# *Lecture Four* – Statistics

## *Section* 4.1 – *Random Variable & probability*

A ***random variable*** is a function that assigns a numerical value to each simple event in a sample space ***S***.

### The probability distribution for a random variable *X*

The probability distribution for a random variable $X$, denoted $P(X = x) = p(x)$ satisfies

1. $0 \le p(x) \le 1, \quad x \in \left\{x_1,\ x_2,\ \cdots,\ x_n\right\}$
2. $p\left(x_1\right) + p\left(x_2\right) + \ldots + p\left(x_n\right) = 1$

*where* $\left\{x_1,\ x_2,\ \cdots,\ x_n\right\}$ are the (range) values of $X$.

### *Expected value* of the random variable *X*

The expected value of the random variable $X$ is the sum of the products of the values of $X$ and their respective probabilities.

| $x_i$ | $x_1$ | $x_2$ | $\cdots$ | $x_n$ |
|---|---|---|---|---|
| $p_i$ | $p_1$ | $p_2$ | $\cdots$ | $p_n$ |

$$E(X) = x_1 p_1 + x_2 p_2 + \cdots + x_n p_n$$

***Example***

| $x_i$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $p_i$ | .14 | .13 | .18 | .20 | .11 | .24 |

What is the expected value of $X$?

<u>*Solution*</u>

$E(X) = 1(.14) + 2(.13) + 3(.18) + 4(.2) + 5(.11) + 6(.24)$

$= 3.72$

***Example***

A carton of 20 watch batteries contains 2 dead ones. A random sample of 4 is selected from 20 and tested. Let *X* be the random variable associated with the number of dead batteries found in a sample.

*a)* Find the probability distribution of *X*.

*b)* Find the expected number of dead batteries in a sample.

<u>Solution</u>

***a)*** $P(0\ dead\ battery) = \dfrac{C_{18,4}}{C_{20,4}}\ \underline{= .632|}$

$P(1\ dead\ battery) = \dfrac{C_{2,1}C_{18,3}}{C_{20,4}}\ \underline{= .337|}$

$P(2\ dead\ battery) = \dfrac{C_{2,2}C_{18,2}}{C_{20,4}}\ \underline{= .032|}$

| X | 0 | 1 | 2 |
|---|---|---|---|
| P | .632 | .337 | .032 |

$\longrightarrow \approx 1$

***b***) E(*X*) = 0(.632) + 1(.337) + 2(.032) = .4

***Example***

A spinner device is numbered from 0 to 5, and each of the 6 numbers is as likely to come up as any other. A player who bets $1 on any given number wins $4 (gets the bet back) if the pointer comes to rest on the chosen number, otherwise, the $1 bet is lost. What is the expected value of the game?

<u>Solution</u>

S = {0, 1, 2, 3, 4, 5}

$E(X) = 4\left(\frac{1}{6}\right) + (-1)\left(\frac{5}{6}\right) \approx -17$ per game

| *X* | $4 | -$1 |
|---|---|---|
| *P* | $\frac{1}{6}$ | $\frac{5}{6}$ |

The player will lose an average of about 17¢ per game

***Example***

Suppose you are interested in insuring a car stereo system for $500 against theft. An insurance company charges a premium of $60 for coverage 1 year, claiming an empirically determining probability of 0.1 that the stereo will be stolen some time during the year. What is your expected return from the insurance company if you take out this insurance?

<u>Solution</u>

We pay premium $60 to get $500

$\rightarrow$ 500 – 60 = $440 with P = .1

| *X* | $440 | -$60 |
|---|---|---|
| *P* | .1 | .9 |

E(X) = .1(440) + .9(-60) = -$10

# *Exercises* *Section 4.1 – Random Variable & probability*

**1.** Suppose a random sample of 2 light bulbs is selected from a group of 8 bulbs that contain 3 defective bulbs.

*a)* What is the expected value of the number of defective bulbs in the sample?

*b)* Probability Distribution Table

*c)* What is your expected return?

**2.** Suppose 1000 raffle tickets are sold at a price of \$10 each. Two first place tickets will be drawn, 5 second place tickets will be drawn and 10 third place tickets will be drawn. The first place prize is a \$200 VCR, the second place prize is a \$100 printer, and the third place prize is a \$50 gift certificate.

**3.** Find the expected value of each random variable.

*a)*

| $x$ | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| $P(x)$ | 0.1 | 0.4 | 0.3 | 0.2 |

*b)*

| $x$ | 4 | 6 | 8 | 10 |
|---|---|---|---|---|
| $P(x)$ | 0.4 | 0.4 | 0.05 | 0.15 |

*c)*

| $x$ | 9 | 12 | 15 | 18 | 21 |
|---|---|---|---|---|---|
| $P(x)$ | 0.14 | 0.22 | 0.38 | 0.19 | 0.07 |

*d)*

| $x$ | 30 | 32 | 36 | 38 | 44 |
|---|---|---|---|---|---|
| $P(x)$ | 0.31 | 0.29 | 0.26 | 0.09 | 0.05 |

**4.** A delegation of 3 selected from a city council made up of 5 liberals and 6 conservatives.

*a)* What is the expected number of liberals in the delegation?

*b)* What is the expected number of conservatives in the delegation?

**5.** From a group of 3 women and 5 men, a delegation of 2 is selected. Find the expected number of women in the delegation.

**6.** In a club with 20 senior and 10 junior members, what is the expected number of junior members on a 4-member committee?

**7.** If 2 cards are drawn at one time from a deck of 52 cards, what is the expected number of diamonds?

**8.** A local used-car dealer gets complaints about his car as shown in the table below. Find the expected number of complaints per day

| *Number of Complaints per day* | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| *Probability* | 0.02 | 0.06 | 0.16 | 0.25 | 0.32 | 0.13 | 0.06 |

**9.** An insurance company has written 100 policies for \$100,000, 500 policies for \$50,000, and 100 policies for \$10,000 for people of age 20. If experience shows that the probability that a person will die at age 20 is 0.0012, how much can the company expect to pay out during the ear policies were written?

**10.** An insurance policy on an electrical device pays a benefit of $4,000 if the device fails during the first year. The amount of the benefit decreases by $1,000 each successive year until it reaches 0. If the device has not failed by the beginning of any given year, the probability of failure that year is 0.4. What the expected benefit under the policy? (Choose the appropriate answer)

*a.* $2,234 *b.* $2,400 *c.* $2,500 *d.* $2,667 *e.* $2,694

**11.** A tour operator has a bus that can accommodate 20 tourists. The operator knows that tourists may not show up, so he sells 21 tickets. The probability that an individual tourist will not show up is 0.02, independent of all other tourists. Each ticket costs $50, and is non-refundable if a tourist fails to show up. If a tourist shows up and a seat is not available, the tour operator has to pay $100 (ticket cost + $50 penalty) to the tourist. What is the expected revenue of the tour operator? (Choose the appropriate answer)

*a.* $935 *b.* $950 *c.* $967 *d.* $976 *e.* $985

# Section 4.2 – Frequency Distribution; Measure of Central Tendency

## Frequency Distribution

Frequency Distributions are presented as lists ordered by quantity showing the number of times each value appears.

