

# Homework #8, Mei Maddox.

Mei Maddox

Submit the solution on Canvas into the corresponding assignment (e.g. “Homework #1”) in the form of R Markdown report, knitted into either of the available formats (HTML, pdf or Word). Provide only code and output. NO NEED TO COPY THE PROBLEM FORMULATION (!)

## Problem 1

```
my.t.test <- function(vec, mu, ci.lvl=0.95){
  n.obs <- length(vec)
  xbar <- mean(vec)
  s <- sd(vec)
  se <- s/sqrt(n.obs)
  alpha <- 1 - ci.lvl

  ts <- (xbar-mu)/se

  p.value <- pt(abs(ts), df=n.obs-1, lower.tail = F)*2

  return( list(t=ts,
               df=n.obs-1,
               p.value=p.value,
               CI=c(xbar - qt(p=alpha/2, df=n.obs-1,lower.tail=F)*se,
                     xbar + qt(p=alpha/2, df=n.obs-1,lower.tail=F)*se),
               sample.estimates=mean(vec)) )
}

t.test(c(1:10), mu=5) # Comparative t.test() call.
```

```
##
## One Sample t-test
##
## data:  c(1:10)
## t = 0.52223, df = 9, p-value = 0.6141
## alternative hypothesis: true mean is not equal to 5
## 95 percent confidence interval:
##  3.334149 7.665851
## sample estimates:
## mean of x
##      5.5
```

```
my.t.test(c(1:10), mu=5)
```

```
## $t
## [1] 0.522233
##
## $df
## [1] 9
##
## $p.value
## [1] 0.6141173
##
## $CI
## [1] 3.334149 7.665851
##
## $sample.estimates
## [1] 5.5
```

```
# Example for non-typical CI level.
# Compare output with similar t.test() call.
my.t.test(c(1:10), mu=5, ci.lvl=0.99)
```

```
## $t
## [1] 0.522233
##
## $df
## [1] 9
##
## $p.value
## [1] 0.6141173
##
## $CI
## [1] 2.388519 8.611481
##
## $sample.estimates
## [1] 5.5
```

```
t.test(c(1:10), mu=5, conf.level=0.99)
```

```
##
## One Sample t-test
##
## data: c(1:10)
## t = 0.52223, df = 9, p-value = 0.6141
## alternative hypothesis: true mean is not equal to 5
## 99 percent confidence interval:
## 2.388519 8.611481
## sample estimates:
## mean of x
## 5.5
```

The results of the t-tests for the typical and non-typical CI level is the same between my custom version and the R-function.

## Problem 2

- a)  $H_0$  :Average working hours for adults is 40 ( $H_0 : \mu = 40$ ),  $H_a$  :Average working hours for adults is NOT 40 ( $H_a : \mu \neq 40$ ).

```
working.hours <- read.csv("https://img1.wsimg.com/blobby/go/bbca5dba-4947-4587-b40a-db346c01b1b3/download")
t.results <- t.test(working.hours$Hours, mu=40)
```

We fail to reject the null hypothesis due to the p-value 0.5478529 being larger than standard alpha 0.05. Therefore, we lack sufficient evidence to conclude that the mean number of hours worked is not equal to 40. The 95% confidence interval agrees with this assertion as the true population parameter we are testing for lies within the interval. The results also have little practical significance as the Cohen's effect size is extremely small ( $d = 0.0176051$ ).

