4.29 3.67 4.69  
## [46] 4.12 4.56 4.26 4.66 4.28 3.83 4.20 5.24 4.02 4.33 3.76 4.81 4.17 3.96 3.27  
## [61] 4.61 4.26 3.96 4.23 3.76 4.01 4.29 3.67 3.39 4.12 4.27 3.61 4.98 4.24 3.83  
## [76] 4.20 3.71 4.03 4.34 4.69 3.62 4.18 4.26 4.36 5.28 4.21 4.42 4.36 3.66 4.02  
## [91] 4.31 4.83 3.59 3.97 3.96 4.49 5.11 4.20 4.36 4.54 3.72 3.97 4.28 4.76 3.21  
## [106] 4.04 4.56 4.25 4.92 4.23 4.47 3.60 5.23 4.02 4.32 4.68 4.76 3.69 4.61 4.26  
## [121] 3.89 4.21 4.36 3.42 5.01 4.01 4.29 3.68 4.71 4.13 4.57 4.26 4.03 5.46 4.16  
## [136] 3.64 4.16 3.76

## 频数描述

|  |  |
| --- | --- |
| 函数 | 描述 |
| length() | 显示对象中元素的数量 |
| max() | 计算最大值 |
| min() | 计算最小值 |
| mean() | 计算平均值 |
| median() | 计算中位数 |
| sd() | 计算标准差 |
| var() | 计算方差 |
| sd()/sqrt() | 计算标准误 |
| range() | 计算值域 |
| quantile(,) | 求分位数 |

### Codes:

length(E02\_01$x) # 显示对象中元素的数量  
max(E02\_01$x) # 计算最大值  
min(E02\_01$x) # 计算最小值  
mean(E02\_01$x) # 计算平均值  
median(E02\_01$x) # 计算中位数  
sd(E02\_01$x) # 计算标准差  
var(E02\_01$x) # 计算方差  
sd(E02\_01$x)/sqrt(length(E02\_01$x)) # 计算标准误  
range(E02\_01$x) # 计算值域  
quantile(E02\_01$x,c(0.025, 0.25, 0.5, 0.75, 0.975)) # 求分位数， 2.5%, 25%, 50%, 75%, 97.5 %的分位点

## 频数分布图

|  |  |
| --- | --- |
| 函数 | 描述 |
| qqnorm() | Q-Q图 |
| hist() | 直方图 |
| boxplot() | 箱式图 |
| plot() | 散点图 |
| stem() | 茎叶图 |
| ggplot()+  geom\_line(, stat = "density") | 概率密度曲线 |

### Codes:

shapiro.test(E02\_01$x) # 夏皮洛-威尔克正态性检验  
qqnorm( E02\_01$x ) # 画出E02\_01的Q-Q图  
hist(E02\_01$x) # 直方图  
boxplot(E02\_01$x) # 箱式图  
plot(E02\_01$x) # 散点图  
stem(E02\_01$x) # 将数据序列x从x轴到数据值按照茎状形式画出，以圆圈终止。杆图  
  
X <- pretty( c(2.5, 6), 300 ) # pretty()函数用于创建美观的分割点  
# dnorm()函数返回正态分布概率密度函数，求出X点的概率值  
TranData <- dnorm( X, mean=mean(E02\_01$x), sd = sd(E02\_01$x) )  
# X与X点的概率值之间的散点图  
plot( X, TranData)  
  
ggplot(E02\_01)+ #ggplot2做图  
 geom\_line(aes(x), stat = "density")

## 清洗数据

|  |  |
| --- | --- |
| 函数 | 描述 |
| summary() | 查看统计 |
| str() | 查看数据结构 |
| head() | 查看开头几行数据（n=行数） |
| tail() | 查看结尾几行数据（n=行数） |
| dplyr::slice() | 查看指定行数（4:9） |
| dplyr::select() | 选择指定列 |
| dplyr::filter() | 筛选行 |
| dplyr::arrange() | 排序 |
| dplyr::group\_by() | 分组 |
| dplyr::summarise() | 汇总 |

### Code:

summary(E02\_01)  
str(E02\_01)  
head(E02\_01[[1]])  
tail(E02\_01[[1]])  
E02\_01 %>% slice(2:5)

接下来我们用内置的mtcars集来演示：

head(mtcars) #只取前几行  
mtcars %>% select(mpg,cyl) %>% head()  
mtcars %>% filter(cyl == 4) %>% arrange(-mpg)#筛选 cyl为4的行并按mpg从大到小排序  
mtcars %>% summarise(mean\_mpg = mean(mpg)) #不分组统计  
mtcars %>% group\_by(cyl) %>% summarise(mean\_mpg = mean(mpg)) #分组统计

# 计量资料常用检验

## T检验

* 适用范围：
  + 样本含量**较少**
  + 均值参数比较
* T检验需要满足条件：
  + 总体为正态分布
  + 两样本对应的总体的方差相等 即

### 单样本T检验

用例子 *3-5* 中数据估计样本和总体的均值是否有区别，其中已知总体均值 codes:

df <- read\_xls("../医学统计学第五版/03章/09 R数据文件/E03\_05.xls")  
df$hb  
t.test(df$hb, mu = 140) # 已知总体均值 mu = 140

### 配对样本T检验

用例子 *3-6* 中数据比较两种实验方法是否有区别 codes:

df <- read\_xls("../医学统计学第五版/03章/09 R数据文件/E03\_06.xls")  
df

|  |  |  |
| --- | --- | --- |
| no | x1 | x2 |
| 1 | 0.840 | 0.580 |
| 2 | 0.591 | 0.509 |
| 3 | 0.674 | 0.500 |
| 4 | 0.632 | 0.316 |
| 5 | 0.687 | 0.337 |
| 6 | 0.978 | 0.517 |
| 7 | 0.750 | 0.454 |
| 8 | 0.730 | 0.512 |
| 9 | 1.200 | 0.997 |
| 10 | 0.870 | 0.506 |

t.test(df$x1, df$x2, paired = T) #若是长表格则可以用 ~ 匹配关系

##   
## Paired t-test  
##   
## data: df$x1 and df$x2  
## t = 7.926, df = 9, p-value = 2.384e-05  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 0.1946542 0.3501458  
## sample estimates:  
## mean of the differences   
## 0.2724