***Example***

Given the data:

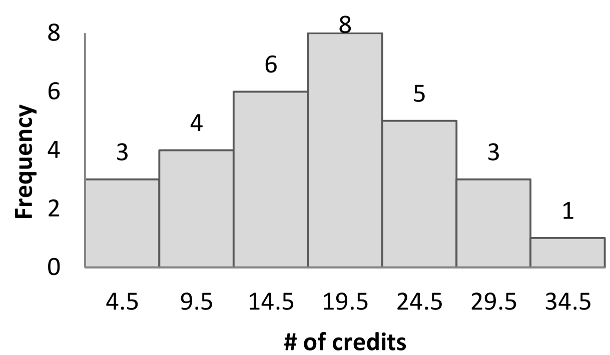
| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 3 | 25 | 33 | 16 | 0 | 9 | 14 | 8 | 34 | 21 |
| 15 | 12 | 9 | 3 | 8 | 15 | 20 | 12 | 28 | 19 |
| 17 | 16 | 23 | 19 | 12 | 14 | 29 | 13 | 24 | 18 |

Group the data into intervals and find the frequency of each interval.

<u>*Solution*</u>

| *Interval* | *Frequency* |
|---|---|
| 0 – 4 | 3 |
| 5 – 9 | 4 |
| 10 – 14 | 6 |
| 15 – 19 | 8 |
| 20 – 24 | 5 |
| 25 – 39 | 3 |
| 30 – 34 | 1 |
| ***Total*** | 30 |

*Histogram*



## *Mean* – Ungrouped Data

The mean is the quantitative data is equal to the sum of all the measurements in the data set divided by the total number of measurements in the set.

$\bar{X}$ : Sample mean

μ : Population mean **($\mu$: mu)**

## Definition

If $x_1, x_2, \ldots, x_n$ is a set of n measurements, then the mean of the set of measurements is given by

$$\left[mean\right] = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

$$\bar{x} = \frac{\sum x}{n}$$

***Example***

The number of bankruptcy petitions (in thousands) filed in the U.S. in the years 2000-2005 are given in the table.

| *Year* | *Petitions* |
|---|---|
| 2000 | 1253 |
| 2001 | 1492 |
| 2002 | 1578 |
| 2003 | 1660 |
| 2004 | 1597 |
| 2005 | 2078 |

Find the mean number of bankruptcy petitions filed annually during this period.

*Solution*

$$\bar{x} = \frac{1253 + 1492 + 1578 + 1660 + 1597 + 2078}{6}$$

$$\approx 1610$$

The mean number is 1,610,000

***Example***

Find the mean for the sample measurements 3.2, 4.5, 2.8, 5.0, and 3.6.

*Solution*

$$\bar{X} = \frac{3.2 + 4.5 + 2.8 + 5 + 3.6}{5} \approx 3.8$$

## *Mean* – Grouped Data

A data set of n measurements is grouped into $k$ classes in a frequency table. If $x_i$ is the midpoint of the $i^{th}$ class interval and $f_i$ is the $i^{th}$ class interval.

The mean for grouped data is given by:

$$[mean] = \frac{\sum_{i=1}^{n} x_i f_i}{n} = \frac{x_1 f_1 + x_2 f_2 + \ldots + x_n f_n}{n}$$

$$n = \sum_{i=1}^{n} f_i \text{ = total number of measurements}$$

***Example***

Find the mean for the data shown in the following frequency distribution

| ***Value (x)*** | ***Frequency (f)*** | ***x . f*** |
|---|---|---|
| 30 | 6 | 30.6 =180 |
| 32 | 9 | 288 |
| 33 | 7 | 231 |
| 37 | 12 | 444 |
| 42 | 6 | 252 |
| Total | 40 | 1395 |

*Solution*

$$\overline{x} = \frac{1395}{40} = 34.875$$

***Example***

Find the mean from the grouped frequency distribution

| *Interval* | *Midpoint, **x*** | *Frequency, **f*** | ***x . f*** |
|---|---|---|---|
| 0 – 4 | 2 | 3 | 6 |
| 5 – 9 | 7 | 4 | 28 |
| 10 – 14 | 12 | 6 | 72 |
| 15 – 19 | 17 | 8 | 136 |
| 20 – 24 | 22 | 5 | 110 |
| 25 – 39 | 27 | 3 | 81 |
| 30 – 34 | 32 | 1 | 32 |
| ***Total*** | | 30 | 465 |

*Solution*

$$\overline{x} = \frac{465}{30} = 15.5$$

## *Median*

*Arrange the data in increasing order.*

1. If the number of measurements in a set is ***odd*** $\Rightarrow$ the **median** is the middle measurement.
2. If the number of measurements in a set is ***even*** $\Rightarrow$ the **median** is the mean of the two middle measurements

| *Odd Number of Entries* | *Even Number of Entries* |
|---|---|
| 8 | 2 |
| 7 | 3 |
| **Median = 4** | $\left.\begin{matrix}4\\7\end{matrix}\right\} Median = \frac{4+7}{2} = 5.5$ |
| 3 | 9 |
| 1 | 11 |

***Example***

Find the median of each list of numbers

*a)* 11, 12, 17, 20 , 23, 28, 29
*b)* 15, 13, 7, 11, 19, 30, 39, 5, 10
*c)* 47, 59, 32, 81, 74, 153

*Solution*

***a)*** ***Median*** = 20

***b)*** $\{5,7,10,11,13,15,19,30,39\}$

***Median*** = 13

***c)*** $\{32,47,59,74,81,153\}$

***Median*** = $\frac{59+74}{2} = 66.5$

***Example***

Find the median and mean salary for:

$17,000, $18,000, $18,000, $20,000, $24,000, $28,000, $100,000, $120,000

*Solution*

$$Median = \frac{20000+24000}{2} = \$22,000$$

$$Mean = \frac{17000+18000+18000+20000+24000+28000+100000+120000}{8} = \$43,125$$

## *Mode*

The mode is the most frequency occurring measurement in a data set. There may be a unique mode, several modes, or, if no measurement occurs more than once, essentially no mode.

