CWRU DSCI351-451: Homework 5 Inference

Roger H. French, JiQi Liu

18 October, 2018

Contents

Inference Guide

There is a useful Inference Cheat Sheet in your readings folder

- os2_extra_inference_guide.pdf

There is Hadley Wickham's book on ggplot in your readings textboosk folder

- Elegant Graphics for Data Analysis (Wickham 2016)

5.0.0.1 1) Twitter users and News

5.0.0.1.1 Part I. (OIS 4.8)

A poll conducted in 2013 found that

- 52% of U.S. adult Twitter users get at least some news on Twitter.(Mitchell and Guskin 2013)
- The standard error for this estimate was 2.4%,
  - and a normal distribution may be used to model the sample proportion.

Construct a 99% confidence interval for

- the fraction of U.S. adult Twitter users
- who get some news on Twitter,

and interpret the confidence interval in context.

Answer:

5.0.0.1.2 Twitter users and news, Part II. (OIS 4.10)

Identify each of the following statements as true or false. Provide an explanation to justify each of your answers.

(a) The data provide statistically significant evidence that

- more than half of U.S. adult Twitter users
  - get some news through Twitter.
- Use a significance level of $\alpha = 0.01$.

Answer:

(b) Since the standard error is 2.4%,

- we can conclude that 97.6% of all U.S. adult Twitter users
  - were included in the study.

Answer:

(c) If we want to reduce the standard error of the estimate,

- we should collect less data.

Answer:

(d) If we construct a 90% confidence interval

- for the percentage of U.S. adults Twitter users
  - who get some news through Twitter,
- this confidence interval will be wider
  - than a corresponding 99% confidence interval.

Answer:

5.0.0.2 4.24 Gifted children,

5.0.0.2.1 Part I. (OIS 4.24)

Researchers investigating characteristics of gifted children

- collected data from schools in a large city
  - on a random sample
- of thirty-six children
  - who were identified as gifted children
  - soon after they reached the age of four.

The following histogram shows

- the distribution of the ages (in months)
- at which these children first counted to 10 successfully.

Also provided are some sample statistics.(Graybill and Iyer 1994)

(a) Are conditions for inference satisfied?

Answer:

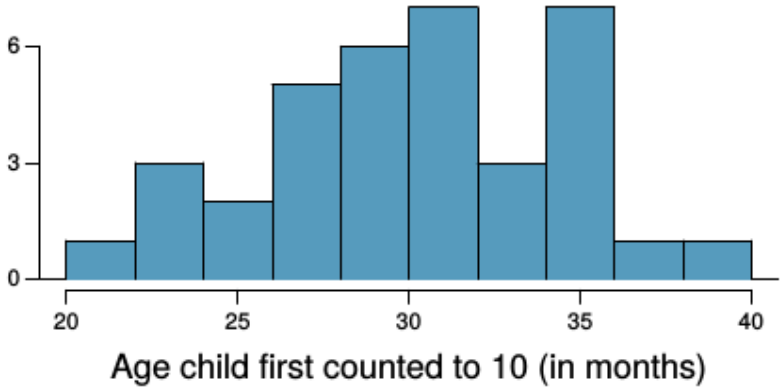
(b) Suppose you read online that children

- first count to 10 successfully when they are 32 months old, on average.

Perform a hypothesis test to evaluate

- if these data provide convincing evidence that
- the average age at which gifted children fist count to 10 successfully
  - is less than the general average of 32 months.
- Use a significance level of 0.10.

Answer:



n	36
min	21
mean	30.69
sd	4.31
max	39

Figure 1: Age Child First Counted to 10

(c) Interpret the p-value in context

- of the hypothesis test
- and the data.

Answer:

(d) Calculate a 90% confidence interval

- for the average age at which gifted children
  - first count to 10 successfully.

Answer:

(e) Do your results from

- the hypothesis test and
- the confidence interval agree?

Explain.

