Ch6. The Normal Probability Distribution

In Chapters 5, you learned about discrete random variables and their probability distributions. In this chapter, you will learn about continuous random variables and their probability distributions and about one very important continuous random variable - the normal. You will learn how to calculate normal probabilities and, under certain conditions, how to use the normal probability distribution to approximate the binomial probability distribution. Then, in Chapter 7 and in the chapters that follow, you will see how the normal probability distribution plays a central role in statistical inference.

윤 연 옥

Continuous Random Variables

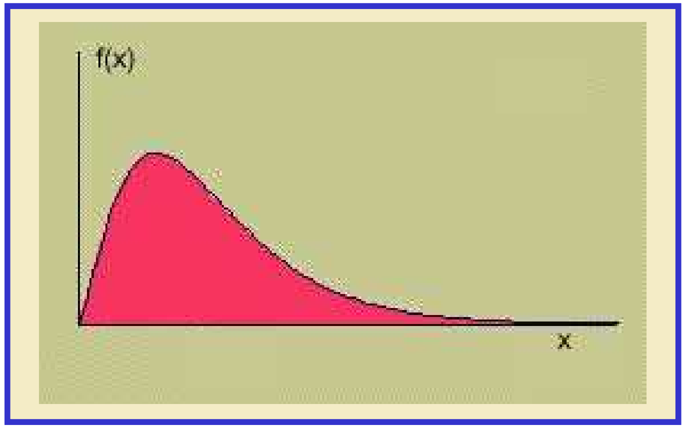
- Continuous random variables can assume the infinitely many values corresponding to points on a line interval.

Example

- ✓ Heights, weights
- ✓ length of life of a particular product

Continuous Random Variables

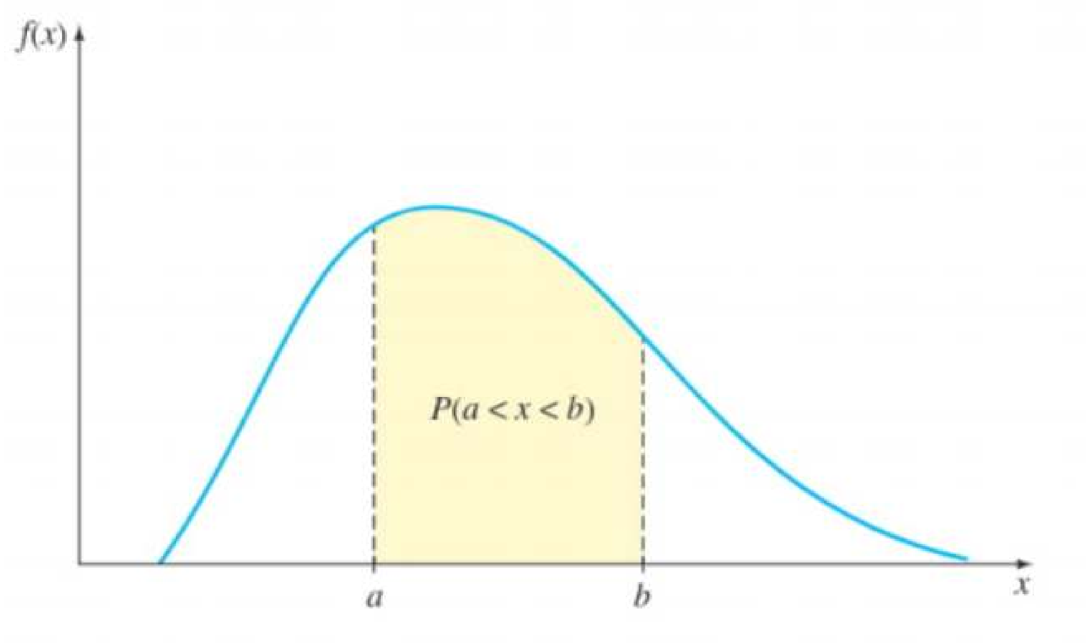
- A **smooth curve** describes the probability distribution of a continuous random variable



- The depth or density of the probability, which varies with x, may be described by a mathematical formula $f(x)$, called the probability distribution or probability density function for the random variable x .

