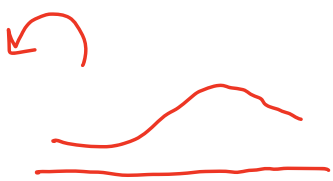
Here is a stemplot of the scores of Bob the Bowler's last 18 bowling games, made by Minitab. **Use this graph for questions 1-3**

Stem-and-leaf of Bowling N = 18
Leaf Unit = 1.0

1	9	4
1	10	
1	11	
2	12	8
4	13	12
7	14	346
(6)	15	147799
5	16	01445

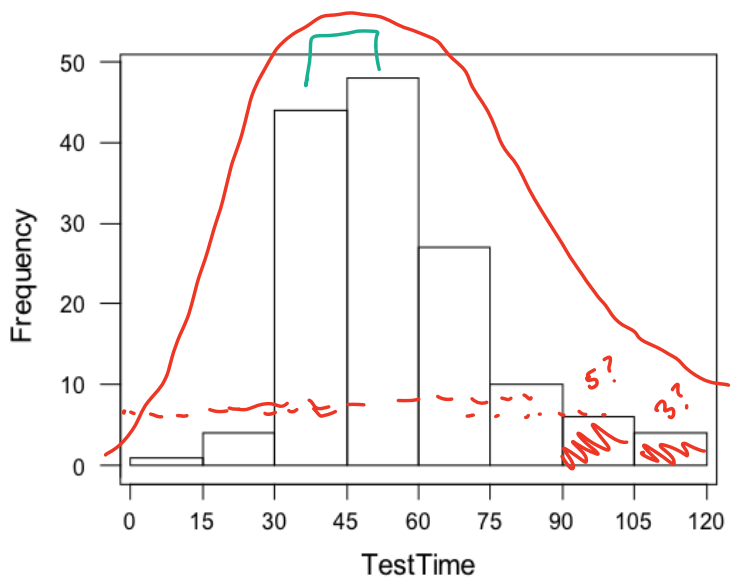


1. What is the best description for the shape of this graph?
   a) Bell/Mound-shaped
   b) Skewed to the left
   c) Skewed to the right
   d) Uniform

2. Where is the center of this graph?
   a) Around 105
   b) Around 135
   c) Around 155
   d) Around 165

3. Does this graph have any outliers?
   a) Yes, 94 is an outlier
   b) Yes, 165 is an outlier
   c) No, the numbers are too close together
   d) No, N is too small to have outliers

One semester, the TA for a math class decided to time how long it took her students to finish their Final Exam. Below is a histogram of the results. **Use this graph for questions 4-6**



4. About how many students took longer than 90 minutes to finish their test?
   a) About 5
   b) About 10
   c) About 20
   d) About 105

5. Where is the center of this graph?
   a) Between 15 and 30 minutes
   b) Between 45 and 60 minutes
   c) Between 75 and 90 minutes
   d) Between 105 and 120 minutes

6. How will the mean and median of these test times relate, based on the graph?
   a) The mean will be slightly larger than the median
   b) The mean will be slightly smaller than the median
   c) The mean and median will be exactly equal.
   d) It is impossible to tell from this graph

skew right: median < mean

skew left: median > mean

The Gators' Woman Basketball team has played 19 games as of 1/25/02. Here is a stemplot, made by Minitab, of the number of points the Lady Gators have scored in each game so far. **Use this graph for questions 7-10**

Stem-and-leaf of Scores N = 19
Leaf Unit = 1.0

1	4	9
3	5	19
7	6	1248
→ (5)	7	03583
7	8	01348
2	9	03

75

7. What is the median of this data?
   a) 10
   b) 73
   c) 75
   d) 78

8. What is the IQR (Interquartile Range) of this data?
   a) 10
   b) 13
   c) 21
   d) 62

$Q_3 - Q_1$

$83 - 62$

21

9. What is the standard deviation of this data?
   a) 12.78
   b) 44
   c) 93
   d) 163.33

$\frac{1387}{19} \rightarrow \bar{x} = 73$

$$\sqrt{\frac{(x_1-\bar{x})^2 + (x_2-\bar{x})^2 + (x_3-\bar{x})^2 + \cdots + (x_{19}-\bar{x})^2}{19-1}}$$

10. If z is a standard normal random variable, what is the variance of z?
    a) 0
    b) 1
    c) 4
    d) It depends on what kind of standard normal variable you have

$\mu = 0 \quad \sigma = 1 \quad$ variance $= \sigma^2$

A particular type of 4th grade Achievement Test provides overall scores that are normally distributed with a mean of 50 and a standard deviation of 10. **Use this information for questions 11-14**

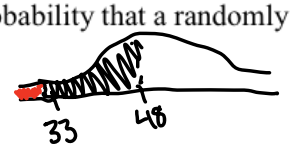
11. What is the probability that a randomly selected student earns a score of at least 42?
    a) .7881
    b) .2881
    c) .2119
    d) .1921

