

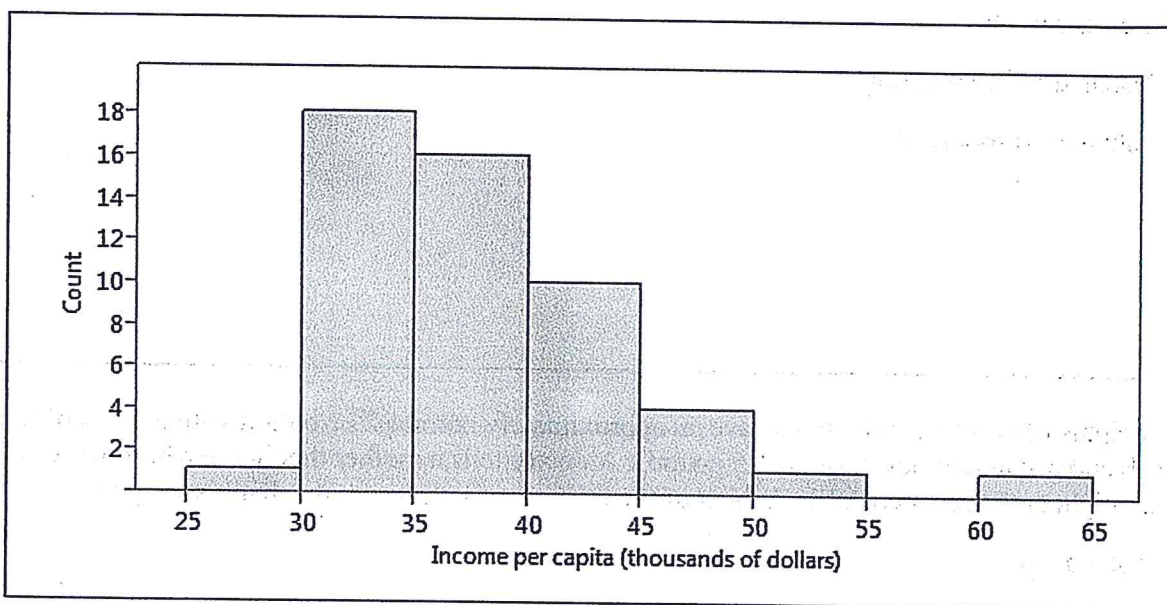
**STATISTICS  
SECTION I**

**Time --- 1 hour and 30 minutes**  
**Number of questions --- 40**  
**Percent of total grade --- 50**

Answer  
Key

**Directions:** Solve each of the following problems, using the available space for scratch work. Decide which is the best of the choices given and fill in the corresponding box on the answer sheet. Do not spend too much time on any one problem.

1. The histogram below summarizes the distribution of personal income per capita (in thousands of dollars) for each of the 50 states and the District of Columbia.



Which of the following best describes the shape of this distribution and the most likely relationship between the mean and median of this distribution?

- (A) Skewed left; mean < median
- (B) Skewed left; mean > median
- (C) Skewed right; mean < median
- (D) Skewed right; mean > median
- (E) Symmetric; mean = median

2. A researcher wants to know the percentage of villages in a certain African country that have access to a clean drinking water source less than 1/4 mile from the center of the village. The country is divided into 12 districts and each district has many villages in it, as indicated in the table below. The researcher selects a random sample of 10% of the villages from each district.

District	A	B	C	D	E	F	G	H	I	J	K	L
Number of Villages	50	40	30	80	50	30	80	90	20	20	40	70
Number Selected	5	4	3	8	5	3	8	9	2	2	4	7

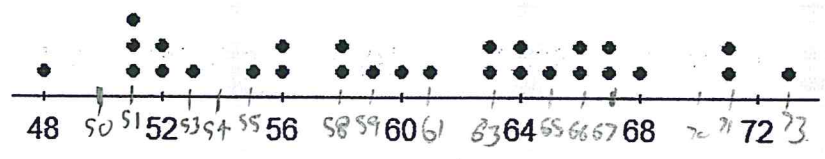
Which of the following terms best describes this sampling method?

- (A) Simple random sampling
- (B) Stratified random sampling *(subgroups look like each other)*
- (C) Cluster sampling
- (D) Matched-pairs sampling
- (E) Voluntary response sampling

3. The weights of newborn baby boys have an approximately normal distribution with a mean of 8.0 pounds and a standard deviation of 1.5 pounds. A doctor tells a mother that her newborn baby boy has a weight at the 25<sup>th</sup> percentile. Which of the following is closest to this baby's weight?

- (A) 2.00 pounds
- (B) 5.00 pounds
- (C) 7.00 pounds *inv Norm (area: 0.25,  $\mu$ : 8,  $\sigma$ : 1.5)*
- (D) 7.63 pounds
- (E) 8.90 pounds

4. The dotplot below displays the daily high temperature, in degrees Fahrenheit, for a city in the southeastern United States during the 28 days in the month of February 2014.



Daily High Temperature (degrees Fahrenheit)

probably an easier way to do this

Which of the following is closest to the percentile rank of a day in that month where the daily high temperature in this city was 70 degrees Fahrenheit?

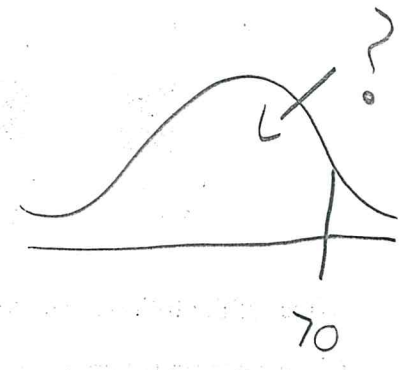
- (A) 10
- (B) 25
- (C) 70
- (D) 90**
- (E) 95

First I made a list for the temperatures and calculated the mean of 60.32

and standard deviation of 7.07

then:

