5 The Normal Distribution

(Ch 4.3)

The Normal Distribution

The <u>normal distribution</u> is probably the most important distribution in all of probability and statistics.

Many populations have distributions that can be fit very closely by an appropriate normal (or Gaussian, bell) curve.

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

The Normal Distribution

Definition

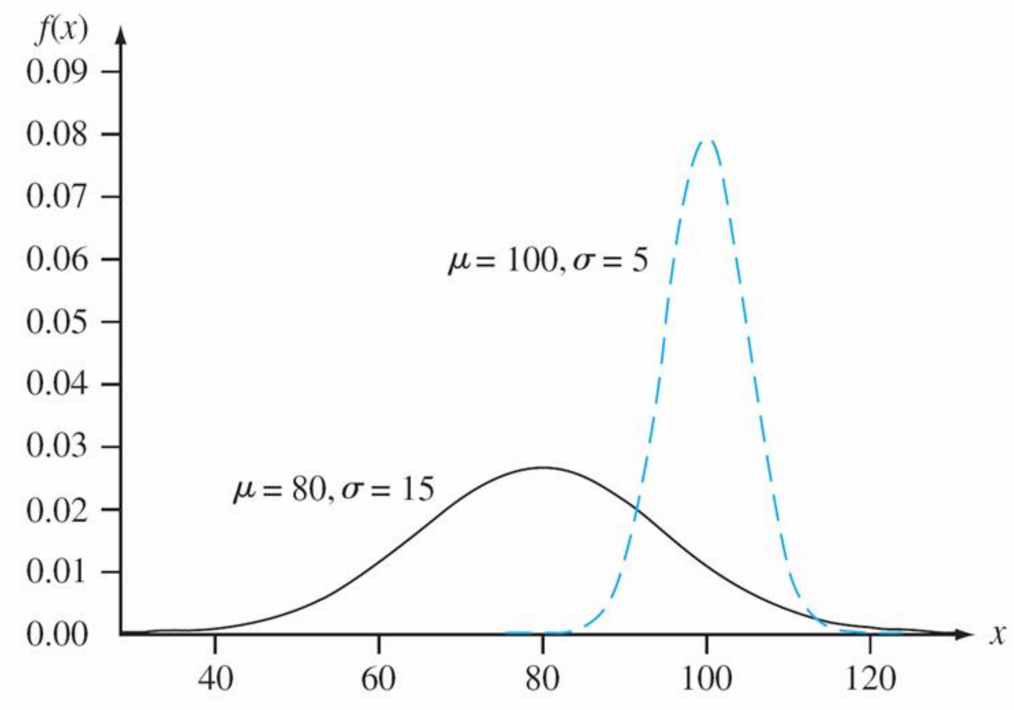
A continuous r.v. X is said to have a **normal distribution** with parameters μ and $\sigma > 0$ (or μ and σ^2), if the pdf of X is

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} \text{ where } -\infty < x < \infty$$

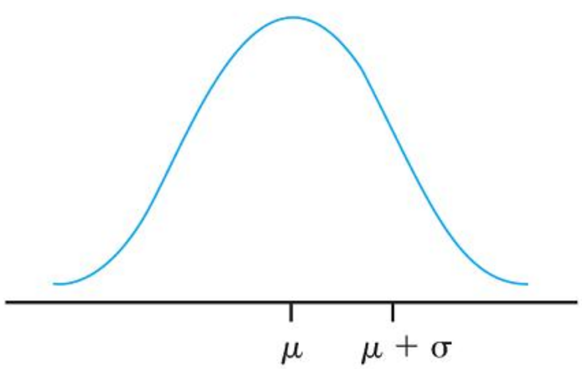
The statement that X is normally distributed with parameters μ and σ^2 is often abbreviated $X \sim N(\mu, \sigma^2)$.

The Normal Distribution

Figure below presents graphs of $f(x; \mu, \sigma)$ for several different (μ, σ) pairs.



Two different normal density curves



Visualizing μ and σ for a normal distribution

The Standard Normal Distribution

The normal distribution with parameter values $\mu = 0$ and $\sigma = 1$ is called the **standard normal distribution**.