$z = \frac{x-\mu}{\sigma} = \frac{42-50}{10} = \frac{-8}{10} = -0.8$

$1 - 0.2119$

0.7881

12. What is the probability that a randomly selected student earns a score between 33 and 48?
    a) .3761
    b) .4207
    c) .4653
    d) .0446



$z_{48} = \frac{48-50}{10} = -0.20$ $\quad z_{33} = \frac{33-50}{10} = -1.7$

$P_1 = 0.4207$ $\quad P_2 = 0.0446$

$0.4207 - 0.0446$

$= 0.3761$

13. One state wants to allow all students with scores in the top 3% into a special advanced program. What will be the minimum score required to be admitted into this program?
    a) 1.88
    b) 31.2
    c) -1.88
    d) 68.8

$P = 0.03$

$1 - 0.03$

$= 0.97$

$z = 1.88 = \frac{x-50}{10}$

$18.8 = x - 50$

$x = 68.8$

14. Suppose that after the first exam you compute the z-score that corresponds to your exam score on Exam 1. Your z-score was -1.34. Which of the following can you say about your score on Exam 1?
    a) You scored above the mean on Exam 1.
    b) You scored exactly the mean on Exam 1.
    c) You scored below the mean on Exam 1.
    d) This can not be determined from the above information.

mean: $\frac{50-50}{10} = 0$

less: $\frac{40-50}{10} = -0.1$

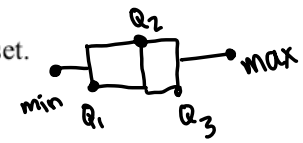
more: $\frac{60-50}{10} = 0.1$

17. You find old research on this species of whales that only reports the z-scores of lengths and not the actual lengths of whales. One particular whale stands out to you because it has a z-score of 5.2. What does the z-score tell you?

a) This whales is much larger than the mean.
b) This whale is much smaller than the mean
c) The whale is an average length
d) Not enough information is given.

18. Which of the following is a **true** statement? (Assume that the first quartile and minimum for this data set are not equal.)

a) the IQR is always bigger than the range for a given data set.
b) the IQR is always smaller than the range.
c) the IQR is equal to the range.
d) this cannot be determined from the given information.



range = max - min
IQR = Q3 - Q1

19. Which of the following is not true about normal curves?

a) They all have mean 0.
b) They are all symmetric.
c) The area underneath the curve is equal to 1.
d) They are all bell shaped.

20. What is the biggest advantage of the standard deviation over the variance?

a) The standard deviation is always smaller than the variance.
b) The standard deviation is calculated with the median instead of the mean.
c) The standard deviation is better for describing skewed distributions.
d) The standard deviation is in the same units as the original data.

var: $\frac{(x_i - \bar{x})^2}{n-1}$ → unit²

sd: $\sqrt{\frac{(x_i - \bar{x})^2}{n-1}}$ → unit

For questions 24 An orthopedic surgeon treats many women for back pain. She suspects that one common carried item, the woman's purse, might contribute to this, especially if the purse was heavy. She sampled the purses of 44 women with back pain who were clients at the clinic and got these statistics:

Variable	N	Mean	Median	TrMean	StDev	SE Mean
pursepai	48	4485	4000	4143	2958	427

Variable	Minimum	Maximum	Q1	Q3
pursepai	1500	16000	2900	4875

24. What can we say about the shape of this distribution by looking at the output?

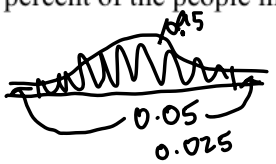
a) symmetric
b) skewed right
c) skewed left
d) It cannot be determined from the information given.

mean > median

For questions 27-30 A popular news magazine wants to write an article on how much Americans know about geography. They devise a test that lists 100 cities in the US, all of them mentioned in the news magazine in the last year. Each respondent must guess the state in which the city can be found. Some examples were: (Los Angeles, Tuscon, Biloxi.) Each correct answer earns one point, for a maximum of 100. The random sample of 5000 people had a distribution of scores that was normally distributed with mean 62 and standard deviation 12. $N=5000$ $\bar{x}=62$ $s=12$ $s=\frac{\sigma}{\sqrt{N}}$

27. The central ninety-five percent of the people in this sample can identify how many states correctly?

a) 38-86
b) 50-86
c) 50-74
d) 26-98



$p=0.025$ $z=\pm 1.96$

$1-0.025$
0.975

$-1.96=\frac{x-62}{12}$
$-23.52=x-62$
$x=38.48$

$1.96=\frac{x-62}{12}$
$23.52=x-62$
$85.52=x$

Questions 31-34 use the following scenario:
Suppose that you have decided to buy an ice cream truck to go into the ice cream business this summer instead of getting a summer job. You collected data every day last summer while working for an ice cream company about the temperature (in °F) and sales (in dollars) for that day as a way to research for your new business. You decided to fit a regression line and get the following based off of your data

