CHAPTER 5.5

The Normal Distribution

A random variable X is **continuous** if for some function $f: \mathbb{R} \to \mathbb{R}$ and for any numbers a and b with $a \leq b$,

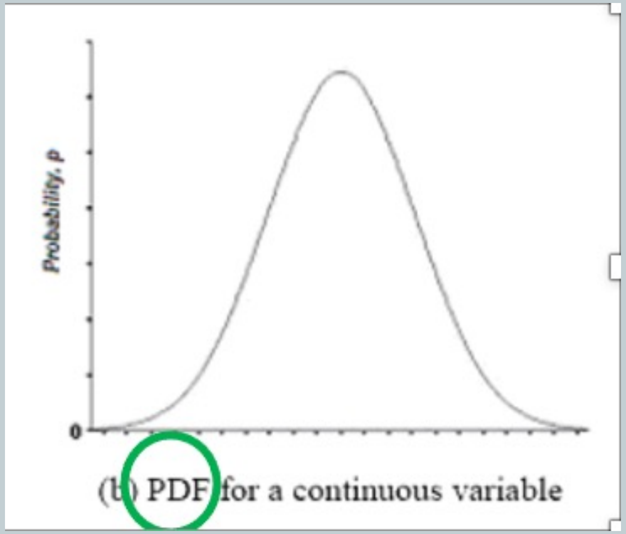
$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

The function f has to satisfy:

[1] $f(x) \geq 0$ for all x

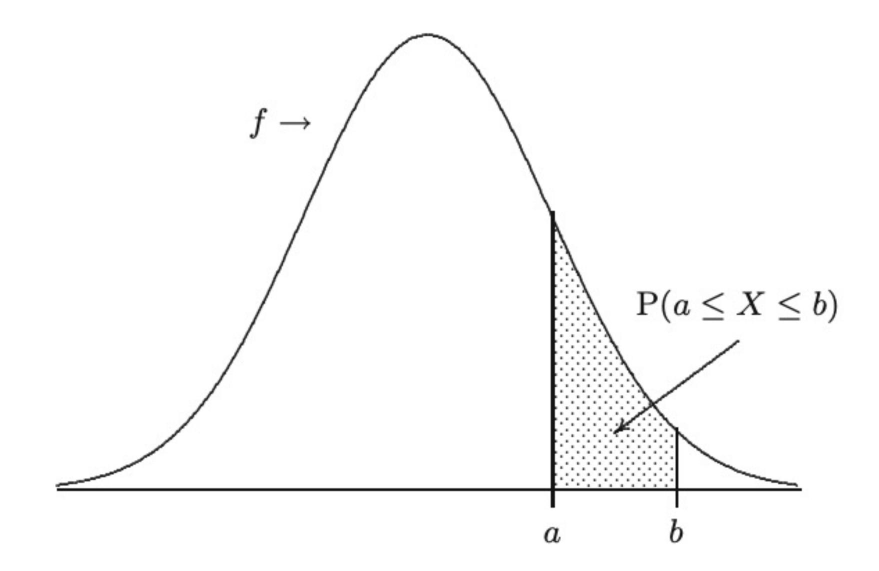
[2] $\int_{-\infty}^{\infty} f(x)dx = 1.$

We call f the probability density function, or probability density of X.



Recall, that in a continuous distribution we can express the probability that X lies in an interval $(a, b]$ directly in terms of F:

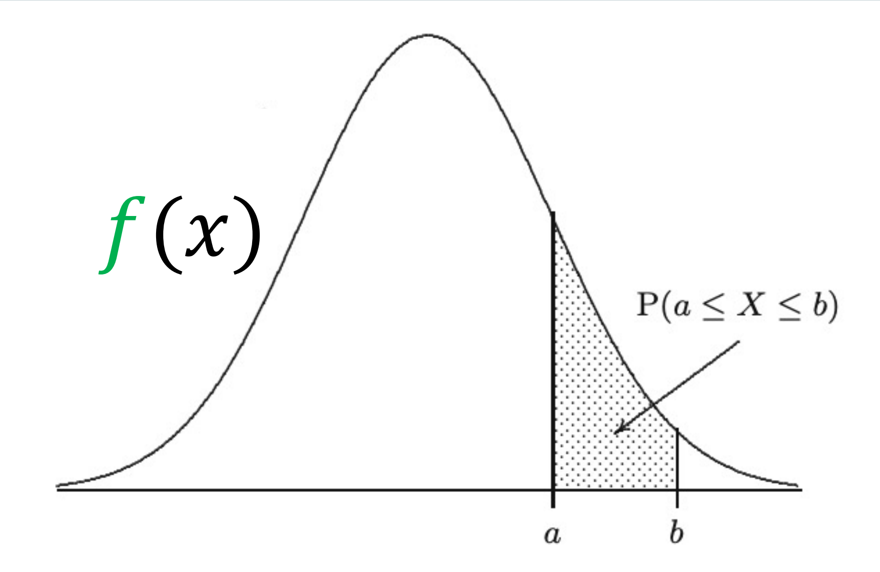
$$P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$$



The relation between the cumulative distribution function, F, and the probability density function, f, of a continuous random variable:

$$F(b) = \int_{-\infty}^{b} f(x)dx \qquad \text{and} \qquad f(x) = \frac{d}{dx}F(x)$$

The CDF is the integral of the PDF
The PDF is the derivative of the CDF

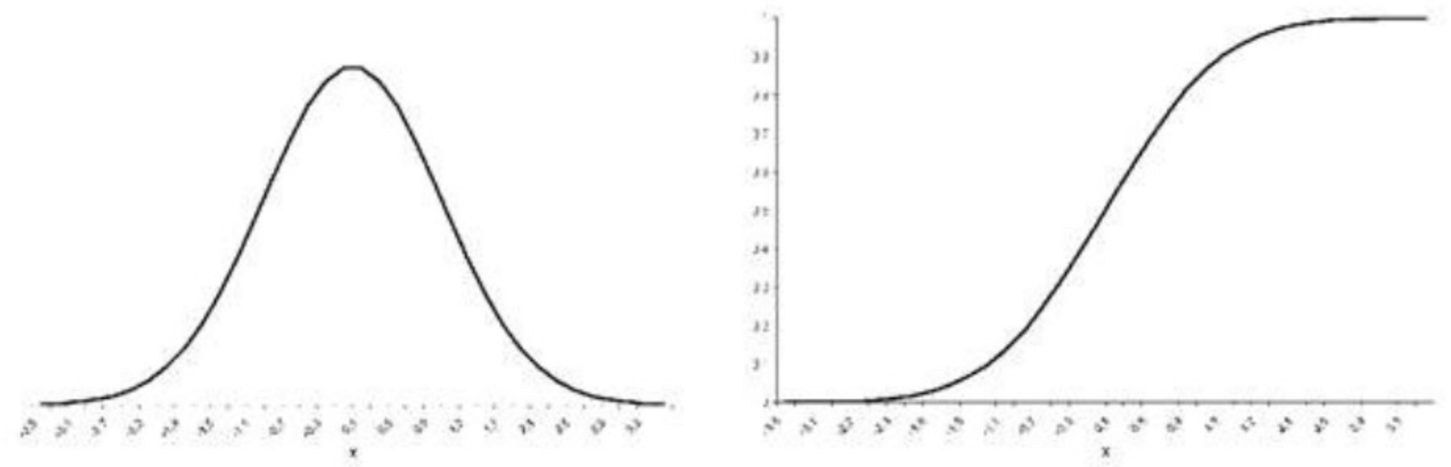


- A continuous random variable has a **normal distribution** with parameters μ and $\sigma^2 > 0$ if its probability density function f is given by:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

for $-\infty < x < \infty$.

