ST 314 - Data Analysis 5

Ivan Timothy Halim

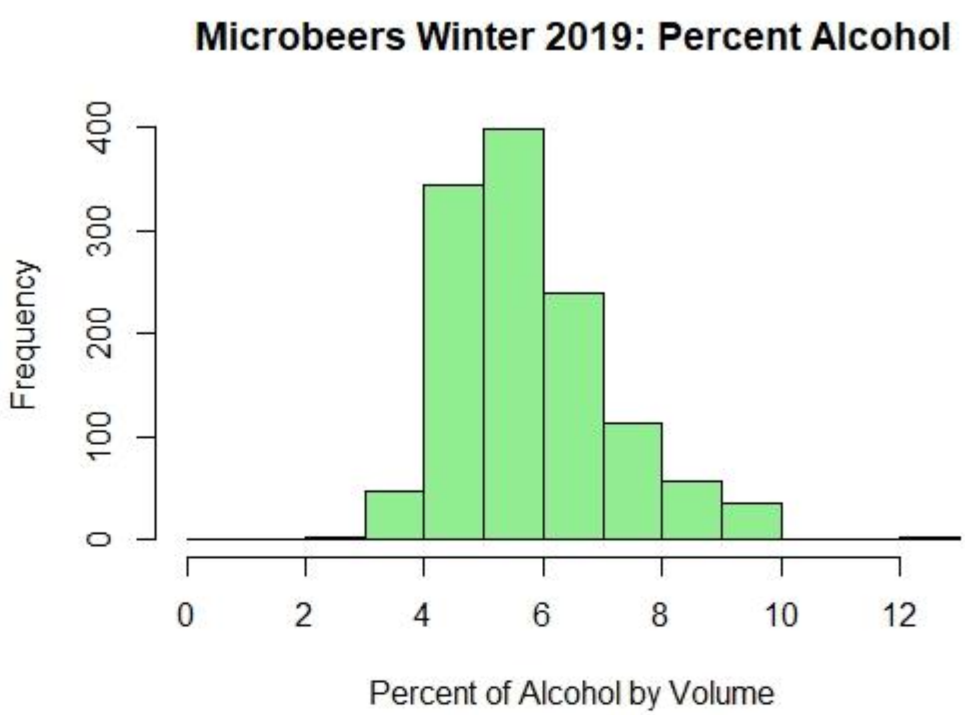
February 5, 2019

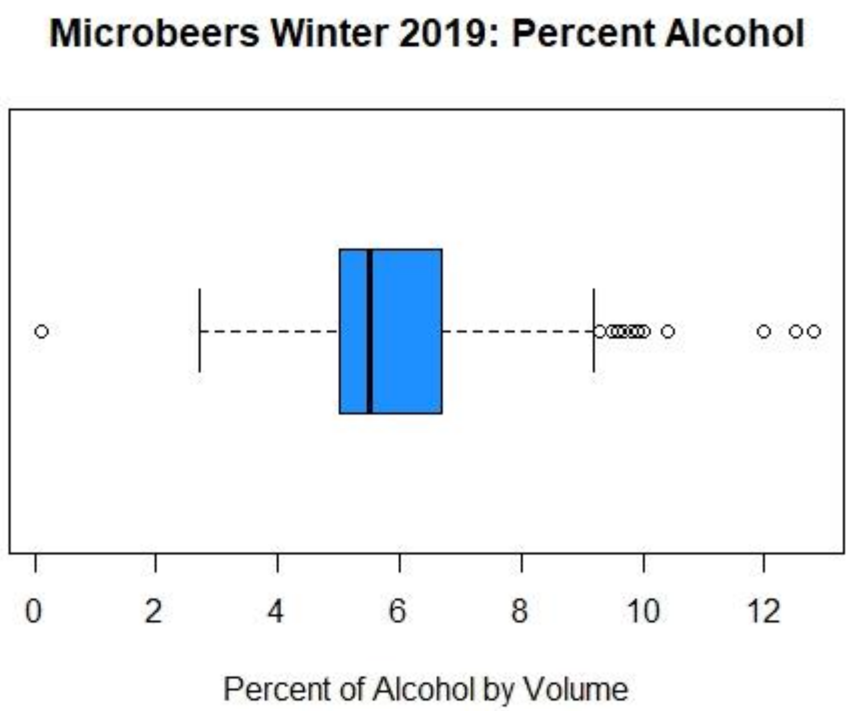
Part 1

a. The parameters that we're interested in are:
   - The Sample Mean, $\overline{x}$
   - The Sample Standard Deviation, s
   - The Degrees of Freedom, df

b. $H_0: \mu = 5$
   $H_a: \mu \neq 5$

c.





The median is greater than 5 and the data is positively skewed. When the data is positively skewed, the mean is greater than the median. Since the median is greater than 5 then the mean must be greater than 5.

d. $\overline{x} = 5.904$
$s = 1.374$

e. Samples are taken randomly from around the United States, the first condition is met.
n is sufficiently large for $\overline{x}$ to be approximately normal ($n = 1244$).
We don't know the population standard deviation, so the third condition is met.

f. $t = \frac{\overline{x}-\mu}{\frac{s}{\sqrt{n}}} = \frac{5.904-5}{\frac{1.374}{\sqrt{1244}}} = 23.21$

g. The p-$value$ is a very small number approaching zero. It is two-sided.

```
2*(1-pt(23.21,1243))

## [1] 0
```

h. $t_{1243,95\%} \approx 1.960$

$$SE_{\overline{x}} = \frac{s}{\sqrt{n}} = \frac{1.374}{\sqrt{1244}} = 0.039$$

$$95\%\ CI = \overline{x} \pm t_{n-1,\frac{\alpha}{2}} SE_{\overline{x}}$$

$$= 5.904 \pm (1.96 \times 0.039) = (5.828, 5.980)$$

i.

```
t.test(microbeers$abv, mu = 5,
       alternative = "two.sided",
       conv.level = 0.95)

##
##  One Sample t-test
##
## data:  microbeers$abv
## t = 23.21, df = 1243, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 5
## 95 percent confidence interval:
##  5.827753 5.980607
## sample estimates:
## mean of x
##   5.90418
```