$\text{slope}=\frac{y}{x}$

$$\text{Sales} = -762 + 18.53*\text{Temperature} \qquad R^2 = 47.1\%$$

(y = Sales, x = Temperature)

for every one unit of x increased, y will inc. by slope units

31. Which of the following is the proper interpretation of the slope?

a) For every one dollar increase in Sales, Temperature will increase on average by 18.53 degrees.
b) For every one degree increase in Temperature, Sales will increase on average by 18.53 dollars.
c) When the Temperature is 0 degrees, Sales will be 18.53 dollars, on average.
d) When the Sales are 0 dollars, Temperature will be 18.53 degrees, on average.

$r=\sqrt{R^2}$ $0\le R^2\le 1$

32. What is the correlation between these two variables?

a) 0.6862944
b) -0.6862944
c) 0.221841
d) -0.221841

$R^2=0.471$
$r=\sqrt{0.471}=0.686229$
$-1\le r\le 1$

33. The range of the variable Temperature that you observed was 72°F - 100°F. You hear a weather report saying a massive heat wave is coming your way and the high in your town will be 120°F tomorrow. You decide that you would like a prediction of your sales tomorrow since you presume you will make so much money. You use your regression equation to predict your sales for tomorrow and get a predicted value of $1461. What error have you made?

a) restricted range problem
b) misinterpretation of the slope and intercept
c) misuse of cause and effect
d) extrapolation

34. Let's say that on July 4 the temperature outside was 90°F and you sell 1100 dollars worth of ice cream. Which of the following is the residual for that day?

a) $905.70
b) -$194.30
c) $194.30
d) -$905.70

$y = -762 + 18.53x$
$y = -762 + 18.53(90)$
$-762 + 1667.7$
$= 905.7 \leftarrow$ pred

residual = actual − pred
$= 1100 - 905.7$
$= 194.3$

35. Which of the following points will always lie in a Least Squares Regression line?

a) (x, x)
b) $(\bar{x}, \bar{y})$
c) (s_x, s_y)
d) $(0, 0)$

min. distance of each pt. to the line

36. Suppose you are designing an experiment with one factor and that factor has 3 levels. You have 12 people in your experiment and assign each one a treatment by pulling a piece of paper out of a hat with either an "A", "B", or "C" on it. Which part of the experimental design process have you just completed?

a) Control
b) Randomization
c) Replication
d) Matched pairs

37. You manufacture consumer electronics and want to get feedback from your customers about their perception of your company. To do this you include a small survey in every one of your products sold and ask that your customers send it back to you for their feedback. Which of the following best describes what kind of sample this is?

a) Simple Random Sample
b) Probability Sample
c) Voluntary Response Sample
d) Stratified Random Sample

NOT A RANDOM SAMPLE

38. You are a biologist and got to the jungles of Central America to gather data about the species of mammals native to that region in their natural environment. What kind of study are you conducting?

a) Experiment
b) Survey
c) Completely Randomized Design
d) Observational Study

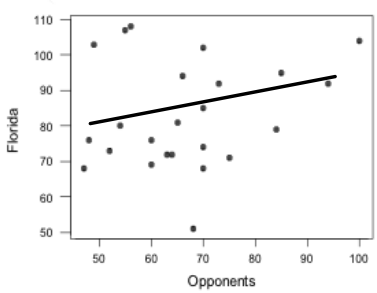
39. Suppose you are measuring the effect of two fertilizers, X and Y. You decide to design an experiment that involves two plant species, A and B. In your design you decide to make it easy on the lab technician and always give plant species A fertilizer X and plant species B fertilizer Y. Which error have you made in your design?

a) Confounding the Variables Effects
b) Undercoverage ← not enough variation
c) Restricted Rage Problem
d) Lack of Realism

40. You decide to test out a new teaching method by splitting up 10 pairs of identical twins into two groups, so that one of each pair of twin is in each group. You then apply your new teaching method to the first group and the standard teaching method to the second group. After a six week period you give both groups a test and compare the results on the test for each set of twins. Which of the following best describes the type of experiment that you have done?

a) Matched Pairs experiment
b) Observational study
c) Double Blind study
d) Simple random Sample

41. Consider the following scatterplot of results of the Gator Men's Basketball game scores as of February 19, 2002 (with Opponent's scores as x and Florida's scores as y). How can you best describe the relationship between Florida's scores and their opponent's scores?



a) There is no real relationship between x and y in this scatterplot.
b) There is a strong, positive linear relationship between x and y.
c) There is a strong, negative linear relationship between x and y.
d) There is a strong, curved relationship between x and y.

For Questions 42-44, consider the following data set.

