Section 15-1

Presenting Statistical Data

Objectives

- to draw a stem and leaf plot for a distribution of data
- to find the mean, median and mode for a distribution of data
- to use a frequency distribution table for data
- to read a histogram for a distribution of data

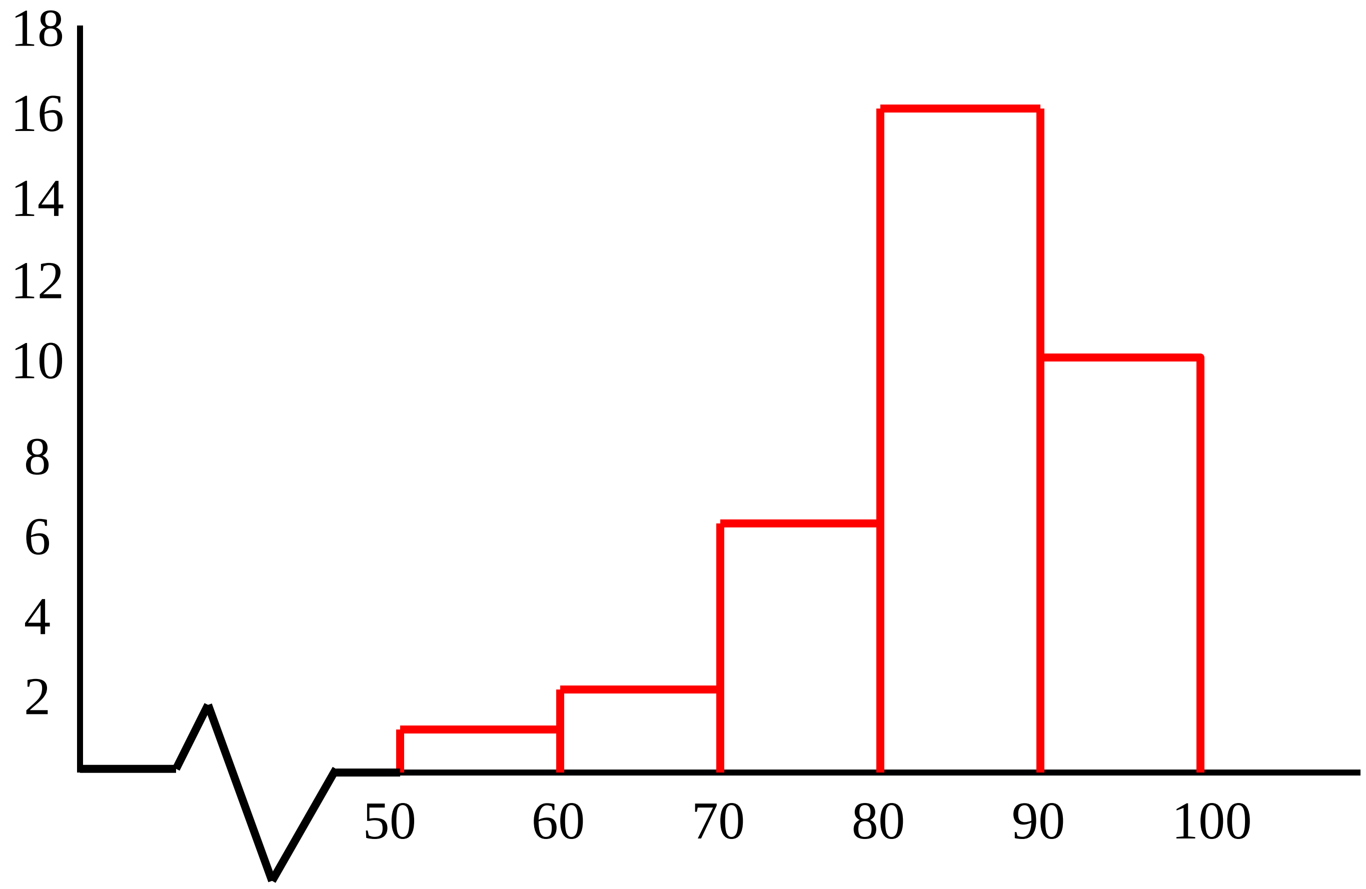
Presenting Statistical Data

- frequency distribution: a table showing each element in a data set and the frequency with which it occurred

score	#	score	#
99	2	82	1
98	1	80	2
95	3	78	1
94	1	75	2
93	1	74	1
90	2	73	2
88	2	68	1
87	3	65	1
84	5	56	1
83	3		

