13.2) The SD and the Normal Curve

Vitor Kamada

December 2019

Reference

Tables, Graphics, and Figures from

Computational and Inferential Thinking: The Foundations of Data Science

Adhikari & DeNero (2019): Ch 14.3 The SD and the Normal Curve

https://www.inferentialthinking.com/

The Distribution of Heights of Mothers

```
from datascience import *
path_data = 'https://github.com/data-8/textbook/raw/gh-pages/data/'
baby = Table.read_table(path_data + 'baby.csv')
```

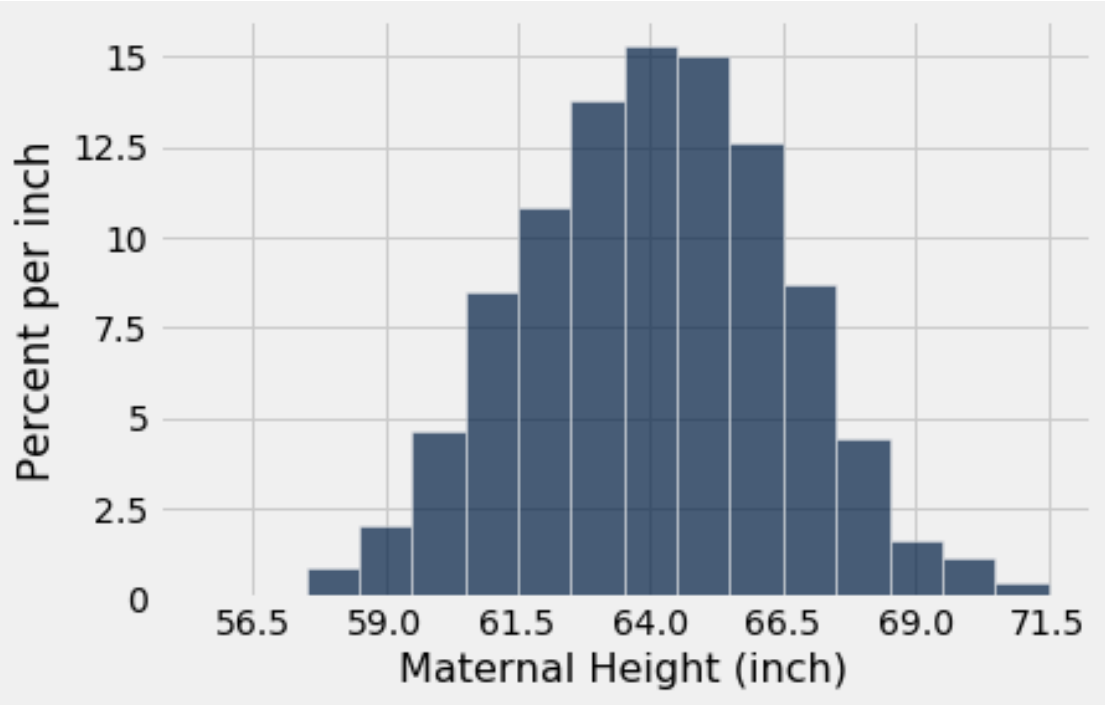
```
import numpy as np
heights = baby.column('Maternal Height')
mean_height = np.round(np.mean(heights), 1)
```

