Chapter 1 - 6

Categorical variable
Quantitative variable
Distribution of a variable
Bar graph
Pie Chart
Dotplot
Center, Spread, Shape, Outlier
Stemplot
Histogram
Symmetric, Skewed Right, Skewed Left
Relative frequency graph
Percentile
Cumulative frequency graph (ogive)
Timeplot
Mean
Median
Mode
Range
Q1 (25^{th} percentile)
Q3 (75^{th} percentile)
IQR
Five Number Summary
Boxplot
Modified Boxplot
Standard deviation (s)
Variance (s^2)
Parallel Boxplots
Back-to-Back Stemplot

Mathematical model
Density curve (symmetric, skewed right, skewed left).
Mean
Median

Activities:

(1) Make density curve from frequency histogram. Flip coin eight times, record X (number of heads). X has possible values 0, 1, 2, 3, 4, 5, 6, 7, 8. From frequency histogram for X values, create empirical probability histogram by dividing by total number of trials. Fit a curve to the histogram (approximately normal). Discuss concept of area under curve. [NOTE: Can simulate coin flipping by rolling 8 dice. Odds represent heads and evens represent tails (much quicker than actually flipping coins).]

(2) Roll 8 dice and let X = number of 6's obtained. X has possible values 0, 1, 2, 3, 4, 5, 6, 7, 8. As in (1), create frequency histogram, empirical probability histogram, and fit a curve to the histogram (skewed right). Discuss area under curve.

NOTE: Using the empirical probability histograms, it is possible to informally introduce the concept of **significance**. For instance, one could define an event to be significant if the probability that it will happen is less then some percent (say 5%). With this definition, one can ask question like "Is the event x = 8 significant?"

Other density curves, including uniform distribution.
Normal distributions.
The 65-95-99.7 Rule
Standard deviation (σ for population, s for sample)
μ is standard symbol for mean of a population.
Standard normal distribution ($\mu = 0$, $\sigma = 1$)
z-scores: If x is a score from a population with mean μ and standard deviation σ, then
$z_x = (x - \mu)/\sigma$.
Practice using the standard normal distribution table.

- Find proportion of observations (normal distribution) between two z-scores
- Find proportion of observations (normal distribution) greater than on less than a given z-score.
- Find z-score (normal distribution) for given percentile.

Again, can informally talk about **significance** here. For instance, one can define z-score to be significant if the probability of obtaining it is less than 5%.

Chapter 11 - 13

Simulations
Observational study
Experiment
Population
Sample
Census
Sample design
Voluntary response sample
Convenience sampling
Biased sample design
Simple random sample (SRS)
Stratified random sample
Undercoverage
Nonresponse

Large random samples give more accurate results than smaller samples.

Experimental units (if human, called subjects)
Treatment
Placebo effect
Control group

BASIC PRINCIPLES OF EXPERIMENTAL DESIGN
1. Control
2. Randomize
3. Replicate
Statistically significant (meaning of phrase)

Chapter 14 - 17

Probability
Sample space
Event
Probability model
Tree diagram
Sampling with replacement
Sampling without replacement
Probability rules
Disjoint (mutually exclusive) events
Independent events
Conditional probability

Random variable
Discrete random variable
Continuous random variable
Uniform distribution
Probability distribution of random variable x
Normal distributions… type of continuous probability distribution
Expected value of a random variable x
Mean of discrete random variable x: $\mu_X = \sum x_i p(x_i)$
Variance of discrete random x: $\sigma_X^2 = \sum (x_i - \mu)^2 \, p(x_i)$
Law of large numbers
If x and y are any two random variables, then
$\mu_{x \pm y} = \mu_x \pm \mu_y$
If x and y are independent random variables, then
$(\sigma_{x \pm y})^2 = \sigma_x^2 + \sigma_y^2$

Binomial distributions
Binomial setting
1. YES,NO (only two categories)
2. Fixed number of observations
3. Observations independent
4. probability YES same for each observation

Binomial random variable
Probability distribution function
Cumulative probability distribution function
Binomial probability

$$\text{Probability } (x = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Mean and standard deviation of binomial random variable x

$\mu_x = np$

$\sigma_x = \sqrt{np(1-p)}$

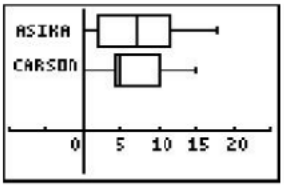
As n gets large, the binomial distribution approaches a normal distribution.
Rule of thumb: OK to use normal approximation for binomial distribution if $np \geq 10$ and $n(1-p) \geq 10$.

