

# PDF and CDF



QL 1.1

By the end of today, you should be able to...

1. Compute probability density functions and cumulative density functions
2. Use the `scipy.stats` package to compute the Survival Value or CDF Value for a known distribution

[Click here for warm up activity.](#)

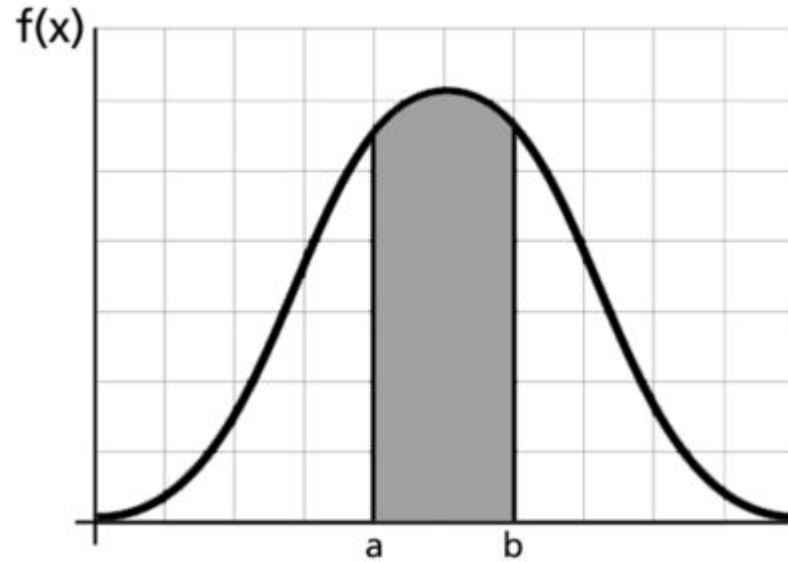
# Warm up

7 mins

First let's review the difference between discrete and continuous random variables:

- Discrete: takes on a finite or countable number of values.
- Continuous: takes on an infinite number of values
  - What's the probability of having  $x$  to be exactly 3.0000 if  $0.0 < x < 10.0$  ?
  - what's the probability that New York City gets **exactly** 4 inches of snow on December 17th?

- PDF is used to specify the probability of the random variable falling within a particular range of values, as opposed to taking on any one value.





# Plotting histogram

7 min

Using 'seaborn' library, plot the  
histogram of titanic passengers age.

**What percent of passengers are younger than 40?**

1 min

```
how_many_younger_40 = df[???]

pr_below_40 =
len(How_many_younger_40)/len(df[ 'Age' ].dropna(
))

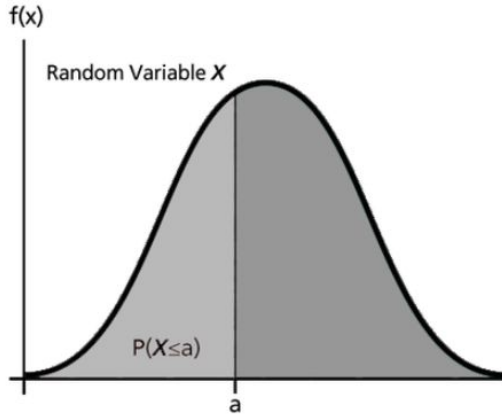
print(pr_below_40)
```



In the above example, we could not easily obtain the percentage from a PDF, although it is possible.

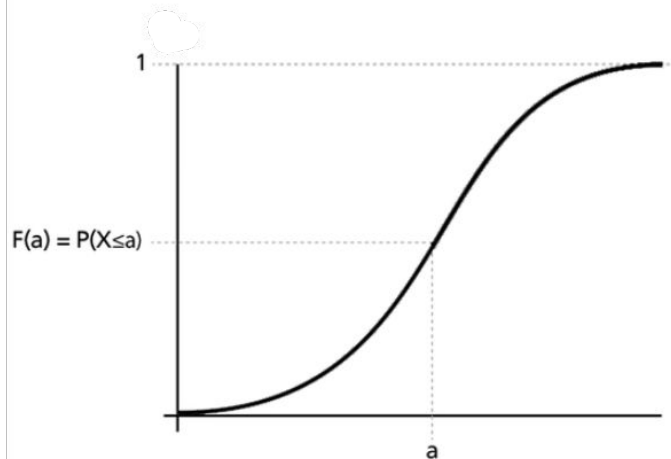
This is much easier if we use a CDF.

- A CDF calculates the **probability that a random variable is less than a threshold value**



$$f(x) = \frac{d(F(x))}{dx}$$

$$\int_{-\infty}^{+\infty} f(x) dx = 1$$



## Probability Density Function

$$F(x) = P(a \leq x \leq b) = \int_a^b f(x) dx \geq 0$$

- Let's hop to [this jupyter notebook](#).