SA1

Joseph Raphael M. Espiritu, Harneyyer Clores

II Test I:

```
# (from Table 10.6)
males <- c(
  12, 4, 11, 13, 11,
  7, 9, 10, 10, 7,
  7, 12, 6, 9, 15,
  10, 11, 12, 7, 8,
  8, 9, 11, 10, 9,
  10, 9, 9, 7, 9,
  11, 7, 10, 10, 11,
  9, 12, 12, 8, 13,
  9, 10, 8, 11, 10,
  13, 13, 9, 10, 13
)

females <- c(
  11, 9, 7, 10, 9,
  10, 10, 7, 9, 10,
  11, 8, 9, 6, 11,
  10, 7, 9, 12, 14,
  11, 12, 12, 8, 12,
  12, 9, 10, 11, 7,
  12, 7, 9, 8, 11,
  10, 8, 13, 8, 10,
  9, 9, 9, 11, 9,
  9, 8, 8, 12, 11
)

combined <- c(males, females)

#Descriptive Summary for Entire Dataset
describe(combined)
```

```
##    vars   n mean   sd median trimmed  mad min max range skew kurtosis  se
## X1    1 100 9.75 1.97     10    9.74 1.48   4  15    11    0    -0.13 0.2
```

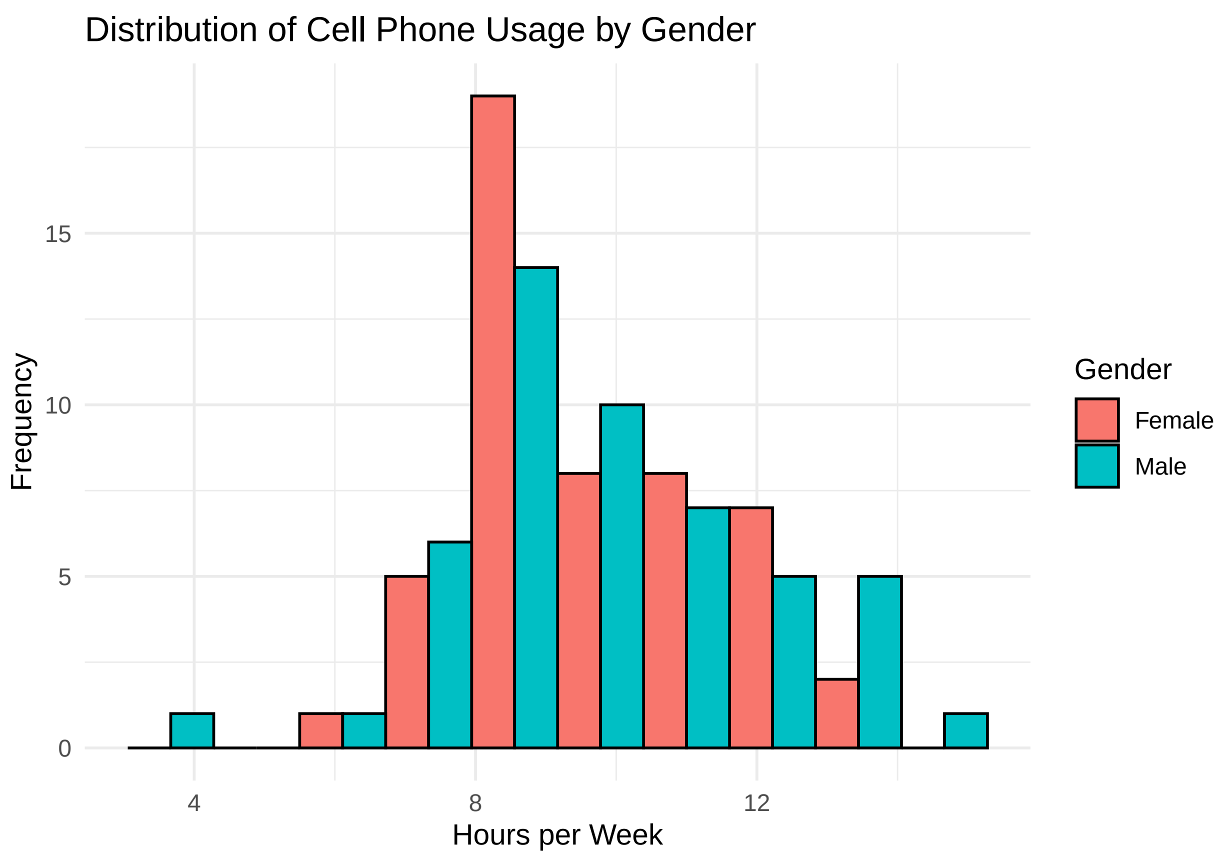
```
#Descriptive Summary by Gender
describeBy(c(males, females), group = rep(c("Male","Female"), each=50))
```

```
##
##  Descriptive statistics by group
## group: Female
##    vars  n mean   sd median trimmed  mad min max range skew kurtosis   se
## X1    1 50 9.68 1.79    9.5    9.68 2.22   6  14     8 0.14    -0.67 0.25
```