*Example*

Find the mode for each list of numbers

*a)* 57, 38, 55, 55, 80, 87, 98, 55, 57

*b)* 182, 185, 183, 184, 187, 187, 185

*c)* 10708, 11519, 10972, 17546, 13905, 12182

<u>*Solution*</u>

- ***a)*** ***Mode***: 55
- ***b)*** ***Mode***: 185; 187
- ***c)*** ***Mode***: No mode

# Exercises Section 4.2 – Frequency Distribution; Measure of Central Tendency

**1.** Find the mean: 9.4, 11.3, 10.5, 7.4, 9.1, 8.4, 9.7, 5.2, 1.1, 4.7

**2.** Find the median: 28.4, 9.1, 3.4, 27.6, 59.8, 32.1, 47.6, 29.8

**3.** Find the mode: 16, 15, 13, 15, 14, 13, 11, 15, 14

**4.** Find the mean and median: 8, 10, 16, 21, 25

**5.** Find the mean and median: 67, 89, 78, 86, 100, 96

**6.** Find the mean and median: 30,200; 23,700; 33,320; 29,410; 24,600; 27,750; 27,300; 32,680

**7.** Find the mean and median: 15.3, 27.2, 14.8, 16.5, 31.8, 40.1, 18.9, 28.4, 26.3, 35.3

**8.** The number of nations participating in the winter Olympic Games, is given below.

| *Year* | *Participating* |
|---|---|
| 1968 | 37 |
| 1972 | 35 |
| 1976 | 37 |
| 1980 | 37 |
| 1984 | 49 |
| 1988 | 57 |
| 1992 | 64 |
| 1994 | 67 |
| 1998 | 72 |
| 2002 | 77 |
| 2006 | 85 |

Find: Mean, Media, and Mode

**9.** Compute the mean for the grouped sample data listed in below table.

| **Class Interval** | **Frequency** |
|---|---|
| 0.5 – 5.5 | 6 |
| 5.5 – 10.5 | 20 |
| 10.5 – 15.5 | 18 |
| 15.5 – 20.5 | 4 |

**10.** Compute the mode(s), median, and mean for each data set:

*a)* 2, 1, 2, 1, 1, 5, 1, 9, 4
*b)* 2, 5, 1, 4, 9, 8, 7
*c)* 8, 2 , 6, 8, 3, 3, 1, 5, 1, 8, 3

**11.** U.S. wheat prices and production figures for a recent decade are given in the following table.

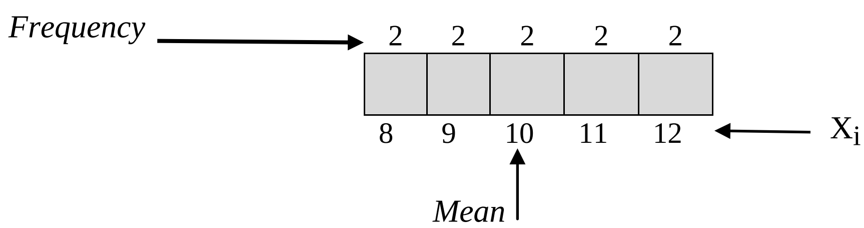
| ***Year*** | ***Price* (*$ per bushel*)** | ***Production* (*millions of bushels*)** |
|---|---|---|
| 1996 | 4.30 | 2277 |
| 1997 | 3.38 | 2481 |
| 1998 | 2.65 | 2547 |
| 1999 | 2.48 | 2296 |
| 2000 | 2.62 | 2228 |
| 2001 | 2.78 | 1947 |
| 2002 | 3.56 | 1606 |
| 2003 | 3.40 | 2345 |
| 2004 | 3.40 | 2158 |
| 2005 | 3.45 | 2105 |

Find the mean and median for the following

*a)* Price per bushel of wheat

*b)* Wheat production

# *Section* 4.3 – Measures of Variation / Dispersion




## *Range*

Range is the difference between the largest and the smallest number in a sample.

***Example***

Find the range

*a)* 12, 27, 6, 19, 38, 9, 42, 15
*b)* 74, 112, 59, 88, 200, 73, 92, 175

<u>*Solution*</u>

***a)*** ***Range*** = 42 – 6 = 36

***b)*** ***Range*** = 200 – 59 = 141

## Deviation from the Mean

Sample Mean $\overline{X} = \frac{\sum x_i}{n}$

The deviation from the mean is the difference between the mean and each number: $\left(x_i - \overline{x}\right)$

## *Variance*

The variance of a sample of $n$ numbers $x_1, x_2, x_3, \ldots, x_n$, with mean $\overline{x}$, is

$$[Variance] = s^2 = \frac{\sum x^2 - n\overline{x}^2}{n-1}$$

## *Standard Deviation*

$$[\text{Standard Deviation}] = s = \sqrt{\frac{\sum x^2 - n\bar{x}^2}{n-1}}$$

*TI-8x*

*Store data: create a list*

*STAT menu* $\rightarrow$ *CALC* $\rightarrow$ ***1: 1-Var Stats***

***Example***

Find the standard deviation of the numbers: 7, 9, 18, 22, 27, 29, 32, 40.

<u>*Solution*</u>

$$Mean = \frac{7+9+18+22+27+29+32+40}{8}$$

$$\underline{= 23}$$

| Number, $x$ | Square of the Number, $x^2$ |
|---|---|
| 7 | 49 |
| 9 | 81 |
| 18 | 324 |
| 22 | 484 |
| 27 | 729 |
| 29 | 841 |
| 32 | 1024 |
| 40 | 1600 |
| | Total: 5132 |

$$s = \sqrt{\frac{\sum x^2 - n\bar{x}^2}{n-1}}$$

$$= \sqrt{\frac{5132 - 8(23)^2}{8-1}}$$

$$\underline{\approx 11.34}$$

## Standard Deviation for a Grouped Distribution

$$\textit{Standard Deviation: } s = \sqrt{\frac{\sum fx^2 - n\bar{x}^2}{n-1}}$$

***Example***

Find ***s*** for the grouped

| *Interval* | *Frequency* |
|---|---|
| 0 – 4 | 3 |
| 5 – 9 | 4 |
| 10 – 14 | 6 |
| 15 – 19 | 8 |
| 20 – 24 | 5 |
| 25 – 39 | 3 |
| 30 – 34 | 1 |
| ***Total*** | 30 |

<u>*Solution*</u>

| *Interval* | ***x*** | $x^2$ | ***f*** | $fx^2$ |
|---|---|---|---|---|
| 0 – 4 | 2 | 4 | 3 | 12 |
| 5 – 9 | 7 | 49 | 4 | 196 |
| 10 – 14 | 12 | 144 | 6 | 864 |
| 15 – 19 | 17 | 289 | 8 | 2312 |
| 20 – 24 | 22 | 484 | 5 | 2420 |
| 25 – 39 | 27 | 729 | 3 | 2187 |
| 30 – 34 | 32 | 1024 | 1 | 1024 |
| ***Total*** | | | 30 | 9015 |

$$s = \sqrt{\frac{\sum fx^2 - n\bar{x}^2}{n-1}}$$

$$= \sqrt{\frac{9015 - 30(15.5)^2}{30-1}}$$

$$\approx 7.89$$

***TI-8x***

*Store data: create 2 lists*

*STAT menu* $\rightarrow$ *CALC* $\rightarrow$ ***1: 1-Var Stats L1, L2***

***Example***

Since 1996, Nathan's Famous Hot Dogs has held an annual hot dog eating contest, in which each contest attempts to consume as many hot dogs with burns as possible in a 12-minute period.

