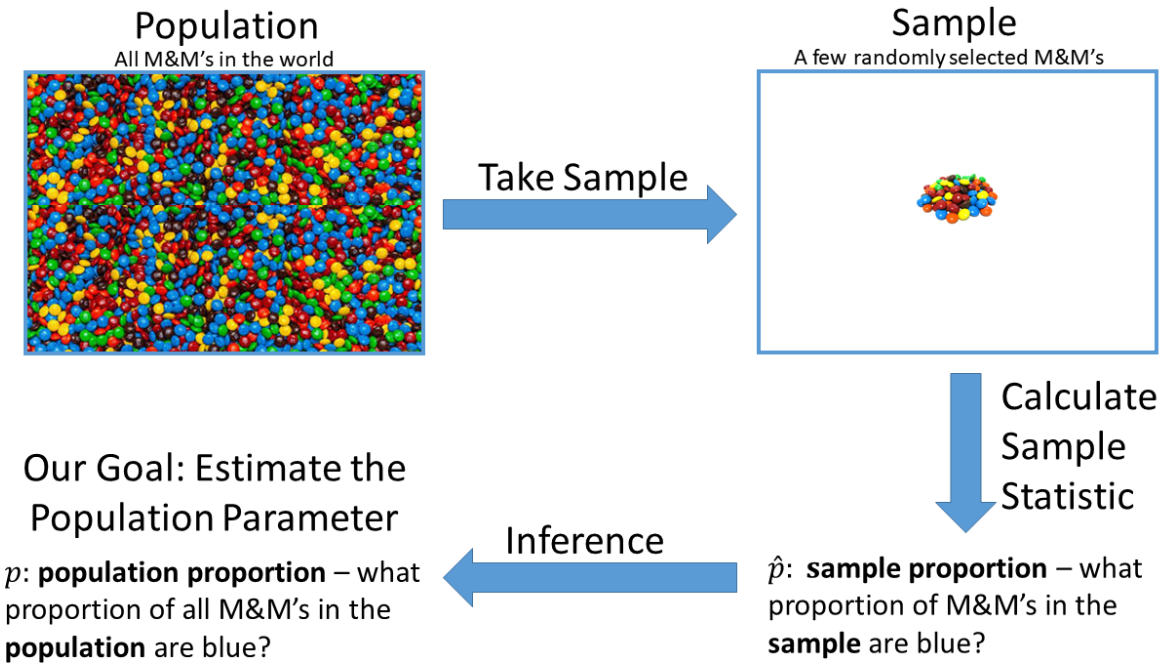
Intro to Inference for a Population Mean

Previously: Estimating a Population Proportion with the Sample Proportion

What Proportion of M&M's are blue?