A recent study was done to try to determine if a student's grade in a class can be used to help predict the evaluation of the teacher as given by the student. Ten students were randomly selected from a class, and each student's grade and overall teacher evaluation (both out of 100 points) were recorded. Minitab reports that the correlation is 0.755. Here is that data:

↑ closer to 1 R ↑

Student	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*mean*	*stdev*
Grade	94	85	57	78	81	91	62	55	70	74	74.7	13.66
Evaluation	91	88	85	77	79	95	66	60	71	72	78.4	11.35
	3	-4	-28	1	2	-4	-4	-5	-1	2		

42. How would we interpret the correlation of this data?

a) 75.5% of the variability in grades is explained by evaluations.
b) 75.5% of the variability in evaluations is explained by grades.
c) There is a fairly strong, positive linear relationship between evaluations and grades.
d) There is a fairly strong, negative linear relationship between evaluations and grades.

43. What is the intercept of the least squares regression line for this data?

a) –0.63
b) 16.33
c) 31.5
d) 57.0

$b = 0.755 \cdot \frac{11.35}{13.66}$

$b = 0.6273$

$b = r\frac{s_y}{s_x}$ $\quad a = \bar{y} - b\bar{x}$

$a = 78.4 - (0.6275)(74.7)$

$a = 31.54$

44. If the data point (57, 85) were removed from the study, what would happen to the least squares regression line?

a) There would be little change since this point falls in line with the others.
b) Since this point is an outlier but not influential, it would only strengthen the correlation of the line.
c) Since this point is an outlier but not influential, it would only weaken the correlation of the line
d) Since the point is an ~~influential outlier~~, it would change the direction of the slope of the line.

For Questions 45-48, consider the following situation:

Alcohol abuse researchers wanted to determine if the number of alcoholic drinks per week drunk by a successful college student had any impact on his/her studies (and in particular, on his/her GPA). Sixty graduating seniors were selected at random and asked what their GPA was and how many drinks they had, per week, throughout their college career. Here is Minitab's analysis of the least squares regression line for this data:

N=60

The regression equation is
gpa = 3.45140 - 0.0592606 drinks

S = 0.386810 R-Sq = 31.8 % R-Sq(adj) = 30.6 %

45. How would we interpret the slope of this equation?
   a) For each one additional drink per week, the student's GPA should increase by .059 points, on average.
   b) For each one additional drink per week, the student's GPA should decrease by .059 points, on average.
   c) For each 3.45 additional drinks per week, the student's GPA should decrease by .059 points, on average
   d) It is inappropriate to interpret the slope for this equation, since R^2 is so small.

x=0

46. How would we interpret the intercept of this equation?
   a) If a student did not drink, we expect his/her GPA to be 3.45.
   b) If a student did not drink, we expect his/her GPA to be .059.
   c) If a student's GPA was 0.0, we expect that he/she consumed 3.45 drinks per week.
   d) It is inappropriate to interpret the intercept for this equation, since graduating seniors cannot have a GPA of 0.0.

47. How would we interpret R^2 for this equation?

variability in "y" explained by "x"

   a) 31.8% of the variability in GPA is explained by the number of drinks per week.
   b) 31.8% of the variability in the number of drinks per week is explained by GPA.
   c) 31.8% of the variability in GPA cannot be explained.
   d) It is inappropriate to interpret R^2 for this equation, since it is so small.

48. What is the correlation between GPA and number of drinks per week?
   a) 0.318
   b) 0.564
   c) –0.564
   d) It is impossible to determine from the information given.

$R = \sqrt{R^2} = \sqrt{0.318}$ $= -0.56439$

↓ neg slope

For Questions 49-50 A study was conducted several years ago in the military to determine if shaving men's heads was effective in decreasing insubordination among new recruits. Two hundred new recruits were randomly selected, and 100 of them had their head shaved, while the other 100 were free to choose whichever hairstyle they wanted. Among the 100 shaved men, six had disciplinary problems; amongst the 100 unshaved men, there were 22 with a disciplinary problem.

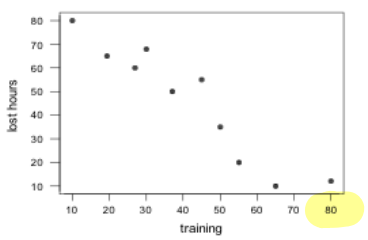
Cat22 = 28

49. What percentage of the soldiers had disciplinary problems in this study?
   a) 6/100 b) 22/100 c) 28/200 d) 100/200

50. This study is an example of a(n):
   a) experiment ← changing something b) observational study ← just watching
   c) survey ← for questions d) None of the above

For questions 51-53 A car manufacturing plant is striving to decrease the number of injuries occurring on the assembly line. They hope that training workers on safety measures and proper operation of the machines will reduce the number of work hours lost to injury. (For instance, if an injury causes a worker to leave at lunch to see a physician, the factory might lose 3-4 man-hours of labor.) Each of 10 divisions received safety training appropriate for their department. Then the number of hours lost to injuries was tabulated. The scatterplot appears below.

