Advanced Data Analysis

The Standard Deviation

Definition

Standard = "For a typical value in the list"
Deviation = "Distance from the mean"
Thus, the standard deviation describes a "typical plus / minus distance from the mean."

Representation

Putting mean in middle and counting three SD's to the left and right will capture all or nearly all values.

Ex: "In a professional bowling tournament, ladies' scores have a mean of 200 and a standard deviation of 10."

z-score = # of SD's from the mean

$$z = \frac{x - M}{SD}$$

What is the *z*-score for a bowling score of 185?

What bowling score is 2.5 standard deviations above the mean?

Percentiles, quartiles, and IQR

Percentiles

A percentage of a distribution is best represented by area on its graph.
Percentile of x = Percentage of scores below x
Percentiles are cumulative; they add / subtract.

If Alex is at the 30th percentile and Bobby is at the 80th percentile,

1. What percent of students scored worse than Alex?

2. What percent of students scored better than Bobby?

3. What percent of students scored between Alex and Bobby?

Scores are one-dimensional and don't give this information.

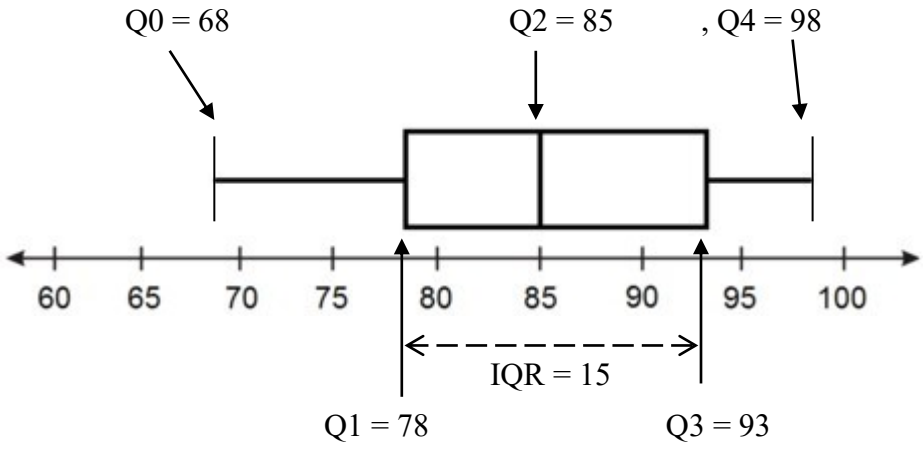
Quartiles

Quartiles separate the data into four quarters containing equal numbers of data points (25% each)

The GRE uses the term "quartile" sometimes as a point (*at* the 3rd quartile) and sometimes as an interval (*in* the 3rd quartile)

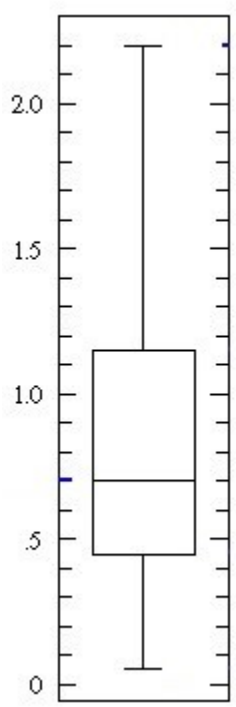
Quartile	"At" the point	"In" the interval
"0th"	Q0 = Min	
1st / lowest	Q1 = 25th percentile	From min to 25th percentile
2nd	Q2 = 50th percentile (median)	From 25th to 50th percentile
3rd	Q3 = 75th percentile	From 50th to 75th percentile
4th / highest	Q4 = Max	From 75th to 100th percentile

A “box and whisker plot” uses this 5-point summary to represent the distribution.



Inter-quartile range = IQR = Q3 – Q1
(15 in this example)

Note that this plot shows us at a glance where the “heart” of the data is and the approximate shape of the data (skew follows the long whisker).



Problems: The vertical box-and-whisker plot represents the weights (in pounds) of 1,000 packages in a mail room.

