Practice Test

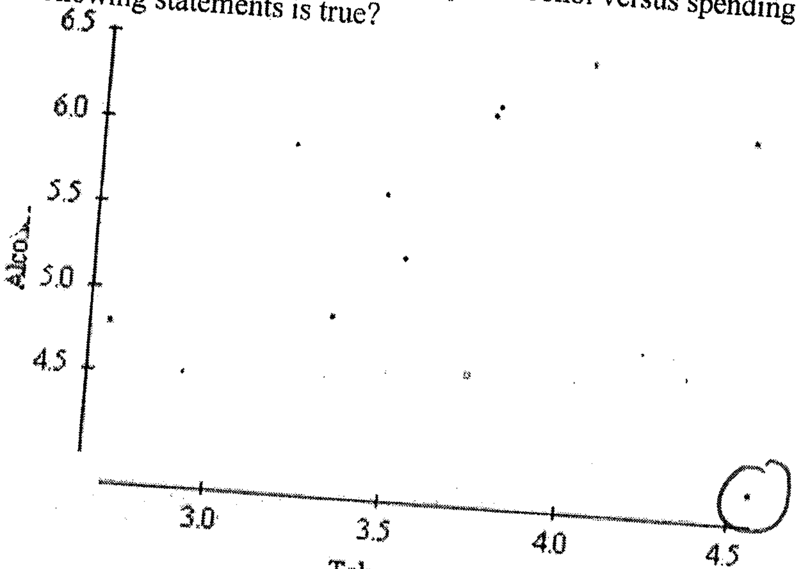
Name Key

Date ______ Hour ______

1. A school guidance counselor examines the number of extracurricular activities that students do and their grade point average. The guidance counselor says, "The evidence indicates that the correlation between the number of extracurricular activities a student participates in and his or her grade point average is close to zero." A correct interpretation of this statement would be that

(a) active students tend to be students with poor grades, and vice versa.
(b) students with good grades tend to be students who are not involved in many extracurricular activities, and vice versa.
(c) students involved in many extracurricular activities are just as likely to get good grades as bad grades; the same is true for students involved in few extracurricular activities.
(d) there is no linear relationship between number of activities and grade point average for students at this school.
(e) involvement in many extracurricular activities and good grades go hand in hand.

2. The British government conducts regular surveys of household spending. The average weekly household spending (in pounds) on tobacco products and alcoholic beverages for each of 11 regions in Great Britain was recorded. A scatterplot of spending on alcohol versus spending on tobacco is shown below. Which of the following statements is true?



Outliers in the x direction are influential. they pull the line in its direction

(a) The observation (4.5, 6.0) is an outlier.
(b) There is clear evidence of a negative association between spending on alcohol and tobacco.
(c) The equation of the least-squares line for this plot would be approximately $\hat{y} = 10 - 2x$.
(d) The correlation for these data is $r = 0.99$.
(e) The observation in the lower-right corner of the plot is influential for the least-squares line.

3. The fraction of the variation in the values of y that is explained by the least-squares regression of y on x is
(a) the correlation.
(b) the slope of the least-squares regression line.
(c) the square of the correlation coefficient. (coefficient of determination)
(d) the intercept of the least-squares regression line.
e) the residual.

% of variation of the response variable explained by x (explanatory variable)

4. An AP Statistics student designs an experiment to see whether today's high school students are becoming too calculator dependent. She prepares two quizzes, both of which contain 40 questions that are best done using paper-and-pencil methods. A random sample of 30 students participates in the experiment. Each student takes both quizzes—one with a calculator and one without—in a random order. To analyze the data, the student constructs a scatterplot that displays the number of correct answers with and without a calculator for each of the 30 students. A least-squares regression yields the equation

Calculator = −1.2 + 0.865(Pencil) $r = 0.79$

Which of the following statements is/are true? Switching x and y doesn't change r, just LSRL Equation

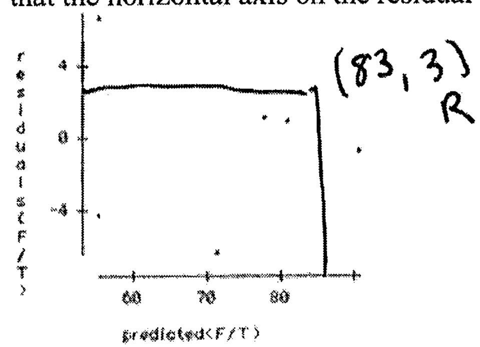
True I. If the student had used Calculator as the explanatory variable, the correlation would remain the same.

II. If the student had used Calculator as the explanatory variable, the slope of the least-squares line would remain the same. no LSRL equation is not the same when switched

III. The standard deviation of the number of correct answers on the paper-and-pencil quizzes was larger than the standard deviation on the calculator quizzes.

