STAB22F: 2001 Midterm Exam-Solutions

Time: 2 hours
Instructions: Only use No. 2 black lead pencil on the Answer Sheet. First fill-in your student number and name on the Answer Sheet. Indicate your answer to each question by thoroughly filling in the appropriate circle on the Answer Sheet. Each question has the same weight. The exam paper must be turned in at the end of the exam.
Aids: The exam is open book. Students may use any notes, books and calculators in writing this exam.
Correct answers indicated by a *.

1. Suppose that X denotes the number of checks written in a given month by a randomly selected bank customer. Let the mean value of X be 22.2 and the standard deviation be 16.5. If a random sample of 100 customers is selected and $\bar{X}$ denotes the sample average number of checks written during a particular month, then which of the following statements is correct.

a)* The sampling distribution of $\bar{X}$ has mean 22.2.
b) The sampling distribution of $\bar{X}$ has standard deviation 16.5.
c) The sampling distribution of $\bar{X}$ is $N(22, 16.5)$.
d) None of the above.

2. What is the primary characteristic of a set of data for which the standard deviation is zero?

a) All values of the variable appear with equal probability.
b)* All values of the variable have the same value.
c) The mean of the values is zero.
d) All of the above are correct.
e) None of the above is correct.

3. Which of the following is true?

a) The probability that a continuous random variable lies in the interval 4 to 7, inclusively, is the sum of $P(4) + P(5) + P(6) + P(7)$.
b) If A and B are two events, then $P(A \cap B) = P(A)P(B)$, no matter what the relation between A and B.
c)* If two events are mutually exclusive, then $P(A \cup B) = P(A) + P(B)$.
d) All of the above are true.
e) None of the above are true.

4. A sample of 100 IQ scores produced the following statistics.

mean = 100	lower quartile = 70
median = 95	upper quartile = 120
mode = 75	standard deviation = 30

Which statement is correct?

a) Half of the scores are less than 100.
b) The most common score is 95.
c) The standard deviation is a better measure of dispersion than the quartiles.
d) One quarter of the scores are greater than 100.
e)* None of the above.

5. A sample of 100 students is randomly selected from the population of students at the University of Toronto and a variable X is measured for each student. Suppose the mean of the sample is 5.2 and the sample standard deviation is .5. Which one of the following intervals contains the true value of the population mean with virtual certainty and is shortest?

a) 5.2 ± 10
b) 5.2 ± 3
c) $5.2 \pm .098$
d)* $5.2 \pm .15$
e) 5.2 ± 2.1

6. A sample of pounds lost in a given week by individual members of a weight reducing clinic produced a mean equal to 95.6 pounds and a median of 98.2 pounds From these statistics, one could conclude that the shape of the distribution of pounds of lost weight is

a)* skewed to the left.
b) symmetric.
c) skewed to the right.
d) of any shape whatsoever.

7. A retail establishment accepts either the American Express or the VISA credit card. A total of 24% of it's customers carry an American Express card, 61% carry a VISA card, and 11% carry both. The percentage of it's customers that carry at least one of these credit card is

a) 11%.
b) 15%.
c)* 74%.
d) 85%.
e) 96%.

8. Suppose that you know that the distribution of heights of trees in a certain region is Chisquare(4) where the unit of measurement is meters. You are asked to determine the proportion of trees with heights less than 6 meters. Which Minitab command would you use?

a) `invcdf`
b) `pdf`
c)* `cdf`
d) `sample`
e) `random`

9. Pearson's coefficient of linear correlation r was computed to be 0.53 for a particular data set. Which of the following values of r represents a stronger linear relationship than 0.53?

a) 0.0
b) 0.5
c)* - 0.7
d) 20.3
e) None of the above.

10. A census is carried out on a population of size 700 and the values of two categorical variables X and Y are measured. Suppose that the following table of counts is obtained.

	$X = 1$	$X = 2$	$X = 3$
$Y = 1$	40	60	100
$Y = 2$	80	120	200
$Y = 3$	20	30	50

Which one of the following statements is correct?

a) A cause-effect relationship exists between X and Y.
b) A relationship exists between X and Y.
c) Whether or not a relationship exists depends upon which variable we select as the predictor and which variable we select as the response.
d)* No relationship exists between X and Y.
e) We really need to do a statistical test to conclude anything.