64.0

```
sd_height = np.round(np.std(heights), 1)
```

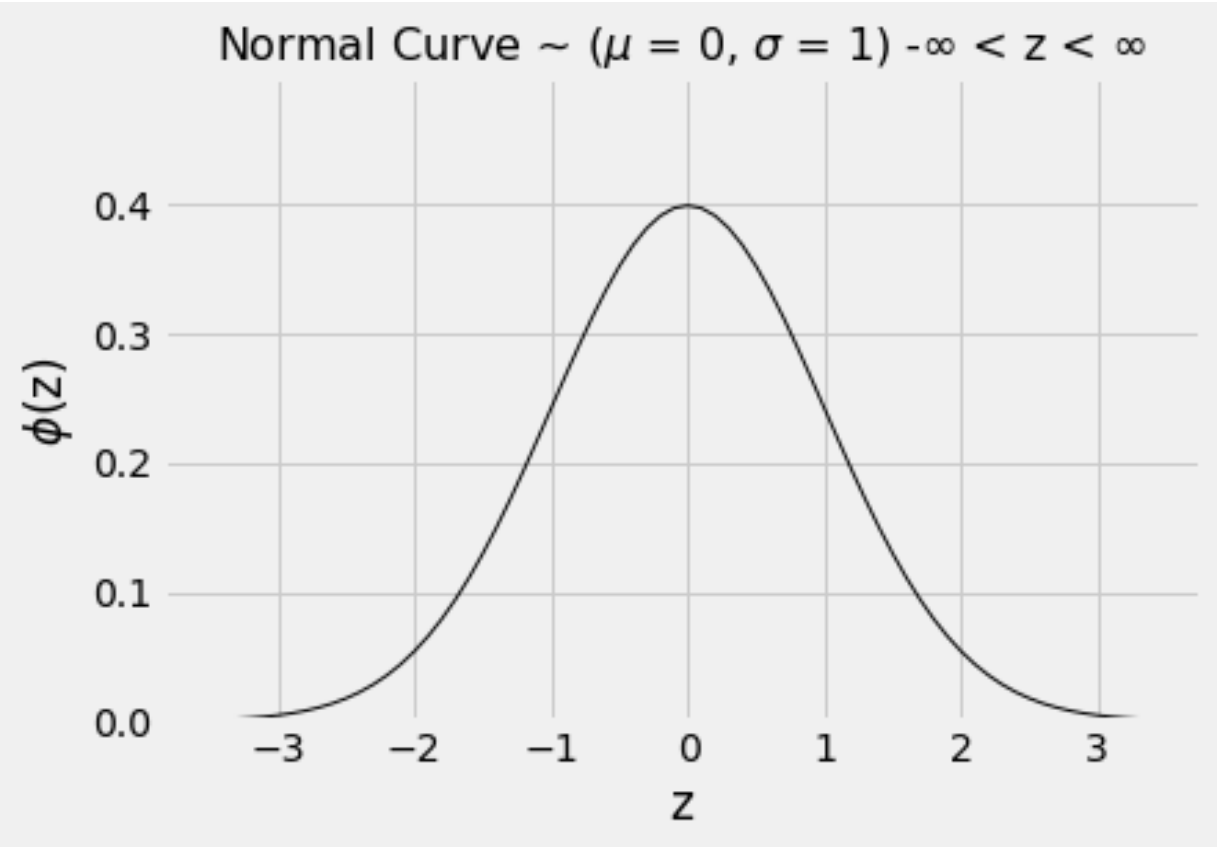
2.5

```
%matplotlib inline
import matplotlib.pyplot as plots
plots.style.use('fivethirtyeight')
baby.hist('Maternal Height', bins=np.arange(55.5, 72.5, 1),
          unit='inch')
positions = np.arange(-3, 3.1, 1)*sd_height + mean_height
plots.xticks(positions);
```