Geometric distribution
Geometric setting
1. YES, NO (only two categories)
2. Variable of interest x = number of trials required to obtain first YES
3. Observations independent
4. probability YES same for each observation

$$\text{Probability } (x = n) = (1-p)^{n-1} p$$

$$\mu_X = \frac{1}{p} \qquad \sigma_X = \sqrt{\frac{q}{p^2}}$$

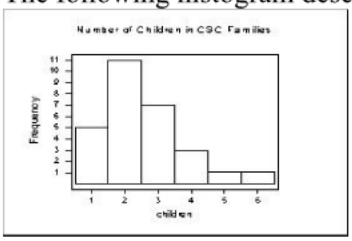
1.



The following side-by-side boxplots represent the rushing yards gained by the Woodward Academy starting running backs in the opening game of 1997 against Westminster.

Compare and contrast their performance.

2. The following histogram describes the number of children in a study of Clayton State College families.



(a) How many families are included in the study ?
(b) Find the median number of children.
(c) Find the modal number of children.
(d) Find the range of the number of children.

3. Classify the following variables as measurement (M) or categorical (C). If they are measurement, also specify whether they are discrete (D) or continuous (C):
(a) the number of students in an AP Statistics class _____
(b) the amount of Coke in a liter bottle _____
(c) the temperature of the inside of a refrigerator _____
(d) the color of ducks swimming in Lake Woodward _____
(e) the difficulty level of this test _____

4. In 1987 Reggie Jackson retired from baseball after a career of 21 years in the major leagues. He had a reputation for playing particularly well in crucial situations such as World Series games. The following table gives the number of Reggie's times at bat and hits for regular-season play and World Series games.

	Hits	At Bats
Regular Season	2584	9864
World Series	35	98

A player's batting average is the proportion of times that a player gets a hit, that is, "Hits" divided by "At-bats". Using appropriate techniques, comment on his reputation.

5. Consumer Reports magazine (June 1986) presented the following data on the number of calories in a hot dog for each of 17 brands of meat hot dogs.

173	191	182	190	172	147	146	139	175	136	179	153	107	195	135	140	139

Match the numerical statistical descriptions with the appropriate value for this data.

Description	Letter of Choice	Value	
Mean	_____	A. 138.5	H. 159.8
Median	_____	B. 24.2	I. 25.2
Mode	_____	C. 158.8	J. 153
Standard Deviation	_____	D. 634.3	K. 180.5
Variance	_____	E. 151	L. 88
Range	_____	F. 41.5	M. 2699
IQR	_____	G. 139	N. 44
Quartile 1	_____		
Quartile 3	_____		

6. Twenty men and twenty women with high blood pressure were subjects in an experiment to determine the effectiveness of a new drug in lowering blood pressure. Ten of the twenty men and ten of the twenty women were chosen at random to receive the new drug. The remaining ten men and ten women received a placebo. The change in blood pressure was measured for each subject. The design of this experiment is:
(a) completely randomized with one factor, drug
(b) completely randomized with one factor, gender
(c) randomized block, blocked by drug and gender
(d) randomized block, blocked by drug
(e) randomized block, blocked by gender

7. The primary reason for using blocking when designing an experiment is to reduce:
(a) the sensitivity of the experiment (b) variation (c) the need for randomization
(d) bias (e) confounding

8. Suppose you had twenty tomato plants and wanted to know if fertilizing them helped them produce more fruit. You randomly assign ten of them to receive fertilizer and the remaining ten receive none. You otherwise treat the plants in an identical manner.
(a) What is the explanatory variable and what is the response variable?
(b) Illustrate the appropriate design for this situation.
(c) Explain whether this would be an observational study or an experiment.
(d) If the fertilized plants produced 30% more fruit than the unfertilized plants, could you conclude that the fertilizer caused the plants to produce more? Explain.

