The test will have five sections:

- (10 × 1 marks = 10 marks) Matching
- (10 × 1 marks = 10 marks) True or False
- (10 × 2 marks = 20 marks) Multiple Choice
- (6 × 5 marks = 30 marks) Short Answer
- (3 × 10 marks = 30 marks) Long Answer

You will be provided the following formulas:

$$\text{margin of error} = 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$\text{mean} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Section 1. (Matching Match term or quantity in left column to the one description that best applies from the numbered columns.

variables 11

confounded variables 3

bias 4

response variable 14

negative association 5

positive association 12

prospective study 6

sample 7

simple random sample 9

disjoint events 10

1. there is no linear relationship between two variables.
2. neither the experimental subjects nor the people interacting with them know which treatment each subject received.
3. the effects of the variables on the outcome of a study cannot be distinguished from each other.
4. a systematic error that tends to cause the observation to deviate in the same direction from the truth about the population whenever a sample or experiment is repeated.
5. an increase is one variable tends to accompany a decrease in the other variable.
6. an observational study that follows two or more groups of subjects forward in time.
7. a part of the population that is actually observed and used to draw conclusions about the entire population.
8. a sample of people who select themselves by responding to a general invitation to give their opinions.
9. a sample chosen by chance, so that every possible sample of the same size has an equal chance of being selected.
10. events that have no outcomes in common.
11. the objects described by a data set.
12. an increase is one variable tends to accompany a increase in the other variable.
13. a variable that attempts to justify the observed outcomes.
14. a variable that measures an outcome of a study.
15. a line that describes how a response variable y changes as we change an explanatory variable x.

Section 2. True or False Circle True (T) or False (F):

(1) It is impossible to tell the skewness of a distribution from the distributions stemplot ..... T (F)

(2) The five number summary of a distribution is made up of the mean, the standard deviation, the first quartile, the third quartile, and the range of the distribution. ..... T (F)

(3) For a symmetric distribution, the mean and median will be the same. ..... (T) F

(4) Distributions with a large number of observations are very well represented by stemplots. ..... T (F)

(5) Stemplots are incapable of displaying any skewness in a distribution. ..... T (F)

(6) The mean and standard deviation can be used to describe the skewness in a distribution. ..... T (F)

(7) The mean of a distribution is greatly affected by outliers in the distribution. ..... (T) F

(8) For a scatterplot, a correlation near 1 means the data points are not very well described by a straight line with positive slope ..... T (F)

(9) Voluntary response samples rarely exhibit bias ..... T (F)

(10) A randomized comparative experiment is a special type of an observational study ..... T (F)

(11) Nonresponse is a typical difficulty in surveys that a simple random sample is designed to overcome T (F)

(12) The probability of rolling seven ones in a row for a fair die is $\frac{7}{6}$. ..... T (F)

(13) The following is a valid probability model

Outcome	0	1	2	3	4	5
Probability	0.10	0.15	0.35	0.35	0.05	0.10

..... T (F)

Section 3. Multiple Choice Circle the most appropriate answer:

(1) Given the set of data below, find the median.

9, 2, 6, 4, 5, 4

(A) 4.5 B) 5 C) 4 D) 30

2, 4, 4, 5, 6, 9

$\frac{4+5}{2} = 4.5$

(2) The regression equation $y = 50 + 0.1x$ gives the yield in bushels per acre of corn when x pounds of fertilizer are applied. Predict the yield for a farmer who plans to use 320 pounds of fertilizer.