1. Describe the distribution of these weights

2. What is the median weight?

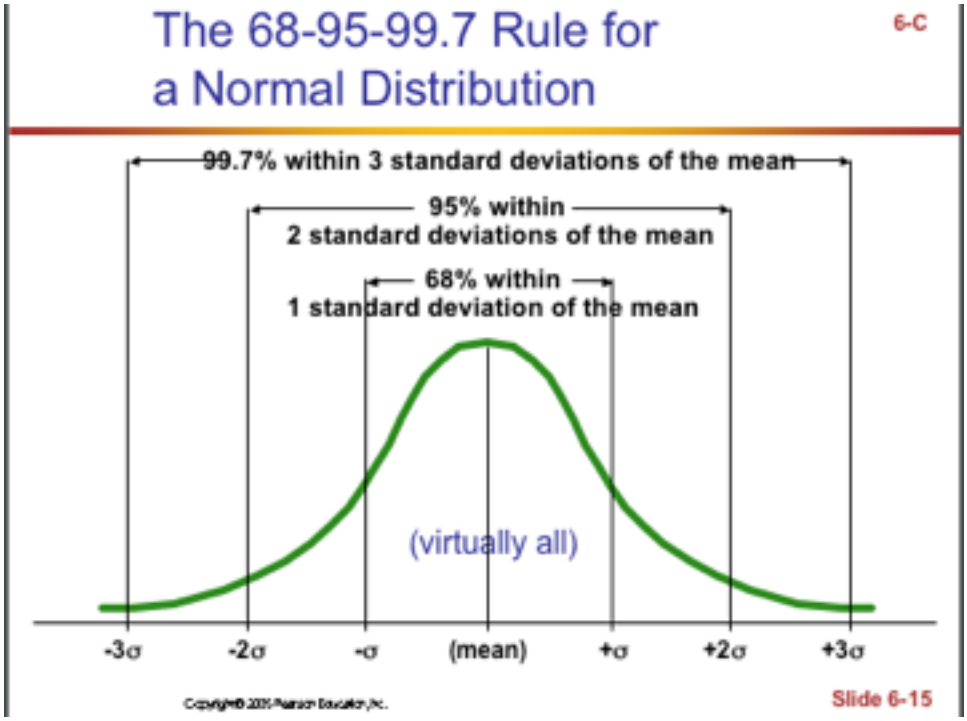
3. Find the inter-quartile range

4. What percent of the packages have weights between 0.45 and 0.7 pounds?

5. How many packages weigh less than 1.15 pounds?

The Normal Distribution

This is a particular type of "bell curve" with a known relationship between *z*-scores and percentiles.



Problem: If checkout times at a grocery store are normally distributed with mean = 3.6 minutes and SD = 0.9 minutes, what percent of checkouts are

Between 2.7 and 4.5 minutes?

Between 4.5 and 5.4 minutes?

Greater than 6.3 minutes?

Probability

The key terms and their interpretations

Mix-and-match

_____ Basic probability	A. Addition
_____ AND	B. Subtraction
_____ OR	C. Multiplication
_____ NOT	D. Division

Venn Diagram Visualizations

NOT = Outside

	A	**NOT A**
B		
NOT B		

AND = Intersection or overlap

	A	**NOT A**
B	A AND B	
NOT B		

OR = Union or all-inclusive

	A	**NOT A**
B	A OR B	
NOT B		

2931 S. Sepulveda Blvd. Upstairs Ste. I
Los Angeles, CA, 90064
(310) 478-6146
scot@NthDegreeTutoring.com

Basic problems

Patterned-English question:
If you select a choice, what is the probability that it will be a success?

Answer: $\dfrac{n(\text{Successes Within the Choices})}{n(\text{Choices})}$

Empirical probability table: Make row / column totals!

The following table shows the number of website subscribers who chose various kinds of subscription options.

	Basic membership	**Premium membership**	**Gold membership**
Monthly	8,000	6,000	2,000
Annual	5,000	2,000	1,000

If you choose a subscriber, what is the probability she chose a premium membership?

If you choose a gold member, what is the probability that he subscribes monthly?

If you choose a monthly member, what is the probability that she has a gold membership?

AND, OR, and NOT (Compound Probability)

The adding / subtracting principles

	Basic membership	Premium membership	Gold membership	Total
Monthly	8,000	6,000	2,000	16,000
Annual	5,000	2,000	1,000	8,000
Total	13,000	8,000	3,000	24,000

If you choose a subscriber, what is the probability that he is NOT a gold member?