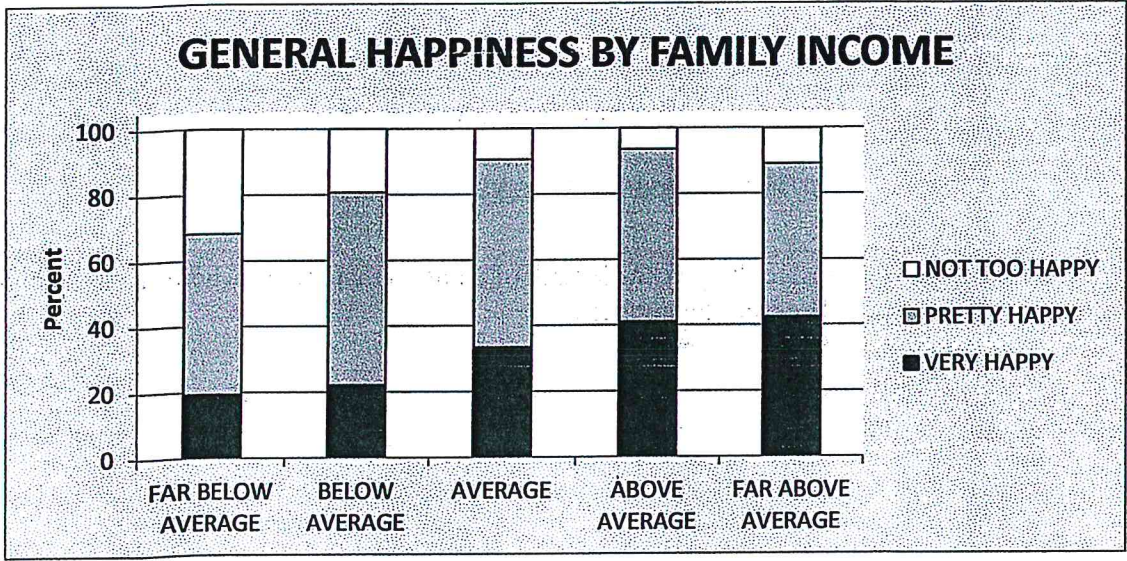
normalcdf (lower: 10000, upper: 70,  $\mu$ : 60.32,  $\sigma$ : 7.07)



5. The U.S. Census Bureau's 2013 American Community Survey (ACS) estimates that the proportion of Americans between the ages of 18 and 24 who have not received a high school diploma (or equivalent) is 0.1467. Wilmington, NC (population 112,067) is conducting a study in 2015 to see if their proportion is significantly different from the 2013 national proportion. Eight hundred residents between the ages of 18 and 24 were selected at random and it was found that 97 of them had not received a high school diploma or equivalent. If a hypothesis test were carried out to investigate the town's claim, to which of the following groups could the results be generalized?

- (A) Americans between the ages of 18 and 24 in 2013.
- (B) The 112,067 residents of Wilmington, NC in 2015.
- (C) The residents of Wilmington, NC between the ages of 18 and 24 in 2015.**
- (D) The 800 randomly selected residents of Wilmington, NC in 2015.
- (E) The 112,067 residents of Wilmington, NC in 2013.

6. Each year, the General Social Survey asks respondents to rate both their general level of happiness and their relative family income. The segmented bar graph below displays the results from the most recent General Social Survey.



Which of the following statements is supported by the information presented in this graph?

- (A) If a person is very happy, he or she most likely has an income far above average.
- (B) If a person is not too happy, he or she most likely has an income far below average.
- (C) If a person has a below average income, he or she is most likely not too happy.
- (D) Higher levels of income tend to be associated with higher levels of happiness.
- (E) Lower levels of income tend to be associated with higher levels of happiness.

most likely pretty happy

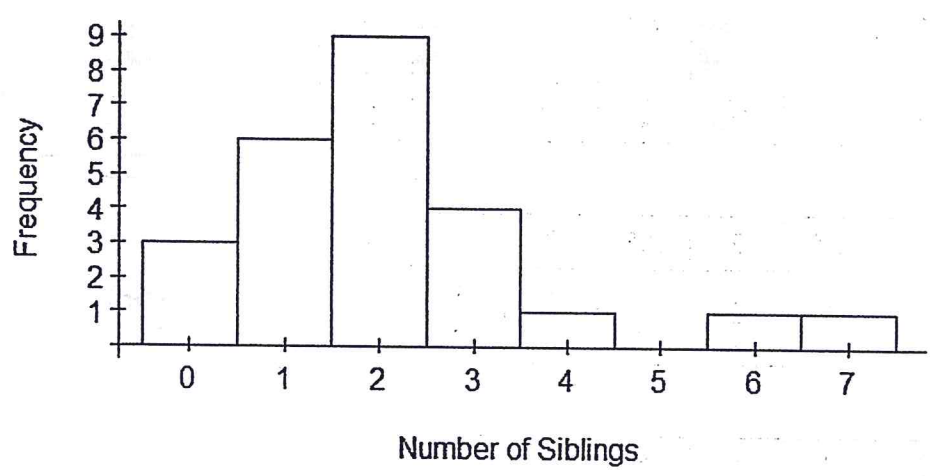
7. When manufacturing a certain smartphone, defects occur at different rates depending on the type of defect. Specifically, screen-related defects occur in 1% of phones, defects related to the charging port occur in 2% of phones, and both defects occur in 0.5% of phones. What is the probability a phone has a screen-related defect, given that it has a defect related to the charging port?

- (A) 0.0300
- (B) 0.0302
- (C) 0.2000
- (D) 0.2500
- (E) 0.5000

$P(A|B) = \frac{P(A \cap B)}{P(B)}$  ← formula sheet! conditional probability

$P(\text{screen defect} / \text{charging defect}) = \frac{P(S \cap C)}{P(C)} = \frac{0.005}{0.02} = 0.25$

8. The 25 students in one professor's introductory statistics course at a large university were surveyed and asked to report the number of siblings that they have. The histogram below displays the data collected in the survey.



Which of the following is the median number of siblings for these 25 students?