This is denoted $N(\mu, \sigma^2)$.



a) Normal Probability Density Function

b) Normal Cumulative Distribution Function

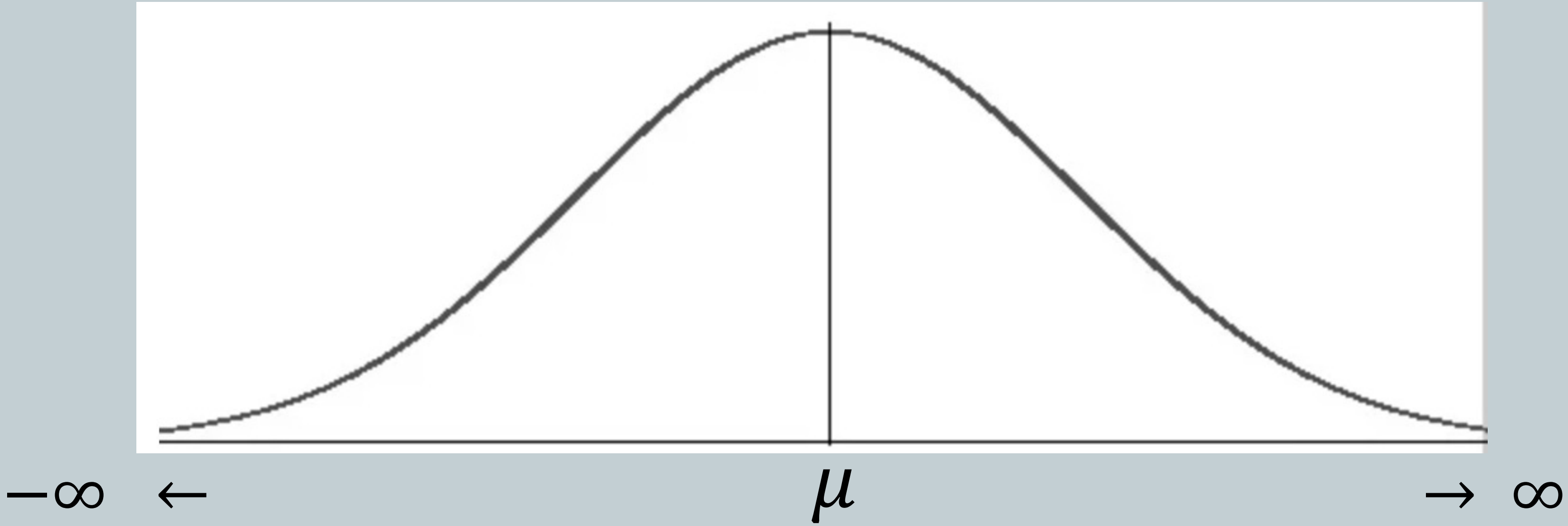
Suppose X has an $N(\mu, \sigma^2)$ distribution, then its CDF is given by:

$$F(a) = \int_{-\infty}^{a} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \text{ for } -\infty < a < \infty.$$

Unfortunately, f doesn't have an antiderivative!
Therefore, there is no explicit expression for F.

Luckily though, for any μ and σ^2, a $N(\mu, \sigma^2)$ distribution can easily be changed to a $N(0,1)$ distribution.

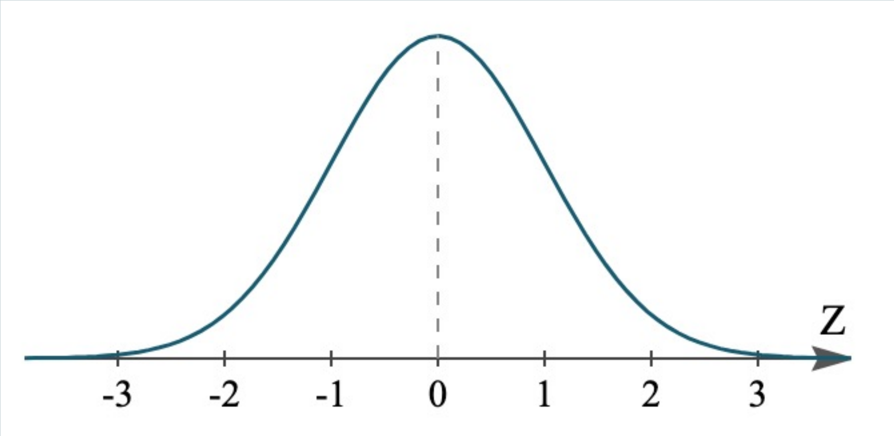
Area under the curve = 1



Consider a Normal Distribution with any given μ and σ: $N(\mu, \sigma^2)$

This distribution can be changed into a Standard Normal Distribution:

$z = \dfrac{x - \mu}{\sigma}$ or $x = \sigma z + \mu$.



