

# t-Test Examples for Math 338

## Example 1

The alcohol by volume (ABV) content of a brand of beer is advertised to be 5%. However, a lawsuit claims that the beer is watered down (lower ABV content than advertised). Twelve independent labs test a can of the beer and measure the ABV content of their can. The ABV content of each can is found in the file beer.csv. Using a default significance level of 0.05, can we claim that the beer is watered down?

## Solution

First we define the cases and variables of interest in our study. It often helps to import and view the dataset to see what was recorded:

```
library(readr)
beer <- read_csv("~/Math 338 Master Examples/t Test/beer.csv")
View(beer) # view the data
```

- Each case is an independent measurement of the beer. There are 12 cases in the dataset.
- The only variable of interest is the measured ABV content. It is numerical and the units are percent.

Since we have a single numerical variable, this suggests we should use a t-test to test the claim that the beer is watered down.

Since we do not have a suggested alternative value for the ABV content of the beer, we will use the NHST framework. The claim that the beer is watered down is equivalent to saying that the mean ABV content is less than 5%. This leads to the null-alternative hypothesis pair shown below:

$$H_0 : \mu = 5$$

$$H_a : \mu < 5$$

Now we are ready to perform exploratory analysis on the data:

```
summary(beer)

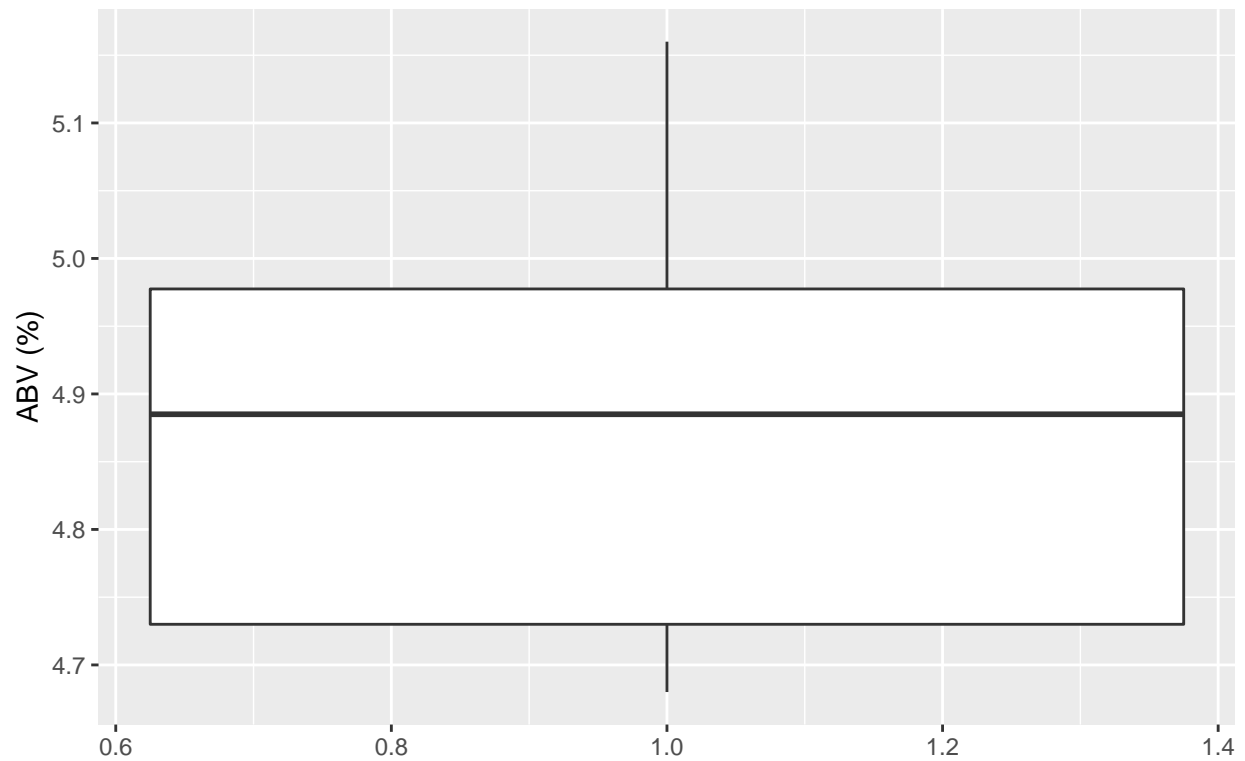
##           beer
##  Min.      :4.680
## 1st Qu.:4.730
##  Median :4.885
##   Mean  :4.875
## 3rd Qu.:4.978
##   Max.  :5.160

sd(beer$beer) # get the standard deviation

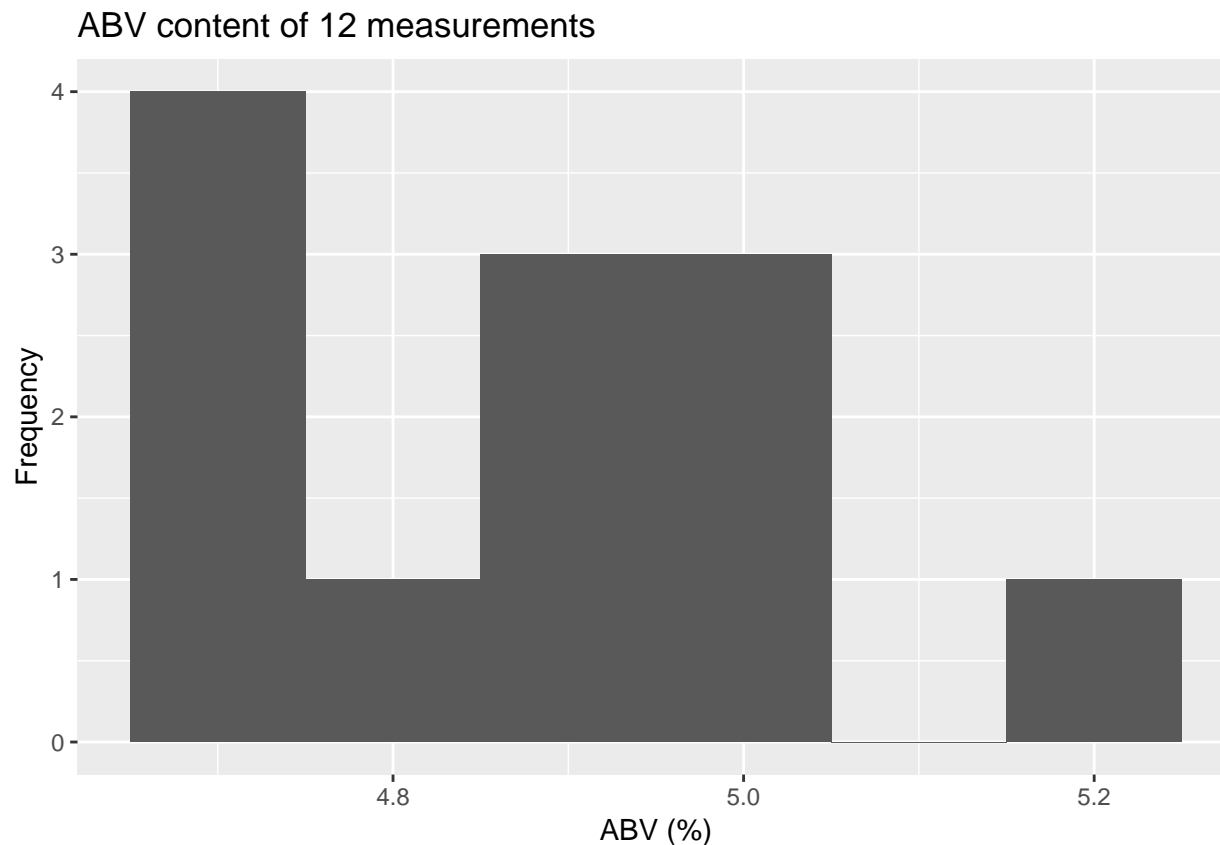
## [1] 0.1524646

library(ggplot2)
# make a boxplot
beer_boxplot <- ggplot(beer, aes(y = beer)) +
  geom_boxplot(aes(x = 1)) + labs(title = "ABV content of 12 beers",
    x = "", y = "ABV (%)")
print(beer_boxplot)
```

ABV content of 12 beers



```
# make a histogram
beer_histogram <- ggplot(beer, aes(x = beer)) +
  geom_histogram(binwidth = 0.1, center = 5) +
  labs(title = "ABV content of 12 measurements",
        x = "ABV (%)", y = "Frequency")
print(beer_histogram)
```