Properties of Continuous Probability Distributions

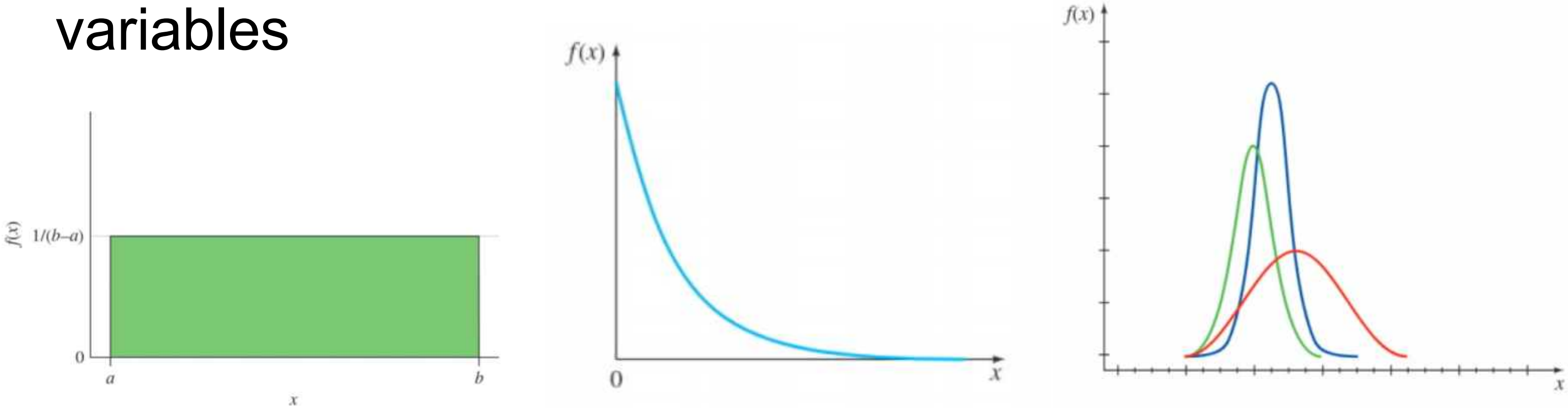
- The area under the curve is equal to **1.**
- $P(a \le x \le b)$= area under the curve between a and b.



- There is no probability attached to any single value of x .
  - ✓ That is, $P(x = a) = 0,\ P(x \ge a) = P(x > a)$

Continuous Probability Distributions

- There are many different types of continuous random variables



- We try to pick a model - probability distribution f(x)- that
  - Fits the data well
  - Allows us to make the best possible inferences using the data.
- One important continuous random variable is the **normal random variable.**

The Normal Probability Distribution

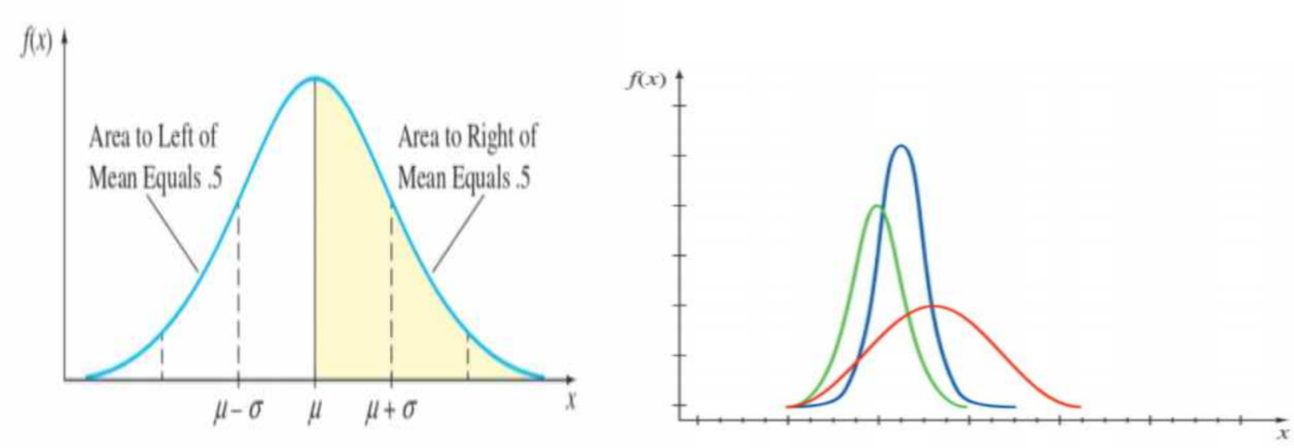
- The formula that generates the normal probability distribution is:

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

$$e = 2.7183 \qquad \pi = 3.1416$$

μ and σ are the population mean and standard deviation

- The shape and location of the normal curve changes as the mean and standard deviation change.



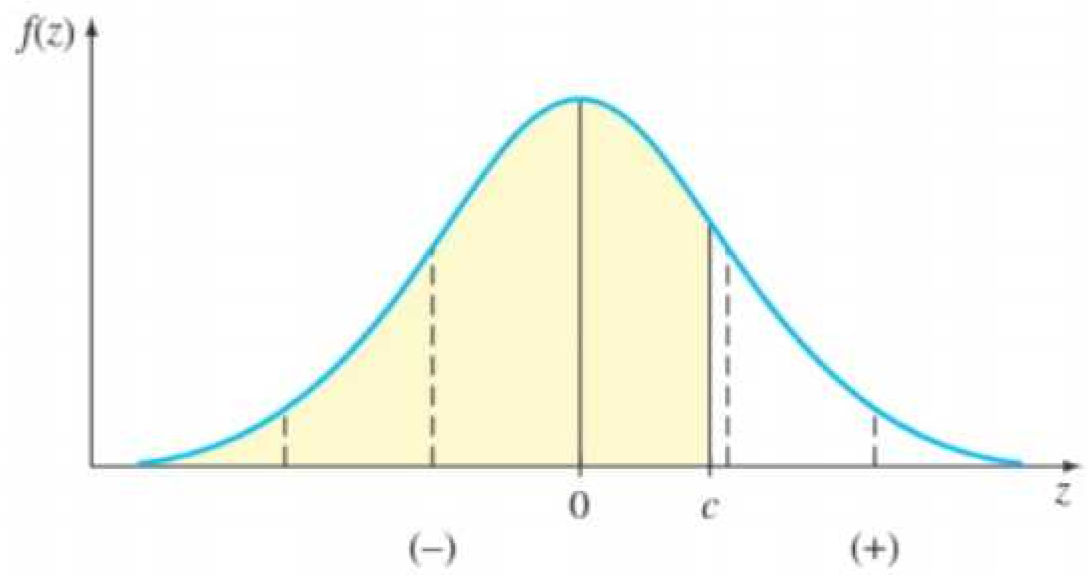
The Standard Normal Distribution

- 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 .

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

The Standard Normal(z) Distribution

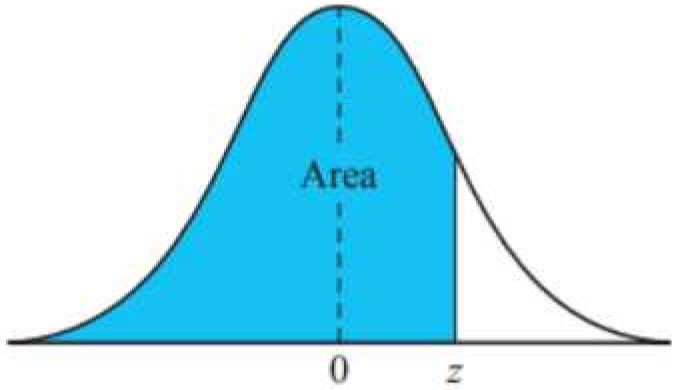
- Mean = 0; Standard deviation = 1
- When $x = \mu,\ z = 0$
- Symmetric about $z = 0$
- Values of z to the left of center are negative
- Values of z to the right of center are positive
- Total area under the curve is 1.



