

Midterm, MA 331-B

Daniel Detore

March 13, 2025

Problem 1

```
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 4.4.3
```

```
GpaGender <- read_excel("GpaGender.xls")
```

```
M_SC <- GpaGender[which(GpaGender$Gender == 2), "SelfConcept"]; M_SC
```

```
## # A tibble: 47 x 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
```

```
F_SC <- GpaGender[which(GpaGender$Gender == 1), "SelfConcept"]; F_SC
```

```
## # A tibble: 31 x 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
```

```
# Male summary:
```

```
summary(M_SC)
```

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

```
# Female summary:
summary(F_SC)
```

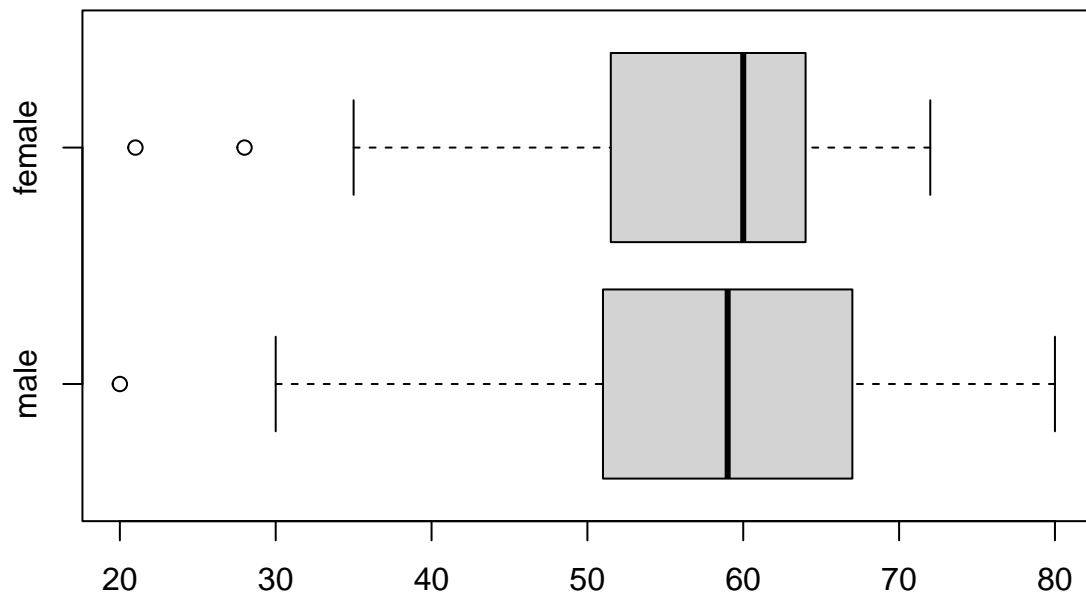
```
## SelfConcept
## Min. :21.00
## 1st Qu.:51.50
## Median :60.00
## Mean :55.52
## 3rd Qu.:64.00
## Max. :72.00
```

Problem 2

```
both_SC <- c(M_SC, F_SC); both_SC
```

```
## $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(both_SC, names = c("male", "female"), horizontal = TRUE)
```



Yes, their means (from Problem 1) are similar and, because of their interquartile ranges, they probably have similar variances.

Problem 3

Male interval is [54.31379, 61.51600] according to R:

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

```
##
## One Sample t-test
##
## data: M_SC
## 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
```

Female interval is [50.85916, 60.17310]:

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

```
##
## One Sample t-test
##
```

```
## data: F_SC
## 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
```

Problem 4

```
M_SC <- M_SC$SelfConcept
F_SC <- F_SC$SelfConcept
nm <- length(M_SC); nm
```

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

```
nf <- length(F_SC); nf
```

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

We assume $H_0 : S_M^2 = S_F^2$, with null distribution $F_{n_M-1, n_F-1} \equiv F_{46,30}$. For H_a , we check the sample variances:

```
SSM <- var(M_SC); SSM
```

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

```
SSF <- var(F_SC); SSF
```

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

Since $S_M^2 < S_F^2$, $H_a : \sigma_M^2 < \sigma_F^2$. Our observed test statistic F is

```
f <- SSM/SSF; f
```

```
## [1] 0.9332221
```

which, when we do a left-tail test, gives

```
pf(f, nm - 1, nf - 1)
```

```
## [1] 0.4087687
```

which is greater than our $\alpha = 0.1$, therefore we cannot reject H_0 and must assume $S_M^2 = S_F^2$. We could use the R function for the same result:

```
var.test(M_SC, F_SC, ratio = 1, alternative = c("less"), conf.level = 1 - 0.1)
```

```
##
```

```
## F test to compare two variances
```

```
##
```

```
## data: M_SC and F_SC
```

```
## F = 0.93322, num df = 46, denom df = 30, p-value = 0.4088
```

```
## alternative hypothesis: true ratio of variances is less than 1
```

```
## 90 percent confidence interval:
```

```
## 0.00000 1.41427
```

```
## sample estimates:
```

```
## ratio of variances
```

```
## 0.9332221
```

We can get confidence interval [0.000000, 1.593047] by

```
var.test(M_SC, F_SC, ratio = 1, alternative = c("less"), conf.level = 0.95)
```

```
##
## F test to compare two variances
##
## data: M_SC and F_SC
## 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
```

