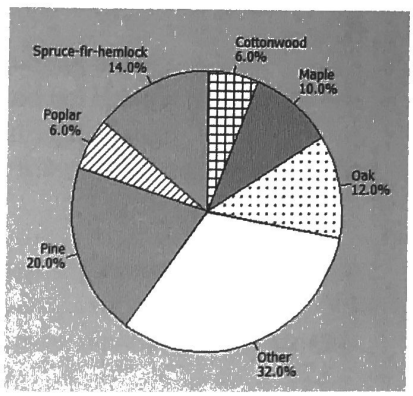
Review

Test 1A AP Statistics Name: Key

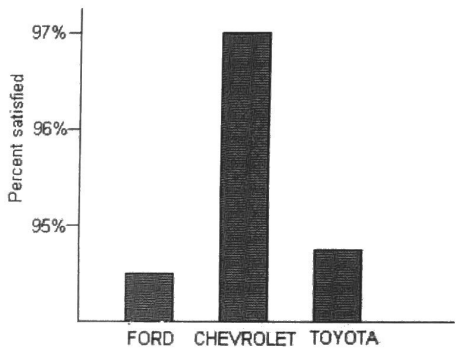
Part 1: Multiple Choice. *Circle the letter corresponding to the best answer.*

1. The pie chart at right describes the distribution of state tree types for the 50 states in the United States. The category "Other" include all trees that are the state tree for two or fewer states. Which of the following conclusions can we draw from this chart?



(a) Some states have not designated a "state tree."
(b) The cottonwood is the state tree for 12 states.
(c) Taken together, oak, pine, and maple are the state trees for more than half the states.
(d) There are 10 states that have designated a pine as their state tree.
(e) There is no state that has designated the Eastern Red Cedar as its state tree.

2 The following bar graph gives the percent of owners of three brands of trucks who are satisfied with their truck. From this graph, we may conclude that



(a) owners of other brands of trucks are less satisfied than the owners of these three brands.
(b) Chevrolet owners are substantially more satisfied than Ford or Toyota owners.
(c) there is very little difference in the satisfaction of owners for the three brands.
(d) Chevrolet probably sells more trucks than Ford or Toyota.
(e) a pie chart would have been a better choice for displaying these data.

3. Here are the IQ test scores of 10 randomly chosen fifth-grade students:

145 139 126 122 125 130 96 110 118 118

To make a stemplot of these scores, you would use as stems
(a) 0 and 1.
(b) 09, 10, 11, 12, 13, and 14.
(c) 96, 110, 118, 122, 125, 126, 130, 139, and 145.
(d) 0, 2, 3, 5, 6, 8, 9.
(e) None of the above is a correct answer.

4. If a distribution is skewed to the right, which of the following is true?
(a) The mean must be less than the median.
(b) The mean and median must be equal.
(c) The mean must be greater than the median.
(d) The mean is either equal to or less than the median,
(e) It's impossible to tell which of the above statements is true without seeing the data.

7. A small company that prints custom t-shirts has 6 employees, one of whom is the owner and manager. Suppose the owner makes $120,000 per year and the other employees make between $40,000 and $50,000 per year. One day, the owner decides to give himself a $30,000 raise. Which of the following describes how the company's mean and median salaries would change?
   (a) The mean and median would both increase by $5,000.
   (b) The mean would increase by $5,000 and the median would not change.
   (c) The mean would increase by $6,000 and the median would not change.
   (d) The median would increase by $6,000 and the mean would not change.
   (e) The mean would increase by $6,000, but we cannot determine the change in the median without more information.

8. The mean speed of vehicles in the "cars only" lanes of the New Jersey turnpike is 68 miles per hour. The mean speed of vehicles in the "any vehicle" lanes is 64 miles per hour. What must be true about the mean speed of all vehicles on the turnpike, assuming these are the only types of lanes?
   (a) It could be any number between 64 and 68 miles per hour.
   (b) It must be larger than the median speed.
   (c) It must be larger than 66 miles per hour.
   (d) It must be 66 miles per hours.
   (e) We don't have enough information to draw any conclusion about the mean speed of all vehicles.

