The Normal Distribution

The Normal Distribution

Administrivia

- **Homework 3** due Friday. Good **Milestones**:
  - Problems 1, 2, and 3 done last week
  - Problems 4 and 5 done this week
- **Midterm** coming up **in-class** on Wednesday **October 18th**
  - Mix of Multiple Choice and Free-Response Questions
  - Allowed one 8.5 x 11in sheet of handwritten notes (no magnifying glasses)
  - Allowed a calculator that can’t connect to internet or store large large data

Previously on CSCI 3022

Def: A random variable X is **continuous** if for some function $f : \mathbb{R} \rightarrow \mathbb{R}$ 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 $f(x) \geq 0$ for all x and $\int_{-\infty}^{\infty} f(x)\ dx = 1$. We call f the probability density function of X.

Def: The cumulative distribution function of X is defined such that

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)\ dy$$

The Normal Distribution

The normal distribution (aka Gaussian distribution) is probably the most important distribution in probability and statistics.

Many populations have distributions well-approximated by a normal distribution

Examples: weight, height, and other physical characteristics, scores on tests, etc

The Normal Distribution

Def: a continuous random variable has a normal (or Gaussian) distribution with parameters μ and σ^2 if its probability density function is given by

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

If a random variable X is normally distributed we say $X \sim N(\mu, \sigma^2)$

Exploration! https://academo.org/demos/gaussian-distribution/

The Standard Normal Distribution

Def: a normal distribution with parameter values $\mu = 0,\ \sigma^2 = 1$ is called the **standard normal distribution**

Question: What is the pdf of the standard normal distribution

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

The Standard Normal Distribution

Def: a normal distribution with parameter values $\mu = 0,\ \sigma^2 = 1$ is called the **standard normal distribution**

A standard normal random variable is typically called Z

Recall: The normal distribution does not have a closed form cumulative distribution function

We use a special notation to denote the CDF of the **standard** normal distribution

$$\Phi(z) = P(Z \leq z)$$

The Standard Normal Distribution

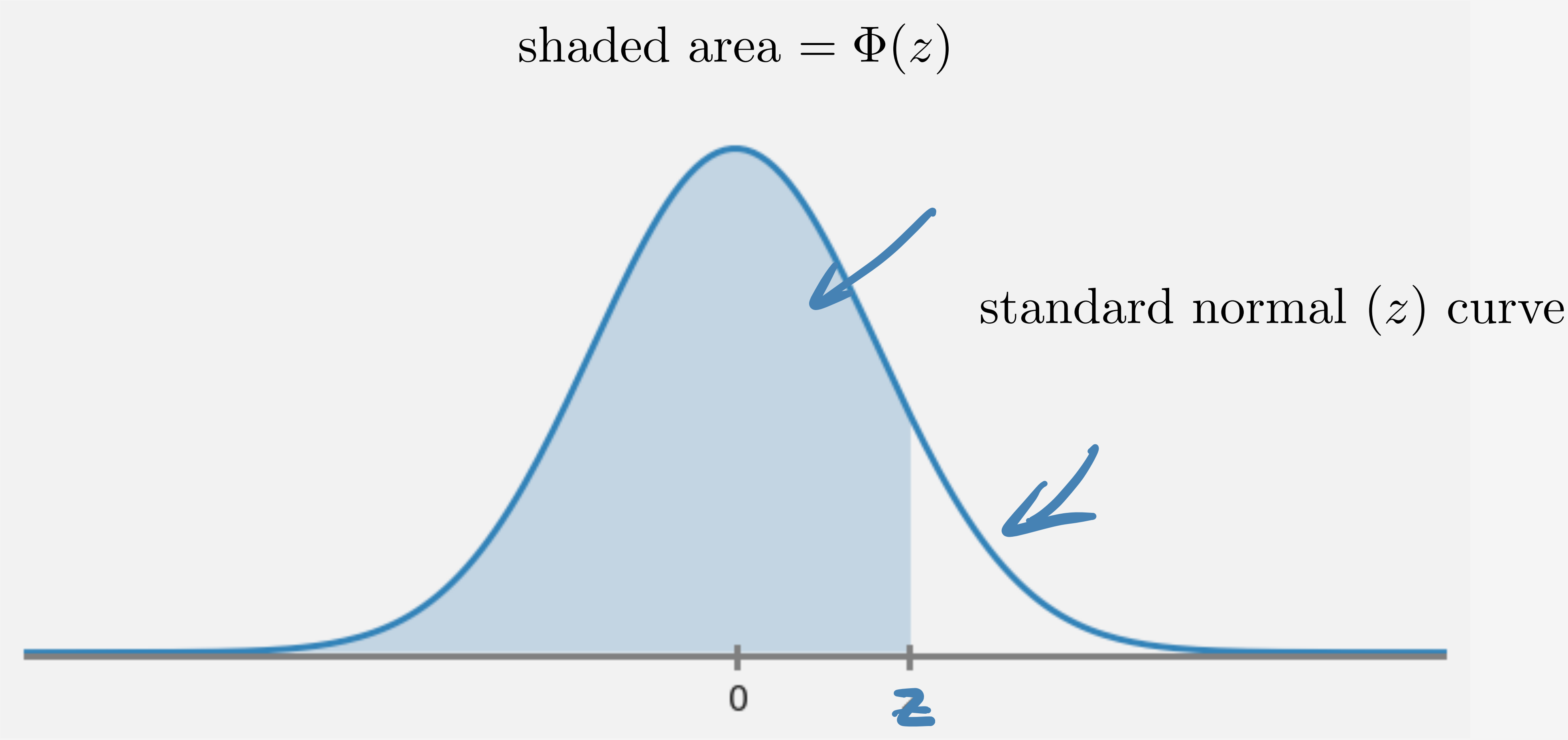
The standard normal distribution **rarely** occurs in real life