N=10



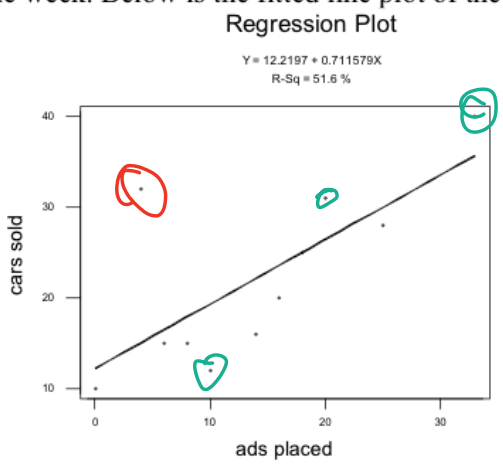
51. This relationship can best be described as…
   a) Moderate and positive X b) Weak and negative
   c) Strong and negative d) Nonlinear but strong X

52. What can you say about the relationship between safety training and productivity?
   a) Training makes little difference in how many injuries people suffer. X
   b) People who have more training seem to have more injuries on the job. X
   c) In order to decrease ~~training~~ lost time, managers should focus on safety.
   d) none of the above

53. If we give 100 hours of safety training to all divisions, lost hours due to accidents will be totally eliminated. This is an example of:
   a) misuse of cause and effect b) restricted range problem
   c) extrapolation d) potential lurking variables

data: [10, 80] 100 not in range

Questions 54-57 EasyCar Used Vehicles is analyzing the effectiveness of its advertising strategy to see if it can improve car sales. The dealer kept a record of how many television ads she placed each week, along with the number of cars sold in that same week. Below is the fitted line plot of the data:



54. Which of the following are true statements that characterize the unusual observation(s) in this scatterplot?

I. The point at (33, 40) is an influential outlier.
II. The observation at (4,32) has a large residual.
III. The point at (4, 32) is an outlier.

a) only I
b) only III
c) II and III
d) I, II, and III

55. Consider the observation at (4, 32). Which sentence below could be true about that week?

a) Since so few cars were sold, we might guess that the ads were ineffective.
b) The dealer placed many ads that week, so it makes sense that she sold many cars.
c) Although few ads were placed that week, sales were highly successful.
d) The dealer may have placed the advertisements on the wrong channel that week.

56. Which statement describes the overall relationship between these two variables?

a) There is a moderate, positive correlation.
b) The relationship is strong and negative.
c) The relationship is weak but positive.
d) There is no linear relationship between these two variables.

57. The time of day and the channel of broadcast are examples of:

a) Explanatory variables
b) Lurking variables
c) Response variables
d) Influential variables

→ Variable not controlled for

Questions 63-66 Can aspirin help prevent heart attacks? The Physician's Health Study involved a group of 22,000 male physicians in answering this question. Half of the doctors (11,000) took an aspirin tablet a day, while the rest took a placebo. After 4 years, those who received aspirin had significantly fewer heart attacks than those who took no aspirin.

test subjects

63. What are the experimental units in this scenario?
   a) Heart attacks
   b) Aspirin and Placebo
   c) Years
   d) Doctors

64. How many factors are there?
   a) one, with two levels
   b) two, with one level each
   c) 22,000 with two levels
   d) 11,000 with one level each

65. What are the treatments? factors x levels
   a) Heart attacks
   b) Aspirin and Placebo
   c) Years
   d) Doctors

66. What is the response variable?
   a) Heart attacks
   b) Aspirin and Placebo
   c) Years
   d) Doctors

Questions 67-70 Rogaine is a pharmaceutical product developed by Upjohn, which is advertised as "The only product ever proven to regrow hair." Advertisements in national magazines outline the results of clinical tests, in which 347 women with diffuse hair loss were randomly assigned to use either Rogaine or a placebo (a similar solution without the active ingredient.) To avoid biasing the results, patients were not aware of which product they were using. The product was applied by rubbing it into clean, dry scalp twice a day. After 8 months participants reported whether they had experienced moderate, minimal or no hair regrowth. The following table contains the data:

N=347

Regrowth	Moderate	Minimal	None	Total
Rogaine	33	70	72	175
Placebo	12	57	103	172
Total	45	127	175	347

67. What are the experimental units?
   a) women
   b) women with hair loss
   c) men and women with hair loss
   d) men with hair loss
   e) Americans

68. If those 347 people were a random sample, we could extend the conclusions of this study to:
   a) all women with diffuse hair loss
   b) all adults with hair loss
   c) all people who participated in the study
   d) all people who use Rogaine

69. We include a placebo group in this experiment to account for the reported hair regrowth that can be attributed to:

a) the physiological effect of rubbing your head twice a day, which may stimulate the scalp.

b) the psychological effect of participating in a study and being more aware of small changes.

c) the desire of all participants to see some hair regrowth, whether real or imagined.

d) all of the above.