The histogram of sample data suggests that the population distribution might not be normal - it might be skewed right - and there might be an outlier around 5.2. However, we can assume that the measurement error is normally distributed, so this is more likely to be small-sample-size wackiness rather than truly a skewed-right population distribution. So we can assume that a t-test will work.

```
t.test(x = beer$beer, mu = 5, alternative = "less")
```

```
##
## One Sample t-test
##
## data: beer$beer
## t = -2.8401, df = 11, p-value = 0.00804
## alternative hypothesis: true mean is less than 5
## 95 percent confidence interval:
##      -Inf 4.954042
## sample estimates:
## mean of x
##      4.875
```

From the output, we observe that the test statistic  $t = -2.84$  comes from a t-distribution with 11 degrees of freedom, and the p-value is 0.008. Thus, at a significance level of 0.05, we can reject our null hypothesis. We have sufficient evidence to suspect that the mean alcohol content of the beer is lower than advertised.

## Example 2

Is a sample of 12 measurements enough to detect that the true mean alcohol content of the beer is 4.9%? Assume your sample standard deviation would be 0.15 percentage points.

### Solution

Now we are in Neyman-Pearson framework because we now have a specific alternative value of 4.9% to test against. Thus we write:

$$H_0 : \mu = 5$$
$$H_1 : \mu = 4.9$$

The command to compute power is `power.t.test()`. This command takes five or six arguments:

- 1) `n` is the sample size.
- 2) `delta` is the difference between the hypothesized value under  $H_0$  and the hypothesized value under  $H_1$ . In the code below I write it as `abs( $\mu_0 - \mu_1$ )`, but the important part is to make it a non-negative number.
- 3) `sd` is the estimated sample standard deviation of the sample you will obtain. (If you have real data already, just use the sd of the real data)
- 4) `type` is the type of t-test. It should be “one.sample” for a one-sample test, “paired” for a matched pairs test, and “two.sample” for a two independent samples test.
- 5) `alternative` is “one.sided” for a one-sided test and “two.sided” for a two-sided test (where the critical region is split half on the left side and half on the right side).
- 6) `sig.level` is the significance level. Note that `power.t.test()` uses `sig.level` interchangeably with  $\alpha$ . By default this value is 0.05, so we only need to include it if we use a different  $\alpha$  value.

```
power.t.test(n = 12, delta = abs(5 - 4.9),
             sd = 0.15, type = "one.sample", alternative = "one.sided",
             sig.level = 0.05)
```

```
##
##      One-sample t test power calculation
##
##              n = 12
##            delta = 0.1
##             sd = 0.15
##          sig.level = 0.05
##             power = 0.6981908
##      alternative = one.sided
```

About 69.8% of the time, when the beer is watered down to 4.9% ABV, we can randomly sample a twelve-pack of beer cans and reject the null hypothesis that the mean ABV is 5%.

Note that the Rguroo values are slightly different due to how the sd is treated.

## Example 3

Researchers want to determine whether their exercise program builds stronger bones. They measure the total bone body mineral content (TBBMC) of 25 participants before and after six weeks of the program. A simulated TBBMC for each participant (in grams) before and after the program can be found in the file `tbbmc.csv`. Determine whether the researchers can claim that the exercise program increases TBBMC levels.

## Solution

First we define the cases and the variables in our study. To help us, we can import our dataset and look to see what was recorded.

```
# since we already loaded the readr
# package we don't need to load it again
tbbmc <- read_csv("~/Math 338 Master Examples/t Test/tbbmc.csv")
View(tbbmc) # view the data
```

- Each case is a person. There are 25 cases in the dataset.
- The gender is a categorical variable and seems not to have any relevance to the researchers' question.
- The before variable represents the TBBMC value before the program and the after variable represents the TBBMC value after the program.
- Both before and after are numerical variables and measure the same thing under different conditions.

Since we have the two numerical variables measuring the same real-world thing on the same participants under two different conditions ("before" and "after"), we can use a matched-pairs t-test to test the claim that the program increases TBBMC levels.