Instead, it's a reference distribution that allows us to learn about other (non-standard) normal distributions using a simple transformation

Recall: For computing probabilities, having a CDF is just as good (or better) as having a pdf

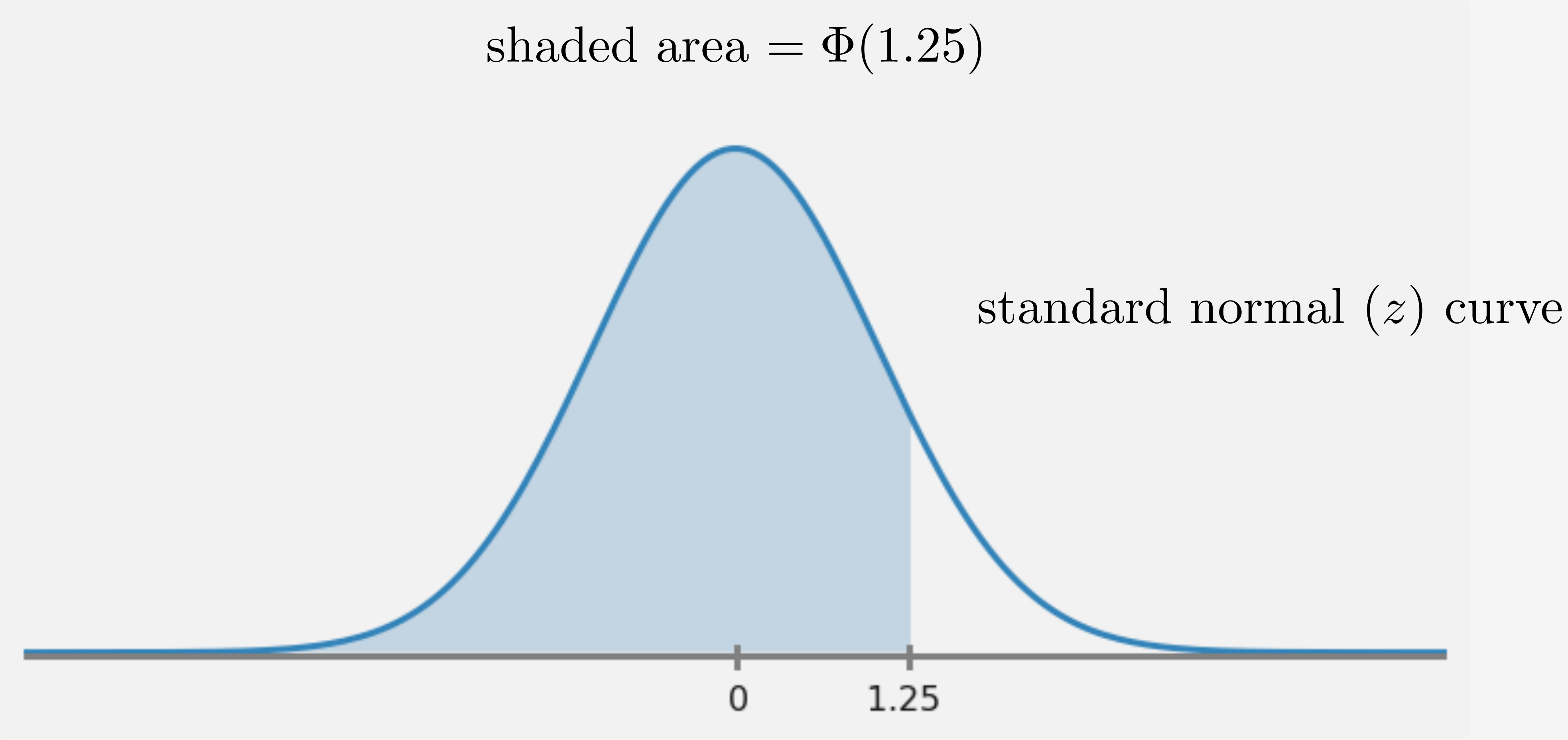
Back in the day, you looked up values of the standard normal CDF in **normal tables** in the back of probability books