(A) 82 B) 16, 032 C) 37 D) 50.32

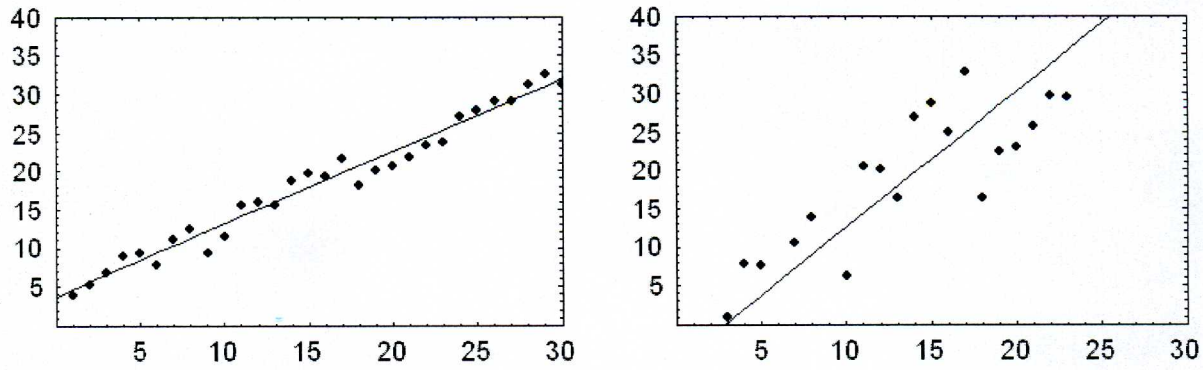
$y = 50 + 0.1(320)$
$= 50 + 32$
$= 82$

(3) Given the stemplot below, which description is true?

0	1
1	9
2	267
3	12
4	0259
5	258
6	05
7	
8	
9	6

A) There are no outliers in the distribution.
B) The number 6 is the only outlier in the distribution.
C) The number 96 is the only outlier in the distribution.
D) The numbers 1 and 6 are both outliers in the distribution.

(4) The correlation for the following two scatterplots

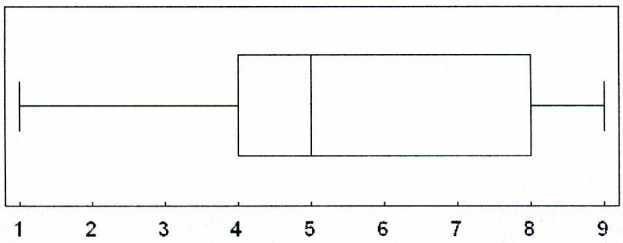


A) will have the same sign
B) will be larger for the plot on the left than the plot on the right
C) will have opposite signs
D) Both A) and B)

(5) A study gathered data on 1000 randomly selected students and showed that students who took Latin in high school had much higher scores on a test of verbal English skills than those who did not take Latin. The study cannot conclude that taking Latin improves verbal English skills because

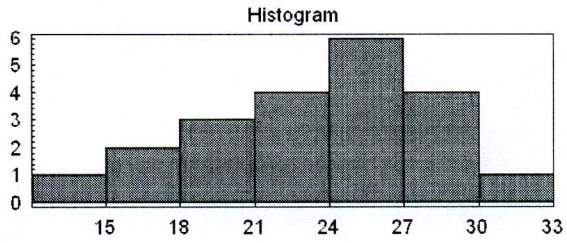
A) the study was not double blind
B) of the placebo effect
C) the study was not an experiment
D) the verbal English test was faulty

(6) The following boxplot represents a distribution which



A) has a mean of 6.
B) has a median of 5.
C) has 75% of its measurements between 4 and 8.
D) both B and C.

(7) The following histogram represents a distribution which



A) is skewed to the right.
B) is skewed to the left.
C) is symmetric.
D) both A and C.

(8) Suppose a game has four outcomes, A, B, C, and D with probabilities $P(A) = 0.2$, $P(B) = 0.1$, $P(C) = 0.4$ and $P(D) = 0.3$. A player will receive \$1 when outcome A occurs, \$4 when outcome B occurs, and have to pay \$2 if outcomes C or D occur. What is the mean value of one trial of this game?