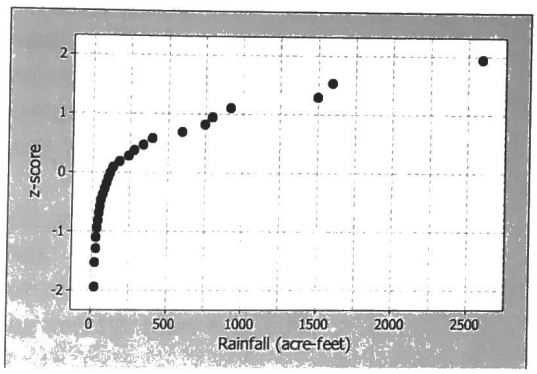
9. The mean birth weight of infants born at a certain hospital in the month of April was 128 oz. with a standard deviation of 10.2 oz. Which of the following is a correct interpretation of standard deviation?
   (a) All the infants born in April weighed between 117.8 oz. and 138.2 oz.
   (b) About half the infants born in April weighed between 117.8 oz. and 138.2 oz.
   (c) The difference between the mean weight and the median weight of infants born in April was 10.2 oz.
   (d) The distance between the weight of each infant bon in April and the mean weight was, on average, about 10.2 oz.
   (e) The mean weight of infants born in subsequent months is likely to be within 10.2 oz. of the mean weight in April.

10. A medical researcher collects health data on many women in each of several countries. One of the variables measured for each woman in the study is her weight in pounds. The following list gives the five-number summary for the weights of adult women in one of the countries.

    Country A: 92, 110, 120, 160, 240

    About what percent of Country A women weigh between 110 and 240 pounds?
    (a) 50%
    (b) 65%
    (c) 75%
    (d) 85%
    (e) 95%

6. The graph at right is a Normal probability plot for the amount of rainfall (in acre-feet) obtained from 26 randomly selected clouds that were seeded with silver oxide. Which of the following statements about the shape of the rainfall distribution is true?

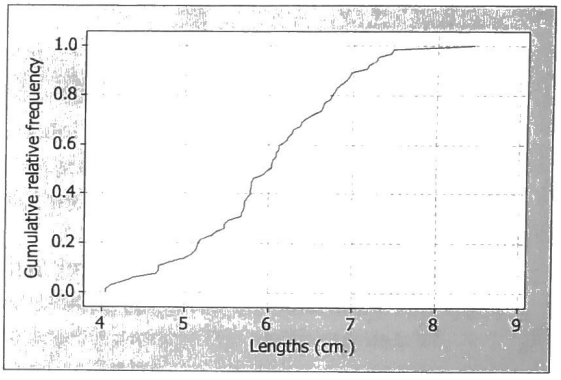


(a) The distribution is Normal.
(b) The distribution is approximately Normal.
(c) The distribution is skewed left.
(d) The distribution has no potential outliers.
(e) The distribution is skewed right.

7. The distribution of the time it takes for different people to solve a certain crossword puzzle is strongly skewed to the right, with a mean of 30 minutes and a standard deviation of 15 minutes. The distribution of z-scores for those times is

(a) Normally distributed, with mean 30 and standard deviation 15.
(b) Skewed to the right, with mean 30 and standard deviation 15.
(c) Normally distributed, with mean 0 and standard deviation 1.
(d) Skewed to the right, with mean 0 and standard deviation 1.
(e) Skewed to the right, but the mean and standard deviation cannot be determined without more information.

8. The cumulative relative frequency graph at right shows the distribution of lengths (in centimeters) of fingerlings at a fish hatchery. The interquartile range for this distribution is approximately:



(a) 0.18 to 0.85 centimeters
(b) 5 to 7 centimeters
(c) 5.5 to 6.7 centimeters
(d) 1.2 centimeters
(e) 2 centimeters

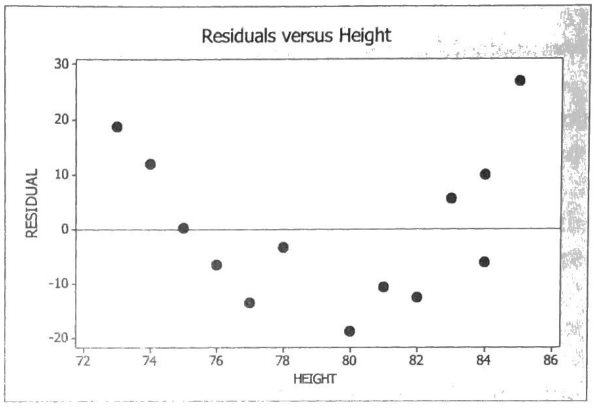
9. Which of the following properties is true for all Normal density curves?

