

作业二

题一：

- 数据分析：单样本计量数据且样本数据较少($n=10$)
- 符合正态分布用单样本t检验，不符合正态性用Wilcox
- 正态性检验(Shapiro-Wilk normality test) `shapiro.test()`
- 正态性检验 $p = 0.625 > 0.05$ ，符合正态性，但是 n 较小，用该方法检验存在假阳性
- 先用单样本t检验 `t.test()`，再用 `wilcox.test()`，都用单侧

检验	原假设 H_0	备择 H_1	得到的 p	结论 ($\alpha = 0.05$)
t 检验	均值 ≥ 400	均值 < 400	0.7311	不拒绝 $H_0 \rightarrow$ 无证据表明均值 < 400
Wilcoxon	中位数 ≥ 400	中位数 < 400	0.2783	不拒绝 $H_0 \rightarrow$ 无证据表明中位数 < 400

结论: 无论是均值还是中位数，都没有足够证据认为 CD4 水平 低于 400 (p 值均 > 0.05)。当前数据与“ $CD4 \geq 400$ ”兼容。

```
# 1)data0205 cd4是一组艾滋病人的CD4细胞水平，采用合适的统计方法（说明选用依据），检验艾滋病人的CD4细胞水平
# 1.计量资料；2.判断正态性和方差齐 -> 样本量为10 -> ; 3.单样本t检验
library(haven)

d_cd4 <- read_sav("./DATA/data0205+cd4.sav"); head(cd4)

cd4 <- d_cd4$cd4

# description info
summary(cd4)
mean(cd4)
sd(cd4) # 标准差
# hist(cd4, col="black", main = "cd4 histogram", xlab = "cd4") # 频数直方图

# 正态性检验
shapiro.test(cd4) # p > 0.05, 服从正态分布

t.test(cd4, mu=400, alternative = "less") # p < 0.05, 拒绝原假设, cd4 < 400

wilcox.test(cd4, mu=400, alternative = "less") # p < 0.05, 拒绝原假设, cd4 < 400
```

题二：

- 数据分析：单样本计量数据($n=50$)
- 符合正态分布用单样本t检验，不符合正态性用Wilcox
- 正态性检验(Shapiro-Wilk normality test) `shapiro.test()`
- 正态性检验 $p = 3.456e-05 < 0.05$ ，不符合正态性
- 用 `wilcox.test()`，双侧

结论: 零假设为与15无差异，在99%置信水平下，该栖息地所有植物叶片蛋白质干重比例的**中位数介于12.6与17.4**之间，wilcox检验 $p\text{-value} = 0.4961$ ，不拒绝零假设，则植物的叶片蛋白质干重比例与15无明显差异。

```
# 2)data0206 protein是某种灵长类动物栖息地内几种植物的叶片蛋白质干重比例，采用合适的统计方法（说明选
# 该物种栖息地内所有种类植物叶片蛋白质干重比例的99%置信区间是多少

library(haven)
d_protein <- read_sav('./DATA/data0206+protein.sav'); head(d_protein)

dim(d_protein)

t.test(d_protein$protein, mu=15, conf.level = 0.99, alternative = "two.sided") # "two.sided" (c

shapiro.test(d_protein$protein) # p = 3.456e-05 < 0.05, 不服从正态分布

wilcox.test(d_protein$protein, mu=15, conf.int = TRUE, conf.level= 0.99, alternative = "two.side
# data: d_protein$protein
# V = 566.5, p-value = 0.4961
# alternative hypothesis: true location is not equal to 15
# 99 percent confidence interval:
# 12.56998 17.44001
# sample estimates:
# (pseudo)median
# 14.31494
```