(a) I only (b) ~~II only~~ (c) III only (d) I and III only ~~(e) I, II, and III~~

Questions 5 and 6 refer to the following setting. Scientists examined the activity level of fish at 7 different temperatures. Fish activity was rated on a scale of 0 (no activity) to 100 (maximal activity). The temperature was measured in degrees Celsius. A computer regression printout and a residual plot are given below. Notice that the horizontal axis on the residual plot is labeled "predicted (F/T)."



Dependent variable is: Fish Activ
No Selector
R squared = 91.0% R squared (adjusted) = 89.2%
s = 4.785 with 7 − 2 = 5 degrees of freedom

$\hat{y} = 148.517 - 3.21667x$

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	148.617	10.71	13.9	≤ 0.0001
Temp	−3.21667	0.4533	−7.1	0.0009

$\hat{y} = 148.517 - 3.21667(20.4) = 82.9 \approx 83$

5. What was the activity level rating for the fish at a temperature of 20.4°C?

(a) 86 (b) 83 (c) 80 (d) 66 (e) 3

$R = y - \hat{y}$ $3 = y - 83$ $\boxed{y = 86}$

6. Which of the following gives a correct interpretation of s in this setting? $s = 4.785$ the average error (residual) from LSRL is 4.78

(a) For every 1°C increase in temperature, fish activity is predicted to increase by 4.785 units.

(b) The average distance of the temperature readings from their mean is about 4.785°C.

(c) The average distance of the activity level ratings from the least-squares line is about 4.785 units.

(d) The average distance of the activity level readings from their mean is about 4.785.

(e) At a temperature of 0°C, this model predicts an activity level of 4.785.

s: the average error (residual)

7. Which of these is *not* true of the correlation r between the lengths in inches and weights in pounds of a sample of brook trout? false

(a) r must take a value between -1 and 1. true

(b) r is measured in inches. (r has no units) false

(c) if longer trout tend to also be heavier, then $r > 0$.

(d) r would not change if we measured the lengths of the trout in centimeters instead of inches. true

(e) r would not change if we measured the weights of the trout in kilograms instead of pounds. true

8. When we standardize the values of a variable, the distribution of standardized values has mean 0 and standard deviation 1. Suppose we measure two variables X and Y on each of several subjects. We standardize both variables and then compute the least-squares regression line. Suppose the slope of the least-squares regression line is -0.44. We may conclude that

$b = -.44$

(a) the correlation will be $1/-0.44$.

(b) the intercept will also be -0.44.

(c) the intercept will be 1.0.

(d) the correlation will be 1.0.

(e) the correlation will also be -0.44.

r is the slope of the regression line when both X and Y are expressed as z-scores.

$\hat{y} = a + bx \quad b = r\frac{s_y}{s_x} \quad -.44 = r\left(\frac{1}{1}\right)$ so $r = -.44$

9. There is a linear relationship between the number of chirps made by the striped ground cricket and the air temperature. A least-squares fit of some data collected by a biologist gives the model $\hat{y} = 25.2 + 3.3x$, where x is the number of chirps per minute and $\hat{y}$ is the estimated temperature in degrees Fahrenheit. What is the predicted increase in temperature for an increase of 5 chirps per minute?

(a) 3.3°F (b) 16.5°F (c) 25.2°F (d) 28.5°F (e) 41.7°F

Let $x = 10$

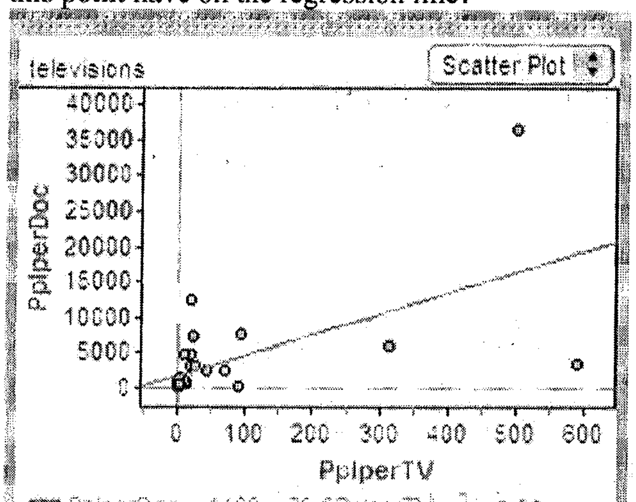
$\hat{y} = 25.2 + 3.3(10) = 58.2\,°F$

Let $x = 15$

$y = 25.2 + 3.3(15) = 74.7\,°F$

$\boxed{74.7 - 58.2 = 16.5\,°F}$

10. A data set included the number of people per television set and the number of people per physician for 40 countries. The Fathom screen shot below displays a scatterplot of the data with the least-squares regression line dded. In Ethiopia, there were 503 people per TV and 36,660 people per doctor. What effect would removing this point have on the regression line?



remember outlier in direction of x is influential, pulls line in its direction

(a) Slope would increase; y intercept would increase.