Using Table 3

- The four digit probability in a particular row and column of Table 3 gives the area under the z curve to the left that particular value of z.

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



$P(z < 1.36)$

Area for z = 1.36

Using Table 3

z	.00	.01	.02	.03	.04	.05	.06	.07
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011
−2.9	.0019	.0018	.0017	.0017	.0016	.0016	.0015	.0015
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192

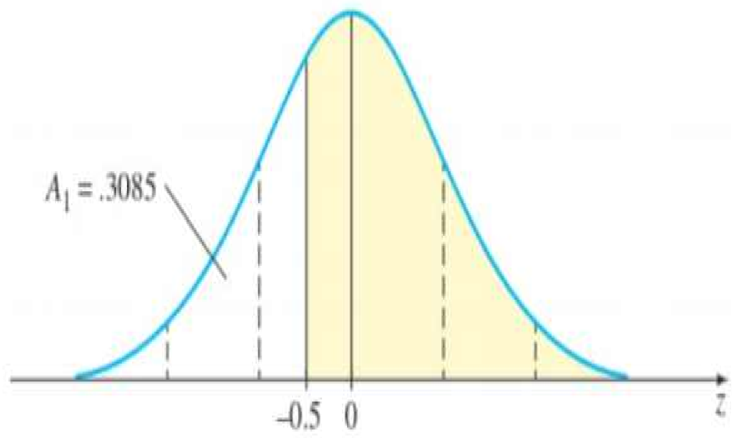
Example

- $P(z \geq -0.5)$

  $= 1 - A_1$

  $= 1 - .3085 = .6915$

  or $\mathrm{P}(\mathrm{z} \geq -0.5) = P(z \leq 0.5)$

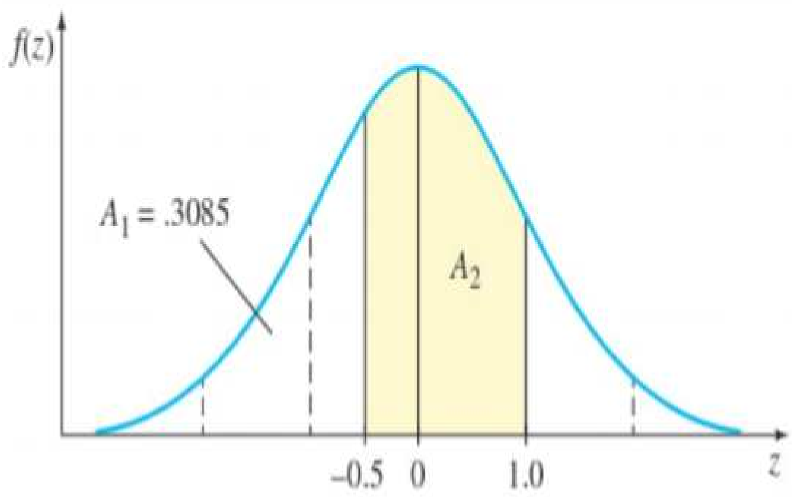


- $P(-.5 \leq z \leq 1.0)$

  $= P(z \leq 1.0) - P(z \leq -0.5)$

  $= (A_1 + A_2) - A_1$

  $= .8413 - .3085 = .5328$



Example

- Find the probability that a normally distributed random variable will fall within these ranges:

  - One, two, and three standard deviations of its mean

Sol)

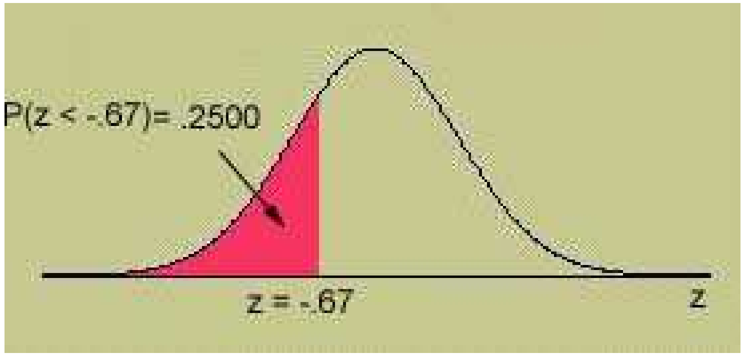
1. $P(-1 \leq z \leq 1) = P(z \leq 1) - P(z \leq -1) = .8413 - .1587 = .6826$
2. $P(-2 \leq z \leq 2) = P(z \leq 2) - P(z \leq -2) = .9772 - .0228 = .9544$
3. $P(-3 \leq z \leq 3) = P(z \leq 3) - P(z \leq -3) = .9987 - .0013 = .9974$

