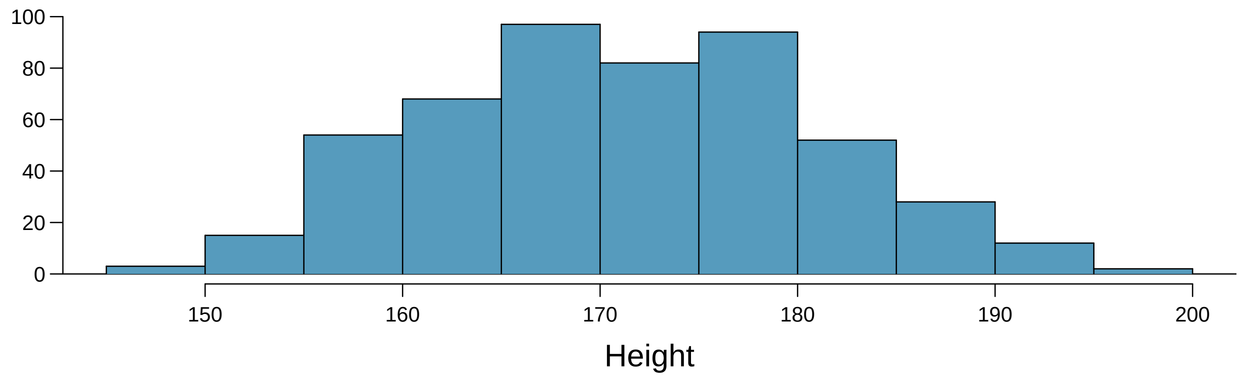
Chapter 5 - Foundations for Inference

Adam Gersowitz

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?

```
## [1] 171.1438
```

```
## [1] 170.3
```

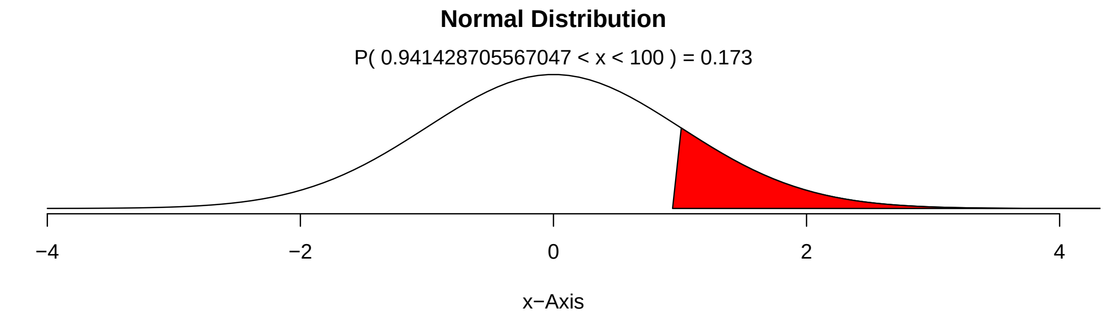
(b) What is the point estimate for the standard deviation of the heights of active individuals? What about the IQR?

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

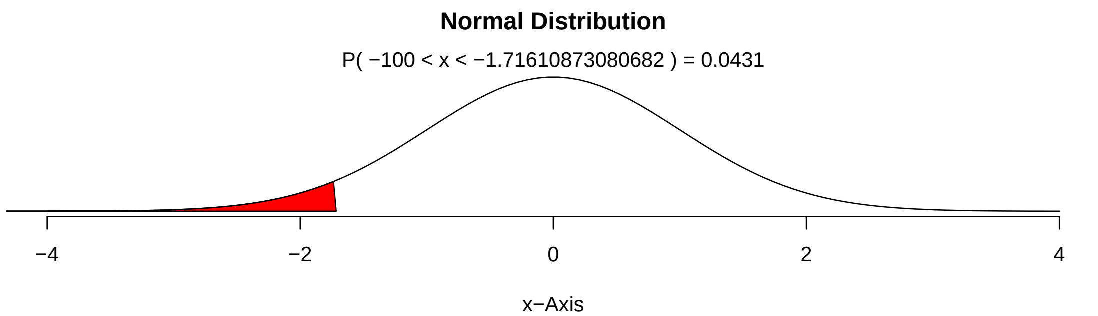
```
## [1] 14
```

(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.

```
## [1] 0.9414287
```




```
## [1] -1.716109
```



No, someone who is 180 cm is not abnormally tall (they ae stil shorter then 17% of the population). Yes, someone who is 155 cm is unusually short as they are shorter than 95% of the population.

(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.

