Hypothesis Testing Homework Problem

George Paul

2024-02-14

```
rm(list = setdiff(ls(), lsf.str()))
```

```
# install.packages('readxl')
library(readxl)
```

```
excel_path <- 'D:\\FILES\\BRSMHypDataset.xlsx'
```

```
qdata <- read_excel(excel_path)
```

```
## New names:
## * `` -> `...1`
## * `GPA` -> `GPA...2`
## * `GPA` -> `GPA...6`
## * `` -> `...7`
## * `` -> `...8`
## * `` -> `...12`
## * `` -> `...13`
## * `` -> `...15`
## * `` -> `...16`
## * `` -> `...18`
## * `` -> `...19`
## * `` -> `...20`
## * `` -> `...22`
## * `` -> `...23`
```

```
qdata
```

```
## # A tibble: 78 x 23
##     ...1 GPA...2    IQ GENDER `Placement \r\nTESTSCORE` GPA...6 ...7  ...8
##    <dbl>   <dbl> <dbl>  <dbl>                     <dbl>   <dbl> <lgl> <lgl>
##  1     1    7.94   111      2                        67    7.94 NA    NA
##  2     2    8.29   107      2                        43    8.29 NA    NA
##  3     3    4.64   100      2                        52    4.64 NA    NA
##  4     4    7.47   107      2                        66    7.47 NA    NA
##  5     5    8.88   114      1                        58    8.88 NA    NA
##  6     6    7.58   115      2                        51    7.58 NA    NA
##  7     7    7.65   111      2                        71    7.65 NA    NA
##  8     8    2.41    97      2                        51    2.41 NA    NA
##  9     9    6      100      1                        49    6    NA    NA
## 10    10    8.83   112      2                        51    8.83 NA    NA
```

```
## # i 68 more rows
## # i 15 more variables: Exerice_Times <dbl>, `Exercise code` <dbl>,
## #   Anxiety <dbl>, ...12 <lgl>, ...13 <lgl>, `anxiety scores` <chr>,
## #   ...15 <chr>, ...16 <lgl>, `t-Test: Paired Two Sample for Means` <chr>,
## #   ...18 <chr>, ...19 <chr>, ...20 <lgl>,
## #   `t-Test: Two-Sample Assuming Equal Variances` <chr>, ...22 <chr>,
## #   ...23 <chr>
```

```
split_data <- split(qdata, qdata$GPA...2 < 7)
split_data
```

```
## $`FALSE`
## # A tibble: 56 x 23
##     ...1 GPA...2    IQ GENDER `Placement \r\nTESTSCORE` GPA...6 ...7  ...8
##    <dbl>   <dbl> <dbl>  <dbl>                     <dbl>   <dbl> <lgl> <lgl>
##  1     1    7.94   111      2                        67    7.94 NA    NA
##  2     2    8.29   107      2                        43    8.29 NA    NA
##  3     4    7.47   107      2                        66    7.47 NA    NA
##  4     5    8.88   114      1                        58    8.88 NA    NA
##  5     6    7.58   115      2                        51    7.58 NA    NA
##  6     7    7.65   111      2                        71    7.65 NA    NA
##  7    10    8.83   112      2                        51    8.83 NA    NA
##  8    11    7.47   104      1                        35    7.47 NA    NA
##  9    13    7.17   104      2                        54    7.17 NA    NA
## 10    14    7.57   102      1                        64    7.57 NA    NA
## # i 46 more rows
## # i 15 more variables: Exerice_Times <dbl>, `Exercise code` <dbl>,
## #   Anxiety <dbl>, ...12 <lgl>, ...13 <lgl>, `anxiety scores` <chr>,
## #   ...15 <chr>, ...16 <lgl>, `t-Test: Paired Two Sample for Means` <chr>,
## #   ...18 <chr>, ...19 <chr>, ...20 <lgl>,
## #   `t-Test: Two-Sample Assuming Equal Variances` <chr>, ...22 <chr>,
## #   ...23 <chr>
##
## $`TRUE`
## # A tibble: 22 x 23
##     ...1 GPA...2    IQ GENDER `Placement \r\nTESTSCORE` GPA...6 ...7  ...8
##    <dbl>   <dbl> <dbl>  <dbl>                     <dbl>   <dbl> <lgl> <lgl>
##  1     3    4.64   100      2                        52    4.64 NA    NA
##  2     8    2.41    97      2                        51    2.41 NA    NA
##  3     9    6      100      1                        49    6    NA    NA
##  4    12    5.53    89      1                        54    5.53 NA    NA
##  5    15    4.7     91      1                        56    4.7  NA    NA
##  6    19    4      106      2                        40    4    NA    NA
##  7    20    6.23   105      1                        66    6.23 NA    NA
##  8    22    1.76   109      2                        20    1.76 NA    NA
##  9    24    6.42   108      1                        56    6.42 NA    NA
## 10    50    3.65    97      2                        52    3.65 NA    NA
## # i 12 more rows
## # i 15 more variables: Exerice_Times <dbl>, `Exercise code` <dbl>,
## #   Anxiety <dbl>, ...12 <lgl>, ...13 <lgl>, `anxiety scores` <chr>,
## #   ...15 <chr>, ...16 <lgl>, `t-Test: Paired Two Sample for Means` <chr>,
## #   ...18 <chr>, ...19 <chr>, ...20 <lgl>,
## #   `t-Test: Two-Sample Assuming Equal Variances` <chr>, ...22 <chr>,
## #   ...23 <chr>
```

```r
lt7data <- split_data[["TRUE"]]
gt7data <- split_data[["FALSE"]]
```

```r
# "Placement \r\nTESTSCORE"
t_test_result <- t.test(lt7data$"Placement \r\nTESTSCORE", gt7data$"Placement \r\nTESTSCORE")
t_test_result
```

```
##
##  Welch Two Sample t-test
##
## data:  lt7data$"Placement \r\nTESTSCORE" and gt7data$"Placement \r\nTESTSCORE"
## t = -4.3771, df = 29.822, p-value = 0.0001357
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -20.164774  -7.331979
## sample estimates:
## mean of x mean of y
##  47.09091  60.83929
```

Considering the standard value for $\alpha = 0.05$ when evaluating at p-values, the calculated p-value is 0.0001357 which is $< \alpha$. With this we can conclude that the test is statistically significant and that students with a $GPA \leq 7$ have lower $placementTESTSCORES$ than those with a $GPA > 7$.