Presenting Statistical Data

- histogram: a graphical representation of the information contained in a frequency distribution



Presenting Statistical Data

- stem-and-leaf plot: a diagram that displays the data in a distribution, the stem is information common to many elements in the distribution

stem	leaf
5	6
6	5, 8
7	3, 3, 4, 5, 5, 8
8	0, 0, 2, 3, 3, 3, 4, 4, 4, 4, 4, 7, 7, 7, 8, 8
9	0, 0, 3, 4, 5, 5, 5, 8, 9, 9

Statistics

- statistics: numbers used to describe a data set
- measures of central tendency
  - mode: number that occurs most frequently
  - median: the middle number when the elements of a distribution are put in numerical order; if there is an even number of elements then the median is the average of the middle two
  - mean; the arithmetic average of the data set

Example for 1-4

16, 54, 23, 38, 22, 22, 40, 46, 52, 19, 20

stem	leaf
1	6, 9
2	0, 2, 2, 3
3	8
4	0, 6
5	2, 4

Example for 5-16

16, 54, 23, 38, 22, 22, 40, 46, 52, 19, 20

16, 19, 20, **22**, **22**, 23, 38, 40, 46, 52, 54

mode = 22

16, 19, 20, 22, 22, **23**, 38, 40, 46, 52, 54

median = 23

$$\frac{16+19+20+22+22+23+38+40+46+52+54}{11}$$

$$\frac{352}{11}$$

mean = 32

Section 15-2

Analyzing Statistical Data

Objectives

- to find median, 1st quartile and 3rd quartile of a distribution
- to make and use box-and-whisker plots
- to find the mean, standard deviation and variance of a distribution

Divisions of Data

- 1st quartile: the median of the lower half of a distribution
- 3rd quartile: the median of the upper half of a distribution
- box-and-whisker plot: a diagram of the divisions of the data
- measures of dispersion
  - range; the difference between the highest element and the lowest element in a data set
  - variance: the average of the squares of the deviations from the mean

