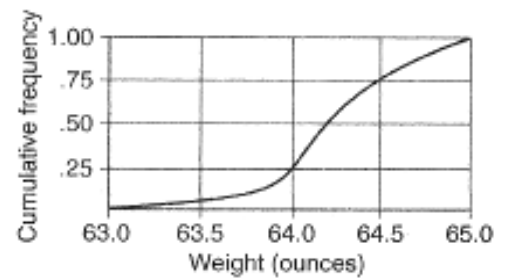


MULTIPLE CHOICE Decide which is the best of the choices given. You may wish to write your answers to these multiple choice questions on a separate sheet (i.e., an answer sheet).

- A distribution of 6 scores has a median of 21. If the highest score increases 3 points, the median will become:
A) 21 B) 21.5 C) 24 D) cannot be determined E) none of these
- If you are told a population has a mean of 25 and a variance of 0, what must you conclude?
A) Someone has made a mistake D) All the elements in the population are 25.
B) There is only one element in the population. E) none of the above
C) There are no elements in the population

- A cumulative relative frequency plot of the number of ounces in a sample of 200 half-gallon orange juice containers is shown at the right.



Which of the following is a valid conclusion?

- Which of the following is a valid conclusion?
A) The median number of ounces is 64.0 ounces.
B) The interquartile range is about 0.5 ounces.
C) 25% of the containers had more than 64.0 ounces.
D) None of the above are valid conclusions.
- A sports columnist wishes to portray basketball players as *overpaid*. Which measure of center should he report as the average salary of NBA players? (Many players make large salaries and a small handful make **ASTRONOMICAL** salaries.)
A) the mean
B) the median
C) either the mean or median since they will be equal
D) neither the mean nor the median, since both will be much lower than the actual average salary
E) none of the above

- The following are all resistant statistics (resistant to the effects of an outlier)
A) mode, median, mean B) Q_1 , median, Q_3 , IQR C) median, mean
D) range, mean, standard deviation E) none of the above

- Consider the following data: 53, 61, 38, 65, 72, 58, 52, 63, 69, 74, 66
You are given: $\sum(x - \bar{x})^2 = 1082$ and $\sum x = 671$

The sample mean and the sample variance are

- A) 98.4, 67.1 B) 61.0, 108.2 C) 67.1, 98.36 D) 108.2, 67.1

- A distribution is _____ if the portions greater and less than its center are mirror images of each other.
A) skewed right B) bimodal C) truncated D) skewed left E) symmetric

8. Ethnicity is a _____ variable and height is a _____ variable.
- A) quantitative, categorical B) categorical, quantitative
C) quantitative, response D) response, explanatory E) none of these
9. The heights of American men aged 18 to 24 are approximately normally distributed with mean 68 inches and standard deviation 2.5 inches. Only about 5% of young men have heights outside the heights from
- A) 65.5 to 70.5 inches B) 63 to 73 inches
C) 60.5 to 75.5 inches D) 58 to 78 inches
10. By definition, a simple random sample is one such that
- A) every possible sample of size n has an equal chance of being selected
B) each n th person has an equal chance of being selected.
C) each element in every stratum has an equal chance of being selected.
D) None of these are a part of the definition.
11. National test A has a normal distribution with a mean of 500 and a standard deviation of 100, while national test B has a normal distribution with a mean of 18 and a standard deviation of 6. A student scored 677 on test A and 29 on test B. Relative to the respective distributions, which score is better?
- A) The score of 677 on Test A is better.
B) The score of 29 on Test B is better.
C) The scores are equally strong.
D) A comparison cannot be made without knowing the number of students who took each exam.
12. The Empirical Rule can be used when assessing a distribution if
- A) the distribution is approximately normal
B) the distribution is skewed
C) the distribution is heavy tailed
D) the standard deviation is close to the interquartile range
E) the mean is equal to 0, and the standard deviation is equal to 1.
13. Which of the following are important in the design of comparative experiments?
- I. Randomization in assigning subjects to different treatments
II. Control of sources of variation other than the factors that are being tested
III. Replication of the experiment using sufficient numbers of subjects
- A) I and II
B) I and III
C) II and III
D) I, II, and III
E) none of the above
14. What does the correlation coefficient, r , represent?
- A) the strength of the relationship between any two variables
B) the strength of the relationship between any two numerical variables
C) the strength of the linear relationship between any two variables
D) the strength of the linear relationship between any two numerical variables
E) the proportion of variation in Y that is explained by gravity