- b)  $H_0$  :Average working hours for men is 40 ( $H_0 : \mu = 40$ ),  $H_a$  :Average working hours for men is NOT 40 ( $H_a : \mu \neq 40$ ).

```
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

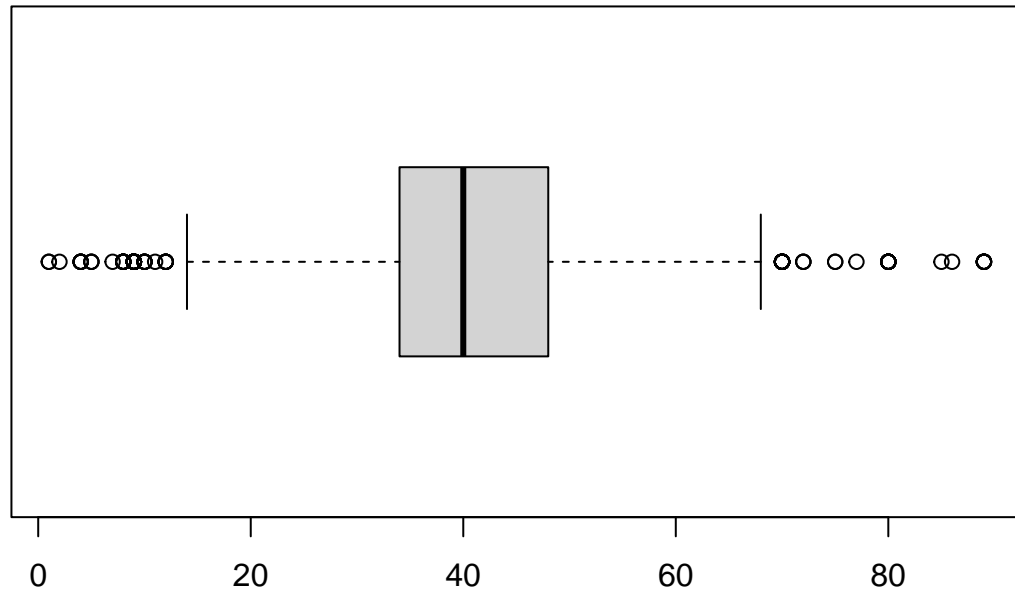
working.hours.gender <- working.hours %>% split(working.hours$Gender) %>% sapply(function(x) return (x))
t.results <- working.hours.gender$Male.Hours %>% t.test(mu=40)
```

We reject the null hypothesis in favor of the alternative hypothesis because the p-value  $4.668033 \times 10^{-8}$  is smaller than the standard alpha 0.05. Therefore, we can conclude that the average working hours for men is NOT equal to 40. The 95% confidence interval agrees with this assertion as the true population parameter we are testing for does NOT lie within the interval. The results have moderate practical significance as the Cohen's effect size is medium ( $d = 0.2293088$ ). The confidence interval also allows the judgement practical importance through gauging "how far" the true parameter is from the interval.

- c)

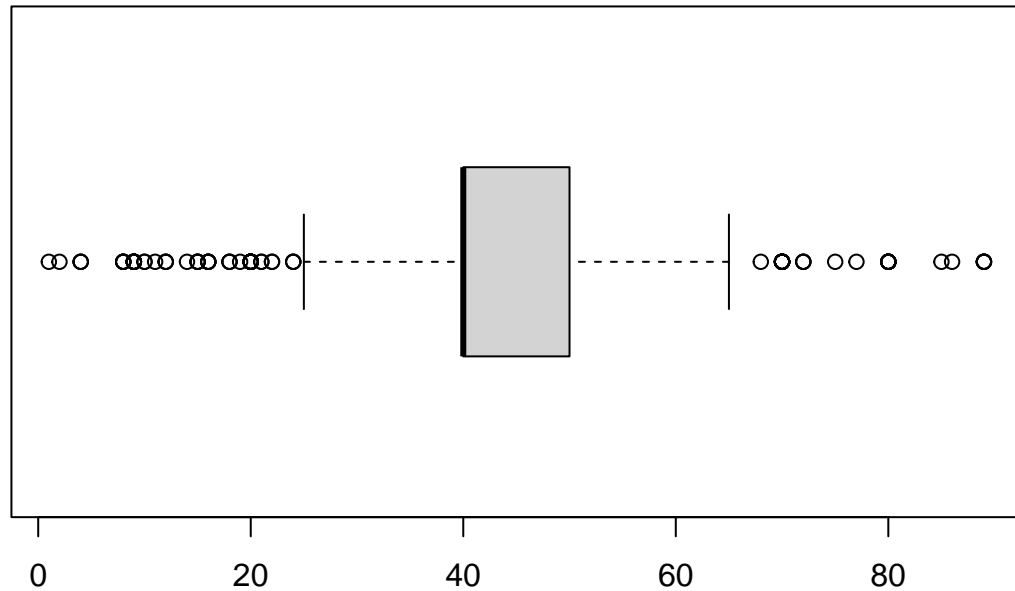
```
boxplot(c(working.hours.gender$Male.Hours, working.hours.gender$Female.Hours), horizontal = T)
title("Working Hours")
```

## Working Hours



```
boxplot(working.hours.gender$Male.Hours, horizontal = T)
title("Male Working Hours")
```

## Male Working Hours



Although the IQR method for detecting potential outliers provides several suggestions, none of these seem particularly outlandish for either boxplot. Therefore, there is no need to be concerned about the legitimacy of the t-results as outliers are not heavily affecting the results.

