Interval estimate

Protim Ganguly

16 November 2018

```
library(infer)
library(tidyverse)

## -- Attaching packages ----------------------------------------------------
---------------------------- tidyverse 1.2.1 --

## v ggplot2 3.0.0     v purrr   0.2.5
## v tibble  1.4.2     v dplyr   0.7.7
## v tidyr   0.8.1     v stringr 1.3.1
## v readr   1.1.1     v forcats 0.3.0

## -- Conflicts -------------------------------------------------------------
----------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

library(tolerance)

chipspeed.df <- read.csv("C:/Users/prodi/Desktop/Books/LABS/Statistics 1/chip
speed.csv")
glimpse(chipspeed.df)

## Observations: 100
## Variables: 2
## $ Speed <dbl> 505.19, 489.58, 491.83, 495.40, 497.53, 505.72, 496.39, ...
## $ Chip  <fct> New, New, New, New, New, New, New, New, New, New, New, N...
```

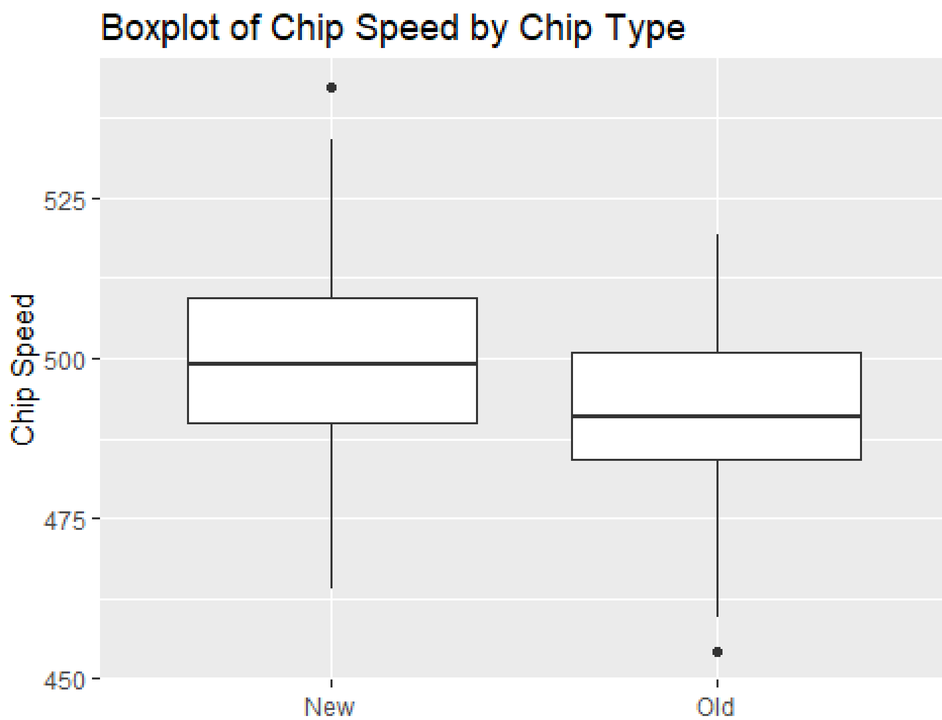
Summary Statistics

```
chipspeed.df %>% group_by(Chip) %>%
  summarize(Mean=mean(Speed), SD= sd(Speed))

## # A tibble: 2 x 3
##   Chip   Mean    SD
##   <fct> <dbl> <dbl>
## 1 New    500.  17.2
## 2 Old    491.  13.5

chipspeed.df %>% filter(Chip %in% 'New') %>%
  summarize(A.Mean=mean(Speed), A.SD= sd(Speed))

##     A.Mean     A.SD
## 1 499.7442 17.17434
```

```
chipspeed.df %>% filter(Chip %in% 'Old') %>%
  summarize(B.Mean=mean(Speed), B.SD= sd(Speed))

##     B.Mean    B.SD
## 1 491.4836 13.51641
```

Boxplot

```
ggplot(chipspeed.df, aes(x = Chip, y = Speed)) +
        geom_boxplot() +
  ggtitle("Boxplot of Chip Speed by Chip Type") +
  ylab("Chip Speed") +
  xlab("")
```