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{n}$$

  - standard deviation: the principal square root of the variance

$$\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{n}}$$

Example for 1&2

0	2, 3, 7
1	0, 4, 4, 8
2	2, 5, 5, 5, 9, 9
3	0, 4, 4, 8, 8
4	0, 1, 7

median = 25

1st quartile = 14

3rd quartile = 34

range = 47 - 2 **= 45**

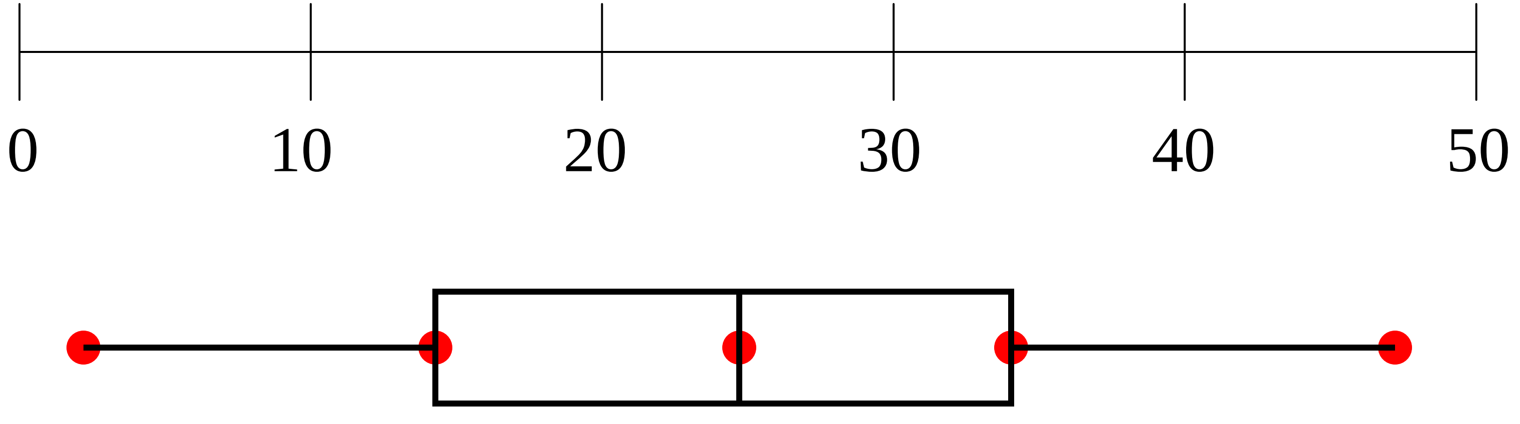
Example for 3&4

median = 25

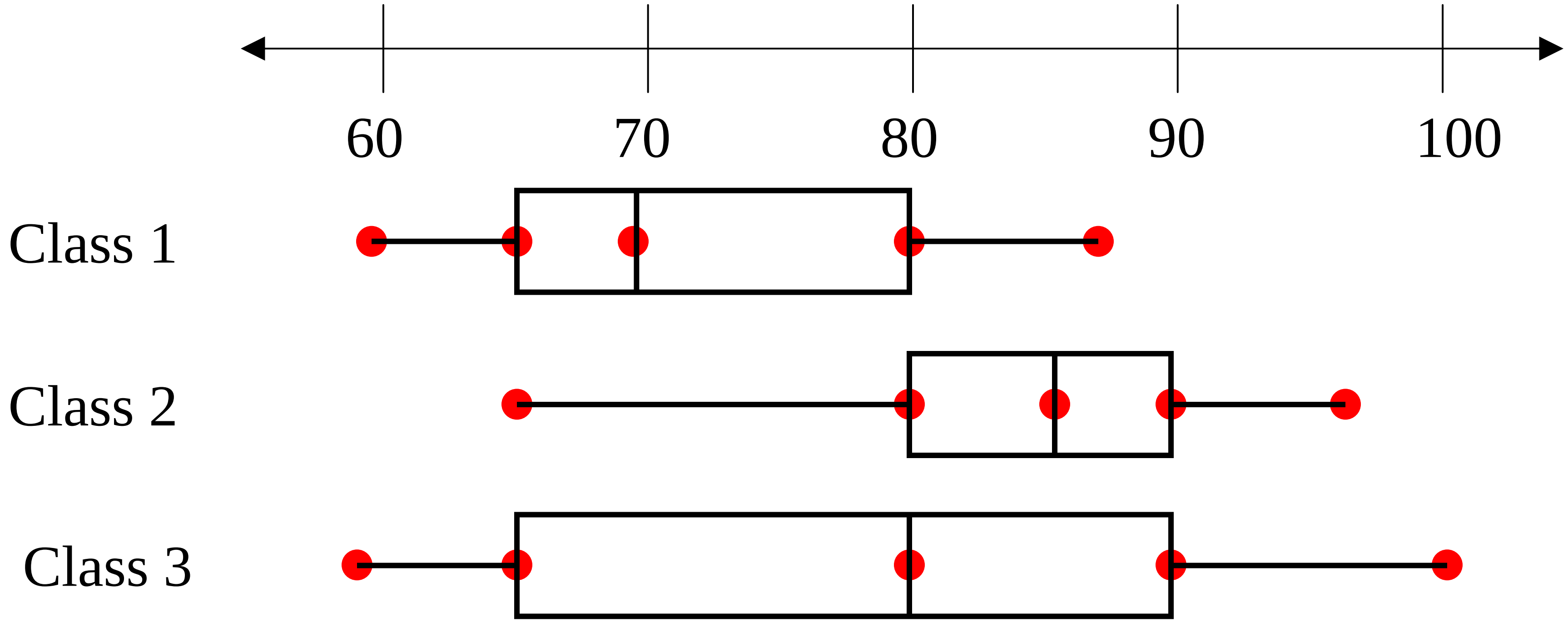
1st quartile = 14

3rd quartile = 34

range = 45



Example for 5-8



Which class has the highest median? **2**

For which class are the scores in the middle half closest together? **2**

Example for 9-18

1, 4, 6, 6, 7, 8, 8, 8

$$\text{mean} = \frac{1+4+6+6+7+8+8+8}{8} = \frac{48}{8} = 6$$

$$\sigma^2 = \frac{(1-6)^2 + (4-6)^2 + 2(6-6)^2 + (7-6)^2 + 3(8-6)^2}{8}$$

$$\sigma^2 = \frac{(-5)^2 + (-2)^2 + 2(0)^2 + (1)^2 + 3(2)^2}{8}$$

$$\sigma^2 = \frac{25+4+0+1+12}{8} = \frac{42}{8} = 5.3$$

$$\sigma = \sqrt{5.3} = 2.3$$

Section 15-3

The Normal Distribution

Objectives

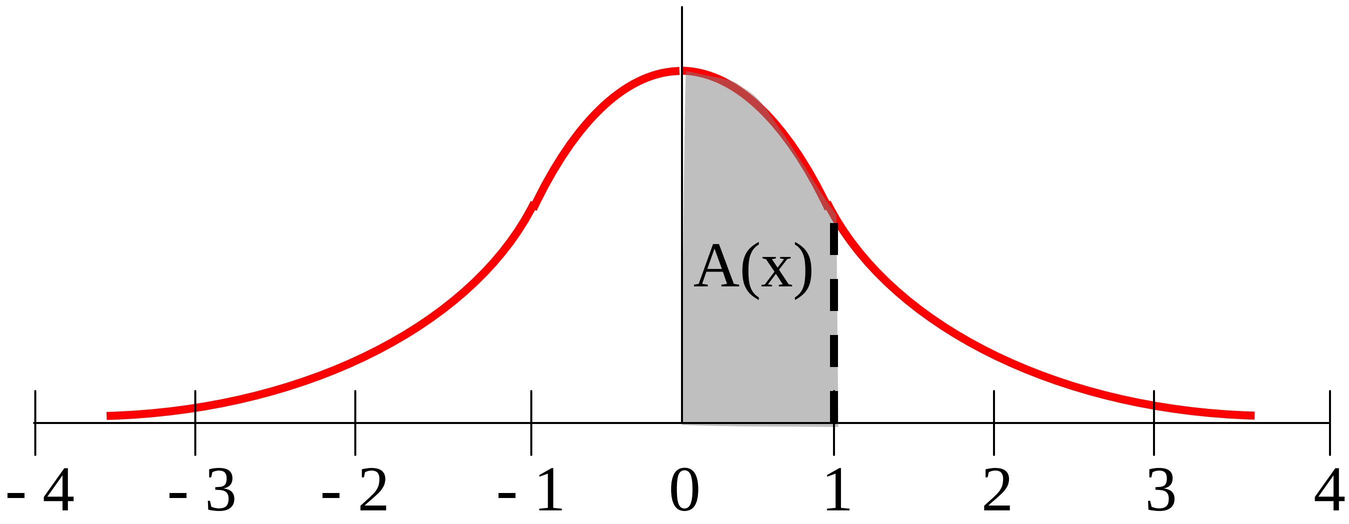
- to find the percentage of data between standardized scores

The Normal Distribution Curve

- normal distribution: data with a distribution graph creating a bell curve