| Year | Winner | Hot Dogs Eaten |
|---|---|---|
| 1997 | Hirofumi Nakajima | 24.5 |
| 1998 | Hirofumi Nakajima | 19 |
| 1999 | Steve Keiner | 20.25 |
| 2000 | Kazutoyo Arai | 25.125 |
| 2001 | Takeru Kobayashi | 50 |
| 2002 | Takeru Kobayashi | 50.5 |
| 2003 | Takeru Kobayashi | 44.5 |
| 2004 | Takeru Kobayashi | 53.5 |
| 2005 | Takeru Kobayashi | 49 |
| 2006 | Takeru Kobayashi | 53.75 |
| 2007 | Joey Chestnut | 66 |

In what percent of the contests did the number of hot dogs eaten by the winner fall within one standard deviation of the mean number of hot dogs?

*Solution*

$$\bar{x} = \frac{24.5+19+20.25+25.125+50+50.5+44.5+53.5+49+53.75+66}{11}$$

$$\approx 41.47$$

$$s \approx 16.21$$

***Lower***: $\bar{x} - s = 41.47 - 16.21 = 25.26$

***Upper***: $\bar{x} + s = 41.47 + 16.21 = 57.68$

$\therefore$ 6 out of 11 (node: between the lower and upper) $\rightarrow$ = 55%

# *Exercises* *Section* 4.3 – Measures of Variation / Dispersion

**1.** Find the range and standard deviation for: $\{3, 7, 4, 12, 15, 18, 19, 27, 24, 11\}$

**2.** Find the range and standard deviation for: $S = \{1.2, 1.4, 1.7, 1.3, 1.5\}$

**3.** Find the range and standard deviation for: 72, 61, 57, 83, 52, 66, 85

**4.** Find the range and standard deviation for: 241, 248, 251, 257, 252, 287

**5.** Find the range and standard deviation for: 122, 132, 141, 158, 162, 169, 180

**6.** Find the standard deviation for the following data

| *Interval* | *Frequency* |
|---|---|
| 30 – 39 | 1 |
| 40 – 49 | 6 |
| 50 – 59 | 13 |
| 60 – 69 | 22 |
| 70 – 79 | 17 |
| 80 – 89 | 13 |
| 90 – 99 | 8 |

**7.** Find the standard deviation for the following data

| *Interval* | *Frequency* |
|---|---|
| 0 – 24 | 4 |
| 25 – 49 | 8 |
| 50 – 74 | 5 |
| 75 – 99 | 10 |
| 100 – 124 | 4 |
| 125 – 149 | 5 |

**8.** Forever Power Company analysis conducted tests on the life of its batteries and those of a competitor (Brand X). They found that their batteries have a mean life (in hours) of 26.2, with a standard deviation of 4.1. Their results for a sample of 10 Brand X were as follows: 15, 18, 19, 23, 25, 25, 28, 30, 34, 38.

*a)* Find the mean and standard deviation for the sample of Brand X batteries.

*b)* Which batteries have a more uniform life in hours?

*c)* Which batteries have the highest average life in hours?

# Section 4.4 – Bernoulli Trials & Binomial Distributions

## *Bernoulli Trials* (Jacob 1654-1705)

- show *or* not, False *or* True
  E *or* E′ $\Rightarrow$ 2 possible outcomes $\Rightarrow$ Bernoulli Experiment *or* trial
  Success ($S$) or Failure ($F$)

Probability of success: P(S) = $p$

Probability of Failure: P(F) = $1 - p = q$ $\qquad \Rightarrow\ p + q = 1$

***Example***

Find $p$ and $q$ for a single roll of a fair die, where a success is a number divisible by 3 turning up
Divisible by 3: {3, 6}

<u>*Solution*</u>

$$p = \frac{2}{6} = \frac{1}{3} \Rightarrow q = 1 - p = 1 - \frac{1}{3} = \frac{2}{3}$$

## Bernoulli Trials

1. Only 2 outcomes are possible or each trial
2. Probability of success = $p$ and P(*Failure*) = $q \rightarrow p + q = 1$
3. All trials are independent

***Example***

If we roll a fair die five times and identify a success in a single roll with a 1 turning up, what is the probability of the sequence FSSSF occurring?

<u>*Solution*</u>

$$P(FSSSF) = p^3q^2 = \left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2 \approx 0.003$$

*Outcome*: 3 out of 5 $\Rightarrow C_{5,3} = 10$

$$P(exactly\ 3\ success) = C_{5,3}\,p^3q^2$$

## Probability of *x* success in *n* Bernoulli Trials

The probability of exactly $x$ success in $n$ independent repeated Bernoulli trials, with the probability of success of each trial $p$ (and of failure $q$), is

$$P(x\ success) = C_{n,x}\, p^{x} q^{n-x}$$

***Example***

If a fair die is rolled five times, what is the probability of rolling

*a)* Exactly one 3?

*b)* At least one 3?

<u>*Solution*</u>

***a)*** Exactly one 3?

$$P(exactly\ 1\text{-}3's \to x = 1) = C_{5,1}\left(\frac{1}{6}\right)^{1}\left(\frac{5}{6}\right)^{4}$$

$$\approx .402$$

***b)*** At least one 3?

$$\begin{aligned} P(x \ge 1) &= P(x=1) + P(x=2) + \cdots + P(x=5) \\ &= 1 - P(x<1) \\ &= 1 - P(x=0) \\ &\approx 1 - .402 \\ &\approx .598 \end{aligned}$$

## **Binomial Formula**: ***Brief Review***

$$(a+b)^1 = a+b$$
$$(a+b)^2 = a^2 + 2ab + b^2$$
$$(a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$$

$$(a+b)^n = C_{n,0}a^n + C_{n,1}a^{n-1}b + C_{n,2}a^{n-2}b^2 + \cdots + C_{n,n}b^n$$

## **Binomial Distribution**

*In 3 Trials*

| ***Simple Event*** | ***Pr of E*** | ***$X_3$*** | |
|---|---|---|---|
| FFF | $q^3$ | 0 | $q^3$ |
| FFS | $q^2p$ | 1 | |
| FSF | $q^2p$ | 1 | $3q^2p$ |
| SFF | $q^2p$ | 1 | |
| FSS | $p^2q$ | 2 | |
| SFS | $p^2q$ | 2 | $3p^2q$ |
| SSF | $p^2q$ | 2 | |
| SSS | $p^3$ | 3 | $p^3$ |
| | | | $(p+q)^3$ |

1- $0 \le P(X_3 = x) \le 1 \qquad \therefore X \in \{0,1,2,3\}$

2- $1 = 1^3 = (p+q)^3 = C_{3,0}q^3 + C_{3,1}q^2p + C_{3,2}qp^2 + C_{3,3}p^3$ **1 = *p* + *q***

$$= q^3 + 3q^2p + 3qp^2 + p^3$$
$$= P(X_3 = 0) + P(X_3 = 1) + P(X_3 = 2) + P(X_3 = 3)$$

## *Binomial Distribution*

$$\Rightarrow P(X_n = x) = P(x \text{ } success \text{ in } n \text{ } trials) = C_{n,x} p^x q^{n-x}$$

***Example***

Suppose a fair die is rolled two times and a success on a single roll is considered to be rolling a number divisible by 3.