A) \$1 B) \$0.80 C) −\$0.80 D) \$2 $0.2(\$1) + 0.1(\$4) - 0.4(\$2) - 0.3(\$2)$

(9) The mean length of time, per week, that students at a school spend on their homework is 24.3 hours, with a standard deviation of 1.4 hours. Assuming the distribution of study times is normal, what percent of students study between 22.9 and 25.7 hours?

$24.3 + 1.4 = 25.7$ $24.3 - 1.4 = 22.9$ use 68-95-99.7 rule.

A) 99.7% B) 68% C) 95% D) 50%

(10) The scores of students on a standardized test are normally distributed with a mean of 300 and a standard deviation of 40. Between what two values do 68% of the test scores lie?

68% -> 1 standard deviation in 68-95-99.7 rule

A) 260–340 B) 220–380 C) 297–303 D) 180–420

(11) The scores of students on a standardized test are normally distributed with a mean of 300 and a standard deviation of 40. Between what two values do 95% of the test scores lie?

A) 260–340 B) 220–380 C) 297–303 D) 180–420

(12) Strictly speaking, what is the best interpretation of a 95% confidence interval for the mean?

A) If repeated samples were taken and the 95% confidence interval was computed for each sample, 95% of the intervals would contain the population mean.
B) A 95% confidence interval has a 0.95 probability of containing the population mean.
C) 95% of the population distribution is contained in the confidence interval.

(13) Suppose that we conduct a survey of 190 people to find out what percent of their income the average person donates to charity. We discover that the mean is 15%. Find a 95% confidence interval for the mean percentage donated to charity. Assume that the distribution of all charity percents denoted is approximately normal. use formula for margin of error. $2\sqrt{0.15(1-0.15)/190}$

A) 9.8% - 20.2% B) 12.4% - 17.6% C) 14.8% - 15.2% D) 14.6% - 15.4%

(14) A bag contains 13 numbered balls. One at a time, a ball is drawn from the bag, noted, and not returned to the bag. The number of ways you can choose 4 unordered numbers from the bag in this manner is

A) 28,561 B) 52 C) 17,160 D) 715 $\frac{13\times12\times11\times10}{4\times3\times2\times1}$ ← divide by 4! since order does not matter.

(15) Two fair dice are rolled and the sum rolled is recorded. The probability that the sum is not 3 is

A) 1/18 B) 15/16 C) 5/6 D) 17/18 Whoops! Make sure to read question.

(16) The regression equation $y = 50 + 0.1x$ gives the yield in bushels per acre of corn when x pounds of fertilizer are applied. If a farmer hopes to yield 70 bushels of corn per acre, how much fertilizer should she use?

A) 200 B) 20 C) 57 D) 120 $70 = 50 + 0.1x$ $20 = 0.1x \longrightarrow x = 200$

(17) The regression equation $y = 3.15x - 20.68$ gives the stopping distance in feet when a car traveling x miles per hour applies the brakes. How much will the stopping distance increase between a car going 30 mph and one going 40 mph, using this regression equation? $[3.15(40) - 20.68] - [3.15(30) - 20.68]$

A) 105.32 ft B) 73./82 ft C) 31.5 ft D) -31.5 ft E) 20.68 ft

(18) The regression equation $y = 3.15x - 20.68$ gives the stopping distance in feet when a car traveling x miles per hour applies the brakes. How fast should a driver go if they want to be able to stop within 90 ft, using this regression equation? $90 = 3.15x - 20.68$ $x = 35.1365$

A) 35 mph B) 22 mph C) 590 ft D) 0.02846 mph E) 262 mph

(19) Which of the following properties are <u>not</u> properties of the normal distribution?

A) The mean is the same as the median.
B) The distribution is completely described by the mean and standard deviation.
C) The distribution is symmetric.
D) The tails of the distribution fall off rapidly.
E) None of the above.
F) All of the above.

(20) Which of the following modifications to a fair die will result in a fair die?