- standard normal distribution: a bell curve with a mean of zero and a standard deviation of one
- standard normal curve: bell curve
  - symmetric with respect to the y-axis
  - approaches the x-axis asymptotically as $|x|$ increases
  - the total area under the curve above the x-axis is equal to 1
- standardized value (z-score): for any raw score x with a mean raw score of $\bar{x}$ and a standard deviation of σ is given by the formula:

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

The Normal Distribution Curve



x	Area, A(x)	x	Area, A(x)	x	Area, A(x)
0.0	0.0000	1.4	0.4192	2.8	0.4974
0.2	0.0793	1.6	0.4452	3.0	0.4987
0.4	0.1554	1.8	0.4641	3.2	0.4993
0.6	0.2257	2.0	0.4772	3.4	0.4997
0.8	0.2881	2.2	0.4861	3.6	0.4998
1.0	0.3413	2.4	0.4918	3.8	0.4999
1.2	0.3849	2.6	0.4953	4.0	0.5000

Example for 1-10

The mean weight of a loaf of bread was found by sampling to be 455g, with a standard deviation of 5g. Assuming a normal distribution, find the percent of loaves with weights that are:

less than 450g

$$z = \frac{450 - 455}{5}$$

$$z = \frac{-5}{5}$$

$$z = -1$$

34.13% of loafs are between - 1 standard deviations and the mean; therefore 50.00% - 34.13% = **15.87%** weigh less than 450g.