I. They are symmetric.
II. The curve reaches its peak at the mean.
III. 95% percent of the area under the curve is within one standard deviation of the mean.

(a) I only
(b) II only
(c) I and II only
(d) I and III only
(e) All three statements are correct.

6. A set of data describes the relationship between the size of annual salary raises and the performance ratings for employees of a certain company. The least squares regression equation is $\hat{y} = 1400 + 2000x$ where y is the raise amount (in dollars) and x is the performance rating. Which of the following statements is *not necessarily* true?
   (a) For each one-point increase in performance rating, the raise will increase on average by $2000.
   (b) The actual relationship between salary raises and performance rating is linear.
   (c) A rating of 0 will yield a predicted raise of $1400.
   (d) The correlation between salary raise and performance rating is positive.
   (e) If the average performance rating is 1.2, then the average raise is $3800.

7. A least-squares regression line for predicting weights of basketball players on the basis of their heights produced the residual plot below.



What does the residual plot tell you about the linear model?
   (a) A residual plot is not an appropriate means for evaluating a linear model.
   (b) The curved pattern in the residual plot suggests that there is no association between the weight and height of basketball players.
   (c) The curved pattern in the residual plot suggests that the linear model is not appropriate.
   (d) There are not enough data points to draw any conclusions from the residual plot.
   (e) The linear model is appropriate, because there are approximately the same number of points above and below the horizontal line in the residual plot.

5. To test the effect of music on productivity, a group of assembly line workers are given portable mp3 players to play whatever music they choose while working for one month. For another month, they work without music. The order of the two treatments for each worker is determined randomly. This is

(a) an observational study.
(b) a matched pairs experiment.
(c) a completely randomized experiment.
(d) a block design, but not a matched pairs experiment.
(e) impossible to classify unless more details of the study are provided.

6. A survey was done in the town of Mechanicsville to estimate the proportion of cars that are red and made by companies based in Japan. A random sample of 25 cars from a student parking lot at Lee-Davis High School was taken. Which of the following statements is <u>not</u> correct?

(a) This sample may not be representative of the cars in Mechanicsville because mainly students park at Lee-Davis High School.
(b) If the particular parking space is vacant, we can simply select another parking space at random because it is unlikely that a space being vacant is related to the color or manufacturer of the car.
(c) It would an error to simply select the first 25 parking spaces in the lot closest to the auditorium because there are a number of parking spaces there reserved for Drivers Ed vehicles, whose primary color is white.
(d) A different team doing the sampling independently would obtain different answers for their sample proportions.
(e) The results will be the same regardless of the time of day that the sample is taken.

7. A nutritionist wants to study the effect of storage time (6, 12, and 18 months) on the amount of vitamin C present in freeze dried fruit when stored for these lengths of time. Vitamin C is measured in milligrams per 100 milligrams of fruit. Six fruit packs were randomly assigned to each of the three storage times. The treatment, experimental unit, and response are respectively:

(a) A specific storage time, amount of vitamin C, a fruit pack
(b) A fruit pack, amount of vitamin C, a specific storage time
(c) Random assignment, a fruit pack, amount of vitamin C
(d) A specific storage time, a fruit pack, amount of vitamin C
(e) A specific storage time, the nutritionist, amount of vitamin C

8. A researcher observes that, on average, the number of divorces in cities with Major League Baseball teams is larger than in cities without Major League Baseball teams. The most plausible explanation for this observed association is that the

(a) presence of a Major League Baseball team causes the number of divorces to rise (perhaps husbands are spending too much time at the ballpark).
(b) high number of divorces is responsible for the presence of Major League Baseball teams (more single men means potentially more fans at the ballpark, making it attractive for an owner to relocate to such cities).
(c) association is due to the presence of a lurking variable (Major League teams tend to be in large cities with more people, hence a greater number of divorces).
(d) association makes no sense, since many married couples go to the ballpark together.