15. The data on frying time (seconds) and moisture content (%) of tortilla chips appeared in the Food Processing and Preservation magazine. Since there was a distinct curvature in the scatterplot, the observations were re-expressed and the following linear model was deemed appropriate.

$$\text{predicted log(moisture)} = 2.01 - 1.05 \text{ log(time)}$$

Predict the moisture content when the time is 20 seconds.

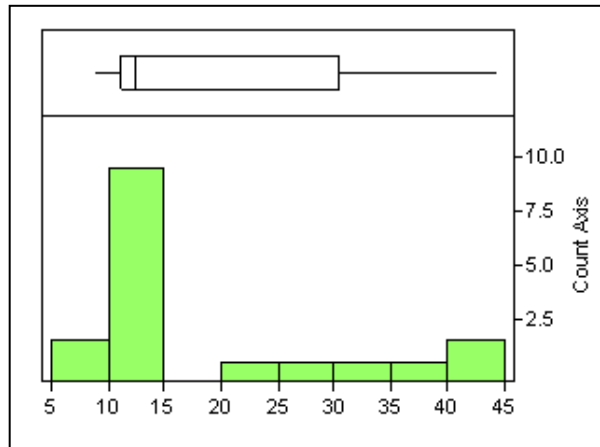
- A) 4.40% B) 1.90% C) 0.64% D) -18.99%
16. Coach F.T. Ball thinks that ballet training will improve the performance of his football team! He decides to randomly select half of the team to take six weeks of ballet training before the football season begins while the other half uses the same amount of time to study past football game films. He will then compare agility scores of group A (with ballet training) and group B (without ballet training) by comparing the mean of group A with the mean of group B. This study would be classified as:
- A) an experiment with a control group
B) an experiment with blocking
C) an observational study with comparison and randomization
D) an observational study with little if any bias
17. Using the information in **the previous question**, ballet training is
- A) blocking B) a treatment C) a control D) a placebo
18. Suppose that a Normal model described student scores in a history class. Sarah Bellum has a standardized score (z-score) of +2.5. This means that Sarah
- A) is 2.5 points above average for the class.
B) is 2.5 standard deviations above average for the class.
C) has a standard deviation of 2.5.
D) has a score that is 2.5 times the average for the class.
E) None of the above.
19. In regression analysis, an influential point is:
- A) an observation with a large residual
B) an observation with the largest Y value
C) an observation which, if deleted, leads to a dramatic change in the slope of the least squares line
D) an observation that is far from the bulk of the data
E) an observation that is right on the least squares line
20. You must choose a simple random sample of five from a lot of 64 vials of a pharmaceutical product for study. You label the vials beginning with 01 and use the random digit table copied below beginning with the start of the first line reading left to right to choose your sample.
- 68417 35013 15509 72765 85089 57067 50211 47487 82739 57890
- Your sample contains the vials labeled
- A) 41, 50, 13, 15, 50 B) 41, 50, 13, 15, 27
C) 68, 41, 73, 13, 15 D) 41, 73, 50, 13, 15

21. The issue in research design that relates to the researcher's lack of awareness of the experimental or control status of the subjects is referred to as
- A) quota sampling B) blinding C) placebo effect
 D) statistical regression E) selection bias
22. A distribution can have more than one
- A) mean D) mode
 B) interquartile range E) median
 C) standard deviation

Use the following information to answer questions # 23 - #24.