$x - \mu$ centers the distribution at 0.

$\dfrac{x - \mu}{\sigma}$ scales to a standard deviation of 1.

The $N(\mu, \sigma^2)$ distribution has been changed to a $N(0, 1)$ distribution.

The $N(0,1)$ distribution is so important that it gets its own name and symbol!
The $N(0,1)$ distribution is called the **Standard Normal Distribution, $\phi(x)$**.

The Standard Normal Distribution, $\phi(x)$

Since $N(0,1)$ means $\mu = 0$ and $\sigma = 1$, $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ becomes:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

for $-\infty < x < \infty$.

Instead of PDF f, we have PDF ϕ.
Instead of CDF F, we have CDF Φ.

$$\Phi(a) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$$

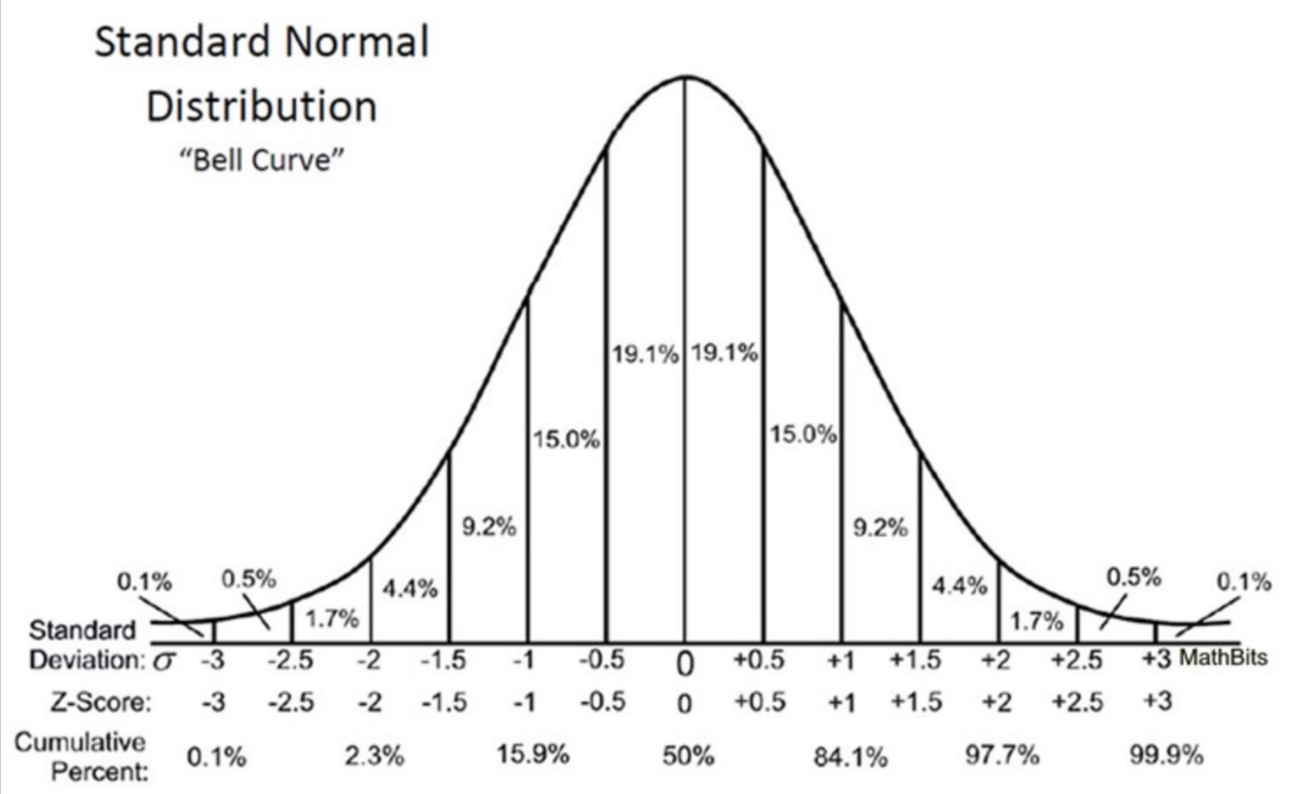
for $-\infty < a < \infty$.