9. For each of the following situations, state which type of sampling plan was used. Explain whether or not you think the sampling plan would result in a biased sample.
I. In order to find out how its employees felt about higher student fees imposed by the legislature, a university divided employees into three categories: staff, faculty, and student employees. A random sample was selected from each group and they were telephoned and asked for their opinions.
II. A large variety store wanted to know if consumers would be willing to pay slightly higher prices in order to have computers available throughout the store to help them locate items. They posted an interviewer at the door and told her to collect a sample of 100 opinions by asking the next person who came in the door each time she had finished an interview.

10. An educational researcher wishes to compare three teaching styles : lecture, group discovery, and computer-assisted instruction. He theorizes that there will be a significant difference in the effect on rural versus suburban schools. The following schools have agreed to participate in the study:

Rural	**Suburban**
Cattle County High	Trendy Heights High
Pecan Tech	Sport Utilities Academy
Tractor Academy	Mall Hangers Institute

Describe, in detail, the design of the experiment including the assignment of treatments and the manner in which results will be measured.

11. A blood disease is found in 2% of the persons in a certain population. A new blood test will correctly identify 96% of the persons with the disease and 94% of the persons without disease.
(a) What is the probability that a person does not have the disease?
(b) What is the probability that a person is correctly identified?

12. If 30 dice are rolled, find:
(a) the expected number of fours
(b) the probability of exactly 7 fours
(c) the probability of between 5 and 20 fours (inclusive)
(d) Sketch the graph of the distribution of fours.

13. If the probability that the 1998 War Eagle football team will lose each game next year is 0.125, determine:
(a) the probability that their first loss will be on the 7th game.
(b) the expected number of games it will take to get their first loss.

14. A box contains 8 orange balls and 2 yellow balls. Five balls are drawn from the box with replacement. Find the probability of drawing exactly 3 yellow balls.

15. A pair of dice are tossed 4 times. Find the probability of throwing a sum less than 10 more than twice.

16. A dice is rolled 20 times. The number of times 2 will come out is like the sum of draws with replacement from which number group?
(A) {0, 1} (B) {-1, 1} (C) {0, 0, 0, 0, 0, 1} (D) {-1, -1, -1, -1, -1, 1}

17. The average number of books in the homes of all LHHS students is 1000. You have selected 25 homes at random and the first two you look at have 900 books and 950 books respectively. What do you expect the mean number of books to be for the entire sample (numerical answer).

18. The scores on the first midterm follows a normal curve. The average and the SD of the midterm was 63 and 8 respectively. If 10% of the class got A's, what score approximately does Sherry have to get to obtain an A.
(A) 63 (B) 73 (C) 83 (D) 93

19. There are 52 people in a class. The average on the test was 60 points with a standard deviation of 10. If everyone above 1 SD gets an A, about how many people would receive an A?
(A) 5 (B) 8 (C) 9 (D) 10

20. A coin is tossed 25 times, what is the chance of getting 10 heads and 15 tails?
(A) 6.75 % (B) 9.74 % (C) 13.11 % (D) None of the above

21. A test was done to see how tall men between the ages 18 - 35 are. The average of the test was 65 inches and the standard deviation is 6, what percentage of the men are between 68 -77 inches tall?
(A) 28.6% (B) 47.73% (C) 56.16% (D) None of the above

22. In a chemistry class, the scores on the final follows the normal curve. given the average is 64 and the SD is 12, approximately what percentage of the students had scores within the range 46 -76?
(A) 77.46 % (B) 86.64 % (C) 87.42 % (D) None of the above

23. Which of the following pairs of events are mutually exclusive?
a. A: the odd numbers; B: the number 5
b. A: the even numbers; B: the numbers greater than 10
c. A: the numbers less than 5; B: all negative numbers
d. A: the numbers above 100; B: the numbers less than -200
e. A: negative numbers; B: odd numbers

24. One card is drawn from a standard 52 card deck. In describing the occurrence of two possible events, an Ace and a King, these two events are said to be:
(a) independent (b) mutually exclusive (c) random variables (d) randomly independent.

25 True or False? If False, correct it.
The law of large numbers makes it possible to predict long run relative frequencies but not particular chance events.