- (A) 1
- (B) 1.5
- (C) 2**
- (D) 2.5
- (E) 3

Handwritten solution for Q8:  
 median  
 ↓  
 000111111 2222222 3333 4 6 >  
 ~~~~~  
 3 6 9 7 1 1 1



9. Let the random variable  $X$  be the high temperature ( $^{\circ}\text{F}$ ) in a randomly selected city in the U.S. on the first day of summer. Suppose  $X$  has an approximately normal distribution with  $\mu = 85$  and  $P(X > 96) = 0.0139$ . Which of the following changes would lead to the greatest reduction in this probability?

- (A) decrease  $\mu$  to 82
- (B) decrease  $\sigma$  to 4
- (C) increase 96 to 100**
- (D) increase  $\mu$  to 96
- (E) change  $>$  to  $<$

B  
 Normalcdf  
 lower: 96  
 upper: 10000  
 $\mu$ : 85  
 $\sigma$ : 4  
 = 0.0030  
 Decrease:  
 0.0139  
 0.0030  
 0.0109

A Normalcdf (lower: 96, upper: 10000,  $\mu$ : 85,  $\sigma$ : 5) = 0.0026  
 Decrease: 0.0139 - 0.0026 = 0.0113  
 C Normalcdf (lower: 100, upper: 10000,  $\mu$ : 85,  $\sigma$ : 5) = 0.0013  
 Decrease: 0.0139 - 0.0013 = 0.0126  
 biggest reduction → 0.0126

To find  $\sigma$ :  
 invNorm (area: 0.9861,  $\mu$ : 0,  $\sigma$ : 1)  
 this gives a  $z = 2.2$   
 $z = \frac{96 - 85}{\sigma}$   
 $2.2 = \frac{11}{\sigma}$   
 $\sigma = 5$

D Normalcdf (lower: 96, upper: 10000,  $\mu$ : 96,  $\sigma$ : 5) = 0.5  
 not a reduction

E Normalcdf (lower: -10000, upper: 96,  $\mu$ : 85,  $\sigma$ : 5) = 0.9861  
 X not a reduction

10. Suppose you toss two fair, four-sided dice whose faces are labeled 1, 2, 3, and 4. Let  $\bar{X}$  represent the average of the two sides that are facing up. Which of the following is the sampling distribution of  $\bar{X}$ ?

(A)

|             |                |               |                |               |                |               |                |
|-------------|----------------|---------------|----------------|---------------|----------------|---------------|----------------|
| $\bar{X}$   | 1              | 1.5           | 2              | 2.5           | 3              | 3.5           | 4              |
| Probability | $\frac{1}{16}$ | $\frac{1}{8}$ | $\frac{3}{16}$ | $\frac{1}{4}$ | $\frac{3}{16}$ | $\frac{1}{8}$ | $\frac{1}{16}$ |

(B)

|             |               |               |               |               |               |               |               |
|-------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| $\bar{X}$   | 1             | 1.5           | 2             | 2.5           | 3             | 3.5           | 4             |
| Probability | $\frac{1}{7}$ | $\frac{1}{7}$ | $\frac{1}{7}$ | $\frac{1}{7}$ | $\frac{1}{7}$ | $\frac{1}{7}$ | $\frac{1}{7}$ |

(C)

|             |                |               |                |               |                |               |                |
|-------------|----------------|---------------|----------------|---------------|----------------|---------------|----------------|
| $\bar{X}$   | 2              | 3             | 4              | 5             | 6              | 7             | 8              |
| Probability | $\frac{1}{16}$ | $\frac{1}{8}$ | $\frac{3}{16}$ | $\frac{1}{4}$ | $\frac{3}{16}$ | $\frac{1}{8}$ | $\frac{1}{16}$ |

(D)

|             |               |               |               |               |               |               |               |
|-------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| $\bar{X}$   | 2             | 3             | 4             | 5             | 6             | 7             | 8             |
| Probability | $\frac{1}{7}$ | $\frac{1}{7}$ | $\frac{1}{7}$ | $\frac{1}{7}$ | $\frac{1}{7}$ | $\frac{1}{7}$ | $\frac{1}{7}$ |

(E)

|             |               |                |               |               |               |                |               |
|-------------|---------------|----------------|---------------|---------------|---------------|----------------|---------------|
| $\bar{X}$   | 2             | 3              | 4             | 5             | 6             | 7              | 8             |
| Probability | $\frac{1}{9}$ | $\frac{5}{36}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{5}{36}$ | $\frac{1}{9}$ |

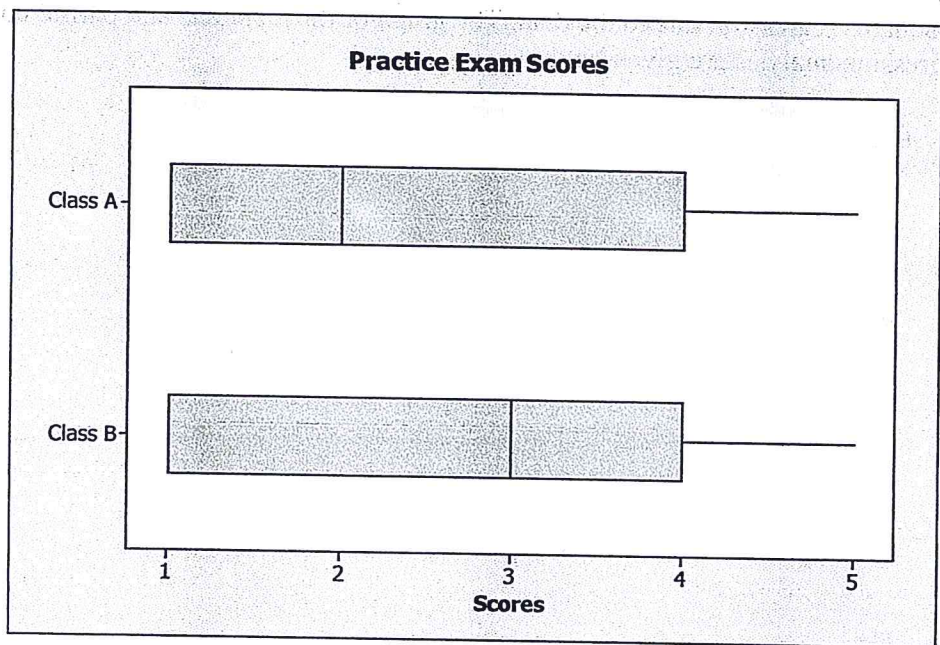
combinations

| Die 1 | Die 2 | $\bar{X}$ | outcome     |
|-------|-------|-----------|-------------|
| 1     | 1     | 1         | occurs once |
| 1     | 2     | 1.5       | twice       |
| 1     | 3     | 2         | three times |
| 1     | 4     | 2.5       | four times  |
| 2     | 1     | 1.5       |             |
| 2     | 2     | 2         |             |
| 2     | 3     | 2.5       |             |
| 2     | 4     | 3         | three times |
| 3     | 1     | 2         |             |
| 3     | 2     | 2.5       |             |
| 3     | 3     | 3         |             |
| 3     | 4     | 3.5       | twice       |
| 4     | 1     | 2.5       |             |
| 4     | 2     | 3         |             |
| 4     | 3     | 3.5       |             |
| 4     | 4     | 4         | occurs once |

|                        |                |               |                |               |                |               |                |
|------------------------|----------------|---------------|----------------|---------------|----------------|---------------|----------------|
| outcomes ( $\bar{X}$ ) | 1              | 1.5           | 2              | 2.5           | 3              | 3.5           | 4              |
| $P(\bar{X})$           | $\frac{1}{16}$ | $\frac{1}{8}$ | $\frac{3}{16}$ | $\frac{1}{4}$ | $\frac{3}{16}$ | $\frac{1}{8}$ | $\frac{1}{16}$ |

16 total outcomes

11. An AP Biology teacher gave a practice AP Exam in her two classes. The boxplots below illustrate the distribution of exam scores from the two classes, Class A and Class B.



Based on these boxplots, which of the following is the most plausible conclusion about the data collected?

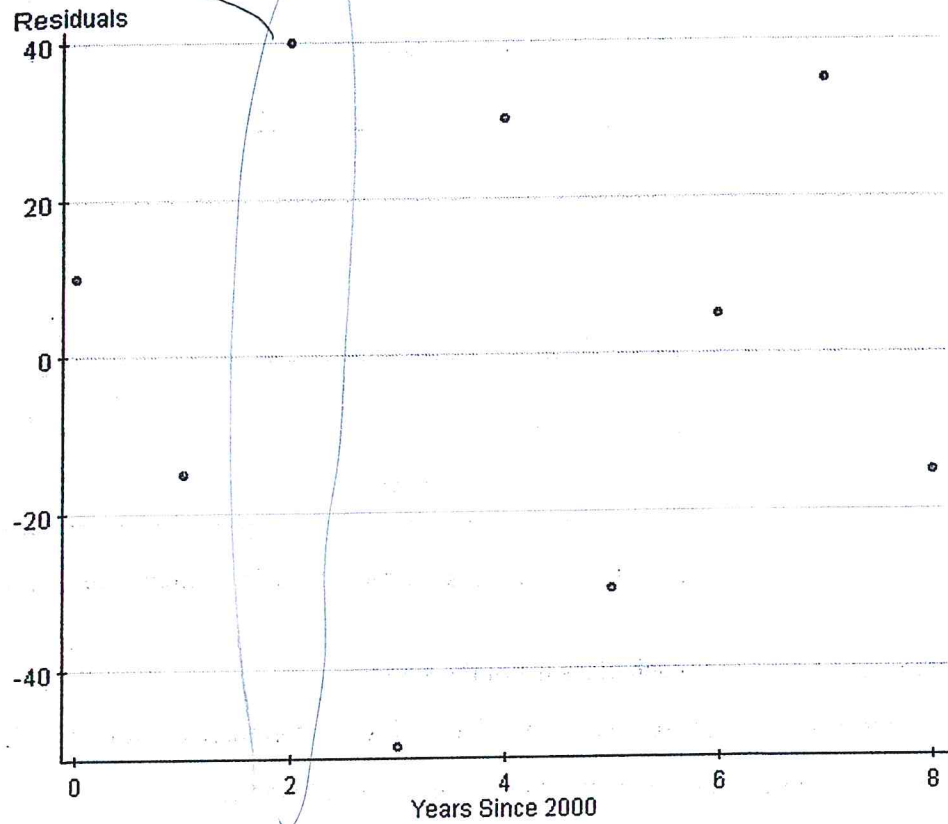