Recall for continuous random variables we have notation:

X is the random variable

f is the probability density function

F is the cumulative distribution function



When we are speaking of the Standard Normal random variable, then:

Z is the random variable

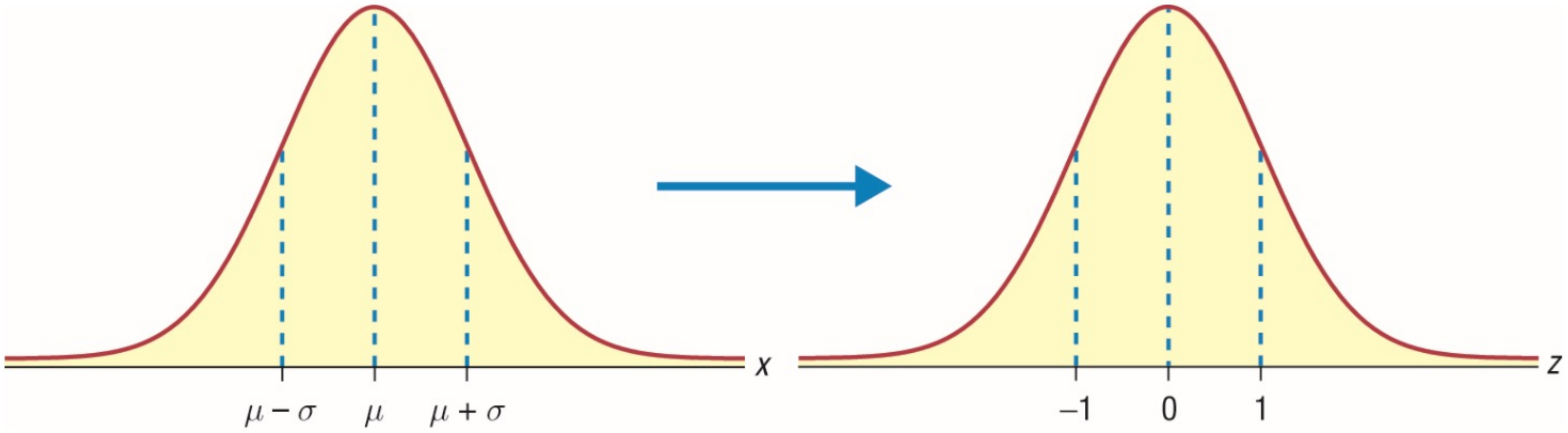
ϕ is the probability density function

Φ is the cumulative distribution function

```
import scipy.stats as stats
```

So $\Phi(z) = P(Z \leq z)$ is found with `stats.norm.cdf(z)`

$$N(\mu,\sigma) \xrightarrow{z\text{-}transformation} N(0,1)$$



Since the Standard Normal Distribution rarely (if ever) occurs in practice, we take non-standard normal distributions and standardize them.

Z implies we are talking about the standard normal distribution.

