

第一次练习课参考答案

黄嘉平

2025-03-28

第一题

1. 命令的含义

根据正态分布 $N(5, 0.5^2)$ 生成 20 个伪随机数，四舍五入至小数点后 1 位，然后保存至 x 。

2.

```
# (i) 将 x 中的值从小到大排序后代入变量 y
y <- x |> sort()

# (ii) 将 y 中的值标准化后代入变量 y_stand
y_stand <- (y - mean(y)) / sd(y)

# (iii) 计算 y_stand 的经验分布函数 (ECDF) 值，并保存至变量 y_psample
y_psample <- c() # 准备一个空向量
for (i in 1:length(y_stand)) {
  y_psample[i] <- sum(y_stand <= y_stand[i]) / length(y_stand)
  # 针对每一个 i，首先比较 y_stand <= y_stand[i]，这会返回一个逻辑向量
  # 逻辑值 TRUE 对应 1，FALSE 对应 0，因此对逻辑向量求和可得真值个数
}

# (iv) 计算 y_stand 对应的正态分布函数值，并保存至变量 y_pnorm
y_pnorm <- pnorm(y_stand)

# (v) 计算 y_psample 和 y_pnorm 之间的均方误差
y_mse <- (y_psample - y_pnorm)^2 |> mean()
y_mse
```

```
## [1] 0.004127242
```

均方误差约为 0.004。

3. 均值的假设检验

```
# 均值的置信区间
x_test <- t.test(x, mu = 5, conf.level = 0.95)
x_test$p.value
```

```
## [1] 0.09824025
```

```
x_test$conf.int
```

```
## [1] 4.603336 5.036664
```

```
## attr(,"conf.level")
```

```
## [1] 0.95
```

p -值为 0.098，因此在 5% 水平下无法拒绝零假设。95% 置信区间为 (4.603336, 5.036664)。

第二题

1. women 的数据类型

```
class(women)
```

```
## [1] "data.frame"
```

2. 转变为 tibble 数据

```
women_tb <- as_tibble(women)
class(women_tb)
```

```
## [1] "tbl_df"      "tbl"        "data.frame"
```

3. 计算 BMI 并添加

```
women_tb <- women_tb |> mutate(bmi = weight / height^2 * 703)
# BMI 的计算公式: 体重 / 身高的平方
# 当体重的单位是 lbs 磅, 身高的单位是 in 英寸时, 在计算 BMI 时应乘以 703
head(women_tb)
```

```
## # A tibble: 6 x 3
```

```
##   height weight  bmi
```

```
##   <dbl> <dbl> <dbl>
```

```
## 1     58    115  24.0
```

```
## 2     59    117  23.6
```

```
## 3      60      120  23.4
## 4      61      123  23.2
## 5      62      126  23.0
## 6      63      129  22.8
```

4. 针对身高小于 66 的观测值，计算体重和 BMI 的均值

```
women_tb |> filter(height < 66) |> summarize(mean(weight), mean(bmi))
```

```
## # A tibble: 1 x 2
##   `mean(weight)` `mean(bmi)`
##           <dbl>       <dbl>
## 1           125.         23.2
```

平均体重为 125 lbs，平均 BMI 为 23.2。