4.3 The Normal Distribution

DEFINITION

A continuous rv X is said to have a **normal distribution** with parameters μ and σ (or μ and σ^2), where $-\infty < \mu < \infty$ and $0 < \sigma$, if the pdf of X is

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/(2\sigma^2)} \qquad -\infty < x < \infty \tag{4.3}$$

e ≈ 2.71828, π ≈ 3.14159.

Notation: $N(\mu,\sigma^2)$

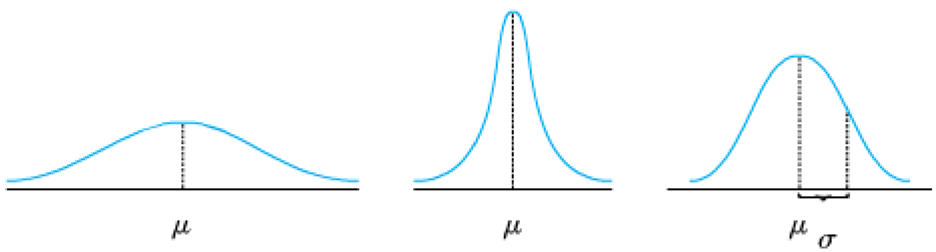


Figure 4.13 Normal density curves

Theorem: If X has normal distribution with parameters μ and σ then

$E(X) = \mu$

$V(X) = \sigma^2$

COMPUTATION

TI-83: TI-83 has normal probilities and percentiles programed in DISTR menu ([2ND] [VARS])

P(a ≤ X ≤ b) = **normalcdf(a,b,μ,σ)** (use a = - 10^99 if a = - ∞, and b = 10^99 if b = ∞)

100p%-th percentile = **invNorm(p,μ,σ)**

OR Table A3

DEFINITION

The normal distribution with parameter values $\mu = 0$ and $\sigma = 1$ is called the **standard normal distribution.** A random variable having a standard normal distribution is called a **standard normal random variable** and will be denoted by Z. The pdf of Z is

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

The graph of $f(z; 0, 1)$ is called the *standard normal* (or z) curve. Its inflection points are at 1 and -1. The cdf of Z is $P(Z \leq z) = \int_{-\infty}^{z} f(y; 0, 1)\, dy$, which we will denote by $\Phi(z)$.

Notation: N(0,1)

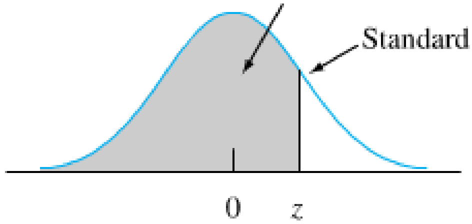


Figure 4.14 Standard normal cumulative areas tabulated in Appendix Table A.3

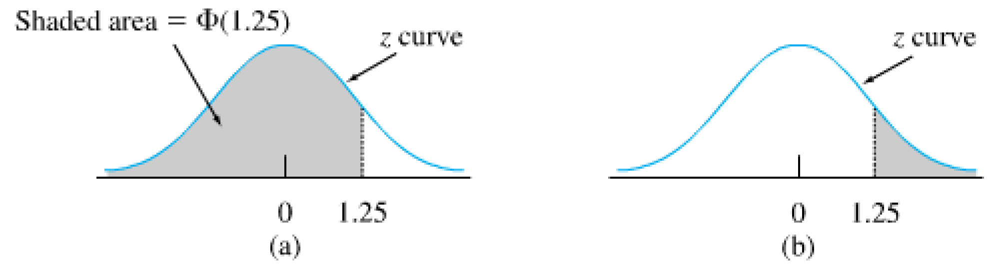
Example 4.13 Let's determine the following standard normal probabilities: (a) $P(Z \leq 1.25)$, (b) $P(Z > 1.25)$, (c) $P(Z \leq -1.25)$, and (d) $P(-.38 \leq Z \leq 1.25)$.