Let's define population 1 to be the "after exercise" population and population 2 to be the "before exercise" population. Then we have:

$$H_0 : \mu_{after-before} = 0$$

$$H_a : \mu_{after-before} > 0$$

(we want to show that the "after exercise" group has a higher mean TBBMC than "before exercise")

Alternatively, if we defined population 1 to be the "before exercise" population and population 2 to be the "after exercise" population, we would have:  $H_0 : \mu_{before-after} = 0$

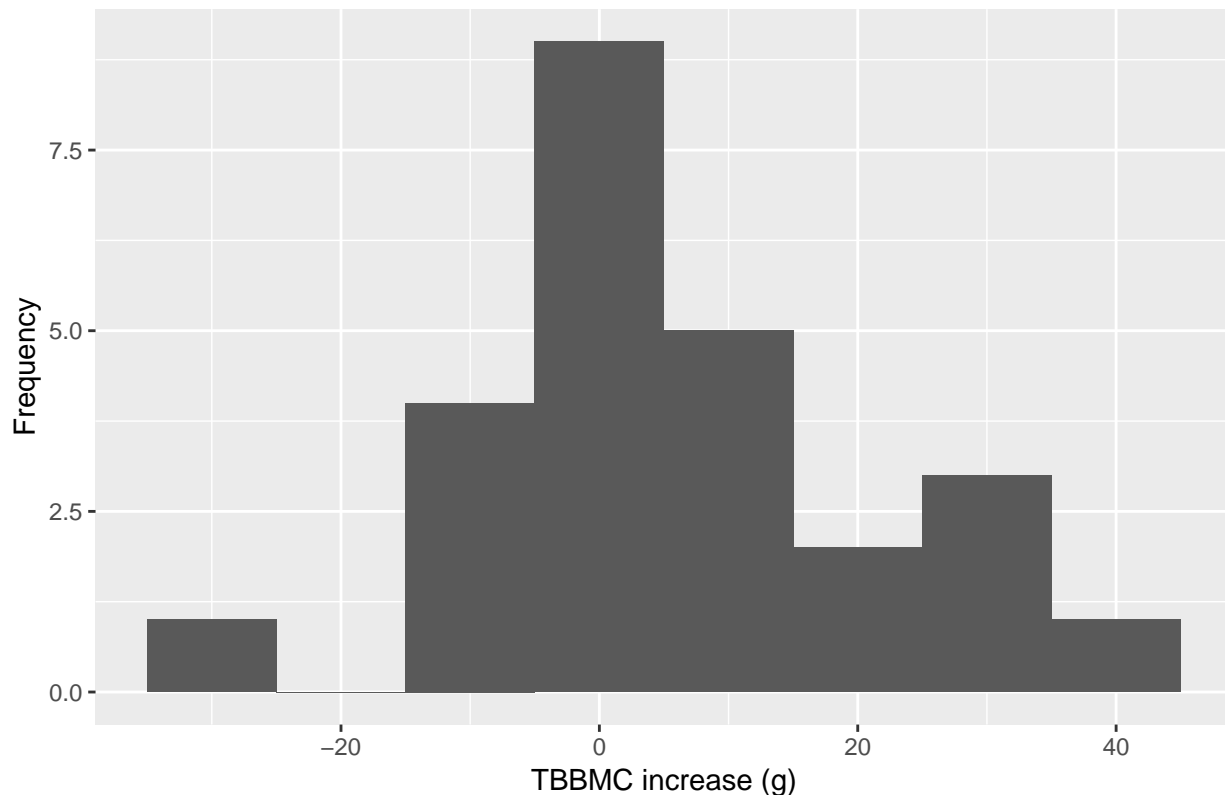
$$H_a : \mu_{before-after} < 0$$

(we want to show that the "before exercise" group has a lower mean TBBMC than "after exercise")

Let's first do some exploratory data analysis. Since this is a paired data set, we are less interested in the distribution of the individual variables "before" and "after" and more interested in their difference.

```
library(dplyr)
tbbmc <- tbbmc %>% mutate(diff = after -
  before)
tbbmc_histogram <- ggplot(tbbmc, aes(x = diff)) +
  geom_histogram(binwidth = 10, center = 500) +
  labs(title = "Increase in TBBMC of 25 Subjects",
    x = "TBBMC increase (g)", y = "Frequency")
print(tbbmc_histogram)
```

### Increase in TBBMC of 25 Subjects



This distribution looks a little bit skewed to the right, but not so terribly skewed that a t-test would obviously not be good with a sample size of 25. So we are probably okay to run the matched pairs t-test.