70. This is an example of:

a) a blind, controlled experiment

b) a stratified, convenience sample

c) an biased, observational study

d) a multistage, randomized design

blind: test subject doesn't know treatment

double blind: test subject & researcher don't know which treatment

73. Which of the following is a true statement about probabilities?

a) A Probability can be negative.

b) Probabilities must be less than one half

c) A probability can not be greater than 1.

d) You can never have a probability of 0.

$0 \leq p \leq 1$

74. Why do we use inferential statistics?

a. to help explain the outcomes of random phenomena

b. to make informed predictions about parameters we don't know

c. to describe samples that are normal and large enough (n>30)

d. to generate samples of random data for a more reliable analysis

probability

75. Parameters and statistics…

a. Are both used to make inferences about $\bar{x}$

b. Describe the population and the sample, respectively.

c. Describe different groups of individuals.

d. Describe the same group of individuals.

statistic

Question 80-82 Suppose 20 donors come to a blood drive. Assume that the blood donors are not related in any way, so that we can consider them independent. The probability that the donor is O- blood is 0.06, which is constant from donor to donor. Let X = the number of donors that have O- blood.

$$P(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

X = # successes
two outcomes
$X \sim B(n, p)$
$n = 20$
$p = 0.06$

80. What is the distribution of X?
a.) X~Binomial(20, 0.06)
b.) X~Normal (1.2, 1.06)
c.) X~Normal (0.06, 0.0531)
d.) Can't be determined

$$P(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

81. What is the probability that two donors have type O- blood?
a.) 0.225
b.) 0.290
c.) 0.370
d.) 0.885

$\binom{n}{k} p^k (1-p)^{n-k}$

$\frac{n!}{k!(n-k)!}$

$\frac{20!}{2!(20-2)!} (0.06)^2 (1-0.06)^{18}$

$190(0.00036)(0.94)^{18}$

82. What is the probability that 1 or more donors have type O- blood?
a.) 0.290
b.) 0.370
c.) 0.630
d.) 0.710

1, 2, 3, 4, ... 15, 16, 17, 18

$1 - P(X=0)$

$\binom{20}{0} (0.06)^0 (0.94)^{20}$

$\frac{20!}{0!(20-0)!} (1)(0.29)$

$1 - 0.29$
$= 0.71$

Questions 83-86 For a sales promotion the manufacturer places symbols under the caps of 10% of all Pepsi bottles. You buy 6 random bottles of Pepsi. Assume that the bottles are independent.

83. What is the probability that all of the bottles have the winning symbols?
a.) 0.10^6
b.) 0.90^6
c.) $1-.90^6$
d.) $1-.10^6$

P(one) = .10

$P(b_1) * P(b_2) * P(b_3) * P(b_4) * P(b_5) * P(b_6)$

0.1^6

84. What is the probability that none of the bottles have a winning symbol?
a.) 0.10^6
b.) 0.90^6
c.) $1-.90^6$
d.) $1-.10^6$

P(no) = 1 − 0.10 = 0.9

85. What is the probability that at least one of the bottles has a winning symbol?
a.) 0.10^6
b.) 0.90^6
c.) $1-.90^6$
d.) $1-.10^6$

P(1), P(2), P(3), P(4), P(5), P(6)

$1 - P(0)$

$1 - 0.90^6$

86. What is the probability that at least one of the bottles does not have a winning symbol?
a.) 0.10^6
b.) 0.90^6
c.) $1-.90^6$
d.) $1-.10^6$

$P(\text{all}) = 0.1^6$

Questions 87 -90 It is though that the number of damaged cans in a boxcar shipment (Y) is a function of the speed of the boxcar (X) upon impact. The speed of the boxcar goes from 0 to 16 ft/s. The summary statistics of thirteen (13) boxcars appears below.
$\bar{x} = 4.615$, $\bar{y} = 71.8$, $S_x = 1.895$, $S_y = 42.9$, $r = 0.522$

87. For now let's say the slope is an arbitrary *positive* number *k*. How do we interpret the slope for this regression equation? for every 1 inc. in x (speed), y inc. by k

a) The slope is the number we get from plugging the numbers into the equation.
b) For every one unit increase in the speed of the boxcars, the number of damaged cans decreases by k.
c) For every one unit increase in the speed of the boxcars, the number of damaged cans increases by k.
d) There is not enough information to answer this question.

