Instructor: Dr. Jennifer Murdock
Duration: 120 minutes (2:00 – 4:00 pm). You must stay in the test room for at least 90 minutes.
Part I: 20 multiple choice questions with answers recorded on SCANTRON form
Part II: 3 problems with answers written directly on test papers
Point values for Part I: Each multiple question worth 2 points. There are 40 total possible points for Part I. There are 100 total possible points overall for Parts I and II combined.
Allowed aids: A non-programmable calculator (and attached aid sheets, which you may detach)

<u>**INSTRUCTIONS FOR PART I:**</u>
Do NOT write your answers to the multiple choice questions on these test papers
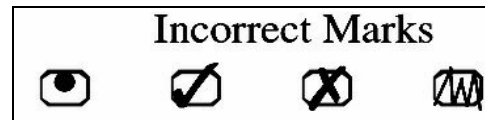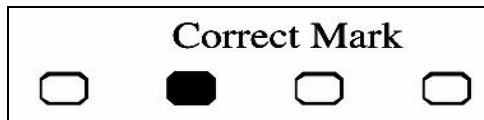ONLY those answers correctly marked on the SCANTRON form can earn marks
You <u>MAY</u> do scratch work on these pages

<div align="center">

SCANTRON INSTRUCTIONS

</div>

- Use <u>only</u> a blue pen, black pen or pencil

<div align="center">

BLUE OR BLACK | BALL POINT PEN   USE NO. 2 PENCIL ONLY

</div>

- **Pencil recommended**, it can be erased if a mistake is made

- Make dark solid marks that fill the bubble completely

<div align="center">

Correct Mark	Incorrect Marks

</div>

- Select the <u>one</u> best alternative
  - Questions with more than one answer selected will be scored incorrect

- Erase completely any marks you want to change
  - Do <u>not</u> use correction fluid

- Do <u>not</u> make stray marks on the form

- Answer every question: there is no penalty for guessing

1st : Print your **LAST NAME** and **INITIALS** in boxes provided
  - Use exact name you are <u>officially registered</u> under
  - Darken each letter in the corresponding bracket <u>below each box</u>

2nd : Print your 9 digit **STUDENT NUMBER** in the boxes provided
  - Fill in zeros in front of the number if less than 9 digits
  - Darken each number in the corresponding bracket <u>below each box</u>

**3rd :** Print 2 digit **FORM** number in the boxes provided
  - Your FORM number is 03 if you are in Section L0301 (11:10 – 12:00)
  - Your FORM number is 04 if you are in Section L0401 (12:10 – 1:00)
  - Darken each number in the corresponding bracket <u>below each box</u>

4th : Sign your name in the **SIGNATURE** box

(1) Consider two variables measuring the price and quantity sold of a particular brand of laundry detergent. Which of the following would be an example of <u>panel (longitudinal)</u> data?

    (a) Price and quantity for 50 different stores in a particular day
    (b) Daily price and quantity for a particular store for 7 different days
    (c) Daily price and quantity for 50 different stores for 7 different days
    (d) Average overall price and weekly quantity sold over 50 different stores over 7 different days
    (e) All of the above

(2) Which of the following histograms is <u>positively skewed</u>?
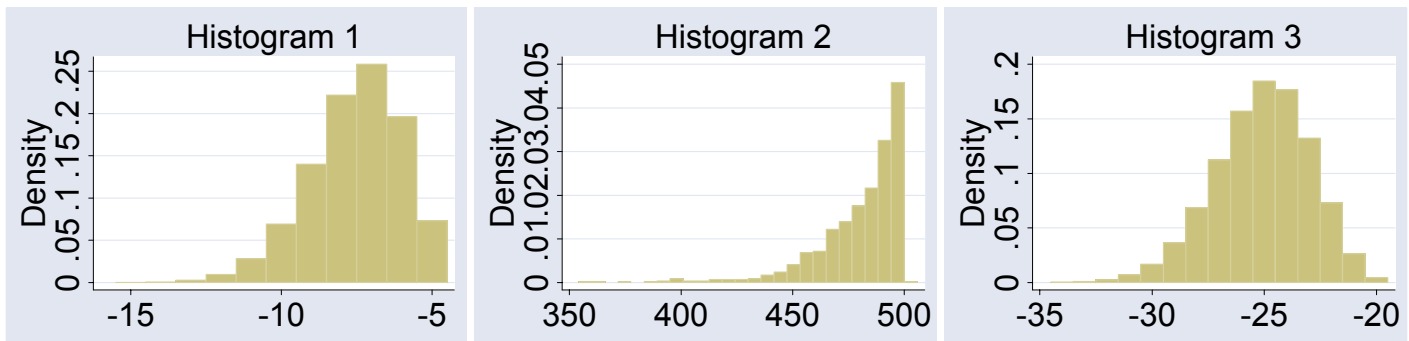
    (a) Histogram 1
    (b) Histogram 2
    (c) Histogram 3
    (d) All of the above
    (e) None of the above

(3) Consider a sample taken from a normal population. With numbers recorded to the second decimal place, which of the following statements about the percent of observations within 1 standard deviation (s.d.) of the mean is <u>true</u>?