In a study of male / female differences in carnivores, the height of the canine teeth in the lower jaws were measured. The data below are graphic representations of these data.

Boxplot and histogram of Lower canine tooth height (mm)



Stemplot of Lower canine tooth height (mm) n = 18

```

0|99
1|0111223344      1|4 represents 14 mm
2|4
3|03
4|014
  
```

23. The median of the lower canine tooth heights is:
- A) 10 B) 11 C) 12 D) 13 E) 14
24. Considering the graphic displays, the best description of these data would be:
- A) Skewed to the lower tooth heights
 B) Skewed to the upper tooth heights
 C) Symmetric
 D) Bimodal
 E) Light tailed

25. Among a dozen eggs, three are rotten. A cookie recipe calls for two eggs; they will be selected randomly from that dozen. (Why, you ask? Answer: no clue!) Which plan could be used to simulate the number of rotten eggs that might be chosen?
- A) Randomly generate a 0, 1, or 2 to represent the rotten eggs you get.
 B) Since 25% of the eggs are rotten, let 0 = rotten, 1, 2, 3 = good. Generate two random numbers 0 – 3 and see how many 0's you get.
 C) Let 0, 1, and 2 represent the rotten eggs, and 3, 4, ..., 11 the good eggs. Generate two random numbers 0 – 11, ignoring repeats.
 D) Let 1 and 2 represent the rotten eggs and 3 – 8 represent the good eggs. Randomly generate two numbers from 1 – 8.
26. McNeil High School has decided to try to attract more students to eat in its cafeteria. To do so, it decided to put little trinkets (toys) on each plate lunch it serves that students could play with in their statistics class. (Of course their teachers are so good-natured, they don't mind.) MHS has decided to include 5 different trinkets. The trinkets are randomly distributed in all the plate lunches. To run a simulation to find out how many lunches have to be eaten in order to collect all five, we give the following single digit assignments:

Trinket #1: **1** Trinket #2: **2** Trinket #3: **3**
 Trinket #4: **4** Trinket #5: **5** ignore digits 0, 6 - 9

Use the random numbers below to run one trial. What is the **count** for this trial?

1 9 2 2 6 9 5 0 5 4 0 5 7 5 6 2 8 7 1 3 2 7 7 5 4 2 9 4 8 5 8 2 2 2 6
 9 0 0 0 5 9 8 6 2 4 6 6 9 7 9 3 5 4 3 5 6 2 2 2 4 5 4 1 0 2 2 2 2 0 2

- A) 9 B) 11 C) 12 D) 13 E) 14

You draw an M&M from a large bag of M&Ms with outcomes governed by the probabilities given below:

Color	brown	red	yellow	green	orange	blue
Probability	0.3	0.2	0.2	0.1	0.1	0.1

27. The probability of choosing a single M&M that is brown or red is:

- A) 0 B) 0.06 C) 0.5 D) 0.6

28. The probability of choosing a single M&M that is brown and red is:

- A) 0 B) 0.06 C) 0.5 D) 0.6

29. The mean height (x) of American women in their early twenties is about 65.5 inches and the standard deviation is about 2.5 inches. The mean height (y) of men the same age is about 68.5 inches with standard deviation about 2.7 inches. If the correlation between the heights of husbands and wives is about $r = 0.5$, what is the **slope of the regression line** for predicting husband's height using the wife's height for young couples?

- A) 0.46 B) 0.50 C) 0.52 D) 0.54

30. A correlation between college entrance exam grades and scholastic achievement was found to be -1.08 . On the basis of this you would tell the university that:
- A) the entrance exam is a good predictor of success
 - B) they should hire a new statistician
 - C) the exam is a poor predictor of success
 - D) students who do best on this exam will make the worst students
 - E) students at this school are underachieving

A local trade union consists of plumbers and electricians.

