NAME:________________________

NAME:________________________

NAME:________________________

NAME:________________________

Suppose you have an infinite population and you sample from it n times. The random variable associated to the sampling is called x and its mean and variance are respectively μ and σ^2 (the same as for the population). Now consider the two random variables

$$\overline{x} = \frac{\sum x}{n} \quad \text{and} \quad y = \frac{\sum (x - \overline{x})^2}{n}.$$

Hint: you need to think about expressions like

$$\sum x = x_1 + x_2 + \ldots + x_n$$

where the x_i's for $i = 1, \ldots, n$ are independent random variables with same probability distribution, since they are generated by sampling from the same population.

1. Compute the expected value of the random variable $\overline{x}$. How does it compare to the expected value of x?

2. Compute the variance of the random variable $\overline{x}$. How does it compare to the variance of x?

3. Prove that the random variable y is also equal to $\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2$.

4. Define $z = \overline{x}^2 = \left(\frac{\sum x}{n}\right)^2$. What is the expected value of z? (Hint: you need to use the third foundamental mystery of probability)

5. Compute the expected value of the random variable y. Is this smaller or greater than σ^2?

6. How would you modify the formula for the random variable y so that the expected value is exactly σ^2?

7. What was most surprising about what you discovered?

8. **Bonus question** What is the minimum number of people chosen at random such that the probability of at least two of them being born the same day of the year is more than 50%? (Assume that nobody was born February 29th of a leap year)