Correlation\_Test

Youngjin\_Lee

2022 2 18

Table of Contents

통계분석 차이검정 : 상관분석(Correlation Sample) test 문제 K의류에서는 새로운 옷을 디자인하려고 하는데, 키와 몸무게가 어떤 관계가 있는지를 보고자 한다.

# 1.기본 package 설정, library 로드

# 2.데이터 불러오기

corr\_tb <- read\_csv('data/CORR.csv',   
 col\_names = TRUE,  
 locale=locale('ko', encoding='euc-kr'), # 한글  
 na=".") %>%  
 mutate\_if(is.character, as.factor)

## Rows: 30 Columns: 2  
## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## dbl (2): 몸무게, 키  
##   
## 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(corr\_tb)

## spec\_tbl\_df [30 x 2] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ 몸무게: num [1:30] 72 72 70 43 48 54 51 52 73 45 ...  
## $ 키 : num [1:30] 176 172 182 160 163 165 168 163 182 148 ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. 몸무게 = col\_double(),  
## .. 키 = col\_double()  
## .. )  
## - attr(\*, "problems")=<externalptr>

# 3.기본통계치 확인

skim(corr\_tb)

Data summary

|  |  |
| --- | --- |
| Name | corr\_tb |
| Number of rows | 30 |
| Number of columns | 2 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| numeric | 2 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: numeric**

| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 몸무게 | 0 | 1 | 62.70 | 11.52 | 43 | 54.00 | 63 | 72.00 | 88 | ▇▆▇▇▂ |
| 키 | 0 | 1 | 170.33 | 8.68 | 148 | 164.25 | 170 | 175.75 | 188 | ▁▅▇▅▃ |

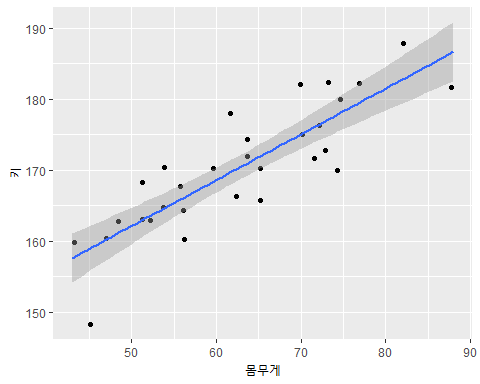
corr\_tb %>%   
 get\_summary\_stats(몸무게, 키)

## # A tibble: 2 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 몸무게 30 43 88 63 54 72 18 13.3 62.7 11.5 2.10  
## 2 키 30 148 188 170 164. 176. 11.5 8.90 170. 8.68 1.58  
## # ... with 1 more variable: ci <dbl>

# 4.그래프 그리기(산점도)

corr\_tb %>%  
 ggplot(mapping = aes(x = 몸무게,  
 y = 키)) +  
 geom\_jitter() +  
 geom\_smooth(method = "lm")

## `geom\_smooth()` using formula 'y ~ x'



# 5.상관분석

corr\_tb %>%  
 cor\_test(몸무게, 키,   
 method = "pearson")

## # A tibble: 1 x 8  
## var1 var2 cor statistic p conf.low conf.high method   
## <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <chr>   
## 1 몸무게 키 0.86 8.79 0.00000000154 0.718 0.930 Pearson