Probability, 2

Frank Edwards

The expectation of a random variable

The expectation of a random variable $E(X)$ is the mean of a random variable.

Be careful not to confuse $E(X)$ and $\overline{x}$.

The expectation of a random variable

The expectation of a random variable $E(X)$ is the mean of a random variable.

Be careful not to confuse $E(X)$ and $\bar{x}$.

For a discrete variable, the expectation is the sum of all values of x weighted by their probability, given by the PDF $f(x)$.

$$E(X) = \sum_x x \times f(x)$$

Variance and standard deviation of a random variable

The standard deviation sd is

$$sd = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

And the sample variance is sd^2

Variance and standard deviation of a random variable

The standard deviation sd is

$$sd = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

And the sample variance is sd^2

Variance defined

For a random variable X, the variance is defined via the expectation instead of sample mean

$$V(X) = E[\{X - E(X)\}^2]$$

Variance defined

For a random variable X, the variance is defined via the expectation instead of sample mean

$$V(X) = E[\{X - E(X)\}^2]$$

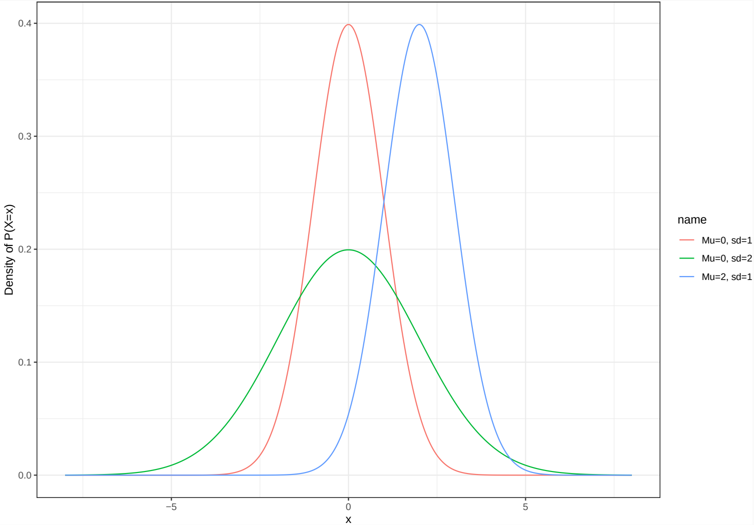
Note the similarities in the two equations

$$V(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

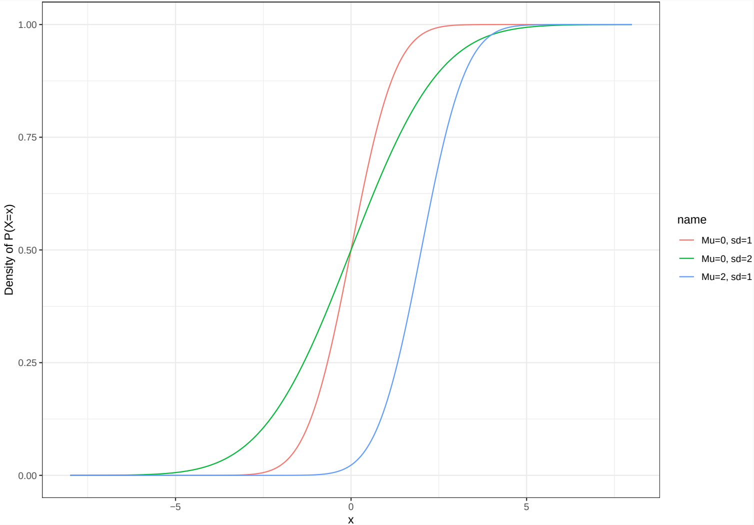
The Normal (Gaussian) distribution is continuous, and takes on values from $[-\infty, \infty]$. It has two parameters, the mean μ and standard deviation σ (or variance σ^2).