If you choose a subscriber, what is the probability that he is a basic member AND is an annual subscriber? (AND = "overlap") Note that sloppy English won't always use the word "AND".

If you choose a subscriber, what is the probability that she is a basic member OR a premium member? (OR = "all-encompassing")

$$P(A \text{ OR } B) = P(A) + P(B)$$

Must subtract overlap:

If you choose a subscriber, what is the probability that she is a basic member OR a monthly subscriber?

$$P(A \text{ OR } B) = P(A) + P(B) - P(A \text{ AND } B)$$

The multiplication principles

Draw a card and roll a die. What is the probability that both are a 5?

Reword: The die is a 5 AND the card is a 5.

Multiply: $P(A \text{ AND } B) = P(A) \times P(B)$

Adjust if the first selection affects the second one ("conditional" probability)

Draw two cards from a deck. What is the probability that both are 5's?

Practice # 15 with class. Have them do #'s 10 and 17 alone.

c) What's the probability that a man with high blood pressure has high cholesterol?
d) What's the probability that a man has high blood pressure if it's known that he has high cholesterol?

Death penalty. The table shows the political affiliations of American voters and their positions on the death penalty.

10.

		Death Penalty	
		Favor	Oppose
Party	Republican	0.26	0.04
	Democrat	0.12	0.24
	Other	0.24	0.10

a) What's the probability that a randomly chosen voter favors the death penalty?
b) What's the probability that a Republican favors the death penalty?
c) What's the probability that a voter who favors the death penalty is a Democrat?
d) A candidate thinks she has a good chance of gaining the votes of anyone who is a Republican or in favor of the death penalty. What portion of the voters is that?

Global survey, take 2. Look again at the table summarizing the Roper survey in Exercise 5.
a) If we select a respondent at random, what's the probability we choose a person from the United States who has done post-graduate study?
b) Among the respondents who have done post-graduate study, what's the probability the person is from the United States?
c) What's the probability that a respondent from the United States has done post-graduate study?
d) What's the probability that a respondent from China has only a primary-level education?
e) What's the probability that a respondent with only a primary-level education is from China?

Birth order, take 2. Look again at the data about birth order of Intro Stats students and their choices of colleges shown in Exercise 6.
a) If we select a student at random, what's the probability the person is an Arts and Sciences student who is a second child (or more)?
b) Among the Arts and Sciences students, what's the probability a student was a second child (or more)?
c) Among second children (or more), what's the probability the student is enrolled in Arts and Sciences?
d) What's the probability that a first or only child is enrolled in the Agriculture College?
e) What is the probability that an Agriculture student is a first or only child?

Sick kids. Seventy percent of kids who visit a doctor have a fever, and 30% of kids with a fever have sore throats. What's the probability that a kid who goes to the doctor has a fever and a sore throat?

14. Sick cars. Twenty percent of cars that are inspected have faulty pollution control systems. The cost of repairing a pollution control system exceeds $100 about 40% of the time. When a driver takes her car in for inspection, what's the probability that she will end up paying more than $100 to repair the pollution control system?

3

15. Cards. You are dealt a hand of three cards, one at a time. Find the probability of each of the following.
3, 8
a) The first heart you get is the third card dealt.
b) Your cards are all red (that is, all diamonds or hearts).
c) You get no spades.
d) You have at least one ace.

16. Another hand. You pick three cards at random from a deck. Find the probability of each event described below.
a) You get no aces.
3, 8
b) You get all hearts.
c) The third card is your first red card.
d) You have at least one diamond.

17. Batteries. A junk box in your room contains a dozen old batteries, five of which are totally dead. You start picking batteries one at a time and testing them. Find the probability of each outcome.
3, 8
a) The first two you choose are both good.
b) At least one of the first three works.
c) The first four you pick all work.
d) You have to pick 5 batteries in order to find one that works.

18. Shirts. The soccer team's shirts have arrived in a big box, and people just start grabbing them, looking for the right size. The box contains 4 medium, 10 large, and 6 extra-large shirts. You want a medium for you and one for your sister. Find the probability of each event described.
3, 8
a) The first two you grab are the wrong sizes.
b) The first medium shirt you find is the third one you check.
c) The first four shirts you pick are all extra-large.
d) At least one of the first four shirts you check is a medium.