The Standard Normal Distribution



The Standard Normal Distribution

Example: What is $P(Z \leq 1.25)$?



The Standard Normal Distribution

Example: What is $P(Z \geq 1.25)$?

$$P(Z \geq 1.25) = 1 - P(Z \leq 1.25) = 1 - \Phi(1.25)$$

Example: What is $P(Z \leq -1.25)$?

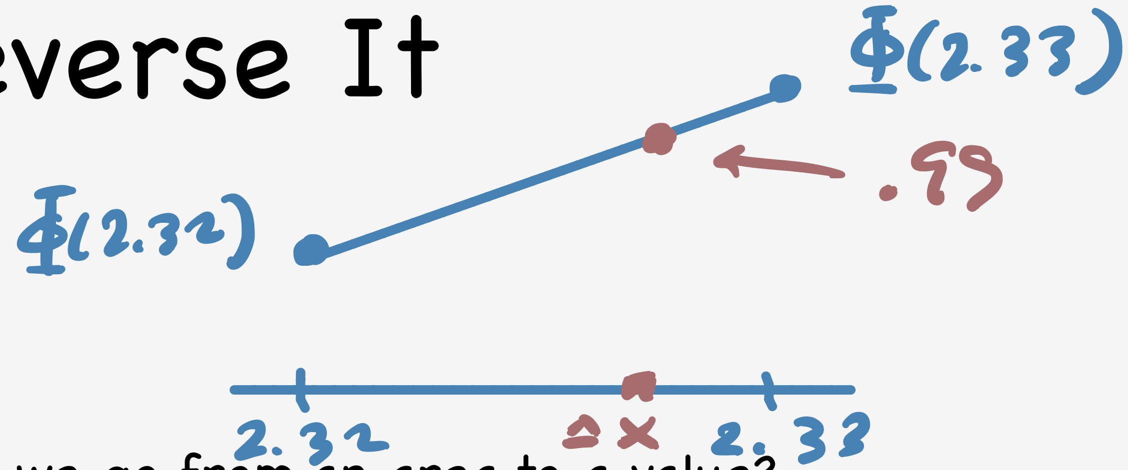
$$P(Z \leq -1.25) = 1 - \Phi(1.25) = \Phi(-1.25)$$

Example: How can we compute $P(-0.38 \leq Z \leq 1.25)$?

$$P(-0.38 \leq Z \leq 1.25) = \Phi(1.25) - (1 - \Phi(.38))$$

$$= \Phi(1.25) - \Phi(-.38)$$

Flip It and Reverse It



Example: What is the 99$^{\text{th}}$ percentile of $N(0,1)$?

Hmm: We have tables that tell us areas. How can we go from an area to a value?

This is the inverse **problem** to $P(Z \leq z) = 0.99$

How could you do this with a table?

