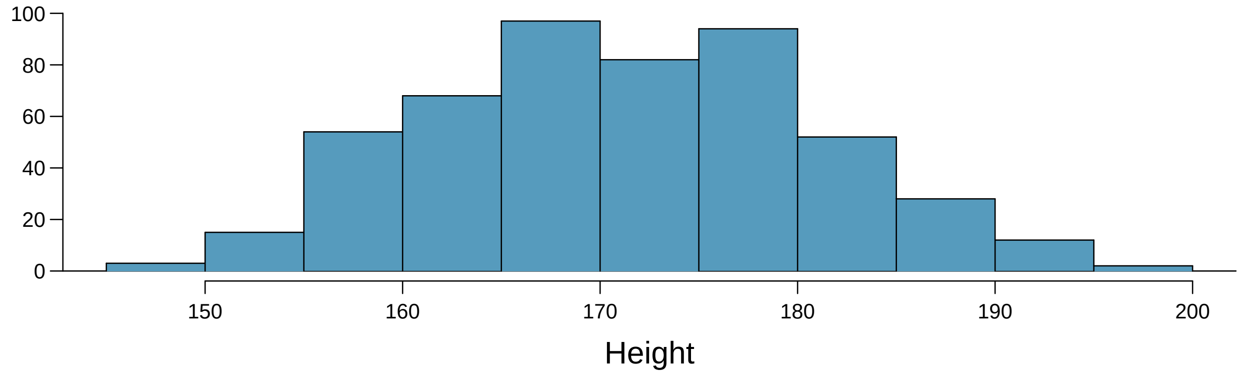
Homework 5

Irene Jacob

2020-10-04

Heights of adults

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?

```
mean(bdims$hgt)
```

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

```
median(bdims$hgt)
```

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

(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
```

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

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

IQR = 163.8 - 177.8 cms

(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.

A person who is 180 cm tall is only 1SD away from the mean of 171.1 as the sd is 9.4 so he/she is not unusually tall. But a 155 cm tall person is 2SDs away from mean so he/she is unusually short.

(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.

The new sample is totally different so the values will vary but may be close to the values in this sample.

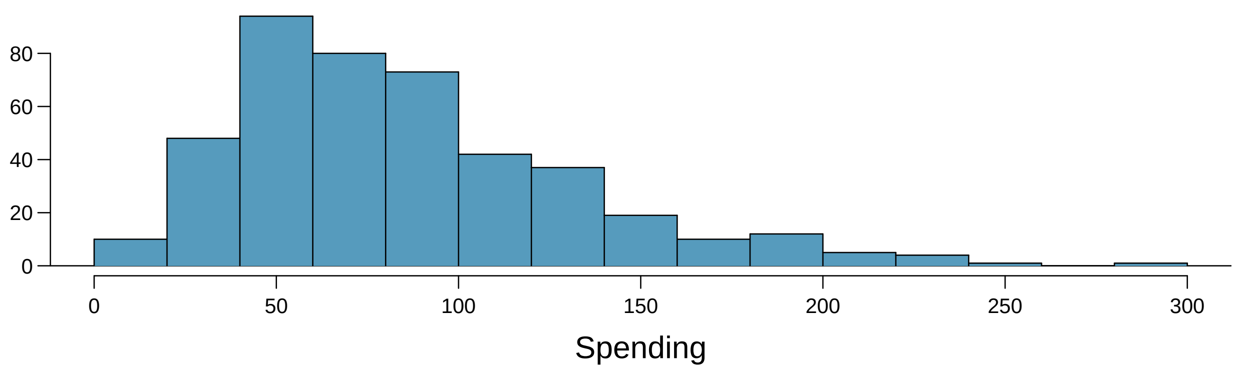
(e) The sample means obtained are point estimates for the mean height of all active individuals, if the sample of individuals is equivalent to a simple random sample. What measure do we use to quantify the variability of such an estimate (Hint: recall that $SD_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$)? Compute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

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

```
## [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

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

The sample is large enough so FALSE.

