

Problems Chosen: \_\_\_\_\_

Extra Credit Chosen: \_\_\_\_\_

Directions: Choose 8 problems. Each will be worth 25 points for a total of 200 points. You must include two from  $\{1,2,3,4\}$ , two from  $\{5,6,7,8\}$ , two from  $\{11,12,14,15\}$  and two from  $\{16,17,18,20\}$ . You cannot include more than one from  $\{2,5,11,16\}$ .

Extra problems can be done when you have completed your chosen 8. They can be chosen from  $\{9,10,13,19\}$

1. The Get-A-Grip tire company claims that the mean lifetime of tires sold on new cars is 23,000 miles and the standard deviation is 2500 miles.
  - a) If the claim by Get-A-Grip is true, what is the mean of the sampling distribution of  $\bar{x}$  for samples of size  $n = 4$ ?
  - b) If the claim by Get-A-Grip is true, what is the standard deviation of the sampling distribution of  $\bar{x}$  for samples of size  $n = 4$ ?
  - c) If the distribution of tire life is approximately normal, what is the probability that the mean of a random sample of  $n = 4$  tire lifetimes will be less than 20,000 miles?
2. Identify the following statements as true or false. Write a sentence to explain your answer.
  - a. As  $n$  increases, the mean of the sampling distribution of  $\bar{x}$  gets closer to  $\mu$ .
  - b. The standard deviation of the distribution of  $\bar{x}$  decreases as  $n$  increases.
  - c. The sampling distribution of  $\hat{p}$  tends to be more spread out for larger sample sizes than for smaller sample sizes.
  - d. For  $n$  sufficiently large, the distribution of  $\frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$  is approximately a standard normal distribution.
3. One method for estimating abundance of animals is known as line-intercept sampling. The theory of this method, when applied to Alaskan wolverines, predicts that the proportion  $p = 0.453$  of attempts to locate wolverine tracks should be successful. Suppose that biologists will make 100 attempts to locate wolverine tracks in random locations in Alaska.
  - a) Show that this sample size is large enough for the sampling distribution of  $\hat{p}$  to be approximately normal.
  - b) What is the mean of the sampling distribution of  $\hat{p}$  if the proportion predicted by line-intercept sampling is correct?

- c) What is the standard deviation of the sampling distribution of  $\hat{p}$  if the proportion predicted by intercept sampling is correct?
- d) If the proportion predicted by line-intercept sampling is correct, what is the probability that a sample proportion,  $\hat{p}$ , would differ from  $p = 0.453$  by as much as 0.05?
- 4.** The first large-scale study of the human sex ratio involved over 6,000 families each having 12 children. (This was done in 19th Century Germany – large families were more common.) 52% of the children they observed were boys. Suppose that 21st Century researchers wish to replicate this observational study to see if the proportion of boys might have changed in the intervening years. Further suppose the researchers track down 50 families with 12 children. From these 600 children, a random sample of 50 children is taken. 30 of the 50 children were boys ( $\hat{p} = 0.60$ .)
- a) If the modern true population proportion of newborn boys is  $p = 0.52$ , show that it is reasonable to approximate the sampling distribution of  $\hat{p}$  using a normal distribution.
- b) If the modern true population proportion of newborn boys is  $p = 0.52$ , what is the probability of observing a sample proportion of at least  $\hat{p} = 0.6$ ?
- 5.** Identify each statement as true or false and write a sentence to explain your answer.
- a. For a continuous random variable  $x$ , the area under the density curve over an interval  $a$  to  $b$  represents the probability that  $x$  is between  $a$  and  $b$ .
- b. For every random variable, both discrete and continuous,  $P(a \leq x \leq b) = P(a < x < b)$ .
- c. If  $x$  is a random variable, and random variable  $y$  is defined as  $y = a + bx$ , then  $m_y = |b|s_x$ .
- d. The probability of the first success at trial  $x$  of a sequence of independent trials, each with equal probability of success  $p$ , is:  $p^{x-1}(1-p)$ .
- 6.** At the College of Warm & Fuzzy, good grades in math are easy to come by. The grade distribution is given in the table below:

Grade	A	B	C	D	F
Proportion	0.30	0.35	0.25	0.09	0.01

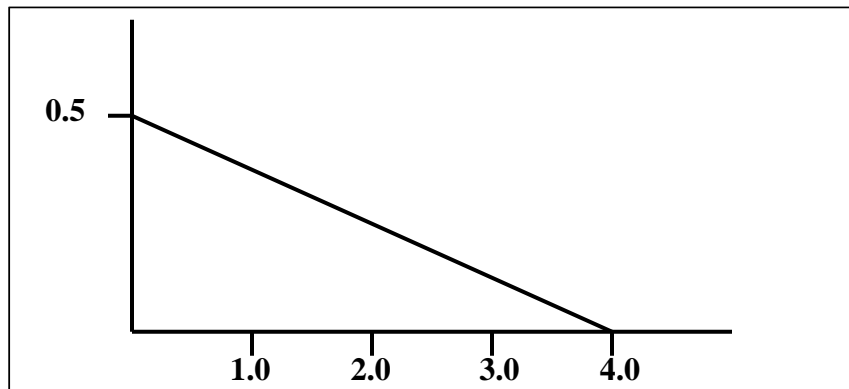
Suppose three students' grade reports are selected at random with replacement. (The math grade is written down and the report replaced in the population of reports before the next report is selected.) Three possible outcomes of this experiment are listed below. Calculate the probabilities of these sequences appearing.

a) **BAC**

b) **CFF**

7. The density curve for a continuous random variable is shown below. Use this curve to find the following probabilities:

- a)  $P(x < 1)$   
 b)  $P(2 < x < 3)$   
 c)  $P(x \text{ is at least } 3)$



You may use any of the following area formulas in your calculations:

Area of a rectangle:  $A = lw$

Area of a trapezoid:  $A = \frac{1}{2}h(b_1 + b_2)$

Area of a right triangle:  $A = \frac{1}{2}ab$

8. Suppose that the distribution of maximum daily temperatures in Hacienda Heights, CA, for the month of December has a mean of  $17^\circ\text{Celsius}$  with a standard deviation of  $3^\circ\text{Celsius}$ . Let the random variable  $F$  be the maximum daily temperature in degrees Fahrenheit. The relation between

Fahrenheit and Celsius degrees is: Degrees  $F = \frac{9}{5}C + 32$ .

a) What is the mean of  $F$ ?

b) What is the standard deviation of  $F$ ?

**9.** A gasoline tank for a certain model car is designed to hold 12 gallons of gas. Suppose that the actual capacity of the gas tank in cars of this type is well approximated by a normal distribution with mean 12.0 gallons and standard deviation 0.2 gallons. What is the approximate probability that a randomly selected car of this model will have a gas tank that holds at most 11.7 gallons?

**10.** For a variable that has a standard normal distribution,

- a) What is the probability that  $z < -1.34$ ?
- b) What is the probability that  $z < +2.56$ ?
- c) What is the probability that  $z$  is between  $-1.5$  and  $+1.5$ ?
- d) What value of  $z$  separates the smaller 5% of the standard normal distribution from the larger 95%?
- e) What values of  $-z$  and  $+z$  separate the middle 90% of the standard normal distribution from the extreme 10%?

**11.** Identify each statement as true or false and write a sentence to explain your answer.

- a. Any collection of possible outcomes of a chance experiment is called a sample space.
- b. The probability of an event  $E$  can always be computed using the formula,

$$P(E) = \frac{\text{number of outcomes favorable to } E}{\text{number of outcomes in the sample space}}$$

- c. Two events are independent if they cannot occur simultaneously.
- d. Two events,  $E$  and  $F$ , are independent if  $P(E|F) = P(F|E)$ .

**12.** Each school day as it gets close to the time when classes begin, automobiles start to arrive on campus. Consider an experiment in which a vehicle arriving at the school is selected at random. Define events  $V_1, V_2$ , and  $V_3$  as follows:

$V_1$  = the vehicle belongs to a school staff member

$V_2$  = the vehicle is less than 5 years old

$V_3$  = the vehicle is made in Detroit

For each problem below sketch a Venn diagram with 3 circles and shade the regions that correspond to the following events:

- a) An vehicle made in Detroit that is less than 5 years old is selected.
- b) A staff member's vehicle is selected.
- c) A staff member's Detroit-made vehicle at least 5 years old is selected.