(e) observed association is purely coincidental. It is implausible to believe the observed association could be anything other than accidental.

9. Control groups are used in experiments in order to
(a) control the effects of outside variables on the outcome.
(b) control the subjects of a study to ensure that all participate equally.
(c) guarantee that someone other than the investigators, who have a vested interest in the outcome, controls how the experiment is conducted.
(d) achieve a proper and uniform level of randomization.
(e) reduce the variability in results.

10. A survey is to be administered to recent graduates of a certain nursing school in order to compare the starting salaries of women and men. For a random sample of graduates, three variables are to be recorded: sex, starting salary, and area of specialization. Which of the follow best describes a conclusion that can be drawn from this study?
(a) Whether being female causes graduates of this nursing school to have lower (or higher) starting salaries than males.
(b) Whether being female causes graduates in this sample to have lower (or higher) starting salaries than males.
(c) Whether choosing certain area of specialization causes females graduates of this nursing school to have lower (or higher) starting salaries than males.
(d) Whether there is an association between sex and starting salary among graduates of this nursing school.
(e) Whether there is an association between sex and starting salary at all nursing schools similar to this one.

6. Event A occurs with probability 0.2. Event B occurs with probability 0.8. If A and B are disjoint (mutually exclusive), then
(a) $P(\text{A or B}) = 1.0$.
(b) $P(\text{A and B}) = 0.16$.
(c) $P(\text{A and B}) = 1.0$.
(d) $P(\text{A or B}) = 0.16$.
(e) both (a) and (b) are true.

7. If $P(\text{A}) = 0.24$ and $P(\text{B}) = 0.52$ and A and B are independent, what is $P(\text{A or B})$?
(a) 0.1248
(b) 0.28
(c) 0.6352 .24 + .52 − (.24)(.52)
(d) 0.76
(e) The answer cannot be determined from the information given.

8. People with type O-negative blood are universal donors. That is, any patient can receive a transfusion of O-negative blood. Only 7.2% of the American population has O-negative blood. If 10 people appear at random to give blood, what is the probability that at least 1 of them is a universal donor?
(a) 0
(b) 0.280 $1 - (.9280)^{10}$
(c) 0.526
(d) 0.720
(e) 1

9. Of people who died in the United States in a recent year, 86% were white, 12% were black, and 2% were Asian. (We will ignore the small number of deaths among other races.) Diabetes caused 2.8% of deaths among whites, 4.4% among blacks, and 3.5% among Asians. The probability that a randomly chosen death was due to diabetes is about
(a) 0.96.
(b) 0.107. (.86)(.028) + (.12)(.044) + (.02)(.035)
(c) 0.042.
(d) 0.038.
(e) 0.030.

10. In your top dresser drawer are 6 blue socks and 10 grey socks, unpaired and mixed up. One dark morning you pull two socks from the drawer (without replacement, of course!). What is the probability that the two socks match?
(a) 0.075 (b) 0.375 (c) 0.450 (d) 0.500 (e) 0.550

$P(B \cap B) + P(G \cap G) = \left(\frac{6}{16} \cdot \frac{5}{15}\right) + \left(\frac{10}{16} \cdot \frac{9}{15}\right)$

5. X and Y are independent random variables, and a and b are constants. Which one of the following statements is true?

(a) $\sigma_{X+Y} = \sigma_X + \sigma_Y$

(b) $\text{Var}(X-Y) = \text{Var}(X) + \text{Var}(Y)$

(c) $\text{Var}(a+bX) = b\text{Var}(X)$

(d) $\sigma_{X-Y} = \sigma_X - \sigma_Y$

(e) $\text{Var}(X+Y) = \sqrt{\text{Var}(X^2) + \text{Var}(Y^2)}$

6. Let the random variable X represent the profit made on a randomly selected day by a certain store. Assume that X is Normal with mean \$360 and standard deviation \$50. What is $P(X >$ \$400)?

z-score: $\frac{400-360}{50} = .8$ Table A! →

(a) 0.2119 (b) 0.2881 (c) 0.5319 (d) 0.7881 (e) 0.8450

7. A dealer in the Sands Casino in Las Vegas selects 40 cards from a standard deck of 52 cards. Let Y be the number of red cards (hearts or diamonds) in the 40 cards selected. Which of the following best describes this setting?