    (a) About 68.00% of the observations will be within 1 s.d. of the mean
    (b) About 68.26% of the observations will be within 1 s.d. of the mean
    (c) About 68.30% of the observations will be within 1 s.d. of the mean
    (d) About 68.34% of the observations will be within 1 s.d. of the mean
    (e) None of the above

(4) Based on the tabulation of the variable X below, what is the sample <u>median</u> of X?

```
       x |      Freq.      Percent         Cum.
---------+-----------------------------------------
       1 |      2,570         4.60          4.60
       2 |     20,216        36.17         40.77
       3 |     19,944        35.69         76.46
       4 |      9,041        16.18         92.64
       5 |      2,896         5.18         97.82
       6 |        859         1.54         99.35
       7 |        253         0.45         99.81
       8 |         71         0.13         99.93
       9 |         27         0.05         99.98
      10 |          8         0.01        100.00
      11 |          2         0.00        100.00
---------+-----------------------------------------
   Total |     55,887       100.00
```

    (a) 2
    (b) 2.5
    (c) 3
    (d) 5.5
    (e) 6

(5) Consider data measuring the grades (X) of a sample of 132 University of Toronto students and their salaries 5 years after graduation (Y). You compute a sample covariance of 249.92 (cov(X,Y) = 249.92). What can you reasonably conclude about the sample based on this statistic?

    (a) There is a strong linear relationship between grades (X) and salary (Y)
    (b) There is a weak linear relationship between grades (X) and salary (Y)
    (c) There is no linear relationship between grades (X) and salary (Y)
    (d) There is a positive relationship between grades (X) and salary (Y)
    (e) None of the above

(6) Suppose X is a normally distributed variable with mean 10 and variance 20: X ~ N(10, 20). What is the probability that X is negative (rounded to nearest hundredth)?

    (a) 0.00
    (b) 0.01
    (c) 0.04
    (d) 0.48
    (e) None of the above

► For Questions **(7)** and **(8)**, consider <u>ALL</u> of the following information:

A survey is administered to a sample of people who have flown to a destination within North America from Toronto in the past 6 months. Question #12 in the survey asks: "With which airline have you flown with most often in the past 12 months: (1) Air Canada, (2) American, (3) U.S. Airways, (4) Continental, (5) West Jet?" A tabulation of the results for this question yields:

```
      q12 |      Freq.      Percent         Cum.
----------+-----------------------------------------
        1 |      1,548       47.81         47.81
        2 |        703       21.71         69.52
        3 |        548       16.92         86.44
        4 |        378       11.67         98.12
        5 |         61        1.88        100.00
----------+-----------------------------------------
    Total |      3,238      100.00
```

(7) What can you reasonably conclude about the sample based only on the above information?

  (a) The airlines flown by the sampled passengers are positively skewed
  (b) Assuming everyone answered Question #12, the sample size is 3,238
  (c) The answers to Question #12 are interval data with a mean of 1.98
  (d) Of the sampled passengers, 86.44% answered "U.S. Airways" to Question #12
 **(e)** All of the above

(8) What can you reasonably infer about the population based only on the above information?

  (a) About 50.3% of travelers in North America fly via American, U.S. Airways, or Continental
  (b) West Jet serves less than 2% of air travel passengers in Canada
  (c) Air Canada serves less than 50% of air travel passengers originating in Toronto
  (d) All of the above
  (e) None of the above

(9) Consider taking a sample of size 40 from a uniform population where a = 0 and b = 1: U[0,1]. Which of the following statements about the <u>sample range</u> and <u>population range</u> is true?

  (a) The expected value of the sample range is less than the population range
  (b) The sample range will be equal to the population range
  (c) According to the Central Limit Theorem (CLT) the sample range will be normally distributed and centered at the population range
  (d) The population range is infinite because the uniform distribution has unbounded support
  (e) The expected value of the sample range is the population mean (μ)

► For Questions **(10)** and **(11)**, consider the histogram below:
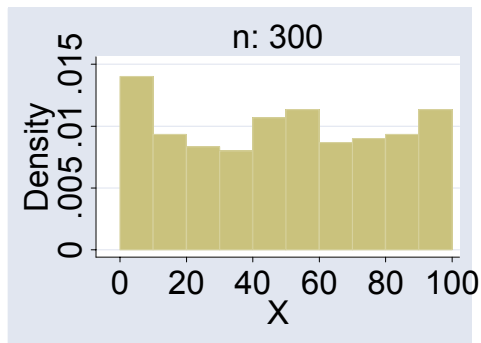
(10) Which is closest to the sample <u>standard deviation</u>?

