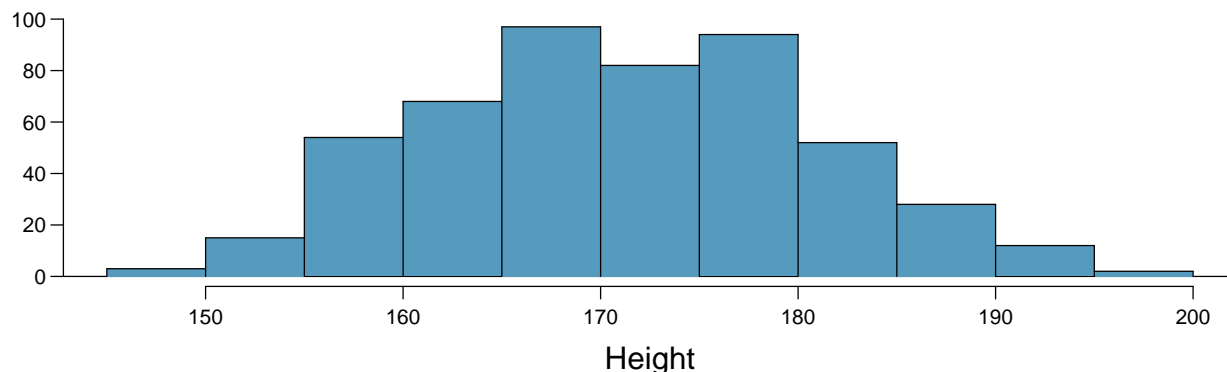# Chapter 5 - Foundations for Inference

Samuel Kigamba

Heights of adults. (7.7, p. 260) Researchers studying anthropometry collected body girth measurements and skeletal diameter measurements, as well as age, weight, height and gender, for 507 physically active individuals. The histogram below shows the sample distribution of heights in centimeters.

(a) What is the point estimate for the average height of active individuals? What about the median?

> The average height or the mean is 171.1cm and the median is 170.3cm.

```
summary(bdims$hgt)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   147.2   163.8   170.3   171.1   177.8   198.1
```

(b) What is the point estimate for the standard deviation of the heights of active individuals? What about the IQR?

```
sd(bdims$hgt)
```

```
## [1] 9.407205
```

```
IQR(bdims$hgt)
```

```
## [1] 14
```

(c) Is a person who is 1m 80cm (180 cm) tall considered unusually tall? And is a person who is 1m 55cm (155cm) considered unusually short? Explain your reasoning.

Both individuals would not be considered to be unusually tall or short since they are within two standard deviations of the mean.

```
tall <- (mean(bdims$hgt) - 180) / sd(bdims$hgt)
short <- (mean(bdims$hgt) - 155) / sd(bdims$hgt)

c(short, tall)
```

```
## [1]  1.7161087 -0.9414287
```

(d) The researchers take another random sample of physically active individuals. Would you expect the mean and the standard deviation of this new sample to be the ones given above? Explain your reasoning.

```
I would expect the mean and standard deviation to be close to but not exactly as the ones above due
```

(e) The sample means obtained are point estimates for the mean height of all active individuals, if the sample of individuals is equivalent to a simple random sample. What measure do we use to quantify the variability of such an estimate (Hint: recall that $SD_x = \frac{\sigma}{\sqrt{n}}$)? Compute this quantity using the data from the original sample under the condition that the data are a simple random sample.

```
This is the sample error (SE) computed as a product of the standard deviation (sd) of the sample ar
```

```
n <- 507
SE_height <- sd(bdims$hgt)/sqrt(n)
SE_height
```

```
## [1] 0.4177887
```

