Introduction to Statistical Modelling

STAT2507D

Chapter 6-2

The Normal Probability Distributions

The Normal Distribution

- Continuous probability distributions can assume a variety of shapes.
- However, a large number of random variables observed in nature possess a frequency distribution that is approximately mound-shaped or approximately normal distribution

THE NORMAL DISTRIBUTION

- The formula that generates the normal probability distribution is:

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

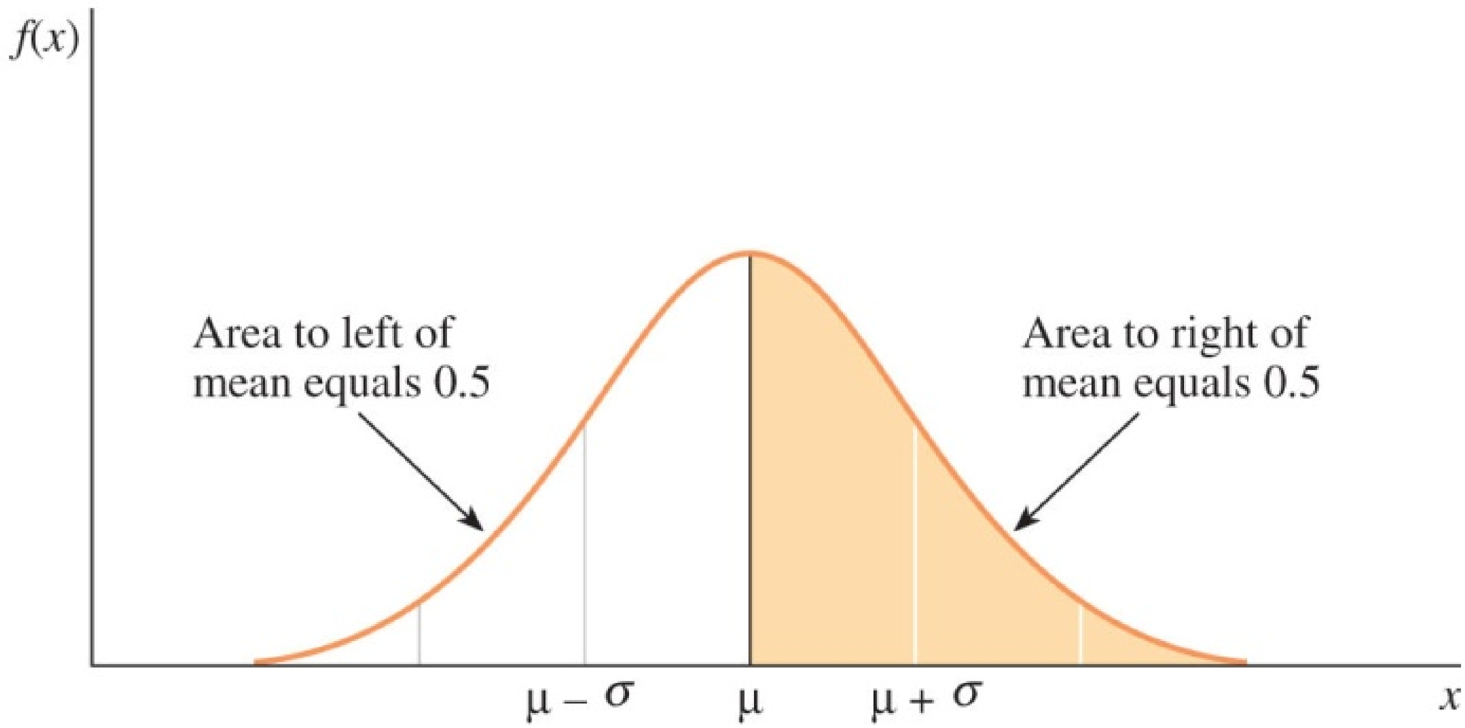
$$-\infty < \mu < \infty, \sigma > 0, e = 2.7183, \pi = 3.1416$$

  - μ and σ are the population mean and standard deviation

THE NORMAL DISTRIBUTION WITH MEAN μ, STD σ

- The mean μ locates the centre of the distribution and distribution is symmetric about mean μ
- Total area under the curve is 1. Symmetry implies area right of μ is 0.5 and left of μ is 0.5.
- The shape of the distribution is determined by σ, the population standard deviation.