Example for 1-10

The mean weight of a loaf of bread was found by sampling to be 455g, with a standard deviation of 5g. Assuming a normal distribution, find the percent of loaves with weights that are:

greater than 445g

$$z = \frac{445 - 455}{5}$$

$$z = \frac{-10}{5}$$

$$z = -2$$

47.72% of loafs are between - 2 standard deviations and the mean; therefore 50.00% + 47.72% = **97.72%** weigh greater than 445g.

Example for 1-10

The mean weight of a loaf of bread was found by sampling to be 455g, with a standard deviation of 5g. Assuming a normal distribution, find the percent of loaves with weights that are:

greater than 470g

$$z = \frac{470 - 455}{5}$$

$$z = \frac{15}{5}$$

$$z = 3$$

49.87% of loafs are between 3 standard deviations and the mean; therefore 50.00% - 49.87% = **0.13%** weigh greater than 470g.

Example for 1-10

The mean weight of a loaf of bread was found by sampling to be 455g, with a standard deviation of 5g. Assuming a normal distribution, find the percent of loaves with weights that are:

between 450g and 460g

$$z = \frac{450 - 455}{5} \qquad z = \frac{460 - 455}{5}$$

$$z = \frac{-5}{5} \qquad z = \frac{5}{5}$$

$$z = -1 \qquad z = 1$$

34.13% of loafs are between - 1standard deviations and the mean; 34.13% of loafs are between 1standard deviations and the mean; therefore 34.13% + 34.13% = **68.26%** weigh between 450g and 460g.

Section 15-4

Correlation

Objectives

- to draw a scatter plot and identify in general terms the correlation between two sets of data
- to find the correlation coefficient
- to write a regression equation
- to use a regression equation to make predictions about unknown values in a data set

Correlation

- scatter plot: a diagram of the mathematical relationship between two sets of data
- correlation coefficient: ranges between - 1 and 1 inclusive and characterizes how closely the points in a scatter plot cluster about a line. Correlation is NOT a cause and effect relationship.
  - high positive correlation : the points in the scatter plot cluster about a line with positive slope indicating that the variables increase or decrease together
  - high negative correlation: the points of the scatter plot cluster about a line with negative slope indicating that one variable increases as the other decreases
  - zero correlation coefficient: points are randomly distributed indicating no correlation between the variables.

Calculating the Correlation Coefficient

- given a set of ordered pairs (x, y), the correlation coefficient r_{xy} or just r is:

$$r = \frac{\overline{xy} - \overline{x} \bullet \overline{y}}{\sigma_x \sigma_y}$$

  - where $\overline{xy}$ is the mean of the products of the elements
  - $\overline{x}$ and σ_x are the mean and standard deviation of the independent values, respectively
  - $\overline{y}$ and σ_y are the mean and standard deviation of the dependent values, respectively

Regression Equations

- regression line: when the correlation between variables in a scatter plot is high you can draw a regression line through the data that best fits the known variables .
- the regression line stays relatively close to all of the points in the data sets; therefore, it can be used to estimate values not appearing the original data set
- the regression line of a data set for ordered pairs (x, y)
  - passes through the point $\left(\bar{x}, \bar{y}\right)$
  - and has a slope: $r\left(\frac{\sigma_y}{\sigma_x}\right)$

Example for 1-10

Given $M_{xy} \approx 3226.33$, $M_x \approx 65.17$, $M_y \approx 49.33$, $\sigma_x \approx 2.85$, and $\sigma_y \approx 4.19$, find (a) the value of the correlation coefficient for the ordered pairs and (b) an equation of the regression line relating x and y. $M_{xy} = \overline{xy},\ M_x = \overline{x},\ M_y = \overline{y}$

$$r = \frac{\overline{xy} - \overline{x} \bullet \overline{y}}{\sigma_x \sigma_y}$$

$$y - y_1 = m(x - x_1)$$

$$r = \frac{3226.33 - (65.17)(49.33)}{(2.85)(4.19)}$$

$$y - 49.33 = (0.96)\left(\frac{4.19}{2.85}\right)(x - 65.17)$$

r = 0.96

y = 1.41x - 42.56

Section 15-5

Fundamental Counting Principles

Objectives

- to use the fundamental counting principles

Fundamental Counting Principles

- If one selection can be made m ways, and for each of these a second selection can be made n ways, then the number of ways the two selections can be made is (m)(n).
- mutually exclusive: two or more sets with no common elements.
- If the possibilities being counted can be grouped into mutually exclusive cases, then the total number of possibilities is the sum of the number of possibilities in each case.