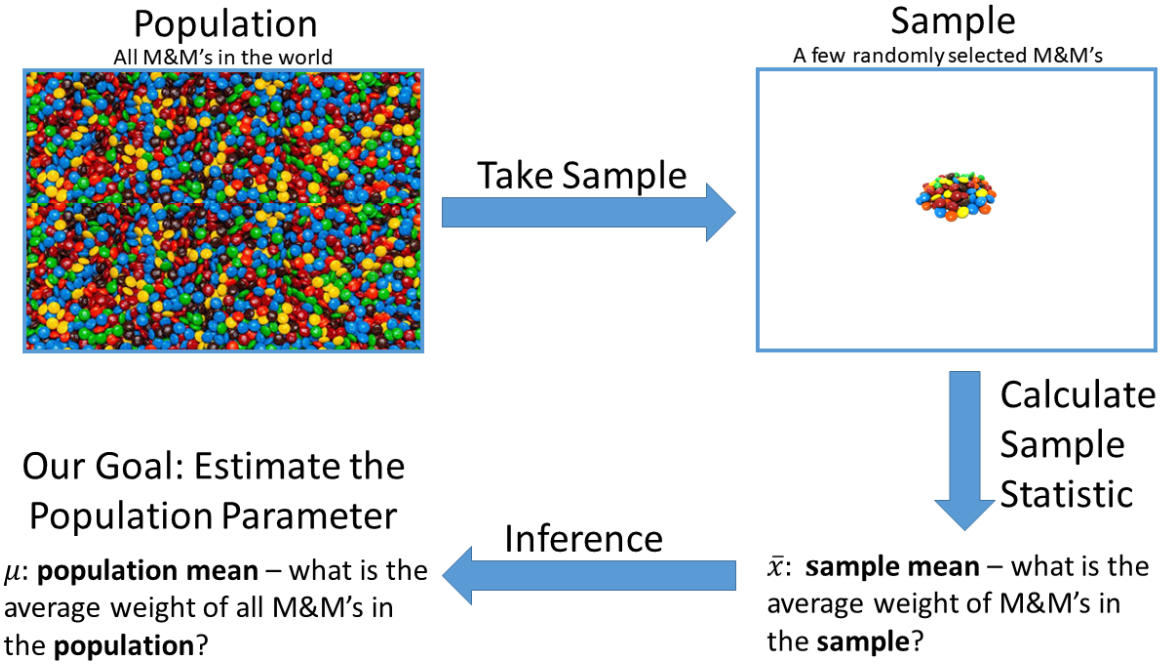
Key point: The variable we measured on each M&M was a categorical variable with two categories (or at least, two categories we cared about): is the M&M blue?

The population parameter we are interested in is **what proportion of the population have a characteristic**.

Next: Estimating a Population Mean with the Sample Mean

What is the average weight of all M&M's?

What is the average gestation time of all babies born in the United States?



Key point: Now, the variable we measure on each M&M is a quantitative variable: what is the M&M's weight in grams?

The population parameter we are interested in is a measure of **the center of the distribution of that quantitative variable**.

In this class, we will only discuss methods for making inference about the **mean** of a quantitative variable.

Example: Body Temperatures

It's generally believed that the average body temperature is 98.6 degrees Farenheit (37 degrees Celsius).

Let's investigate with measurements of the temperatures of 130 randomly selected adults. The following R code reads the data in and makes an initial plot:

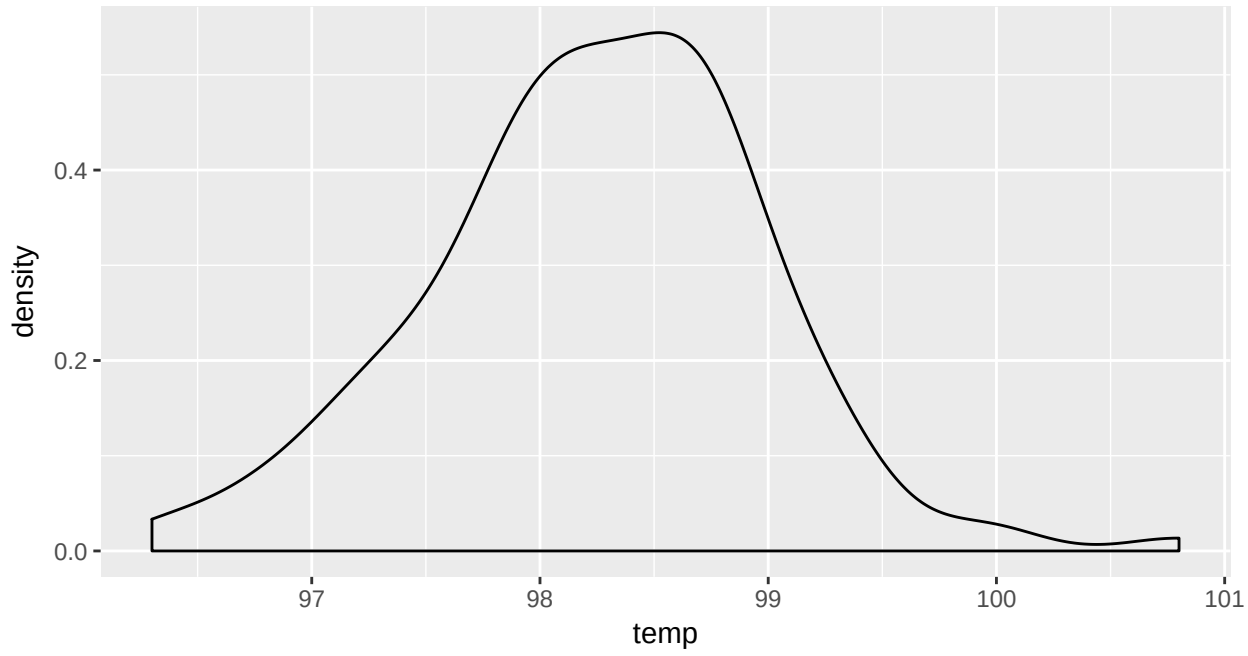
```
library(readr)
library(dplyr)
bodytemp <- read_csv("http://www.evanlray.com/stat140_s2018/data/jse/normtemp/bodytemp.csv")
```

```
## Parsed with column specification:
## cols(
##   temp = col_double(),
##   sex = col_character(),
##   hr = col_integer()
## )
```

```
bodytemp$sex <- factor(bodytemp$sex)
head(bodytemp)
```

```
## # A tibble: 6 x 3
##    temp    sex    hr
##   <dbl> <fctr> <int>
## 1  96.3   Male    70
## 2  96.7   Male    71
## 3  96.9   Male    74
## 4  97.0   Male    80
## 5  97.1   Male    73
## 6  97.1   Male    75
```

```
ggplot(data = bodytemp, mapping = aes(x = temp)) +
  geom_density()
```