(b) Slope would increase; y intercept would decrease.

(c) Slope would decrease; y intercept would increase.

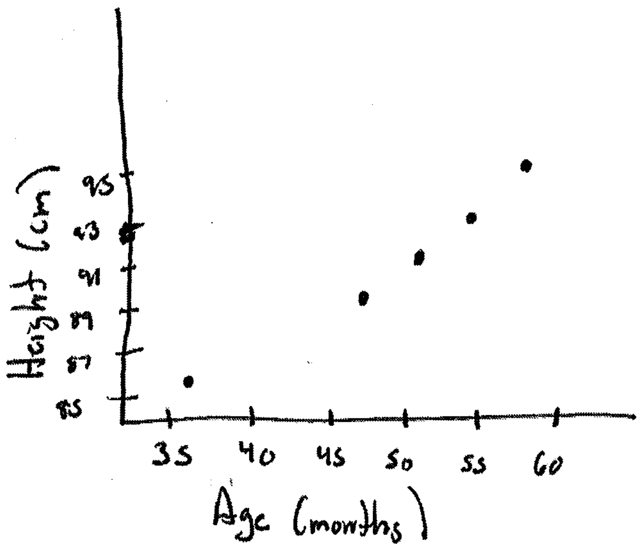
(d) Slope would decrease; y intercept would decrease.

(e) Slope and y intercept would stay the same.

11. Sarah's parents are concerned that she seems short for her age. Their doctor has the following record of Sarah's height:

Age (months):	36	48	51	54	57	60
Height (cm):	86	90	91	93	94	95

(a) Make a scatterplot of these data.



(b) Using your calculator, find the equation of the least-squares regression line of height on age.

$\hat{y} = 71.95 + .383(x)$ $\quad$ x = # months (age) $\quad$ y = height (cm)

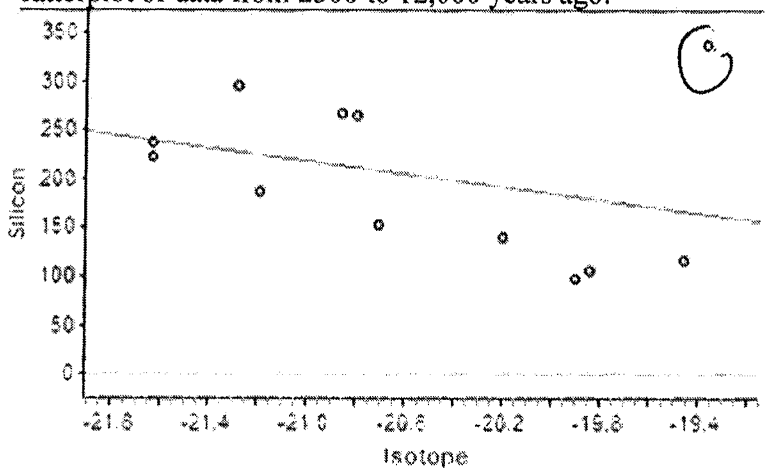
(c) Use your regression line to predict Sarah's height at age 40 years (480 months). Convert your prediction to inches (2.54 cm = 1 inch).

$\hat{y} = 71.95 + .383(480) = 255.79$ cm $\quad \frac{255.79\text{cm}}{1} \cdot \frac{1\text{in}}{2.54\text{cm}} = \boxed{100.70\text{ in}}$

(d) The prediction is impossibly large. Explain why this happened.

This height is impossibly large (about 8 ft 4 inches) because we use extrapolation. Our data was based on only the 1st 5 years of life. The linear trend does not carry all the way out to 40 years old.

12. Drilling down beneath a lake in Alaska yields chemical evidence of past changes in climate. Biological silicon, left by the skeletons of single-celled creatures called diatoms, is a measure of the abundance of life in the lake. A rather complex variable based on the ratio of certain isotopes relative to ocean water gives an indirect measure of moisture, mostly from snow. As we drill down, we look further into the past. Here is a scatterplot of data from 2300 to 12,000 years ago:



(a) Identify the unusual point in the scatterplot. Explain what's unusual about this point.

The unusual point is the one in the upper-right corner with isotope value ≈ -19.4 and silicone value ≈ 345. This point is unusual in that it has such a high silicone value for the given isotope value.