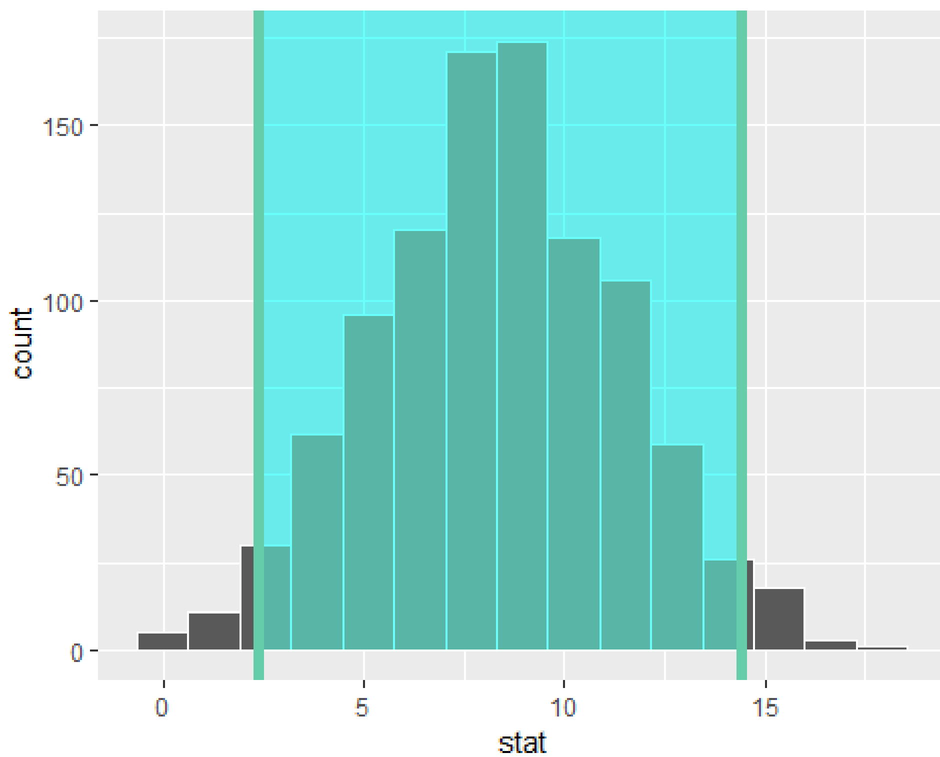
t-tes tand 95% Confidence Interval for difference in the population mean

```
# t test

chipspeed.df %>%
t_test(Speed ~ Chip, order = c("New", "Old"),
       alternative = "two_sided",
       mu=0,
       conf_level = 0.95,
       paired=FALSE,
       var.equal=FALSE)
```

```
## # A tibble: 1 x 6
##   statistic  t_df p_value alternative lower_ci upper_ci
##       <dbl> <dbl>   <dbl> <chr>          <dbl>    <dbl>
## 1      2.67  92.9 0.00889 two.sided       2.12     14.4
```

95% Bootstrap CI for difference in means

```
chipspeed.boot <- chipspeed.df %>%
  specify(response = Speed, explanatory = Chip) %>%
  generate(reps = 1000, type = "bootstrap") %>%
  calculate(stat = "diff in means", order = c("New", "Old"))

percentile_ci <- get_ci(chipspeed.boot)

percentile_ci

## # A tibble: 1 x 2
##   `2.5%` `97.5%`
##    <dbl>   <dbl>
## 1   2.39    14.4

chipspeed.boot %>% visualize(endpoints = percentile_ci, direction = "between"
)
```



Tolerance Intervals

Speed New

```
New.Speed <- chipspeed.df %>% filter(Chip %in% 'New') %>% select(Speed) %>% a
s.data.frame()

normtol.int(New.Speed$Speed, alpha = 0.05, P = 0.95, side = 2)

##   alpha    P    x.bar 2-sided.lower 2-sided.upper
## 1  0.05 0.95 499.7442      458.8391      540.6493
```

Speed Old

```
Old.Speed <- chipspeed.df %>% filter(Chip %in% 'Old') %>% select(Speed) %>% a
s.data.frame()

normtol.int(Old.Speed$Speed, alpha = 0.05, P = 0.95, side = 2)

##   alpha    P    x.bar 2-sided.lower 2-sided.upper
## 1  0.05 0.95 491.4836      459.2908      523.6764
```

Conclusion -

We have a 2 sample pair problem. Here the Mean ChipSpeed for New Chips is more than Mean ChipSpeed for Old Chips which shows that NewChips have higher speed on average. From Box Plot we can see the median in both the New and Old Chip sets as well as the outliers for both the groups. From the t_test we can infer that from the sample of 50 New Chip sets + 50 Old Chip sets, that over the whole population the new chips will have a higher speed than the old chips on average which is given by the Confidence Interval (2.12 , 14.4). However our t-test is valid with the assumption that our population has a normal distribution which we will never know in advance. Also we have two outliers in both the groups of chips. Hence we do bootstrapping to take samples with replacement to get BootStrap Interval (BI) which is nearly same as the Interval for CI and hence our assumptions are valid. There is evidence that there is around 1.6% improvement in speed in New Chips on average. However not all New chip speeds will have more speed than the old chips. From Tolerance interval we can be 95% confident that around 95% of the new chips will have a speed between 458 and 540.

Null Hypothesis (H0)- The mean speed of both New and Old Chips are Equal.

Alternate Hypothesis (H1) - The mean speed of both New and Old Chips are different.

Since we have a p-value of 0.008 which is less than 0.05 we have evidence to rule out the null hypothesis and accept our alternate hypothesis.