*a)* Write the function defining the distribution

*b)* Construct a table for the distribution

*c)* Construct a histogram for the distribution

<u>Solution</u>

***a)*** 3 & 6 are divisible by 3 $\Rightarrow p = \frac{2}{6} = \frac{1}{3}$

$p = \frac{1}{3}, \quad q = \frac{2}{3}$

$(Rolled \text{ } twice) \rightarrow n = 2$

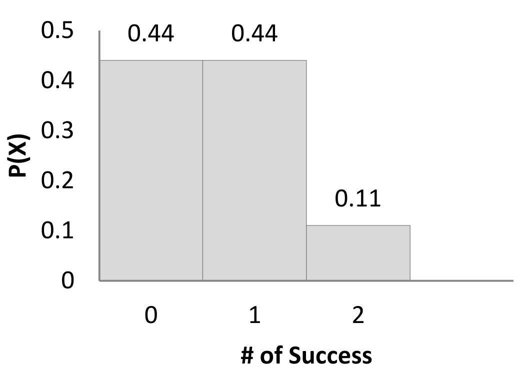
Probability function for the binomial distribution

$$P(X) = P(x \text{ } success \text{ } in \text{ } 2 \text{ } trials) = C_{2,x} \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{2-x} \qquad x \in \{0,1,2\}$$

***b)*** Distribution Table

| $x$ | P($x$) |
|---|---|
| 0 | $C_{2,0}\left(\frac{1}{3}\right)^0\left(\frac{2}{3}\right)^{2-0} = \frac{4}{9} = .44$ |
| 1 | $C_{2,1}\left(\frac{1}{3}\right)^1\left(\frac{2}{3}\right)^{2-1} = 2\frac{1}{3}\frac{2}{3} = .44$ |
| 2 | $C_{2,2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^{2-2} = .11$ |

***c)*** Histogram



## Expected Value

The expected value is denoted by: $E(X) = x_1 p_1 + x_2 p_2 + \cdots + x_n p_n$

E: called Mean of Random Variable X

Standard Deviation: $\sigma = \sqrt{(x_1 - \mu)^2 p_1 + (x2 - \mu)^2 p_2 + \cdots}$

Mean: $\mu = np$

Standard Deviation: $\sigma = \sqrt{npq}$

***Example***

Suppose a fair die is rolled two times and a success on a single roll is considered to be rolling a number divisible by 3. Compute the mean and standard deviation

*<u>Solution</u>*

$n = 2 \rightarrow p = \frac{1}{3}, q = \frac{2}{3}$

$\mu = np = 2\frac{1}{3} = \frac{2}{3} \approx .67$

$\sigma = \sqrt{npq} = \sqrt{2\frac{1}{3}\frac{2}{3}} \approx .67$

***Example***

The probability of recovering after a particular type of operation is 0.5. Let us investigate the binomial distribution involving four patients undergoing this operation

*a)* Write the function defining the distribution
*b)* Construct a table for the distribution
*c)* Construct a histogram for the distribution

<u>Solution</u>

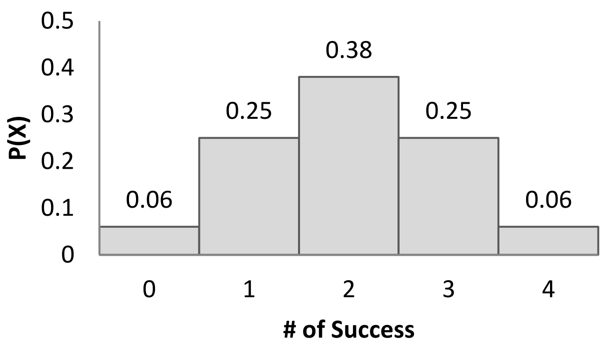
$$n = 4 \;\rightarrow p = .5, \;\; q = .5$$

***a)*** $P(X) = C_{4,x}(.5)^x(.5)^{4-x} = C_{4,x}(.5)^4$

***b)*** Distribution table

| ***x*** | **P(*x*)** | |
|---|---|---|
| 0 | .06 | $P(x=0) = C_{4,0}(.5)^0(.5)^{4-0}$ |
| 1 | .25 | $P(x=1) = C_{4,1}(.5)^1(.5)^{4-1}$ |
| 2 | .38 | $P(x=2) = C_{4,2}(.5)^2(.5)^{4-2}$ |
| 3 | .25 | $P(x=3) = C_{4,3}(.5)^3(.5)^{4-3}$ |
| 4 | .06 | $P(x=4) = C_{4,4}(.5)^4(.5)^{4-4}$ |

***c)*** Histogram




***d)*** $\mu = np = 4(.5) = 2$

$\sigma = \sqrt{npq} = \sqrt{4(.5)(.5)} = 1$

# *Exercises* *Section 4.4 – Bernoulli Trials & Binomial Distributions*

**1.** If a baseball player has a batting average of 0.350, what is the probability that the player will get the following number of hits in the next four times at bat?

*a*) Exactly 2 hits

*b*) At least 2 hits.

**2.** If a true-false test with 10 questions is given, what is the probability of scoring

*a)* Exactly 70% just by guessing?

*b)* 70% or better just by guessing?

**3.** If 60% of the electorate supports the mayor, what is the probability that in a random sample of 10 voters, fewer than half support her?

**4.** Each year a company selects a number of employees for a management training program given by nearby university. On the average, 70% of those sent complete the program. Out of 7 people sent by the company, what is the probability that

*a*) Exactly 5 complete the program?

*b*) 5 or more complete the program?

**5.** If the probability of a new employee in a fast-food chain still being with the company at the end of 1 year is 0.6, what is the probability that out of 8 newly hired people?

*a*) 5 will still be with the company after 1 year?

*b*) 5 or more will still be with the company after 1 year?

**6.** A manufacturing process produces, on the average, 6 defective items out of 100. To control quality, each day a sample of 10 completed items is selected at random and inspected. If the sample produces more than 2 defective items, then the whole day's output is inspected and the manufacturing process is reviewed. What is the probability of this happening, assuming that the process is still producing 6% defective items?

