ECON 453, Spring 2023, Pset 2
Minh Duong

Problem Set 2 Report

Minh Duong

2023 – 03 – 21

a. 95% confidence interval for the overall average test score.

Using the t-test function in R, the 95% confidence interval for the overall average test score is calculated to be [59.11, 65.63]

b. Test the null hypothesis that the overall average test scores (for both schools and for both years combined) is 63 against the alternative that the average is not 63. Use a 5% level of significance.

Using the t-test function in R with mu (true mean) being 63, we calculated that the p-value of the test is 0.7034 > 0.05, so we failed to reject the null hypothesis. So, with 5% level of significance, the null hypothesis is true, and the overall average test score is 63.

c. The School District, which encompasses both schools, claims that average test scores increased for both schools between 2014 and 2016. Does the data support the School District's claim? Verify the claim one school at a time.

- Lincoln

  The result of the t-test (ran by R) has a p-value of 0.009 < 0.05, so we reject the null hypothesis that the average test score for Lincoln school for 2016 is not greater than that for 2014. In other words, the score for Lincoln school for 2016 is (statistically significantly) greater than that for 2014.

- Kennedy

  The result of the t-test (ran by R) has a p-value of 0.005 < 0.05, so we reject the null hypothesis that the average test score for Kennedy school for 2016 is not greater than that for 2014. In other words, the score for Kennedy school for 2016 is (statistically significantly) greater than that for 2014.

d. **A county official claims that the performance of fourth graders does not differ between both schools. Does the data support the claim? Verify the claim for 2014 and 2016 separately.**

- 2014:

  The result of the t-test (ran by R) has a p-value of 0.03 < 0.05, so we reject the null hypothesis that the performance of the two schools is not different in 2014. In other words, the performance of the two schools is statistically different in 2014.

  As a piece of additional information, running the one-sided t-test with the alternative hypothesis that the Lincoln school has smaller score results in p-value = 0.01 < 0.05, which indicates that the Kennedy school had statistically significantly greater test scores compared to the Lincoln school in 2014.

- 2016:

  The result of the t-test (ran by R) has a p-value of 0.03 < 0.05, so we reject the null hypothesis that the performance of the two schools is not different in 2016. In other words, the performance of the two schools is statistically different in 2016.

  As a piece of additional information, running the one-sided t-test with the alternative hypothesis that the Lincoln school has smaller score results in p-value = 0.01 < 0.05, which indicates that the Kennedy school had statistically significantly greater test scores compared to the Lincoln school in 2016.

e. **A score of 40 is considered a pass. Test the hypothesis that the proportion of students who pass the exam is the same for both schools against the alternative that the proportions are not the same.**

For the test between the proportion of students passing the exam between the two schools, the t-test performed by R gives the p-value of 0.0003 < 0.05, so we reject the null hypothesis. In other words, the proportion of students passing the exam is statistically significantly different between the two schools.

As a piece of additional information, running the one-sided t-test with the alternative hypothesis that the Kennedy school has a larger proportion of students passing the exam gives out the p-value of $0.0001 < 0.005$, which indicates that the Kennedy school had a statistically significantly greater proportion of students passing the test, compared to the Lincoln school.

f. Write a small report (4 or 5 sentences) based on the results from the above questions and any other inference you can deduct from the data. This report is written for a local newspaper and should be non-technical.

A comparison of test scores between Lincoln and Kennedy Primary Schools has revealed some interesting findings. The overall average test score for both schools is 63. Both schools have shown an increase in test scores from 2014 to 2016. Additionally, Kennedy Primary School has statistically higher scores than Lincoln Primary School, with a larger proportion of students passing the test. These results suggest that parents looking for high-performing schools may want to consider Kennedy Primary School over Lincoln Primary School.