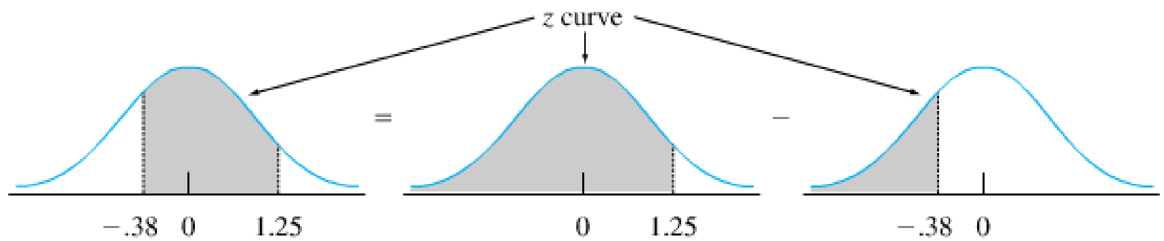
(a) $P(Z \le 1.25) = \Phi(1.25) = 0.8944$

(b) $P(Z > 1.25) = 1 - \Phi(1.25) = 0.1056$

(c) $P(Z \le -1.25) = \Phi(-1.25) = 0.1056$

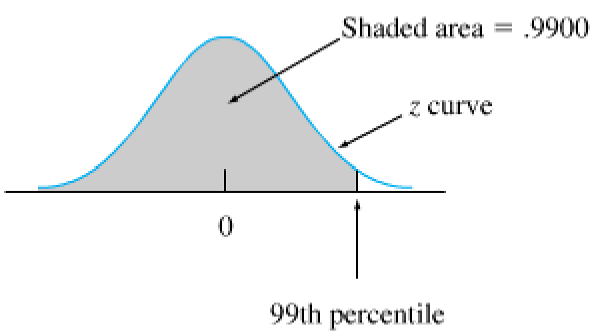
(d) $P(-.38 \le Z \le 1.25) = \Phi(1.25) - \Phi(-.38)$

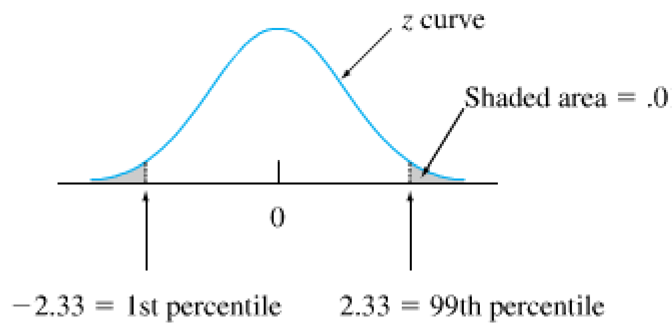




Finding percentiles of the Standard Normal Distribution:

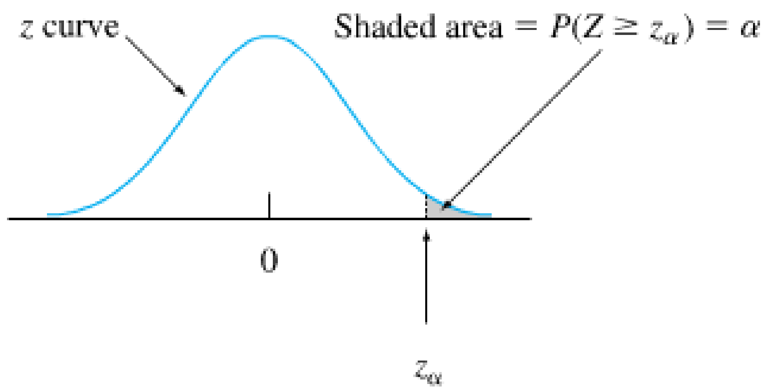
Example 2. Find the 99th percentile and the 1st percentile of the standard normal distribution.





Answer: 99th percentile ≈ 2.33, 1st percentile ≈ -2.33

z_α Value (critical value): z_α is the value of Z for which the area under the z-curve and to the right of z_α is α

