Last Name:

First Name:

Student ID #:

Instructor: Prof. Murdock
Duration: 50 minutes
Format: 5 questions
Total Points: 50 possible points total
Allowed Aids: A non-programmable calculator and distributed aid sheets

INSTRUCTIONS:

Write your answers clearly, concisely, and completely on these test papers.

You do not have to fill all of the blank space after each question: a generous amount is provided for your convenience.

For questions that include a writing component, I give you a requested length for your response in brackets. For example, "Would any non-sampling errors affect your answer? [2 – 3 sentences]"

Show all work. If you are solving a problem mathematically make sure you clearly show *how* you are approaching the problem: connect your mathematical arguments to your overall argument. (This becomes especially important if you make a quantitative mistake.)

You are not required to draw a graph unless one is specifically requested.

For questions with multiple parts, attempt each part even if you had trouble with previous parts.

If you run out of room you may continue your answer on Page 9, but indicate you have done so and clearly label your additional responses (for example: "Question (2) (b) continued:").

Students writing this test from 10:10 – 11:00 must stay in the test room for the entire 50 minutes. Students writing this test from 11:10 – 12:00 may not arrive late.

Marks	Q1	Q2	Q3	Q4	Q5	Total
Maximum possible	12	6	12	14	6	50
Marks earned						

Percent Mark:

(1) [12 pts] For 70 percent of a large population X is 0 and for 30 percent X is 60. For a sample size of 3, find the sampling distribution of the sample mean. (Use three decimal places: 0.000.)

(2) [6 pts] A skewed population has a mean of 1.65 and a standard deviation of 2. A simulation to find the sampling distribution of the sample mean yields this STATA summary and graph.

```
                          Sample Mean
     --------------------------------------------------------------
          Percentiles        Smallest
    1%       .6737117         .3481812
    5%       .8396572         .3592617
   10%       .9489565         .3662439       Obs             500000
   25%       1.172941         .3738413       Sum of Wgt.     500000

   50%       1.501314                        Mean            1.648705
                              Largest        Std. Dev.       .7070909
   75%        1.95057         13.5546
   90%       2.505256         15.42728       Variance        .4999776
   95%       2.940229         16.66789       Skewness        2.15083
   99%       4.071384         16.76013       Kurtosis        14.58178
```
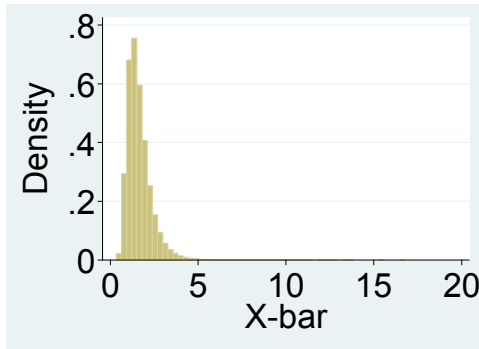

(a) [3 pts] Find the sample size.

(b) [3 pts] What is the most important reason that we needed this simulation? [2 – 3 sentences]

(3) [12 pts] A sample size of 52 is drawn from a Uniform population from 0 to 100, $X \sim U[0, 100]$.

(a) [7 pts] Find the sampling distribution of the sample mean. [analysis & 1 – 2 sentences]

(b) [5 pts] Find these probabilities: $P(45 < \overline{X} < 55)$ and $P(45 < X < 55)$.

(4) [14 pts] A university administration claims that only 40 percent of students support the Student Union (SU). The SU decides to survey students.

(a) [6 pts] In a random sample of 50 students, 25 support SU. If the administration's claim is true, what is the chance that 25 or more students support SU? [analysis & 1 – 2 sentences]

(b) [8 pts] In a random sample of 6 students, 4 support SU. If the administration's claim is true, is sampling error a plausible explanation for such a large number supporting SU? [analysis & 1 – 2 sentences]

(5) [6 pts] Recall this multiple-choice question from Quiz #2. The correct answer is **(B)**.

To study the effect of caffeine on brain activity, researchers randomly select 500 students at different times and locations. They measure caffeine concentration (X) with a blood test and brain activity (Y) with electrodes. The slope of the least squares line will <u>not</u> measure the causal effect of caffeine on brain activity. Which is the best explanation of why not?

    (A) Students' caffeine sensitivities vary because of physiological differences

    (B) Factors such as fatigue affect caffeine consumption and affect brain activity

    (C) Brain activity is an endogenous variable that cannot be randomly set by the researchers

    (D) Researchers failed to measure intelligence (e.g. IQ), which certainly affects brain activity

    (E) An unobserved variable such as coffee availability (e.g. Second Cup nearby) can affect caffeine consumption

---

(a) [3 pts] Why is **(C)** an incorrect answer to the question? [2 – 3 sentences]

(b) [3 pts] Why is **(E)** an incorrect answer to the question? [2 – 3 sentences]

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