88. Compute the *actual* slope.
a) 42.9
b) 22.6
c) 11.8
d) 15.6
e) -11.8

$b = r\frac{s_y}{s_x}$ $\quad a = \bar{y} - b\bar{x}$

$b = 0.522\left(\frac{42.9}{1.895}\right) = 11.8$

89. The y-intercept for this equation is 17.343, can this be interpreted given the context of the problem? x=0, speed=0

a) Yes, it turns out that when we have a speed of 0 from the boxcar, we have about 17 damaged cans.
b) Yes, it turns out that when we have a speed of 0 from the boxcar, we have about 72 damaged cans.
c) No, there is no such thing as a speed of 0, we cannot interpret the y-intercept.
d) There is not enough information to answer this question.

Questions 91 – 93 A popular fashion magazine has just recently published a poll saying that 45% of women who, if given the chance, would want to marry their first boyfriend. A random sample of 20 women from across the United States was taken. Answer the following questions.

91. This is clearly a binomial experiment. Which of the following is not a property of the Binomial? p=0.45 n=20

a) The observation for each woman is independent, since the women were randomly sampled. ✓
b) Each woman has the same probability they would like to marry their first boyfriend (45%). ✓
c) Each woman in the survey has based their response upon another woman's response. ✗
d) There is not enough information to answer this question.

92. What is the probability that half of the women sampled would like to marry their first boyfriend?

a) 0.4500
b) 0.1593
c) 0.5513
d) 0.6312
e) Not enough information to answer this question.

$n=20$ $x=10$ $p=0.45$ $1-p=0.55$

$\frac{20!}{10!\,10!}(0.45)^{10}(0.55)^{10} = 0.1593$

$P(x)=\binom{n}{x}p^x(1-p)^{n-x}$

93. Let's say we increased our sample to 100 women, what is the expected number of women in this new sample that would say that they would like to marry their first boyfriend?

a) 45
b) 55
c) 65
d) 75
e) 35

yes $E(x)=np = 100(0.45) = 45$ would marry

no $n(1-p) = 100(0.55) = 55$ would not marry

Questions 94-95 When parking a car in a downtown parking lot, drivers pay according to the number of hours. The probability distribution for the number of hours a car is parked has been given below:

x	1	2	3	4	5	6	7	8
P(x)	0.24	0.18	0.13	0.10	0.07	0.04	0.04	0.20

94. What is the probability that a car will be parked in this parking lot for at least 6 hours?

a) 0.04
b) 0.28
c) 0.24
d) 0.10
e) 0.18

$0.04+0.04+0.2 = 0.28$

95. What is the expected number of hours a given car will be parked in the parking lot?

a) 3.86
b) 2.84
c) 5.63
d) 10.23
e) 4.52

$1(0.24)+2(0.18)+3(0.13)+4(0.1)+5(0.07)+6(0.04)+7(0.04)+8(0.2)$
$= 3.86$

96.The data set below gives the length of a bear in inches and the weight of the bear in pounds. Without calculating the value of r, what is the linear correlation coefficient between length and weight?

Length	Weight
45	65
47.5	70
57	74

$-1 < r < 1$

a) r=.92 pounds ✗
b) r=.92 inches ✗
c) r= 9.2 ✗
d) None of the above

r has no units

97.If we divide the length of the bears by 12 so that the length is now in feet, will the linear correlation coefficient between length and weight change?

a) yes, it will change
b) no, it will not change
c) not enough information
d) maybe it will change

r has no units

Change scale of graph, but not r

98.Which of these statements is false?

a) A parameter, in practice, is an unknown number describing the population. ✓
b) A statistic is used to estimate an unknown parameter. ✓
c) A parameter is used to estimate an unknown statistic. ✗
d) Statistics can change from sample to sample. ✓

99.For X~N(1, 12), find the probability P(X=5).

a) 0
b) .1
c) .01
d) .98

continuous

$P(X=\#) = 0$

101. A player flips two fair coins and counts the number of heads. We must find the probability distribution for X={number of heads}. What should the value of "Q" be to complete the probability distribution?

X	P(x)
0	.25
1	Q
2	.25

$P(0) + P(1) + P(2) = 1$
$0.25 + P(1) + 0.25 = 1$
$0.5 + P(1) = 1$
$P(1) = 0.5$

a) .5 b) .25 c) .05 d) 1

102. Which of the following is/are true about a skewed right distribution with extreme outliers?

I) The mean is greater than the median

II) The median should be used as the measure of center because it is more resistant to extreme observations than the mean

III) The standard deviation should be used as the measure of spread because it is more resistant to extreme observations than the range or inter-quartile range

IQR is more resistant bc doesn't include end observation

a) I and II only
b) I and III only
c) II only
d) I, II, and III

103. Suppose a survey of US adults reports a correlation of r=0.92 between the amount of junk food a person consumes and their weight. Which of the following can we say about the results?