**7.** A manufacturing process produces, on the average, 3% defective items. The company ships 10 items in each box and wishes to guarantee no more than 1 defective item per box. If this guarantee accompanies each box, what is the probability that the box will fail to satisfy the guarantee?

**8.** A manufacturing process produces, on the average, 5 defective items out of 100. To control quality, each day a random sample of 6 completed items is selected and inspected. If a success on a single trial (inspection of 1 item) is finding the item defective, then the inspection of each of the 6 items in the sample constitutes a binomial experiment, which has a binomial distribution.

*a)* Write the function defining the distribution

*b)* Construct a table and histogram for the distribution.

*c)* Compute the mean and standard deviation.

**9.** Each year a company selects 5 employees for a management training program given at a nearly university. On the average, 40% of those sent complete the course in the top 10% of their class. If we consider an employee finishing in the top 10% of the class a success in a binomial experiment, then for the 5 employee entering the program there exists a binomial distribution involving P(x success out of 5).

*a)* Write the function defining the distribution
*b)* Construct a table and histogram for the distribution.
*c)* Compute the mean and standard deviation.

**10.** A person with tuberculosis is given a chest *x*-ray. Four tuberculosis *x*-ray specialists examine each *x*-ray independently. If each specialist can detect tuberculosis 80% of the time when it is present, what is the probability that at least 1 of the specialists will detect tuberculosis in this person?

**11.** A pharmaceutical laboratory claims that a drug it produces causes serious side effects in 20 people out of 1,000 on the average. To check this claim, a hospital administers the drug to 10 randomly chosen patients and finds that 3 suffer from serious side effects. If the laboratory's claims are correct, what is the probability of the hospital obtaining these results?

**12.** The probability that brown-eyed parents, both with the recessive gene for blue, will have a child with brown eye is .75. If such parents have 5 children, what is the probability that they will have

*a*) All blue-eyed children?
*b*) Exactly 3 children with brown eyes?
*c*) At least 3 children with brown eyes?

**13.** The probability of gene mutation under a given level of radiation is $3\times10^{-5}$. What is the probability of the occurrence of at least 1 gene mutation if $10^5$ genes are expected to this level of radiation?

**14.** If the probability of a person contracting influenza on exposure is .6 consider the binomial distribution for a family of 6 that has been exposed.

*a)* Write the function defining the distribution.
*b)* Compute the mean and standard deviation.

**15.** The probability that a given drug will produce a serious side effect in a person using the drug is .02. In the binomial distribution for 450 people using the drug, what are the mean and standard deviation?

**16.** An opinion poll based on a small sample can be unrepresentative of the population. To see why, let us assume that 40% of the electorate favors a certain candidate. If a random sample of 7 is asked their preference, what is the probability that a majority will favor this candidate?

**17.** A multiple choice test is given with 5 choices only one is correct, for each of 5 questions. Answering each of the 5 questions by guessing constitutes a binomial experiment with an associated binomial distribution

*a)* Write the function defining the distribution.

*b)* Compute the mean and standard deviation.

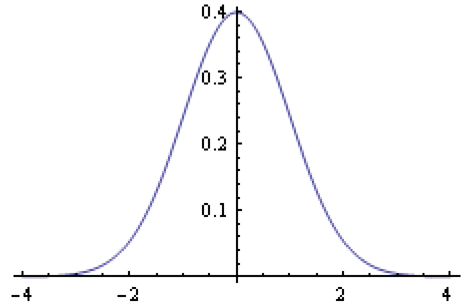
**18.** Suppose a die is rolled 4 times.

*a)* Find the probability distribution for the number of times 1 is rolled.

*b)* What is the expected number of times 1 is rolled

# *Section 4.5 – Normal Distribution*

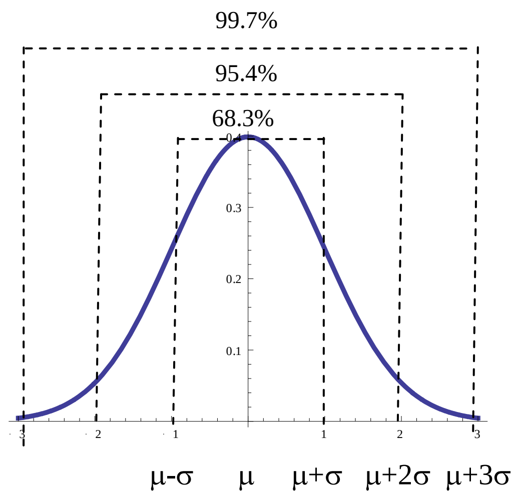
***Braham De Moivre*** (1667 – 1754), *Laplace*, *Gauss*




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

μ: Mean

σ: Standard deviation




***Properties***:

1. Symmetrical
2. $\mu$ Mean @ axis of Symmetry
3. Shape determined by $\mu$ & σ

## Area under Normal Curves

Area between μ & σ: μ → μ + σ *or* μ - σ (are the same)

To Find how many standard deviations a measurement $x$ is from a mean μ, first, determine the distance between $x$ and μ then divide by σ.

$z-scores$

$$z = \frac{\text{distance between } x \text{ and } \mu}{\text{standard deviation}} = \frac{x-\mu}{\sigma}$$

### *Properties*

$P(a \le x \le b)$ = Area under the normal curve from $a \to b$

$P(-\infty < x < \infty) = 1$ = Total Area

$P(x = c) = 0$

***Example***

A manufacturing process produces light bulbs with life expectancies that are normally distributed with a mean of 500 hours and a standard deviation of 100 hours. What *percentage* of the light bulbs can be expected to last 500 to 750 hours?

<u>*Solution*</u>

$\mu = 500,\ \sigma = 100$

$\Rightarrow 500\ \&\ 750$

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

$$= \frac{750-500}{100}$$

$$= 2.5$$

$$\Rightarrow \boxed{A = 49.38\%}$$

***Example***

What is the *probability* of the light bulbs can be expected to last 400 to 500 hours?

<u>*Solution*</u>

$400 \to 500$

$$z = \frac{400-500}{100}$$

$$= -1$$

$$\Rightarrow \boxed{A = .3413}$$

***Example***

If a normal distribution has mean 50 and standard deviation 4, find the following $z-scores$.

*a)* The $z-scores$ for $x = 46$

*b)* The $z-scores$ for $x = 60$

<u>*Solution*</u>

***a)*** $z = \dfrac{x-\mu}{\sigma}$

$$= \frac{46-50}{4}$$

$$\underline{= -1}$$

***b)*** $z = \dfrac{60-50}{4}$

$$\underline{= 2.5}$$

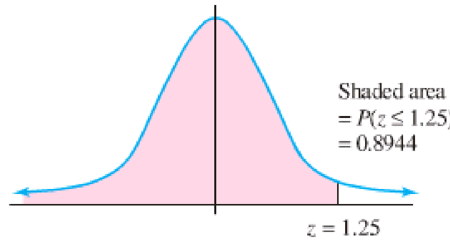
The area of the shaded region under the normal curve from *a* to *b* is the ***probability*** that an observed data value will be between *a* to *b*.