- (A) The interquartile range for each class is not the same. *both 3*
- (B) There are more scores that are greater than 2 for Class A than there are for Class B.
- (C) There are more scores that are greater than 3 for Class A than there are for Class B.
- (D) All students in the lower 25% of both classes had a score of 1.
- (E) More students in Class B scored a 1 than in Class A.

*same amount*

*No low whisker indicates the lowest score is the first quartile*

*we do not know the quantity of students in these classes just the percentage in each quartile*

12. Bob and Joan were searching the internet to find information on air travel in the United States. They found data on the number of commercial aircraft flying in the United States during the years 2000-2008. The dates were recorded as years since 2000 (thus, the year 2000 is recorded as year 0). They fit a least squares regression line to the data. The graph of the residuals and partial computer output for the regression analysis are given below.



| Variable | Coefficient | Std Dev | T      | P     |
|----------|-------------|---------|--------|-------|
| Constant | 2939.93     | 20.55   | 143.09 | 0.000 |
| Years    | 233.517     | 4.316   | 54.11  | 0.000 |

What is the actual number of commercial aircraft flying in 2002 (rounded to the nearest whole number)?

- (A) 2,472 flights
- (B) 2,513 flights
- (C) 3,022 flights
- (D) 3,367 flights

(E) 3,447 flights

$$\hat{y} = 2939.93 + 233.517x$$

number of commercial aircraft flights =  $2939.93 + 233.517(\text{years since } 2000)$

$$\hat{y} = 2939.93 + 233.517(2)$$

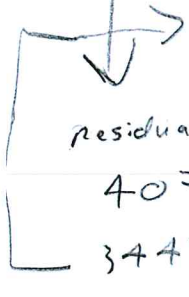
$$\hat{y} = 3406.967$$

-8 = predicted

residual = actual - predicted

$$40 = y - 3407$$

$$3447 = y$$





13. Researchers are interested in the effect of distractions on work efficiency. Forty college students volunteered to be part of an experiment in which half were randomly assigned to a quiet room and the other half to a room in which loud music was playing. The students were asked to complete an easy Sudoku puzzle and their completion times were recorded. A 95% confidence interval for the difference between mean completion times for the puzzle (music-no music) was found to be (5, 11) minutes. Could this information be used to conclude that we should reject a null hypothesis of no difference in the average completion time between these two groups at a significance level of 0.05?

- (A) Yes, because the confidence interval contains zero.
- (B) Yes, because the confidence interval does not include zero.
- (C) No, because the confidence interval does not include zero.
- (D) No, there is no relationship between confidence intervals and hypothesis tests.
- (E) No, this was not a properly designed experiment.

---

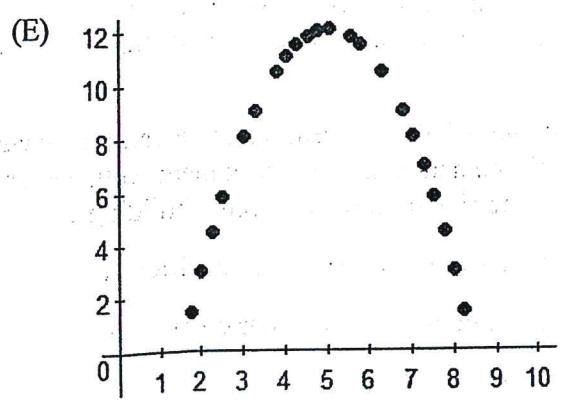
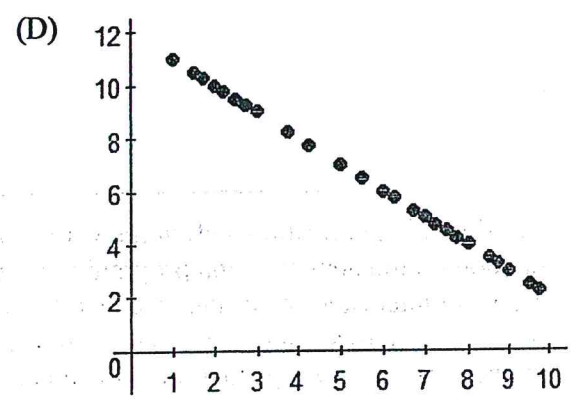
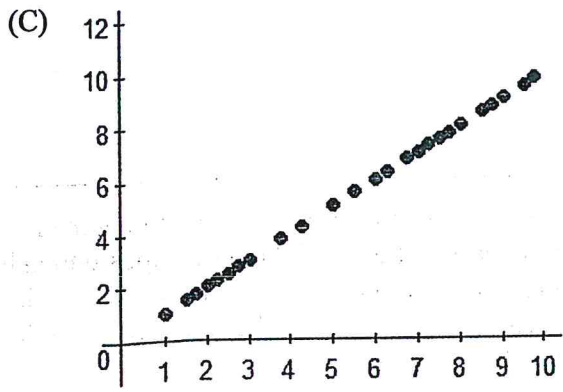
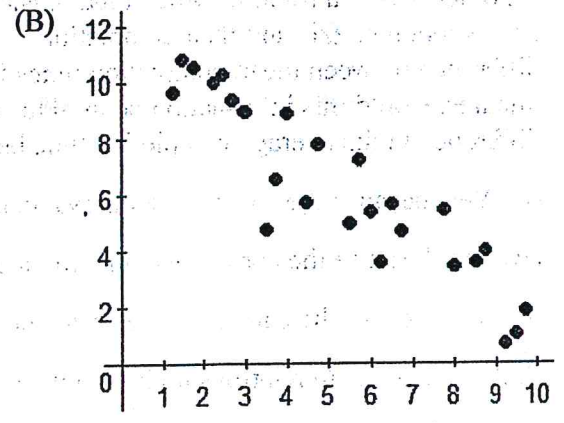
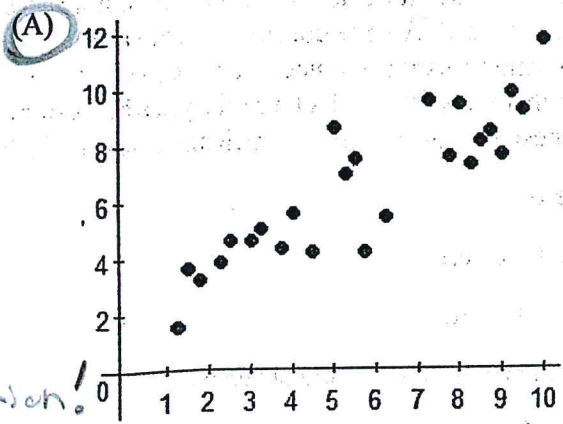
14. The dining and nutrition staff at the University of Georgia plans to survey students to get their opinion on the new nutrition program introduced this semester at each of the on-campus dining halls. They are interested in getting feedback from students living both on-campus and off-campus about the new gluten-free and vegetarian options offered at each meal. Which of the following sampling methods is the most appropriate for accomplishing this?

- (A) Hand out a survey to every 10<sup>th</sup> student that enters each dining hall on a specified day.
- (B) Group students by housing status, one group representing those living on campus and the other representing those living off campus. Email a survey to 100 randomly selected students from each group.
- (C) On equally sized slips of paper, write down the names of all the dormitories on campus as well as all the apartment complexes off campus. Put all the names in a hat, mix them well, and draw out five of them. Email a survey to all students in the five randomly selected buildings.
- (D) Hand out a survey to the first 50 students that enter each dining hall on a specified day.
- (E) Create a Facebook page for each dining hall where students can post their comments.

open to undercoverage: not all off-campus students live in apartment complexes

15. Which of the following scatterplots could represent a data set with a correlation coefficient of  $r = 0.85$ ?

Guess the correlation!



16. If the mean of a data set is 12.3 and the standard deviation is 0, then

- (A) all of the observations are equal to 12.3.
- (B) there must have been a mistake in the calculation, as it is not possible to have a standard deviation of 0.
- (C) a histogram of the data set has a single peak at zero.
- (D) a histogram of the data set will be strongly skewed.
- (E) there is a large amount of variability in the data set.

17. A biologist is interested in predicting the brain weight (in grams) of a certain species of bird from its body weight (in grams). A least squares regression line was fit to data collected from a random sample of 28 birds of this species. The equation of the line is

$$\hat{y} = 3.79 + 0.08x$$

$\frac{0.08}{1} \quad \frac{\Delta \hat{y}}{\Delta x} \leftarrow \begin{array}{l} \text{predicted brain} \\ \text{weight} \\ \text{body weight} \end{array}$

where  $\hat{y}$  is the predicted brain weight and  $x$  is the body weight of the bird. Which of the following gives the best interpretation of the slope of the regression line?

- (A) There is an increase of 0.08 grams in the predicted brain weight of this species of bird for every increase of 1 gram in body weight.
- (B) There is an increase of 0.08 grams in the predicted body weight of this species of bird for every increase of 1 gram in brain weight.
- (C) There is an increase of 3.79 grams in the predicted brain weight of this species of bird for every increase of 1 gram in body weight.
