Midterm

Harry Wang

2024-03-19

I pledge my honor that I have abided by the Stevens Honor System

Problem 1:

```
library(readxl)
GpaGender = read_excel("./GpaGender.xls")
n = length(GpaGender$SelfConcept)
n
```

```
## [1] 78
```

```
significance = 0.05

#males
X = GpaGender[which(GpaGender$Gender == 2), "SelfConcept"]
X
```

```
## # A tibble: 47 × 1
##    SelfConcept
##          <dbl>
##  1          67
##  2          43
##  3          52
##  4          66
##  5          51
##  6          71
##  7          51
##  8          51
##  9          54
## 10          40
## # ℹ 37 more rows
```

```
male_summary = summary(X)
male_summary
```

```
##   SelfConcept
##  Min.   :20.00
##  1st Qu.:51.00
##  Median :59.00
##  Mean   :57.91
##  3rd Qu.:67.00
##  Max.   :80.00
```

```
#females
Y = GpaGender[which(GpaGender$Gender == 1), "SelfConcept"]
Y
```

```
## # A tibble: 31 × 1
##    SelfConcept
##          <dbl>
##  1          58
##  2          49
##  3          35
##  4          54
##  5          64
##  6          56
##  7          69
##  8          55
##  9          65
## 10          66
## # ℹ 21 more rows
```

```
female_summary = summary(Y)
female_summary
```

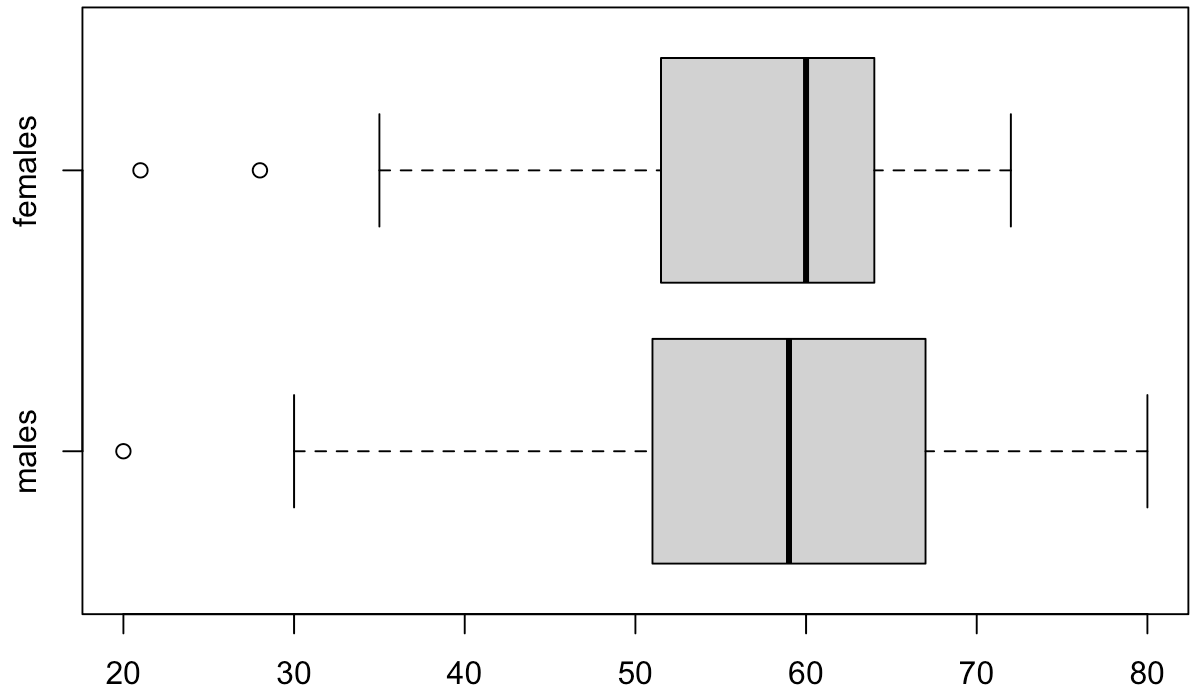
```
##   SelfConcept
##  Min.   :21.00
##  1st Qu.:51.50
##  Median :60.00
##  Mean   :55.52
##  3rd Qu.:64.00
##  Max.   :72.00
```

Problem 2:

```
together = c(X,Y)
together
```

```
## $SelfConcept
##  [1] 67 43 52 66 51 71 51 51 54 40 55 20 68 70 80 53 65 71 59 64 71 64 58 70 72
## [26] 70 47 52 66 67 63 53 67 61 63 30 54 66 44 49 67 73 59 36 42 51 56
##
## $SelfConcept
##  [1] 58 49 35 54 64 56 69 55 65 66 56 69 67 62 39 60 72 54 46 54 60 60 44 64 37
## [26] 63 64 28 60 70 21
```

```
boxplot(together,horizontal = TRUE, names = c("males","females"))
```



