Homework 4

Nora Quick

2021-10-16

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.

If we understand the behavior of a statistic over many samples than we probably have a relatively good understanding of the subject within the sample. Understanding the behavior of something can help in a lot of ways in determening future actions or reactions. If we know actions and reactions based on the sampling distribution we can make educated inferences on the population from a single sample because we will already understand the subjects in the population and how they'll act and react.

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?**

   We would take the sample mean of 5 minutes. It's an estimate of how long the population took to complete it.

   b) **Why is there uncertainty in this estimate?**

   There is uncertainty because we 1. only observed a small number of students solving the puzzle and 2. because depending on the student there is a difference in how long it will take causing a variation.

   c) **How much the standard error of the estimate will change if we increase the sample size to 1000 graduates?**

   sqrt(1000/10) It would decrease 10 times.

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?**

15 is not within the 95% range, therefore, we would reject the null hypothesis.

b) **What can you say about the p-value from the hypothesis test?**

The p-value would be 0.05 (based on 95%).

c) **What can you say about the sample mean from the measured bags?**

It will the the middle of the range they found so the smaple mean would be 14.7 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?**

14.8 lbs is within the 95% interval so we can say the p-value is larger that 0.05.

e) **What would happen to the interval, if they increased the level of confidence to to 99%?**

If it was increased to 99% then the interval/range will be larger.

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?**

My first concern would be the sample size. How many bags did they measure? Personally, I would want more bags over the course of more months to get a better idea of the weight. Additionally, what does the supplier promise? Do they say exactly 15 lbs or about 15 lbs.

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.**

Null Hypothesis: $\mu = 100$ The population mean of the IQ of mothers with gifted children in equal to 100.

Alternative hypothesis: $\mu > 100$ The population mean of the IQ of mothers with gifted children is greatere than 100.

b) **Describe the test statistic you will use, and calculate it.**

We would want to find the z-statistic.

```r
x_bar <- mean(gifted$motheriq)
mu <- 100
s <- sd(gifted$motheriq)
sq_n <- sqrt(nrow(gifted))

Z <- ((x_bar - mu) / (s / sq_n))

Z
```

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

c) **Find the p-value for the test.**

```
p_val <- pnorm(Z, mean = 0, sd = 1, lower.tail = FALSE)

p_val
```

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

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

```
#PE
x_bar
```

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

```r
#95% CI
upper <- x_bar + qnorm(0.975) * (s / sq_n)
lower <- x_bar - qnorm(0.975) * (s / sq_n)
c(lower, upper)
```

```
## [1] 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.

There is convincing evidence that the IQ of mothers with gifted children is higher than 100 so we should reject the null hypothesis (100 is not within the 95% range we found). We found that the estimated IQ is about 118 and the 95% confidence interval would be between 116 and 120.