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

The confidence interval is not about the sample mean so FALSE.

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

TRUE

(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

(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.

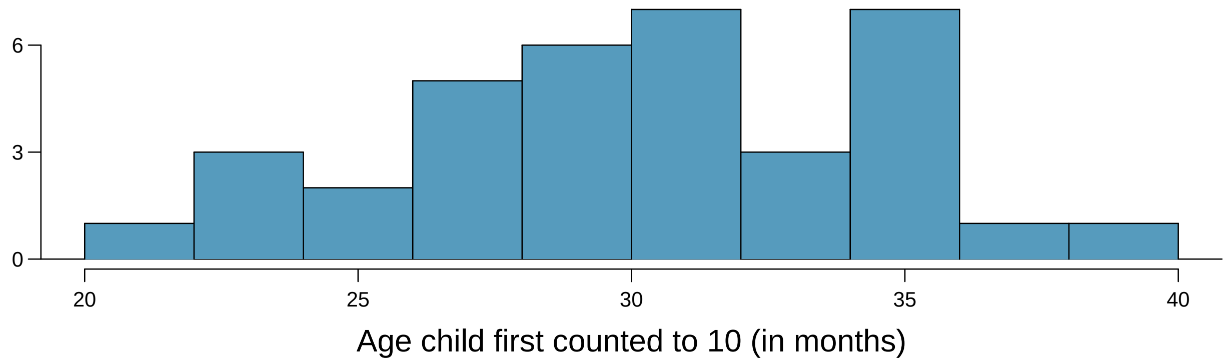
FALSE as the same needs to 3^2 times larger to decrease the error to a third

(g) The margin of error is 4.4.

TRUE

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 as the sample is selected at random and the size of the sample is large enough

(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.

The confidence level is 0.90 as the significance level is 0.10.

```
n <-36
m <-30.69
sd <-4.31

x<- m - 1.645 * sd / sqrt(n)
y<- m + 1.645 * sd / sqrt(n)

c(x, y)
```

```
## [1] 29.50834 31.87166
```

This shows that the general average of gifted children counting to 10 is between the months of 29.50834 and 31.87166.

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

```
z <- (m - 32) / (sd / sqrt(n))
z
```

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

```
pnorm(z)
```

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

as p value is less than 0.10 this hypothesis is rejected.

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

```
n <-36
m <-30.69
sd <-4.31

x<- m - 1.645 * sd / sqrt(n)
y<- m + 1.645 * sd / sqrt(n)

c(x, y)
```

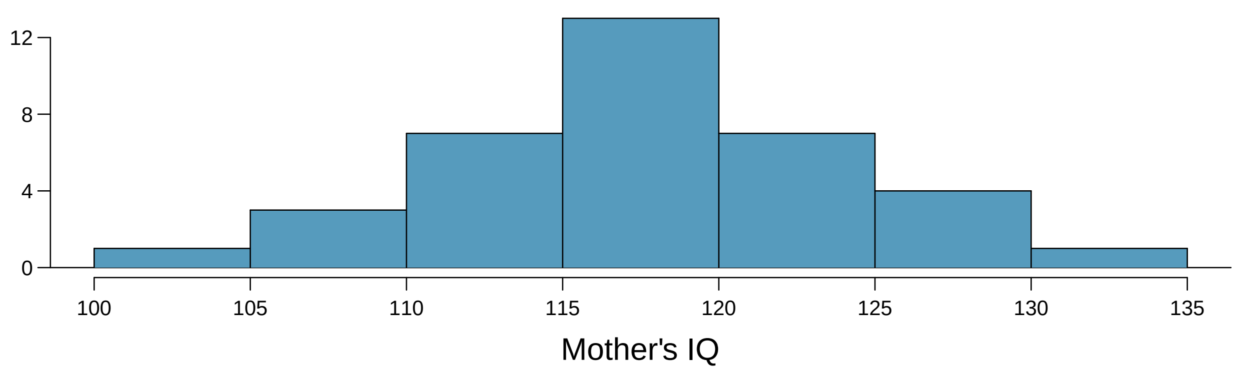
```
## [1] 29.50834 31.87166
```

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

Yes. 1-confidence level = 1-0.90=0.10 = significance level

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) 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.

```
n1 <- 36
sd1 <- 6.5
```

```r
ho <- 118.2
ha <- 100

pnorm(ha,ho,sd1)
```

```
## [1] 0.00255513
```

