CR_Portfolio_2

Seunghun Lee

Oct 12, 2020

Contents

Research Question

Researchers were interested in whether the correlation between first mathematics exam score (G1) and final mathematics exam score (G3) was stronger for male students than female students in a secondary education of Portuguese school. They obtained a random sample of 395 students (208 females, 187 males) from an independent dataset and investigated their research question using an alpha of 0.05 (P. Cortez and A. Silva, 2008).

```
# https://archive.ics.uci.edu/ml/datasets/student+performance
dat <- read.table('./data/student-mat.csv',sep=";",header=TRUE)
# Subset necessary attributes
dat_f <- dat %>%
  select(sex, G1, G3) %>%
  filter(sex == 'F')
summary(dat_f)
```

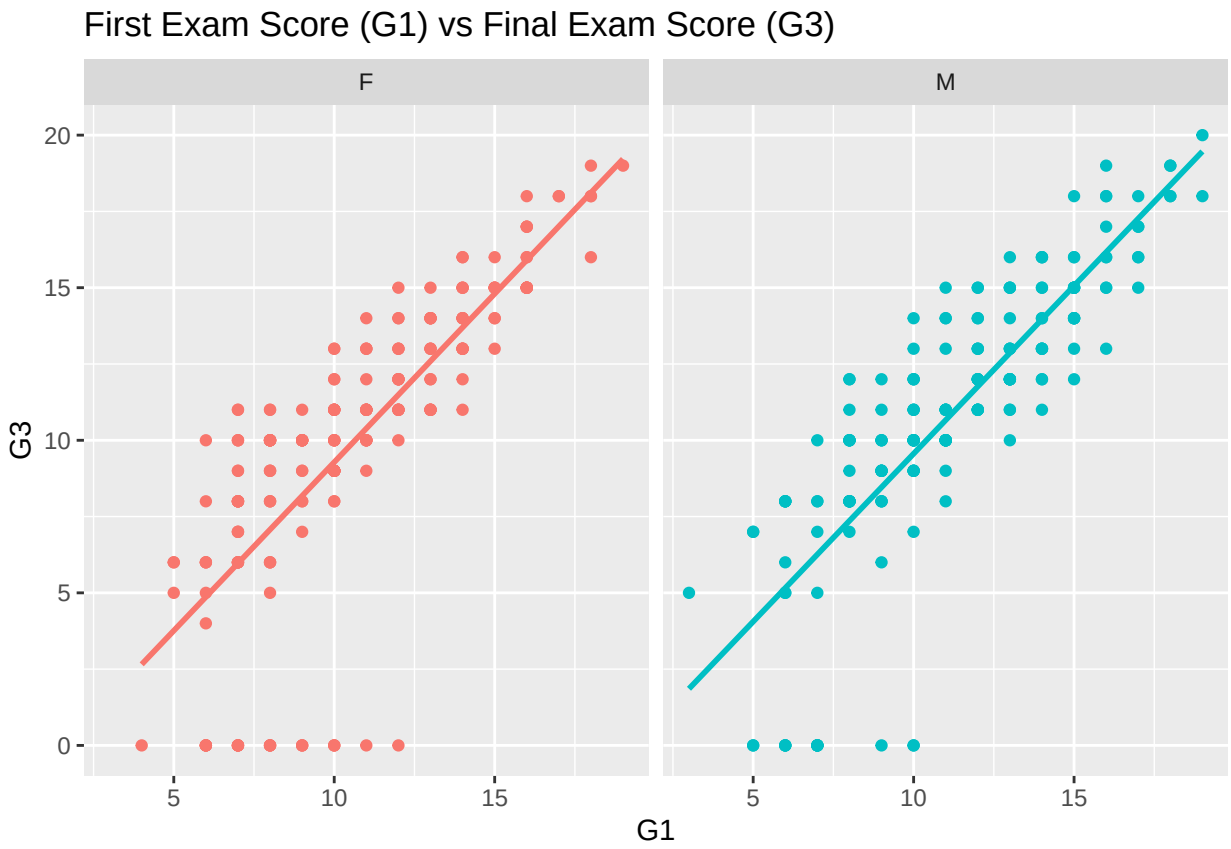
```
##  sex           G1              G3
##  F:208   Min.   : 4.00   Min.   : 0.000
##  M:  0   1st Qu.: 8.00   1st Qu.: 8.000
##          Median :10.00   Median :10.000
##          Mean   :10.62   Mean   : 9.966
##          3rd Qu.:13.00   3rd Qu.:13.000
##          Max.   :19.00   Max.   :19.000
```

```
dat_m <- dat %>%
  select(sex, G1, G3) %>%
  filter(sex  == 'M')
summary(dat_m)
```

```
##  sex          G1              G3
##  F:  0   Min.   : 3.00   Min.   : 0.00
##  M:187   1st Qu.: 9.00   1st Qu.: 9.00
##          Median :11.00   Median :11.00
##          Mean   :11.23   Mean   :10.91
##          3rd Qu.:14.00   3rd Qu.:14.00
##          Max.   :19.00   Max.   :20.00
```

```
dat %>%
  select(sex, G1, G3) %>%
  ggplot(aes(G1, G3, color = sex)) +
  geom_point() +
  geom_smooth(method = 'lm', se = FALSE) +
  facet_grid(cols = vars(sex)) +
  labs(title = 'First Exam Score (G1) vs Final Exam Score (G3)') +
  theme(legend.position = "none")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



Hypotheses

H_1: The correlation between first exam score (G1) and final exam score (G3) was stronger for male than female ($\rho_M - \rho_F > 0$)
H_0: The correlation between first exam score (G1) and final score (G3) was not stronger for male than female ($\rho_M - \rho_F \leq 0$)

Critical test statistic