$P(Z > 1.25) = \ ?$

a	0	1	2	3	4	5	6	7	8	9
0.0	5000	4960	4920	4880	4840	4801	4761	4721	4681	4641
0.1	4602	4562	4522	4483	4443	4404	4364	4325	4286	4247
0.2	4207	4168	4129	4090	4052	4013	3974	3936	3897	3859
0.3	3821	3783	3745	3707	3669	3632	3594	3557	3520	3483
0.4	3446	3409	3372	3336	3300	3264	3228	3192	3156	3121
0.5	3085	3050	3015	2981	2946	2912	2877	2843	2810	2776
0.6	2743	2709	2676	2643	2611	2578	2546	2514	2483	2451
0.7	2420	2389	2358	2327	2296	2266	2236	2206	2177	2148
0.8	2119	2090	2061	2033	2005	1977	1949	1922	1894	1867
0.9	1841	1814	1788	1762	1736	1711	1685	1660	1635	1611
1.0	1587	1562	1539	1515	1492	1469	1446	1423	1401	1379
1.1	1357	1335	1314	1292	1271	1251	1230	1210	1190	1170
1.2	1151	1131	1112	1093	1075	1056	1038	1020	1003	0985
1.3	0968	0951	0934	0918	0901	0885	0869	0853	0838	0823

$P(Z \leq 1.25) = ?$

a	0	1	2	3	4	5	6	7	8	9
0.0	5000	4960	4920	4880	4840	4801	4761	4721	4681	4641
0.1	4602	4562	4522	4483	4443	4404	4364	4325	4286	4247
0.2	4207	4168	4129	4090	4052	4013	3974	3936	3897	3859
0.3	3821	3783	3745	3707	3669	3632	3594	3557	3520	3483
0.4	3446	3409	3372	3336	3300	3264	3228	3192	3156	3121
0.5	3085	3050	3015	2981	2946	2912	2877	2843	2810	2776
0.6	2743	2709	2676	2643	2611	2578	2546	2514	2483	2451
0.7	2420	2389	2358	2327	2296	2266	2236	2206	2177	2148
0.8	2119	2090	2061	2033	2005	1977	1949	1922	1894	1867
0.9	1841	1814	1788	1762	1736	1711	1685	1660	1635	1611
1.0	1587	1562	1539	1515	1492	1469	1446	1423	1401	1379
1.1	1357	1335	1314	1292	1271	1251	1230	1210	1190	1170
1.2	1151	1131	1112	1093	1075	1056	1038	1020	1003	0985
1.3	0968	0951	0934	0918	0901	0885	0869	0853	0838	0823

$$P(Z \leq -1.25) = \ ?$$

$P(-.38 \leq Z \leq 1.25) = \,?$

a	0	1	2	3	4	5	6	7	8	9
0.0	5000	4960	4920	4880	4840	4801	4761	4721	4681	4641
0.1	4602	4562	4522	4483	4443	4404	4364	4325	4286	4247
0.2	4207	4168	4129	4090	4052	4013	3974	3936	3897	3859
0.3	3821	3783	3745	3707	3669	3632	3594	3557	3520	3483
0.4	3446	3409	3372	3336	3300	3264	3228	3192	3156	3121
0.5	3085	3050	3015	2981	2946	2912	2877	2843	2810	2776
0.6	2743	2709	2676	2643	2611	2578	2546	2514	2483	2451
0.7	2420	2389	2358	2327	2296	2266	2236	2206	2177	2148
0.8	2119	2090	2061	2033	2005	1977	1949	1922	1894	1867
0.9	1841	1814	1788	1762	1736	1711	1685	1660	1635	1611
1.0	1587	1562	1539	1515	1492	1469	1446	1423	1401	1379
1.1	1357	1335	1314	1292	1271	1251	1230	1210	1190	1170
1.2	1151	1131	1112	1093	1075	1056	1038	1020	1003	0985
1.3	0968	0951	0934	0918	0901	0885	0869	0853	0838	0823