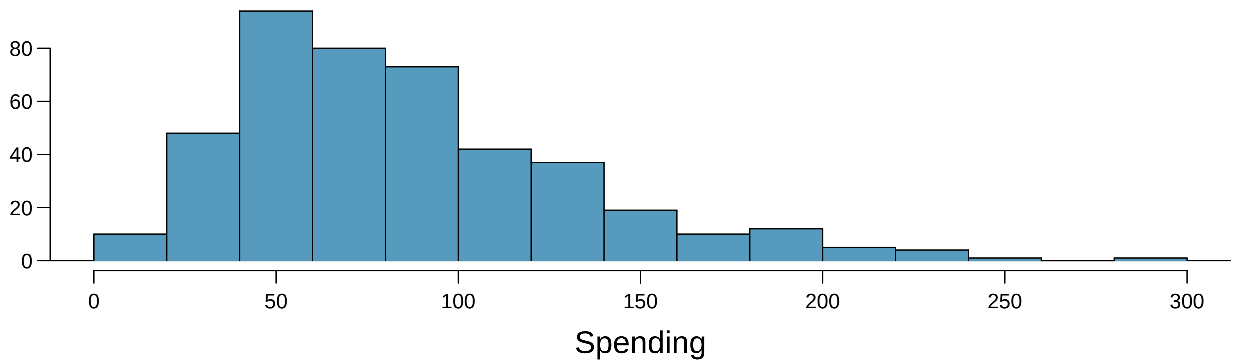
No, they will likely be similar but they will not be the same. They aer unlikely to be close unless the samlpling size is close to the population size.

(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.

We would use standard error.

```
## [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, we are 95% confident that the population mean falls within this range based on our sample data.

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

False, it is valid because the sample size is large and the observations are independent so we can use the confidence interval.

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

False, the 95% CI refers to the population mean not of random samples.

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

True, see answers above.

(e) A 90% confidence interval would be narrower than the 95% confidence interval since we don't need to be as sure about our estimate.

True. it would be only contianing 90% of the values and ignoring tails of 5% rather than 2.5%.

(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 it would be need to be increased by a factor of 3ˆ2 = 9

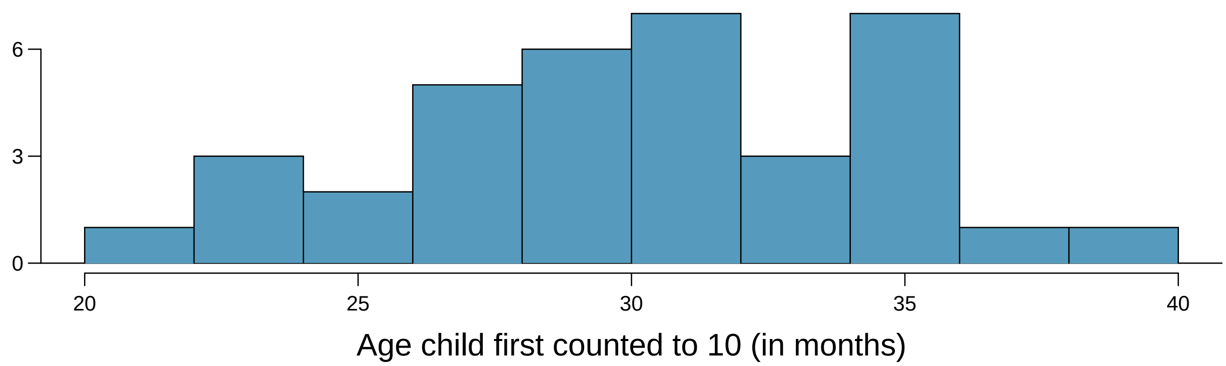
(g) The margin of error is 4.4.

TRUE

```
## [1] 2.247468
```

```
## [1] 4.405038
```

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.



n	36
min	21
mean	30.69
sd	4.31
max	39

(a) Are conditions for inference satisfied?

Yes, it is a random smaple with a sample size > 30 and the observations are independent.

(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.

```
## [1] -1.823666
```

```
## [1] 0.0341013
```

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

The p-value (0.03) is less than the significance level of 0.10 which means we reject the null hypothesis and conclude that the mean for gifted children is statistically significant difference from the average child.

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

The 90% confidence interval range is 29.5 to 31.9.

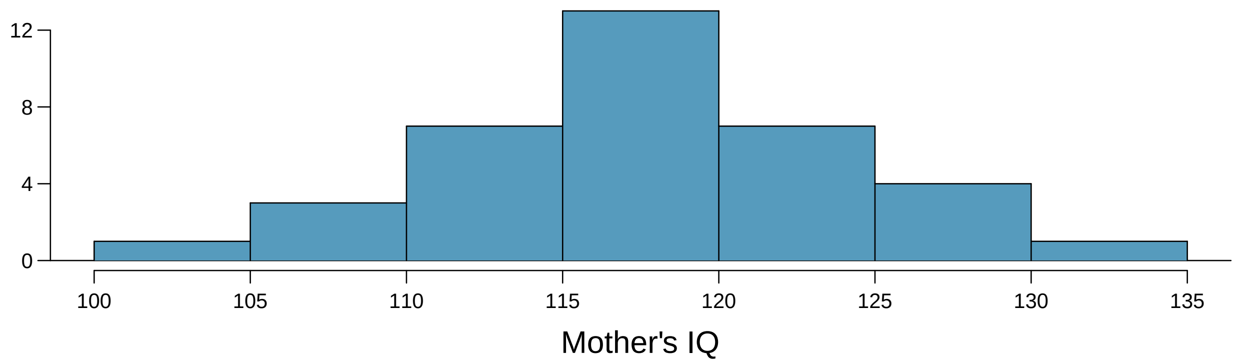
```
## [1] 29.507
```