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

31. A member of the union is selected at random. Given that the person selected is a plumber, the probability that he/she is a journeyman is:
- A) $1/2$
 - B) $1/3$
 - C) $4/15$
 - D) $2/15$
 - E) none of these
32. What is the probability that a randomly selected local trade union member is a master plumber?
- A) $1/6$
 - B) $1/5$
 - C) $1/4$
 - D) $1/3$
 - E) $1/2$
33. Which of the following is a true statement?
- A) Being an apprentice and being a plumber are dependent events.
 - B) Being an apprentice and being a plumber are independent events.
 - C) More information is needed to decide about independence of being an apprentice and being a plumber.

In a statistics course a linear regression equation was computed to predict the final exam from the score on the first test. The equation was

$$(\text{predicted score on final exam}) = 10 + .9(\text{score on first test}).$$

Eve scored 95 on the first test.

34. On the final exam Eve actually scored 98. What is the value of her residual?
- A) 98
 - B) 2.5
 - C) -2.5
 - D) 0.5
 - E) -0.5
35. A least square regression line was computed to be $\hat{y} = 16.1 - 6.7x$. The correlation coefficient r is
- A) positive
 - B) negative
 - C) either positive or negative
 - D) zero

For #36 - 40, use the following information:

In order to ease parking problems in a community containing a university, university officials propose purchasing one acre of community parkland that is adjacent to the university to build a parking garage. The officials believe community members will overwhelmingly support this proposal, and they would like to conduct a survey of 100 community members to confirm their belief.

Use the following words to identify the **type of sampling** proposed below for this survey. Indicate your choice by using the **letter** of the best choice.

- (A) convenience sampling (C) simple random sampling (E) stratified random sampling
(B) cluster sampling (D) voluntary response sampling (F) systematic random sampling

36. Using the latest census data from the community, numbering the residents and using a random number table to choose 100 people.
37. Selecting 100 people from the local phone directory.
38. Using the latest census data from the community and randomly choosing 25 residents ages 18 – 25, 25 residents ages 26 – 39, 25 residents ages 40 – 64, and 25 residents age 65 and over.
39. Using the latest census data from the community, randomly choose the first person from among the first 50 on the list and then select every 50th person after that.
40. Recording the opinion of the first 100 people who call the university regarding this issue.
-

For #41 – 42, use the following information:

A commuter must pass through five traffic lights on her way to work and will stop at each one that is red. She estimates the probability model for the number of red lights she hits, as shown below.

x = # of red	0	1	2	3	4	5
P(x)	.05	.25	.35	.14	.16	.05

41. How many red lights should she expect to hit each day?
A) 2.25 B) 2.26 C) 3.00 D) 3.25 E) 3.26
42. What is the variance for this distribution?
A) 1.00 B) 1.06 C) 1.61 D) 1.86 E) 1.27 F) 2.00
-
43. Which of the following is true about independent events?
A) $P(A | B) = 0$ B) $P(A | B) = P(B)$ C) $P(A | B) = P(A)$ D) $P(A | B) = 1$
44. Five Republicans and five Democrats have applied for two open positions on a City planning committee. Since all the applicants are qualified to serve, the City Council decides to pick the two new members randomly from the group of applicants. What is the probability that both of the applicants that are chosen are Republicans?
A) 1/4 B) 2/9 C) 1/3 D) 1/2
45. The probability that a certain virus infects a random person is 0.2. If a random sample of 15 people is taken, what is the probability that at least one person is infected?
A) 0.13 B) 0.23 C) 0.30 D) 0.87 E) 0.96