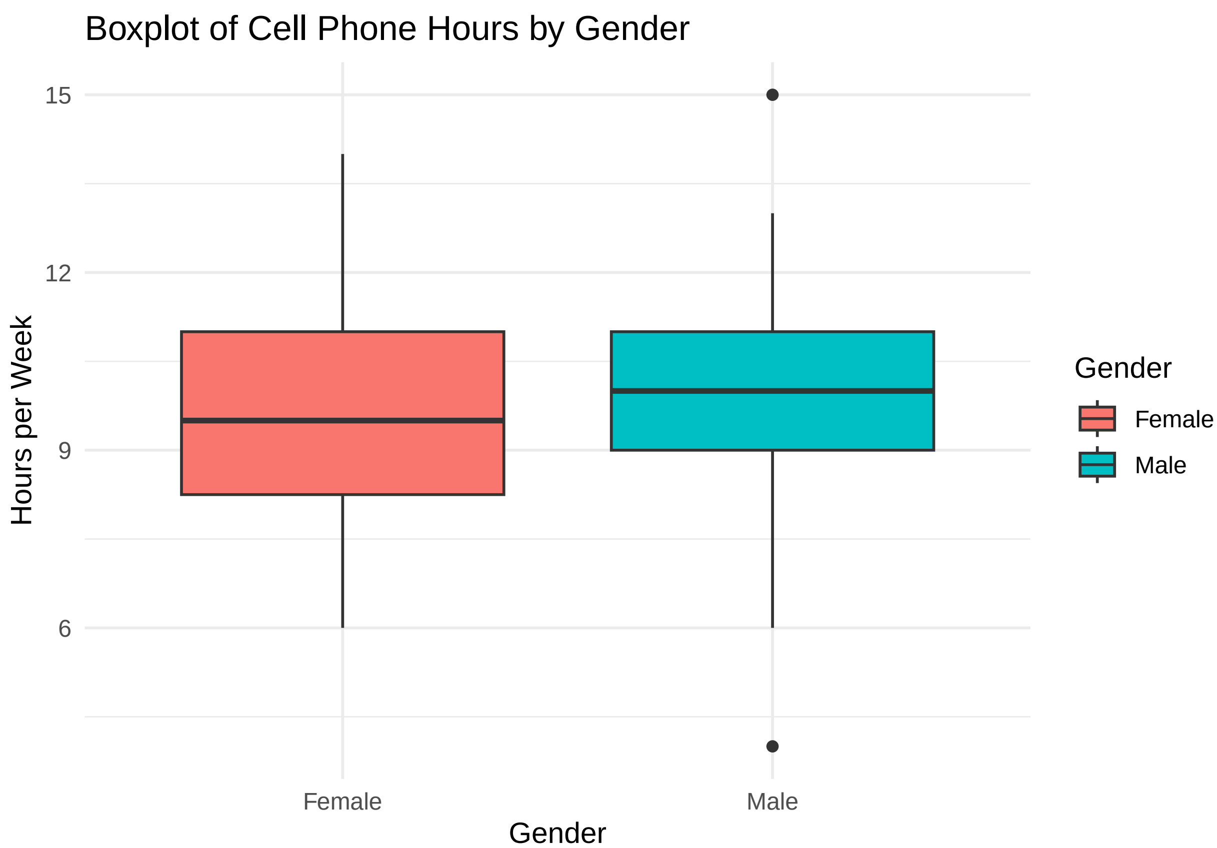
```
## ------------------------------------------------------------
## group: Male
##    vars  n mean   sd median trimmed  mad min max range  skew kurtosis  se
## X1    1 50 9.82 2.15     10    9.82 1.48   4  15    11 -0.11    -0.08 0.3
```

```r
df <- data.frame(
  Hours = c(males, females),
  Gender = rep(c("Male", "Female"), each = 50)
)

# Histogram
ggplot(df, aes(x = Hours, fill = Gender)) +
  geom_histogram(position = "dodge", bins = 10, color = "black") +
  labs(title = "Distribution of Cell Phone Usage by Gender",
       x = "Hours per Week", y = "Frequency") +
  theme_minimal()
```




```r
# Boxplot
ggplot(df, aes(x = Gender, y = Hours, fill = Gender)) +
  geom_boxplot() +
  labs(title = "Boxplot of Cell Phone Hours by Gender",
       y = "Hours per Week") +
  theme_minimal()
```




```r
t_test_result <- t.test(males, females, var.equal = TRUE)
t_test_result
```

```
##
##  Two Sample t-test
##
## data:  males and females
## t = 0.35351, df = 98, p-value = 0.7245
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.6459113  0.9259113
## sample estimates:
## mean of x mean of y
##      9.82      9.68
```