### Problem 3

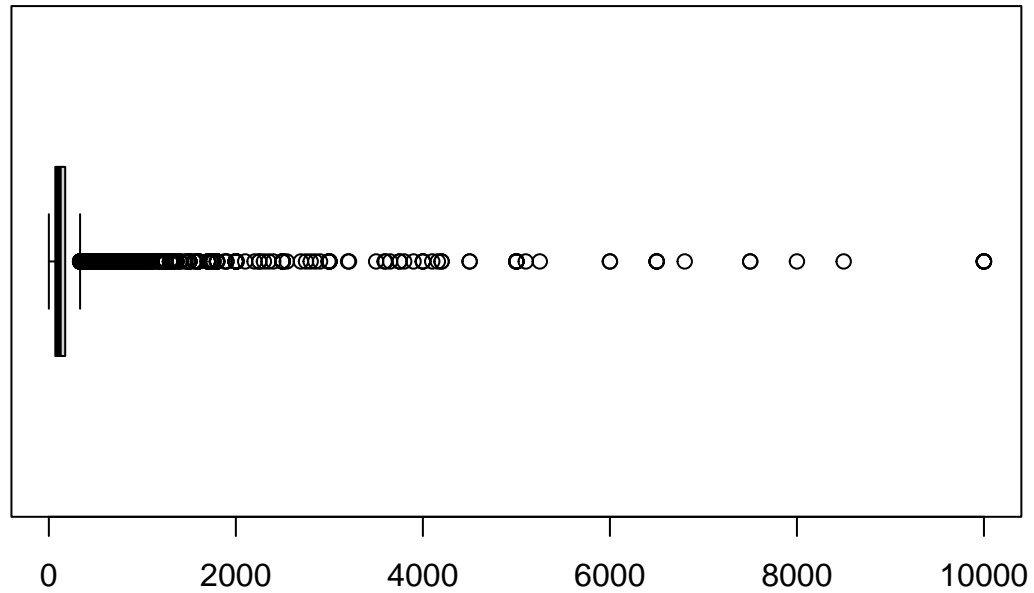
1. We are inferring to all NYC listings when conducting the t-test.  $H_0$  :Average NYC listings price is 150 ( $H_0 : \mu = 150$ ),  $H_a$  :Average NYC listings price is NOT 150 ( $H_a : \mu \neq 150$ ).

```
listings <- read.csv(sprintf("https://docs.google.com/uc?id=%s&export=download", "1au2tYZUtxb1AHckSTkdT"))
t.results <- t.test(listings$price, mu=150)
```

We fail to reject the null hypothesis due to the p-value 0.1745445 being larger than standard alpha 0.05. Therefore, we lack sufficient evidence to conclude that the average NYC listing is not equal to 150. The results also have little practical significance as the Cohen's effect size is extremely small ( $d = 0.0061423$ ). The 95% confidence interval agrees with this assertion as the true population parameter we are testing for lies within the interval.