Answer:

5.0.0.2.2 Part II. (OIS 4.26)

4.26 Gifted children, Part II. Exercise 4.24 describes a study on gifted children.

In this study, along with variables on the children,

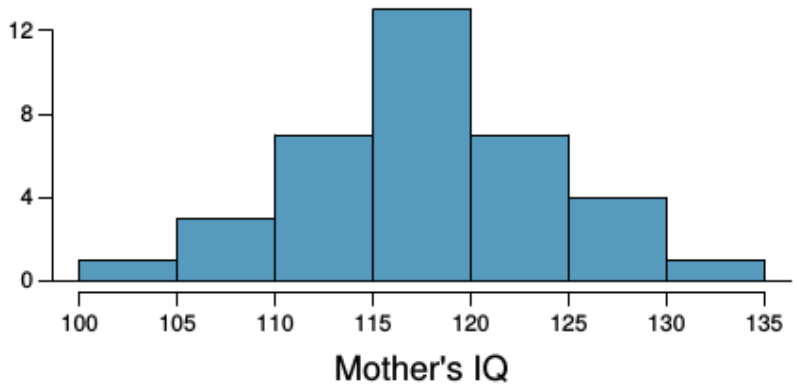
- the researchers also collected data
  - on the mother's and father's IQ
  - of the 36 randomly sampled gifted children.

The histogram below shows the distribution of mother's IQ. Also provided are some sample statistics.

(a) Perform a hypothesis test

- to evaluate if these data provide convincing evidence
  - that the average IQ of mothers of gifted children
- is different than the average IQ for the population at large,
  - which is 100.
- Use a significance level of 0.10.

Answer:



n	36
min	101
mean	118.2
sd	6.5
max	131

Figure 2: Mother's IQ

(b) Calculate a 90% confidence interval

- for the average IQ of mothers of gifted children.

Answer:

(c) Do your results from

- the hypothesis test
- and the confidence interval agree?

Explain.

Answer:

5.0.0.3 Spray Paint (OIS 4.42)

Suppose the area that can be painted using a single can of spray paint

- is slightly variable
- and follows a nearly normal distribution
  - with a mean of 25 square feet
  - and a standard deviation of 3 square feet.

(a) What is the probability that

- the area covered by a can of spray paint
- is more than 27 square feet?

Answer:

(b) Suppose you want to spray paint

- an area of 540 square feet
- using 20 cans of spray paint.

On average, how many square feet

- must each can be able to cover
- to spray paint all 540 square feet?

Answer:

(c) What is the probability

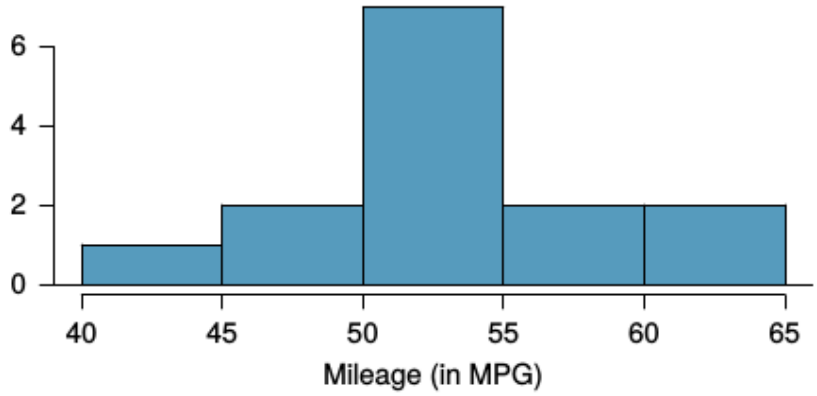


Figure 3: Mileage in MPG

- that you can cover a 540 square feet area
- using 20 cans of spray paint?

Answer:

(d) If the area covered by a can of spray paint

- had a slightly skewed distribution,
- could you still calculate the probabilities in parts (a) and (c)
  - using the normal distribution?

Answer:

5.0.0.4 Fuel efficiency of Prius. (OIS 5.8)

Fueleconomy.gov,

- the official US government source
  - for fuel economy information,
- allows users to share gas mileage information on their vehicles.