✓ Remember the Empirical Rule:

- Approximately 99.7%(95%, 68%) of the measurements lie within 3(2, 1) standard deviations of the mean.

Example : working backwards

- Find the value of z that has area .25 to its left.



steps

1. Look for the four digit area closest to .2500 in Table 3.
2. What row and column does this value correspond to?
3. Z=-.67

✓ 25th percentile, or 1st quartile (Q_1)

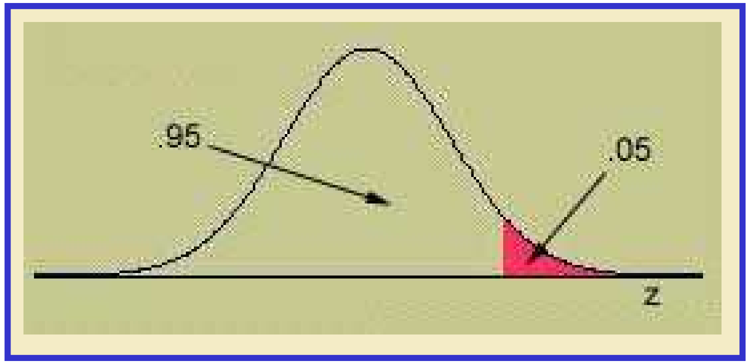
z	.00	.01	.02	.03	.04	.05	.06	.07	.08
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	$\boxed{0.2514}$	0.2483
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810

Example : working backwards

- Find the value of z that has area .05 to its right.

<u>steps</u>

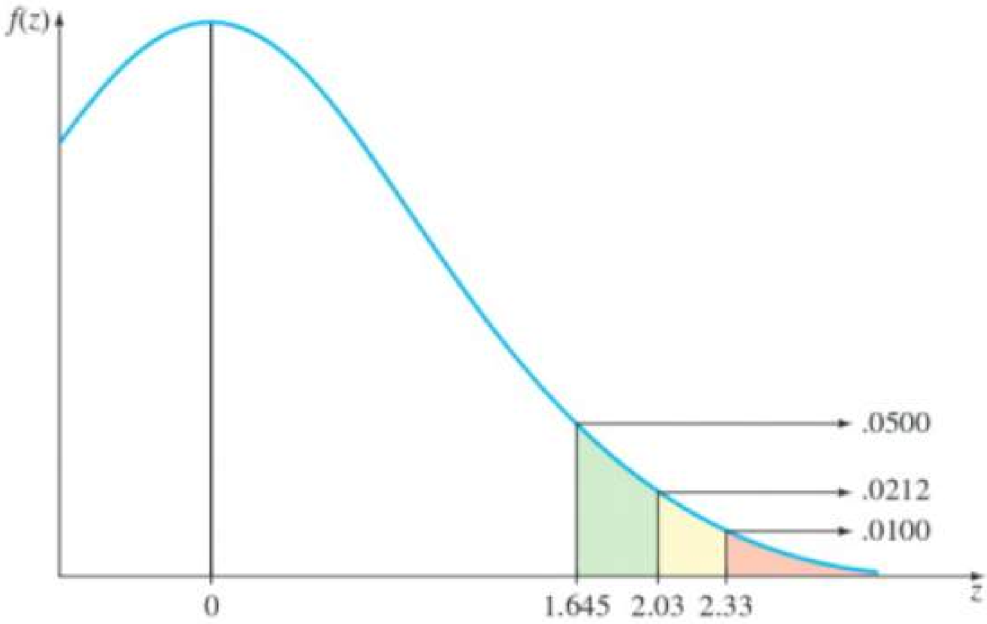
1. The area to its left will be
   1 - .05 = .95
2. Look for the four digit area closest to .9500 in Table 3.
3. Since the value .9500 is halfway between .9495 and .9505, we choose z halfway between 1.64 and 1.65.
4. z = 1.645