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

```r
SE = 1.645*sd1/(n1)**0.5
SE
```

```
## [1] 1.782083
```

```r
CI = ho + c(-SE,SE)
CI
```

```
## [1] 116.4179 119.9821
```

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

Yes. The probability of the mothers having IQ less than or equal to 100 is less than the significance level also 100 falls outside the confidence interval.

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.

Sampling distribution is the random selection of samples from a population to calculate the mean and then creating a distribution from the results of the mean values. As the sample size increases, the spread gets smaller and the shape will get taller and more sharp. The center will stay relatively the same as that is the population mean.

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?

```r
pnorm(10500, 9000, 1000, lower.tail = FALSE)
```

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

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

```r
sd3 <- 1000
m3 <- 9000

s <- sd3/sqrt(15)
s
```

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

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

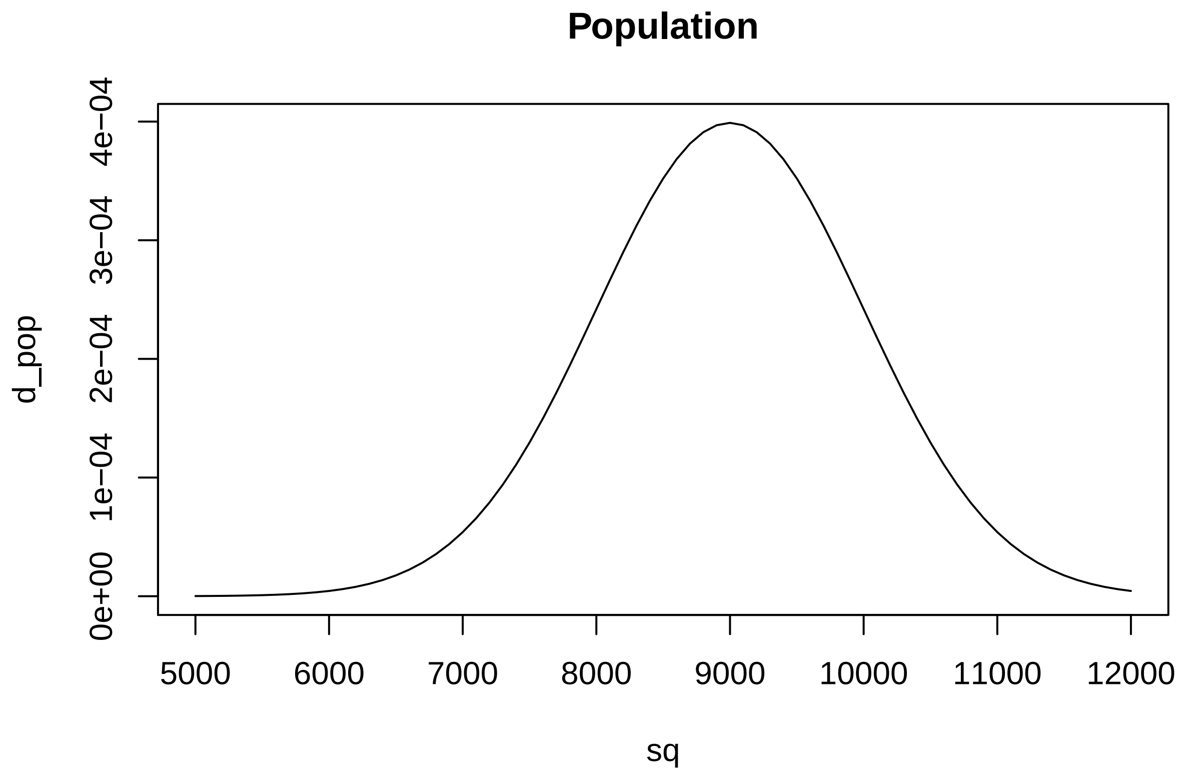
```r
pnorm(10500, 9000, 1000/sqrt(15), lower.tail = FALSE)
```