- (D) There is an increase of 3.79 grams in the predicted body weight of this species of bird for every increase of 1 gram in brain weight.
- (E) Approximately 8% of the variability in brain weight is predicted by its linear relationship with body weight.

18. Suppose study participants are asked two questions. Based on a large number of participants, the probability distribution for the responses to Questions 1 and 2 is given below.

|            |           | Question 1 |           | Total |
|------------|-----------|------------|-----------|-------|
|            |           | Correct    | Incorrect |       |
| Question 2 | Correct   | 0.25       | 0.20      | 0.45  |
|            | Incorrect | 0.08       | 0.47      | 0.55  |
| Total      |           | 0.33       | 0.67      | 1     |

For one participant, let A = Question 1 is answered correctly and B = Question 2 is answered correctly. Which of the following must be true?

- (A) A and B are independent because each question is correct with probability 0.50, and the probability that both questions are answered correctly is  $(0.5)(0.5) = 0.25$ .
- (B) A and B are not independent because the probability in each cell is not 0.25.  
 *$P(A)P(B) \neq P(A \cap B)$*
- (C) A and B are not independent because  $(0.33)(0.45) \neq 0.25$ .
- (D) A and B are independent because one participant's answers do not affect another participant's answers.
- (E) There is not enough information provided to determine whether A and B are independent.

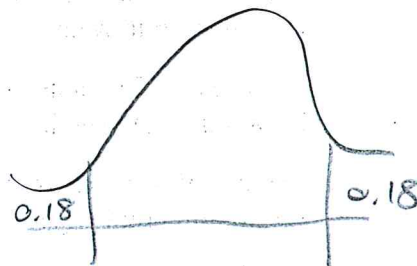
multiplication rule for independent events (see S.3 notes)

If independent  $P(A \cap B) = P(A) \cdot P(B)$

$P(\text{question 1 and question 2 correct}) = P(\text{question 1 correct}) \cdot P(\text{question 2 correct})$   
 $0.25 \neq (0.33)(0.45)$

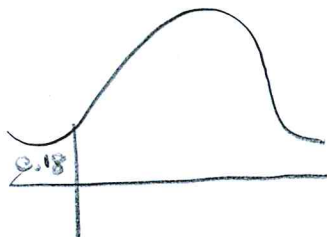
19. A researcher uses a set of sample data to test the hypothesis  $H_0: \mu = 2$  versus  $H_a: \mu \neq 2$  and the resulting  $p$ -value is 0.36. What would be the  $p$ -value if this same data were instead used to test  $H_0$  against the alternative hypothesis  $H_a: \mu < 2$ ?

- (A) The  $p$ -value would be 0.18.
- (B) The  $p$ -value would be 0.72.
- (C) The  $p$ -value would be 0.82.
- (D) The  $p$ -value could be 0.18 or 0.82.
- (E) The  $p$ -value would remain unchanged.

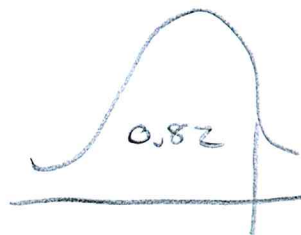


$\mu \neq 2$

Don't know whether  $t$  is positive or negative



$\mu < 2$   
if  $t$  is negative



$\mu < 2$   
if  $t$  is positive

20. A large-sample 99% confidence interval for the proportion of American adults that believe climate change is a serious problem is (0.61, 0.69). What is the sample proportion of American adults that believe climate change is a serious problem from which this interval was constructed?

- (A) 0.04
- (B) 0.08
- (C) 0.65**
- (D) 0.66
- (E) It cannot be determined from the information given.

$$\frac{0.61 + 0.69}{2} \rightarrow \text{point estimate}$$

---

21. According to data collected by the U.S. Census Bureau, in 1970 the average age of a mother at the birth of her first child was 21 years. A sociologist suspects that the average age of first-time mothers is much higher today and surveys a random sample of 35 mothers to test the hypotheses

$$H_0: \mu = 21$$
$$H_a: \mu > 21.$$

In these hypotheses, what does  $\mu$  represent?

- (A) The sample mean age of a mother at the birth of her first child in 1970.
- (B) The population mean age of a mother at the birth of her first child in 1970.
- (C) The mean age at the birth of their first child for the 35 mothers surveyed by this sociologist.
- (D) The current population mean age of a mother at the birth of her first child.**
- (E) The population proportion of mothers having their first child when they are over the age of 21.

22. A university statistics professor wants to know if including review problems in each set of homework problems (treatment I) is more effective than including only new problems (treatment II). He teaches three sections of the course: a morning, an afternoon, and an evening section, each with 30 students. Within each section the professor randomly assigns 15 students to treatment I and 15 students to treatment II. Compared to randomly assigning 45 students to each treatment, what is the advantage of randomly assigning 15 students to each treatment within each section?

- (A) Random assignment within section eliminates the placebo effect.
- (B) Random assignment within section allows the professor to generalize the results to all sections.
- (C) Random assignment within section permits the professor and students to be blinded as to the treatment group assignment.
- (D) Random assignment within section accounts for possible differences in performance due to the time of day the class meets.
- (E) Random assignment within section reduces the effect of nonresponse bias.

maybe students in night class are more tired; within the section though everyone would be equally tired

23. The Department of Health for a Midwestern state conducted an observational study in which users of public restrooms at several sites throughout the state were discreetly observed. Of the 634 females observed, 476 washed their hands. Of the 561 males observed, 326 washed their hands. What significance test should be used to determine whether these data provide sufficient evidence to reject the hypothesis that the proportion of females in this state that wash their hands when using a public restroom is the same as the proportion of males in this state that wash their hands when using a public restroom?

- (A) One proportion z-test  $H_0: p_F - p_M = 0$
- (B) Two proportion z-test  $H_a: p_F - p_M > 0$
- (C) One sample z-test for a mean
- (D) One sample t-test for a mean
- (E) Two sample t-test for means