Example for 1-12

How many odd three digit positive integers can be written using the digits 2, 3, 4, 5, and 6?

There are three positions to be filled: the hundreds, tens and ones positions; therefore, this is an example of the first fundamental counting principle.

All five digits can be used in the hundreds position, all five digits can be used in the tens position, but only the 3 and the 5 can be used in the ones position.

(5)(5)(2) = **50 numbers**

Example for 13-16

How many license plate of 3 symbols can be made using at least two letters for each?

You have these mutually exclusive groups: L-L-#, L-#-L, #-L-L, L-L-L. So you must use the second fundamental counting principle and add their totals, but to find the total number of each kind of license you must use the first fundamental counting principle.

L-L-# = (26)(26)(10) = 6760

L-#-L = (26)(10)(26) = 6760

#-L-L = (10)(26)(26) = 6760

L-L-L = (26)(26)(26) = 17,576

total number of license plates = 37,856

Section 15-6

Permutations

Objectives

- to find the number of permutations of a set when all elements are used
- to find the number of permutations of a set when only some of the elements are used
- to find the number of permutations of a set when some elements are identical

Permutations

- permutations: an arrangement of the elements of a set in a specific order
- the number of permutations of a set with n elements when all of the elements will be used is n!
- the number of permutations of a set of n elements when only r elements will be used is given by the formula:

$$ {}_{n}P_{r} = \frac{n!}{(n-r)!} $$

- the number of permutations of a set with n elements where some of the elements are identical is given by the formula:

$$ P = \frac{n!}{n_1!n_2!\ldots} $$

where n_1 is the frequency of the first identical element and n_2 is the frequency of the second identical element

Example for 5-8, 11, 12, 14

Find ${}_{n}P_{r}$ if n = 6 and r = 3

$$_{n}P_{r} = \frac{n!}{(n-r)!}$$

$$_{6}P_{3} = \frac{6!}{(6-3)!}$$

$$_{6}P_{3} = \frac{6!}{3!}$$

$$_{6}P_{3} = \frac{(6)(5)(4)(3)(2)(1)}{(3)(2)(1)}$$

$$_{6}P_{3} = (6)(5)(4) = 120$$

Example for 9, 10, 13

In how many ways can 6 different books be arranged on a shelf?

$P = 6!$

$P = 720$

Example for 15-24

Find the number of ways the letters of the word MISSISSIPPI can be arranged.

There are three elements of the set that have duplicates. The letter "I" which occurs with a frequency of 4, the letter "S" which also has a frequency of 4 , and the letter "P" which has a frequency of 2. This makes the formula for calculating the number of permutations:

$$P = \frac{11!}{4!4!2!}$$

$$P = \frac{(11)(10)(9)(8)(7)(6)(5)(4)(3)(2)(1)}{(4)(3)(2)(1)(4)(3)(2)(1)(2)(1)}$$

P = (11)(10)(9)(7)(5) = **34,650**

Example for 25-28

Show that ${}_nP_5 - {}_nP_4 = (n - 5)\,{}_nP_4$

$$\frac{n!}{(n-5)!} - \frac{n!}{(n-4)!} = (n-5)_n P_4$$

$$\frac{n!}{(n-5)!} - \frac{n!}{(n-4)(n-5)!} = (n-5)_n P_4$$

$$\frac{(n-4)n!}{(n-4)(n-5)!} - \frac{n!}{(n-4)(n-5)!} = (n-5)_n P_4$$

$$\frac{(n-4)n! - n!}{(n-4)(n-5)!} = (n-5)_n P_4$$

$$\frac{n![(n-4)-1]}{(n-4)(n-5)!} = (n-5)_n P_4$$

$$\frac{n!(n-5)}{(n-4)!} = (n-5)_n P_4$$

Section 15-7

Combinations

Objectives

- to list the subset of a set
- to evaluate combinations
- to apply the concept of combinations to counting subsets

Combinations

- subset of set A: every member (element) of the subset is a member of A. the empty set (ϕ) is a subset of all sets.
- combination: a subset with r elements taken from a set with n elements. Order is not important in combinations.
- ${}_nC_r$: is the number of combinations of r elements taken from a set with n elements

$$ {}_nC_r = \frac{n!}{r!(n-r)!} $$

Example for 1&2

For the 2-letter set {J, K} find all the subsets.