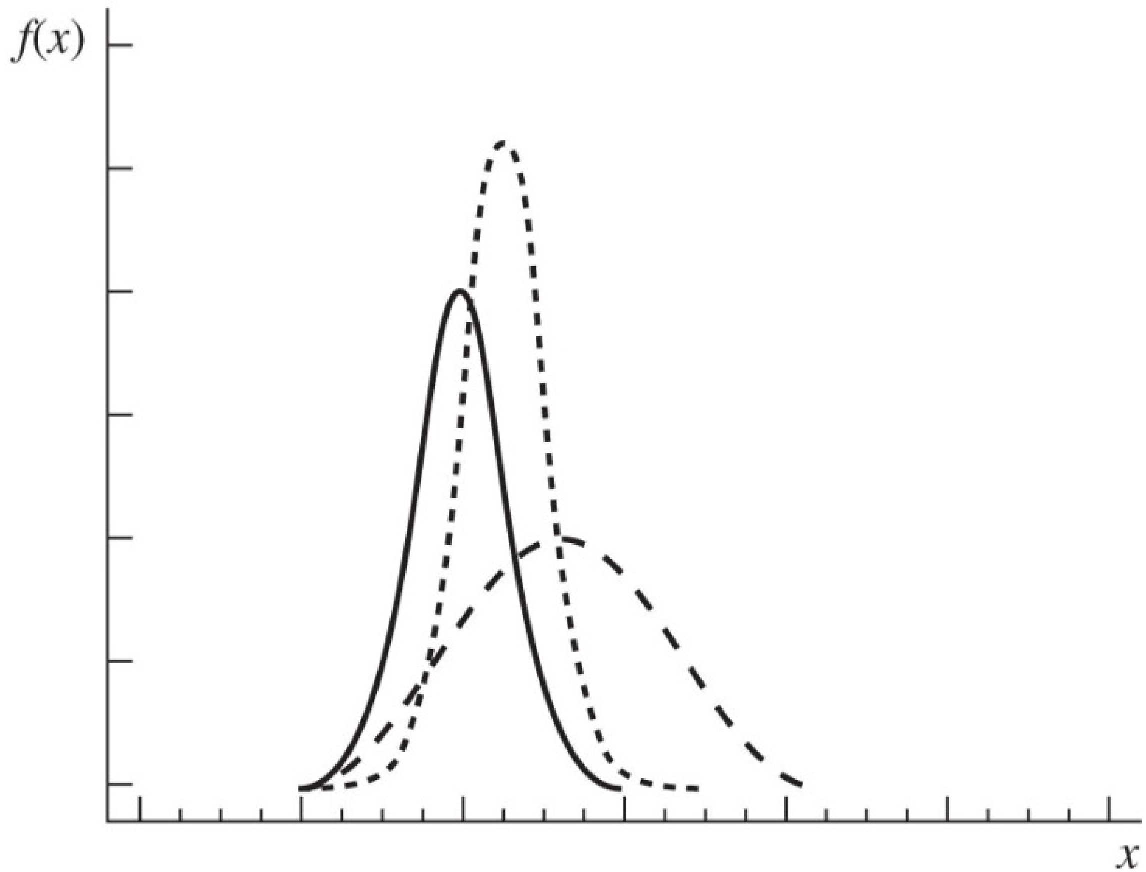
THE NORMAL DISTRIBUTION WITH MEAN μ, STD σ



THE NORMAL DISTRIBUTION WITH MEAN μ, STD σ

- Larger values of σ reduce the height of the curve and increases the spread;
- Small values of σ increase the height of the curve and reduce the spread
- According to Empirical rule, almost all values of a normal random variable lie in the interval $\mu \pm 3\sigma$

NORMAL PROBABILITY DISTRIBUTION WITH DIFFERING VALUES OF μ AND σ



THE NORMAL DISTRIBUTION WITH MEAN μ, STANDARD DEVIATION σ

- There are infinitely large number of normal distributions – one for each different mean and standard deviation.
- Mean could take any values and σ could be any value greater than zero

  (i.e. $-\infty < \mu < \infty, \sigma > 0$)
- We use a standardization procedure that allows us to use the same table for all normal distributions.