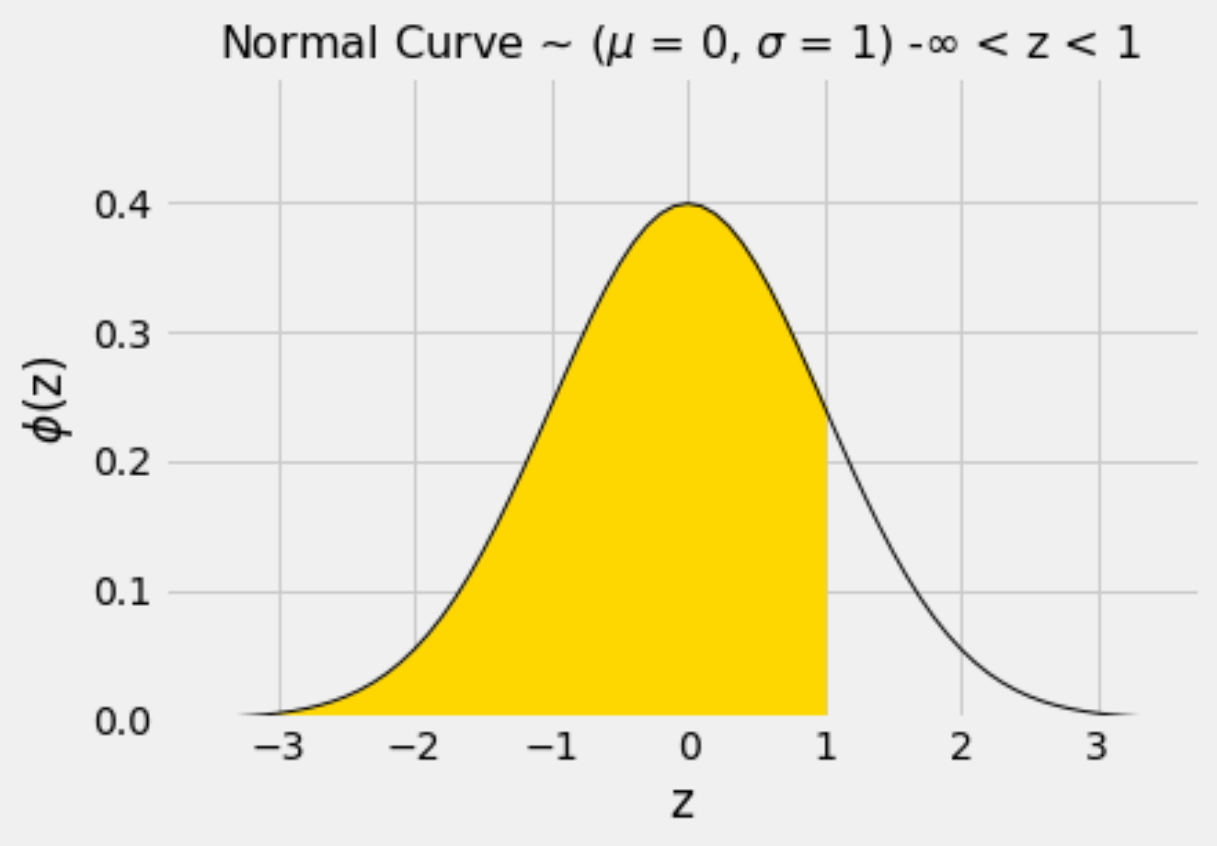


plot_normal_cdf()

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, \quad -\infty < z < \infty$$