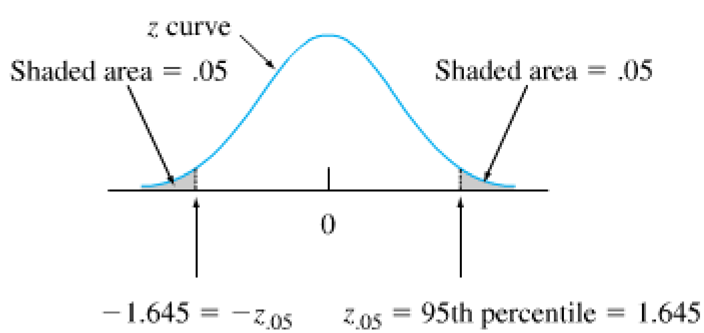


z_α = (1-α)100% percentile of the standard normal curve

Table 4.1 Standard Normal Percentiles and Critical Values

Percentile	90	95	97.5	99	99.5	99.9	99.95
α (tail area)	.1	.05	.025	.01	.005	.001	.0005
$z_\alpha = 100(1-\alpha)$th percentile	1.28	1.645	1.96	2.33	2.58	3.08	3.27

Example 3. Find $z_{0.05}$. Answer: $z_{0.05}$ = 95th percentile of standardnormal distribution = 1.645



Relation between nonstandard and standard normal distributions:

PROPOSITION

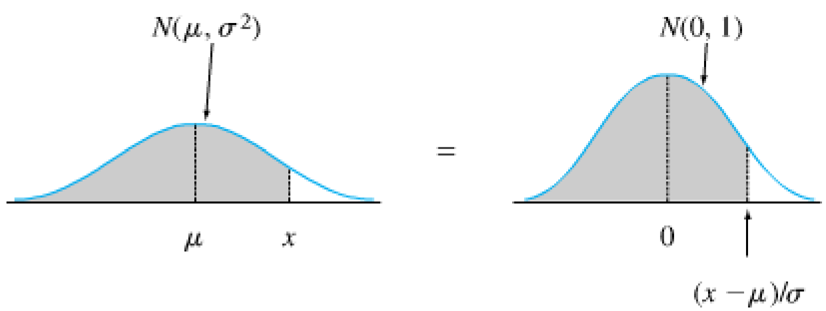
If X has a normal distribution with mean μ and standard deviation σ, then

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

has a standard normal distribution. Thus

$$P(a \le X \le b) = P\left(\frac{a - \mu}{\sigma} \le Z \le \frac{b - \mu}{\sigma}\right)$$
$$= \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)$$

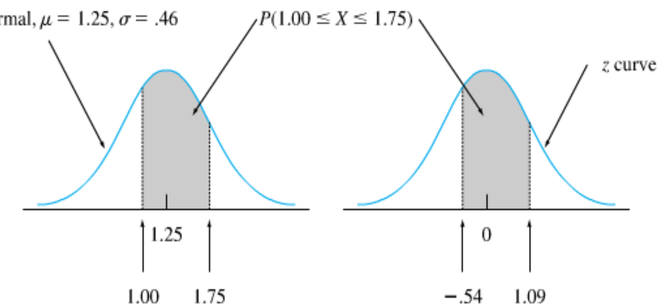
$$P(X \le a) = \Phi\left(\frac{a - \mu}{\sigma}\right) \qquad P(X \ge b) = 1 - \Phi\left(\frac{b - \mu}{\sigma}\right)$$



Example 4.16 Suppose that X is normally distributed with mean $\mu = 1.25$ and standard deviation $\sigma = 0.46$, that is X is N(1.25, .2116).

1. Find the probability $P(1.00 \le X \le 1.75)$.

$$P(1.00 \le X \le 1.75) = P\left(\frac{1.00 - 1.25}{.46} \le Z \le \frac{1.75 - 1.25}{.46}\right)$$
$$= P(-.54 \le Z \le 1.09) = \Phi(1.09) - \Phi(-.54)$$
$$= .8621 - .2946 = .5675$$



