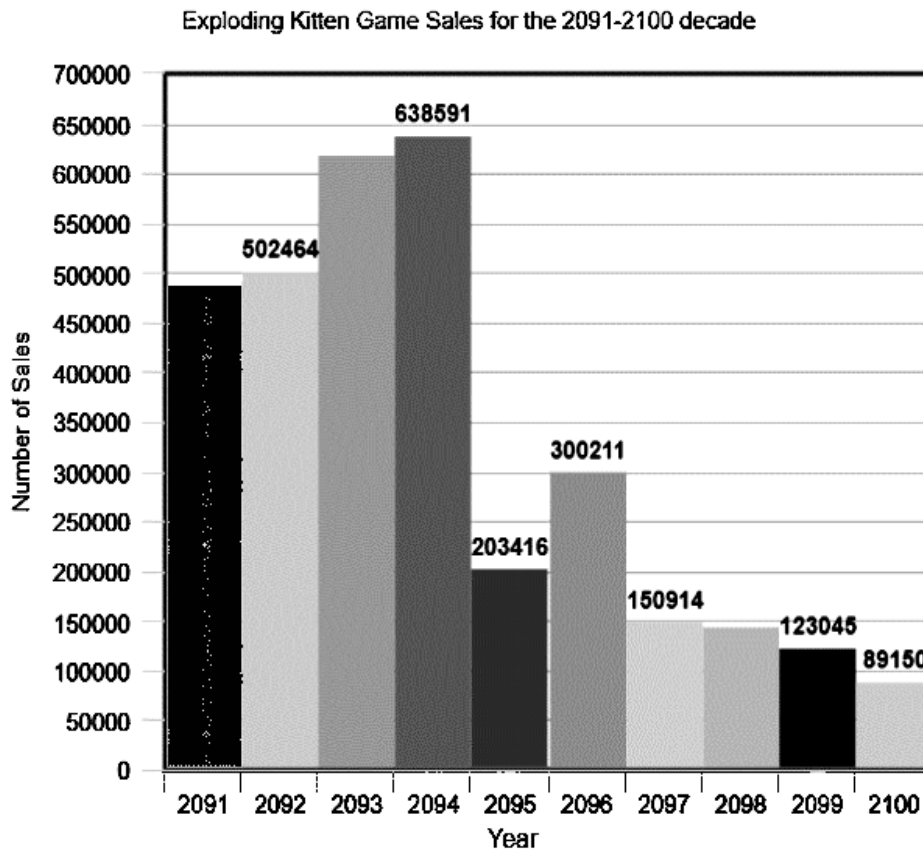


**Important Instructions for this Test:** Please pay close attention to and carefully follow all rounding instructions. Round any intermediate steps as indicated or as necessary to make the final answer as accurate as possible. Good luck, have fun, and as always: “NOTA” stands for “None of These Answers is correct.”

Use the following information for questions 1 – 4:

Aniché leaked the number of sales for the popular card game “Exploding Kittens” publicly, which is shown in the graph below. Unfortunately, he was not able to obtain 3 years’ worth of information. Your job is to fill in those blanks! Don’t worry, I will give you hints as to what these values are. (Disclaimer: I made up all of this data, it’s not real 😊)



1. Between the years 2091 through 2100, inclusive, what is the median number of sales of the Exploding Kittens game?

- A) 300,211    B) 251,813.5    C) 249,930.5    D) 394,624    E) NOTA

2. What is the third quartile,  $Q_3$ , of the number of sales of the Exploding Kittens game between 2091 and 2100, inclusive, rounded to the nearest integer?

- A) 502,464    B) 496,232    C) 445,202    D) 520,392    E) NOTA

3. Sashta couldn’t figure out a reasonable estimate for the first quartile,  $Q_1$ , of the number of sales of the Exploding Kittens game between 2091 and 2100, inclusive. Which of the following is the most reasonable estimate for  $Q_1$ ? Only choose a value from the choices A, B, C, or D. There is no reason to choose E unless you believe this question is completely flawed or does not have an answer among choices A through D.

- A) 145,000    B) 150,000    C) 115,000    D) 160,000    E) NOTA

4. Grandpa Tomer wrote to his grandson Francisco some things about the Exploding Kitten game sales from the years 2091-2100 inclusive, but Francisco is quite skeptical of his grandfather’s assertions. Can you tell Francisco which of the following statements below that his grandfather wrote are true?

