

## 9.1: Mean

The mean is defined as the sum of values divided by the number of values being summed:

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$$

Let's say that we want to obtain the mean height for adults in the NHANES database (contained in the data `Height`). We would sum the individual heights (using the `sum()` function) and then divide by the number of values:

```
sum(NHANES$Height)/length(NHANES$Height)
```

```
## [1] NA
```

This returns the value NA, because there are missing values for some rows, and the `sum()` function doesn't automatically handle those. To address this, we could filter the data frame using `drop_na()` to drop rows with NA values for this variable:

```
height_noNA <- NHANES %>%  
  drop_na(Height) %>%  
  pull(Height)  
  
sum(height_noNA)/length(height_noNA)
```

```
## [1] 160
```

There is, of course, a built-in function in R called `mean()` that will compute the mean. Like the `sum()` function, `mean()` will return NA if there are any NA values in the data:

```
mean(NHANES$Height)
```

```
## [1] NA
```

The `mean()` function includes an optional argument called `na.rm` that will remove NA values if it is set to TRUE:

```
mean(NHANES$Height, na.rm=TRUE)
```

```
## [1] 160
```

This page titled [9.1: Mean](#) is shared under a [CC BY-NC 4.0](#) license and was authored, remixed, and/or curated by [Russell A. Poldrack](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.