26 Among twenty-five articles, nine are defective, six having only minor defects and three having major defects. Determine the probability that an article selected at random has major defects given that it has defects.
a. 1/3 b. .25 c. .24 d. .08

27 Suppose a certain ophthalmic trait is associated with eye color. 300 randomly selected individuals are studied with results as follows:

	Eye Color			
Trait	**Blue**	**Brown**	**Other**	**Total**
Yes	70	30	20	120
No	20	110	50	180
Total	90	140	70	300

A. What is the probability that a person has blue eyes?

B. What would you expect to be the value P(having the trait and blue eyes) if eye color and trait status were independent?

C. Which of the following expressions describes the relationship between the events A = a person has brown eyes and B = a person has blue eyes? (circle the correct answer)

i. independent ii. Exhaustive

iii. simple iv. mutually exclusive

28 The checking accounts of Save-More Bank are categorized by age of account and balance in account. We are going to select an account at random from this group of 2000 accounts.

	Balance		
Age of Account	**$0-99**	**$100-499**	**>$500**
< 3 years	700	100	400
=> 3 years	200	400	200

i) Then P($500 or more given 0 - 3 years) =

a) 6/7 b) 2/3 c) 1/3 d) 1/5 e) none of these

ii) Then P[($0 - $99) or (3 years or more)] =

a) 9/10 b) 9/20 c) 3/4 d) 2/5 e) none of these

iii) Then P($100 or more) =

a) 1/4 b) 3/20 c) 11/20 d) 3/10 e) none of these

iv) What is the conditional probability that the account has a balance under $100, given that it is less than 3 years old?

a) 7/9 b) 9/20 c) 3/4 d) 7/12 e) none of these

v) Are age of account and balance in account independent at Save-More Bank? Why or why not?

vi) Suppose fourteen accounts are drawn at random from this bank. Let G be the event: "At least five accounts are less than 3 years old". State G', the complement of G.

29 The depositors at Save-More Bank are categorized by age and sex. We are going to select an individual at random from this group of 2000 depositors.

	Sex	
Age	Male	Female
30 or less	800	600
31 or more	400	200

i) Then P(Female and 30 or less) =

a) 2/5 b) 3/4 c) 3/7 d) 3/10 e) none of these

ii) Then P[Male or (31 or more)] =

a) 1/5 b) 3/10 c) 1/2 d) 7/10 e) none of these

iii) Then P(Female) =

a) 3/10 b) 2/5 c) 3/5 d) 2/3 e) none of these

iv) What is the conditional probability that the depositor drawn is 30 or less, given that he is a male?

a) 2/3 b) 7/10 c) 4/7 d) 2/5 e) none of these

v) Are age of depositor and sex of depositor independent at Save-More Bank? Why or why not?

30. An epidemiologist feels that railroads have something to do with the development of a new disease because the probability of a person's living within a mile of railroad tracks, given that he has the disease, is .80. Do you agree with him? Why or why not?

31. Among twenty-five articles eight are defective, six having only minor defects and two having major defects. Determine the probability that an article selected at random has major defects given that it has defects.

(a) .08 (b) 1/3 (c) .25 (d) .24

32. A dormitory on campus houses 200 students. 120 are male, 50 are upper division students, and 40 are upper division male students. A student is selected at random. The probability of selecting a lower division student, given the student is a female, is:

(a) 7/8 (b) 7/20 (c) 7/15 (d) 1/4 (e) 2/5

33 400 adult males with angina pectoris are classified by age and weight as follows:

	Weight in Pounds				
Age (years)	**130-149**	**150-169**	**170-189**	**>=190**	**Total**
30-39	10	20	20	40	90
40-49	10	15	50	70	145
50-59	5	15	50	40	110
60-69	5	10	15	25	55
Total	30	60	135	175	400

Using the table, find for a randomly selected individual from this population the probability that he or she:

a) Is in the age interval 40-49
b) Is in the age interval 40-49 and weighs 170-189 lb.
c) Is in the age interval 40-49 or 60-69
d) Is in the age interval 40-49 or 60-69 and weighs 150-169 lb.
e) Is in the age interval 40-49 given a weight between 150-169 lb.
f) Weighs less than 170 lb.
g) Weighs less than 170 lb. and is less than 50 years
h) Weighs less than 170 lb. given that he is less than 50 years

34 A local trade union consists of plumbers and electricians. Classified according to rank:

	Apprentice	**Journeyman**	**Master**	Totals
Plumbers	25	20	30	75
Electricians	15	40	20	75
Totals	40	60	50	150

A member of the union is selected at random. Given that the person selected is a plumber, the probability that he is a journeyman is:

a. 1/2 b. 1/3 c. 4/15 d. 2/15 e. none of these.