A) Shaving one side down a little bit.
B) Rounding off a few of the corners more than others.
C) Rounding off all the corners the same amount.
D) Adding (or removing) weight to the inside of the die near one of the faces.
E) None of the above.

Section 4. Short Answer

(1) The scores of students on a standardized test are normally distributed with a mean of 300 and a standard deviation of 40.

(i) Between what two values do 95% of the test scores lie? 220 - 380

(ii) What percentage of students scored above 380 on the test? 2.5%

(iii) Between what two values do 68% of the test scores lie? 260 - 340

(iv) What percentage of students scored above 340 on the test? 16%

(v) What percentage of students scored above 341 on the test? need calculus to answer.

(2) For the following probability model:

Event	A	B	C	D	E	F
Probability	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{2}{7}$	$\frac{1}{7}$

(i) The probability of events A, B, or C occurring is 3/7

(ii) The probability of event A not occurring is 6/7

(iii) The probability of event A and then B occurring is $1/49 = \frac{1}{7} \times \frac{1}{7}$

(iv) The probability of event A occurring 5 consecutive times is $(1/7)^5 = \frac{1}{16807}$

(v) The probability of event E not occurring in two consecutive trials is $(5/7)(5/7) = 25/49$

Answers at end.

(3) A national poll asked 1581 adults whether they were satisfied with their jobs, and 54% said they were. Estimate a 95% confidence interval for the actual percentage among all adults.

(4) In order to determine the support for building a new stadium in a city, a local news reporter goes to the city's football team's home game and interviews fans. Explain why this is a poor sampling technique.

(5) A high school principal wishes to determine the proportion of the school's students who like the school's new mascot. The principal decides to survey every 25th name from the school enrollment records (an alphabetical list of all students at the school). Is this a valid random sample technique? Why or why not?

(6) A farmer believes playing classical music to his chickens will cause them to produce more eggs. Describe how the farmer may design a randomized comparative experiment to test this theory.

(7) Three fair coins are flipped. What is the probability of getting exactly two heads?

(8) Explain why we need probability rules to figure out the probability model for rolling two unfair dice–why can't we just write out the sample space and use counting?

(9) What is the sample space for the following game:

Roll a fair 6-sided red die, and if the result is less than or equal to 2 you squirt your friend twice with a water gun, otherwise you give your friend a hug.

(10) The Regional Manager of Startbucks wants to do unannounced site visits to six of the 54 Starbucks Coffee Shops in a neighbourhood. The Manager wishes to use a simple random sample to determine which particular coffee shops to visit.

(a) What should be the labels for the coffee shops? The numbers 01, 02, ..., 54 are the labels.
(b) Using the Table below starting at row 1, which labels should be selected for the site visits?

Visit sites associated with labels
54
01
04
37
30 and 27

Row	Random Digits									
1	845	474	010	437	302	787	936	238	913	855
2	239	542	157	614	383	792	478	913	709	908
3	199	114	672	017	056	854	820	874	750	199
4	306	054	894	563	970	231	227	117	624	515
5	946	397	276	708	563	378	059	163	274	123
6	765	304	672	922	636	863	871	305	808	938

Section 5. Long Answer

Answers at end

(1) What are the odds of winning a lottery where 5 integer numbers are drawn between 1 and 1000 (once a number is drawn it cannot be drawn again, and winning means matching all 5 numbers drawn in any order)?

(2) Create a stemplot for the distribution of test scores: 56, 90, 77, 78, 77, 70, 88, 82, 83, 71, 89.

(3) Find the five number summary for the distribution of test scores: 56, 90, 77, 78, 77, 70, 88, 82, 83, 71, 89.

(4) The regression equation $y = 50 + 0.1x$ gives the yield in bushels per acre of corn when x pounds of fertilizer are applied. Sketch this regression equation, and explain what the meaning of the y-intercept is.