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767

- Several important z-values have right-tail areas

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

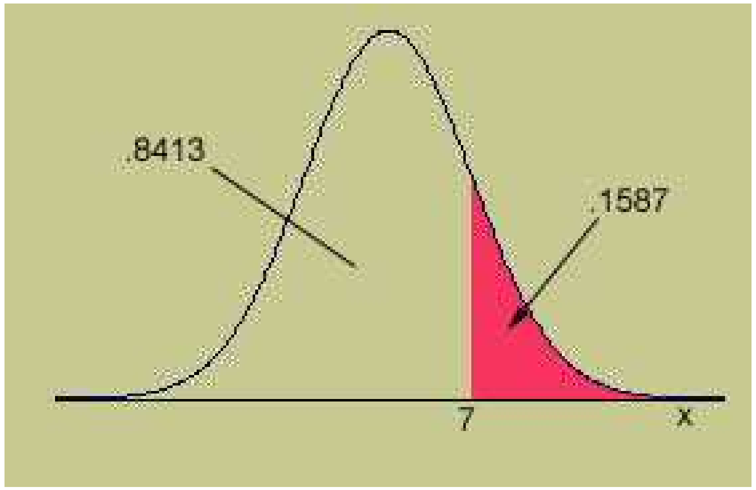


- To find an area for a normal random variable x with mean μ and standard deviation σ, standardize or rescale the interval in terms of z.
- Find the appropriate area using Table 3.

Example: x has a normal distribution with $\mu = 5$ and $\sigma = 2$.

Find $P(x > 7)$.

$$P(x > 7) = P\left(z > \frac{7-5}{2}\right)$$
$$= P(z > 1) = 1 - .8413 = .1587$$



Example

(1) Studies show that gasoline use for compact cars sold in the United States is normally distributed, with a mean of 35.5 miles per gallon (mpg) and a standard deviation of 4.5 mpg. What percentage of compacts get 40 mpg or more?

Sol)

$$P(x \geq 40) = 1 - P\left(z \leq \frac{40 - 35.5}{4.5}\right)$$
$$= 1 - P(z \leq 1) = 1 - .8413 = .1587$$

The percentage exceeding 40 mpg is 100(.1587) = 15.87%

Example

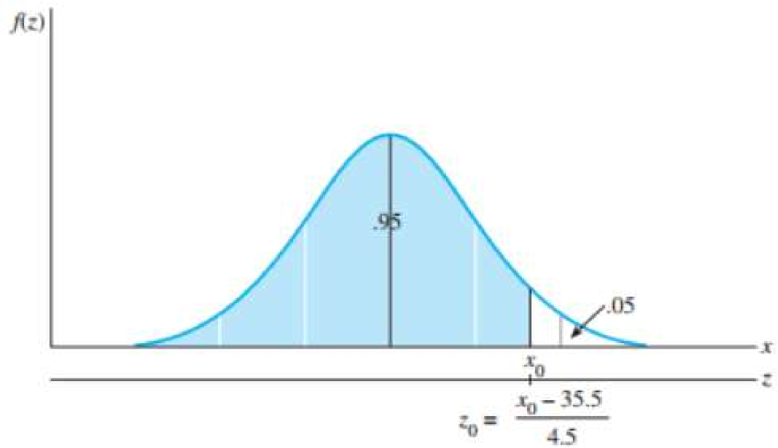
(2) If a manufacturer wishes to develop a compact car that outperforms 95% of the current compacts in fuel economy, what must the gasoline use rate for the new car be?