I) That the junk food causes weight gain

Correlation ≠ causation

II) That the results are from an experiment

III) That there is a strong positive linear relationship between the amount of junk food a person consumes and their weight

a) III only
b) I and III only
c) I and II only
d) I, II, and III

104. A study investigated the relationship between the number of hours of sleep that a person got the night before and their response time(in seconds) on a manual dexterity test. The correlation coefficient was equal to 0.7183 and the least squares regression equation was response time=12.22+0.71hours. What of the statements below is a true statement?

$r = 0.7183 \rightarrow r^2 = 0.7183^2$

$= 0.51596$

51.6%

a) That 71.1579% of the variability in the number of hours slept is explained by response time
b) That 71.1579% of the variability in response time is explained by the number of hours slept
c) That 51.6% of the variability in the number of hours slept is explained by response time
d) That 51.6% of the variability in response time is explained by the number of hours slept

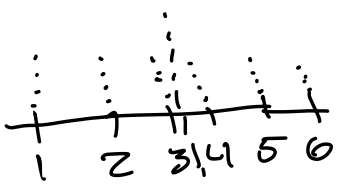
variation in y explained by x

105.Which of the following is/are true about a normal distribution?

a) If the same area is shaded in the lower tail and in the upper tail, the probabilities of the lower and upper tail are equal.
b) The mean is equal to the median
c) It is a continuous distribution
d) All of the above are true

106.Suppose that you had a distribution of rolling a fair 6 sided dice. What can we say about the distribution?

a) It is uniform.
b) Each of the observations are equally likely.
c) It has a mean equal to 3.5.
d) All of the above.



109.In studying the incidence of surf wax toe-burn among Florida's east coast skin boarders, over several years Shelley the Shands medical student has interviewed a sample of 100 people that skin boards on the east coast of Florida. Shelley wants to publish her astonishing results - that the mean pain level (on a scale of 1 to 10) of toe-burn among Florida's skin boarders is 7.5. In the discussion portion of Shelley's research paper, include the word _____ to describe the mean value she discovered.
Fill in the blank.
a) "statistic"
b) "parameter"
c) Neither (a) nor (b) have anything to do with Shelley's medical research.
d) Shelley can use either (a) and (b) because these words are interchanged freely.

110.Phlebotomists(those who draw blood for analysis) at the NW 12 Street blood plasma donation center have found that 15% of their donors eat more than three cookies following their plasma donation. On a given day, 13 donors give plasma. What is the probability that less than two of the day's donors will eat more than three cookies after they donate blood plasma?
a) .398
b) .692
c) .121
d) .277

$P(x) = \binom{n}{x} p^x (1-p)^{n-x}$

$P = 15\%$
$1 - P = 85$

$P(0) + P(1)$

$\frac{13!}{0!(13-0)!}(0.15)^0(0.85)^{13} + \frac{13!}{1!(13-1)!}(0.15)^1(0.85)^{12}$

$= 0.121 + 0.277$

= 0.5416

111.The University of Florida's zip code is 32612, is this a categorical or quantitative variable?

a) Categorical
b) Quantitative
c) Neither

112. Gubernatorial candidates Charlie Crist and Jim Davis each roll a die. Let the random variable X be the sum of the numbers on the two dice. Below is the probability distribution for all possible sums of the rolls of the dice. What is the missing probability?

X=2	X=3	X=4	X=5	X=6	X=7	X=8	X=9	X=10	X=11	X=12
1/36	1/18	1/12	1/9		1/6	5/36	1/9	1/12	1/18	1/36

a) 1/36
b) 4/36
c) 5/36
d) 7/36

$\frac{1}{36} + \frac{1}{18} + \frac{1}{12} + \frac{1}{9} + P(X=6) + \frac{1}{6} + \frac{5}{36} + \frac{1}{9} + \frac{1}{12} + \frac{1}{18} + \frac{1}{36} = 1$

$\frac{2}{12} + \frac{7}{36} + \frac{2}{18} + \frac{2}{9} + \frac{1}{6} + P(X=6) = 1$

$\frac{6}{36} + \frac{7}{36} + \frac{4}{36} + \frac{8}{36} + \frac{6}{36} + P(X=6) = \frac{36}{36}$ $\quad \frac{31}{36} + P(X=6) = \frac{36}{36}$

$P(X=6) = \frac{5}{36}$

113.A bored student during class flips a fair coin 5 times. Assuming the flips are independent. What is the probability he gets exactly 3 heads and 2 tails?

a) ½
b) ¾
c) 5/16
d) 5/64

independent trials & constant probability(p)

$p = 0.5$
$n = 5$
$k = 3$

$\binom{5}{3}(0.5)^3(1-0.5)^2$

$\frac{5!}{3!2!} 0.5^3 0.5^2$

10×0.5^5

$10 \times \left(\frac{1}{2}\right)^5$

$10 \left(\frac{1}{64}\right) = \frac{10}{32} = \frac{5}{16}$