(5) A pizza can be made with any of the following toppings: cheese, pepperoni, sausage, mushrooms, anchovies, green peppers, or olives. How many different three topping pizzas can be made?

(6) A game consists of tossing three fair coins. Find the probability model for the number of coins showing heads. What is the mean number of heads for this game?

(7) Use the probability rules to find the probability model when rolling two unfair dice, with probabilities given by the following:

Die 1:

Outcome	1	2	3	4
Probability	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{6}$

Die 2:

Outcome	1	2	3	4
Probability	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{3}$

short Answer

(3) margin of error $= 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ $\hat{p} = 0.54$, $n = 1581$

$= 2\sqrt{\frac{0.54(1-0.54)}{1581}}$

$= 0.02507$

95% confidence interval $0.54 - 0.02507$ to $0.54 + 0.02507$

0.51493 to 0.56507

51.5% to 56.5%

(4) Fans at the game are going to more strongly support building a new stadium than the general population. This is a convenience sample, and is not going to be able to adequately sample the diversity of the population. People who are not fans of football or cannot afford to attend games are not being represented in the sample, so the sample suffers from undercoverage.

(5) A random sample requires that every possible sample of the same size has an equal chance of being selected, but in this sample two students who are side-by-side on the list cannot be in the sample. This alone makes it not a simple random sample.

(6) Divide chickens into two groups of same size using random sampling technique. For one group, play classical music and see how many eggs are laid over a time period. For second group, play no classical music and see how many eggs are laid in same time period. Try to keep all other variables the same between the two groups (amount they are fed, time they are fed, how they a fed, etc).

(7) Possible outcomes:

TTT THH HHH
TTH HTH
THT HHT
HTT

Since coins are fair, each outcome is equally likely, so we get exactly 2 heads $\frac{3}{8}$ of the time

(8) We can't just use counting since each of the outcomes does not have the same probability of occuring.

(9) The sample space is: two squirts with water gun, hag

(these are the ultimate outcomes of the game).

Long Answer

(1) $1000 \times 999 \times 998 \times 997 \times 996 = 990,034,950,024,000$

Divide by 5! to remove fact that the order does not matter.

$5! = 120$

$$\frac{990,034,950,024,000}{120} = 8,250,291,250,200$$

odds of winning $\frac{1}{8,250,291,250,200} \sim 1.2 \times 10^{-13}$

(2)

5	6
6	
7	01778
8	2389
9	0

(3)

56 ← Min = 56
70
71 ← Q = 71
77
78
78 ← Median 78
82
83
88 ← $Q_3 = 88$
89
90 ← Max = 90

(4) Sketch straight line by finding two points

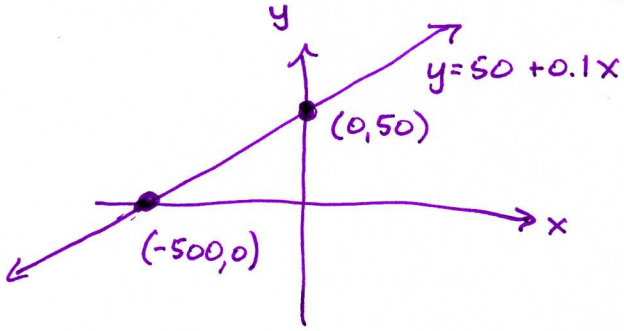
$y = 50 + 0.1x$

x	y
0	50
−500	0

$0 = 50 + 0.1x$

$-50 = 0.1x$

$-500 = x$



The y-intercept gives the yield in bushels per acre when no fertilizer is used.

(5) There are 7 different toppings, from which we want to choose 3. The order we choose the 3 toppings does not matter, so we should divide by $3! = 3 \times 2 \times 1 = 6$.

$$\frac{7 \times 6 \times 5}{6} = 35 \text{ different 3 topping pizzas.}$$

(6) For single fair coin:

Event	H	T
Prob	$\frac{1}{2}$	$\frac{1}{2}$

For 3 coins:

Event	0H	1H	2H	3H
Prob	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$
	0.125	0.375	0.375	0.125

(number of heads)

← must be $\frac{1}{8}$ since sum of probabilities must be 1.