19. Eligibility. A university requires its biology majors to take a course called BioResearch. The prerequisite for this course is that students must have taken either a Statistics course or a computer course. By the time they are juniors, 52% of the Biology majors have taken Statistics, 23% have had a computer course, and 7% have done both.
1, 4, 5
a) What percent of the junior Biology majors are ineligible for BioResearch?
b) What's the probability that a junior Biology major who has taken Statistics has also taken a computer course?
c) Are taking these two courses disjoint events? Explain.
d) Are taking these two courses independent events? Explain.

20. Benefits. Fifty-six percent of all American workers have a workplace retirement plan, 68% have health insurance, and 49% have both benefits. We select a worker at random.

7

Counting Principles

Overlapping sets

300 diners went through a Thanksgiving buffet with a choice of ham and / or turkey.
100 had ham.
150 had turkey.
20 had both ham and turkey.
How many diners had exactly one kind of meat?

	Ham	No ham	Total
Turkey			
No turkey			
Total			

300 diners went through a Thanksgiving buffet with a choice of ham and / or turkey.
100 had ham.
150 had turkey.
What is the least / greatest amount of diners who could have had no meat?

	Ham	No ham	Total
Turkey			
No turkey			
Total			

Multiplication / Powers:

To select "one of each" from multiple sets, multiply the sizes of the sets.

Jimmy has 8 shirts, 4 pairs of pants, 3 coats, and 5 pairs of shoes in his closet. How many outfits could he create out of his closet, if an outfit consists of one shirt, one pair of pants, one coat, and one pair of shoes?

Sometimes "multiple sets" are really all the same set, if you keep selecting from the whole set ("replacement")

You select a card, put it back in the deck, and select again and again, three times. What is the probability that you draw a red card all three times? How would you write this with an exponent?

How many CA license plates can be made of the form #AAA### ?

___ ___ ___ ___ ___ ___ ___

Combinatorics

How many ways are there to select a license plate if you are not allowed to repeat characters?

___ ___ ___ ___ ___ ___ ___

Multiplying descending integers is the basic counting principle behind "combinatorics", or counting the number of ways to make selections from one set without repetition.

First, recap difference between "set" and "sequence".

Set = The grocery bag with all your grocery items lumped inside.
Sequence = The individual grocery items lined up on the checkout line.
A "sequence" can be indicated by any situation involving order, rank, labels, or distinctions of any kind. (Different orderings = different sequences)

One set can be arranged into many sequences.

Factorial

Orderings of an entire set into a sequence.
If your grocery basket has six items in it, in how many different orders can they be rung up on the checkout line?
Memorize factorials up to 6! = 720

$1! = 1$	$2! = 2 \times 1 = 2$	$3! = 3\times2\times1 = 6$
$4! = 4\times3\times2\times1 = 24$	$5! = 5\times4\times3\times2\times1 = 120$	$6! = 6\times5\times4\times3\times2\times1 = 720$

Permutation

Orderings of part of a set (a subset) into a sequence.
If your grocery basket has six items in it but you can only afford four of them, in how many orders can you ring up four of them on the checkout line?

${}_6P_4 =$ ____ ____ ____ ____ =

Combination

Selections of subsets (no order, rank, distinction, etc)
If your grocery basket has eight items in it but it's too heavy and you have to remove three of them into a different bag, how many combinations of items can be in that bag?

$${}_8C_3 = \frac{{}_8P_3}{3!} = \frac{\_\_\_\_\times\_\_\_\_\times\_\_\_\_}{\_\_\_\_\times\_\_\_\_\times\_\_\_\_} =$$

What do you notice about the number of combinations left for the original bag? What does this tell you about ${}_8C_3$ and ${}_8C_5$?

Set or sequence?

Students have a lot of trouble identifying a sequence when "order" is not specifically invoked.
"Order" can be represented by rank, labels, or distinctions of any kind.
Key question: If I scrambled the items, would I be creating a new selection?
Alternatively, can each item be identified in two different ways? (Like name and order?)
If the answer to one of these key questions is "Yes", then it's a permutation / factorial / sequence.

In a classroom of 20 students, 6 will be selected to wear identical costumes for a parade.
Solution: We are choosing a set of students, because each selected student can only be identified in one way, by name.