(a) Y has a binomial distribution with $n = 40$ observations and probability of success $p = 0.5$.

(b) Y has a binomial distribution with $n = 40$ observations and probability of success $p = 0.5$, provided the deck is shuffled well.

(c) Y has a binomial distribution with $n = 40$ observations and probability of success $p = 0.5$, provided that after selecting a card it is replaced in the deck and the deck is shuffled well before the next card is selected.

(d) Y has a geometric distribution with $n = 40$ observations and probability of success $p = 0.5$.

(e) Y has a geometric distribution with n = 52 observations and probability of success $p = 0.5$.

8. In a large population of college students, 20% of the students have experienced feelings of math anxiety. If you take a random sample of 10 students from this population, the mean and standard deviation of the number of students in the sample who have experienced math anxiety is:

(a) $\mu = 1.6$; $\sigma = 1.414$

(b) $\mu = 1.6$; $\sigma = 1.265$

(c) $\mu = 2$; $\sigma = 1.6$

(d) $\mu = 2$; $\sigma = 1.414$

(e) $\mu = 2$; $\sigma = 1.265$

$\mu_X = (10)(.2) = 2$

$\sigma_X = \sqrt{(10)(.2)(.8)} = 1.265$

Extra problem on optional material (text pp30-32)

To help consumers make informed decisions about health care, the government releases data about patient outcomes in hospitals. You want to compare Hospital A and Hospital B, which serve your community. Not all surgery cases are equally serious, however. Patients are classified as being in either "poor" or "good" condition before surgery. Here are the data broken down by patient condition.

	Good Condition	
	Hospital A	Hospital B
Died	6	8
Survived	594	592
Total	600	600

	Poor Condition	
	Hospital A	Hospital B
Died	57	8
Survived	1443	192
Total	1500	200

(a) Calculate the percent of patients in good condition who survive at Hospital A and Hospital B. Do likewise for the patients in poor condition.

good condition - Hosp A = 99%. Hosp B = 98.6%.

poor condition - Hosp A = 96.2%. Hosp B = 96%.

(b) Use the tables above to create a two-way table that shows the relationship between which hospital a patient goes to and survival status (that is, combine the results for good and bad condition).

	Hos. A	Hos. B
Died	63	16
survived	2037	784

(c) Calculate the percent of patients that survive at each hospital.

A: 97%.

B: 98%.

(d) Explain your findings in (a) and (c).

While a slightly higher percentage survived at Hospital B, condition upon arriving is important. Many more patients were in poor condition at Hosp. A before surgery.

14. "Normal" body temperature varies by time of day. A series of readings was taken of the body temperature of a subject. The mean reading was found to be 36.5°C with a standard deviation of 0.3°C. If you wanted to convert the temperatures to the Fahrenheit scale, what would the new mean and standard deviation be? (Note: °F = °C(1.8) + 32).

Mean = $36.5(1.8) + 32 = 97.7°F$

SD = $0.3(1.8) = 0.54°F$

A local post office weighs outgoing mail and finds that the weights of first-class letters is approximately Normally distributed with a mean of 0.69 ounces and a standard deviation of 0.16 ounces.

15. What is the 60th percentile of first-class letter weights?

z for 60th % is .25

$= .69 + .25(0.16)$

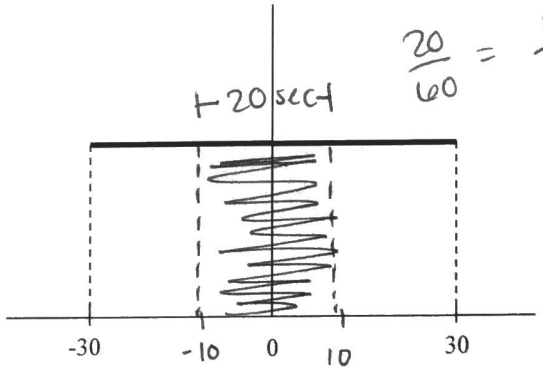
$= .73$ oz

16. First-class letters weighing more than 1 ounce require additional postage. What proportion of first-class letters at this post office require additional postage?