(b) If this point was removed, describe the effect on

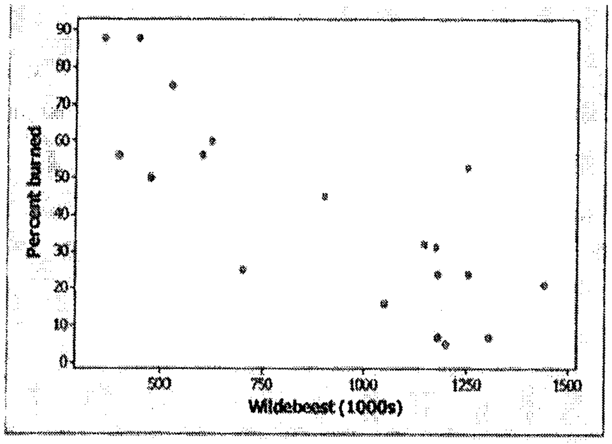
i. the correlation.

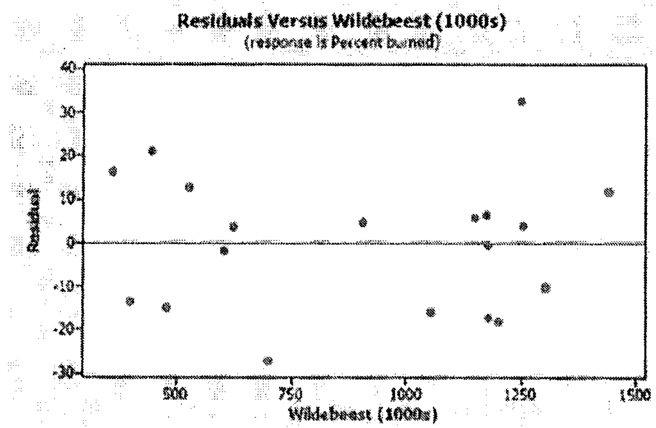
If the point were removed, the correlation would grow stronger, closer to -1, b/c the point does not follow the linear pattern of the other points.

ii. the slope and y intercept of the least-squares line.

Since this point has a higher silicon value, if it were removed, the slope of the regression line would increase in a negative direction and y intercept would increase.

13. Long-term records from the Serengeti National Park in Tanzania show interesting ecological relationships. When wildebeest are more abundant, they graze the grass more heavily, so there are fewer fires and more trees grow. Lions feed more successfully when there are more trees, so the lion population increases. Researchers collected data on one part of this cycle, wildebeest abundance (in thousands of animals) and the percent of the grass area burned in the same year. The results of a least-squares regression on the data are shown here.[27]





Predictor	Coef	SE Coef	T	P
Constant	92.29	10.06	9.17	0.000
Wildebeest (1000s)	-0.05762	0.01035	-5.56	0.000

S = 15.9880 R-Sq = 64.6% R-Sq(adj) = 62.5%

(a) Give the equation of the least-squares regression line. Be sure to define any variables you use.

x: # of wildebeest (in 1000s) $\hat{y} = 92.29 - 0.05762(x)$
y: % of grass burned

(b) Explain what the slope of the regression line means in this setting.

The slope -0.05762 means that for an increase in 1000 wildebeest, we predict the grassy area burned will decrease by .05762 %.

(c) Find the correlation. Interpret this value in context.

$r^2 = .646$ $r = \pm\sqrt{.646}$ $r = -.804$. This tells us the overall pattern is moderately linear.

(d) Is a linear model appropriate for describing the relationship between wildebeest abundance and percent of grass area burned? Support your answer with appropriate evidence.

The linear model is appropriate for describing the relationship between wildebeest abundance and % of grass area burned.
The residual plot shows a fairly "random" scatter of points around the residual line y = 0. (There is one large positive residual at ≈ 1250 thousand wildebeest.) Since $r^2 = .646$, 64.6% of the variation in % of grass burned is explained by LSRL (or 64.6% of variation in % of grass burned can be explained by abundance of wildebeests) (This leaves 35.4% of variation unexplained)

Chapter 4 REVIEW HW

AP Statistics Practice Test (page 274)

T4.1 c. A census is defined to be measuring all individuals in the population.

T4.2 e. Ignore numbers that are larger than 816 or are duplicate numbers.

T4.3 d. In order to infer cause and effect, we must run a well-designed experiment. This was an observational study.

T4.4 c. This is the definition of a Simple Random Sample.

T4.5 b. By randomly assigning treatments we are attempting to make the different groups look as similar as possible so that we can reduce the likelihood of a confounding variable.

T4.6 b. It is very difficult to show cause and effect using observational studies. It is much easier in an experiment where the researcher has control over how the treatments are applied.

T4.7 d. By stratifying we can control how many people we survey in each of the different kinds of areas.

T4.8 d. Bias in the responses means that you are getting responses that are systematically different from the truth.

T4.9 d. This is a completely randomized design because you randomly assign subjects to one of the four groups. There are two factors: Length of ad (30 seconds or 60 seconds) and Repeat (1 time or 3 times).

