# Statistics Workshop (STA-2023)

## Probability & Discrete Probability Distribution

**Probability Rules**

What is probability and how do we find it?

Probability is the chance or likelihood of an event (E) occurring and is most often written as a decimal instead of a percentage. In statistics, we define an event as a collection of outcomes or results from an experiment or survey.

For example, say I have a bag filled with candy. In this bag there are 6 gummy candies, 8 chocolates, and 7 hard candies. What would be the probability of me selecting a chocolate candy from the bag?

In this example, the event would be selecting a chocolate candy and the number of all possible outcomes would be the total number of candies in a bag.
To the probability of selecting a chocolate candy, we would use the formula

$$P(event) = \frac{the\ number\ of\ ways\ the\ event\ can\ occur}{number\ of\ possible\ outcomes}$$

How many ways can we select chocolate? Since there are 8 chocolate candies, it would be 8 ways.

So, the probability would be $P(E) = \frac{8}{21} = 0.3810$, this means that there is an approximate chance of 38.1% that I would select a chocolate candy.

1) A bag of 100 tulip bulbs purchased from a nursery contains 45 red tulip bulbs, 20 yellow tulip bulbs, and 35 pink tulip bulbs.

(A) what is the probability that a randomly selected tulip bulb is pink?

$$P(E) = \frac{\quad}{100} =$$

(B) What is the probability that a randomly selected tulip bulb is yellow?

$$P(E) = \frac{\quad}{\quad} =$$

2) From a survey of 1300 female adults (18 years or older), it was determined that 410 volunteered at least once in the past year. What is the probability that a randomly selected adult female volunteered at least once in the past year?

$$P(E) = \frac{\quad}{\quad} =$$

**Addition Rule and Complements**

Now let's consider what happens when we take the probability of more than one event at a time, from our example of the bag of candies, what would be the probability of randomly selecting a chocolate candy or a gummy candy? Since a gummy candy and a chocolate candy are to distinct types of candy these events would be disjoint, meaning that they don't have any outcomes in common. In other words, there is no chance to select a candy which is chocolate and gummy at the same time.

When two events (let's call them events E and F) are disjoint, the probability of randomly selecting E or F would be found by the addition rule (indicated with the word "or"),

$$P(E\ or\ F) = P(E) + P(F)$$

So, for our example, the probability of randomly selecting a gummy candy or a chocolate would be $P(E\ or\ F) = \frac{6}{21} + \frac{8}{21} = 0.6667$.

If the events E and F do have outcomes in common, then we use the General addition rule

$$P(E\ or\ F) = P(E) + P(F) - P(E\ and\ F)$$

Notice that in our example of the bag of candies, $P(E\ and\ F) = 0$ since they are disjoint (or mutually exclusive) events.

Sometimes we need to find the probably of an event NOT occurring, in this case we would use the complement rule

$$P(E^C) = 1 - P(E)$$

where $E^C$ is the complement of E.

The complement $E^C$ is all of the outcomes in the sample space that are NOT in event E.

For example, in our bag of candies, what is the probably of me randomly selecting a candy that is not chocolate? Well since the probably of selecting a chocolate is $0.381$, which we found before, the probability of selecting a candy that is not chocolate would be $P(E^C) = 1 - 0.3810 = 0.6190$.

3) A bag of 100 tulip bulbs purchased from a nursery contains 45 red tulip bulbs, 20 yellow tulip bulbs, and 35 pink tulip bulbs.

(A) Find the probability of randomly selecting a pink or yellow tulip?

$$P(E) = \frac{\quad}{\quad} + \frac{\quad}{\quad} =$$

(B) What is the probability of not selecting a yellow tulip?

$$P(E) =$$

4) Find the probability of the following, excluding leap years from the calculations.