```
## [1] 31.873
```

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

Yes, the upper bound of the 90% CI for the average time it takes a gifted child to count to ten is 31.873. This is still lower than the time for an average child 32. This means I am 95% confident that the average time it takes the population of gifted children to count to ten is less than 31.873 months.

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 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.



n	36
min	101
mean	118.2
sd	6.5
max	131

(a) Performahypothesistesttoevaluateifthesedataprovideconvincingevidencethattheaverage 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.

```
## [1] 16.8
```

```
## [1] 0
```

We reject the null hypothesis that the IQ of their mother is not statistically sifferent that the average IQ for the population at large because the p-value of 0 is much lower then the signficance value of .10. Additionally, the z-score of 16.8 is much higher then 1.960.

(b) Calculate a 90% confidence interval for the average IQ of mothers of gifted children.

The 90% confidence interval is from 116.42 to 119.98

```
## [1] 116.4179
```

```
## [1] 119.9821
```

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

Yes they strongly agree. The lower bound of the 90% confidence interval(116.4) is much higher than the average for the general population (100). This would suggest that the rejection of the null hypothesis as the result of the hypothesis test was the correct conclusion.

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.

The sampling distribution of the mean is an unbiased estimator that is centered at the populaiton averagee. The spread is othe variability that is introduced when sampling a certain number of observations form the population. When the sample size is larger the center gets closer to the population mean, the shape becomes more normally distributed and the spread lessens and gets more compact.

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?

```
## [1] 0.0668072
```

There is about a 6.7% chance that the lightbulb lasts more than 10,500 hours.

(b) Describe the distribution of the mean lifespan of 15 light bulbs.

The distribution will be normally distributed like the population. We can assume that with enough samples that the mean will be the same and the sd will be smaller then the population

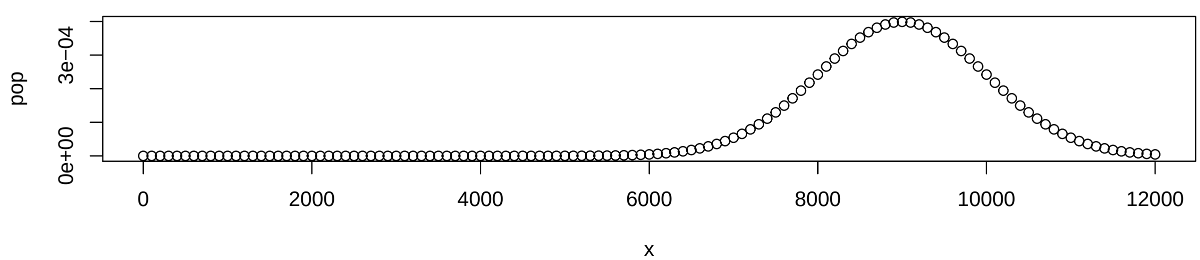
(c) What is the probability that the mean lifespan of 15 randomly chosen light bulbs is more than 10,500 hours?

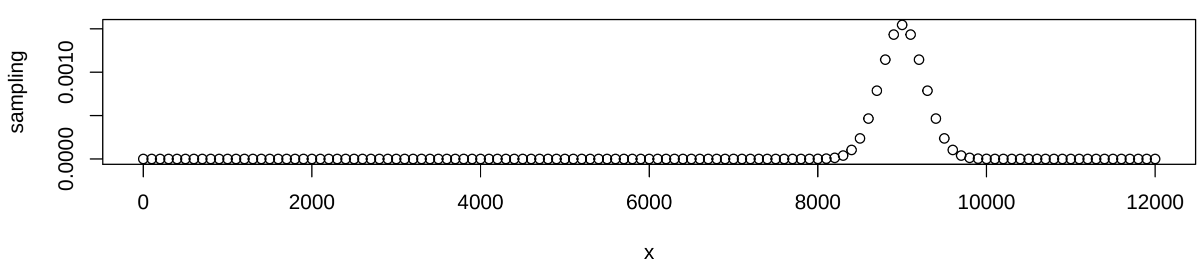
```
## [1] 258.1989
```

```
## [1] 3.133452e-09
```

The probability that the mean lifespan of 15bulbs will be over 10500 is nearly 0.

(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?

Yes, but we would need the sample size to be very large.

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.

This will depend on the null and alternate hypothesis. If the null hypothesis is flase then The p-value would become smaller. As the sample size increases the uncertainty about the population decreases so out p-value will chang ein way that will make it easier to reject or support a null hypothesis.