FROM TABLE

$\Phi(2.32) = 0.98983$

$\Phi(2.33) = 0.99010$

$\text{SLOPE} = \dfrac{\Phi(2.33) - \Phi(2.32)}{.01}$

$\Rightarrow \quad 0.99 = \Phi(2.32) + \text{SLOPE} \cdot \Delta x \Rightarrow$

$\Delta x = \dfrac{0.99 - \Phi(2.32)}{\text{SLOPE}} = .0063$

$z = 2.3263$

Flip It and Reverse It

Example: What is the 99th percentile of $N(0, 1)$?

Hmm: We have tables that tell us areas. How can we go from an area to a value?

This is the inverse **problem** to $P(Z \le z) = 0.99$

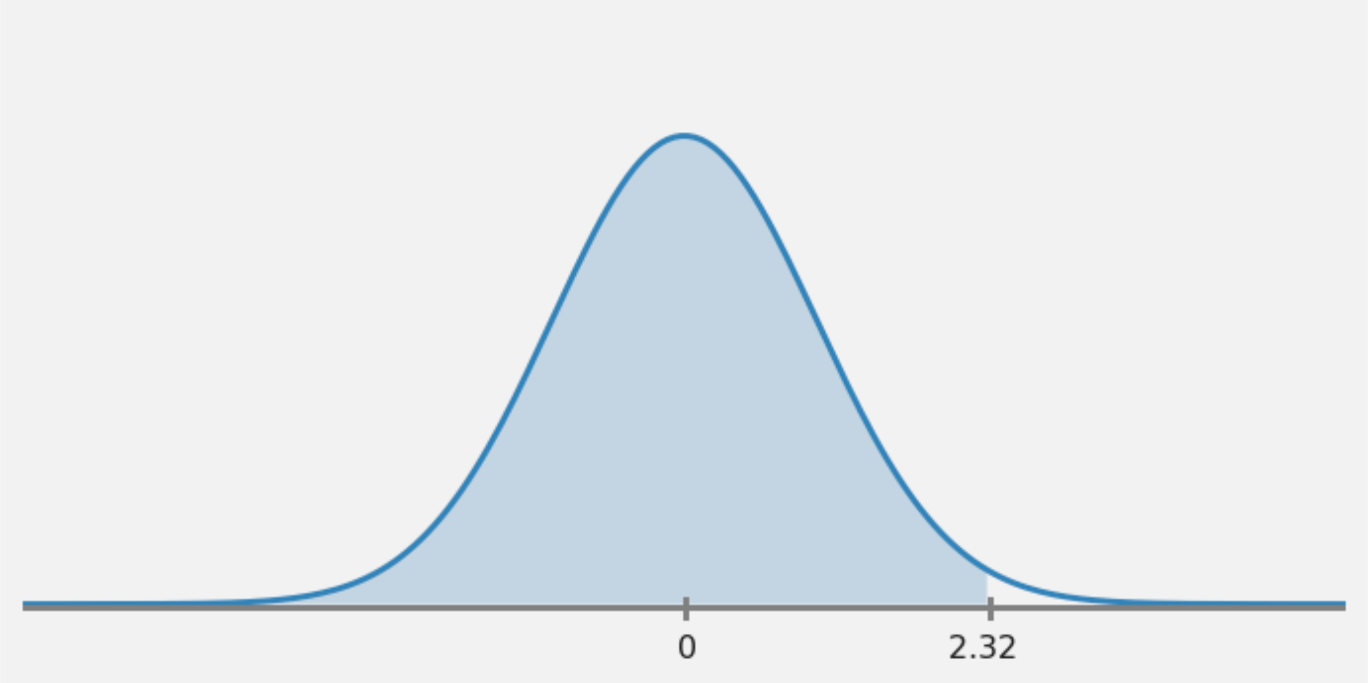
pdf, cdf

How could we do this with Python?

scipy.stats.norm.ppf(0.99)

Flip It and Reverse It

In [37]:
```
from scipy import stats
stats.norm.ppf(0.99)
```
Out[37]: 2.3263478740408408



stats.norm has lots of good functions related to normal distributions: pdf, cdf, ppf, etc