(A) Compute the probability that a randomly selected individual has a birthday on the 1st of a month or was born in April.

$$P(E) =$$

(B) Find the probability that a randomly selected individual was not born in November.

$$P(E) =$$

**Independence and the Multiplication Rule**

Independent events E and F means that the probability of E occurring does NOT affect the probability of event F occurring. If events E and F are dependent that indicates that the probability of event E occurring has an effect on the probability of event F occurring. This is an important distinction when using the multiplication rule.

For independent events, we can find the probability of occurring event E and F by the multiplication rule (indicated with the word "and")

$$P(E \text{ and } F) = P(E) \cdot P(F)$$

If there are more than two independent events, such as

$$P(E_1 \text{ and } E_2 \text{ and } E_3 \text{ and } \dots \text{and } E_n) = P(E_1) \cdot P(E_2) \cdot P(E_3) \cdot \dots \cdot P(E_n).$$

For example, let's go back to the bag filled with candy. What is the probability of selecting a chocolate and a hard candy?

Well, plugging our numbers into the formula, we have $P(E \text{ and } F) = \frac{8}{21} \cdot \frac{7}{21} = \frac{8}{63} = 0.1270.$

**Conditional Probability and General Multiplication Rule**

The conditional probability is indicated by the word "given". The word, conditional (like dependent) indicates that the occurrence of one event affects the occurrence of another. The formula for conditional probability is

$$P(F|E) = \frac{P(E \text{ and } F)}{P(E)},$$

$P(F|E)$ is read as "the probability of F given E".

For example, suppose that E and F are two events and that $P(E \text{ and } F) = 0.22$ and $P(E) = 0.35$. Then, $P(F|E) = \frac{P(E \text{ and } F)}{P(E)} = \frac{0.22}{0.35} = 0.6286.$

The general multiplication rule for the probability that two events E and F both occur is

$$P(E \text{ and } F) = P(E) \cdot P(F|E).$$

In words, the probability of E and F both occur is the probability of event E occurring times the probability of event F occurring given the occurrence of event E.

5) A bag of 100 tulip bulbs purchased from a nursery contains 45 red tulip bulbs, 20 yellow tulip bulbs, and 35 pink tulip bulbs.

(A) What is the probability that both bulbs are pink?

$$P(E) = \frac{\quad}{\quad} \cdot \frac{\quad}{\quad}$$

(B) The probability that the first bulb is red and the second is yellow?

$$P(E) = \frac{\quad}{\quad} \cdot \frac{\quad}{\quad}$$

(C) What is the probability that the first is yellow and the second is red?

$$P(E) = \frac{\quad}{\quad} \cdot \frac{\quad}{\quad}$$

(D) What is the probability that one bulb is red and one is yellow?

$$P(E) = \frac{\quad}{\quad} \cdot \frac{\quad}{\quad} + \frac{\quad}{\quad} \cdot \frac{\quad}{\quad}$$

6) The following data represent the number of different communication activities (e.g., cell phone, text messages, e-mail, and so on) used by a random sample of teenagers over the past 24 hours.

| **ACTIVITIES** | **0** | **1-2** | **3-4** | **5+** | **TOTAL** |
|---|---|---|---|---|---|
| **MALE** | 22 | 85 | 47 | 46 | 200 |
| **FEMALE** | 22 | 58 | 56 | 64 | 200 |
| **TOTAL** | 44 | 143 | 103 | 110 | 400 |

(A) What is the probability that a randomly selected individual is male and has done 1-2 activities?
$P(E) =$

(B) What is the probability that a randomly selected individual has done 3-4 activities given that they are female?
$P(E) =$

(C) Are the events “male” and 0 activities independent? Why?

7) The local golf store sells an "onion bag" that contains 35 "experienced" golf balls. Suppose that the bag contains 20 Titleists, 8 Maxflis, and 7 Top-Flites.

(A) What is the probability that two randomly selected golf balls are both Titleists?

$P(E) =$

(B) What is the probability that four randomly selected golf balls are all Titleists?