Dear grandson,

- 1) The shape of the distribution of the number of sales of the Exploding Kittens game from the years 2091 to 2100 is strongly skewed to the left.
- 2) The range in the number of sales of the Exploding Kitten game from the 2091-2100 decade, inclusive, is 549,441.
- 3) The graph displayed on the previous page depicting the number of sales of the Exploding Kitten game between the years 2091 to 2100 (inclusive) is a histogram.

Love, Tomer

- A) 1 and 2      B) 1 and 3      C) 2 and 3      D) All three      E) NOTA

Use the following information for questions 5 – 10:

*Zaxbee’s* is a popular franchise restaurant. In fact, many wealthy individuals eat there. Federico, Vismay, Venkata, and Ritesh each randomly select a different restaurant location with an astronaut pen and analyze their assigned location’s food quality according to a customer satisfaction survey of the food eaten. They wanted to determine whether the average food quality rating decreases throughout the day by randomly surveying customers at each restaurant location each hour the restaurant is open. The restaurant is open every day from 8:00 a.m. to 9:00 p.m., and the value of “X = 0” in the table below corresponds to the hour from 8:00 a.m. to 9:00 a.m. etc. The response variable is the average food quality rating of all the randomly selected customers surveyed at each of the randomly selected *Zaxbee’s* locations during the indicated hour. The data is shown in the table below. You may assume all least squares regression analysis conditions are satisfied.

<i>X = # of Hours after 8 AM</i>	0	1	2	3	4	5	6	7	8	9	10	11	12
<i>Y = Average Food Quality Rating</i>	8.9	8.7	8.8	8.7	8.6	8.7	8.3	8.4	8.5	8	8.6	7.3	6.9

5. Conduct a least squares linear regression analysis on the data above. Using the LSRL equation, what is the sum of the squared deviations between the predicted *Average Food Quality Rating* for each of the given number of hours past 8 AM and the overall sample mean of the *Average Food Quality Ratings* of the original data? This is also known as the “sum of the squares from the regression” or the “SSR.”  $SSR = \sum(\hat{y} - \bar{y})^2$ . Round your final answer to the nearest hundredth.

- A) 1.57      B) 2.78      C) 4.35      D) 2.76      E) NOTA

6. If it is appropriate to do so, determine the predicted *Average Food Quality Rating* during the hour from 11:00 p.m. to 12:00 a.m. using the LSRL equation based upon the data in the table above rounded to the nearest hundredth.

- A) 7.48      B) 7.35      C) 7.23      D) 7.72      E) NOTA

7. What is the standard deviation of the set of differences of the abscissa and the ordinate of the ordered pairs (*# Hours after 8 AM, Average Food Quality Rating*)? That is, what is  $S_{x-y}$  for the data in the table above? Round your final answer to four decimal places.

- A) 6.7007      B) 3.4345      C) 10.8393      D) 4.3889      E) NOTA

8. Let P = the number of outlier data points in the scatterplot of this bivariate regression data set with *X = The Number of Hours Past 8 AM* and its corresponding *Y = Average Food Quality Rating*. Let R = the sum of the tenths place digit and the hundredths place digit of the highest positive residual value from the LSRL analysis of this same bivariate data set. For the purposes of this question, consider a point an outlier in the scatterplot of the regression of Y on X as any point with an unusually large residual according to the 1.5(IQR) rule for outliers when applied to the residuals. What is P + R?

- A) 12      B) 13      C) 15      D) 18      E) NOTA

Use the Following Stemplot and Boxplot for Question 9:

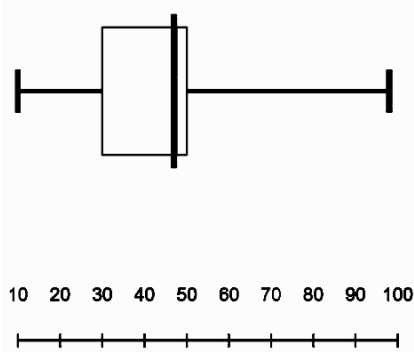
*Ritesh's Data*

Key:

$1|0$  is 1.0

1	0
2	89
3	024
4	3789
5	006
6	0
7	
8	
9	8

*Vismay's Data*



9. Observing the stemplot and the boxplot displaying the distributions for the two arbitrary data sets above and the bivariate data set in the table on the previous page, which of the following univariate data sets definitely has a median that is greater than its respective mean?