FREE RESPONSE QUESTIONS

1. The stem-and-leaf display measures the salary of the employees at the MAVS Manufacturing company.

- State the 5 number summary.
- What shape best describes the distribution?
- Construct a box plot for the distribution.
Show your 1.5IQR test for outliers.

```

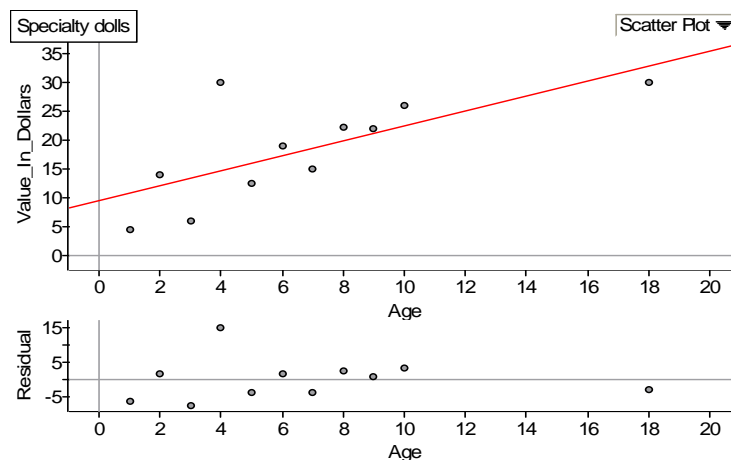
1 | 2578889
2 | 34467889
3 | 034799
4 | 124
5 | 0
6 |
7 | 2
8 | 3
9 | 4
    |
    +----- 7|2 represents $72,000
  
```

2. You are asked to design a randomized comparative experiment which studies the effect of the drug gemfibrozil in preventing heart attacks in middle-aged men with high blood cholesterol. There are a total of 2500 men and both the drug gemfibrozil and a placebo will be used.

- Describe how you would conduct a completely randomized design for this study.
- About half of the 2500 men in the study participate in regular exercise. Let us suppose that regular exercise is known to have an association with the effectiveness of the drug gemfibrozil. Describe the changes you would make to your design in part (a) if you wish to incorporate blocking.

3. A doll enthusiast owns a collection of specialty dolls, and the age and value of each is noted in the chart below.

Specialty dolls		
	Age	Value_In_Dollars
1	1	4.6
2	2	14
3	3	6.1
4	4	30
5	5	12.5
6	6	19
7	7	15
8	8	22.4
9	9	22
10	10	26
11	18	30



- If the least squares regression equation is $\hat{y} = 9.7 + 1.3x$, rewrite this equation in context.
- What does the residual plot tell you about the appropriateness of a linear model for this data? Explain.
- State the value of the slope (don't forget units!) and interpret the slope in context.
- For the value (4, 30), what is the residual for that value? Estimate this from the graph **AND** calculate it.

4. A random sample of 11 high school students produced the following results for number of hours of television watched per week and GPA. A computer printout of the regression analysis is shown below.

Predictor	Coef	StdDev	T	P
Constant	3.8	2.0426	1.86	0.05
Hours	-0.0558	0.01769	-3.154	0.012

$s = 0.355$ $R\text{-sq} = 53\%$

What is the least squares regression equation in context?

5. A die is weighted so that the probability of rolling a "6" is 0.48. The die is rolled 18 times.
- Find the probability that the die lands on a "6" exactly 11 times.
 - Find the probability that the die lands on a "6" either 7 or 8 times.
 - Find the probability that the die lands on a "6" no more than 5 times (*this means 5 times or fewer*).
 - Find the probability that the die lands on a "6" at least 4 times.

6. X and Y are two independent random variables with the following attributes:

$$E(X) = 13 \quad SD(X) = 4 \quad E(Y) = 22 \quad SD(Y) = 6$$

Find the mean and standard deviation of each of these random variables:

- $3Y$
 - $X_1 + X_2 + X_3$
 - $3X + Y$
7. At a charity fundraiser, a game of chance is designed such that the average payout on each play of the game has an average of \$1.52, with a standard deviation of \$8. The charity decides to charge \$2.00 to play the game each time. Based on last year's event, the charity anticipates that this game will be played 1000 times.
- If the game is played exactly 1000 times, what are the mean and standard deviation of the charity's **profits** from this game?
 - Based on the mean and standard deviation calculated in part (a), what is the probability that the charity will NOT make a profit from this game for 1000 plays? Assume that the profits for 1000 plays of this game are normally distributed.