ST 371 MIDTERM 1

September 14, 2020

NAME:

- You have **48 hours** to complete this exam.

1. Suppose $x_1, \ldots, x_n$ with $n = 36$ is a simple random sample from some population, and suppose that these observed data values are described by the following density table. Note that the density values have been rounded to three decimal places.

bin	$0.0- < 1.5$	$1.5- < 3.0$	$3.0- < 4.5$	$4.5- < 6.0$	$6.0- < 7.5$	$7.5- < 9.0$
density	.072	.161	.120	.028	.253	.033

   (a) (3 points) Which bin contains the sample median?

   (b) (3 points) Assuming that $x_i \in \{0.9, 2.1, 3.3, 5.0, 6.2, 7.7\}$ for all $i \in \{1, \ldots, 36\}$, what is the sample variance of $x_1, \ldots, x_{36}$? Due to the issue of rounding, use the relative frequencies (rather than the frequencies) of each bin in your calculation.

2. Suppose that two players bet on the outcomes of successive coin tosses. Player 1 bets 1 dollar that the coin will land heads up, while Player 2 bets 1 dollar that the coin will land tails up. Each player begins with 6 dollars. Assume that there is .6 probability that the coin will land heads up.

   (a) (3 points) What is the probability that both players have exactly 6 dollars after the coin is tossed 6 times?

   (b) (3 points) What is the probability that Player 1 wins all the money on the tenth coin toss?

3. Suppose that a rare disease has been linked to 8 identified risk factors, where a risk factor is some feature that an individual either does or does not exhibit (e.g., an individual either uses tobacco or does not use tobacco). Assume that a randomly selected individual exhibits each of the 8 risk factors independently and with probability .5. Further, assume that an individual is considered to be at high risk for the disease if they exhibit any combination of at least 3 of the 8 risk factors. More precisely, the disease is observed in 30 percent of high risk individuals, whereas it is only observed in 5 percent of individuals that are not high risk.

(a) (3 points) Within these 8 risk factors, how many outcomes exist that categorize an individual to be at high risk for the disease?

(b) (3 points) What is the probability that a randomly selected individual is at high risk for the disease?

(c) (3 points) Alternatively, if it is known that a particular individual has the disease, what is the probability that they are at high risk?

4. Let $A_1, \ldots, A_k$ be events of outcomes contained in some sample space $\mathcal{S}$, and assume that $A_i \subseteq A_j$ for $i \leq j$. Recall that the set minus operation, $A_j \setminus A_i$, denotes the event consisting of all outcomes in A_j that are not also contained in A_i. Prove the following statements.

   (a) (3 points) $P(A_j \setminus A_i) = P(A_j) - P(A_i)$ for $i \leq j$.

   (b) (3 points) $P\Big(\bigcup_{i=2}^{k}(A_i \setminus A_{i-1})\Big) = P(A_k) - P(A_1)$.