Standard Normal Random Variable

- To find P(a < x < b), we need to find the area under the appropriate normal curve
- To simplify the tabulation of these areas, we **standardize** each value of x by expressing it as a z-score, the number of standard deviations - σ it lies from the mean - μ

$$z = \frac{x - \mu}{\sigma}$$

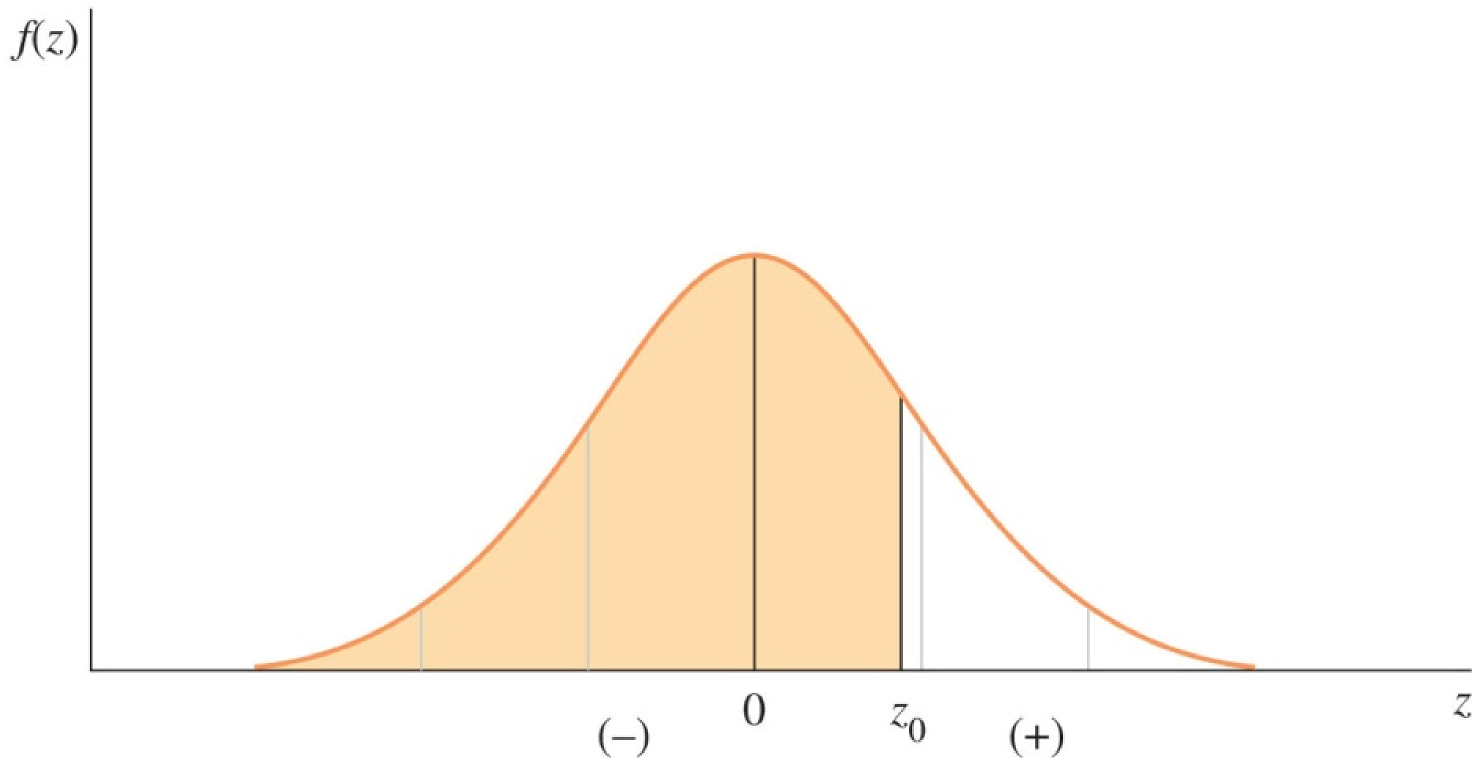
- Or equivalently

$$x = \mu + z\sigma$$

The Standard Normal Random Variable (z)

