Midterm

Daniel Detore

March 13, 2025

```
library(readxl)
```

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

```
GpaGender <- read_excel("GpaGender.xls")
GPA <- GpaGender$GPA; GPA
```

```
##  [1]  7.940  8.292  4.643  7.470  8.882  7.585  7.650  2.412  6.000  8.833
## [11]  7.470  5.528  7.167  7.571  4.700  8.167  7.822  7.598  4.000  6.231
## [21]  7.643  1.760  6.419  9.648 10.700 10.580  9.429  8.000  9.585  9.571
## [31]  8.998  8.333  8.175  8.000  9.333  9.500  9.167 10.140  9.999 10.760
## [41]  9.763  9.410  9.167  9.348  8.167  3.647  3.408  3.936  7.167  7.647
## [51]  0.530  6.173  7.295  7.295  8.938  7.882  8.353  5.062  8.175  8.235
## [61]  7.588  7.647  5.237  7.825  7.333  9.167  7.996  8.714  7.833  4.885
## [71]  7.998  3.820  5.936  9.000  9.500  6.057  6.057  6.938
```

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

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

```
alpha <- 0.05
```

Problem 1

```
B_GPA <- ifelse(GPA > 7, 1, 0); B_GPA
```

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

```
phat <- sum(B_GPA)/length(B_GPA); phat
```

```
## [1] 0.7179487
```

```
p0 <- 0.30
```

$H_0 : p = 0.30$ with null distribution $N(0, 1)$

$H_a : p > 0.30$

We are testing one population proportion, so we use observation z of Z given by:

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

```
## [1] 8.054909
```

which makes the p-value for the left-tail test:

```r
pnorm(z)
```

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

which is greater than our $\alpha = 0.05$. We cannot disprove H_0.

Using the R function:

```r
prop.test(sum(B_GPA), length(B_GPA), p = p0, alternative = c("less"), correct = FALSE)
```

```
##
##  1-sample proportions test without continuity correction
##
## data:  sum(B_GPA) out of length(B_GPA), null probability p0
## X-squared = 64.882, df = 1, p-value = 1
## alternative hypothesis: true p is less than 0.3
## 95 percent confidence interval:
##  0.0000000 0.7933579
## sample estimates:
##         p
## 0.7179487
```

Where $p = 1 > \alpha = 0.05$. We cannot disprove H_0.

Problem 2

```r
M_GPA <- GpaGender[which(GpaGender$Gender == 2), "GPA"]; M_GPA
```

```
## # A tibble: 47 x 1
##      GPA
##    <dbl>
##  1  7.94
##  2  8.29
##  3  4.64
##  4  7.47
##  5  7.58
##  6  7.65
##  7  2.41
##  8  8.83
##  9  7.17
## 10  4
## # i 37 more rows
```

```r
F_GPA <- GpaGender[which(GpaGender$Gender == 1), "GPA"]; F_GPA
```

```
## # A tibble: 31 x 1
##      GPA
##    <dbl>
##  1  8.88
##  2  6
##  3  7.47
##  4  5.53
##  5  7.57
##  6  4.7
##  7  8.17
##  8  7.82
```

```
## 9  7.60
## 10  6.23
## # i 21 more rows
```

```r
# Male summary:
summary(M_GPA)
```

```
##       GPA
##  Min.   : 0.530
##  1st Qu.: 6.556
##  Median : 7.825
##  Mean   : 7.282
##  3rd Qu.: 9.000
##  Max.   :10.760
```

```r
# Female summary:
summary(F_GPA)
```