{J, K} {J} {K} ϕ

For the 2-letter set {J, K} find all the subsets containing fewer than 2 letters.

{J} {K} ϕ

Example for 3-10

$${}_{10}C_8$$

$$\frac{10!}{8!(10-8)!}$$

$$\frac{10!}{8!2!}$$

$$\frac{(10)(9)}{2}$$

45

Example for 11-20

Seven points lie on a circle. How many inscribed triangles can be constructed having any three of these points as vertices.

$${}_{7}C_{3}$$

$$\frac{7!}{3!(7-3)!}$$

$$\frac{7!}{3!4!}$$

$$\frac{(7)(6)(5)}{(3)(2)}$$

35

Example for 21-26

How many 5-card hands can be dealt having exactly 3 aces and 2 other cards?

There are two parts to the hand the ace part and the rest of the hand. These must be counted separately and the total multiplied like we did for problems in section 15-5.

ace part ${}_4C_3$ and the rest of the hand ${}_{48}C_2$

${}_4C_3 = 4$ and ${}_{48}C_2 = 1128$

(4)(1128) = 4512

Section 15-8

Sample Spaces and Events

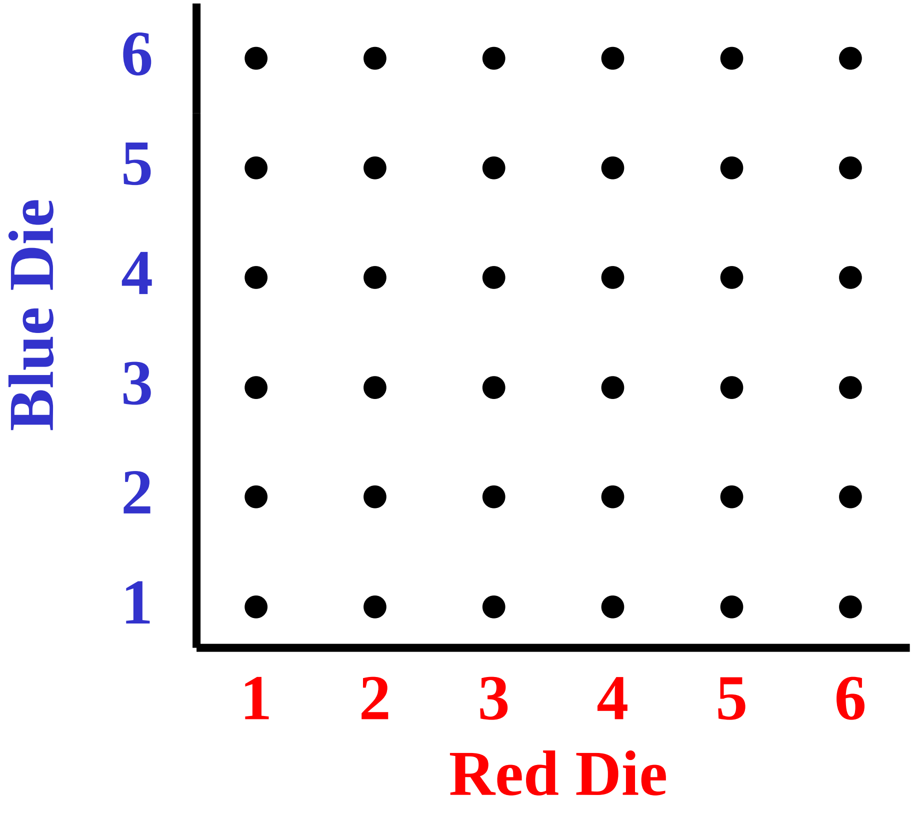
Objectives

- to specify a sample space
- to specify an event

Sample Spaces and Events

- random experiment: an experiment where you do not necessarily get the same outcome when you repeat it under the same conditions
- sample space: the set of all possible outcomes of a random experiment
- event: a subset of possible outcomes for an experiment
- simple event: an event with a single element

Dice Experiment



Example for 1-8

The product of the numbers showing on the two dice is 12.