T4.10 b. In a matched pairs design, the two observations in the pair should be as similar as possible. So use a subjective method for pairing the plots. Once the pairs are chosen, then randomly assign the two treatments to the two plots in the pair.

T4.11 d. The teachers who responded likely feel more strongly about the issue and shouldn't be considered to be representative of the entire population of teachers under consideration.

4R HW

T4.12

EXPERIMENTAL UNITS		TREATMENTS		RESPONSE VARIABLE
72 acacia trees	→	T1 ACTIVE BEEHIVES (N=24)	→	COMPARE DAMAGE CAUSED BY ELEPHANTS
	→	T2 EMPTY BEEHIVES (N=24)	→	
	→	T3 NO BEE HIVE (N=24)	→	

random assignments

(B) COMPLETELY RANDOMIZED DESIGN

1. RANDOMLY ASSIGN THE ACADIA TREES TO THE 3 TREATMENSS. EACH TREATMENT WOULD BE ASSIGNED 24 TREES.
2. USE HAT, RANDOM DIGIT TABLE, OR TECHNOLOGY TO RANDOMLY ASSIGN TREES. TO DO SO
   - ASSIGN EACH TREE A NUMBERS 01-72
   - USE A RANDOM NUMBER TABLE TO PICK 24 2-DIGIT NUMBERS IN THE RANGE 01-72 EXCLUDING REPEATS.
   - THE 1ST 24 numbers will get beehives in these trees.
   - THE NEXT 24 numbers get empty beehives
   - THE REMAINING trees will remain empty
3. AT THE END, COMPARE THE DAMAGES CAUSED BY ELEPHANTS TO THE TREES WITH ACTIVE BEE HIVES, EMPTY BEE HIVES, AND NO BEE HIVE.

4R HW Cont

T4.13 (a) This is NOT a simple random sample (SRS) because not all samples were possible. For example, given their method, they could not have all respondents from the east Coast.

(b) One adult was chosen at random to control for lurking variables. Perhaps household members who generally answer the phone have a different opinion than those who don't generally answer the phone.

(c) There was undercoverage in this survey. Those who do not have telephones, or those who have only cell phones were not part of the sampling frame. So their opinions would not have been measured. Since cell-phone-only users tend to be younger, the results of the survey may not accurately reflect the entire population's opinion

4R cont

T4.14 (a) • This is a matched pairs design.

• EACH OF THE 11 individuals will be blocks.

• EACH PARTICIPANT WILL TAKE THE CAFFEINE ON ONE DAY AND THE PLACEBO ON THE OTHER.

• THE ORDER IN WHICH THEY TAKE THE PLACEBO OR CAFFEINE IS DECIDED RANDOMLY.

• THE TAPPING TEST IS ADMINISTERED AT THE END OF EACH 2 DAY TRIAL.

• THE RESULTS TO BE COMPARED ARE THE DIFFERENCES BETWEEN THE CAFFEINE AND PLACEBO SCORES ON THE TAPPING TEST.

THE BLOCKING WAS DONE TO CONTROL FOR INDIVIDUAL DIFFERENCES IN DEXTERITY.

(b) The order was randomized to control for any possible influence of the order in which the treatments were administered on the subject's tapping speed.

(c) A DOUBLE-BLIND MANNER CAN BE DONE BY ENSURING NEITHER THE SUBJECT OR ADMINISTRATOR OF THE TREATMENT HAS KNOWLEDGE OF THE ORDER IN WHICH THE CAFFEINE OR PLACEBO WAS ADMINISTERED

AP STATS
Chapter 5 Practice Test

Name Key
Date ____________ Hour ______

T5.1. Dr. Stats plans to toss a fair coin 10,000 times in the hope that it will lead him to a deeper understanding of the laws of probability. Which of the following statements is true?

(a) It is unlikely that Dr. Stats will get more than 5000 heads. no expected over long run prob 1/2
(b) Whenever Dr. Stats gets a string of 15 tails in a row, it becomes more likely that the next toss will be a head.
(c) The fraction of tosses resulting in heads should be close to 1/2.
(d) The chance that the 100th toss will be a head depends somewhat on the results of the first 99 tosses.
(e) All of the above statements are true.

T5.2. China has 1.2 billion people. Marketers want to know which international brands they have heard of. A large study showed that 62% of all Chinese adults have heard of Coca-Cola. You want to simulate choosing a Chinese at random and asking if he or she has heard of Coca-Cola. One correct way to assign random digits to simulate the answer is:

(a) One digit simulates one person's answer; odd means "Yes" and even means "No." This is 50/50
(b) One digit simulates one person's answer; 0 to 6 mean "Yes" and 7 to 9 mean "No." 70% is yes
(c) One digit simulates the result; 0 to 9 tells how many in the sample said "Yes." this is 100%
(d) Two digits simulate one person's answer; 00 to 61 mean "Yes" and 62 to 99 mean "No." yes 62% yes
(e) Two digits simulate one person's answer; 00 to 62 mean "Yes" and 63 to 99 mean "No."
63% yes

T5.3. Choose an American household at random and record the number of vehicles they own. Here is the probability model if we ignore the few households that own more than 5 cars:

Number of cars:	0	1	2	3	4	5
Probability:	0.09	0.36	0.35	0.13	0.05	0.02

A housing company builds houses with two-car garages. What percent of households have more cars than the garage can hold?