35 Suppose a loaded die has the following model:

Face	1	2	3	4	5	6
Probability	0.3	0.1	0.1	0.1	0.1	0.3

If this die is thrown and the top face shows an odd number,

a. What is the probability that the die shows a four?
b. What is the probability that the die shows a 1?

36 There are two urns marked H and T. Urn H contains 2 red marbles and 1 blue marble. Urn T contains 1 red and 2 blue marbles. A coin is to be tossed. If it lands heads, a marble is drawn from Urn H. If it lands tails a marble is drawn from Urn T. Find the following probabilities:

	H	**T**	
R	1/3	1/6	1/2
B	1/6	1/3	1/2
	1/2	1/2	1

a. P(heads and red) b. P(tails) c. P(red) d. P(blue) e. P(heads|red)

37 How may the standard deviation be expressed?

a. A point on a z-score scale.
b. A distance on a z-score scale.
c. An index on a squared numerical scale.
d. (a) and (c) are both correct.
e. (b) and (c) are both correct.

38 Suppose that college students are asked to identify their preferences in political affiliation (Democrat, Republican, or Independent) and in ice cream (chocolate, vanilla, or strawberry). Suppose that their responses are represented in the following two-way table (with some of the totals left for you to calculate):

	Chocolate	Vanilla	Strawberry	Total
Democrat	26.00	43	13	82
Republican	45.00	12	8	65
Independent	9.00	13	4	26
Total	80.00	68	25	173

a. What proportion of the respondents prefer chocolate ice cream?
b. What proportion of the respondents are Independents?
c. What proportion of Independents prefer chocolate ice cream?
d. What proportion of those who prefer chocolate ice cream are Independents?

39 If the manufacturer wants to keep the mean at 12.8 ounces but adjust the standard deviation so that only 1% of the bags weigh less than 12 ounces, how small does he/she need to make that standard deviation?

40 Consider the question of whether women tend to pay more for a haircut than do men. Students were asked to report the total cost of their most recent haircut. A total of 17 men and 15 women responded. Their results follow with the amounts recorded in dollars; notice that these values have been ordered.

Men	Women
0.00	11.00
4.00	12.00
4.25	14.00
6.00	15.00
7.00	15.00
7.00	15.00
8.00	15.00
8.00	15.00
8.00	18.00
10.00	18.00
10.00	20.00
10.00	20.00
10.00	25.00
12.00	25.00
15.00	50.00
15.00	
17.00	

Calculate the median haircut price for the men who responded and the median haircut price for the women who responded.

The lower and upper quartiles for the men's haircut prices are \$6.50 and \$11.00; the lower and upper quartiles for the women's haircut prices are \$15.00 and \$20.00. Use this information to construct (on the same scale) modified boxplots of the haircut prices for both sexes.

Write a few sentences comparing and contrasting the distributions of haircut prices between men and women. Indicate whether the data support the proposition than women tend to pay more for haircuts than do men. Also comment on whether every woman pays more for a haircut than does every man.

41 Suppose that a company employs five men and five women. Construct a hypothetical example which demonstrates that even though the mean salary for men is much higher than the mean salary for women, it is possible for most of the women in the company to earn more than most of the men. List five hypothetical men's salaries and five hypothetical women's salaries for which the following conditions hold:

- the mean salary for men is higher than the mean salary for women
- four of the five lowest salaries in the company belong to men
- four of the five highest salaries in the company belong to women