Sol) $P(x \leq x_0) = .95$

$$P(x \leq x_0) = \left(z \leq \frac{x_0 - 35.5}{4.5} = z_0\right) = .95$$

$$z_0 = \frac{x_0 - 35.5}{4.5} = 1.645$$

$$\Rightarrow x_0 = 35.5 + 1.645(4.5) = 42.9$$



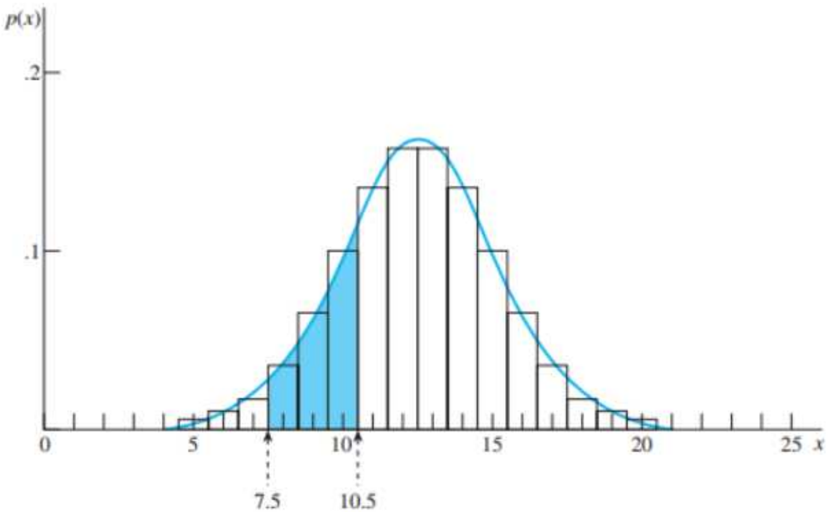
The Normal Approximation to the Binomial

- We can calculate binomial probabilities using
  - The binomial formula
  - The cumulative binomial tables

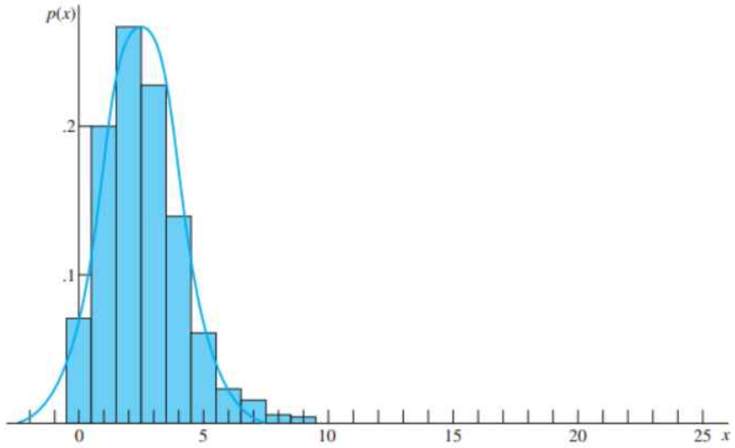
✓ *Poisson approximation can be used when $np < 7$ (n is large, p is small)*

- When n is large, and p is not too close to zero or one, areas under the normal curve with mean np and variance npq can be used to approximate binomial probabilities.

B(n=25, p=.5)



B(n=25, p=.1)



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

  - correct the value of x by $\pm .5$

- Standardize the values of x using

$$z = \frac{x \pm .5 - np}{\sqrt{npq}}$$

- Make sure that ***np > 5* and *nq >5*** to avoid inaccurate approximations!

Example

- Suppose x is a binomial random variable with $n = 25$ and $p = .5$. Using the normal approximation to find $P(8 \le x \le 10)$