    (a) 10
    (b) 20
    (c) 30
    (d) 40
    (e) 50

(11) Which is closest to the sample <u>90th percentile</u>?

    (a) 0.01
    (b) 0.09
    (c) 10
    (d) 80
    (e) 90

(12) Suppose hours per year employees spend surfing the internet at work is normally distributed. If 13.4% of employees spend more than 160 hours, how many standard deviations above the mean is 160?

    (a) 1.11
    (b) 1.34
    (c) 1.50
    (d) Cannot be computed with information provided
    (e) None of the above

► For Questions **(13)** and **(14)**, consider the joint probability table below. A company has two types of sales representatives (reps): junior and senior. Dissatisfied customers can register a complaint against their sales rep. A manager would like to study whether there is any relationship between the junior or senior status of a sales rep and the number of customer complaints.

Number of Complaints	Junior Sales Rep	Senior Sales Rep
0	0.10	0.35
1	0.10	0.20
2	0.05	0.10
3	0.05	0.05
4	0	0

(13) Which is the expected number of complaints for each type of sales representative (round answer to nearest tenth)?

    (a) Junior: 0.3; Senior: 0.7
    (b) Junior: 0.4; Senior: 0.6
    (c) Junior: 0.5; Senior: 1.1
    (d) Junior: 0.9; Senior: 0.9
   **(e)** Junior: 1.2; Senior: 0.8

(14) If the manager estimates that each complaint results in $250 in lost sales and there are 100 sales reps in total, how much money in total does the manager expect to lose due to complaints (round answer to nearest dollar)?

    (a) $22,500
    (b) $25,000
    (c) $50,000
    (d) $55,500
    (e) None of the above

(15) Suppose a random variable X is binomially distributed. In which of the following cases can the probabilities associated with the binomial distribution be reasonably approximated by the probabilities associated with the normal distribution?

    (a) n = 200 and p = 0.1
    (b) n = 100 and p = 0.2
    (c) n = 200 and p = 0.2
    (d) All of the above
    (e) None of the above

(16) Consider the population density function given below. If 1 observation is drawn you can be 90% confident that the <u>sample mean</u> will fall in what interval (round to the nearest tenth)?
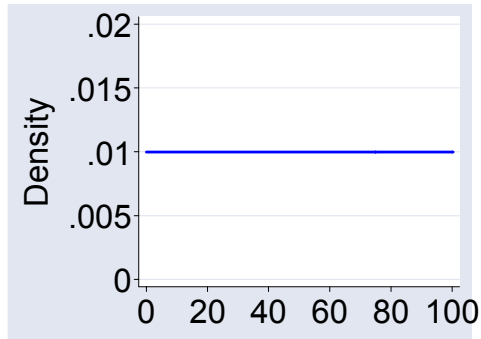
(a) (2.5, 97.5)
(b) (2.8, 97.2)
(c) (5.0, 95.0)
(d) (10.0, 90.0)
(e) None of the above

(17) Suppose X1_bar measures the average profit margin for a sample of monopolized markets. Suppose X2_bar measures the average profit margin for a sample of duopoly markets. Given the sampling distribution of the difference between these means shown below, what is the probability that due to sampling noise we obtain two samples where the average profit margins of the duopolists are actually higher than the monopolists? (Round to nearest hundredth.)
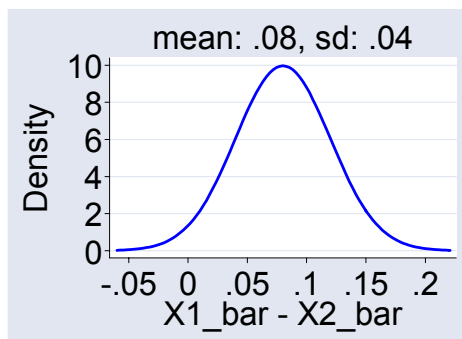
(a) 0
(b) 0.02
(c) 0.04
(d) 0.05
(e) None of the above

► For Question **(18)**, suppose a researcher conducts a Monte Carlo simulation where a sample of 10 observations is taken from a known population and computes the sample standard deviation. This sampling experiment is repeated 100,000 times and the simulation results are summarized below.

```
                     Sample Standard Deviation
------------------------------------------------------------
     Percentiles        Smallest
  1%    1.577621        .4830459
  5%    1.988858        .5163978
 10%    2.162817        .5163978        Obs              100000
 25%     2.54733        .6992059        Sum of Wgt.      100000

 50%     2.82056                        Mean            2.782978
                        Largest         Std. Dev.        .4376082
 75%    3.098387        3.977716
 90%     3.29309        4.049691        Variance         .1915009
 95%    3.403429        4.056545        Skewness        -.6313736
 99%    3.597839        4.056545        Kurtosis         3.43377
```

(18) Using the simulation results, compute the interval that should contain the sample standard deviation with 90% confidence. Which show the correct values rounded to nearest tenth?