{(2, 6), (3, 4), (4, 3), (6, 2)}

Example for 9-12

The ordered pair (r, b) where r + b = 9.

{(3, 6), (4, 5), (5, 4), (6, 3)}

Section 15-9

Probability

Objectives

- to calculate the probability of an event

Probability

- If $\{a_1, a_2, a_3, ...a_n\}$ is a sample space containing n equally likely outcomes, then the probability of each simple event is $\frac{1}{n}$:

$$P(a_1) = P(a_2) = P(a_3) = .. = P(a_n) = \frac{1}{n}$$

- If the sample space for an experiment consists of n equally likely outcomes, and if k of them are in event E, then:

$$P(E) = \underbrace{\frac{1}{n} + \frac{1}{n} + ... + \frac{1}{n}}_{\textbf{k addends}} = \frac{k}{n}$$

Example for 1-12

A bag contains 2 red, 4 yellow and 6 blue marbles. Two marbles are drawn at random. Find the probability of each event.

a) both are red

b) both are yellow

c) both are blue

d) one is red and one is yellow

e) neither is red

f) neither is blue

a. $\frac{{}_{2}C_{2}}{{}_{12}C_{2}} = \frac{1}{66}$ $\quad \frac{1}{66}$

b. $\frac{{}_{4}C_{2}}{{}_{12}C_{2}} = \frac{6}{66}$ $\quad \frac{1}{11}$

c. $\frac{{}_{6}C_{2}}{{}_{12}C_{2}} = \frac{15}{66}$ $\quad \frac{5}{22}$

d. $\frac{({}_{2}C_{1})({}_{4}C_{1})}{{}_{12}C_{2}} = \frac{(2)(4)}{66}$ $\quad \frac{4}{33}$

e. $\frac{{}_{10}C_{2}}{{}_{12}C_{2}} = \frac{45}{66}$ $\quad \frac{15}{22}$

f. $\frac{{}_{6}C_{2}}{{}_{12}C_{2}} = \frac{15}{66}$ $\quad \frac{5}{22}$

Example for 13&14

For a certain math class, the mean time to complete homework assignments is 36 minutes with a standard deviation of 10 minutes. Assuming the completion times are normally distributed, what is the probability that the time required to complete a given homework assignment in this class will be:

a) less than 20 minutes?

b) more than one hour?

a. $\frac{20-36}{10} = \frac{-16}{10} = -1.6$

b. $\frac{60-36}{10} = \frac{24}{10} = 2.4$

0.5000 - 0.4452 = **0.0548**

0.5000 - 0.4918 = **0.0082**

Section 15-10

Mutually Exclusive and Independent Events

Objectives

- to identify and calculate the probability of mutually exclusive events
- to identify and calculate the probability of independent events

Sets

- intersection ($\cap$): given any sets A and B, the intersection is the set whose members are elements belonging to both A and B
- disjoint sets: sets whose intersection is the null set
- union ($\cup$): given the sets A and B, the union is the set containing all of the elements of set A along with all of the elements of set B
- disjoint sets are mutually exclusive sets
- independent: the occurrence of one event has no effect on the probability of another event
- complement ($\overline{A}$): given a sample space S and an event A, the complement of A ($\overline{A}$) the members of S not in A

Probability and Sets

- For any sample space S, $P(S) = 1$ and $P(\phi) = 0$
- For any two events A and B in a sample space, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
- If A and B are mutually exclusive events, $P(A \cup B) = P(A) + P(B)$
- Two events A and B are independent if and only if: $P(A \cap B) = P(A) \bullet P(B)$
- Given an event A, $P\left(A \cup \overline{A}\right) = P(A) + P(\overline{A}) = 1$

Example for 1-16

A red and a green die are rolled. Let A be the event that the red die shows 2 and B be the event that the sum of the number showing is 8.

a) Find the probability of A, B, $A \cup B$, and $A \cap B$.

b) Are A and B independent events?

S = 36

A(r, g): {(2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6)}

B(r, g): {(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)}

$A \cup B$ (r, g): {(2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6), (3, 5), (4, 4), (5, 3), (6, 2)}

$A \cap B$ (r, g): {(2, 6)}

$$P(A) = \frac{1}{6},\ P(B) = \frac{5}{36},\ P(A \cup B) = \frac{5}{18},\ P(A \cap B) = \frac{1}{36}$$

No