- I. Ritesh's Data set.
  - II. Vismay's Data set.
  - III. The univariate data set of the *Average Food Quality Ratings* from the bivariate data set on the previous page.
- A) I and III    B) II and III    C) I Only    D) I, II, and III    E) NOTA

10. Federico decided to modify his bivariate data set of the number of hours past 8 a.m. and the resulting *Average Food Quality Rating* at *Zaxbee's* for the sample of data on the previous page by altering the number of hours past 8 a.m. and the *Average Food Quality Ratings* such that the linear correlation coefficient strengthens. (Yes, he's trying to fudge his data!). Which of the following transformations will definitely strengthen the correlation when applied to the data?

- A) Multiply each of the number of hours past 8 a.m. by a positive real number constant and add the same positive real number constant to each resulting value.
- B) Multiply each *Average Food Quality Rating* by the same positive real number constant and add the same positive real number constant to each resulting value.
- C) Divide each *Average Food Quality Rating* by the same positive real number constant and subtract the same positive real number constant from each resulting value.
- D) All of the above.
- E) NOTA

Use the following information for questions 11 – 16:

Reshoe tried dog food one day and savored it for a moment... "This is delicious!" he exclaimed and so he decided to conduct the following study to answer the question: "Does eating dog food increase one's stamina?" Reshoe decides to conduct a randomized, comparative, and placebo-controlled double-blind experiment on 600 randomly selected United Kingdom residents where each participant will replace their usual lunchtime meal with either dog food or a placebo food for one week. The placebo food is visually identical to and tastes the same as the dog food but it is human food and has completely different ingredients that do not affect the consumer's stamina in any way. He begins by arbitrarily assigning each one of the 600 participants three-digit ID number between 001 and 600, inclusive. Then, he uses a random number table to determine which participants will receive the dog food and which will receive the placebo food. Here is the first line of the random number table that he used:

36518 36777 89116 05542 29705 83775 21564 81639 27973

11. Reshoe needs help determining who the first five random participants assigned to the dog food group are. Given that he will select numbers less than or equal to the sample size of 600, what are the first five participant numbers selected?

- A) 365, 183, 116, 55, 422      C) 365, 518, 183, 367, 116      E) NOTA  
B) 365, 183, 677, 789, 116      D) 36, 51, 83, 36, 77

12. Reshoe believes that the gender of the participant has an influence on the response variable in his experiment, namely one's stamina. So, if he were to begin his experiment with a group of 300 male participants and a group of 300 female participants and then randomly assign the 300 participants within each gender group to either the dog food treatment or the placebo; then what type of experimental design is Reshoe using?

- A) Stratified Design      C) Homogeneous Design      E) NOTA  
B) Randomized Block Design      D) Matched Pairs Design

13. Reshoe measures each participant's initial and final stamina level by recording the amount of time each subject can last running on a treadmill at 10 MPH both at the beginning and at the end of the experiment. He then compares the change in stamina between the two treatment groups within each gender block. Throughout the experiment, Reshoe allows the participants to otherwise eat normally but with the added meal (the dog food or the placebo food) replacing their usual lunchtime meal for one week. Based on all of the information given so far regarding this experiment, which one of the following four principles of experimental design were not properly employed in this experiment?

- A) Comparison      B) Randomization      C) Control      D) Sufficient Replication      E) NOTA

14. Reshoe casually does this experiment; however, like any experiment, there are a multitude of potential confounding variables. Below are some potential confounding variables that could affect the results of this experiment. Which of the following are truly confounding variables in this experiment?

- I. Whether or not the participant is training for a marathon.  
II. The individual participant's regular daily diet.  
III. If the participant decides to give up running on the treadmill out of boredom or needs to get to another appointment of some kind.