**13.** To help ensure the safety of school classrooms, the local Fire Marshall does an inspection at Thomas Jefferson High School every month to check for faulty wiring, overloaded circuits, etc. At this school, the new Academic Wing has 5 math rooms, 10 science rooms, and 10 English rooms. The science rooms are divided into 8 biology and 2 chemistry rooms. The Fire Marshall selects one of the rooms in the new wing at random for inspection. Define the following events:

$S$  = the event the selected room is a science room

$B$  = the event the selected room is a biology room

$M$  = the event the selected room is a math room

$E$  = the event the selected room is an English room

$C$  = the event the selected room is a chemistry room

Calculate the following probabilities:

a)  $P(S)$

b)  $P(M \text{ or } E)$

c)  $P(E \text{ or } B)$

d)  $P(S \text{ and not } C)$

- 14.** For a special report in the Cedar Rapids, IA, *Gazette*, reporters used a radar gun to check the speeds of motorists in two counties. They were shocked (!) to discover that some motorists were driving at a speed greater than the posted limits. They reported the following information in their story:

City	City Speed Limit	Cars Counted	Over Limit	%Over Limit	Speeding data for cities in Linn and Johnson Counties
Cedar Rapids	25	241	233	96.7	
Iowa City	35	292	124	42.5	
Iowa City	20	203	203	100.0	
Marion	30	329	268	81.5	
Coralville	25	174	161	92.5	

Suppose that one of these drivers is selected at random. Define the following events:

F = the event that a driver was in a city with a speed limit above 25 mph.

S = the event that a driver is speeding.

- a) The table below shows the possible combinations of F and S. For each combination, calculate the probability that the combination would be observed for a randomly selected driver. Enter these probabilities into the corresponding cell in the table.

	Speed Limit > 25mph (Fast lane)	Speed Limit 25mph or less (Slow lane)
Driver speeding		
Driver not speeding		

b) Using the information in the table in part (a), calculate  $P(FUS)$ .

c) Using the information in the table in part (a), calculate  $P(F | S)$ .

d) Using the information in the table in part (a), calculate  $P(S)$

**15.** The diamond python (*Morelia spilota*), an Australian snake, is about 3 feet long as an adult. In a multi-year study of the habitats of these creatures, 997 were captured. The following table displays the capture locations of these snakes by season of the year and habitat. The "other" category includes trees, logs, rocks, open ground, and under filtering cover such as shrubs. Suppose a diamond python is selected at random from the study records.

### Diamond Python Habitat

Suppose a diamond python is selected at random from the study records.

Calculate the probability that:

- a) The selected diamond python was located in a building, given that it was spring.
- b) The selected diamond python was located other than in a building given that it was spring or summer.
- c) The selected diamond python was observed in a building in the summer.

**16.** Identify each statement as true or false. Write a sentence to explain your answer.

- a. The value of Pearson's  $r$  is always between 0 and 1.
- b. The slope of the least squares line is the average amount by which  $y$  increases as  $x$  increases by one unit.

	Buildings	Other	Total
Spring	10	343	353
Summer	36	273	309
Autumn	17	157	174
Winter	0	161	161
Total	63	934	997

c. The slopes of the least squares lines for predicting  $y$  from  $x$ , and the least squares line for predicting  $x$  from  $y$ , are equal.

d. The value of the residual plus  $\hat{y}_i$  is equal to  $y_i$ .



- 17.** The early human ancestors were similar in shape to most large primates. The data below are average male hind limb and forelimb lengths for different species of early homonins (humans and their ancestors.) The relatively short hind limbs are thought to represent evolutionary specialization for vertical climbing.

**Hind limb and Forelimb lengths**

<b>Hind limb Length (mm)</b>	<b>Forelimb Length (mm)</b>
471	458
361	514
399	581
557	739
553	553
574	614
857	595
698	762

- a) What is the value of the correlation coefficient for these data?
- b) What is the equation of the least squares line describing the relationship between  $x$  = hind limb length and  $y$  = forelimb length.
- c) Suppose these species are representative of species of early human ancestors. If a new homonin species dating from about the same time were to be discovered with an average hind limb length of 500 mm, what would you predict to be the average forelimb length of this species?