$z = \frac{1 - 0.69}{.16} = 1.94$ $\quad 1 - .9738 = .0262$

17. Old-fashioned mechanical alarm clocks were not very accurate about when the alarm went off. The density curve below describes the distribution of times a certain alarm clock went off. The scale on the x-axis represents when the alarm went off, in seconds, before (negative) or after (positive) the alarm was set to go off. What proportion of the time did the alarm go off within 10 seconds of the time it was set for? Shade the appropriate area on the graph to show how you found the answer.

$\frac{20}{60} = \frac{1}{3} = 33.3\%$



One weekend, a statistician notices that some of the cars in his neighborhood are very clean and others are quite dirty. He decides to explore this phenomenon, and asks 15 of his neighbors how many times they wash their cars each year and how much they paid in car repair costs last year. His results are in the table below:

	Mean	Standard deviation
x = number of car washes per year	6.4	3.78
y = repairs costs for last year	$955.30	$323.50

The correlation for these to two variables is $r = -0.71$

16. Find the equation of the least-squares regression line (with y as the response variable).

$$\text{Slope} = b = r\left(\frac{s_y}{s_x}\right) = -.71\left(\frac{323.5}{3.78}\right) = -60.763$$

$$y\text{-int} = a = \bar{y} - b\bar{x} = 955.3 - (-60.763)(6.4) = 1344.18$$

$$\hat{y} = 1344.18 - 60.763x$$

17. What percentage of the variation in repair costs can be explained by the number of times per year a car is washed?

$r^2 = .5041$ or 50.41%

18. Based on these data, can we conclude that washing your car frequently will reduce repair costs? Explain.

No. This was not a controlled experiment so there could be some confounding variables, for example changing the car's oil.

Part 2: Free Response

Show all your work. Indicate clearly the methods you use, because you will be graded on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

11. Read the following brief article about aspirin and alcohol.

> **Aspirin may enhance impairment by alcohol**
> Aspirin, a long time antidote for the side effects of drinking, may actually enhance alcohol's effect, researchers at the Bronx Veterans' Affairs Medical Center say. In a report on a study published in the *Journal of the American Medical Association*, the researchers said they found that aspirin significantly lowered the body's ability to break down alcohol in the stomach. As a result, five volunteers who had a standard breakfast and two extra-strength aspirin tablets an hour before drinking had blood alcohol levels 30 percent higher than each had when they drank alcohol alone. Each volunteer consumed the equivalent of a glass and a half of wine.
>
> That 30 percent could make the difference between sobriety and impairment, said Dr. Charles S. Lieber, medical director of the Alcohol Research and Treatment Center at the Bronx center, who was co-author of the report with Dr. Risto Roine.

(a) Explain why this is an experiment and not an observational study.

experiment because treatments are imposed on subjects

(b) Identify the explanatory and response variables.

explanatory: aspirin ~~tamp~~ consumption

response: blood alcohol content

(c) Identify the experimental design used in this study. Justify your answer.

Matched pair experiment

(d) In the second sentence above is the phrase, "…researchers said they found that aspirin significantly lowered the body's ability to break down alcohol…" What is the statistical meaning of the word "significantly" in the context of this study?

Significantly means that the difference found in the subjects BAC was large enough that it was not just coincidence.

(e) This was a controlled experiment. Describe how it was controlled and explain the purpose of doing so. Each subject acted as its own control- drinking alcohol alone & with aspirin- to reduce/eliminate confounding variables.

12. High blood pressure adds to the workload of the heart and arteries and may increase the risk of heart attacks. If not treated, this condition can also lead to heart failure, kidney failure, or stroke. We wish to test the effectiveness of Angiotensin-converting enzyme (ACE) inhibitors as a treatment for high blood pressure.