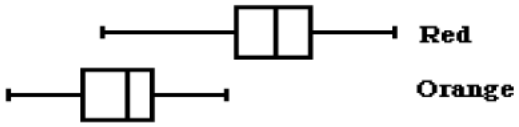
42. Circle the answer in which all entries are *resistant* statistics (value does not respond strongly to changes in a few observations, no matter how large those changes may be. It also limits the influence of outliers)
a) mode, median, mean
b) Q1, median, Q3, IQR
c) range, mean, standard deviation
d) median, mean
e) none of the above

43. Suppose that weights of bags of potato chips coming from a factory follow a normal distribution with mean 12.8 ounces and standard deviation .6 ounces.
   a. What proportion of bags weigh more than 12 ounces?
   b. What proportion of bags weigh between 13 and 14 ounces?
   c. Determine the weight such that 12.5% of the bags weigh more than that weight.

44 Suppose that 80% of all Pennsylvania residents eat turkey on Thanksgiving. Suppose further that you plan to select a simple random sample of 300 Pennsylvania residents and to determine the proportion of them who eat turkey on Thanksgiving.
   a. Is 80% a parameter or a statistic?
   b. What symbol have we used to represent it?
   c. Determine the probability that less than three-fourths of the sample eat turkey on Thanksgiving.

45 A study considers the variable: The weight of an automobile. Circle the true answer:
   a) The variable is automobile and it is a measurement variable.
   b) The variable is automobile and it is a categorical variable.
   c) The variable is weight and it is a measurement variable.
   d) The variable is weight and it is a categorical variable.
   e) None of the above.

46. The two boxplots to the right show the distributions of red and orange M&Ms from the 17 bags of M&Ms which were inspected by 1996-7 AP Stats class. Write a paragraph or two to a knowledgeable statistician at The Mars Candy Company explaining what might be expected about the number of red and orange M&Ms in an 18th bag taken from the same stock.



47 The formula for a z-score is $z = \frac{x_i - \bar{x}}{s_x}$
   a. Explain the purpose of z-scores.
   b. In detail, explain how the formula actually fulfills your answer in part a.
   c. The best male long jumpers for State College since 1973 have averaged a jump of 263.0 inches with a standard deviation of 14.0 inches. The best female long jumpers have averaged 201.2 inches with a standard deviation of 7.7 inches. Which athlete is more impressive within their class, a male with a jump of 275 inches or a female with a jump of 207 inches? Prove your answer with appropriate calculations.

48. A gambler has a special coin that has been flipped so many times that he knows over the long, long run it lands heads 55 out of 100 times.
   a. Determine the probability of a sample of 20 flips having 6 or fewer heads.
   b. Determine the probability of a sample of 20 flips having between 8 and 14 heads, inclusive.

49 Circle the best answer. An enormous vat of red and blue marbles is 50% red. As more and more marbles are pulled from the vat:
   a. The percentage of blue marbles pulled from the vat tends to get closer to 50% and the number of red and blue marbles pulled from the vat start to even out.
   b. The percentage of blue marbles pulled from the vat tends to get closer to 50% and the number of red marbles pulled from the vat tends to move further away from half of the marbles.
   c. The percentage of blue marbles pulled from the vat tends to move further away from 50% and the number of red and blue marbles pulled from the vat start to even out.
   d. The percentage of blue marbles pulled from the vat tends to move further away from 50% and the number of red marbles pulled from the vat tends to move further away from half of the marbles.

50 In a game of chance, you win if a roll of a die is either a 1 or a 6. You decide to roll the die 45 times. X is the number of times you win.

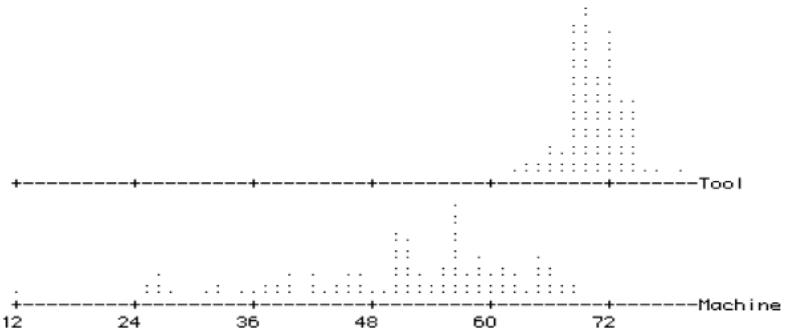
a. Is this a binomial setting? Support your answer.

b. Showing your work, determine the probability of winning EXACTLY 13 times.

c. Approximate the probability of winning 13 or more times. Show your work. Do as a normal approximation.

d. Is this a random phenomenon? If not explain why. If it is, explain why and include a sketch of part or all of a plot labeled with some specific values (numbers) which supports your definition of random phenomenon.