Problem 3:

```
t.test(X,conf.level = 0.95)
```

```
##
##  One Sample t-test
##
## data:  X
## t = 32.372, df = 46, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  54.31379 61.51600
## sample estimates:
## mean of x
##  57.91489
```

The 95% confidence interval for the true male population mean is 54.31 to 61.52.

```
t.test(Y,conf.level = 0.95)
```

```
##
##  One Sample t-test
##
## data:  Y
## t = 24.346, df = 30, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  50.85916 60.17310
## sample estimates:
## mean of x
##  55.51613
```

The 95% confidence interval for the true female population mean is 50.86 to 60.17.

Problem 4:

```
cat("The null hypothesis is H0 : σ2X = σ2Y")
```

```
## The null hypothesis is H0 : σ2X = σ2Y
```

```
nx = length(X$SelfConcept)
nx
```

```
## [1] 47
```

```
ny = length(Y$SelfConcept)
ny
```

```
## [1] 31
```

```
var(X$SelfConcept)
```

```
## [1] 150.4274
```

```
var(Y$SelfConcept)
```

```
## [1] 161.1914
```

The sample variance of sample X (male) is less than that of sample Y (female). So, I set my alternative hypothesis to be Ha : σ2X < σ2Y

```
var.test(X$SelfConcept, Y$SelfConcept, ratio = 1, alternative = "less")
```

```
##
##  F test to compare two variances
##
## data:  X$SelfConcept and Y$SelfConcept
## F = 0.93322, num df = 46, denom df = 30, p-value = 0.4088
## alternative hypothesis: true ratio of variances is less than 1
## 95 percent confidence interval:
##  0.000000 1.593047
## sample estimates:
## ratio of variances
##          0.9332221
```

The testing statistic F is observed as f = 0.93

The p-value 0.41 is not less than the significance level 0.05, so I fail to reject the null hypothesis, meaning that the male population Self Concept variance is equal to the female population Self Concept variance.

Problem 5:

The null hypothesis is H0 : μX = μY The alternative hypothesis is Ha : μX != μY

```
t.test(X$SelfConcept,Y$SelfConcept, alternative = "two.sided", paired = FALSE, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  X$SelfConcept and Y$SelfConcept
## t = 0.8336, df = 76, p-value = 0.4071
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.332460  8.129989
## sample estimates:
## mean of x mean of y
##  57.91489  55.51613
```

The p-value 0.41 is not less than the significance level 0.05, so I fail to reject the null hypothesis, meaning that the male population Self Concept mean is equal to the female population Self Concept mean.

Problem 6:

The null hypothesis is H0 : p = 0.45 The alternative hypothesis is Ha : p > 0.45

```
bin = ifelse(GpaGender$SelfConcept > 50, 1 ,0)
bin
```

```
##  [1] 1 0 1 1 1 1 1 1 0 1 0 1 1 1 1 1 1 0 1 1 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1
## [39] 1 1 1 1 1 1 0 1 0 1 1 1 1 1 1 1 1 1 0 1 1 0 0 0 1 1 1 1 0 1 0 1 0 0 1 1 1
## [77] 0 1
```

```
prop.test(sum(bin), length(bin), p = 0.45, alternative = "greater", correct = FALSE)
```

```
##
##  1-sample proportions test without continuity correction
##
## data:  sum(bin) out of length(bin), null probability 0.45
## X-squared = 34.748, df = 1, p-value = 1.876e-09
## alternative hypothesis: true p is greater than 0.45
## 95 percent confidence interval:
##  0.6964158 1.0000000
## sample estimates:
##         p
## 0.7820513
```

The p-value 1.876e-09 is less than the significance level 0.05, so I reject the null hypothesis, meaning that the population proportion is greater than 0.45.

Problem 7:

The null hypothesis is H0 : p1 = p2 The alternative hypothesis is Ha : p1 !- p2

```
malebin = ifelse(X > 50, 1, 0)
femalebin = ifelse(Y > 50,1,0)

prop.test(c(sum(malebin),sum(femalebin)), c(length(malebin),length(femalebin)), alternative = "two.sided", correc
t = FALSE)
```

```
##
##  2-sample test for equality of proportions without continuity correction
##
## data:  c(sum(malebin), sum(femalebin)) out of c(length(malebin), length(femalebin))
## X-squared = 0.48574, df = 1, p-value = 0.4858
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.1241610  0.2573113
## sample estimates:
##    prop 1    prop 2
## 0.8085106 0.7419355
```

The p-value 0.4858 is greater than the significance level 0.05, so I fail to reject the null hypothesis, meaning that the two groups have the same population proportion of Self Concept through a two tailed test.