ECON 453, Spring 2023, Pset 2
Minh Duong

Addendum

Project Repository: https://github.com/MykeDuong/econ453

R Script code:

```
## Minh Duong - ECON 453 - pset2

rm(list = ls())

library(readxl)

getwd()
setwd("./data/")

score_data<- read_excel("pset1_data.xlsx", sheet="scores")

## A. Using the data for both schools for both years, calculate a 95%
## confidence interval for overall average test score.

# number of rows and degree of freedom
n <- nrow(score_data)
df <- (n - 1)

# Manual way of calculating
# mean(score_data$score) + (sd(score_data$score) /sqrt(n)) * qt(0.025, df)
# mean(score_data$score) + (sd(score_data$score) /sqrt(n)) * qt(0.975, df)

t.test(score_data$score,conf.level = 0.95)

## B. Test the null hypothesis that overall average test scores (for both
## schoolsand for both years combined) is 63 against the alternative that
the
## average is not 63. Use 5% level of significance.


t.test(score_data$score, mu = 63, conf.level = 0.95, alternative =
"two.sided")

# p-value = 0.7034 > 5% => fail to reject the null hypothesis that overall
# average test scores (for both schools and for both years combined) is 63


## C. The School District, which encompasses both schools, claims that
average
## test score increased for both schools between 2014 and 2016. Does the
data
## support the School District,Äôs claim? Verify the claim one school at a
time.

lincoln_2014 = score_data$score[
  score_data$school == 'Lincoln' &
```

```
  score_data$year == 2014
]
lincoln_2016 = score_data$score[
  score_data$school == 'Lincoln' &
  score_data$year == 2016
]

score_2014 = score_data$score[score_data$year == 2014]
score_2016 = score_data$score[score_data$year == 2016]

kennedy_2014 = score_data$score[
  score_data$school == 'Kennedy' &
  score_data$year == 2014
]
kennedy_2016 = score_data$score[
  score_data$school == 'Kennedy' &
  score_data$year == 2016
]

# Custom function to manually calculate one-side t-test
test_claim <- function(data_2014, data_2016) {
  mean_diff = mean(data_2016) - mean(data_2014)
  mean_var_2014 = var(data_2014) / length(data_2014)
  mean_var_2016 = var(data_2016) / length(data_2016)
  t_stat <- mean_diff / sqrt(mean_var_2014 + mean_var_2016)

  t_stat

  print(t_stat)

  df = ((mean_var_2014 + mean_var_2016) ^ 2) /
       (
         (mean_var_2016 ^ 2 / (length(data_2016) - 1)) +
         (mean_var_2014 ^ 2 / (length(data_2014) - 1))
       )

  df

  print(1 - pt(t_stat, df))
}

# Lincoln
t.test(lincoln_2016, lincoln_2014, conf.level = 0.95, alternative =
"greater")
# manual: test_claim(lincoln_2014, lincoln_2016)
# p-value= 0.009 < 5% => reject the null hypothesis that  average test
scores
# for Lincoln for 2016 is no greater than  that for 2014
# => The score for Lincoln for 2016 is greater than that for 2014.

# Kennedy
```

ECON 453, Spring 2023, Pset 2
Minh Duong

```
t.test(kennedy_2016, kennedy_2014, conf.level = 0.95, alternative =
"greater")
# manual: test_claim(kennedy_2014, kennedy_2016)
# p-value= 0.005 < 5% => reject the null hypothesis thataverage test
scores
# for Kennedy for 2016 is no greater than  that for 2014
# => The score for Kennedy for 2016 is greater than that for 2014.

# Both schools
t.test(score_2016, score_2014, conf.level = 0.95, alternative = "greater")
# manual: test_claim(score_2014, score_2016)
# p-value= 0.0004 < 5% => reject the null hypothesis that average test
scores
# for Kennedy for 2016 is no greater than  that for 2014
# => The score for both schools for 2016 is greater than that for 2014.

## D. A county official claims that performance of fourth graders does not
## differ between both schools. Does the data support the claim? Verify
the
## claim for 2014 and 2016 separately.

t.test(lincoln_2014, kennedy_2014, conf.level = 0.95, alternative =
"two.sided")
# p-value= 0.03 < 5% => reject the null hypothesis that the scores of two
# schools are different in 2014
# => The scores of the two schools are not statistically different in 2014

t.test(lincoln_2016, kennedy_2016, conf.level = 0.95, alternative =
"two.sided")
# p-value= 0.03 < 5% => reject the null hypothesis that the scores of two
# schools are different in 2016
# => The scores of the two schools are not statistically different in 2016

## E: A score of 40 is considered as pass. Test the hypothesis that the
## proportion of students who pass the exam is the same for both schools
## against the alternative that the proportions are not the same.

# Define and create pass/fail data
score_data$pass = ifelse((score_data$score >= 40), 1, 0)

kennedy_pass = score_data$pass[score_data$school == 'Kennedy']
lincoln_pass = score_data$pass[score_data$school == 'Lincoln']

t.test(kennedy_pass, lincoln_pass, conf.level = 0.95, alternative =
"two.sided")
# p-value= 0.0003 < 5% => reject the null hypothesis that proportion of
students
# who pass the exam is the same for both schools.

t.test(kennedy_pass, lincoln_pass, conf.level = 0.95, alternative =
"greater")
# # p-value= 0.0001 < 5% => reject the null hypothesis and accept the
```

ECON 453, Spring 2023, Pset 2
Minh Duong

```
# alternative hypothesis that the Kennedy school had a larger proportion
# of students passing the exam, with a confidence level of 95%
```