51 Two companies, The Tool Company and The Machine Company, have made prototype devices to automatically throw softballs a fixed distance. Below are the results of 100 throws for each device. Each device was set to throw each ball a distance of 55 feet.



a. Fill in the chart below with comparisons of The Tool Company and The Machine Company data for the six features that are often of interest when analyzing a distribution. Do this by simply looking at the dotplots. Do not do any counting or calculations.

Feature	Compare The Tool vs. The Machine Company for this Feature
Center	
Shape	
Spread	
Outliers	

b. Each company argued that its prototype is better. In a sentence or two write what you think each company's argument was?

The Tool Company:

The Machine Company:

c. Below Minitab's descriptive statistics for the Machine Company's data. Formally determine if The Machine Company's minimum data value is an outlier. Show your work.

	N	MEAN	MEDIAN	TRMEAN	STDEV	SEMEAN
Machine	100	51.25	53.00	51.92	11.64	1.16

	MIN	MAX	Q1	Q3
Machine	12.00	68.00	45.25	59.00

d. Demonstrate your understanding of the empirical rule for three standard deviations by applying it Machine Company's data and explaining whether or not the rule seems to hold reasonably well. You can use the descriptive statistics displayed above. Show your work.

e. Draw, without counting, a very rough boxplot of The Tool Company's data by looking at the dotplot of that data. There is no need to label values or worry about possible outliers, simply make the relative sizes of the parts clear. Careful: do not use the Machine Company's data.

52. Circle the correct answer:

a. An observational study can show a causal relationship.

b. An experimental study can show a causal relationship.

c. The closer the value of r^2 is to 1, the more evidence there is of a causal relationship between the explanatory variable and the response variable.

d. Both a & b are true.

e. Both b & c are true.

53. Circle the correct answer. The design of an experiment is biased if:
a. A sample has large variability
b. The center of a sample is not close to the population center
c. All samples have large variability
d. The centers of all samples are on the same side of the population center

54 An educational researcher wants to compare the effectiveness of using different computer set-ups to help in reading comprehension. First she gives 12 students a reading comprehension test. Then she randomly assigns them to computers with different set-ups. The computers have one of two different size monitors (13 inch and 17 inch) and they display the text at one of three different speeds (20 words per minute, 40 words per minute & 80 words per minute). She conducts an experiment and then retest the students and compares the increase in reading ability in each group.
a. What are the factors in this experiment?
b. List the treatments in this experiment.
c. Why is this study called an experiment?
d. The 12 students are listed below along with a set of random digits.

Anderson (01), Baxer(02), Cote(03), Fernandex(04),Frank(05), Hicks(06), Klassen(07), Mihalko(08), Rustagi(09), Tomis(10),Ulee(11), Zeg(12)

33063 41842 81068 71035 09001 03367
49497 54580 81507 27102 56027 55892

Demonstrate your understanding of simple random sampling by using the random digits to determine which of the 12 would be the first three randomly assigned. Briefly make it clear how your selections were made.

55. A scientist claims he has performed an experiment in which he both 1) uses a block design, and 2) uses an SRS of ENTIRE population. Explain why this is not possible. Illustrating your point with an example is acceptable.

56. Bill, a statistician, said that the temperature was so cold yesterday at the North pole that it was 3.5 standard deviations BELOW normal. He said that this was a statistically significant event. Clearly demonstrating your understanding of the terms statistically significant and including numeric support to explain if he was correct.

57. Suppose that jumps by Olympic men high jumpers have a normal distribution with mean 2.12 meters and standard deviation 0.12 meters; women's jumps have a normal distribution with mean 1.80 meters and standard deviation 0.09 meters. A man and woman Olympic high jumper are picked at random.
(a) What is the probability the sum of their jumps is over 4 meters?
(b) What is the probability that the man jumped higher than the woman?