- μ determines the location of the distribution
- σ determines the spread of the distribution

The Normal PDF



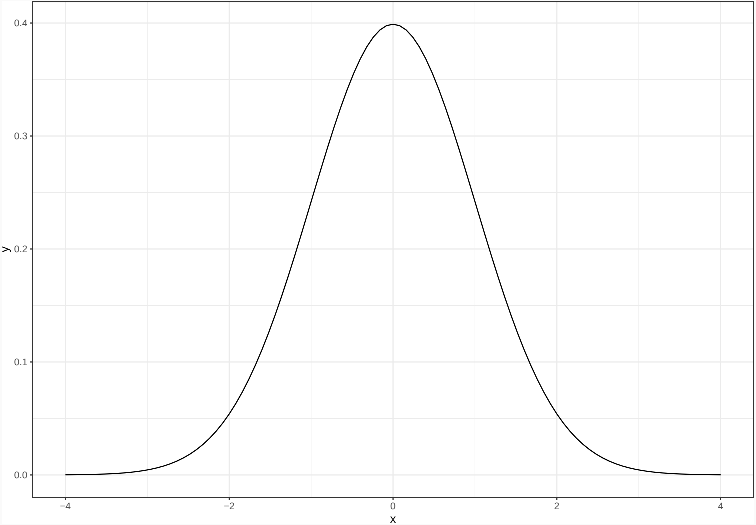
The Normal CDF



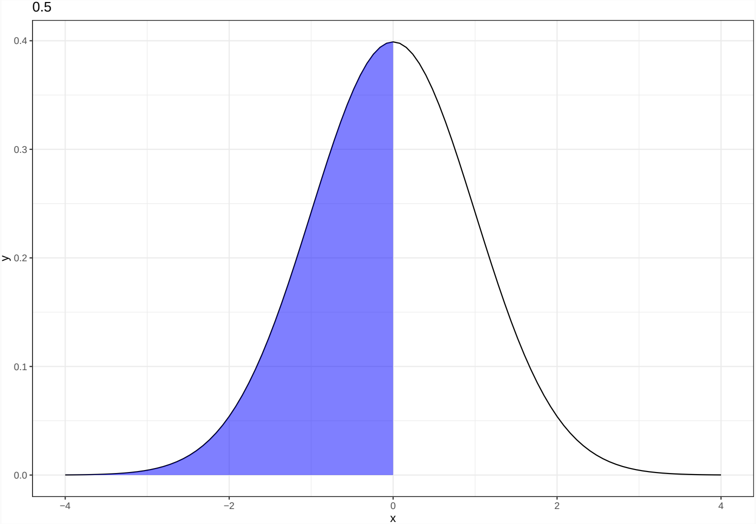
Special features of Normal distributions:

- The sum of many random variables from other distributions are often Normal
- For $X \sim N(\mu, \sigma^2)$, $Z = X + c$ is also Normal: $Z \sim (\mu + c, \sigma^2)$
- $Z = cX$ is distributed $Z \sim N(c\mu, (c\sigma)^2)$
- Z-scores of a Normal random variable are $N(0, 1)$

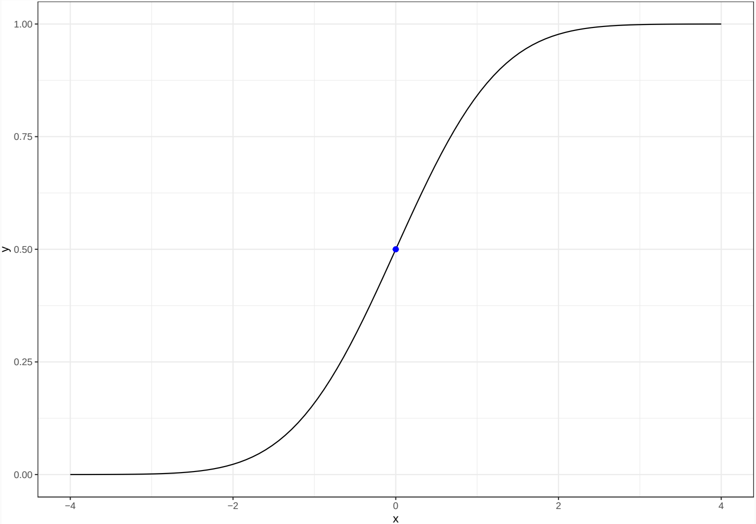
Area under the curve: interpreting the PDF and CDF