***Example***

Find the areas under the standard normal curve

*a)* The area to the ***left*** of $z = 1.25$

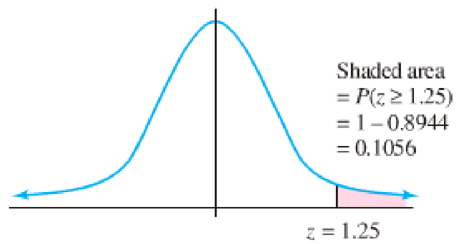
$$A = P(z \le 1.25) \qquad \textbf{(\textit{Using left curve table})}$$
$$= 0.8944$$




$$A = P(z \le 1.25) \qquad \textbf{(\textit{Using right curve table})}$$
$$= .5 + 0.3944$$
$$= 0.8944$$

*b)* The area to the ***right*** of $z = 1.25$

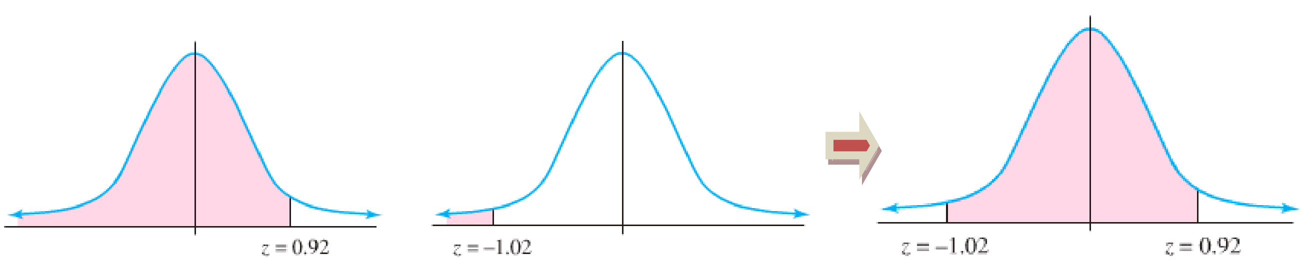
$$A = P(z \ge 1.25) \qquad \textbf{(\textit{Using left curve table})}$$
$$= 1 - 0.8944$$
$$= 0.1056$$




$$A = P(z \ge 1.25) \qquad \textbf{(\textit{Using right curve table})}$$
$$= .5 - 0.3944$$
$$= 0.1056$$

*c)* The area ***between*** $z = -1.02$ and $z = 0.92$

$$A = P(-1.02 \le z \le 0.92) \qquad \textbf{(\textit{Using left curve table})}$$
$$= 0.8212 - 0.1539$$
$$= 0.6673$$




$$A = P(-1.02 \le z \le 0.92) \qquad (\textit{Using right curve table})$$
$$= 0.3461 + 0.3212$$
$$= 0.6673$$

***Example***

Find a value of *z* satisfying the following conditions

*a)* 12.1% of the area is to the left of *z*.

*b)* 20% of the area is to the right of *z*.

<u>Solution</u>

***a)*** $A = 0.121 \quad \Rightarrow \quad \boxed{z = -1.17}$

***b)*** $A = 1 - .2 = .8 \quad \Rightarrow \quad \boxed{z = 0.84}$

***Example***

Dixie Office Supplies finds that it sales force drives an average of 1200 miles per month per person, with a standard deviation of 150 miles. Assume that the number of miles driven by a salesperson is closely approximated by a normal distribution.

*a)* Find the probability that a salesperson drives between 1200 miles and 1600 miles per month

*b)* Find the probability that a salesperson drives between 1000 miles and 1500 miles per month

*c)* Find the shortest and longest distances driven by the middle 95% of the data.

<u>Solution</u>

***a)*** For $x_1 = 1200$

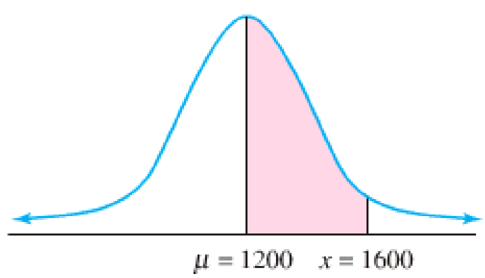
$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{1200 - 1200}{150} = 0$$

$\rightarrow A_1 = 0.5000$

For $x_2 = 1600$

$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{1600 - 1200}{150} \approx 2.67$$

$\rightarrow A_2 = 0.9962$




$A = 0.9962 - 0.5000 = 0.4962$

$P(1200 \le z \le 1600) \underline{= 0.4962}$

***b)*** For $x_1 = 1000$

$$z_1 = \frac{1000 - 1200}{150} \approx -1.33$$

$\rightarrow A_1 = 0.0918$

For $x_2 = 1500$

$$z_2 = \frac{1500 - 1200}{150} \approx 2.00$$

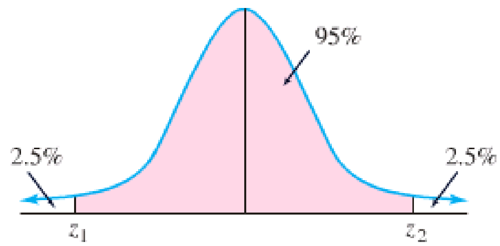
$\rightarrow A_2 = 0.9772$

$A = 0.9772 - 0.0918 = 0.8854$

$P(1000 \le z \le 1500) \underline{= 0.8854}$

***c)*** The lower z value has 2.5% $\Rightarrow z_1 = -1.96$

The higher z value has 97.5% $\Rightarrow z_2 = 1.96$

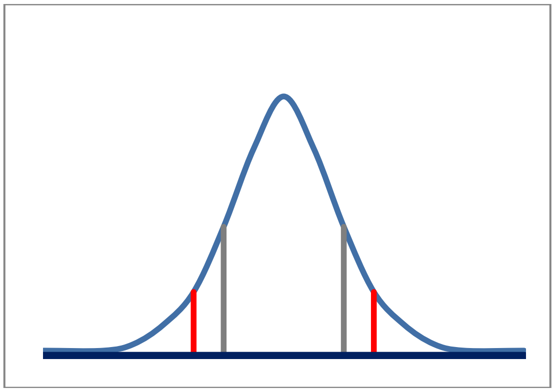


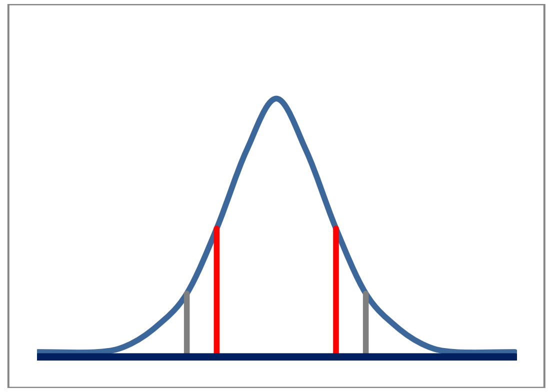

$$\begin{aligned}\text{Shortest distance } &= \mu + z\sigma \\ &= 1200 + (-1.96)(150) \\ &= 906 \textit{ miles}\end{aligned}$$

$$\begin{aligned}\text{Longest distance } &= \mu + z\sigma \\ &= 1200 + (1.96)(150) \\ &= 1494 \textit{ miles}\end{aligned}$$

The distances driven between by the middle 95% of the sales force are between 906 and 1494 miles.

