One Sample t-test

Youngjin\_Lee

2022 2 9

Table of Contents

# 일표본(One Sample) t-test

통계분석 차이검정 : 일표본(One Sample) t-test B아이스크림회사에서 판매하는 아이스크림 중 파인트의 무게는 320g이다. 그러나 K대학 앞에 있는 점포에서 파는 아이스크림의 무게가 320g이 아니라는 소비자들의 불만이 있었다. 이에 따라 소비자단체에서는 B아이스크림회사에서 만든 아이스크림이 320g인지를 검사하고자 한다.

# 1.기본 package 설정, library 로드

library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5 v purrr 0.3.4  
## v tibble 3.1.6 v dplyr 1.0.7  
## v tidyr 1.1.4 v stringr 1.4.0  
## v readr 2.1.1 v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(tidymodels)

## Registered S3 method overwritten by 'tune':  
## method from   
## required\_pkgs.model\_spec parsnip

## -- Attaching packages -------------------------------------- tidymodels 0.1.4 --

## v broom 0.7.12 v rsample 0.1.1   
## v dials 0.0.10 v tune 0.1.6   
## v infer 1.0.0 v workflows 0.2.4   
## v modeldata 0.1.1 v workflowsets 0.1.0   
## v parsnip 0.1.7 v yardstick 0.0.9   
## v recipes 0.1.17

## -- Conflicts ----------------------------------------- tidymodels\_conflicts() --  
## x scales::discard() masks purrr::discard()  
## x dplyr::filter() masks stats::filter()  
## x recipes::fixed() masks stringr::fixed()  
## x dplyr::lag() masks stats::lag()  
## x yardstick::spec() masks readr::spec()  
## x recipes::step() masks stats::step()  
## \* Learn how to get started at https://www.tidymodels.org/start/

library(rstatix)

##   
## 다음의 패키지를 부착합니다: 'rstatix'

## The following objects are masked from 'package:infer':  
##   
## chisq\_test, prop\_test, t\_test

## The following object is masked from 'package:dials':  
##   
## get\_n

## The following object is masked from 'package:stats':  
##   
## filter

library(skimr)

# 2.데이터 불러오기

ost\_tb <- read\_csv('data\\OST.csv',   
 col\_names = TRUE,  
 locale=locale('ko', encoding='euc-kr'), # 한글  
 na=".") %>%  
 round(2) %>% # 소수점 2자리로 반올림  
 mutate\_if(is.character, as.factor)

## Rows: 100 Columns: 1  
## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## dbl (1): weight  
##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

str(ost\_tb)

## spec\_tbl\_df [100 x 1] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ weight: num [1:100] 319 242 291 276 348 ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. weight = col\_double()  
## .. )  
## - attr(\*, "problems")=<externalptr>

glimpse(ost\_tb)

## Rows: 100  
## Columns: 1  
## $ weight <dbl> 319.31, 241.97, 290.98, 276.08, 347.55, 298.84, 292.87, 303.58,~

# 3.기본통계치 확인

skim(ost\_tb)

Data summary

|  |  |
| --- | --- |
| Name | ost\_tb |
| Number of rows | 100 |
| Number of columns | 1 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| numeric | 1 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: numeric**

| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| weight | 0 | 1 | 295.44 | 20.04 | 241.97 | 283.14 | 295.77 | 309.43 | 347.55 | ▂▅▇▇▁ |

ost\_tb %>%  
 get\_summary\_stats(weight)

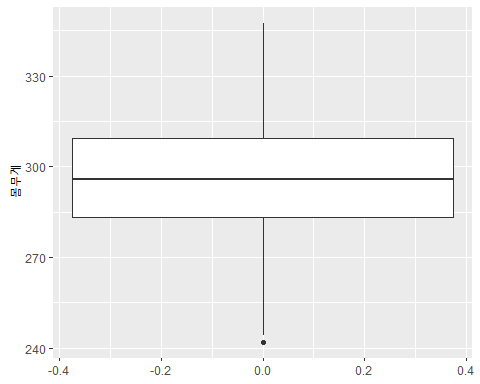
## # A tibble: 1 x 13  
## variable n min max median q1 q3 iqr mad mean sd se  
## <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 weight 100 242. 348. 296. 283. 309. 26.3 19.7 295. 20.0 2.00  
## # ... with 1 more variable: ci <dbl>

ost\_tb %>%   
 summarize(sample\_size = n(),  
 mean = mean(weight),  
 sd = sd(weight),  
 minimum = min(weight),  
 lower\_quartile = quantile(weight, 0.25),  
 median = median(weight),  
 upper\_quartile = quantile(weight, 0.75),  
 max = max(weight))