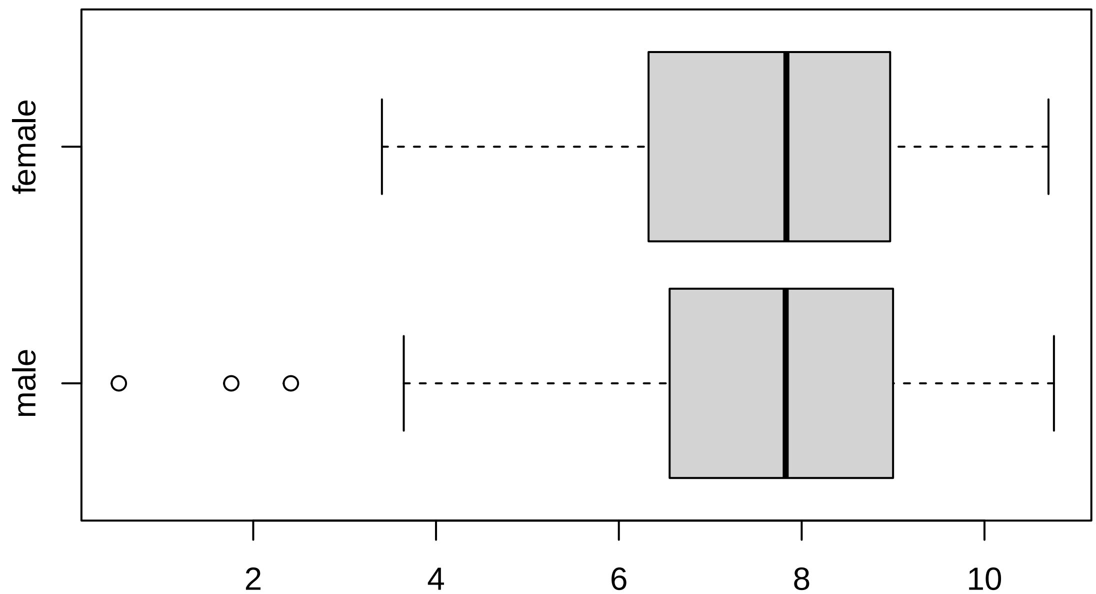
```
##       GPA
##  Min.   : 3.408
##  1st Qu.: 6.325
##  Median : 7.833
##  Mean   : 7.697
##  3rd Qu.: 8.968
##  Max.   :10.700
```

Problem 3

```r
both_GPA <- c(M_GPA, F_GPA); both_GPA
```

```
## $GPA
##  [1]  7.940  8.292  4.643  7.470  7.585  7.650  2.412  8.833  7.167  4.000
## [11]  7.643  1.760  9.648 10.580  9.429  8.000  9.585  8.175  8.000  9.500
## [21]  9.167 10.760  9.763  9.410  9.167  9.348  8.167  3.647  3.936  7.167
## [31]  7.647  0.530  6.173  7.295  8.353  5.062  8.175  8.235  7.588  7.647
## [41]  7.825  9.167  7.996  4.885  3.820  6.057  6.938
##
## $GPA
##  [1]  8.882  6.000  7.470  5.528  7.571  4.700  8.167  7.822  7.598  6.231
## [11]  6.419 10.700  9.571  8.998  8.333  9.333 10.140  9.999  3.408  7.295
## [21]  8.938  7.882  5.237  7.333  8.714  7.833  7.998  5.936  9.000  9.500
## [31]  6.057
```

```r
boxplot(both_GPA, names = c("male", "female"), horizontal = TRUE)
```



Problem 4

```
t.test(M_GPA, conf.level = 0.9)
```

```
##
##  One Sample t-test
##
## data:  M_GPA
## t = 21.527, df = 46, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 90 percent confidence interval:
##  6.713815 7.849462
## sample estimates:
## mean of x
##  7.281638
```

```
t.test(F_GPA, conf.level = 0.9)
```

```
##
##  One Sample t-test
##
## data:  F_GPA
## t = 24.903, df = 30, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 90 percent confidence interval:
##  7.171981 8.221116
```

```
## sample estimates:
## mean of x
##  7.696548
```

90% Confidence interval for males is $[6.713815, 7.849462]$; for females it is $[7.171981, 8.221116]$.