Thanksgiving spending, Part I. The 2009 holiday retail season, which kicked off on November 27, 2009 (the day after Thanksgiving), had been marked by somewhat lower self-reported consumer spending than was seen during the comparable period in 2008. To get an estimate of consumer spending, 436 randomly sampled American adults were surveyed. Daily consumer spending for the six-day period after Thanksgiving, spanning the Black Friday weekend and Cyber Monday, averaged $84.71. A 95% confidence interval based on this sample is ($80.31, $89.11). Determine whether the following statements are true or false, and explain your reasoning.
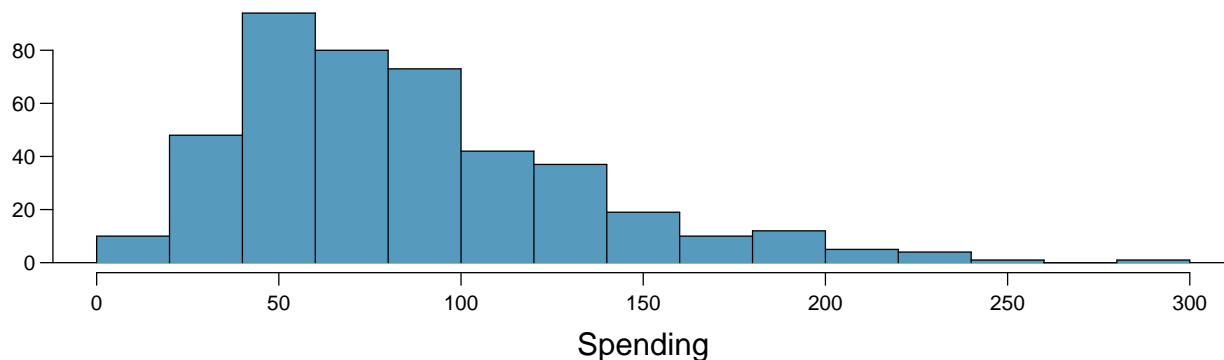
(a) We are 95% confident that the average spending of these 436 American adults is between $80.31 and $89.11.

    `False. A true definition of confidence interval refers to the entire population and not specific to`

(b) This confidence interval is not valid since the distribution of spending in the sample is right skewed.

    `False. As long as the sample was randomly selected and is large enough that is it passes the succes`

(c) 95% of random samples have a sample mean between $80.31 and $89.11.

    `False. This is not a true definition of the confidence interval.`

(d) We are 95% confident that the average spending of all American adults is between $80.31 and $89.11.

    `True. This is the true definition of the confidence interval.`

(e) A 90% confidence interval would be narrower than the 95% confidence interval since we dont need to be as sure about our estimate.

    `True. A smaller confidence interval is narrower, as noted we dont need to be very sure of our estim`

(f) In order to decrease the margin of error of a 95% confidence interval to a third of what it is now, we would need to use a sample 3 times larger.

    `False. we need a sample size equal to the square of the factor by which we need to reduce our error`

(g) The margin of error is 4.4.

    `True. For the given confidence interval of 95% the margin of error is 4.4.`

```r
Lower_inter <- 80.31
Upper_inter <- 89.11
SE_spend <- (Upper_inter - Lower_inter)/2
SE_spend
```

```
## [1] 4.4
```

Gifted children, Part I. Researchers investigating characteristics of gifted children col- lected 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 dis- tribution of the ages (in months) at which these children first counted to 10 successfully. Also provided are some sample statistics.

Age child first counted to 10 (in months)
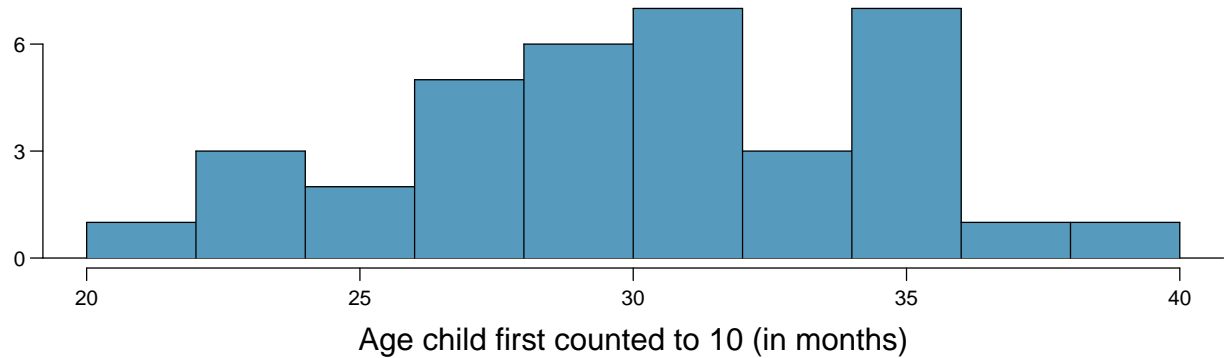
n	36
min	21
mean	30.69
sd	4.31
max	39