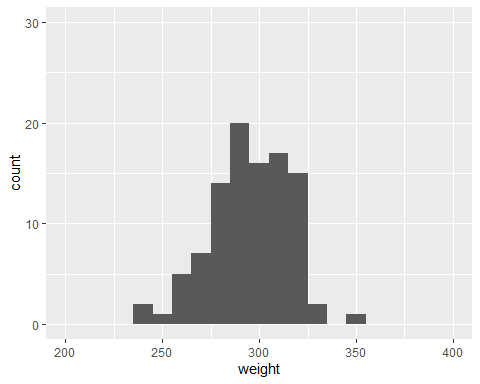
## # A tibble: 1 x 8  
## sample\_size mean sd minimum lower\_quartile median upper\_quartile max  
## <int> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 100 295. 20.0 242. 283. 296. 309. 348.

# 4.그래프 그리기(박스그래프,히스토그램)

ost\_tb %>%   
 ggplot(aes(y = weight)) +  
 geom\_boxplot() +  
 labs(y = "몸무게")



ost\_tb %>%   
 ggplot(mapping = aes(x = weight)) +  
 geom\_histogram(binwidth = 10) +  
 coord\_cartesian(xlim=c(200, 400), # coord\_cartesian:좌표계  
 ylim=c(0, 30))

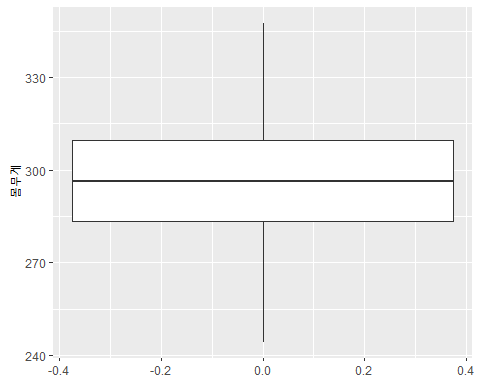


# 5.이상치 제거(이상치 확인, 이상치 제거)

ost\_tb %>%   
 identify\_outliers(weight)

## # A tibble: 1 x 3  
## weight is.outlier is.extreme  
## <dbl> <lgl> <lgl>   
## 1 242. TRUE FALSE

ost\_tb <- ost\_tb %>%  
 filter(!(weight <= 242))  
  
ost\_tb %>%   
 ggplot(aes(y = weight)) +  
 geom\_boxplot() +  
 labs(y = "몸무게")



# 6.정규분포 검정

ost\_tb %>%  
 shapiro\_test(weight)

## # A tibble: 1 x 3  
## variable statistic p  
## <chr> <dbl> <dbl>  
## 1 weight 0.987 0.422

# 7.통계분석

two-sided test: alternative = c(“two.sided”)  
right-sided test: alternative = c(“greater”)  
left-sided test: alternative = c(“less”)

ost\_tb %>%   
 t\_test(formula = weight ~ 1,  
 alternative = "two.sided",  
 mu = 320.0,   
 conf.level = 0.95,  
 detailed = TRUE)

## # A tibble: 1 x 12  
## estimate .y. group1 group2 n statistic p df conf.low conf.high  
## \* <dbl> <chr> <chr> <chr> <int> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 296. weig~ 1 null ~ 99 -12.3 1.27e-21 98 292. 300.  
## # ... with 2 more variables: method <chr>, alternative <chr>

# Cohen’s d(effect size)  
# 0.2 (small effect), 0.5 (moderate effect) and 0.8 (large effect)  
ost\_tb %>%   
 cohens\_d(formula = weight ~ 1,   
 mu = 320.0)

## # A tibble: 1 x 6  
## .y. group1 group2 effsize n magnitude  
## \* <chr> <chr> <chr> <dbl> <int> <ord>   
## 1 weight 1 null model -1.24 99 large

# 8.추론(infer)을 이용한 가설검정 및 그래프

## 8.1 표본평균(x)을 이용한 검정그래프

### 표본평균(x) 계산

x\_bar <- ost\_tb %>%  
 specify(response = weight) %>% # hypothesize 없음  
 calculate(stat = "mean") %>% # stat = "mean"  
 print()