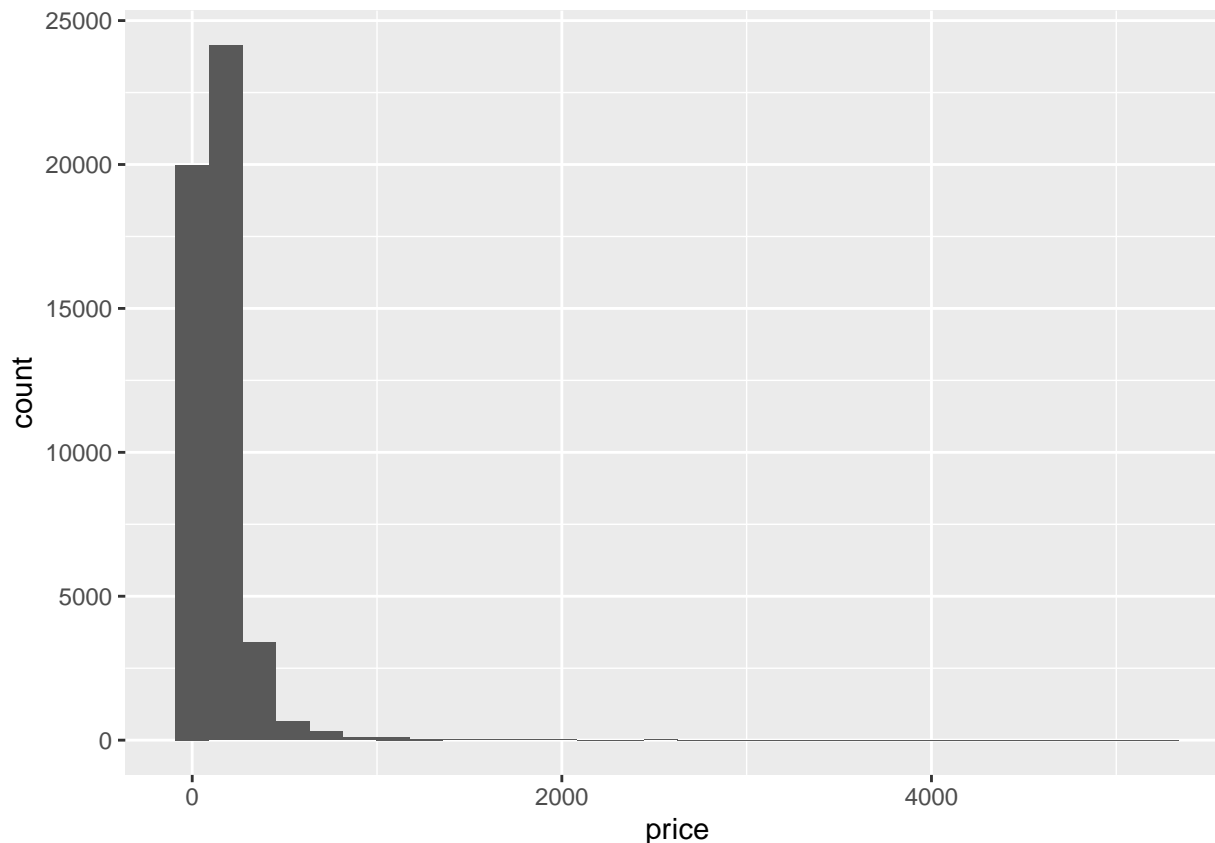
2. There appear to be a few extreme outliers, notably the 10000 value and the other values above ~6000. These values would greatly affect the t-test as it is not robust against outliers, therefore raising concerns regarding the validity of the t-test. Even with the outliers removed, the distribution is still heavily skewed right and does not follow the normal bell-shaped distribution. This lack of normality, however, is not as impactful on the t-test results and therefore is not as much of a concern in regards to t-test validity. This is especially so as there are a large number of samples.

```
library(ggplot2)
boxplot(listings$price, horizontal = T)
```



```
ggplot(listings[listings$price<6000,], aes(price)) + geom_histogram()
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



3.

```
n.sim <- 10000
p.values <- numeric(n.sim)