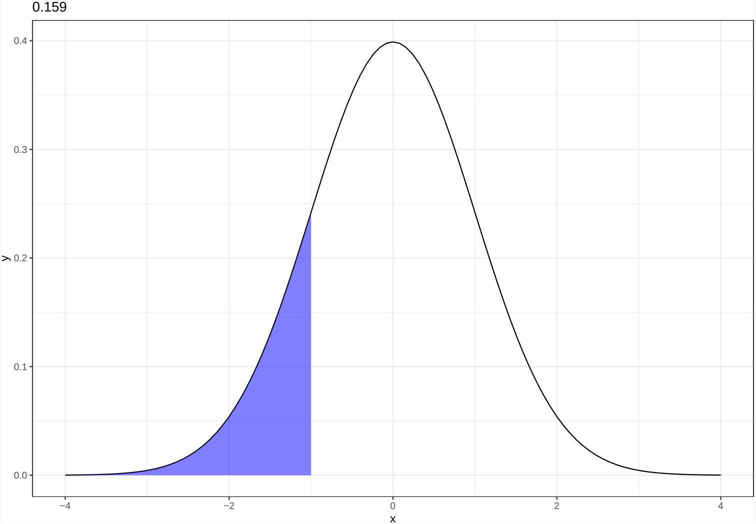
Area under the curve: interpreting the PDF and CDF



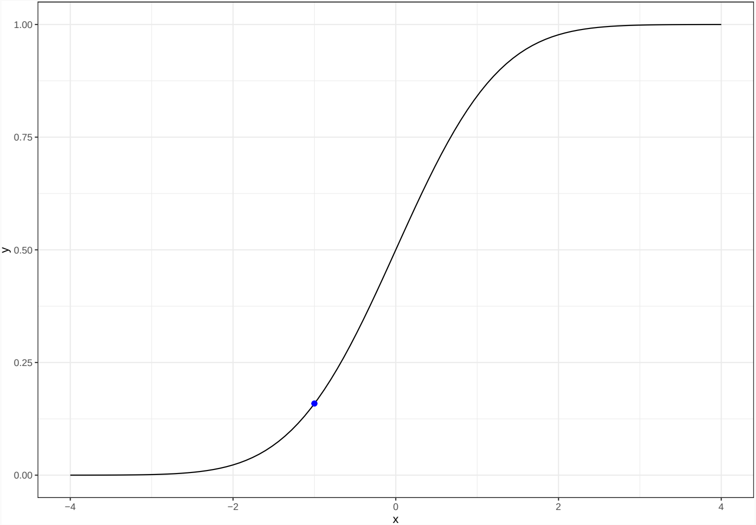
Area under the curve: interpreting the PDF and CDF



Area under the curve: interpreting the PDF and CDF



Area under the curve: interpreting the PDF and CDF



To obtain a z-score, we subtract the mean and divide by the standard deviation:

$$\text{z-score} = \frac{X - \mu}{\sigma}$$

For a Normal variable, z-scores are distributed $z \sim N(0, 1)$

Z-scores and area under the curve

To obtain a z-score, we subtract the mean and divide by the standard deviation:

$$\text{z-score} = \frac{X - \mu}{\sigma}$$

For a Normal variable, z-scores are distributed $z \sim N(0, 1)$

What does a z-score of 0 indicate?

To obtain a z-score, we subtract the mean and divide by the standard deviation:

$$\text{z-score} = \frac{X - \mu}{\sigma}$$

For a Normal variable, z-scores are distributed $z \sim N(0, 1)$

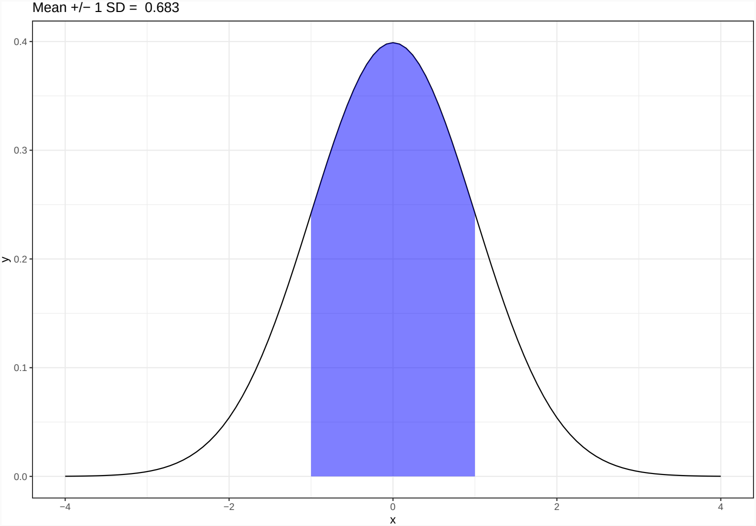
What does a z-score of 0 indicate? -1?