In a classroom of 20 students, 6 will be selected to wear costumes. One will wear red, one orange, one yellow, one green, one blue, and one purple.
Solution: We are selecting a sequence of students, because each selected student can be identified in two separable ways, by name and color. If the students swapped costumes, they would create a different selection.

Practice: Set or sequence?

1. The six students will now dress up as a Skittles bag for Halloween.

2. The six students in colored costumes line up to see if they form a rainbow.

3. You are holding a five-card hand in a poker game.

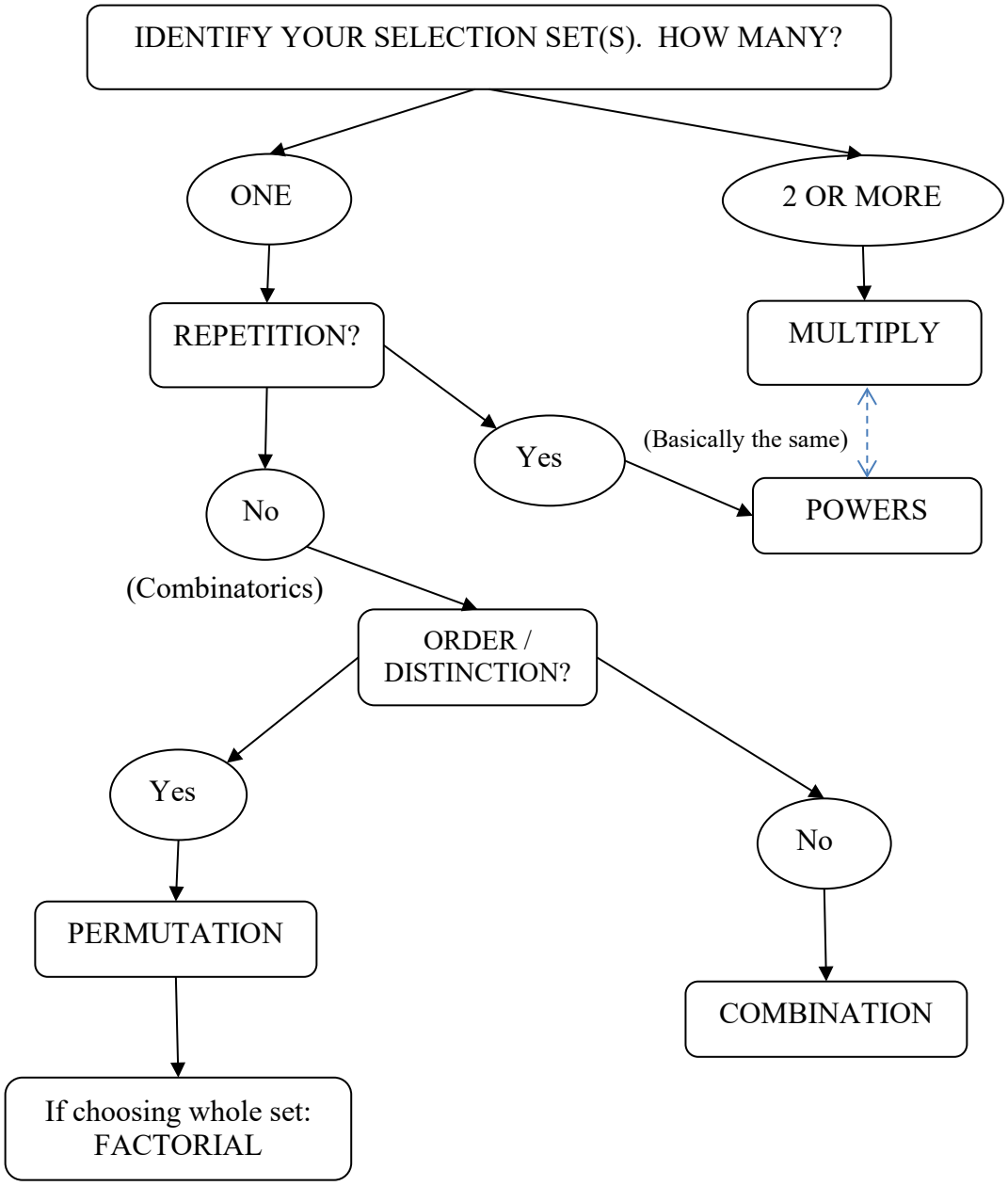
4. In a card game, five players take turns drawing one card from the top of the deck.