A r.v. with this distribution is called a standard normal random variable and is denoted by *Z*. Its pdf is:

$$f(z; 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \text{ where } -\infty < z < \infty$$

The Standard Normal Distribution

The normal distribution with parameter values $\mu = 0$ and $\sigma = 1$ is called the **standard normal distribution**.

We use special notation to denote the cdf of the standard normal curve:

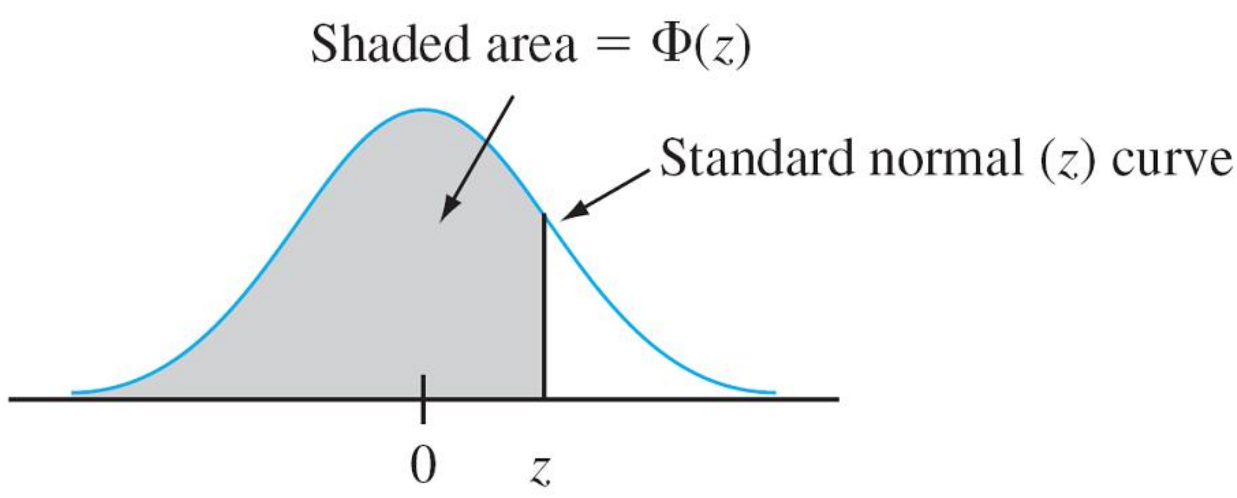
$$\Phi(z) = \int_{-\infty}^{z} f(y; 0, 1)\, dy$$

The Standard Normal Distribution

- The standard normal distribution rarely occurs naturally.
- Instead, it is a **reference distribution** from which information about other normal distributions can be obtained via a simple formula.
- These probabilities can then be found "normal tables".
- This can also be computed with a single command in R.

The Standard Normal Distribution

The figure below illustrates the probabilities found in a normal table (this can easily be found online):



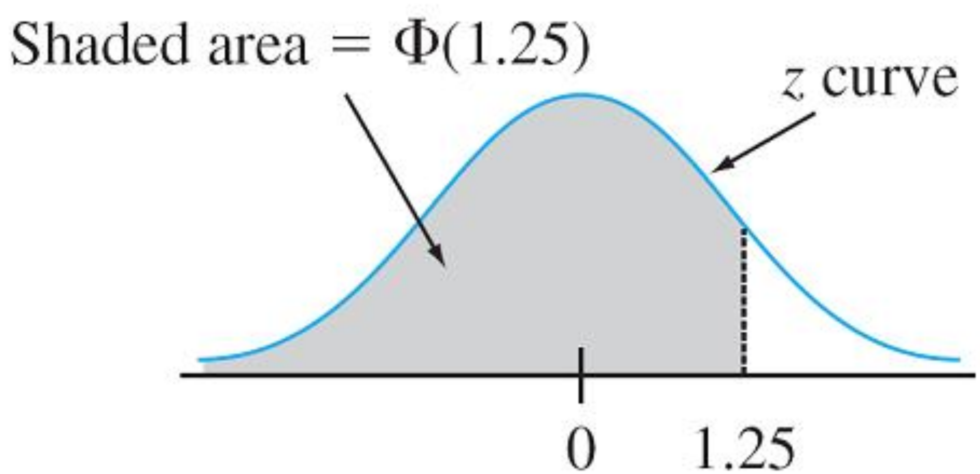
Example

cont'd

$P(Z \leq 1.25) = \Phi(1.25)$, a probability that is tabulated in a normal table.