```
round(abs(qnorm(0.05)), 3)
```

```
## [1] 1.645
```

$\alpha = 0.05$, one-tailed, Critical Z-statistics = 1.645

Test statistic

1. Build a function for independent correlation test

```
indep_cor_test <- function(data, group, var1, var2){
  #Get group levels
  group_levels <- data %>%
    pull({{group}}) %>%
    levels()
  #Correlation in group 1
  data1 <- data %>%
    select({{group}}, {{var1}}, {{var2}}) %>%
    filter({{group}} == group_levels[1])
  cor1 <- data1 %>%
    mutate(var1 = var1 <- {{var1}},
           var2 = var2 <- {{var2}}) %$%
    cor(var1, var2)
  #Correlation in group 2
  data2 <- data %>%
    select({{group}}, {{var1}}, {{var2}}) %>%
    filter({{group}} == group_levels[2])
  cor2 <- data2 %>%
    mutate(var1 = var1 <- {{var1}},
           var2 = var2 <- {{var2}}) %$%
    cor(var1, var2)
  #Independent correlation test
  rtest <- r.test(n = nrow(data1), n2 = nrow(data2), cor1, cor2, twotailed = FALSE)
  #print the results
  print(paste("The correlation in", group_levels[1]))
  print(cor1)
  print(paste("The correlation in", group_levels[2]))
  print(cor2)
  rtest
  }
```

- Computation by the function above

```
indep_cor_test(dat, sex, G1, G3)
```

```
## [1] "The correlation in F"
## [1] 0.7721856
## [1] "The correlation in M"
## [1] 0.8307415
```

```
## Correlation tests
## Call:r.test(n = nrow(data1), r12 = cor1, r34 = cor2, n2 = nrow(data2),
##     twotailed = FALSE)
## Test of difference between two independent correlations
##  z value 1.62    with probability  0.05
```

- Result sample statistic Z value 1.62 < Critical Z-statistic 1.645

2. Computation by hand

```
manual_z <- function(n1, cor1, n2, cor2){
  z1 <- 0.5*log((1+cor1)/(1-cor1))
  z2 <- 0.5*log((1+cor2)/(1-cor2))

  z <- ((z2 - z1) - 0)/sqrt((1/(n1-3))+(1/(n2-3)))
  print(abs(z))
}

manual_z(208, 0.7721856, 187, 0.8307415)
```

```
## [1] 1.622869
```

- Result sample statistic Z value 1.623 < Critical Z-statistic 1.645

Conclusion

- Fail to reject the null hypothesis
- The correlation between G1 and G3 was not significantly stronger for male than for female
- [$r_F = 0.77$, $n_F = 208$, $r_M = 0.83$, $n_M = 187$, Z = 1.62, p > 0.05]