$$\hat{p}_F = \frac{476}{634} = 0.751$$

$$\hat{p}_M = \frac{326}{561} = 0.581$$

24. A random sample of 30 students at Morningside High School was allowed to play a popular cell phone game for an hour in order to score as many points as possible. The sample yielded an average score of 375 and a standard deviation of 100. The resulting 90% confidence interval is  $375 \pm 31$ . Which of the following statements best summarizes the meaning of this confidence interval?
- (A) There is a 90% chance that a randomly selected student at Morningside High School will score between 344 and 406 points.
  - (B) Approximately 90% of the students at Morningside High School will have an average score between 344 and 406 points.
  - (C) If we repeated this procedure many times, about 10% of the confidence intervals constructed would fail to include the mean number of points scored by all students at Morningside High School.
  - (D) If we repeated this procedure many times, the average number of points scored by our sample of 30 students will be between 344 and 406 points approximately 90% of the time.
  - (E) Approximately 10% of all students at Morningside High School will score more than 406 points.

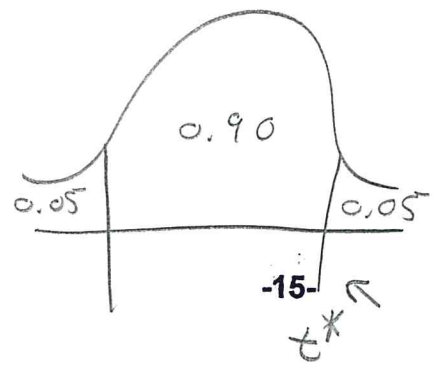
25. A recent survey randomly sampled 25 high school students. One question asked "How many times per day do you access a social media site such as Facebook, Instagram, etc.?" The average number of times reported in the sample was 5.6 with a sample standard deviation of 2.6. A 90% confidence interval for the population mean number of times per day high school students visit social media sites is

- (A)  $5.6 \pm 1.96 \left( \frac{2.6}{\sqrt{25}} \right)$
- (B)  $5.6 \pm 1.645 \left( \frac{2.6}{\sqrt{24}} \right)$
- (C)  $5.6 \pm 1.711 \left( \frac{2.6}{\sqrt{24}} \right)$
- (D)  $5.6 \pm 1.711 \left( \frac{2.6}{\sqrt{25}} \right)$
- (E)  $5.6 \pm 2.064 \left( \frac{2.6}{\sqrt{25}} \right)$

$$\bar{x} \pm t^* \left( \frac{s}{\sqrt{n}} \right)$$

$$5.6 \pm 1.711 \left( \frac{2.6}{\sqrt{25}} \right)$$

$\bar{x} = 5.6$   
 $s = 2.6$   
 $n = 25$   
 $t^* = \text{invT}(\text{area}, \text{df})$   
 $= \text{invT}(0.95, 24)$   
 $= 1.711$



Remember confidence intervals are ALWAYS two-tailed

26. To investigate whether vegetarians have lower cholesterol levels, on average, than non-vegetarians, a nutritionist obtained independent random samples of cholesterol levels for the two groups. The data are summarized below.

Z samp T test  
 $t = -1.845$   
 $p\text{-value} = 0.0337$   
 $\downarrow$   
 reject  $H_0$  at  $\alpha = 0.05$

|                    | Vegetarians | Non-vegetarians |
|--------------------|-------------|-----------------|
| Mean               | 198         | 207             |
| Standard Deviation | 32          | 41              |
| $n$                | 62          | 231             |

$H_0: \mu_V - \mu_N = 0$   
 $H_a: \mu_V - \mu_N < 0$   
 $\downarrow$   
 lower cholesterol

Based on the data, which of the following statements is true?

- (A) At the 5% significance level, vegetarians have a significantly lower mean cholesterol level than non-vegetarians do.
- (B) At the 1% significance level, vegetarians have a significantly lower mean cholesterol level than non-vegetarians do.
- (C) At the 5% significance level, vegetarians have a significantly higher mean cholesterol level than non-vegetarians do.
- (D) At the 1% significance level, vegetarians have a significantly higher mean cholesterol level than non-vegetarians do.
- (E) At the 10% significance level, there is no significant difference in cholesterol levels between vegetarians and non-vegetarians.

27. Historically, only 35% of applicants for a statistical programming internship at a large pharmaceutical company have previous experience with the necessary software. Assume that applicants are selected at random and interviewed sequentially. What is the probability it takes exactly four interviews to find the first applicant with previous experience?

- (A) 0.0150
- (B) 0.0279
- (C) 0.0961
- (D) 0.1115
- (E) 0.3845

the distribution is  
 geometric (we are going until we  
 achieve a success)  
 with  $p = 0.35$

geomet pdf ( $p = 0.35$ ,  $x$  value 4) =  
 0.0961

$\uparrow$   
 exact



28. A researcher selects a random sample of size  $n$  from a population and uses the collected data to compute a 95% confidence interval for the mean of the population. If the same data were used to compute a second confidence interval, which of the following would produce a confidence interval with a larger margin of error?

- (A) Increasing the confidence level → more confident, larger range of values
- (B) Decreasing the confidence level
- (C) Increasing the sample size
- (D) Computing the interval  $n$  times so that approximately 5% of these intervals will be larger
- (E) None of the above can guarantee a larger margin of error

$$\bar{x} \pm t^* \left( \frac{s}{\sqrt{n}} \right)$$

↑  
more confident, larger  $t^*$ , larger MOE

29. In the Atlantic Ocean, the hurricane season runs from June to November. The National Oceanic and Atmospheric Administration (NOAA) has stated that the probability of a hurricane making landfall in the United States in August 2015 is approximately 27%. Which of the following is the most reasonable explanation for how this probability could have been estimated?

- (A) In a large number of years for which hurricane activity was recorded, a hurricane made landfall in the U.S. during the month of August in 27% of these years.
- (B) Twenty-seven percent of the hurricanes that made landfall in the United States in 2014 occurred in August.
- (C) Historically, there have been 27 hurricanes that made landfall in the United States in August.
- (D) There have not been any hurricanes in June or July, which increases the chance of a hurricane to be above 1/6.