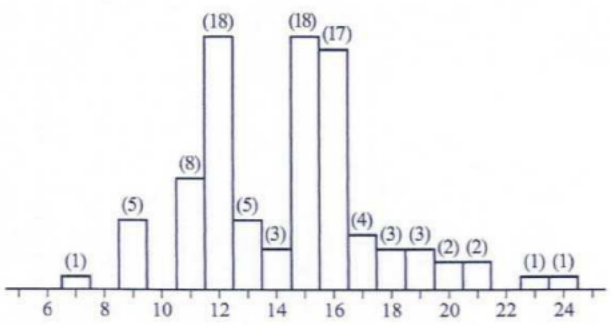
58. The probability distribution for the number of magazine subscriptions to which college students subscribe is as follows:

Number of subscriptions	0	1	2	3	4
Relative frequency	0.12	0.41	0.25	0.15	0.07

(a) Calculate and give a brief interpretation of the mean of this probability distribution.
(b) In a random sample of 10 college students, there are a total of 20 magazine subscriptions. A new random sample of 50 students is planned. How do you expect the average number of subscriptions for this new sample to compare to that of the first sample? Explain.
(c) Find the median of the above distribution, where the median M is defined to be a value such that $P(x \geq M) \geq 0.5$ and $P(x \leq M) \geq 0.5$.

59. A survey is conducted to determine the proportion of young adults who earn less than $15,000. Two sampling methods are being considered. Method A involves standing on a downtown street corner and randomly picking several young adults to interview every 10 minutes during a 12-hour period. Method B involves posting the question on several popular young adult websites; the viewer simply has to click on one of two possible answers to participate.
(a) State a possible source of bias for Method A, and describe how it may affect the results.
(b) State a possible source of bias for Method B, and describe how it may affect the results.
(c) How many young adults should be interviewed to estimate the proportion of the young adult population who earn less than $15,000 to within ±0.03 with 90 percent confidence?

60. A high school math department conducts a study to determine whether a new AP Statistics textbook will lead to higher AP exam scores than the textbook currently in use. Two AP statistics classes are scheduled, each teacher has 18 students, and it is randomly decided which class will use which book. At the end of the academic year, the 18 students in each class take the AP Statistics, exam, and the department notes the scores.
    (a) Identify the response variable, the treatments, and the experimental units.
    (b) Was randomization properly used? Explain.
    (c) Was replication properly used? Explain.
    (d) Teacher is a confounding variable. Explain.
61. An apartment building elevator has a carrying capacity of 2,000 lb. Suppose the men living in the building have a mean weight of 185 lb with a standard deviation of 15 lb, the women have a mean weight of 135 lb with a standard deviation of 10 lb, and both weight distributions are normal.
    (a) What are the mean μ_{sum} and standard deviation σ_{sum} of the combined weight of seven men and five women assuming all weights are independent?
    b) What is the probability that the seven men and five women will overload the elevator?
62. A set of 91 scores has the following histogram:



    (a) Draw a boxplot of the above data.
    (b) What feature does the histogram show that is missed by the boxplot?
    (c) What feature is more clearly distinguished in the boxplot than in the histogram?
63. A psychologist plans a study to determine if high school students read faster with their right eye or left eye alone. Fifty high school students are randomly selected for the study.
    (a) Describe a randomization process and an inference procedure for the study to be conducted with a completely randomized design.
    (b) Describe a randomization process and an inference procedure for the study to be conducted with a matched pairs design.
64. In a random sample of 500 teenagers, 224 answered *yes* to the question "Do you think that you are making a positive difference in your community?" Which of the following is a 90 percent confidence interval for the proportion of all teenagers who would answer yes to this question?
    (A) $0.448 \pm 1.645\sqrt{\dfrac{0.448(0.552)}{500}}$
    (B) $0.448 \pm 1.96\sqrt{\dfrac{0.448(0.552)}{500}}$
    (C) $0.448 \pm 1.645\sqrt{\dfrac{0.5(0.5)}{500}}$
    (D) $224 \pm 1.645\sqrt{500(0.448)(0.552)}$
    (E) $224 \pm 1.96\sqrt{500(0.448)(0.552)}$