Review Section

Bolun Zhang, UCSD Sociology

2/1/2021

Outline

Here we will go through the main points in descriptive statistics, sampling distribution and also point estimator.

- Mean, standard deviation, median, mode, and range of your sample dataset.
- Identify the lower quartile, upper quartile, and interquartile range of your sample dataset.
- Sampling Distribution and Point Estimates
  - What is the estimated standard error of the sample mean?
  - Using the t distribution table (from the textbook appendix or the lecture notes), identify a 95% confidence interval for the population mean.

So, Let's imagine a scenario

We are back in the 1000-population island

```
set.seed(92037)
pop_height <- rnorm(1000, mean = 1.7, sd = 0.15)
pop_height <- round(pop_height, 2)
sample_height <- sample(pop_height, 20)
sample_height
```

```
##  [1] 1.70 2.06 2.01 1.49 1.71 1.60 1.64 1.79 1.36 1.74 1
## [16] 1.64 1.63 1.61 1.43 1.71
```

Mean

We can calculate the mean by

$$Mean = \sum height/20$$

$$= (1.70 + 2.06 + 2.01 + 1.49 ...... + 1.61 + 1.43 + 1.71)/20$$

Mean

The final result is

```
## [1] 1.6815
```

Median

We first arrange the 20 people according to their heights

```
##  [1] 1.36 1.43 1.43 1.49 1.57 1.60 1.61 1.63 1.64 1.64 1
## [16] 1.79 1.90 1.92 2.01 2.06
```

The median is

$$(1.64 + 1.69)/2 = 1.665$$

```
median(sample_height)
```

```
## [1] 1.665
```

Mode

So, what is the mode in this sample?

The Range of the the sample

```
## [1] "the minimum is 1.36"

## [1] "the maximum is 2.06"
```

Standard Deviation

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}}$$

In our case, it equals

$$s = \sqrt{\frac{(1.36 - 1.6815)^2 + (1.43 - 1.6815)^2 ... (2.06 - 1.6815)^2}{20}}$$

Standard Deviation

The result is

```
## [1] "The standard deivation of the sample is 0.18301024
```

If you tried:

```
sd(sample_height)
```

```
## [1] 0.1877646
```

Why are they different?

Quartile

The lower quartile is

$$(1.57 + 1.60)/2 = 1.585$$

and the upper quartile is

$$(1.74 + 1.79)/2 = 1.765$$

You might wonder. . . if you tried

Just want you know, quartile is actually a more complicated issue. If you compare the result from the textbook method and result from the code:

```
quantile(sample_height)
```

```
##     0%    25%    50%    75%   100% 
## 1.3600 1.5925 1.6650 1.7525 2.0600
```

For more, please refer to https://en.m.wikipedia.org/wiki/Quartile. R is using method 4 to calculate the quartile.

Estimated Standard Error

As a reminder, the old-fashioned formula to estimate standard error is

$$\sigma_{\bar{x}} \approx \frac{s}{\sqrt{n}}$$

Thus, we can calculate the standard error $\sigma = 0.183/\sqrt{20}$

```
## [1] 0.04092233
```

Student Distribution

We simulate this in the last workshop.

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

Thus you can understand why the confidence intervals are $\pm T_{0.025} * SE$.

Here the degree of freedom is $20 - 1 = 19$, we can find the t-score in the table.

```
## [1] 0.3936003
```

Thus the confidence intervals of the sample mean is $\pm(0.394 * 0.183)$