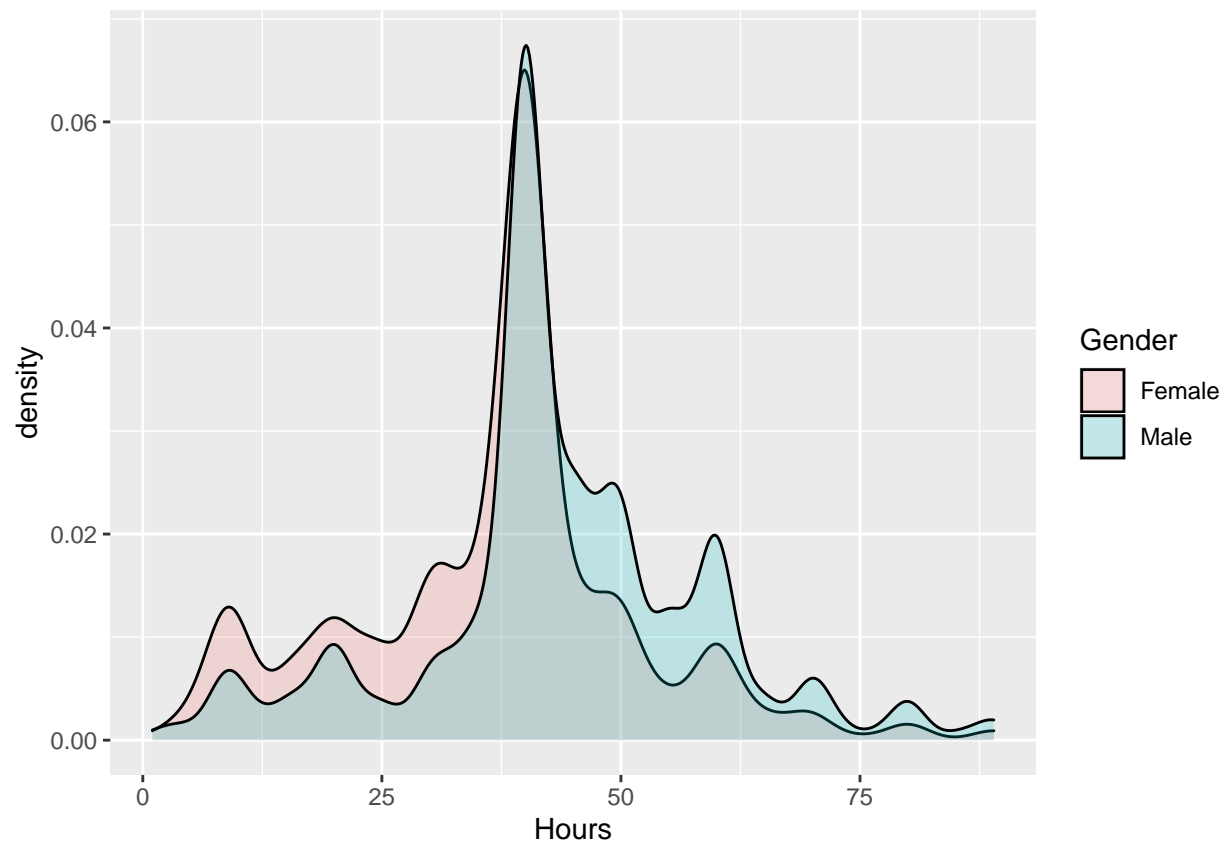
samp.size <- 30
for (j in 1:n.sim){
  x <- sample(listings$price, samp.size)
  p.values[j] <- t.test(x, mu=mean(listings$price, na.rm = T))$p.value
}
```

The percentage of times the p-value was less than 0.05 is much higher than expected with significance level  $\alpha = 0.05$ . We would expect to see a value near 0.05 instead of what we actually have (0.16).

## Problem 4

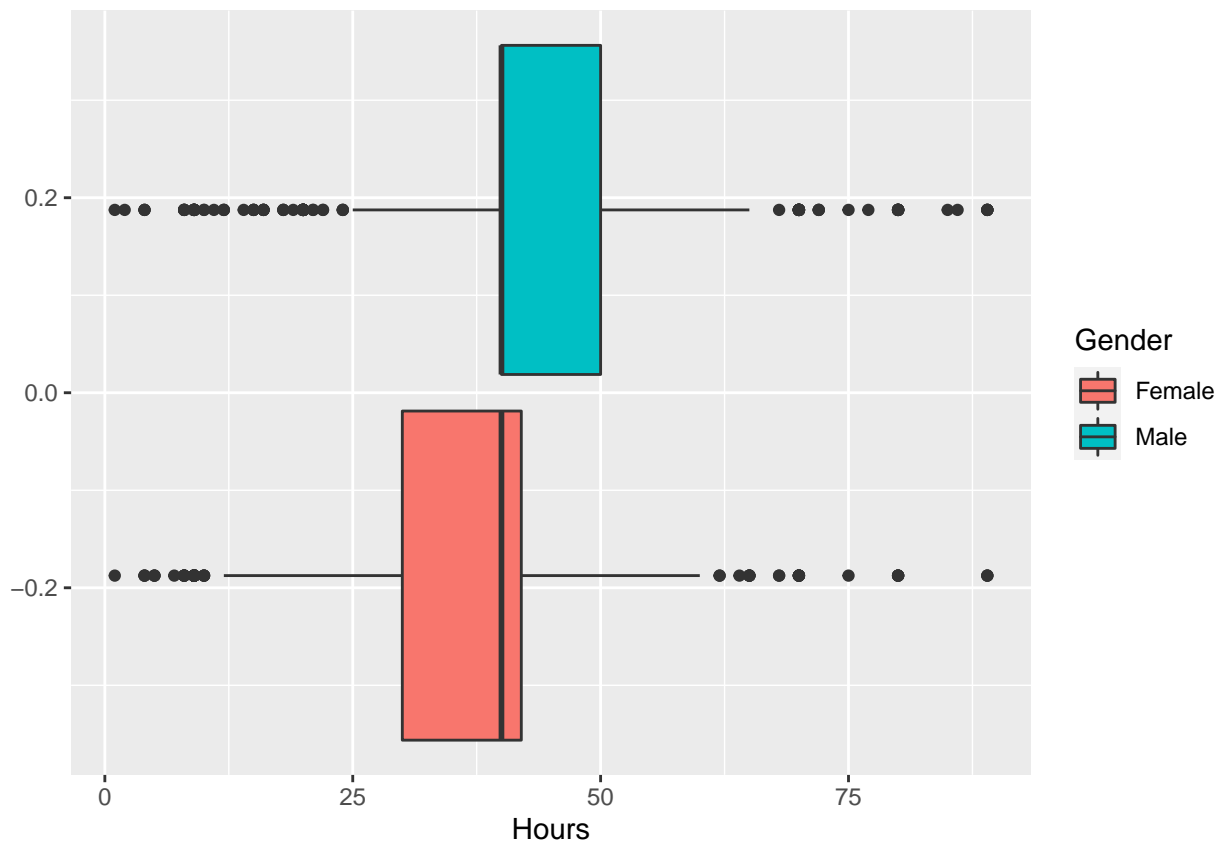
1. The hours for both females and males is not necessarily bell-curved as there is such a large spike in the middle. However, this does not affect the validity of the t-test as it can handle non-normally distributed data, especially when the sample size is large. The 90-ish female work hour is definitely different than the other values but not so far out that I feel it would heavily impact the t-test results.