- Mean = 0; Standard deviation = 1
- Symmetric about z = 0 (Area left of z=0 is 0.5; and area right of z=0 is 0.5 as well)
- Values of z to the left of centre are negative
- Values of z to the right of centre are positive
- Total area under the curve is 1

The Standard Normal Random Variable (z)



USING TABLE 3

- The 4-digit probability in a particular row and column of Table 3 (page 720 – 721) gives the area under the *z* curve to the left that particular value of *z*

6.5) Let $z_0 = 1.00 \Rightarrow P(z \leq 1) = 0.8413$

USING TABLE 3

z	.00	.01	.02	.03	.04	.05	.06
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051
0.9	0.8159	0.8186	0.8212	0.8328	0.8264	0.8289	0.8315
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9278

Area for z_0 = 1.00

EXERCISE

6.6) Find the following:

a. $P(z \leq 1.63)$
b. $P(z \ \geq -0.5)$
c. $P(-0.5 \ \leq z \ \leq 1.0)$
d. $P(-1 \ \leq z \ \leq 1)$ using only positive z-table
e. $P(-2.33 \leq z \leq 2.33)$ using only positive z-table

Exercise

6.7) Find z_0

a. $P(z < z_0) = 0.975$

b. $P(z < z_0) = 0.950$

c. $P(-z_0 < z < z_0) = 0.90$

d. $P(-z_0 < z < z_0) = 0.95$

CALCULATING PROBABILITIES OF GENERAL NORMAL R.VS

- Probability of interest could involve X that is a normal random variable with mean μ and standard deviation σ.
- It is denoted by $X \sim N(\mu, \sigma^2)$

Calculating probabilities of General Normal R.Vs

- Standardize the interval of interest, writing it as the equivalent interval in terms of z, the standard normal random variable. It is denoted by Z ~N(0,1)

$$z = \frac{x - \mu}{\sigma}$$

EXERCISE

6.8) x has a normal distribution with $\mu = 10$ and $\sigma = 2$. Find $P(11 < x < 13.7)$