What is the 88^{th} percentile of a $N(0,1)$ distribution?

```
From scipy import stats
stats.norm.ppf(.88)
```

a	0	1	2	3	4	5	6	7	8	9
0.0	5000	4960	4920	4880	4840	4801	4761	4721	4681	4641
0.1	4602	4562	4522	4483	4443	4404	4364	4325	4286	4247
0.2	4207	4168	4129	4090	4052	4013	3974	3936	3897	3859
0.3	3821	3783	3745	3707	3669	3632	3594	3557	3520	3483
0.4	3446	3409	3372	3336	3300	3264	3228	3192	3156	3121
0.5	3085	3050	3015	2981	2946	2912	2877	2843	2810	2776
0.6	2743	2709	2676	2643	2611	2578	2546	2514	2483	2451
0.7	2420	2389	2358	2327	2296	2266	2236	2206	2177	2148
0.8	2119	2090	2061	2033	2005	1977	1949	1922	1894	1867
0.9	1841	1814	1788	1762	1736	1711	1685	1660	1635	1611
1.0	1587	1562	1539	1515	1492	1469	1446	1423	1401	1379
1.1	1357	1335	1314	1292	1271	1251	1230	1210	1190	1170
1.2	1151	1131	1112	1093	1075	1056	1038	1020	1003	0985
1.3	0968	0951	0934	0918	0901	0885	0869	0853	0838	0823

Suppose further that you took an exam that was $N(496{,}114)$ (SAT)

What would your exam score need to be in order to make the $88th$ percentile ?

(critical value)

remember: $z = \dfrac{x-\mu}{\sigma}$ or $x = \sigma z + \mu.$

We say z_α is the critical value of Z under the standard normal distribution that gives a certain tail area.

In particular, it is the Z value such that exactly α of the area under the curve lies to the right of z_α.

(note, different books have differently arranged tables)

What is the relationship between z_α and the CDF ?

What is the relationship between z_α and percentiles ?

$\Phi(z_\alpha) = 1 - \alpha$

z_α corresponds to the $100 \cdot (1 - \alpha)^{th}$ percentile

Suppose $X \sim N(5, 6)$.

What does $x = 17$ indicate, relative to the norm?

Suppose that on average, with proper diet and exercise, a person can lose 5 pounds per month, give or take 2 pounds.

Let X be the amount of weight lost by a person in a month.

Furthermore, suppose $X \sim N(5, 2)$.

Suppose a person lost ten pounds in a month. The z-score for $x = 10$ is _____ ?

This z-score indicates that $x = 10$ is ______ standard deviations to the ______ (right or left) of the mean ______ (What is the mean)

Suppose a person gained three pounds (a negative weight loss).
Then $z =$ ______. This z-score tells you that $x = -3$ is ______ standard deviations to the _______ (right or left) of the mean.

Suppose $1{,}664{,}479$ students take the SAT exam.
The distribution of scores in the verbal section of the SAT had a mean, $\mu = 496$ and a standard deviation of $\sigma = 114$.
Let X be an SAT exam verbal section score. Then we say $X \sim N(496, 114)$.
Find the z-score for $x_1 = 325$ and $x_2 = 366.21$.
Interpret your findings.

Note: $z_1 = -1.5$ and $\Phi(-1.5) = 0.1469$ 85^{th} or 15^{th} percentile ?

60-year-old males have a body mass index that is normally distributed, $\sim N(29, 6)$. What is the probability that a male aged 60 has a BMI exceeding 35 ?

a	0	1	2	3	4	5	6	7	8	9
0.0	5000	4960	4920	4880	4840	4801	4761	4721	4681	4641
0.1	4602	4562	4522	4483	4443	4404	4364	4325	4286	4247
0.2	4207	4168	4129	4090	4052	4013	3974	3936	3897	3859
0.3	3821	3783	3745	3707	3669	3632	3594	3557	3520	3483
0.4	3446	3409	3372	3336	3300	3264	3228	3192	3156	3121
0.5	3085	3050	3015	2981	2946	2912	2877	2843	2810	2776
0.6	2743	2709	2676	2643	2611	2578	2546	2514	2483	2451
0.7	2420	2389	2358	2327	2296	2266	2236	2206	2177	2148
0.8	2119	2090	2061	2033	2005	1977	1949	1922	1894	1867
0.9	1841	1814	1788	1762	1736	1711	1685	1660	1635	1611
1.0	1587	1562	1539	1515	1492	1469	1446	1423	1401	1379
1.1	1357	1335	1314	1292	1271	1251	1230	1210	1190	1170
1.2	1151	1131	1112	1093	1075	1056	1038	1020	1003	0985
1.3	0968	0951	0934	0918	0901	0885	0869	0853	0838	0823

These three questions are asking the exact same thing:

What is the probability that a male aged 60 has BMI between 30 and 35 ?