$P(0H) = P(TTT)$

$= P(T \text{ and } T \text{ and } T)$

$= P(T)P(T)P(T)$

$= \left(\frac{1}{2}\right)^3$

$= \frac{1}{8}$

$P(1H) = P(TTH \text{ or } THT \text{ or } HTT)$

$= P(TTH) + P(THT) + P(HTT)$

$= \left(\frac{1}{8}\right) + \left(\frac{1}{8}\right) + \left(\frac{1}{8}\right)$

$= 3/8$

$P(2H) = P(THH \text{ or } HTH \text{ or } HHT)$

$= P(THH) + P(HTH) + P(HHT)$

$= \left(\frac{1}{8}\right) + \left(\frac{1}{8}\right) + \left(\frac{1}{8}\right)$

$= \frac{3}{8}$

(7)

Event	2	3	4	5	6	7	8
Prob	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{7}{36}$	$\frac{1}{4}$	$\frac{7}{36}$	$\frac{1}{6}$	$\frac{1}{18}$

$P(\text{roll } 2) = P(1 \text{ and } 1)$

$= P(1)P(1)$

$= \left(\frac{1}{3}\right)\left(\frac{1}{6}\right)$

$= \frac{1}{18}$

1st will always be Die 1, 2nd Die 2.

$P(\text{roll } 3) = P(2 \text{ and } 1 \text{ or } 1 \text{ and } 2)$

$= P(2)P(1) + P(1)P(2)$

$= \left(\frac{1}{6}\right)\left(\frac{1}{6}\right) + \left(\frac{1}{3}\right)\left(\frac{1}{6}\right)$

$= \frac{3}{36} = \frac{1}{12}$

$P(\text{roll } 4) = P(3 \text{ and } 1 \text{ or } 2 \text{ and } 2 \text{ or } 1 \text{ and } 3)$

$= P(3)P(1) + P(2)P(2) + P(1)P(3)$

$= \left(\frac{1}{3}\right)\left(\frac{1}{6}\right) + \left(\frac{1}{6}\right)\left(\frac{1}{6}\right) + \left(\frac{1}{3}\right)\left(\frac{1}{3}\right)$

$= \frac{7}{36}$

$P(\text{roll } 5) = P(4 \text{ and } 1 \text{ or } 3 \text{ and } 2 \text{ or } 2 \text{ and } 3 \text{ or } 1 \text{ and } 4)$

$= \left(\frac{1}{6}\right)\left(\frac{1}{6}\right) + \left(\frac{1}{3}\right)\left(\frac{1}{6}\right) + \left(\frac{1}{6}\right)\left(\frac{1}{3}\right) + \left(\frac{1}{3}\right)\left(\frac{1}{3}\right)$

$= 1/4$

$P(\text{roll } 6) = P(4 \text{ and } 2 \text{ or } 3 \text{ and } 3 \text{ or } 2 \text{ and } 4)$

$= \left(\frac{1}{6}\right)\left(\frac{1}{6}\right) + \left(\frac{1}{3}\right)\left(\frac{1}{3}\right) + \left(\frac{1}{6}\right)\left(\frac{1}{3}\right)$

$= 7/36$

$P(\text{roll } 7) = P(3 \text{ and } 4 \text{ or } 4 \text{ and } 3)$

$= \left(\frac{1}{3}\right)\left(\frac{1}{3}\right) + \left(\frac{1}{6}\right)\left(\frac{1}{3}\right)$

$= \frac{1}{6}$

$P(\text{roll } 8) = P(4 \text{ and } 4)$

$= \left(\frac{1}{6}\right)\left(\frac{1}{3}\right)$

$= \frac{1}{18}$