Problem 5

```r
M_GPA <- GpaGender[which(GpaGender$Gender == 2), "GPA"]$GPA; M_GPA
```

```
##  [1]  7.940  8.292  4.643  7.470  7.585  7.650  2.412  8.833  7.167  4.000
## [11]  7.643  1.760  9.648 10.580  9.429  8.000  9.585  8.175  8.000  9.500
## [21]  9.167 10.760  9.763  9.410  9.167  9.348  8.167  3.647  3.936  7.167
## [31]  7.647  0.530  6.173  7.295  8.353  5.062  8.175  8.235  7.588  7.647
## [41]  7.825  9.167  7.996  4.885  3.820  6.057  6.938
```

```r
F_GPA <- GpaGender[which(GpaGender$Gender == 1), "GPA"]$GPA; F_GPA
```

```
##  [1]  8.882  6.000  7.470  5.528  7.571  4.700  8.167  7.822  7.598  6.231
## [11]  6.419 10.700  9.571  8.998  8.333  9.333 10.140  9.999  3.408  7.295
## [21]  8.938  7.882  5.237  7.333  8.714  7.833  7.998  5.936  9.000  9.500
## [31]  6.057
```

$H_0 : \sigma_M^2 = \sigma_F^2$ where $H_0 \sim F_{n_M - 1,\ n_F - 1} \equiv F_{46,\ 30}$

Let's check our sample variances:

```r
SSM <- var(M_GPA); SSM
```

```
## [1] 5.377723
```

```r
SSF <- var(F_GPA); SSF
```

```
## [1] 2.961198
```

Since $S_M^2 > S_F^2$, $H_a : \sigma_M^2 > \sigma_F^2$

Our observed $F = f =$

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

```
## [1] 1.816063
```

and gives us p-value

```r
1 - pf(f, length(M_GPA) - 1, length(F_GPA) - 1)
```

```
## [1] 0.04321679
```

which is less than $\alpha = 0.05$. We can reject H_0 and assume $\sigma_M^2 > \sigma_F^2$. Using the R function:

```r
var.test(M_GPA, F_GPA, ratio = 1, alternative = c("greater"), conf.level = 1 - alpha)
```

```
##
##  F test to compare two variances
##
## data:  M_GPA and F_GPA
## F = 1.8161, num df = 46, denom df = 30, p-value = 0.04322
## alternative hypothesis: true ratio of variances is greater than 1
## 95 percent confidence interval:
##  1.025001      Inf
```

```
## sample estimates:
## ratio of variances
##          1.816063
```

Problem 6

We know $\sigma_M^2 \neq \sigma_F^2$ and we know neither value. So we will do a general two-sample T-test.

$H_0 : \mu_M = \mu_F$

$H_a : \mu_M \neq \mu_F$

We calculate our observed $T = t$ where

```r
xbar <- mean(M_GPA); xbar
```

```
## [1] 7.281638
```

```r
ybar <- mean(F_GPA); ybar
```

```
## [1] 7.696548
```

```r
n1 <- length(M_GPA); n1
```

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

```r
n2 <- length(F_GPA); n2
```

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

```r
t <- (xbar - ybar)/sqrt(SSM/n1 + SSF/n2); t
```

```
## [1] -0.9055327
```

and get a p-value with

```r
k <- (SSM/n1+SSF/n2)^2/((SSM/n1)^2/(n1-1)+(SSF/n2)^2/(n2-1))
k <- ceiling(k); k
```

```
## [1] 75
```

```r
2 * pt(-abs(t), k)
```

```
## [1] 0.3680831
```

which is greater than 0.05, so we cannot reject H_0, so $\mu_M = \mu_F$.

Using the R function:

```r
t.test(M_GPA, F_GPA, alternative = c("two.sided"), paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  M_GPA and F_GPA
## t = -0.90553, df = 74.862, p-value = 0.3681
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.3277078  0.4978876
## sample estimates:
## mean of x mean of y
##  7.281638  7.696548
```