## Response: weight (numeric)  
## # A tibble: 1 x 1  
## stat  
## <dbl>  
## 1 296.

### Bootstrapping을 이용한 귀무가설 분포 생성

set.seed(123)   
null\_dist\_x <- ost\_tb %>%  
 specify(response = weight) %>%  
 hypothesize(null = "point",   
 mu = 320) %>%  
 generate(reps = 1000,   
 type = "bootstrap") %>%  
 calculate(stat = "mean") %>%  
 print()

## Response: weight (numeric)  
## Null Hypothesis: point  
## # A tibble: 1,000 x 2  
## replicate stat  
## <int> <dbl>  
## 1 1 319.  
## 2 2 317.  
## 3 3 324.  
## 4 4 320.  
## 5 5 320.  
## 6 6 319.  
## 7 7 322.  
## 8 8 318.  
## 9 9 322.  
## 10 10 319.  
## # ... with 990 more rows

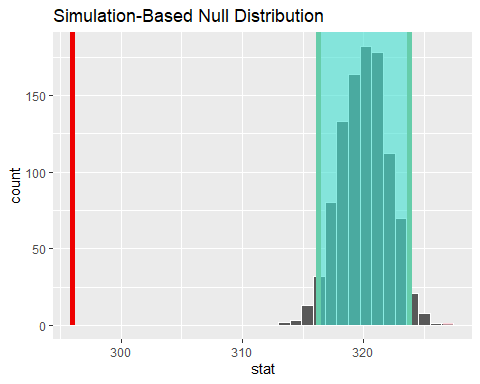
### 신뢰구간 생성

null\_dist\_ci <- null\_dist\_x %>%  
 get\_ci(level = 0.95,   
 type = "percentile") %>%  
 print()

## # A tibble: 1 x 2  
## lower\_ci upper\_ci  
## <dbl> <dbl>  
## 1 316. 324.

###그래프 그리기

null\_dist\_x %>%  
 visualize() + # method 없음  
 shade\_p\_value(obs\_stat = x\_bar,  
 direction = "two-sided") + # x\_bar  
 shade\_confidence\_interval(endpoints = null\_dist\_ci) # CI

 ###p\_value

## 8.2 t값을 이용한 검정그래프

### t\_cal 계산

t\_cal <- ost\_tb %>%  
 specify(response = weight) %>%  
 hypothesize(null = "point", # hypothesize 필요  
 mu = 320) %>%   
 calculate(stat = "t") %>% # stat = "t"   
 print()

## Response: weight (numeric)  
## Null Hypothesis: point  
## # A tibble: 1 x 1  
## stat  
## <dbl>  
## 1 -12.3

### Bootstrapping을 이용한 귀무가설 분포 생성

set.seed(123)   
null\_dist\_t <- ost\_tb %>%  
 specify(response = weight) %>%  
 hypothesize(null = "point",   
 mu = 320) %>%  
 generate(reps = 1000,   
 type = "bootstrap") %>%  
 calculate(stat = "t") %>%  
 print()

## Response: weight (numeric)  
## Null Hypothesis: point  
## # A tibble: 1,000 x 2  
## replicate stat  
## <int> <dbl>  
## 1 1 -0.683   
## 2 2 -1.43   
## 3 3 2.18   
## 4 4 0.198   
## 5 5 -0.0437  
## 6 6 -0.549   
## 7 7 1.02   
## 8 8 -1.12   
## 9 9 1.15   
## 10 10 -0.677   
## # ... with 990 more rows

### 신뢰구간 생성

null\_dist\_ci <- null\_dist\_t %>%  
 get\_ci(level = 0.95,   
 type = "percentile") %>%  
 print()

## # A tibble: 1 x 2  
## lower\_ci upper\_ci  
## <dbl> <dbl>  
## 1 -1.93 2.08

null\_dist\_t %>%  
 visualize(method = "both") + #method = "both": 이론분포+boot분포  
 shade\_p\_value(obs\_stat = t\_cal,  
 direction = "two-sided") +  
 shade\_confidence\_interval(endpoints = null\_dist\_ci)

## Warning: Check to make sure the conditions have been met for the theoretical  
## method. {infer} currently does not check these for you.