What proportion of men aged 60 have BMI between 30 and 35 ?

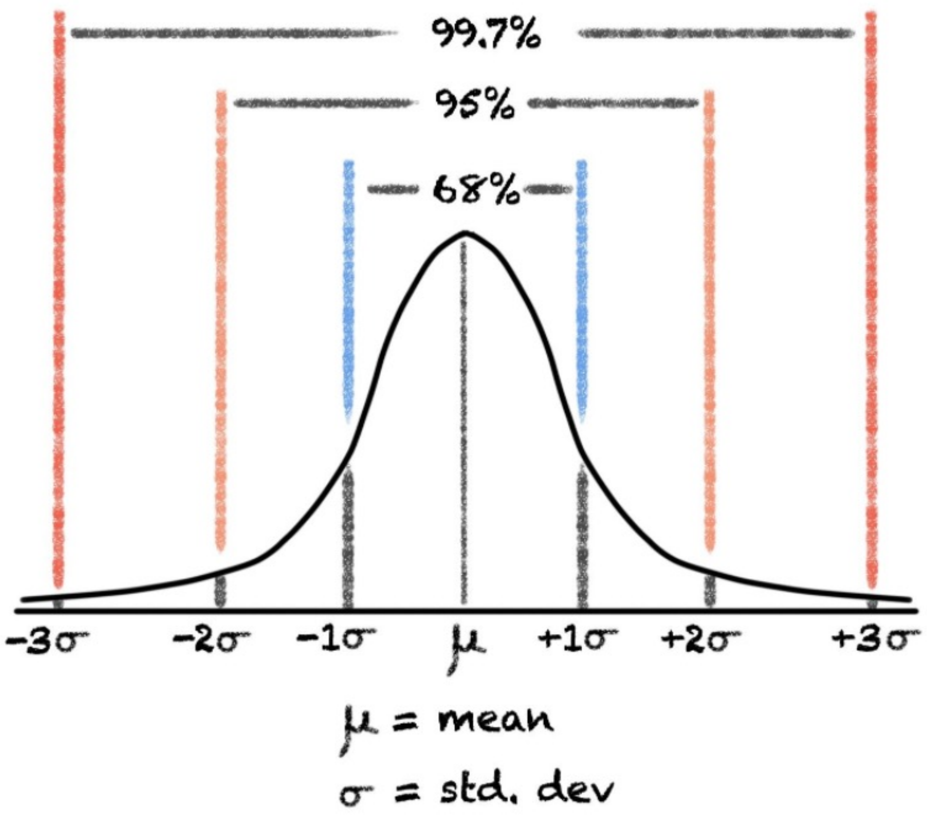
$P(30 < X < 35) = \ ?$

Recall the distribution from the previous slide: $\sim N(29, 6)$

Suppose Jamie is 5’10” tall and makes $47,000 per year.
How often would you expect to meet someone that makes 10X as much money?
How often would you expect to meet someone that is 10X as tall?

The probability depends on the mean AND the standard deviation.

Normal Distribution



$$\int_{-1}^{1} f_Z(z)\,dz \approx 68\%,$$

$$\int_{-2}^{2} f_Z(z)\,dz \approx 95\%,$$

$$\int_{-3}^{3} f_Z(z)\,dz \approx 99.7\%$$

There exists a handy 68-95-99 rule for the standard normal distribution

Suppose the scores on a college entrance exam have an approximate normal distribution with mean, $\mu = 52$ points and a standard deviation, $\sigma = 11$ points.

68% of scores lie between what two values? And what are their z-scores ?

95% of scores lie between what two values? And what are their z-scores ?

99.7% of scores lie between what two values? And what are their z-scores ?

Next, chapter 14 – The central limit theorem