11. A student wishes to generate 40 samples from the $N(10, 3)$ distribution with each sample containing 10 values and then to calculate the standard deviation of each sample. Select the appropriate Minitab commands which do this.

a)
```
MTB > random C1 - C10;
SUBC > normal 10  3.
MTB > rstdev C1 - C10 C11.
```

b)
```
MTB > random 40 C1 - C10;
SUBC > normal 10  3.
MTB > rstdev C1 - C40 C41.
```

c)*
```
MTB > random 40 C1 - C10;
SUBC > normal 10  3.
MTB > rstdev C1 - C10 C11.
```

d)
```
MTB > random 40 C1 - C10;
SUBC > normal 10  3;
MTB > stdev C1 - C40 C41.
```

12. Suppose that we flip a fair coin four independent times and record whether or not exactly 3 heads are observed. What is the long-run proportion of times that we observe 3 heads in repeated performances?

a) 1/16
b) 1/2
c)* 1/4
d) 3/8
e) 3/16

13. The stem and leaf plot for a given data set was obtained as

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

Which of the following statements is NOT TRUE for the above plot?

a) The mean is greater than the median.
b) If the two high scores were deleted the mean of the new data set would be smaller than the mean of the data set displayed above.
c)* Adding 5 points to each score in the data set displayed above would alter both the mean and the standard deviation of the observations.
d) The spread of the data scores is best indicated by IQR.

14. The police in a small municipality know that 25% of the homeowners leave their doors unlocked. Crime records show that of those homes whose doors are left unlocked, 4% are burglarized. Then the probability that a home is both left with it's doors unlocked AND burglarized is

a)* 0.01
b) 0.29
c) 1.0
d) None of the above.

15. Two variables X and Y are measured on the members of a population. To determine if a relationship exists between these variables we must examine

a) the linear regression of Y on X.
b) the distributions of Y and X.
c)* the conditional distribution of Y given X for different values of X.
d) side-by-side boxplots for these variables.

16. The decline of visual sensitivity in adults as age increases is a well known phenomenon. Data on the number of neuron cells per gram of tissue (Y) for $n = 16$ males of various ages (X) in years ranging from 20 to 83 was collected. The following is a partial Minitab output.

The regression equation is
Neurons = 5.21 - 0.0351 Age.

Predictor	Coeff	SECoef	T	P
Constant	5.21300	0.4615	11.29	0.000
Age	-0.03509	0.008	- 4.39	0.001

According to the data analysis which of the following statement is true?

a) Old age causes loss in neuron cells per gram of tissue.
b) The predictor variable is the number of neuron cells per gram of tissue.
c) There is positive linear correlation between number of neuron cells per gram of tissue with age.
d)* Using this least squares line an 18 year old adult can be predicted to have around 4.6 neurons per gram.
e) None of the above.

17. The actual lecture time of a statistics instructor on any particular day has a distribution with mean value 52 minutes and standard deviation 2 minutes. Then for 36 randomly selected lecture times, indicate the appropriate Minitab command to evaluate the probability that the sample mean will be 55 minutes or less.

a)
```
MTB > cdf  55;
SUBC > normal 52  2.
```
b)*
```
MTB > cdf 55;
SUBC > normal  52  0.33.
```
c)
```
MTB > pdf 55;
SUBC > normal 52   2 .
```
d)
```
MTB > pdf  55;
SUBC > normal  52  0.33.
```

18. A sampling error is

a)* the difference between the value of the sample statistic based on a random sample and the value of the corresponding population parameter.
b) the error made while collecting, recording, and tabulating data.
c) the error that occurs because the sample is too small.
d) none of the above.

19. The Vision Apartments are a rather small complex. The owner is studying the number of unoccupied apartments and has gathered the following information.

No of Vacant Units	0	1	2	3
Proportion	.3	.3	.2	.2

The expected number of unoccupied apartments is given by

a) 2.0.
b)* 1.3.
c) 1.6.
d) 3.0.

20. Suppose that the probability of a person having a particular blood group is 0.6. Four individuals who have no relationship to one another are tested for the presence of this blood group. Suppose X is the random variable for the number of individuals having this particular blood group. Select the correct statement.

a) The Minitab command to calculate the probability that at least 2 individuals have this particular blood group is

```
 MTB > cdf  2 ;
SUBC > Bernoulli 0.6.
```

b)* The expected number of individuals having this blood group is 2.4
c) The probability that none of the individuals have this blood group is 0.4.
d) The random variable X is discrete taking values 0 and 1.

21. The following code was used to carry out a simulation in Minitab. What is being estimated?

```
MTB > random 1000 c1-c10;
SUBC> normal 0 1.
MTB > let c11=rssq(c1,c2,c3,c4,c5,c6,c7,c8,c9,c10)
MTB > let c12= c11>12
MTB > let k1=mean(c12)
MTB > print k1
```

a) The mean the distribution of the sum of squares of a sample of 10 from a $N(0,1)$ distribution.
b)* The probability that a the sum of squares of a sample of 10 from a $N(0,1)$ distribution is greater than 12.
c) The distribution function of a $N(0,1)$ distribution.
d) The long-run proportion of times that we get a value greater than 12 when taking samples of 10 from a $N(0,1)$ distribution.