1. Descriptive statistics:

Males (mean   9.82 hours/week)

Females (mean   9.68 hours/week)

Both have a similar median (~10 hours), with slightly variability among males.

The boxplot shows overlap distributions → no strong visible difference.

2. Hypothesis test:

Null hypothesis (H):   =   (no difference between male and female mean hours)

Alternative hypothesis (H):   (there is a difference)

t = 0.35, p = 0.7245 → since $p > 0.05$, we fail to reject H .

Conclusion: There is no statistically significant difference in average cell phone talking time between male and female students at Midwestern University.

Test II:

```r
x <- c(2, 3, 7, 8, 10)

# Moments for the set 2, 3, 7, 8, 10
m1 <- mean(x)
m2 <- mean(x^2)
m3 <- mean(x^3)
m4 <- mean(x^4)

moments_raw <- c(m1, m2, m3, m4)
names(moments_raw) <- c("m1'", "m2'", "m3'", "m4'")
moments_raw
```

```
##    m1'    m2'    m3'    m4'
##    6.0   45.2  378.0 3318.8
```

These moments are calculated directly from the data values themselves — no centering around the mean yet. These values summarize the dataset's overall "power behavior." Higher-order moments (^{3},) show how data values grow when raised to powers — they magnify large numbers

```r
# Central Moments (about mean)
mean_x <- mean(x)
m2_central <- mean((x - mean_x)^2)
m3_central <- mean((x - mean_x)^3)
m4_central <- mean((x - mean_x)^4)

moments_central <- c(m2_central, m3_central, m4_central)
names(moments_central) <- c(" 2", " 3", " 4")
moments_central
```

```
##     2     3     4
##   9.2  -3.6 122.0
```

Variance (= 9.2) means data points differ from the mean by about $\sqrt{9.2}$ 3.03 units on average.

Negative   means the data leans slightly left-skewe).

The excess kurtosis = -1.56, which means:

The distribution is platykurtic → flatter than a normal distribution (not sharply peaked).

```r
# Verification
h <- mean_x
lhs <- m4
rhs <- m4_central + 4*h*m3_central + 6*h^2*m2_central + h^4

verification <- data.frame(LHS_m4prime = lhs, RHS_Expression = rhs)
verification
```

```
##   LHS_m4prime RHS_Expression
## 1      3318.8         3318.8
```

The code checks the mathematical relation between raw and central moments using this identity:

$$m'_4 = \mu_4 + 4h\mu_3 + 6h^2\mu_2 + h^4$$

where =mean=6

When substituted: LHS = 3318.8 RHS = 3318.8

They match → the relationship between raw and central moments holds true.