2. Find the 90th percentile η(.90) of X, that is find a number c such that $P(X \le c) = 0.90$.

   The 90th percentile of Z is 1.28, which means that

$$P(Z \le 1.28) = 0.90$$

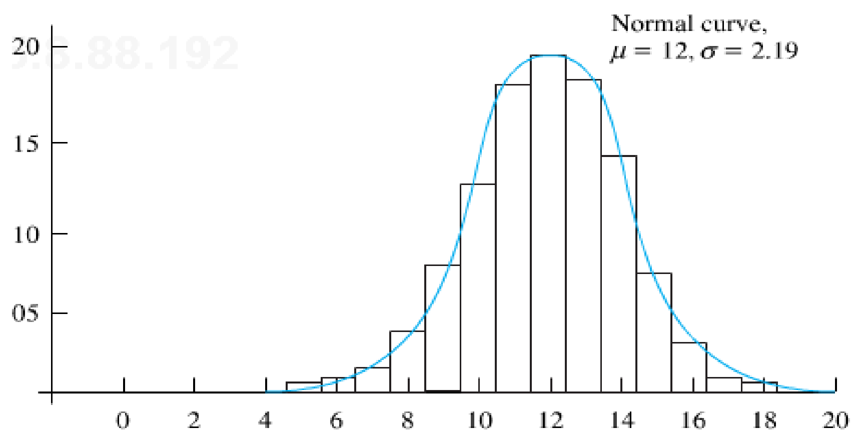
Since $Z = \frac{X-1.25}{.46}$, solving inequality $\frac{X-1.25}{.46} \le 1.28$ for X we obtain that $P(X \le 1.28 \times 0.46 + 1.25) = 0.90$. Hence the 90th percentile c of X is $c = 1.28 \times 0.46 + 1.25 = 1.84$

TI-83:

1. $P(1.00 \le X \le 1.75)$ = normalcdf(1,1.75,1.25,0.46) = 0.5681
2. 90th percentile of X = invNorm(.9,1.25,0.46) = 1.84

Approximating Binomial Distribution:

For large *n* the cdf of binomial distribution with parameters *n* and *p* can be approximated by normal probabilities with $\mu = np$ and $\sigma^2 = n\,p\,(1-p)$



Binomial probability histogram for $n = 20$, $p = .6$ with normal approximation

PROPOSITION

Let X be a binomial rv based on n trials with success probability p. Then if the binomial probability histogram is not too skewed, X has approximately a normal distribution with $\mu = np$ and $\sigma = \sqrt{npq}$. In particular, for $x =$ a possible value of X,

$$P(X \le x) = B(x, n, p) \approx \begin{pmatrix} \text{area under the normal curve} \\ \text{to the left of } x + .5 \end{pmatrix}$$

$$= P\left(\tilde{X} \le x + 0.5\right), \qquad \textit{where } \tilde{X} \textit{ is } N(np, np(1-p))$$

In practice, the approximation is adequate provided that both $np \ge 10$ and $nq \ge 10$, since there is then enough symmetry in the underlying binomial distribution.

EXERCISES 4.3

Verify *the Empirical Rule* (below).

If the population distribution of a variable is (approximately) normal, then

1. Roughly 68% of the values are within 1 SD of the mean.
2. Roughly 95% of the values are within 2 SDs of the mean.
3. Roughly 99.7% of the values are within 3 SDs of the mean.

here SD = standard deviation = σ

35. Suppose the diameter at breast height (in.) of trees of a certain type is normally distributed with $\mu = 8.8$ and $\sigma = 2.8$, as suggested in the article "Simulating a Harvester-Forwarder Softwood Thinning" (*Forest Products J.*, May 1997: 36–41).

a. What is the probability that the diameter of a randomly selected tree will be at least 10 in.? Will exceed 10 in.?

b. What is the probability that the diameter of a randomly selected tree will exceed 20 in.?

c. What is the probability that the diameter of a randomly selected tree will be between 5 and 10 in.?

d. What value c is such that the interval $(8.8 - c, 8.8 + c)$ includes 98% of all diameter values?

e. If four trees are independently selected, what is the probability that at least one has a diameter exceeding 10 in.?

Answers: 0.3341, 0.0000316, 0.5785, 6.52, 0.8034

40. The article "Monte Carlo Simulation—Tool for Better Understanding of LRFD" (*J. Structural Engr.,* 1993: 1586–1599) suggests that yield strength (ksi) for A36 grade steel is normally distributed with $\mu = 43$ and $\sigma = 4.5$.

a. What is the probability that yield strength is at most 40? Greater than 60?

b. What yield strength value separates the strongest 75% from the others?

Answers: a. 0.2525, 0.00008; b. 39.96

54. Suppose that 10% of all steel shafts produced by a certain process are nonconforming but can be reworked (rather than having to be scrapped). Consider a random sample of 200 shafts, and let X denote the number among these that are nonconforming and can be reworked. What is the (approximate) probability that X is

a. At most 30?

b. Less than 30?

c. Between 15 and 25 (inclusive)?

Answers: 0.993, 0.987, 0.792