- A) I and II      B) I and III      C) II and III      D) All three      E) NOTA

15. To potentially further account for more sources of variability in the experiment, Reshoe considers implementing multiple treatments. So, he varies the servings sizes to 100 grams, 200 grams, and 300 grams per serving per day for a week and uses the following four dog food brands: *Blue*, *Eukanuba*, *Freshpet*, and *Royal Canin*. The placebo food is still in place with the same varying serving sizes listed above as well. In this new design, how many treatments does the experiment now have? You may assume that all these dog foods brands and the placebo food all look and taste the same.

- A) 7      B) 8      C) 12      D) 15      E) NOTA

16. Before conducting the experiment, Reshoe wanted to get an idea about what the general population of United Kingdom residents thought about the idea of eating dog food. So, he conducted a twitter poll of his followers and sent out the following question: "Do you think it's okay to eat dog food? Yes, or no?" When reviewing the results that every user of twitter can see, all of Reshoe's followers replied and 98% of them replied with "yes." This scenario primarily suffers from which one of the following types of sampling bias? Pick only the best answer from the choices listed in A through D.

- A) Framing Effect      B) Undercoverage      C) Nonresponse Bias      D) Response Bias      E) NOTA

17. Ryan, Mehul, Salteri, and Lokesh recently joined Captain Vismay's pirate crew. While they are indoors due to heavy thunderstorms, the four of them sneak away from Captain Vismay to play a card game known as "Novock." To begin the game, all four of them flip a fair, two-sided coin exactly once in hopes of obtaining heads on their flip; however, only Salteri (since she outranks the others) gets a second attempt to obtain heads on a flip of the coin if and only if she does not get heads on her first attempt. What is the probability that exactly three out of the four of them obtain a head when they flip the coin given that only Salteri gets a second attempt if she does not get heads on her first attempt?

- A)  $\frac{9}{32}$       B)  $\frac{11}{32}$       C)  $\frac{7}{16}$       D)  $\frac{5}{16}$       E) NOTA

18. Suppose you're trying to estimate the average weight of a SqUiD based on a random sample of 75 observations. It is known that the standard deviation of the distribution of a SqUiD's weight is 5 kg. What is the probability that such a sample mean estimate is within 0.5 kg of the true population mean? Round only your final answer to four decimal places.

- A) 0.6135      B) 0.3829      C) 0.4367      D) 0.5648      E) NOTA

19. The sample mean of an SRS of 10 observations from random variable  $X$  is  $\bar{x} = 0.9$ . The sum of the squares of the elements in the sample is  $\sum x^2 = 183$ . What is the sample standard deviation of this sample of random observations from random variable  $X$ , rounded to three decimal places?

- A) 3.987      B) 15.900      C) 4.408      D) 19.433      E) NOTA

20. Anirude created a valid binomial probability distribution with the constant probability of success being  $p = 0.892$  and with  $n = 10$  trials. Which of the following best describes the shape of Anirude's binomial distribution?

- A) Skewed to the left.      C) Skewed to the right.      E) NOTA  
B) Perfectly symmetric.      D) Not enough information is provided.

**Use the following information for questions 21 – 26:**

A pharmaceutical company, HeRRan, is carrying out a preliminary experimental trial for a potential VIS-HAL-2 vaccine. One statistical analysis they will carry out is measuring the overall efficacy rate of the vaccine in terms of the antibody response rate of the volunteer participants in the study. They plan to split the volunteer participants into four age groups: under 18 years old, 18–54 years old, 55–65 years old, and over 65 years old. They are testing the vaccine at two strengths (low and high) and as either a single dose or as two doses.

21. How many blocks are there in this experiment?

- A) 6      B) 8      C) 4      D) 16      E) NOTA

22. The HeRRan company wants to conduct a hypothesis test to see if there is a significant difference between the overall efficacy rates of the following two age groups: under 18 years old and 55–65 years old. After data collection within each age group consisting of 100 individuals, the company notices that the overall efficacy rate for the under 18 years old age group is 86% and the overall efficacy rate for the 55–65 years old age group is 81%. What is the standard error of the sampling distribution that is used in the calculation of the test statistic in this test, rounded to four decimal places? You may assume all inference assumptions and conditions are met.

- A) 0.0439      B) 3.1902      C) 0.0226      D) 0.0525      E) NOTA

23. Determine the p-value for the appropriate test in the previous question and round it to the hundredths place. What is the sum of the tenths place digit and the hundredths place digit of this rounded p-value?