There is overwhelming evidence to conclude that the average percent of alcohol differs from 5%.

We can reject the null hypothesis based on a significance level of 0.05 ($t = 23.21, df = 1243, p\text{-}value < 2.2 \times 10^{-16}$).

The 95% CI estimates the average percent of alcohol to be between 5.828% and 5.981% with a best guess of 5.904%.

The National Institute of Health estimates wrongly.

Part 2

a. The spread for the online boxplot is from 0 min to 60 min and the spread for the in-class boxplot is from 0 min to 160 min. The in-class range is 160 min which is almost three times more than the online range (60 min). If we don't include the outliers, the in-class range (52 – 0 = 52 min) is still 25% greater than the online range (40 – 0 = 40 min).

   This might suggest that in-class students spend more time doing schoolwork than online students.

b. $H_0: \mu_1 - \mu_2 = 0$
   $H_a: \mu_1 - \mu_2 \neq 0$

   The alternative is two-sided.

c. Samples are taken only from Winter 2019 students, so there's a convenience bias.
   n_1 and n_2 are sufficiently large for $\overline{x_1}$ and $\overline{x_2}$ to be approximately normal ($n_1 = 150 > 30, \mathrm{n}_2 = 67 > 30$).
   Populations are independent (online students and in-class students).

d. There is not enough evidence to conclude that online students and in-class students spend different times doing schoolwork, on average.

   We fail to reject the null hypothesis based on a significance level of 0.05 ($t = 0.956$, $df = 168.89, p\text{-}value = 0.3402$).

   The 95% CI estimates the average difference in time spent to be between -2.128 min and 6.128 min with a best guess of 2.00 min.

   Online students study just as hard as in-class students.