(a) It is well known that men and women may react differently to common cardiovascular drug treatments. What sort of experimental design would you choose for this study, and why?

randomized block design by gender.

reduce diff. in males / females

(b) Explain why an experiment involving 600 men and 500 women is preferable to one involving 60 men and 50 women.

larger # subjects - increased replication - decreases the impact of random variation on experiment results.

(c) Assume that 600 men and 500 women suffering from high blood pressure are available for the study. Describe a design for this experiment. Be sure to include a description of how you assign individuals to the treatment groups.

① create 2 blocks of 600 men / 500 women

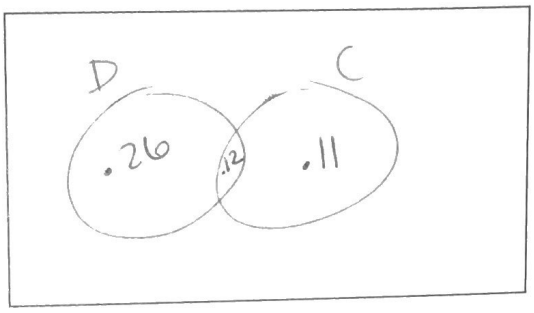
② in each block assign men # 001-600 and women 001-500.

③ choose 3 digit #s from random # table until you have 300 men; follow same procedure & select 250 women. These will be treated with ACE, remaining subjects will receive placebo.

④ compare blood pressure changes blw ACE and control group

12. Meadowbrook School surveys the families of its students and determines the following: if a family is chosen at random, the probability that they own a dog is 0.38, the probability they own a cat is 0.23, and the probability they own both a dog and a cat is 0.12.

(a) Let D = randomly-chosen family owns a dog, and C = randomly-chosen family owns a cat. Sketch a Venn diagram or two-way table that summarizes the probabilities above.



(b) Find each of the following.

i. The probability that a randomly-selected family owns a dog or a cat.

$P(D \cup C) = .49$

ii. The probability that a randomly-selected family owns a dog or doesn't own a cat.

$P(D \cup C^c) = .89$

ii. The probability that a randomly-selected family doesn't own a dog and doesn't own a cat.

$P(D^c \cap C^c) = .51$

13. Suppose your school is in the midst of a flu epidemic. The probability that a randomly-selected student has the flu is 0.35, and the probability that a student who has the flu also has a high fever is 0.90. But there are other illnesses making the rounds, and the probability that a student who doesn't have the flu does have a high fever (as a result of some other ailment) is 0.12. Suppose a student walks into the nurse's office with a high fever. What is the probability that she has the flu?

$$P(\text{Flu} \mid \text{Fever}) = \frac{P(\text{Flu} \cap \text{Fever})}{P(\text{Fever})} = \frac{(.35)(.9)}{(.35)(.9) + (.65)(.12)}$$

$$= .802$$

11. ACT scores for the 1,171,460 members of the 2004 high school graduating class who took the test closely followed the Normal distribution with mean 20.9 and standard deviation 4.8. Choose two students independently and at random from this group.

(a) What is the expected difference in their scores?

0

(b) What is the standard deviation of the difference in their scores?

$\sqrt{(4.8)^2+(4.8)^2} = 6.788$

(c) Find the probability that the difference in the two students' scores is greater than 6.

$P(X>6) = 2 \cdot P(Z>.88) = .3789$

12. A professional soccer player succeeds in scoring a goal on 84% of his penalty kicks. Assume that the success of each kick is independent.

(a) In a series of games, what is the probability that the first time he <u>fails</u> to score a goal is on his fifth penalty kick?

$= (.84)^4(.16)$

$= .0797$

(b) What is the probability that he scores on 5 or fewer of his next 10 penalty kicks?

$P(X \leq 5) = \text{binomcdf}(10, .84, 5)$

(c) Suppose that our soccer player is out of action with an injury for several weeks. When he returns, he only scores on 5 of his next 10 penalty kicks. Is this evidence that his success rate is now less than 84%? Explain.

If the soccer player's success rate were still 84%, the probability that he would score 5 or few goals in 10 penalty kicks is 0.0130. This is so low, we have enough evidence that his penalty rate is below 84%.