- A) 18      B) 5      C) 7      D) 0      E) NOTA

24. In reality, it turns out that there is actually no difference between the efficacy rates of the vaccine for the 55-65 years old population and the under 18 years old population. Based on the conclusion made in the test conducted in question 23 and using a significance level of  $\alpha = 0.01$ ; what type of error, if any, was committed?

- A) A Type I Error      C) A Type III Error      E) NOTA  
B) A Type II Error      D) A Confounding Error

25. Which of the following items listed below are ways to increase the power of the test conducted in question 23?

- I. Increase the treatment group size of the 55-65 years old age group.  
II. Increase the treatment group size of the under 18 years old age group.  
III. Increase the significance level of the test.

- A) I and III      B) I and II      C) II and III      D) All three      E) NOTA

26. The HeRRan company decided to construct a 90% confidence interval for the difference in efficacy rates for the two age groups 18-54 years old and over 65 years old, each consisting of 150 people. Given that the 18-54 years old age group has a sample efficacy rate of 86% and the over 65 years old age group has a sample efficacy rate of 74%, what is the resulting confidence interval with positive limits rounded to the ten-thousandths place (four decimal places)? You may assume all inference assumptions and conditions are met.

- A) (0.0449, 0.1951)      B) (0.1138, 0.1261)      C) (0.0305, 0.2095)      D) (0.0764, 0.1636)      E) NOTA

**Use the following information for questions 27 – 30.**

Hermosillo likes to eat pie. Hermosillo says he eats 0.125 slices of pie per day, on average; however, Giuliano thinks Hermosillo is bluffing and decides to conduct an observational study to test Hermosillo's claim. Giuliano takes an SRS of 45 days throughout the year and found that in these 45 days, Hermosillo, on average, ate 0.8 slices of pie per day with a standard deviation of 1.838 slices per day. You may assume all inference assumptions and conditions are met.

27. A  $C\%$  confidence interval for the true mean number of slices of pie Hermosillo can eat in a day is constructed and the resulting margin of error is 0.306 slices of pie per day. The percentage used in this confidence interval can be expressed as  $C = 100\% * (1 - \alpha)$ , where  $\alpha$  is equivalent to the significance level used in a two-tailed hypothesis test. Given that the margin of error is 0.306 slices of pie per day and that you should round all steps to obtain  $\alpha$  to the thousandths place, what is the sum of the tenths place digit and the hundredths place digit of  $\alpha$  when it is expressed as a decimal rounded to the thousandths place?

- A) 10      B) 9      C) 8      D) 4      E) NOTA

28. After conducting the test, Giuliano decided to monitor how many pies his roommate, Di Gianluca, ate and created the following valid probability distribution for how likely it is that Di Gianluca would eat at least one slice of pie on a given day of the week. Note that  $a = P(\text{Monday})$  and  $b = P(\text{Wednesday})$  are missing.

X	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
P(X)	a	$\frac{1}{11}$	b	$\frac{3}{22}$	$\frac{1}{22}$	$\frac{3}{44}$	$\frac{23}{44}$

What is the probability that Di Gianluca will eat at least one slice of pie on a weekday (Monday through Friday)?

- A)  $\frac{19}{22}$       B)  $\frac{3}{22}$       C)  $\frac{13}{22}$       D)  $\frac{9}{22}$       E) NOTA

29. In the table in question 28, what is the sum of the thousandths place digit and the ten thousandths place digit of  $a$  if it is defined as:  $\frac{1}{100}$  times the probability of obtaining a value between -1 and 2 standard deviations from the mean of a standard normal distribution using the Empirical Rule [i.e.,  $P(-1 < Z < 2)$ ]?

- A) 5      B) 9      C) 10      D) 7      E) NOTA

30. Suppose we let  $X$  represent the enumerator for the day of the week. Thus, Monday is day 0, Tuesday is day 1, Wednesday is day 2, ... and Sunday is day 6. Given this information, what is the expected value of the probability distribution of these enumerated values for  $X$  using the probability distribution from question 28 and your result for the missing probability,  $a$ , from question 29? Round the final answer to the nearest tenth.

- A) 4.2      B) 4.3      C) 4.4      D) 4.5      E) NOTA