1. What is the population?

2. What is the population parameter? What symbol will you use to denote the population parameter?

3. Are the conditions met for using the sample to learn about the population?

There are **four** things to think about:

1. Is the sample **representative** of the population? (always required)
   - Were we careful in **randomly selecting** the people to include in the sample?
   - Are there any forms of **bias** in the sampling? In this case, 23 people did not respond (there is some non-response). Does this matter?

2. Are the the people in the sample **independent** from each other? (required for all methods we will use in this class)
   - Does knowing one person's response change the information we have about another person's response?
   - Closely related to the question of randomization above.

3. Is the variable we have recorded a **quantitative variable for each person or thing in the sample**? (required for inference about a population mean to make sense)

4. Is the **mean a reasonable summary of the center**? (required for inference about a population mean to make sense)
   - Is the distribution approximately symmetric (or at least, not so skewed that the mean will be pulled away from the center)?
   - Is the distribution approximately unimodal (or at least, close enough that the mean is saying something meaningful about the distribution)?

4. Write the null and alternative hypotheses for a test of the claim that the population mean body temperature is 98.6 degrees F.

5. Write the R code to conduct this test and obtain a 99% confidence interval for the population mean body temperature.

```
library(mosaic)
t.test(~temp, mu = 98.6, conf.level = 0.99, data = bodytemp)
```

```
## ~temp
```

```
##
##  One Sample t-test
##
## data:  temp
## t = -5.4548, df = 129, p-value = 2.411e-07
## alternative hypothesis: true mean is not equal to 98.6
## 99 percent confidence interval:
##  98.08111 98.41735
## sample estimates:
## mean of x
##  98.24923
```

Note: Compare to the following output:

```
bodytemp %>%
  summarize(mean_temp = mean(temp))
```

```
## # A tibble: 1 x 1
##   mean_temp
##       <dbl>
## 1  98.24923
```

6. State a point estimate for the population mean body temperature.

7. State and interpret the p-value for the hypothesis test. What is your conclusion, using a significance level of $\alpha = 0.001$?

8. State and interpret a 99% confidence interval for the population mean body temperature.

9. What does the phrase "99% Confident" mean in the context of this example?