The Critical Value

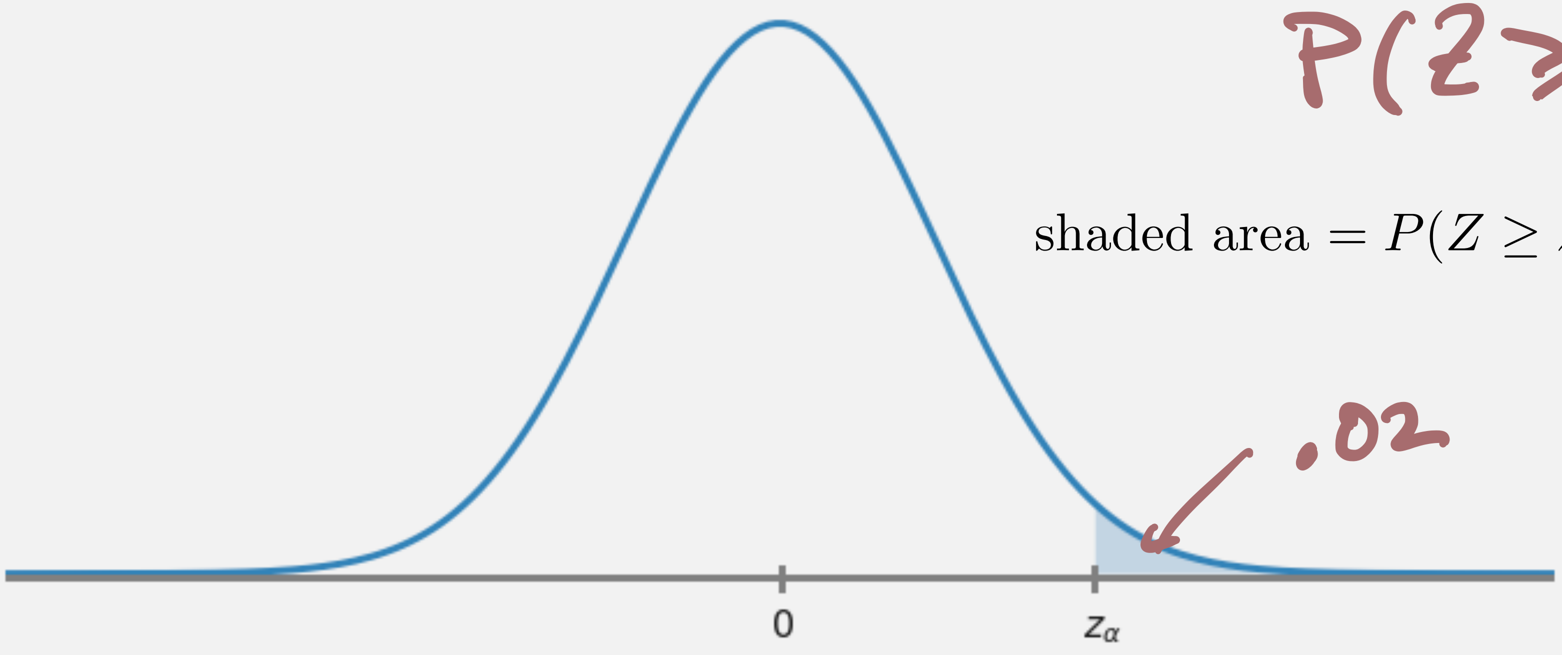
Notation: We say z_α is the value of Z under the standard normal distribution that gives a certain tail area. ~~In~~ particular, it is the z value such that exactly α of the area under the curve lies to the RIGHT of z_α

We call z_α the **critical value**

$z_{.02}$ s.t.