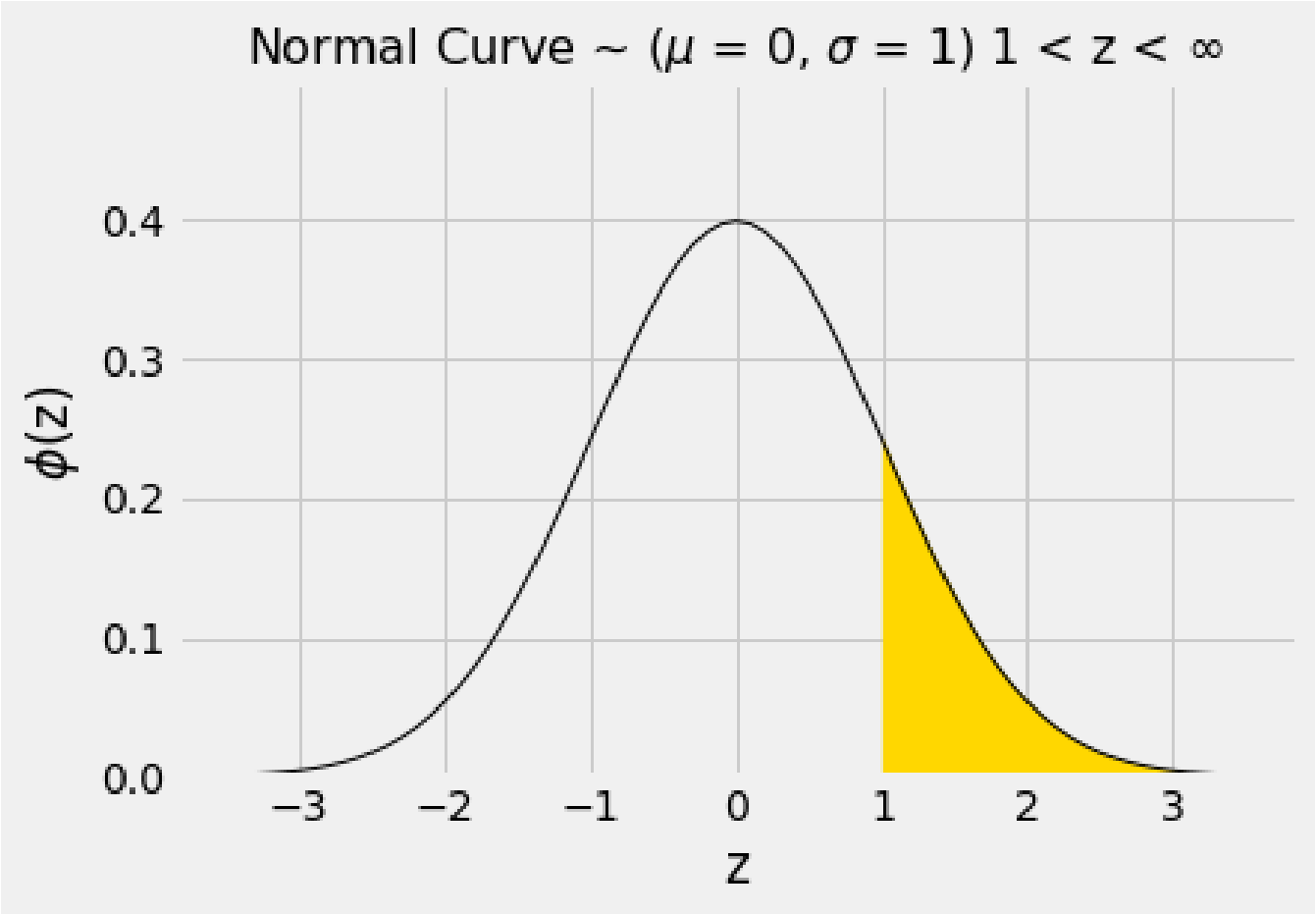
plot_normal_cdf(1)




```
from scipy import stats
stats.norm.cdf(1)
```

```
0.8413447460685429
```