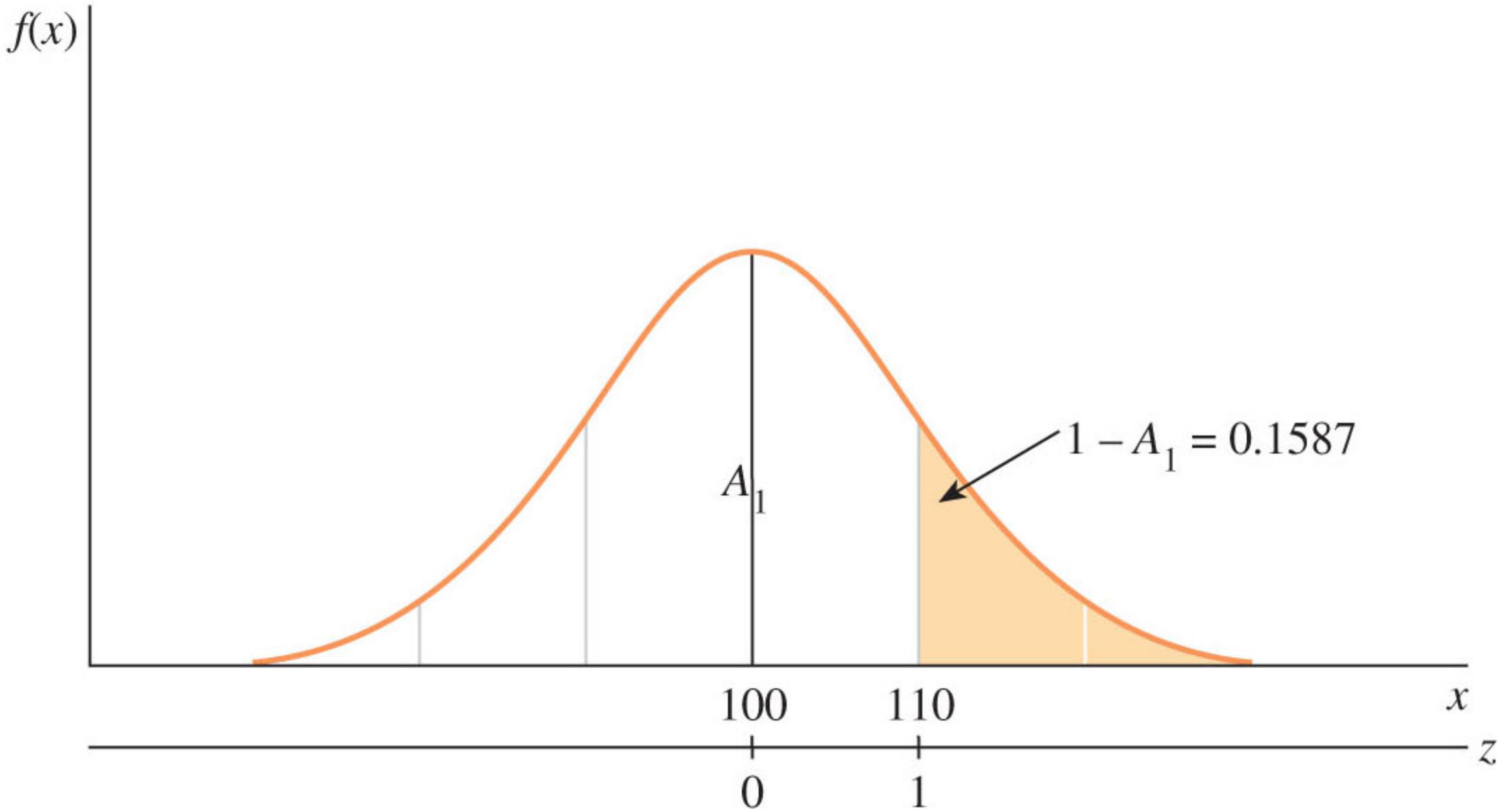
6.9) A normal random variable x has unknown mean μ and standard deviation $\sigma = 2$. If the probability that x exceeds 7.5 is 0.8023, find μ.

EXERCISE

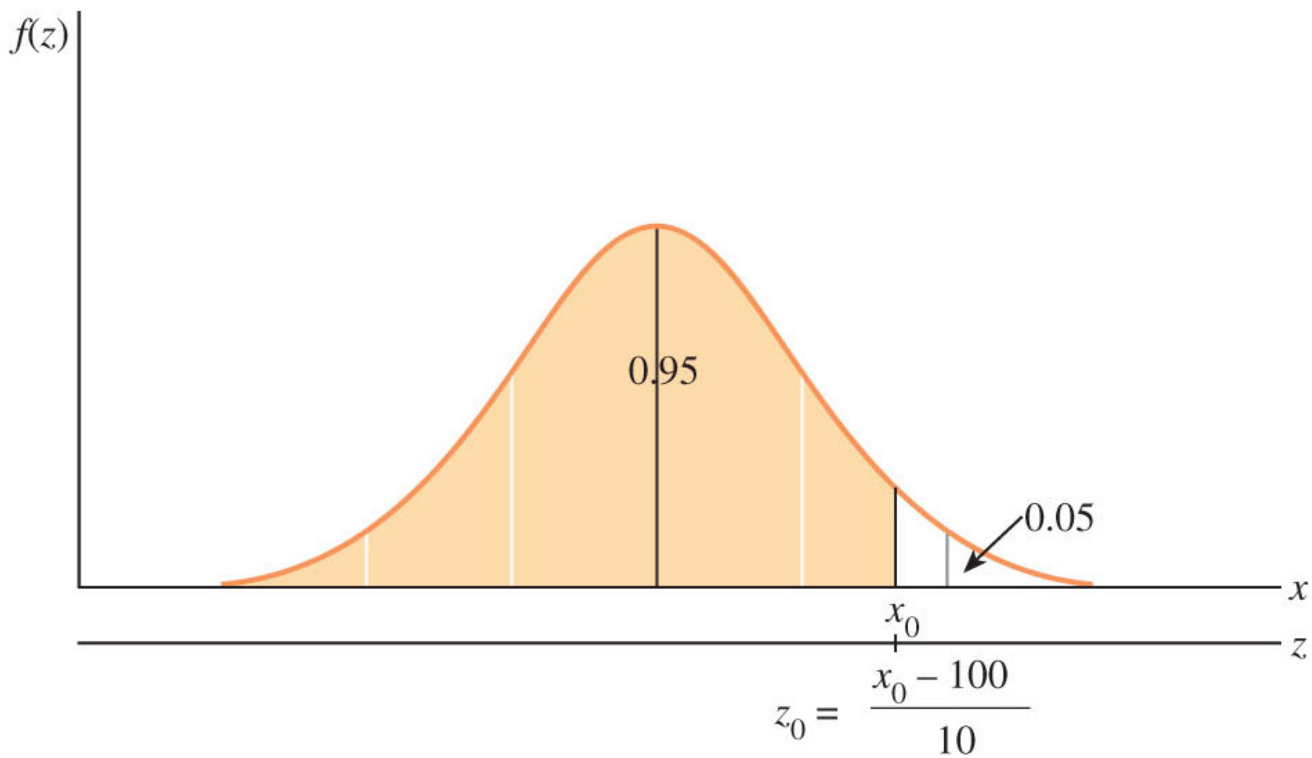
6.10) Gasoline use for compact cars sold in North America is normally distributed, with a mean of 100 km per 9 L (km/L) and a standard deviation of 10 km/9 L.