```
ggplot(working.hours) + geom_density(aes(Hours, fill=Gender), alpha=.2)
```



```
ggplot(working.hours)+ geom_boxplot(aes(Hours, fill=Gender))
```





2.

```
t.results <- t.test(working.hours.gender$Female.Hours, working.hours.gender$Male.Hours)
values <- list(Male=working.hours.gender$Female.Hours, Female= working.hours.gender$Male.Hours)
```

There is a statistically significant difference between the hours worked between female and male adults (p-value =  $5.6802428 \times 10^{-13}$ ), with females working less (females - males  $\in [-8.2394101, -4.7451526]$  with 95% confidence). The results have moderate practical significance because the Cohen's effect size is medium ( $d = 0.4270249$ ).

## Problem 5

**9.49**  $H_0$  : The new drug is safe  $H_a$  : The new drug is not safe

- If  $H_0$  is rejected, we can conclude that the new drug is not safe.
- A Type I error would occur if we concluded that the new drug is not safe wherein fact it is safe.
- If you fail to reject  $H_0$ , we lack the evidence to conclude that the drug is not safe.
- A Type II error would occur if we concluded that the new drug is safe wherein fact it is not safe.

**9.52** Which type of error would usually be considered more serious in the following tests?

- a) A trial to test a murder's defendant's claim innocence, when conviction results in the death penalty: Type I, as you would could be sending an innocent man to death.
- b) Medical diagnostic procedure, such as a mammogram: Type II, as you could fail to pick up on positive cases. With Type I, another test could follow it up. If they receive a negative it is unlikely any further action would occur.

**9.54** Study compare mean score of college entrance exam for students in 2010 vs 1985.  $\mu$ : mean score of all students who took exam in 2010. For random sample of 25,000 students who took exam in 2010,  $\bar{x} = 498$  and  $s = 100$ .

- a)  $TS = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{498 - 500}{100/\sqrt{25000}} \approx -3.1622777$
- b)  $p\text{-value} = P(|T| \geq |TS|) = P(T \geq |TS|) \times 2 \approx 0.0015673$
- c) The test results is statistically significant because the p-value is very small (less than the commonly accepted alphas 0.05 and 0.01). The test results are practically not significant because the Cohen's d value reflects a small effect size.
- d) *Cohen's d*:  $\frac{|\bar{x} - \mu_0|}{s} = \frac{|498 - 500|}{100} = 0.02$

**9.56** A marketing study conducts 60 significance tests about mean and proportions for several groups. Of them, 3 tests are statistically significant at the 0.05 level. It is misleading for the study to only report the significant results and not mention the 57 other tests since 5% (0.05 proportion) of tests are expected to be statistically significant by chance. In this case, it could be assumed that their results were due to chance and are in fact, not significant.

**10.14** 2013 article stated adding energy drinks to alcohol was associated with an elevated blood alcohol content in college students compared to alcohol with no energy drink.

- a) Assuming this conclusion was based on significance test comparing means,  $\mu_A$ : mean blood alcohol content of college students who consume alcohol only,  $\mu_B$ : mean blood alcohol content of college students who consume alcohol mixed with energy drinks.  $H_0: \mu_A - \mu_B = 0$  vs.  $H_a: \mu_A - \mu_B \neq 0$ .
- b) There's a significant difference in blood alcohol content between college students who consume alcohol mixed with energy drinks than those who just consume alcohol only (p-value = 0.01). A confidence interval would provide the plausible range for how different the blood alcohol content actually is. It hits at practical importance and whether  $H_0: \mu = \mu_0$  is badly false.

## 10.24