- (E) Of 500 randomly selected people in the United States, their average prediction was 27%.

30. Nearly 12,000 high school students across 11 different countries were surveyed about both their sleeping habits and their performance in school. Based on the results, researchers concluded that a lack of sleep is linked to students earning poor grades in school. Which of the following statements is true?

- (A) This is an observational study. Therefore, researchers cannot conclude that a lack of sleep causes poor grades.
- (B) This is an observational study. Therefore, researchers can conclude that a lack of sleep causes poor grades.
- (C) This study is a well-designed experiment. Therefore, researchers cannot conclude that a lack of sleep causes poor grades.
- (D) This study is a well-designed experiment. Therefore, researchers can conclude that a lack of sleep causes poor grades.
- (E) This is neither an observational study nor a well-designed experiment.

Cannot make causation claims from observational studies

31. The tires sold at Best Tires last an average of 80,000 miles with a standard deviation of 12,000 miles. Assume the wear of tires is approximately normally distributed, and the wear of each tire is independent. Which of the following describes the sampling distribution of the mean number of miles that a set of 4 randomly selected tires purchased from Best Tires will last?

- (A) Approximately normal with mean 320,000 miles and standard deviation 48,000 miles.
- (B) Approximately normal with mean 80,000 miles and standard deviation 3,000 miles.
- (C) Approximately normal with mean 80,000 miles and standard deviation 6,000 miles.
- (D) Approximately normal with mean 80,000 miles and standard deviation 12,000 miles.
- (E) The shape of the sampling distribution is unknown because the sample size is not more than 30.

$$\mu_{\bar{x}} = \mu = 80000$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{12000}{\sqrt{4}} = 6000$$

(7.3 notes)

11.2

32. A survey conducted by a national news network asked a random sample of U.S. adults whether they get most of their information about current events from newspapers, television, the internet, or some other source. The results, shown in the table below, are reported by age group of the respondents.

|         | Newspapers | Television | Internet | Other | Total |
|---------|------------|------------|----------|-------|-------|
| 18 - 34 | 12         | 35         | 40       | 6     | 93    |
| 35 - 54 | 16         | 55         | 20       | 8     | 99    |
| Over 55 | 33         | 60         | 5        | 5     | 103   |
| Total   | 61         | 150        | 65       | 19    | 295   |

If primary news source is independent of age group, which of the following expressions is equal to the expected number of respondents who are aged 35 to 54, inclusive, and get most of their information about current events from the internet?

- (A)  $\frac{(99)(65)}{295}$
- (B)  $\frac{(99)(150)}{295}$
- (C)  $\frac{(20)(65)}{99}$
- (D)  $\frac{(20)(99)}{295}$
- (E)  $\frac{(20)(65)}{295}$

row total  $\times$  column total  
table total

$$\frac{99 \cdot 65}{295}$$

33. If a statistic has a sampling distribution whose mean is different from the true value of the parameter which the statistic is estimating, the statistic is said to be \_\_\_\_\_.

- (A) biased
- (B) random
- (C) unbiased
- (D) efficient
- (E) consistent

34. Suppose you are responsible for helping an insurance company make a profit equal to 1% of the home value on an insurance policy for a \$215,000 home. In this type of policy, we assume the probability distribution for the loss in any year, expressed as a percentage of the home value, is given below. At what price should you sell the policy?

| Loss        | 0%    | 25%   | 50%   | 75%   | 100%  |
|-------------|-------|-------|-------|-------|-------|
| Probability | 0.980 | 0.011 | 0.006 | 0.002 | 0.001 |

- (A) \$443.44
- (B) \$1773.75
- (C) \$2150.00
- (D) \$3923.75**
- (E) \$4300.00

Expected loss

$$E(\text{loss}) = 0(0.980) + 0.25(0.011) + 0.5(0.006) + 0.75(0.002) + 1(0.001)$$

$$\text{Expected loss} = 0.00825$$

on this home:  $215000(0.00825) = \$1773.75$

required profit:  $215000(0.01) = \underline{\$2150}$

sell policy for:  $\$3923.75$

↑  
after the \$1773.75 expected loss, they still make \$2150 profit

35. Which of the following random variables has a distribution which most closely resembles a normal distribution?

- (A) A random number between 0 and 9 from a table of single-digit random numbers *uniform*
- (B) The number of car accidents in 1 month at a very rural intersection
- (C) The score of a student on a very easy 10-point quiz *left skewed*
- (D) The annual income of a U.S. citizen
- (E) The actual weight of a jar of peanut butter labeled as 16 ounces** *most value in middle*

*as you move away from the mean, values decrease, forming a Normal curve*

8.2 notes

36. A statewide poll is to be conducted using a random sample of high school seniors who plan to attend college to determine if they plan to attend a college in-state or out-of state. Let  $p$  represent the population proportion of such students who plan to stay in-state. How large a sample,  $n$ , will the pollsters need in order to estimate  $p$  with a margin of error of 0.03 with 95% confidence? Historically, 70% of high school graduates stay in-state for college.

- (A) 14 high school seniors
- (B) 632 high school seniors
- (C) 897 high school seniors**
- (D) 1,057 high school seniors
- (E) 1,548 high school seniors

$$z^* \sqrt{\frac{p(1-p)}{n}} \leq \text{MOE}$$

$$1.96 \sqrt{\frac{0.7(0.3)}{n}} \leq 0.03$$

$$\frac{0.21}{\left(\frac{0.03}{1.96}\right)^2} \leq n$$

$z^* : \text{invNorm}(\text{area} = 0.975, \mu = 0, \sigma = 1) = 1.96$

$$\sqrt{\frac{0.21}{n}} \leq \frac{0.03}{1.96}$$

$$\frac{0.21}{n} \leq \left(\frac{0.03}{1.96}\right)^2$$