The histogram below shows

- the distribution of gas mileage in miles per gallon (MPG)
  - from 14 users who drive a 2012 Toyota Prius.
- The sample mean is 53.3 MPG
  - and the standard deviation is 5.2 MPG.

Note that these data are user estimates

- and since the source data cannot be verified,
- the accuracy of these estimates are not guaranteed.("Gas Mileage of 2012 Toyota Prius" n.d.)

(a) We would like to use these data to evaluate

- the average gas mileage of all 2012 Prius drivers.

Do you think this is reasonable?

- Why or why not?

Answer:

(b) The EPA claims that a 2012 Prius gets 50 MPG

- (city and highway mileage combined).

Do these data provide strong evidence against this estimate

- for drivers who participate on fueleconomy.gov?
- Note any assumptions you must make as you proceed with the test.

Answer:

(c) Calculate a 95% confidence interval

- for the average gas mileage of a 2012 Prius
- by drivers who participate on fueleconomy.gov.

Answer:

5.0.0.5 Diamonds

5.0.0.5.1 Diamonds Part I. (OIS 5.28)

Prices of diamonds are determined by what is known as the 4 Cs:

- cut,
- clarity,
- color,
- and carat weight.

The prices of diamonds go up

- as the carat weight increases,
- but the increase is not smooth.

For example, the difference between the size

- of a 0.99 carat diamond and
  - a 1 carat diamond is undetectable to the naked human eye,
- but the price of a 1 carat diamond tends to be much higher
  - than the price of a 0.99 diamond.

In this question we use two random samples of diamonds,

- 0.99 carats and 1 carat,
- each sample of size 23,

and compare the average prices of the diamonds.

In order to be able to compare equivalent units,

-we first divide the price for each diamond - by 100 times its weight in carats.

That is, for a 0.99 carat diamond, we divide the price by 99.

For a 1 carat diamond, we divide the price by 100.

The distributions and some sample statistics are shown below.(Wickham 2016)

(a) Conduct a hypothesis test to evaluate

- if there is a difference between the average standardized prices
- of 0.99 and 1 carat diamonds.

Make sure to

- state your hypotheses clearly,
- check relevant conditions,

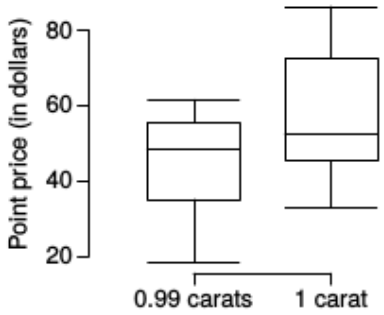


Figure 4: Point Price

	0.99 carats	1 carat
Mean	$ 44.51	$ 56.81
SD	$ 13.32	$ 16.13
n	23	23

Figure 5: Sample Statistics

- and interpret your results in context of the data.

Answer:

5.0.0.5.2 Diamonds Part II. (OIS 5.30)

We discussed diamond prices

- (standardized by weight)
- for diamonds with weights 0.99 carats and 1 carat.

See the table for summary statistics,

- and then construct a 95% confidence interval
- for the average difference
  - between the standardized prices of 0.99 and 1 carat diamonds.

You may assume the conditions for inference are met.

Answer:

5.0.0.6 Links

http://www.r-project.org

http://rmarkdown.rstudio.com/

https://www.openintro.org/stat/textbook.php?stat_book=os

5.0.0.7 References

"Gas Mileage of 2012 Toyota Prius." n.d. Accessed October 21, 2017. https://www.fueleconomy.gov/feg/bymodel/2012_Toyota_Prius.shtml.

Graybill, Franklin A., and Hariharan K. Iyer. 1994. *Regression Analysis: Concepts and Applications*. 1st edition. Belmont, Calif: Duxbury Pr.

Mitchell, Amy, and Emily Guskin. 2013. "Twitter News Consumers: Young, Mobile and Educated." *Pew Research Center's Journalism Project*. http://www.journalism.org/2013/11/04/twitter-news-consumers-young-mobile-and-educated/.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. 2nd ed. 2016 edition. New York, NY: Springer. https://github.com/hadley/ggplot2-book.