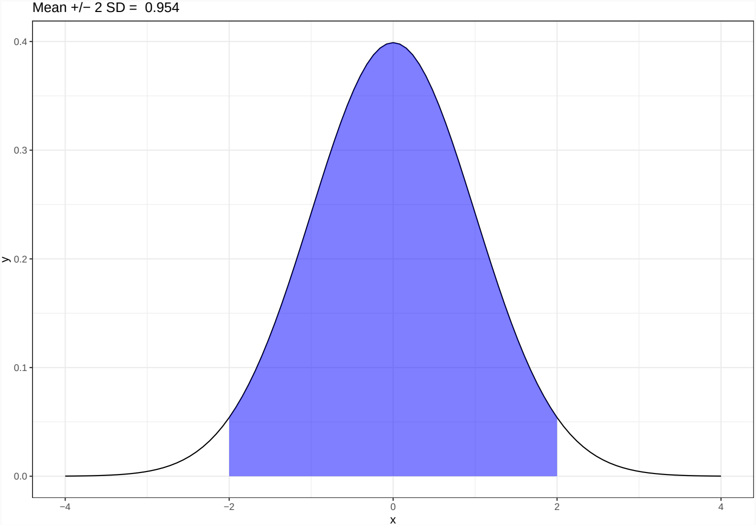
To obtain a z-score, we subtract the mean and divide by the standard deviation:

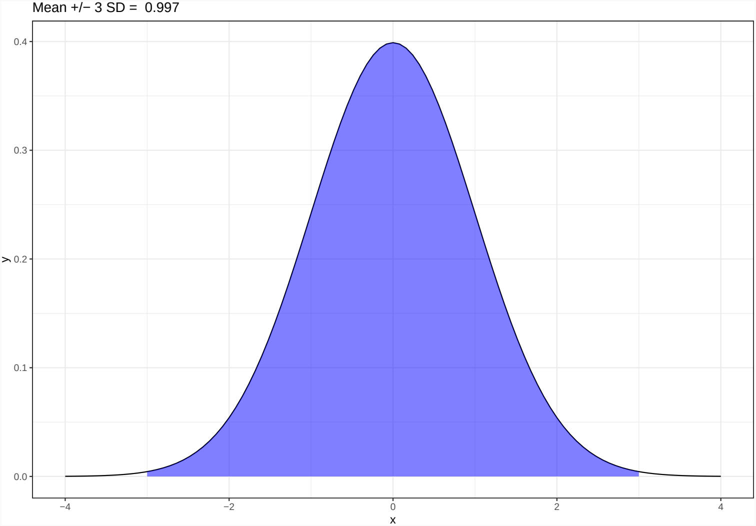
$$\text{z-score} = \frac{X - \mu}{\sigma}$$

For a Normal variable, z-scores are distributed $z \sim N(0, 1)$

What does a z-score of 0 indicate? -1? 2?


Mean +/− 1 SD = 0.683
0.4
0.3
0.2
0.1
0.0
y
−4
−2
0
2
4
x


Mean +/− 2 SD = 0.954
0.4
0.3
0.2
0.1
0.0
y
−4
−2
0
2
4
x



Mean +/− 3 SD = 0.997
0.4
0.3
0.2
0.1
0.0
y
−4
−2
0
2
4
x

Useful probability distribution functions

These will be useful for the homework!

```
### Normal(0,1) probability density function
dnorm(x = 0, mean = 0, sd = 1)
```

```
## [1] 0.3989423
```

```
### Normal(0,1) cumulative distribution function
pnorm(q = 0, mean = 0, sd = 1)
```

```
## [1] 0.5
```

```
### Random draw from a normal(0,1) distribution
rnorm(n = 1, mean = 0, sd = 1)
```

```
## [1] 1.290245
```

```
### CDF position for a given probability (quantile)
qnorm(p = 0.75, mean = 0, sd = 1)
```

```
## [1] 0.6744898
```

RMarkdown basics