What is this probability?

The figure below illustrates this probability:



Example

a) $P(Z \geq 1.25) = ?$

b) Why does $P(Z \leq -1.25) = P(Z >= 1.25)$? What is $\Phi(-1.25)$?

c) How do we calculate $P(-.38 \leq Z \leq 1.25)$?

Example

The *99th* percentile of the standard normal distribution is that value of z such that the area under the z curve to the left of the value is *0.99.*

Tables give for fixed z the area under the standard normal curve to the left of z, whereas now we have the area and want the value of z.

This is the "inverse" problem to $P(Z \leq z) = ?$

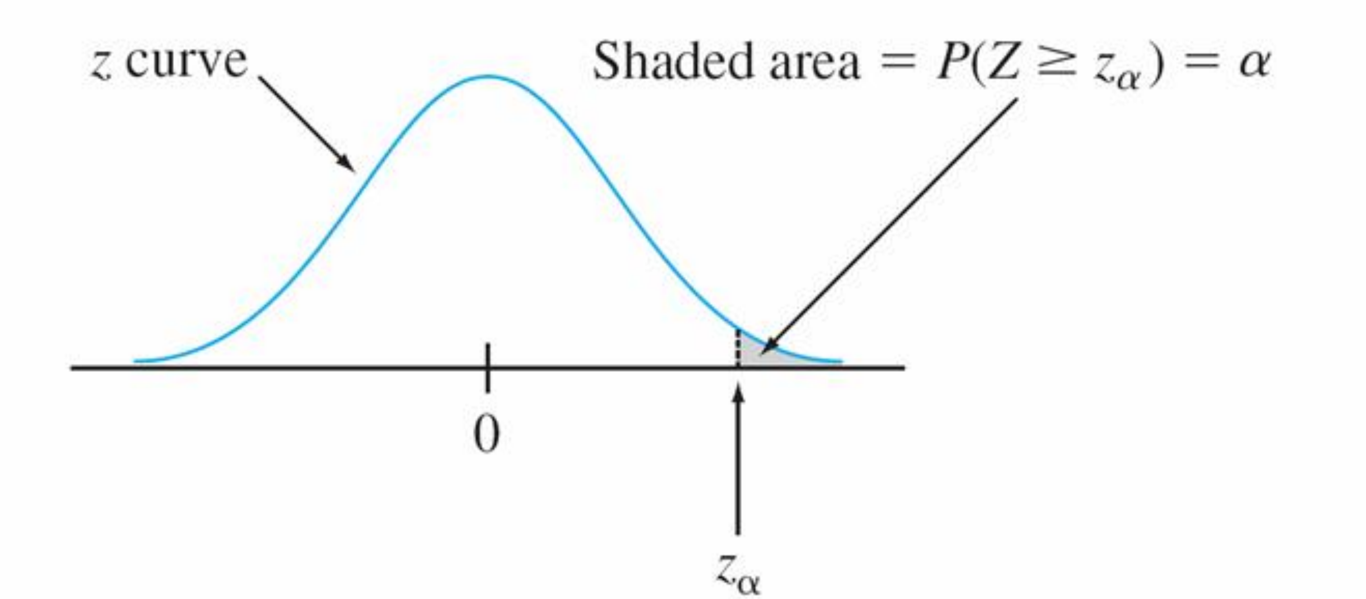
How can the table be used for this?

Notation: z_α

In statistical inference, we need the z values that give certain tail areas under the standard normal curve.

There, this notation will be standard:

z_α will denote the z value for which α of the area under the z curve <u>lies to the right</u> of z_α.



z_α Notation for z Critical Values

For example, $z_{.10}$ captures upper-tail area .10, and $z_{.01}$ captures upper-tail area .01.