(a) 7% (b) 13% (c) 20% (d) 45% (e) 55%

P(3 or 4 or 5) = .13 + .05 + .02 = .2

T5.4. Computer voice recognition software is getting better. Some companies claim that their software correctly recognizes 98% of all words spoken by a trained user. To simulate recognizing a single word when the probability of being correct is 0.98, let two digits simulate one word; 00 to 97 mean "correct." The program recognizes words(or not) independently. To simulate the program's performance on 10 words, use these random digits:

60970 70024 17868 29843 61790 90656 87964 18883

The number of words recognized correctly out of the 10 is

(a) 10 (b) 9 (c) 8 (d) 7 (e) 6

60	97	07	00	24	17	86	82	98	43
1	2	3	4	5	6	7	8	9	10
c	c	c	c	c	c	c	c	~~c~~	c

9/10 correct

Questions T5.5 to T5.7 refer to the following setting. One thousand students at a city high school were classified according to both GPA and whether or not they consistently skipped classes. The two-way table below summarizes the data.

	GPA			
Skipped Classes	**<2.0**	**2.0–3.0**	**>3.0**	Total
Many	80	25	5	110
Few	175	450	265	890
	255	475	270	1000

T5.5. What is the probability that a student has a GPA under 2.0?
(a) 0.227 (b) 0.255 (c) 0.450 (d) 0.475 (e) 0.506

$P(GPA < 2.0) = \frac{255}{1000} = .255$

T5.6. What is the probability that a student has a GPA under 2.0 or has skipped many classes?
(a) 0.080 (b) 0.281 (c) 0.285 (d) 0.365 (e) 0.727

$P(GPA < 2.0 \text{ or many}) = \frac{255}{1000} + \frac{110}{1000} - \frac{80}{1000} = \frac{285}{1000} = .285$

T5.7. What is the probability that a student has a GPA under 2.0 given that he or she has skipped many classes?
(a) 0.080 (b) 0.281 (c) 0.285 (d) 0.314 (e) 0.727

$P(GPA < 2.0 \mid Many) = \frac{80}{110} = .727$

T5.8. For events *a* and *B* related to the same chance process, which of the following statements is true?
(a) If *a* and *B* are mutually exclusive, then they must be independent.
(b) If *a* and *B* are independent, then they must be mutually exclusive.
(c) If *a* and *B* are not mutually exclusive, then they must be independent.
(d) If *a* and *B* are not independent, then they must be mutually exclusive.
(e) If *a* and *B* are independent, then they cannot be mutually exclusive.

T5.9. Choose an American adult at random. The probability that you choose a woman is 0.52. The probability that the person you choose has never married is 0.25. The probability that you choose a woman who has never married is 0.11. The probability that the person you choose is either a woman or has never been married (or both) is therefore about
(a) 0.77 (b) 0.66 (c) 0.44 (d) 0.38 (e) 0.13

$P(W \text{ or never married}) = .52 + .25 - .11 = .66$

T5.10. A deck of playing cards has 52 cards, of which 12 are face cards. If you shuffle the deck well and turn over the top 3 cards, one after the other, what's the probability that all 3 are face cards?
(a) 0.001 (b) 0.005 (c) 0.010 (d) 0.012 (e) 0.02

$P(3 \text{ face cards}) = \underbrace{\frac{12}{52}}_{\text{1st face card}} \cdot \underbrace{\frac{11}{51}}_{\text{2nd face card}} \cdot \underbrace{\frac{10}{50}}_{\text{3rd face card}} = \frac{1320}{132,600} \approx .010$ $.00995$

Section II: 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.*

5.11. Your teacher has invented a "fair" dice game to play. Here's how it works. Your teacher will roll one fair *eight-sided* die, and you will roll a fair *six-sided* die. Each player rolls once, and the winner is the person with the higher number. In case of a tie, neither player wins. The table shows the sample space of this chance process.

	Teacher Rolls							
You Roll	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	1,1	1,2 T	1,3 T	1,4 T	1,5 T	1,6 T	1,7 T	1,8 T
2	2,1	2,2	2,3 T	2,4 T	2,5 T	2,6 T	2,7 T	2,8 T
3	3,1	3,2	3,3	3,4 T	3,5 T	3,6 T	3,7 T	3,8 T
4	4,1	4,2	4,3	4,4	4,5 T	4,6 T	4,7 T	4,8 T
5	5,1	5,2	5,3	5,4	5,5	5,6 T	5,7 T	5,8 T
6	6,1	6,2	6,3	5,5	6,5	6,6	6,7 T	6,8 T

(a) Let *a* be the event "your teacher wins." Find *P*(*a*).