    (a) (2.0, 3.4)
    (b) (2.1, 3.5)
    (c) (2.2, 3.3)
    (d) (2.5, 3.0)
    (e) None of the above

► For Questions **(19)** and **(20)**, suppose that using data from a telephone survey a 95% confidence interval estimate of μ is calculated to be (-13.454, -10.921).

(19) What is the point estimate of μ ?

    (a) -13.454
    (b) -10.921
    (c) -11.428
    (d) -12.591
    (e) None of the above

(20) Which of the following would <u>increase the width</u> of the interval estimate?

    (a) Errors in entering the answers to the survey question
    (b) Misleading language in the survey question
    (c) Failure to roster the individuals associated with each primary telephone number
    (d) Calling a smaller set of randomly generated telephone numbers
    (e) All of the above

Last Name:

First Name:

Student ID #:

Instructor: Dr. Jennifer Murdock
Duration: 120 minutes (2:00 – 4:00 pm). You must stay in the test room for at least 90 minutes.
Part I: 20 multiple choice questions with answers recorded on SCANTRON form
Part II: 3 problems with answers written directly on test papers
Point values for Part II: Point values for problems located next to each. There are 60 total possible points for Part II. There are 100 total possible points overall for Parts I and II combined.
Allowed aids: A non-programmable calculator (and attached aid sheets, which you may detach)

INSTRUCTIONS FOR PART II:

For this part, you must legibly write all answers in the space provided on these test papers.

If you run out of room use the last two pages and clearly indicate which question (number and part) you are continuing. For example, "Answer to Question 2 (a) continued."

Significant points will only be awarded to responses that are correct, clear, complete, and precise.

Failure to show all work will result in 0 points awarded.

DO NOT WRITE IN THIS TABLE:

Marks	Q1	Q2	Q3	Total for Part II
Maximum possible	*22*	*18*	*20*	*60*
Marks earned				

(1) [22 points] A manufacturer of a household appliance runs a television advertisement that claims that only 2% of its products will need a service call for repair during the first year of ownership. A consumer protection association wants to check the claim by surveying 500 households that recently purchased one of the manufacturer's appliances.

(a) [8 points] What is the probability that more than 3% of the sampled products need a service call?

(b) [3 points] State your result in part **(a)** in <u>one sentence</u> that indicates the EXACT interpretation of the number you found. (Simply explain in words what the result is. You will make conclusions and inferences in part **(c)** but <u>not</u> in this part.)

(c) [3 points] If the consumer protection association finds a sample proportion of 3% does that falsify the manufacturer's claim?

(d) [8 points] Suppose that due to budget cutbacks the consumer protection association can only afford to sample 100 customers (instead of 500 as indicated above). If the manufacturer's claim is true, what is the probability that more than 3% of the sampled products will need a service call? (Hint: Think carefully about this part.)

(2) [18 points] Suppose three balls are drawn from this box with replacement.

(a) [12 points] Find the sampling distribution of the sample median.
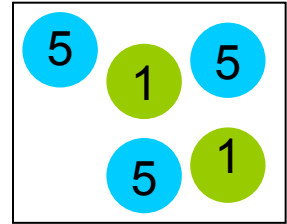
(b) [6 points] For this experiment is the sample median an unbiased estimator of the population median? (Prove any claim you make mathematically.)

(3) [20 points] To conduct a pretest for a larger survey, a researcher selects a random sample of 6 University of Toronto students that do <u>not</u> live on campus. Each student is asked: "In a typical week, how many hours do you spend on campus? Include hours spent in lectures, tutorials, libraries, student common areas and any other space on University property."

Responses: 29   34   33   38   15   19

(a) [12 points] Find the 95% confidence interval estimate of the mean. (Round off to nearest hour.)

(b) [4 points] State your result in part **(a)** in <u>one sentence</u> that indicates the EXACT interpretation of the numbers you found.

(c) [4 points] As part of the pretest the researcher debriefs each interviewee, which means that the researcher goes over with the interviewees how they answered the questions and what factors they considered. The researcher realizes that 2 of the 6 students forgot to include hours they spent working on campus in part-time positions and volunteer organizations. How does this affect the interpretation of the interval estimate?

Extra Space: If you use this space, clearly indicate for which question(s) and part(s).

Extra Space: If you use this space, clearly indicate for which question(s) and part(s).