*Always Add (0.5) to the line limit of the area that you are trying to find.*





***Example***

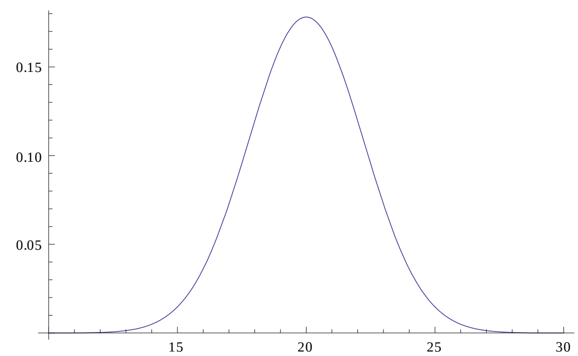
A company manufactures 50,000 ballpoint pens each day. The manufacturing process produces 40 defective pens per 1,000 on the average. A random sample of 400 pens is selected from each day's production and tested. What is the probability that the sample contains

*a)* At least 10 and no more than 20 defective pens?

*b)* 27 or more defective pens?

<u>*Solution*</u>

***a)*** 40 defective pens per 1,000




$$\Rightarrow\ p = \frac{40}{1000} = 0.04$$

$$\mu = np$$
$$= 400(0.04)$$
$$= 16$$

$$\sigma = \sqrt{npq}$$
$$= \sqrt{400(.04)(.96)}$$
$$= 3.92$$

$$z_1 = \frac{x-\mu}{\sigma} = \frac{9.5-16}{3.92} = -1.66 \ \Rightarrow A_1 = .4515$$

$$z_2 = \frac{x-\mu}{\sigma} = \frac{20.5-16}{3.92} = 1.15 \ \Rightarrow A_2 = .3749$$

$$\Rightarrow A = A_1 + A_2 = .4515 + .3749 \approx .8264$$

***b)*** $z = \dfrac{26.5-16}{3.92} \approx 2.68 \Rightarrow A_1 = .4963$

$$\Rightarrow A = 0.5 - A_1 = .5 - .4963 \approx .0037$$

# *Exercises* *Section 4.5 – Normal Distribution*

**1.** A manufacturing process produces light bulbs with life expectancies that are normally distributed with a mean of 500 hours and a standard deviation of 100 hours. What percentage of the light bulbs can be expected to last 500 to 750 hours?

**2.** What is the probability of the light bulbs can be expected to last 400 to 500 hours?

**3.** The average lifetime for a car battery of a certain brand is 170 weeks, with a standard deviation of 10 weeks. If the company guarantees the battery for 3 years, what percentage of the batteries sold would be expected to be returned before the end of the warranty period? Assume a normal distribution

**4.** A manufacturing process produces a critical part of average length 100 millimeters, with a standard deviation of 2 millimeters. All parts deviating by more than 5 millimeters from the mean must be rejected. What percentage of the parts must be rejected, on the average? Assume a normal distribution.

**5.** An automated manufacturing process produces a component with an average width of 7.55 cm, with a standard deviation of 0.02 cm. All components deviating by more than 0.05 cm from the mean must be rejected. What percent of the parts must be rejected, on the average? Assume a normal distribution.

***6.*** A company claims that 60% of the households in a given community uses its product. A competitor surveys the community, using a random sample of 40 households, and finds only 15 households out of the 40 in the sample using the product. If the company's claim is correct, what is the probability of 15 or fewer households using the product in a sample of 40? Conclusion? Approximate a binomial distribution with a normal distribution**.**

**7.** A union representative claims 60% of the union membership will vote in favor of a particular settlement. A random sample of 100 members is polled, and out of these, 47 favor the settlement. What is the approximate probability of 47 or fewer in a sample of 100 favoring the settlement when 60% of all the membership favor the settlement? Conclusion? Approximate a binomial distribution with a normal distribution.

**8.** The average healing time of a certain type of incision is 240 hours, with standard deviation of 20 hours. What percentage of the people having this incision would heal in 8 days or less? Assume a normal distribution.

**9.** The average height of a hay crop is 38 inches, with a standard deviation of 1.5 inches. What percentage of the crop will be 40 inches or more? Assume a normal distribution

**10.** In a family with 2 children, the probability that both children are girls is approximately .25. In a random sample of 1,000 families with 2 children, what is the approximate probability that 220 or fewer will have 2 girls? Approximate a binomial distribution with a normal distribution.

**11.** Aptitude Tests are scaled so that the mean score is 500 and the standard deviation is 100. What percentage of the students taking this test should score 700 or more? Assume a normal distribution

**12.** Candidate Harkins claims a private poll shows that she will receive 52% of the vote for governor. Her opponent, Mankey, secures the services of another pollster, who finds that 470 out of a random sample of 1,000 registered voters favor Harkins. If Harkin's claim is correct, what is the probability that only 470 or fewer will favor her in a random sample of 1,000? Conclusion? Approximate a binomial distribution with a normal distribution.

**13.** An instructor grades on a curve by assuming the grades on a test are normally distributed. If the average grade is 70 and the standard deviation is 8, find the test scores for each grade interval if the instructor wishes to assign grades as follow: 10% A's, 20% B's, 40% C's, 20% D's, and 10% F's.

**14.** At the discount Market, the average weekly grocery bill is $74.50, with a standard deviation of $24.30. What are the largest and smallest amounts spent by the middle 50% of this market's customers?

**15.** A certain type of light bulb has an average life of 500 hours, with a standard deviation of 100 hours. The length of life of the bulb can be closely approximated by a normal curve. An amusement park buys and installs 10,000 such bulbs. Find the total number that can be expected to last for each period of time. Find the shortest and longest lengths of life for the middle 60% of the bulbs.

**16.** A machine that fills quart milk cartons is set up to average 32.2 oz. per carton, with a standard deviation of 1.2 oz. What is the probability that a filled carton will contain less than 32 oz. of milk?

**17.** A machine produces bolts with an average diameter of 0.25 in. and a standard deviation of 0.02 in. What is the probability that a bolt will be produced with a diameter greater than 0.3 in.?