There are two different ways to run a paired t-test in R. First, we can use the original variables.

```
t.test(x = tbbmc$before, y = tbbmc$after,
       alternative = "less", conf.level = 0.95,
       paired = TRUE) # less because we are looking at before - after here
```

```
##
## Paired t-test
##
## data: tbbmc$before and tbbmc$after
## t = -2.0334, df = 24, p-value = 0.02661
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -1.015023
## sample estimates:
## mean of the differences
##                -6.4
```

Second, we can do a one-sample t-test on the difference.

```
t.test(x = tbbmc$diff, alternative = "greater",
       conf.level = 0.95) # greater because we defined diff as after - before
```

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

```
## data:  tbbmc$diff
## t = 2.0334, df = 24, p-value = 0.02661
## alternative hypothesis: true mean is greater than 0
## 95 percent confidence interval:
##  1.015023      Inf
## sample estimates:
## mean of x
##      6.4
```

We get a t test statistic of either -2.033 or 2.033, depending on how we defined the difference. In either case, the t test statistic corresponds to a p-value of 0.0266. Thus, at the 0.05 significance level, we have a significant result and should reject our null hypothesis. We conclude that the exercise program does increase TBBMC, which results in stronger bones.

## Example 4

We wish to determine whether there is GPA difference between biology majors and psychology majors. We obtain a sample of 20 biology majors and, independently, a sample of 30 psychology majors. A simulated GPA for each student can be found in the file gpa.csv. Test the claim that there is no difference in the population mean GPA of the two majors.

### Solution

First we define the cases and the variables in our study. To help us, we can import our dataset and look to see what was recorded.

```
library(readr)
gpa_data <- read_csv("~/Math 338 Master Examples/t Test/gpa.csv")
View(gpa_data) # view the data
```

- Each case is a student. There are 50 cases in the dataset.
- The variable major is a categorical variable that can take levels “bio” and “psych”. We can think of this variable as designating the population the case comes from, or as an explanatory variable that helps explain differences in GPA.
- The variable gpa is a numerical variable and is the response variable of interest, representing the GPA of the student.

Since we have one numerical variable measured on participants in two different groups, we can use a two-sample t-test to test whether there is a difference in the population mean GPA of the two majors.

Let’s define Population 1 to be biology majors and Population 2 to be psychology majors; then our null hypothesis is  $H_0 : \mu_1 = \mu_2$ , or equivalently,  $H_0 : \mu_1 - \mu_2 = 0$ . Since we are simply testing to try to find a difference, our alternative hypothesis in NHST framework is  $H_a : \mu_1 \neq \mu_2$ , or equivalently,  $H_a : \mu_1 - \mu_2 \neq 0$ .

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_a : \mu_1 - \mu_2 \neq 0$$

In general, when defining populations, the R default is to define the order of populations in alphabetical order. Our two levels are “bio” and “psych”; since “bio” comes first in alphabetical order it will be Population 1. This is an easy convention to remember and will make doing the test in R easier.

Let’s do some exploratory data analysis. Here, we will need to split up the data set. We do this in a fancy way using some ggplot2 and dplyr tricks:

```

# Use dplyr library already loaded.
# group_by groups the data frame by the
# values of the categorical variable and
# summarize produces the relevant
# statistic for the indicated numerical
# variable, so this code produces the
# number of observations, minimum,
# maximum, mean, and sd of gpa for each
# major
gpa_data %>% group_by(major) %>% summarize(count = n(),
  min = min(gpa), max = max(gpa), mean = mean(gpa),
  sd = sd(gpa))

```

```

## # A tibble: 2 x 6
##   major count   min   max mean   sd
##   <chr> <int> <dbl> <dbl> <dbl> <dbl>
## 1 bio      20  2.1   3.75  2.92 0.430
## 2 psych    30  2.41  3.83  3.12 0.314