$P(a) = \frac{27}{48}$

(b) Let *B* be the event "you get a 3 on your first roll." Find *P*(*a* ∪ *B*)

$P(a \cup B) = P(a) + P(B) - P(a \cap B)$

$\frac{27}{48} + \frac{8}{48} - \frac{5}{48} = \frac{30}{48} = \frac{5}{8} = .625$

(c) Are events *a* and *B* independent? Justify your answer.

① $P(a) = \frac{27}{48}$

$P(B) = \frac{8}{48}$

$P(B|A) = \frac{5/48}{\frac{27}{48}} = 5/27$ $\quad P(B|A) \neq P(B)$

∴ not independent

or

✓ $P(A \cap B) = P(A) \cdot P(B)$

$\frac{5}{48} \neq \frac{27}{48}\left(\frac{8}{48}\right)$

no, ∴ not independent

or

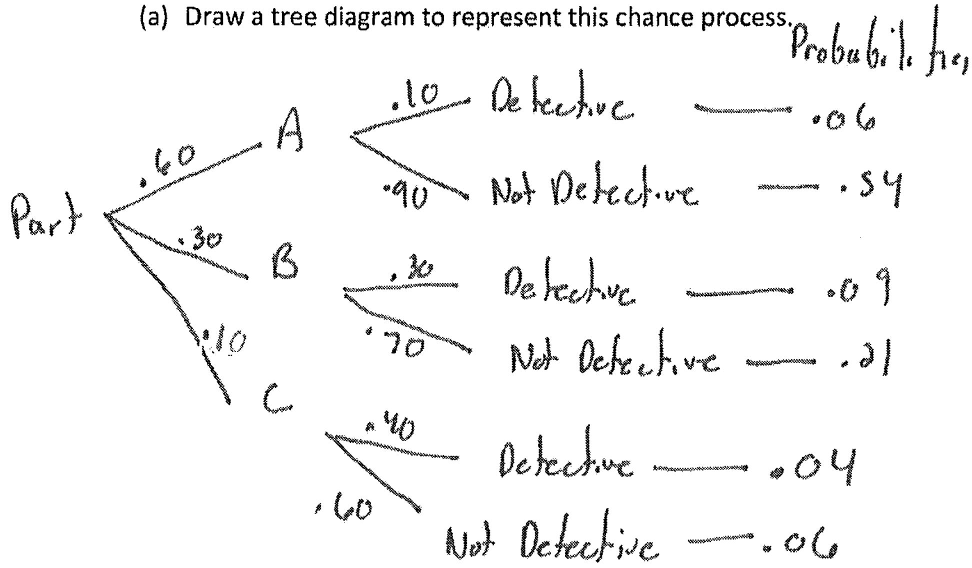
3

$P(A|B) = \frac{5/48}{8/48} = 5/8$ $\quad P(A|B) \neq P(A)$

∴ not independent

T5.12. Three machines—A, B, and C—are used to produce a large quantity of identical parts at a factory. Machine A produces 60% of the parts, while Machines B and C produce 30% and 10% of the parts, respectively. Historical records indicate that 10% of the parts produced by Machine A are defective, compared with 30% for Machine B and 40% for Machine C.

(a) Draw a tree diagram to represent this chance process.



(b) If we choose a part produced by one of these three machines, what's the probability that it's defective? Show your work.

P(Defective) = .06 + .09 + .04 = .19

(c) If a part is inspected and found to be defective, which machine is most likely to have produced it? Give appropriate evidence to support your answer.

$P(A \mid \text{Defective}) = \frac{.06}{.19} = .316$

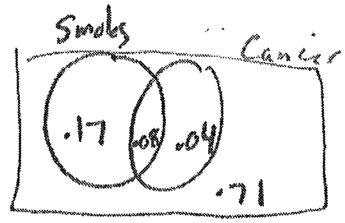
$P(B \mid \text{Defective}) = \frac{.09}{.19} = .474$

$P(C \mid \text{Defective}) = \frac{.04}{.19} = .211$

Machine B is most likely to have produced the part given it was defective.

T5.13. Researchers are interested in the relationship between cigarette smoking and lung cancer. Suppose an adult male is randomly selected from a particular population.The following table shows the probabilities of some events related to this chance process:

Event	Probability
Smokes	0.25
Smokes and gets cancer	0.08
Does not smoke and does not get cancer	0.71



(a) Find the probability that the individual gets cancer given that he is a smoker. Show your work.

