

## 9.4: Variability

Let's first compute the *variance*, which is the average squared difference between each value and the mean. Let's do this with our cleaned-up version of the height data, but instead of working with the entire dataset, let's take a random sample of 150 individuals:

```
height_sample <- NHANES %>%  
  drop_na(Height) %>%  
  sample_n(150) %>%  
  pull(Height)
```

First we need to obtain the sum of squared errors from the mean. In R, we can square a vector using `**2` :

```
SSE <- sum((height_sample - mean(height_sample))**2)  
SSE
```

```
## [1] 63419
```

Then we divide by  $N - 1$  to get the estimated variance:

```
var_est <- SSE/(length(height_sample) - 1)  
var_est
```

```
## [1] 426
```

We can compare this to the built-in `var()` function:

```
var(height_sample)
```

```
## [1] 426
```

We can get the *standard deviation* by simply taking the square root of the variance:

```
sqrt(var_est)
```

```
## [1] 21
```

Which is the same value obtained using the built-in `sd()` function:

```
sd(height_sample)
```

```
## [1] 21
```

This page titled [9.4: Variability](#) 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.