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

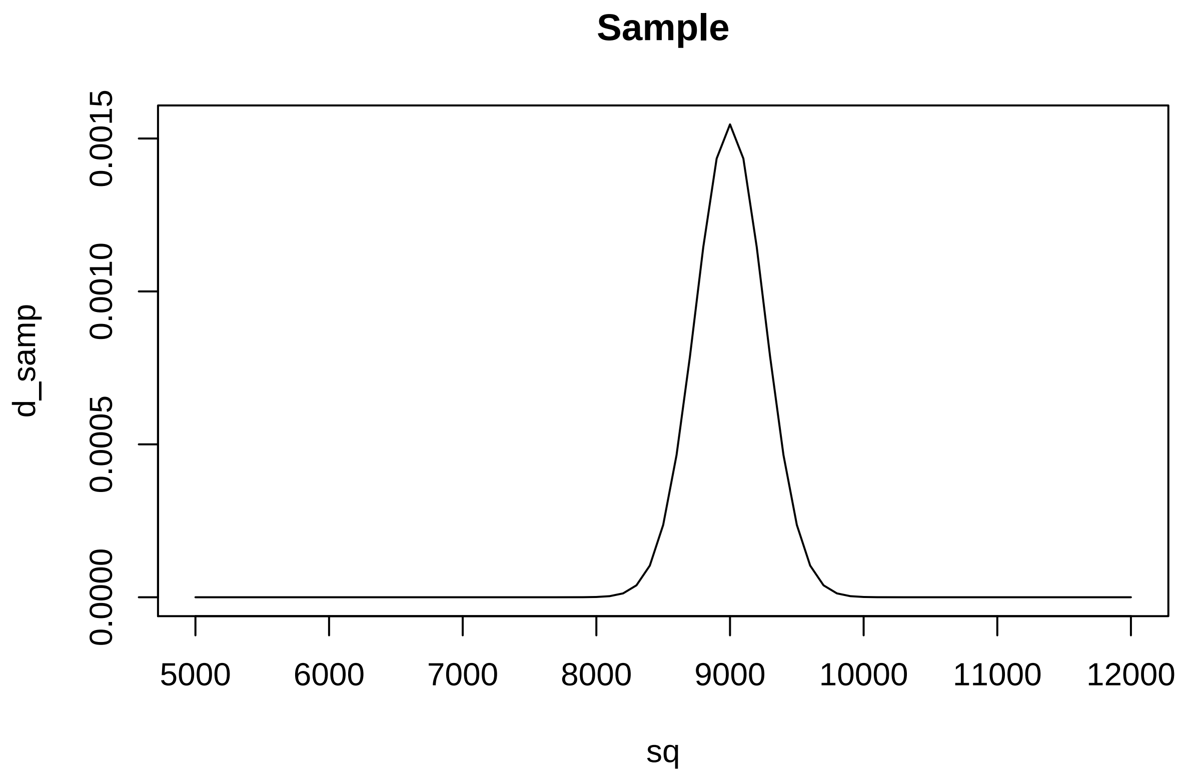
(d) Sketch the two distributions (population and sampling) on the same scale.

```
sq <- seq(5000,12000,100)
d_pop<- dnorm(sq, 9000,1000)
d_samp<- dnorm(sq, 9000, 258)

plot(sq, d_pop, type="l", main="Population")
```




```
plot(sq, d_samp, type="l", main="Sample")
```



(e) Could you estimate the probabilities from parts (a) and (c) if the lifespans of light bulbs had a skewed distribution?

It will be difficult as the sample size is small. If the size is increased it may be possible.

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.

p value decreases as sample size increases.