$P(\text{Cancer} \mid \text{Smoker}) = \frac{.08}{.25} = .32$

(b) Find the probability that the individual smokes or gets cancer. Show your work.

$P(\text{Smoks or Cancer}) = 1 - P(\text{no smoke} \cap \text{no cancer})$

$= 1 - .71 = .29$

or from Venn Diagram $P(S \text{ or } C) = .17 + .08 + .04 = .29$

(c) Two adult males are selected at random. Find the probability that at least one of the two gets cancer. Show your work.

P(at least one of the 2 get cancer) = 1 − P(neither get cancer)

P(getting cancer) = .08 + .04 = .12

for 1 person

so P (not getting cancer) = 1 − .12 = .88

for 1 person

P (not getting cancer for 2 people) = $.88^2 = .7744$

Prob (at least one of the 2 get cancer) = 1 − .7744 = .2256

T5.14. Based on previous records, 17% of the vehicles passing through a tollbooth have out-of-state plates. A bored tollbooth worker decides to pass the time by counting how many vehicles pass through until he sees two with out-of-state plates.[27]

(a) Describe the design of a simulation to estimate the average number of vehicles it takes to find two with out-of-state plates. Explain clearly how you will use the partial table of random digits below to carry out your simulation.

(b) Perform three repetitions of the simulation you described in part (a). Copy the random digits below onto your paper. Then mark on or directly above the table to show your results.

41050	92031	06449	05059	59884	31880
53115	84469	94868	57967	05811	84514
84177	06757	17613	15552	51506	81435
75011	13006	63395	55041	15866	06589

a) Assign #'s 01 through 17 to represent cars with out of state plates and the remaining #'s, 00 and 18-99, to represent the other cars. Start reading two digit numbers from random # table until you have found 2 #'s between 01 and 17. Record how many 2 digit #'s you read in order to get #'s between 01 and 17. Repeat many times for the simulation

b) Trial 1: 41 05 09 3 cars to find 2 out of state plates

Trial 2: 20 31 06 44 90 50 59 59 88 43 18 80 53 11 14 cars to find 2 out of state plates

Trial 3: 58 44 69 94 86 85 79 67 05 81 18 45 14 13 cars to find 3 out of state plates

A and B independent

to check for independence

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = P(B)$$

A and B Mutually Exclusive

to check if events are M.E

$$P(A \cap B) = 0$$

Problem 5.8

A) If a and b M.E, then they must be independent.

Counterexample.

Let a: Red card in deck of cards b: Spade card in deck of cards

a and b are ME ✓

are they independent?

$$P(R|S) = P(R)$$

$$P(R|S) = \frac{P(R \cap S)}{P(S)} = \frac{0}{13} = 0$$

$$P(R) = \frac{13}{52} = \frac{1}{4}$$

$P(R|S) \neq P(R)$ ∴ not independent

Counterexample

B) If a and B are independent, then they must be mutually exclusive.

Let a: Red Card B: Ace

$$P(R \cap A) = 2$$

∴ not mutually exclusive

$(P(A \cap A) \neq 0)$

$$P(R|A) = P(R)$$

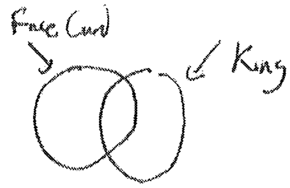
$$P(R|A) = \frac{P(R \cap A)}{P(A)} = \frac{2}{4} = \frac{1}{2}$$

$P(R) = \frac{26}{52} = \frac{1}{2}$ ∴ independent

C) If a and B are not Mutually Exclusive, then they must be independent

Counterexample

a: face card   b: king



not ME

$P(\text{Face Card} \cap \text{King}) = \frac{2}{52} \neq 0$

so not mutually exclusive

are a and b independent?

$P(F|K) = \dfrac{P(F \cap K)}{P(K)} = \dfrac{\frac{4}{52}}{\frac{13}{52}} = \frac{4}{52} \cdot \frac{5}{1} = \boxed{\frac{4}{13}}$

$P(F) = \frac{13}{52} = \frac{1}{13}$

$\frac{1}{13} \neq \frac{4}{13}$

$\therefore$ ~~not independent~~

D) use example above to show D is false

If a and b are not independent, then they must be mutually exclusive

E) If a and B are independent, then they cannot be mutually exclusive.

If a and B are independent, then $P(B|A) = \dfrac{P(A \cap B)}{P(A)} = P(B)$

and $P(A) \neq 0$, so $P(A) > 0$

$P(A|B) = \dfrac{P(A \cap B)}{P(B)} = P(A)$

and $P(B) \neq 0$, so $P(B) > 0$

This means $P(A \cap B) = P(A) \cdot P(B) \neq 0$ since $P(A)$ and $P(B)$ both greater than

$\therefore$ A and B cannot be mutually exclusive