```
x <- data.frame(students=c(237, 95), mean=c(2.9, 0.1), sd=c(3.6, 0.5), row.names = c("Inhalers", "NonInhalers"))
```

```
##           students mean  sd
## Inhalers         237  2.9 3.6
## NonInhalers        95  0.1 0.5
```

- a)  $H_0: \mu_1 = \mu_2$ ,  $H_a: \mu_1 \neq \mu_2$ :  $TS = \frac{\mu_1 - \mu_2 - 0}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} = \frac{2.9 - 0.1}{\sqrt{3.6^2/237 + 0.5^2/95}} \approx 11.6956245$ . If the population means were equal, it would be nearly impossible by random variable to observe this large a statistic because the observed difference is so large.
- b) At common significance levels, the  $H_0$  would be rejected as the large TS leads to a small p-value. The inhaler group had higher mean nicotine dependence as indicated by the positive sign of the TS indicates which group.

- c) Underlying data is normally distributed, there is independence of nicotine dependence between subjects, and the subjects are randomly sampled.

10.49

```
x <- data.frame(subject=1:3, before=c(150, 165, 135), after=c(130, 140, 120))
```

- a) The three observations for before and the three observations for after are dependent because they involve the same subject.
- b)  $\bar{x}_{before} = 150$ ,  $\bar{x}_{after} = 130$ ,  $\bar{x}_{diff} = 20$ : before mean minus the after mean is equivalent to the mean difference when calculated as  $\text{mean}(\text{before} - \text{after})$ .
- c) We are 95% confident that the average blood sugar drop for subjects after walking briskly for at least half an hour a day is between 7.6 and 32.4.

10.58

```
x <- data.frame(matrix(c(5.08, 5.36, 5.99, 5.98, 5.32, 5.62, 6.03, 6.26, 5.44, 5.68), nrow = 2), row.names = c("Before", "After"))
```

```
##           X1    X2    X3    X4    X5
## Before Inhaler 5.08 5.99 5.32 6.03 5.44
## After Inhaler  5.36 5.98 5.62 6.26 5.68
```

- a) The FVC improved by 0.208 after using the inhaler, on average.
- b)

```
ci <- t.test(as.numeric(x[1,]), as.numeric(x[2,]), paired=T)$conf.int
ci
```

```
## [1] -0.36343134 -0.05256866
## attr(,"conf.level")
## [1] 0.95
```

Subjects see an average improvement of FVC after using the inhaler between -0.3634313 and -0.0525687, with 95% confidence.

- c)

```
ci <- t.test(as.numeric(x[1,]), as.numeric(x[2,]))$conf.int
ci
```

```
## [1] -0.7739891  0.3579891
## attr(,"conf.level")
## [1] 0.95
```

If incorrectly assuming independence samples, one would find subjects see an average improvement of FVC after using the inhaler between -0.7739891 and 0.3579891, with 95% confidence.