$$0.21 \leq \left(\frac{0.03}{1.96}\right)^2 n$$

$$896.4 \leq n$$

$\hat{p} = 0.7$   
 $n = ?$   
 $\text{MOE} = 0.03$

37. The American Academy of Pediatrics advises against feeding babies solid foods before they reach 4 months old. A research group obtains a simple random sample of 5,000 mothers nationwide and mails each one a survey asking questions about their feeding practices. Of the 1,000 mothers who completed the survey and mailed it back, 40% indicated that they started feeding their baby solid foods before age 4 months. A newspaper reports on the study with the headline, "Infants are Fed Solid Foods Too Soon." Which of the following is the most serious concern regarding this study?

- (A) There was no control group.
- (B) There could be serious non-response bias.**
- (C) The sample size was too small.
- (D) There was no random assignment.
- (E) The sampling design did not incorporate stratification.

*→ maybe mothers that feed their babies non-solids think the survey is silly and throw it away*

38. An airline with a poor on-time record decides to offer its employees a bonus if, in an upcoming month, the airline's proportion of on-time flights exceeds the overall industry rate of 0.79. A random sample of flights will be selected and used as a basis for choosing between

$$H_0: p = 0.79$$

$$H_a: p > 0.79$$

Which of the following describes a consequence of making a Type I error in this situation?

- (A) The airline employees receive a reward that they do in fact deserve.
- (B) The airline employees receive a reward that they do *not* in fact deserve.
- (C) The airline employees do *not* receive a reward that they do in fact deserve.
- (D) The airline employees do *not* receive a reward that they do *not* in fact deserve.
- (E) A Type I error cannot be committed in this situation.

→ you think  $H_a$  is true, that the proportion of on-time flights is greater than 0.79, which leads to the employees getting a bonus, but it really is *not* true, so the bonus was undeserved seriously???

39. While hiking in the woods, you and your friends are abducted by aliens. The aliens want to study the human race, so they will place you in one of two groups based on your results of a ten question true-false test. Unfortunately, since you don't understand their language, you'll have to guess at the answers. Based on your observations of the people who were tested before you, if you get less than 8 correct, you are placed in the "experimentation" group. What is the probability that you will escape experimentation?

- (A) 0.0107
- (B) 0.9893
- (C) 0.9453
- (D) 0.0547
- (E) 0.0439

this is a binomial distribution with  $p = 0.5$  and  $n = 10$

↓  
meaning find  
 $P(X \geq 8)$

↓  
do  
 $1 - P(X \leq 7)$

$$1 - \text{binomcdf}(\text{trials: } 10, p: 0.5, x \text{ value: } 7) = 0.0547$$

12.1

40. A real estate agent collected information on the selling price (in dollars) and size (in square feet) of 6 recently sold homes in her area. A regression analysis was performed and the partial computer output is given below.

| Model Summary                |        |           |            |         |      |  |
|------------------------------|--------|-----------|------------|---------|------|--|
| S                            | R-sq   | R-sq(adj) | R-sq(pred) |         |      |  |
| 27484.1                      | 97.28% | 96.60%    | 28.80%     |         |      |  |
| Coefficients                 |        |           |            |         |      |  |
| Term                         | Coef   | SE Coef   | T-Value    | P-Value | VIF  |  |
| Constant                     | -19061 | 23090     | -0.83      | 0.455   |      |  |
| Size                         | 112.73 | 9.43      | 11.96      | 0.000   | 1.00 |  |
| Regression Equation          |        |           |            |         |      |  |
| Price = -19061 + 112.73 Size |        |           |            |         |      |  |

Which of the following should be used to compute a 95% confidence interval for the slope of the regression line?

- (A)  $-19061 \pm (2.132)(23090)$
- (B)  $-19061 \pm (2.776)(23090)$
- (C)  $112.73 \pm (1.96)(9.43)$
- (D)  $112.73 \pm (2.132)(9.43)$
- (E)  $112.73 \pm (2.776)(9.43)$

$b \pm t^* SE_b$   
 $112.73 \pm (2.776)(9.43)$   
 $df = n - 2$   
 $6 - 2 = 4$   
 $t^*_{0.025, 4}$   
 $\approx 2.776$

END OF EXAMINATION



Find the area of a square with side length 5 cm.

Area = side  $\times$  side  
Area = 5  $\times$  5  
Area = 25 cm<sup>2</sup>

Find the area of a rectangle with length 8 cm and width 3 cm.

Area = length  $\times$  width  
Area = 8  $\times$  3  
Area = 24 cm<sup>2</sup>

Find the area of a triangle with base 6 cm and height 4 cm.

Area =  $\frac{1}{2}$   $\times$  base  $\times$  height  
Area =  $\frac{1}{2}$   $\times$  6  $\times$  4  
Area = 12 cm<sup>2</sup>

Find the area of a circle with radius 3 cm.

Area =  $\pi r^2$   
Area =  $\pi \times 3^2$   
Area =  $9\pi$  cm<sup>2</sup>

Find the area of a trapezium with parallel sides of length 5 cm and 7 cm, and height 4 cm.

Area =  $\frac{1}{2}$   $\times$  (sum of parallel sides)  $\times$  height  
Area =  $\frac{1}{2}$   $\times$  (5 + 7)  $\times$  4  
Area = 24 cm<sup>2</sup>

Find the area of a parallelogram with base 8 cm and height 5 cm.

Area = base  $\times$  height  
Area = 8  $\times$  5  
Area = 40 cm<sup>2</sup>

Find the area of a rhombus with side length 5 cm and one diagonal of length 6 cm.

Area =  $\frac{1}{2}$   $\times$  (product of diagonals)  
Area =  $\frac{1}{2}$   $\times$  6  $\times$  8  
Area = 24 cm<sup>2</sup>

Find the area of a square with side length 10 cm.

Area = side  $\times$  side  
Area = 10  $\times$  10  
Area = 100 cm<sup>2</sup>

Find the area of a rectangle with length 12 cm and width 5 cm.

Area = length  $\times$  width  
Area = 12  $\times$  5  
Area = 60 cm<sup>2</sup>