Problem 5

We found that $S_M^2 = S_F^2$, but we don't know their values, so we will do a two-sample T-test with equal variances. This means we have $H_0 : \mu_M = \mu_F$, $H_a : \mu_M \neq \mu_F$, null distribution $T_{n_M+n_F-2} \equiv T_{76}$, and our observed test statistic $t =$

```
mbar <- mean(M_SC); mbar
```

```
## [1] 57.91489
```

```
fbar <- mean(F_SC); fbar
```

```
## [1] 55.51613
```

```
SSP <- (((nm - 1)*SSM) + ((nf - 1)*SSF)) / (nm+nf-2); SSP
```

```
## [1] 154.6763
```

```
t <- (mbar - fbar) / sqrt(SSP * ((1/nm) + (1/nf))); t
```

```
## [1] 0.8336009
```

which gives us two-tail significance level

```
2*pt(-abs(t), nm + nf - 2)
```

```
## [1] 0.4071182
```

which is greater than our $\alpha = 0.1$, therefore we cannot reject H_0 and must assume $\mu_M = \mu_F$. We could use the R function for the same result:

```
t.test(M_SC, F_SC, alternative = c("two.sided"), paired = FALSE, var.equal = TRUE,
conf.level = 1 - 0.1)
```

```
##
## Two Sample t-test
##
## data: M_SC and F_SC
## t = 0.8336, df = 76, p-value = 0.4071
## alternative hypothesis: true difference in means is not equal to 0
## 90 percent confidence interval:
## -2.392864 7.190394
## sample estimates:
## mean of x mean of y
```

```
## 57.91489 55.51613
```

We can get confidence interval $[-3.332460, 8.129989]$ by

```
t.test(M_SC, F_SC, alternative = c("two.sided"), paired = FALSE, var.equal = TRUE,
conf.level = 0.95)
```

```
##
## Two Sample t-test
##
## data: M_SC and F_SC
## 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
```

Problem 6

```
SC <- GpaGender$SelfConcept; SC
```

```
## [1] 67 43 52 66 58 51 71 51 49 51 35 54 54 64 56 69 55 65 40 66 55 20 56 68 69
## [26] 70 80 53 65 67 62 39 71 59 60 64 71 72 54 64 58 70 72 70 47 52 46 66 67 63
## [51] 53 67 61 54 60 60 63 30 54 66 44 49 44 67 64 73 59 37 63 36 64 42 28 60 70
## [76] 51 21 56
```

```
B_SC <- ifelse(SC > 52, 1, 0); B_SC
```

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

```
N <- sum(B_SC); N
```

```
## [1] 55
```

```
n <- length(B_SC); n
```

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

```
phat <- N/n; phat
```

```
## [1] 0.7051282
```

```
p0 <- .47
```

We know $H_0 : p = .47$, $H_a : p > .47$, and our null distribution $N(0, 1)$. Our testing statistic and right-tail p-value are given as such:

```
z <- sqrt(n) * (phat - p0) / sqrt(p0 * (1 - p0)); z
```

```
## [1] 4.160688
```

```
1 - pnorm(z)
```

```
## [1] 1.58645e-05
```

Since the p-value is less than our $\alpha = 0.1$, we can reject H_0 and assume $p > .47$.

We get our confidence interval [0.6144804, 1.0000000] by

```
prop.test(N, n, p = p0, alternative = c("greater"), correct = FALSE, conf.level = .95)
```

```
##
## 1-sample proportions test without continuity correction
##
## data:  N out of n, null probability p0
## X-squared = 17.311, df = 1, p-value = 1.586e-05
## alternative hypothesis: true p is greater than 0.47
## 95 percent confidence interval:
##  0.6144804 1.0000000
## sample estimates:
##           p
## 0.7051282
```

Problem 7

```
BM <- ifelse(M_SC > 52, 1, 0); BM
```

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

```
BF <- ifelse(F_SC > 52, 1, 0); BF
```

```
## [1] 1 0 0 1 1 1 1 1 1 1 1 1 1 0 1 1 1 0 1 1 0 1 0 1 1 0 1 1 0
```

```
Nm <- sum(BM); Nm
```

```
## [1] 32
```

```
Nf <- sum(BF); Nf
```

```
## [1] 23
```

```
nm <- length(BM); nm
```

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

```
nf <- length(BF); nf
```

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

```
phat <- (Nm + Nf) / (nm + nf); phat
```

```
## [1] 0.7051282
```

We know $H_0 : p_M = p_F$, $H_a : p_M \neq p_F$, and null distribution $N(0, 1)$. Our testing statistic and two-tail p-value are given as such:

```
z <- (Nm/nm - Nf/nf) / sqrt(phat * (1 - phat) * (1/nm + 1/nf)); z
```

```
## [1] -0.5789783
```

```
2 * pnorm(-abs(z))
```

```
## [1] 0.5626038
```

which is greater than our $\alpha = 0.1$, therefore we cannot reject H_0 and must assume $p_M = p_F$. We could use the R function for the same result:

```
prop.test(c(Nm, Nf), c(nm, nf), alternative = c("two.sided"), correct = FALSE, conf.level = 1-0.1)
```

```
##
## 2-sample test for equality of proportions without continuity correction
##
## data:  c(Nm, Nf) out of c(nm, nf)
## X-squared = 0.33522, df = 1, p-value = 0.5626
## alternative hypothesis: two.sided
## 90 percent confidence interval:
##  -0.2320196  0.1098507
## sample estimates:
##      prop 1      prop 2
## 0.6808511 0.7419355
```

We can find the confidence interval $[-0.2647662, 0.1425974]$ by

```
prop.test(c(Nm, Nf), c(nm, nf), alternative = c("two.sided"), correct = FALSE, conf.level = .95)
```

```
##
## 2-sample test for equality of proportions without continuity correction
##
## data:  c(Nm, Nf) out of c(nm, nf)
## X-squared = 0.33522, df = 1, p-value = 0.5626
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.2647662  0.1425974
## sample estimates:
##      prop 1      prop 2
## 0.6808511 0.7419355
```