22. Suppose X and Y are independently distributed random variables each having a mean μ and variance σ^2. Let $Z = 2X + 2Y + 1$. Then one can conclude that

a) $\text{Var}(Z) = 4\sigma^2 + 1$.
b) $\text{SD}(Z) = 2\sigma + 1$.
c) $\text{E}(Z) = 4\mu$.
d)* none of the above are true.

23. The probability of winning a bet on red at roulette is 18/38. A gambler makes 2 such bets. The probability that she loses both bet is
a) $2 * (1 - 18/38)$.
b) $1 - (2 * 18/38)$.
c)* $(1 - 18/38)^2$.
d) $1 - (18/38)^2$.
e) None of the above.

24. A teacher wants to determine if lecturing is useful for students with respect to learning a particular topic. To assess this he splits his class into two groups where one group will attend lectures and the other group will learn the material on their own. At the end both groups will write the same exam and the distribution of their scores will be compared. The teacher was careful to randomly allocate students to the two groups. Which of the following statements is correct?

a) This is an experiment and so if the teacher decides that there is an effect due to lecturing then he can assert that this is a cause-effect relationship.
b) This is a sampling study and if the teacher decides that there is an effect due to lecturing then he cannot assert that this is a cause-effect relationship.
c)* By randomly allocating students to groups the teacher has avoided his own biases from influencing the outcomes but any inferences he makes based on the outcomes will strictly speaking only apply to this class of students.
d) This is only an observational study and the teacher cannot draw any solid conclusions from the outcomes.

25. Suppose that an investigator computes a sample correlation coefficient between two variables X and Y as $r = .15$. The investigator then changes the units for X by multiplying each value of X by 3. The value of the correlation coefficient between this new variable and Y is

a)* .15
b) .45
c) .05
d) .35

26. Suppose that C1 in a Minitab worksheet contains 50 numbers. What does the following Minitab command do?

```
sample 30 c1 c2
```

a) It generates a random sample with replacement of size 30 from the values in C1 and places this sample in C2.
b)* It generates a random sample without replacement of size 30 from the values in C1 and places this sample in C2.
c) It generates a sample of 30 $N(\mu, \sigma)$ random variables, where μ and σ are contained in C1, and places the sample in C2.

27. Suppose that the probability of a randomly selected student passing course A is .95 and the probability that this same student will pass a follow-up course B (you can't take B unless you pass A) is .8. What is the probability that the student will fail one of these courses?

a) .25
b) .01
c)* .24
d) .5

28. The main reason we employ random selection from a population in a sampling study is because

a) this allows us to use Minitab for the analysis.
b)* this avoids selection effects so that inferences apply to the full population.
c) the central limit theorem can be used to approximate sampling distributions.
d) any relationships that we conclude exist can be interpreted as cause-effect relationships.

29. The Minitab commands

```
MTB > invcdf .6;
SUBC> normal 4 2.2.
```

compute

a) the probability that a N(4,2.2) random variable is less than or equal to .6.
b) the probability that a N(4,2.2) random variable is greater than .6.
c)* the value x such that $P(X <= x) = .6$ when X is a N(4,2.2) random variable.
d) a value X that can be considered a random value from a N(4,2.2) distribution.

30. The reason we quote a confidence interval is that

a) we want a range of plausible values for the unknown population quantity.
b) it is important to feel confident in our statistical results.
c)* the length of the interval helps to quantify our uncertainty about the true value of the population quantity being estimated.
d) it makes efficient use of the data.

31. Suppose we want to know the mean of a variable defined on a population and we do a census. Which statement below is true?

a) The mean is approximately normally distributed when the population size is large.
b) It is necessary to quantify how much error we believe there is in the estimate we quote.
c)* There is no need for statistical inference in this case.
d) A simple random sample could provide greater accuracy.

32. Consider a particular experiment for which a random variable results in one of the possible values, either 1,2 or 3 with probabilities 0.1, 0.3, and 0.6 respectively. A student wishes to generate 20 random samples of size 2 from this distribution using Minitab. He puts 1, 2 and 3 into column C1 and the corresponding probabilities into column C2. Which of the following Minitab commands is correct?

a) MTB> random 20 C1 C2.
b) MTB> random 20 C1 C2 C3 C4.
c) MTB> sample 20 C3 C4;
SUBC> discrete C1 C2.
d) MTB> random 20 C3 C4;
SUBC> pdf C1 C2.
e)* MTB> random 20 C3 C4;
SUBC> discrete C1 C2.