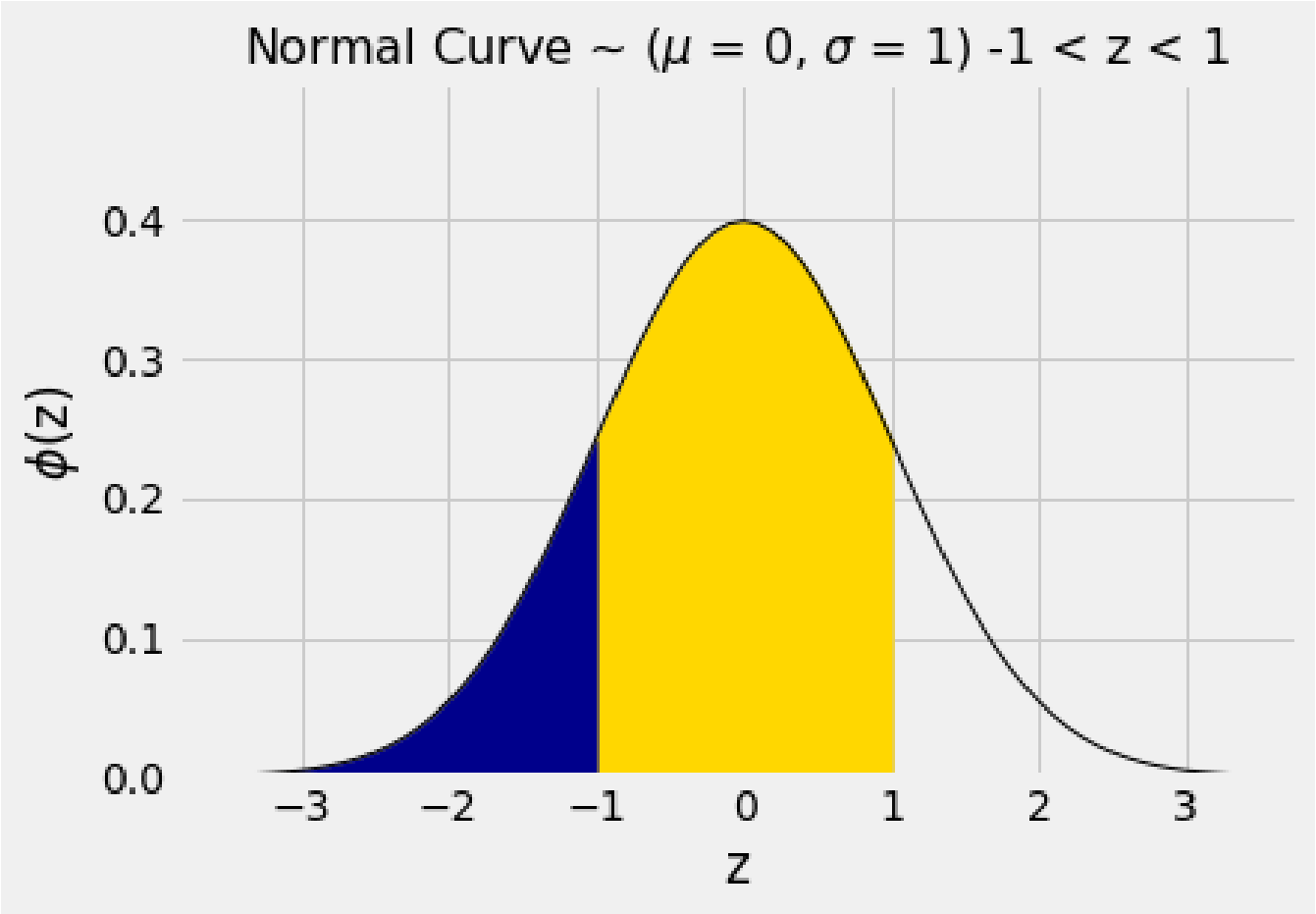
```
plot_normal_cdf(lbound=1)
```




```
1 - stats.norm.cdf(1)
```

```
0.15865525393145707
```

