hw 5

2022-11-04

```
setwd("~/Documents/GitHub/stats100")
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.5
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggplot2)
library(scales)
```

```
##
## Attaching package: 'scales'
##
## The following object is masked from 'package:purrr':
##
##     discard
##
## The following object is masked from 'package:readr':
##
##     col_factor
```

```
library(RColorBrewer)
library(stats)
```

```
drp<- read.csv("drp.csv")
drp
```

```
##    Treatment Response
## 1    Treated       24
## 2    Treated       43
## 3    Treated       58
## 4    Treated       71
## 5    Treated       43
## 6    Treated       49
## 7    Treated       61
```

```
## 8    Treated       44
## 9    Treated       67
## 10   Treated       49
## 11   Treated       53
## 12   Treated       56
## 13   Treated       59
## 14   Treated       52
## 15   Treated       62
## 16   Treated       54
## 17   Treated       57
## 18   Treated       33
## 19   Treated       46
## 20   Treated       43
## 21   Treated       57
## 22   Control       42
## 23   Control       43
## 24   Control       55
## 25   Control       26
## 26   Control       62
## 27   Control       37
## 28   Control       33
## 29   Control       41
## 30   Control       19
## 31   Control       54
## 32   Control       20
## 33   Control       85
## 34   Control       46
## 35   Control       10
## 36   Control       17
## 37   Control       60
## 38   Control       53
## 39   Control       42
## 40   Control       37
## 41   Control       42
## 42   Control       55
## 43   Control       28
## 44   Control       48
```

A treatment class of 21 third-grade students participated in these activities for eight weeks, and a control class of 23 third-graders fol- lowed the same curriculum without the activities. After the eight-week period, students in both classes took a Degree of Reading Power (DRP) test which measures the aspects of reading ability that the treatment is de- signed to improve. The first column is the group (treatment or control), and the second is the DRP score (the higher the better).

(a) Find the 95% confidence interval for the difference in average DRP score.

```
controlt<- drp %>%
  filter(Treatment == "Control") %>%
  summarise(Treatment, Response)
controlt
```

```
##    Treatment Response
## 1    Control       42
```

```
## 2    Control       43
## 3    Control       55
## 4    Control       26
## 5    Control       62
## 6    Control       37
## 7    Control       33
## 8    Control       41
## 9    Control       19
## 10   Control       54
## 11   Control       20
## 12   Control       85
## 13   Control       46
## 14   Control       10
## 15   Control       17
## 16   Control       60
## 17   Control       53
## 18   Control       42
## 19   Control       37
## 20   Control       42
## 21   Control       55
## 22   Control       28
## 23   Control       48
```

```
treatedt<- drp %>%
  filter(Treatment == "Treated") %>%
  summarise(Treatment, Response)
treatedt
```

```
##    Treatment Response
## 1    Treated       24
## 2    Treated       43
## 3    Treated       58
## 4    Treated       71
## 5    Treated       43
## 6    Treated       49
## 7    Treated       61
## 8    Treated       44
## 9    Treated       67
## 10   Treated       49
## 11   Treated       53
## 12   Treated       56
## 13   Treated       59
## 14   Treated       52
## 15   Treated       62
## 16   Treated       54
## 17   Treated       57
## 18   Treated       33
## 19   Treated       46
## 20   Treated       43
## 21   Treated       57
```

```
t.test(controlt$Response, treatedt$Response, conf.level = 0.95)
```

```
##
```

```
##  Welch Two Sample t-test
##
## data:  controlt$Response and treatedt$Response
## t = -2.3109, df = 37.855, p-value = 0.02638
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.67588  -1.23302
## sample estimates:
## mean of x mean of y
##  41.52174  51.47619
```

Answer: (-18.67588, -1.23302)

(b) If the experimenter is interested in if the treatment group has a higher average than the control group, find the test-statistic.

```
t.test(controlt$Response, treatedt$Response, alternative = "less")
```

```
##
##  Welch Two Sample t-test
##
## data:  controlt$Response and treatedt$Response
## t = -2.3109, df = 37.855, p-value = 0.01319
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -2.691293
## sample estimates:
## mean of x mean of y
##  41.52174  51.47619
```

Answer: -2.3109

(c) If the experimenter is interested in if the treatment group has a higher average than the control group, find the p-value.

Answer: 0.01319

(d) Do we fail to reject or reject the null hypothesis if a = 0.05?

Answer: 0.01319<0.05 so p<a and we can reject null

IQ.csv has two columns, the first of which denotes what major a stu- dent is from (A, B, or C). The second is the IQ measured by the Stanford–Binet Intelligence Scales. The goal is to determine if this IQ measures differs on average between majors.

```
iqdf<- read_csv("IQ.csv")
```

```
## Rows: 45 Columns: 2
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): group
## dbl (1): iq
##
## 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.
```

```
iqdf
```

```
## # A tibble: 45 x 2
##    group    iq
##    <chr> <dbl>
##  1 A        44
##  2 A        40
##  3 A        44
##  4 A        39
##  5 A        25
##  6 A        37
##  7 A        31
##  8 A        40
##  9 A        22
## 10 A        34
## # ... with 35 more rows
```

```
iqmodel<- lm(iqdf$iq~iqdf$group)
iqmodel
```

```
##
## Call:
## lm(formula = iqdf$iq ~ iqdf$group)
##
## Coefficients:
## (Intercept)  iqdf$groupB  iqdf$groupC
##    35.80000      0.06667     12.40000
```

```
anova(iqmodel)
```

```
## Analysis of Variance Table
##
## Response: iqdf$iq
##            Df Sum Sq Mean Sq F value    Pr(>F)
## iqdf$group  2 1529.4  764.69  20.016 7.843e-07 ***
## Residuals  42 1604.5   38.20
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(a) Find the value of the F test-statistic. Answer: F stat= 20.016

(b) Find the p-value of the test. Answer: P-value= Pr(>F)=7.843e-07

c) Do we fail to reject or reject the null hypothesis? (Choose your own a) Answer: If a=0.001 then p-value is less than that it, and p<a so we can reject the null

(d) State your conclusion in terms of the problem. Answer: Rejection of null means that average IQ does not vary by major