```

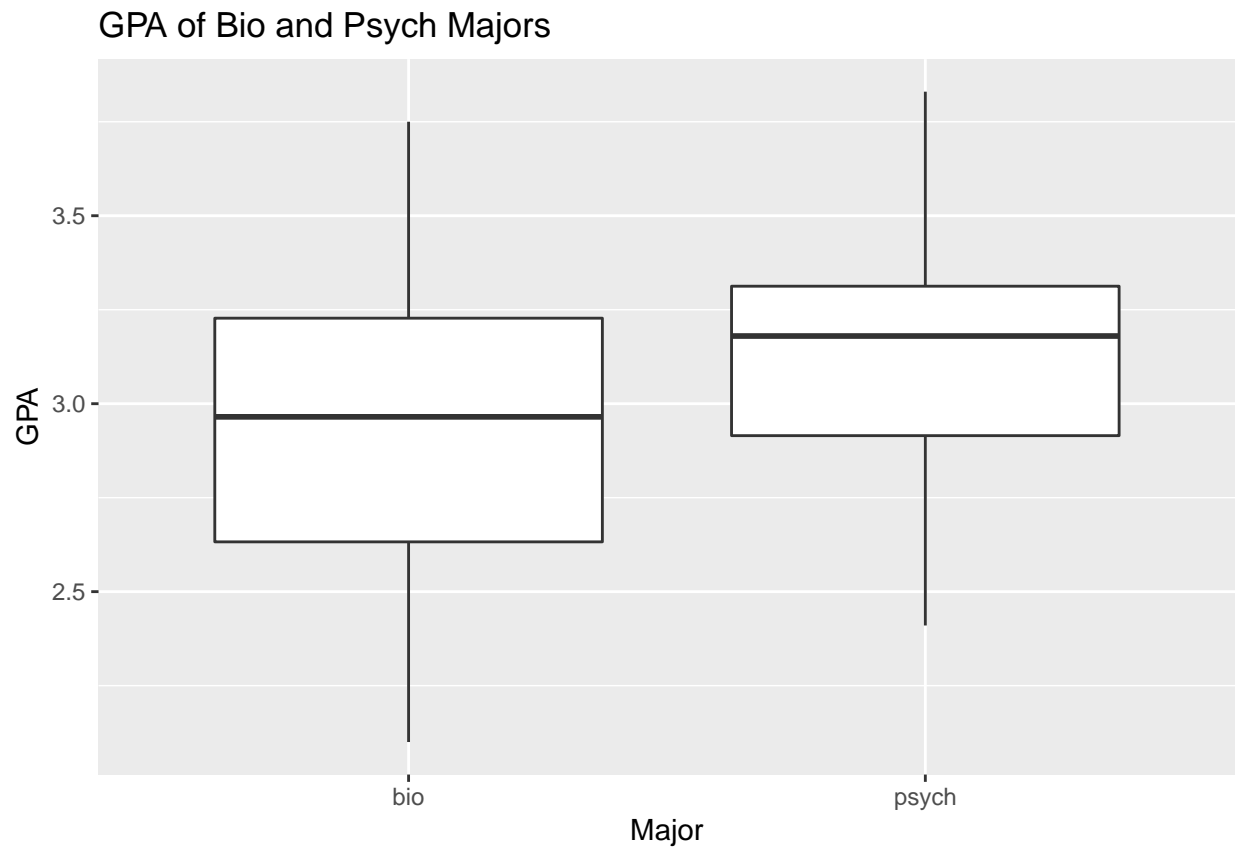
We see that our sample of 20 bio majors have GPAs ranging from 2.10 to 3.75, with a mean of 2.92 and a sd of 0.43. Our sample of 30 psych majors have GPAs ranging from 2.41 to 3.83, with a mean of 3.12 and a sd of 0.31. Let's see if there are any obvious shape problems:

```

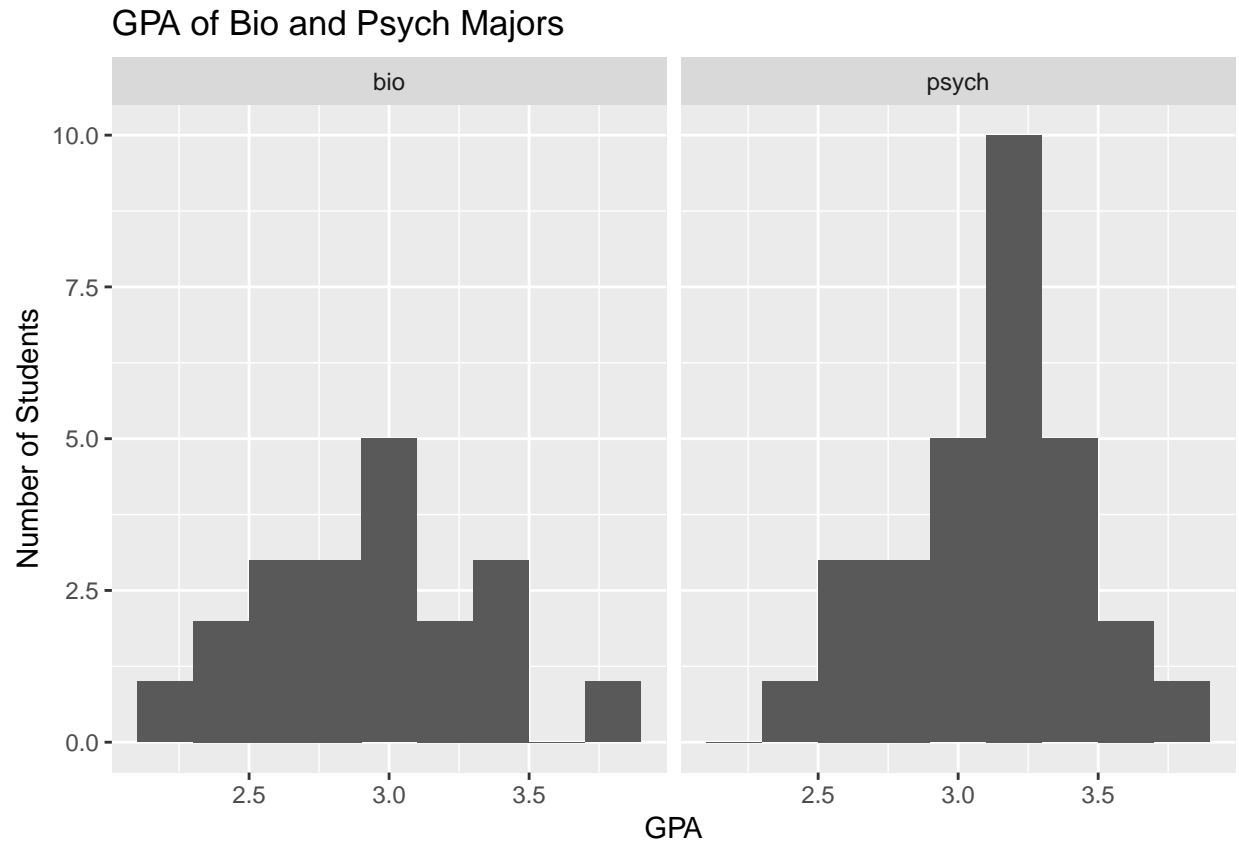
# Use ggplot2 library already loaded
gpa_boxplot <- ggplot(gpa_data, aes(x = major,
  y = gpa)) + geom_boxplot() + labs(title = "GPA of Bio and Psych Majors",
  x = "Major", y = "GPA")
print(gpa_boxplot)