a. What percentage of compacts get 110 km/9 L or more?
b. If a manufacturer wishes to develop a compact car that outperforms 95% of the current compacts in fuel economy, what must the gasoline use rate for the new car be?

EXERCISE



EXERCISE



The Normal Approximation to Binomial

- Let X be a binomial random variable with n trials and probability p of successes.
- The probability distribution of X can be approximated (under certain conditions) using a normal curve with $\mu = np \; and \; \sigma = \sqrt{npq}$
- This approximation is adequate as long as n is large and p is not too close to 0 or 1.
- Rule of Thumb: $np > 5 \; and \; nq > 5$

The Normal Approximation to Binomial

- Binomial random variable X is a discrete random variable.
- We know that $P(X = a) = 0$ for binomial as binomial is a discrete random variable.
- Normal is a continuous random variable.
- **Continuity correction** is a necessary adjustment when approximating binomial with Normal.

THE NORMAL APPROXIMATION TO BINOMIAL

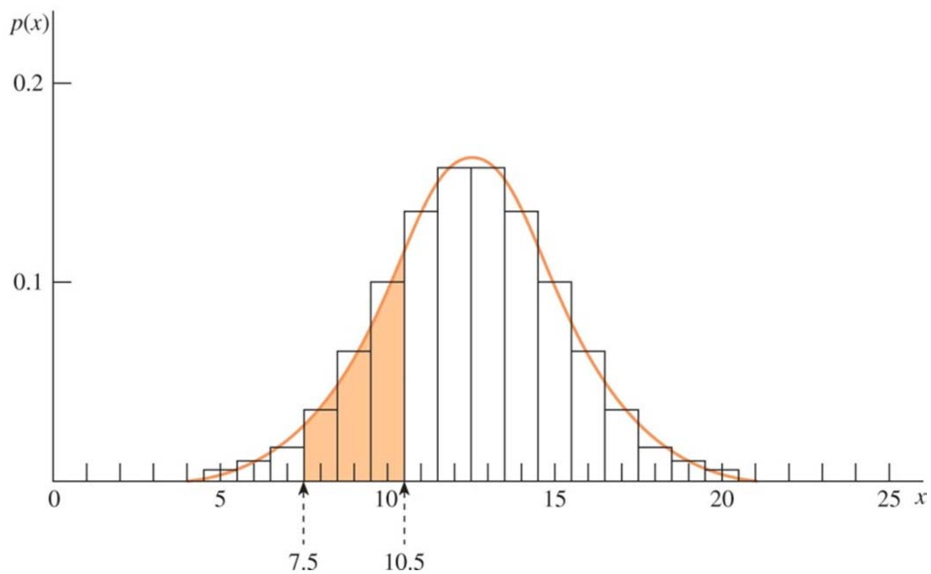
- Continuity correction takes account of the fact that the approximation is from a discrete to a continuous random variable. The continuity correction is applied as follows:

$$z = \frac{x \pm 0.5 - np}{\sqrt{npq}}$$

Required condition: $np > 5 \; and \; nq > 5$

THE NORMAL APPROXIMATION TO BINOMIAL

- Make sure to include the entire rectangle for the values of x in the interval of interest; this is called the continuity correction.



THE NORMAL APPROXIMATION TO BINOMIAL

Binomial	Normal
$P(X = a)$	$P(a - 0.5 < X < a + 0.5)$
$P(X \geq a)$	$P(X > a - 0.5)$
$P(X > a)$	$P(X > a + 0.5)$
$P(X \leq a)$	$P(X < a + 0.5)$
$P(X < a)$	$P(X < a - 0.5)$

EXERCISE

6.11) Let x be a binomial random variable with n=100 and p=0.2. Find the approximation to these probabilities:

a) $P(x > 22)$
b) $P(x \geq 22)$
c) $P(20 < X < 25)$
d) $P(x \leq 25)$

EXERCISE

6.12) A production line produces AA batteries with a reliability rate of 95%. A sample of n = 200 batteries is selected. Find the probability that at least 195 of the batteries work

EXERCISE

6.13) A used-car dealership has found that the length of time before a major repair is required on the cars it sells is normally distributed, with mean equal to 10 months and a standard deviation of 3 months. If the dealer wants only 5% of the cars to fail before the end of the guarantee period, for how many months should the cars be guaranteed?

EXERCISE

6.14) The diameters of Douglas firs grown at a Christmas tree farm are normally distributed with a mean of 10cm and a standard deviation of 3cm.

a. What proportion of trees will have diameters between 8 and 12cm?
b. What proportion of trees will have diameters less than 7cm?
c. You Christmas stand will expand to a diameter of 14cm. What proportion of the trees will not fit in your Christmas tree stand?

DISCRETE AND CONTINUOUS PROBABILITIES WHEN USING CUMULATIVE TABLE

Probabilities	Discrete	Continuous
$P(X < a)$	$P(X \leq (a-1))$	$P(X \leq a)$
$P(X > a)$	$1 - P(X \leq a)$	$1 - P(X \leq a)$
$P(X \geq a)$	$1 - P(X \leq (a-1))$	$1 - P(X \leq a)$
$P(a \leq X \leq b))$	$P(X \leq b) - P(X \leq (a-1))$	$P(X \leq b) - P(X \leq a)$
$P(a < X \leq b)$	$P(X \leq b) - P(X \leq a)$	$P(X \leq b) - P(X \leq a)$
$P(a \leq X < b)$	$P\big(X \leq (b-1)\big) - P(X \leq (a-1))$	$P(X \leq b) - P(X \leq a)$
$P(a < X < b)$	$P\big(X \leq (b-1)\big) - P(X \leq a))$	$P(X \leq b) - P(X \leq a)$

SUMMARY

- Continuous Probability Distributions
  - Continuous random variables
  - Probability distributions of probability density functions
    - Curves are smooth
    - The area under the curve between a and b represents the probability that x falls between a and b
    - $P(x = a) = 0$ for continuous random variables

SUMMARY

- The Normal Probability Distribution
  - Symmetric about its mean μ
  - Shape determined by its standard deviation σ
- The Standard Normal Distribution
  - The standard normal random variable z has mean 0 and standard deviation 1
  - Any normal random variable x can be transformed to a standard normal random variable using:
  - Convert necessary values of x to z
  - Use Table 3 in Appendix I to compute standard normal probabilities

SUMMARY

- Several important z-values have tail areas as follows:

Tail Area:	.005	.01	.025	.05	.10
z-Value:	2.58	2.33	1.96	1.645	1.28

- Normal Approximation to Binomial Distribution