(a) Are conditions for inference satisfied?

Independence is satisfied - This is a random sample selection Sample size is over 30 and no strong slewness is evident. This this sample satisfies inference conditions.

(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.

(c) Interpret the p-value in context of the hypothesis test and the data.

(d) Calculate a 90% confidence interval for the average age at which gifted children first count to 10 successfully.

(e) Do your results from the hypothesis test and the confidence interval agree? Explain.

Gifted children, Part II. Exercise above describes a study on gifted children. In this study, along with variables on the children, the researchers also collected data on the mothers and fathers IQ of the 36 randomly sampled gifted children. The histogram below shows the distribution of mothers IQ. Also provided are some sample statistics.
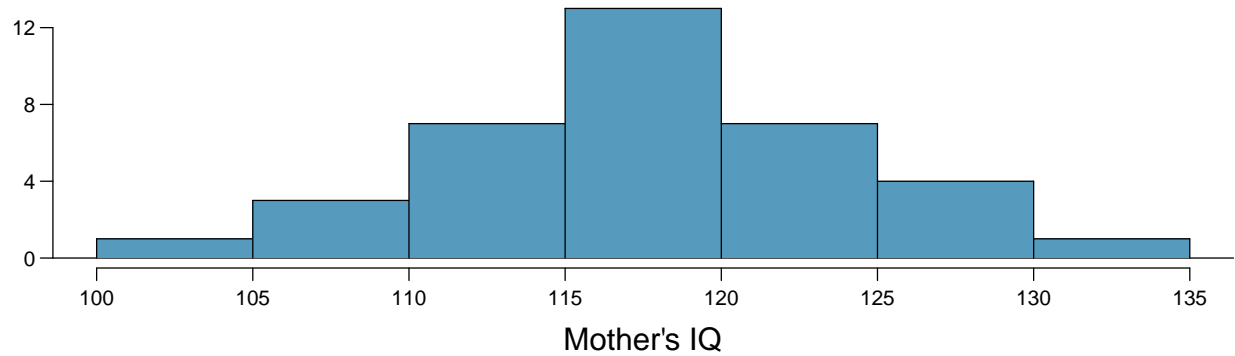
Mother's IQ

n	36
min	101
mean	118.2
sd	6.5
max	131

(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.
(b) Calculate a 90% confidence interval for the average IQ of mothers of gifted children.
(c) Do your results from the hypothesis test and the confidence interval agree? Explain.

CLT. Define the term sampling distribution of the mean, and describe how the shape, center, and spread of the sampling distribution of the mean change as sample size increases.

CFLBs. A manufacturer of compact fluorescent light bulbs advertises that the distribution of the lifespans of these light bulbs is nearly normal with a mean of 9,000 hours and a standard deviation of 1,000 hours.

(a) What is the probability that a randomly chosen light bulb lasts more than 10,500 hours?
(b) Describe the distribution of the mean lifespan of 15 light bulbs.
(c) What is the probability that the mean lifespan of 15 randomly chosen light bulbs is more than 10,500 hours?
(d) Sketch the two distributions (population and sampling) on the same scale.
(e) Could you estimate the probabilities from parts (a) and (c) if the lifespans of light bulbs had a skewed distribution?

Same observation, different sample size. Suppose you conduct a hypothesis test based on a sample where the sample size is n = 50, and arrive at a p-value of 0.08. You then refer back to your notes and discover that you made a careless mistake, the sample size should have been n = 500. Will your p-value increase, decrease, or stay the same? Explain.