```





```
gpa_histogram <- ggplot(gpa_data, aes(x = gpa)) +  
  geom_histogram(center = 3, binwidth = 0.2) +  
  facet_wrap(~major) + labs(title = "GPA of Bio and Psych Majors",  
    x = "GPA", y = "Number of Students")  
print(gpa_histogram)
```



According to the graphs, the bio majors GPA is roughly normal. The psych majors GPA is skewed to the left, but not so skewed as to make our t-procedures unreliable.

There are a couple of different ways to run this test in R. First, we can run a two independent samples t-test using subsets of our gpa variable.

```
t.test(x = gpa_data$gpa[which(gpa_data$major ==
  "bio")], y = gpa_data$gpa[which(gpa_data$major ==
  "psych")], alternative = "two.sided") # by default, the t-test is unpaired
```

```
##
## Welch Two Sample t-test
##
## data: gpa_data$gpa[which(gpa_data$major == "bio")] and gpa_data$gpa[which(gpa_data$major == "psych")]
## t = -1.77, df = 32.245, p-value = 0.08618
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.42650717 0.02984051
## sample estimates:
## mean of x mean of y
## 2.922000 3.120333
```

This kind of subsetting is a bit verbose. When we have data like this, it's usually easier to do a two independent samples t-test using a "formula" interface.

```
# Formula interface: Response ~ Group,
# data = Data_Frame
t.test(gpa ~ major, data = gpa_data, alternative = "two.sided")
```

```
##
## Welch Two Sample t-test
##
## data:  gpa by major
## t = -1.77, df = 32.245, p-value = 0.08618
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.42650717  0.02984051
## sample estimates:
##   mean in group bio mean in group psych
##           2.922000           3.120333
```

We get a t test statistic of -1.77, which corresponds to a p-value of 0.086. At the 0.05 significance level, we fail to reject our null hypothesis. Thus, we conclude that it is reasonable to assume the two programs have equal mean GPAs.