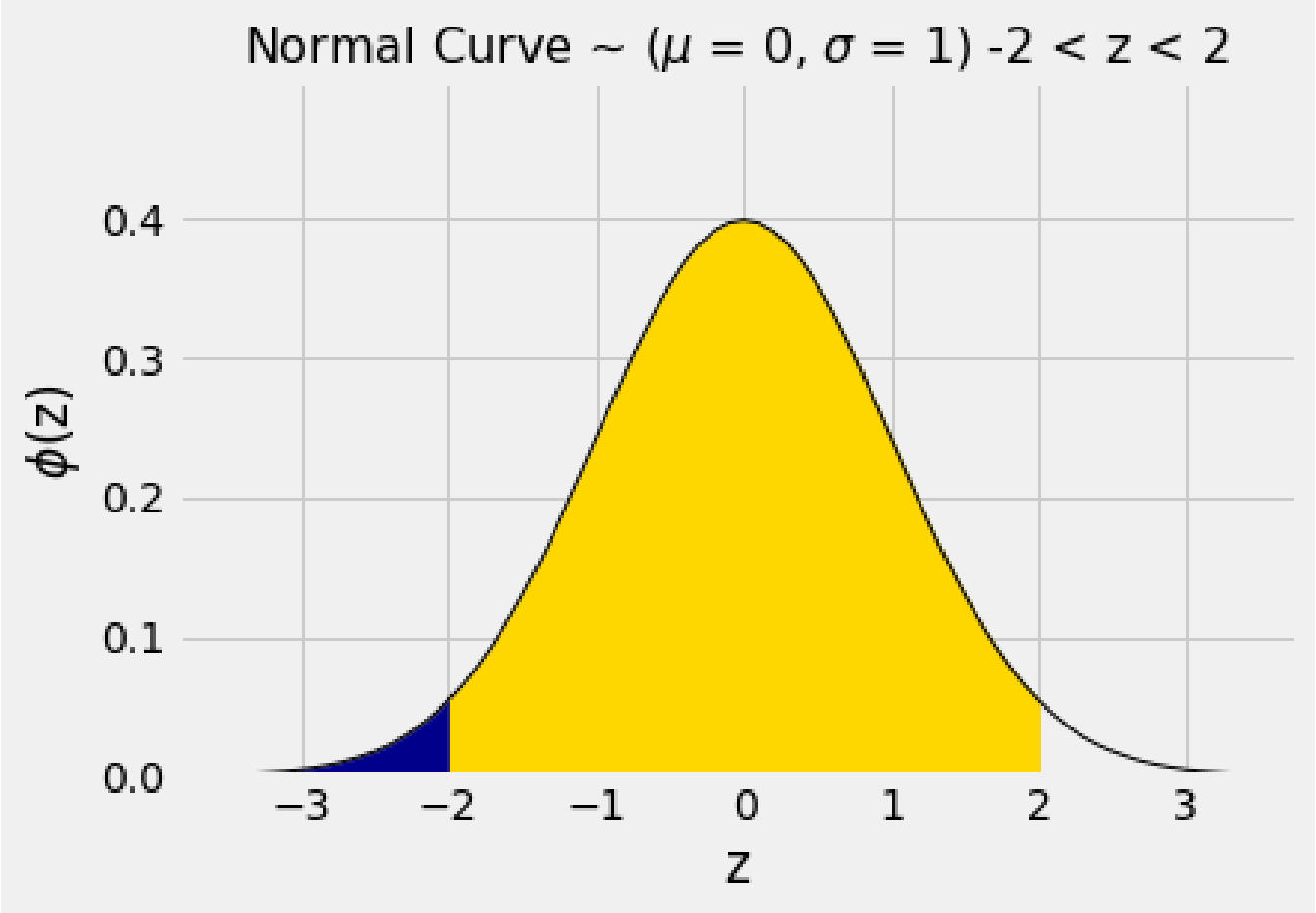
```
plot_normal_cdf(1, lbound=-1)
```




```
stats.norm.cdf(1) - stats.norm.cdf(-1)
```

```
0.6826894921370859
```

```
plot_normal_cdf(2, lbound=-2)
```




```
stats.norm.cdf(2) - stats.norm.cdf(-2)
```

```
0.9544997361036416
```

All Distributions vs Normal Distributions

Percent in Range	All Distributions: Bound	Normal Distribution: Approximation
average ± 1 SD	at least 0%	about 68%
average ± 2 SDs	at least 75%	about 95%
average ± 3 SDs	at least 88.88%	about 99.73%