$P(Z \geq z_{.02}) = .02$

shaded area $= P(Z \geq z_\alpha) = \alpha$



The Critical Value

Notation: We say z_α is the 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

Question: What is the relationship between z_α and the cumulative distribution function?

$$P(Z \geq z_\alpha) = \alpha = 1 - P(Z \leq z_\alpha) = 1 - \Phi(z_\alpha)$$

Question: What is the relationship between z_α and percentiles?

z_α IS THE $100(1-\alpha)$TH PERCENTILE

Nonstandard Normal Distributions

Nonstandard normal distributions can be turned into standard normals really really easily

Proposition: If X is a normally distributed random variable with mean μ and standard deviation σ , then Z is a standard normal distribution if

$$Z = \frac{X-\mu}{\sigma}$$

CHECK: $E[Z] = \frac{1}{\sigma}E[X] - \frac{\mu}{\sigma} = \frac{1}{\sigma}\mu - \frac{\mu}{\sigma} = 0$

$$\mathrm{VAR}(Z) = \mathrm{VAR}\left(\frac{X}{\sigma} - \frac{\mu}{\sigma}\right) = \mathrm{VAR}\left(\frac{X}{\sigma}\right) = \frac{1}{\sigma^2}\mathrm{VAR}(X)$$

$$= \frac{1}{\sigma^2}\cdot\sigma^2 = 1$$

✓✓✓

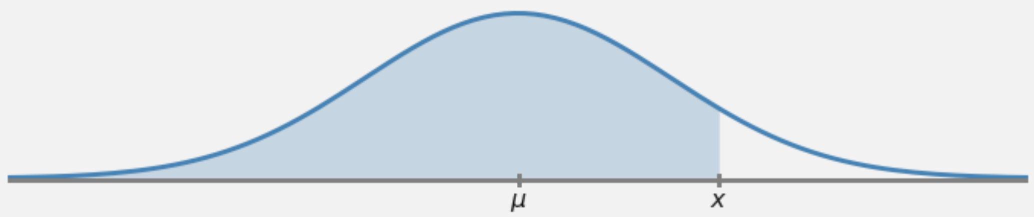
Nonstandard Normal Distributions

Nonstandard normal distributions can be turned into standard normals really really easily

Proposition: If X is a normally distributed random variable with mean μ and standard deviation σ , then Z is a standard normal distribution if

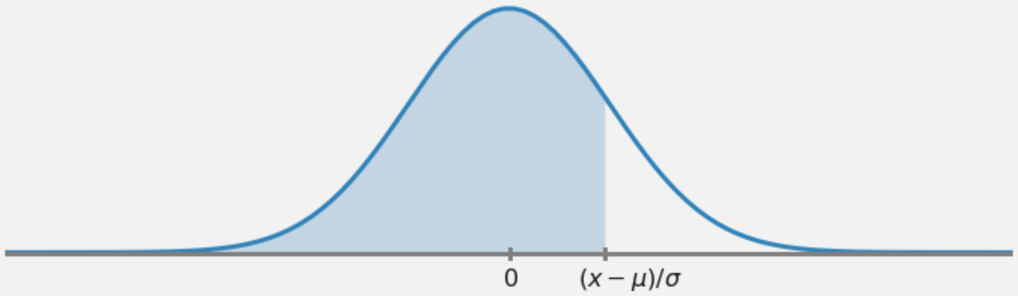
$$Z = \frac{X - \mu}{\sigma} \quad \text{or} \quad X = \sigma Z + \mu$$

$$X \sim N(\mu, \sigma^2)$$



$$\Leftrightarrow$$

$$Z \sim N(0, 1)$$



Brake Lights! (in Grandpa-Voice)

Example: The time it takes a driver to react to brake lights on a decelerating vehicle is critical to helping to avoid rear-end collisions

The article Fast-Rise Brake Lamp as a Collision Prevention Device (Ergonomics, 1993: 391-395) suggests that reaction time for an in-traffic response to a brake signal from standard brake lights can be modeled as a normal distribution having mean value 1.25 sec and standard deviation 0.46 sec.

Question: What is the probability that a reaction time is between 1.0 sec and 1.75 sec?

$$1.0 \rightarrow \frac{(1.0-1.25)}{0.46} \rightarrow -.54$$

$$1.75 \rightarrow \frac{(1.75-1.25)}{0.46} \rightarrow 1.08$$

$$\Phi(1.08) - \Phi(-.54)$$

$$\rightarrow P(1.0 \le X \le 1.75) = P(-.54 \le Z \le 1.08)$$

OK! Let's Go to Work!

Get in groups, get out laptop, and open the Lecture 12 In-Class Notebook

Let's:

- Get some more practice computing normal probabilities in Python
- Look at the way grading curves are often done
- See how we can sample from the standard normal using the Box-Muller method when we don't have stats libraries readily available