- 18.** Assessing the goodness of fit of a regression line involves considering several things, and no single characteristic of data is sufficient for a good assessment. Consider the characteristics below. How does each contribute to an assessment of fit? That is, for each piece of information, what characteristic would indicate a "good" best-fit line?

- a) The shape of the scatter plot

- b) The correlation coefficient
- c) The standard deviation of the residuals
- d) The coefficient of determination

**19.** The Roman Empire minted coins in both Rome and the Eastern provinces. Historians would like to use metallurgical analysis as one tool to identify the source mints of Roman coins. They analyzed 11 coins known to have been produced by Eastern mints in an attempt to identify a trace element profile, and have identified gold and lead as possible factors in identifying other coins' source mint. The gold and lead content of the 11 coins, measured as a percent of weight, is given in the table at right. A scatter plot of these data is displayed below.

- a) What is the equation of the least squares

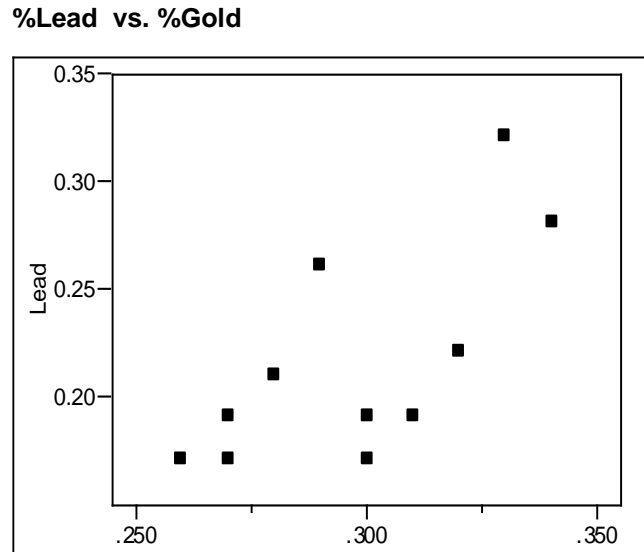
best fit line for predicting lead content  
from gold content?

- b) Sketch the best fit line on the scatter  
plot.

- c) What is the value of the correlation  
coefficient? Interpret this value in the  
context of this problem.

Gold % by Weight	Lead % by Weight
0.30	0.17
0.31	0.19
0.28	0.21
0.32	0.22
0.34	0.28
0.29	0.26
0.33	0.32
0.26	0.17
0.30	0.19
0.27	0.19
0.27	0.17

d) What is the value of the coefficient of determination? Interpret this value in the context of this problem.



Suppose that the coins analyzed in problem 1 are representative of the metallurgical content of coins minted in Rome.

- e) If a coin minted at Rome is selected at random, and its gold content is 0.20% by weight, calculate the predicted lead content.
- f) One of the coins used to calculate the regression equations has a gold content of 0.20%. Calculate the residual for this coin.

**20.** When children are discharged from the hospital their parents may still provide substantial care, such as the insertion of a feeding tube. It is difficult for parents to know how far to insert the tube, especially with rapidly growing infants. Health care professionals believe it may be possible for parents to measure their child's height and by using a regression equation calculate the appropriate insertion length using a regression equation. At a major children's hospital, children and adolescents' heights and esophagus lengths were measured and a regression analysis performed. The data from this analysis is summarized below:

#### Summary statistics from Regression Analysis

Height (cm) and Esophageal Length (cm)

	Esophagus Length ( $E$ )	Height ( $H$ )
$\bar{x}$	34cm	124.5cm
$s$	3.5cm	19cm

$$r = 0.995, \hat{E} = 11.476 + 0.181H$$

- a) For a child with a height one standard deviation above the mean, what would be the predicted esophagus length?
- b) What proportion of the variability in esophagus length is accounted for by the height of the children and adolescents?
- c) From the information presented above, does it appear that the esophagus length can be accurately predicted from the height of young patients? Provide statistical evidence for your response.