5. A club holds an election for president, vice president, secretary, and treasurer.

6. A club selects four people to organize a party.

7. (Hard) To win the lottery, you must select six numbers out of the integers 1 - 40. Winning numbers are always read in increasing order.

“Counting” Flow Chart



9. One person in a stadium filled with 100,000 people is chosen at random to win a free pair of airline tickets. What is the probability that it will *not* be you?

   a. 1 in 100,000 b. 0.99 c. 0.99999

10. There are 365 possible birthdays in a year. In a class of 25 students, the chance of finding 2 students with the same birthday is

    a. 25/365. b. 2 × 25/365. c. greater than 0.5.

Exercises 7E

REVIEW QUESTIONS

1. What are arrangements with repetition? Give an example of a situation in which the n^r formula gives the number of possible arrangements.
2. What do we mean by *permutations*? Explain the meaning of each of the terms in the permutations formula. Give an example of its use.
3. What do we mean by *combinations*? Explain the meaning of each of the terms in the combinations formula. Give an example of its use.
4. Explain what we mean when we say that *some* outcome is much more likely than a particular outcome. How does this idea affect our perception of coincidences?

DOES IT MAKE SENSE?

Decide whether each of the following statements makes sense (or is clearly true) or does not make sense (or is clearly false). Explain your reasoning.

5. I used the permutations formula to determine how many possible relay orders we could make with the 10 girls on our swim team.
6. I used the combinations formula to determine how many different five-card poker hands are possible.
7. The number of different possible batting orders for 9 players on a 25-person baseball team is so large that there's no hope of trying them all out.
8. It must be my lucky day, because the five-card poker hand I got had only about a 1 in 2.5 million chance of being dealt to me.
9. The probability that two people in a randomly selected group will have the same last name is much higher than the probability that someone will have the same last name as I do.
10. Someone wins the lottery every week, so I figure that if I keep playing eventually I will be the one who wins.

BASIC SKILLS & CONCEPTS

11–22: Review of Factorials. Use the skills covered in the Brief Review on p. 459 to evaluate the following quantities *without* using the factorial key on your calculator (you may use the multiplication key). Show your work.

11. $6!$
12. $12!$
13. $\frac{5!}{3!}$
14. $\frac{10!}{8!}$
15. $\frac{12!}{4!\,3!}$
16. $\frac{9!}{4!\,2!}$
17. $\frac{11!}{3!(11-3)!}$
18. $\frac{30!}{29!}$
19. $\frac{8!}{3!(8-3)!}$
20. $\frac{30!}{28!}$
21. $\frac{6!\,8!}{4!\,5!}$
22. $\frac{15!}{2!\,13!}$

23–40: Counting Methods. Answer the following questions using the appropriate counting technique, which may be either arrangements with repetition, permutations, or combinations. Be sure to explain why this counting technique applies to the problem.

23. How many different nine-digit zip codes can be formed?
24. How many different five-character passwords can be formed from the lowercase letters of the alphabet?
25. How many different five-character passwords can be formed from the lowercase letters of the alphabet if repetition is not allowed?
26. A city council with nine members must elect a four-person executive committee consisting of a mayor, deputy mayor, secretary, and treasurer. How many executive committees are possible?
27. How many ways can the nine performances at a piano recital be ordered?
28. A city council with nine members must appoint a three-person subcommittee. How many subcommittees are possible?
29. Suppose you have 20 CDs from which you choose 6 CDs to put in the CD player in your car. If you are not particular about the order, how many 6-CD sets are possible?
30. How many 6-person lineups can be formed from a 15-player volleyball roster, assuming every player can be assigned to any position?
31. How many different birth orders with respect to gender are possible in a family with five children? (For example, BBBGG and BGBGG are different orders.)
32. How many different 5-card hands can be dealt from a 52-card deck?
33. How many license plates can be made of the form *XX–YYYYY*, where *X* is a letter of the alphabet and *Y* is a numeral 0–9?
34. How many different groups of six balls can be drawn from a barrel containing balls numbered 1–36?
35. How many different telephone numbers of the form *aaa-bbb-cccc* can be formed if the area code *aaa* cannot contain 0 and the prefix *bbb* cannot contain 9?
36. How many anagrams (rearrangements) of the letters ILOVEMATH can you make?