$$n = 25,\ p = .5,\ q = .5,\ np = 12.5,\ nq = 12.5$$

Normal approximation is ok

$$\mu = np = 25(.5) = 12.5$$

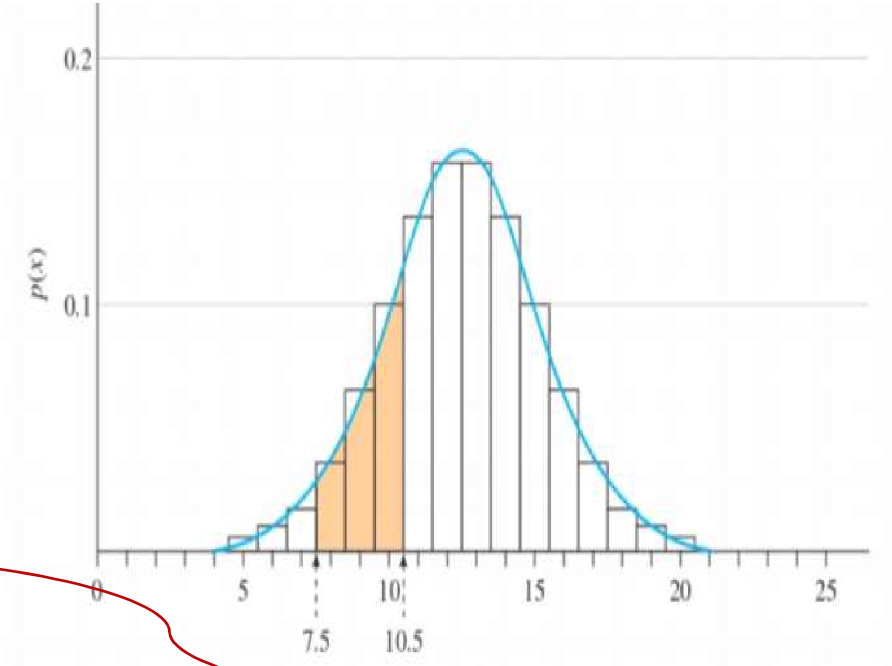
$$\sigma = \sqrt{npq} = \sqrt{25(.5)(.5)} = 2.5$$

$$P(8 \le x \le 10)$$

$$\approx P\left(\frac{7.5 - 12.5}{2.5} \le z \le \frac{10.5 - 12.5}{2.5}\right)$$

$$= P(-2.0 \le z \le -.8)$$

$$= .2119 - .0228 = .1891$$



✓ binomial prob.

$$P(8 \le x \le 10) = P(x \le 10) - P(x \le 7) = .212 - .022 = .190$$

Example

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.

Success = working battery $n = 200$

$p = .95 \qquad np = 190 \qquad nq = 10$

The normal approximation is ok!

$$P(x \geq 195) \approx P\left(z \geq \frac{194.5 - 190}{\sqrt{200(.95)(.05)}}\right)$$

$$= P(z \geq 1.46) = 1 - .9278 = .0722$$

Key Concepts

I. Continuous Probability Distributions

1. Continuous random variables
2. Probability distributions or probability density functions

   a. Curves are smooth.

   b. The area under the curve between a and b represents the probability that x falls between a and b.

   c. $P(x = a) = 0$ for continuous random variables.

II. The Normal Probability Distribution

1. Symmetric about its mean μ.
2. Shape determined by its standard deviation σ.

Key Concepts

III. The Standard Normal Distribution

1. The normal random variable z has mean 0 and standard deviation 1.
2. Any normal random variable x can be transformed to a standard normal random variable using

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

3. Convert necessary values of x to z.
4. Use Table 3 in Appendix I to compute standard normal probabilities.
5. Several important z-values have right-tail areas as follows:

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

Excel 실습

- **NORM.DIST(x,mean,standard_dev,cumulative)**
  : 지정한 평균과 표준편차에 의거 정규 분포 값을 구함
  x : 분포를 구하려는 값
  mean : 분포의 산술 평균
  standard_dev : 분포의 표준 편차
  cumulative : 함수의 폼을 결정하는 논리적 값
  - 누적이 TRUE : 누적
  - FALSE: 확률 밀도 함수

  예) NORM.DIST(0, 0, 1, false) : f(0) 값 when 평균 0, 표준편차 1
  NORM.DIST(0, 0, 1, True) : 0.5 누적 확률

- **NORM.INV(probability,mean,standard_dev)**
  : 주어진 확률에 해당되는 x 의 값
  probability : 정규 분포를 따르는 누적확률
  mean : 분포의 산술 평균
  standard_dev : 표준 편차

예) NORM.INV(0.95, 0, 1) : 1.644854
NORM.INV(0.90, 0, 1) : 1.28
NORM.INV(0.5, 0, 1) : 0