Since α of the area under the z curve lies to the right of z_α, $1 - \alpha$ of the area lies to its left.

Thus z_α *is the* 100($1 - \alpha$)*th percentile of the standard normal distribution*.

Similarly, what does $-z_\alpha$ mean?

Nonstandard Normal Distributions

When $X \sim N(\mu, \sigma^2)$, probabilities involving X are computed by "standardizing." The **standardized variable** is $(X - \mu)/\sigma$.

Subtracting μ shifts the mean from μ to zero, and then dividing by σ scales the variable so that the standard deviation is 1 rather than σ.

Proposition

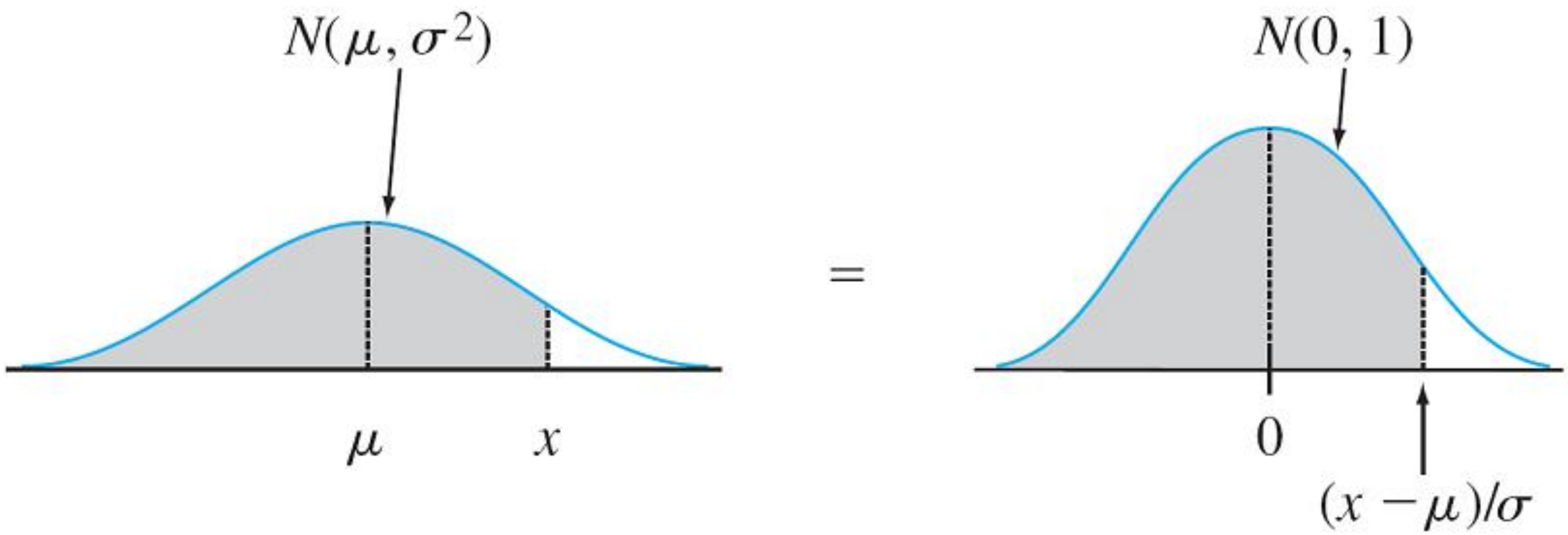
If X has a normal distribution with mean μ and standard deviation σ, then

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

is distributed standard normal.

Nonstandard Normal Distributions

Why do we standardize normal random variables?



Equality of nonstandard and standard normal curve areas

Example

The time that it takes a driver to react to the brake lights on a decelerating vehicle is critical in 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 with a normal distribution having mean value 1.25 sec and standard deviation of .46 sec.

What is the probability that reaction time is between 1.00 sec and 1.75 sec?

The Normal Distribution

We will revisit the normal distribution later on in this class to perform **statistical inference**.

THE NORMAL DISTRIBUTION IN R.