

# Homework 4

Rui Gao

2021-10-24

## Instructions

Use the R Markdown version of this file to complete and submit your homework. Items in **bold** require an answer. Make sure you change the author in the header to your own name.

## Conceptual Questions

1. In your own words, describe how understanding the behavior of a statistic over many samples (the sampling distribution) allows us to make inference about a population from a single sample.

Because the sample can inherit some features from the population, as the number of samples we collect increase, we should have more confidence to conclude features of the population.

2. A random sample of 10 college graduates is selected to play Sudoku, with each graduate attempting the same puzzle. The mean time it takes them to correctly complete the puzzle is 5 min.
  - a) **Which number should we use to estimate the population mean time it takes college graduates to finish this puzzle? Why?** The 5 min. By using the point estimate.
  - b) **Why is there uncertainty in this estimate?** Because there still could exist errors, which is the standard error.
  - c) **How much the standard error of the estimate will change if we increase the sample size to 1000 graduates?** It will decrease after we increase the size of samples.
3. A local restaurant is worried its potato supplier is skimping on the bags of potatoes it supplies. They claim each bag weighs 15lbs. The restaurateur understands that its hard to get a whole number of potatoes to weigh exactly 15lbs, but they feel it should average out. Over a month they weigh every bag they receive (and assume this is a random sample from all bags). They find a 95% confidence interval of the mean weight is 14.5 lbs to 14.9 lbs.
  - a) **Would they reject or fail to reject the null hypothesis (at the 5% level) that the mean weight is 15lbs?** They would reject the null hypothesis.
  - b) **What can you say about the p-value from the hypothesis test?** The p-value should greater than 0.05.
  - c) **What can you say about the sample mean from the measured bags?** The sample mean should stay between 14.5 to 14.9 lbs.
  - d) **What can you say about the p-value from the hypothesis test where the null hypothesis is the mean weight is 14.8 lbs?** Under this condition, the p-value should smaller than 0.05.

- e) What would happen to the interval, if they increased the level of confidence to 99%? The interval should be more wider.
- f) The restaurateur presents this evidence to the supplier and asks for a refund. If you were the supplier, are there any problems you would bring up that cast doubt on the analysis? A result can be statistically significant but not practically significant.

## R Questions

1. This question uses the `gifted` dataset, part of the `openintro` R package. You will need to install the package (do this once and don't store it in your R markdown file):

```
install.packages("openintro")
```

Then load the package:

```
library("openintro")
```

You will then have access to the `gifted` data frame

```
head(gifted) # the first six rows
```

```
## # A tibble: 6 x 8
##   score fatheriq motheriq speak count  read edutv cartoons
##   <int>    <int>    <int> <int> <int> <dbl> <dbl>    <dbl>
## 1   159     115     117   18   26   1.9   3        2
## 2   164     117     113   20   37   2.5   1.75     3.25
## 3   154     115     118   20   32   2.2   2.75     2.5
## 4   157     113     131   12   24   1.7   2.75     2.25
## 5   156     110     109   17   34   2.2   2.25     2.5
## 6   150     113     109   13   28   1.9   1.25     3.75
```

The column `motheriq` contains the mother's IQ for 36 gifted children. You can read more about the data set by looking at the help:

```
?gifted
```

We are interested in whether the mothers of gifted children have an IQ higher than the population at large, which is 100.

- a) **State the null and alternative hypothesis in statistical notation, and in words.**  $H_0: \mu_0 = 100$  vs  $H_a: \mu_a > 100$ . The null hypothesis is the mothers of gifted children have average IQ, which is 100, the alternative hypothesis is the mothers of gifted children have higher IQ.
- b) **Describe the test statistic you will use, and calculate it.** First, extract the mothers' IQ values, Second, calculate the mean of them, Third, calculate the z and p value, and set the confident level, Forth, get the interval and results

```
x <- gifted[, "motheriq"]
typeof(x)
```

```
## [1] "list"
```

```
mom_iq <- unlist(x)
mean(mom_iq)
```

```
## [1] 118.1667
```

c) Find the p-value for the test.

```
z <- (mean(mom_iq) - 100) / (sd(mom_iq)/sqrt(length(mom_iq)))
z
```

```
## [1] 16.75649
```

```
p <- pnorm(abs(z), mean = 0, sd = 1, lower.tail = FALSE)
p
```

```
## [1] 2.538738e-63
```

d) Calculate a point estimate and a 95% confidence interval for the mean IQ of mothers of gifted children.

```
mean(mom_iq) - qnorm(0.975) * sd(mom_iq)/sqrt(length(mom_iq))
```

```
## [1] 116.0418
```

```
mean(mom_iq) + qnorm(0.975) * sd(mom_iq)/sqrt(length(mom_iq))
```

```
## [1] 120.2916
```

The confidence interval is (116.0418, 120.2916).

e) **Give a non-technical summary of your findings.** Make sure to include both a summary of the test result and the confidence interval.

For the gifted children's mothers, we are 95% sure that their IQ values could be between 116 to 120, which means they are smarter than normal.