$P(E) =$

(C) What is the probability that at least one randomly selected golf ball is not a Titlesists?

$P(E) =$

**Counting Techniques**

Now we are going to talk about finding the probability with counting techniques. The *multiplication rule of counting* says if we have p number of selections for the first choice, q number of selections for the second choice, r number of selections for the third choice, and so on, then the number of different ways the choices can be selected would be $p \cdot q \cdot r$.

For example, the fixed-price dinner at Mabenka Restaurant provides the following choices:

Appetizer: soup or salad
Entrée: Baked chicken, broiled beef patty, baby beef liver, or roast beef au jus
Dessert: Ice cream or cheesecake

How many different ways can meals be ordered? Here there are 2 choices for the first option, 4 choices for the second option, and 2 choices for the third. Plugging into the formula we have: $2 \cdot 4 \cdot 2 = 16$ ways a meal can be ordered.

Another important formula for counting techniques would be the *factorial* (symbol is "!") which is the number n multiplies by every number previous, $n! = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 3 \cdot 2 \cdot 1$. We use the factorial in the formulas for both combinations and permutations.

A permutation is an ordered arrangement (order is important!) in which r objects are chosen from n distinct objects so that $r \leq n$ and repetition is NOT allowed, the formula is

$${}_nP_r = \frac{n!}{(n-r)!}\ .$$

A combination is a collection without regard for order (order is NOT important!) in which r objects are chosen from n distinct objects with $r \leq n$ and repetition is NOT allowed, the formula is

$$_nC_r = \frac{n!}{(n-r)!\,r!}$$

A permutation can also be done WITHOUT distinct objects, where $n_1$ are the number of one kind of object and $n_2$ are the number of another kind of the same object. For example, $n_1$ could be blue flags and $n_2$ could be red flags, both are flags but different kinds. The formula for this kind of permutation is

$$\frac{n!}{n_1! \cdot n_2! \cdot \ldots \cdot n_k}$$

8) A social security number is used to identify each resident of the United States uniquely. The number is of the form xxx-xx-xxxx, where each x is a digit from 0 to 9.

(A) How many social security numbers can be formed?

(B) What is the probability of correctly guessing the Social Security number of the mayor in your city?

$P(E) =$

9) The grade appeal process at a university requires that a jury be structured by selecting five individuals from a pool of 7 students and 11 faculty.

(A) What is the probability of selecting a jury of all students?

$P(E) =$

(B) what is the probability of selecting a jury of all faculty?

$P(E) =$

(C) What is the probability of selecting a jury of three students and two faculty?

$P(E) =$

## Binomial Probability Distribution

Binomial probability distribution is a specific type of discrete probability distributions. It describes probabilities for experiments in which there are two mutually exclusive outcomes (success and failure) in a fixed number of trials and there is the same probability of success. The formula for the binomial distribution is

$$P(x) = {}_nC_x \cdot p^x \cdot (1-p)^{n-x} \text{ ,}$$

where:

- n is the number of trials
- p is the probability of success
- 1-p or q is the probability of failure
- X is the number of successes in n number of trials $0 \leq x \leq n$

The criteria for the Binomial Probability Distribution:

1. Experiments are preformed a fixed number of times, called trials
2. The trials are independent of one another, meaning one trial occurring does not affect another trial occurring
3. There are two mutually exclusive outcome: success or failure
4. The probability for success is the same for each trial

The mean and standard deviation for the binomial probability distribution can be calculated (often not given) with the formulas

$$\mu_x = n \cdot p \quad and \quad \sigma_x = \sqrt{n \cdot p \cdot (1-p)}$$

10) According to flightstats.com, American Airlines flights from Dallas to Chicago are on time 74% of the time. Suppose 100 flights are randomly selected.

(A) What is the probability that exactly 80 flights are on time?

$$P(E) =$$

(B) What is the probability that less than 80 flights are on time?

$$P(E) =$$

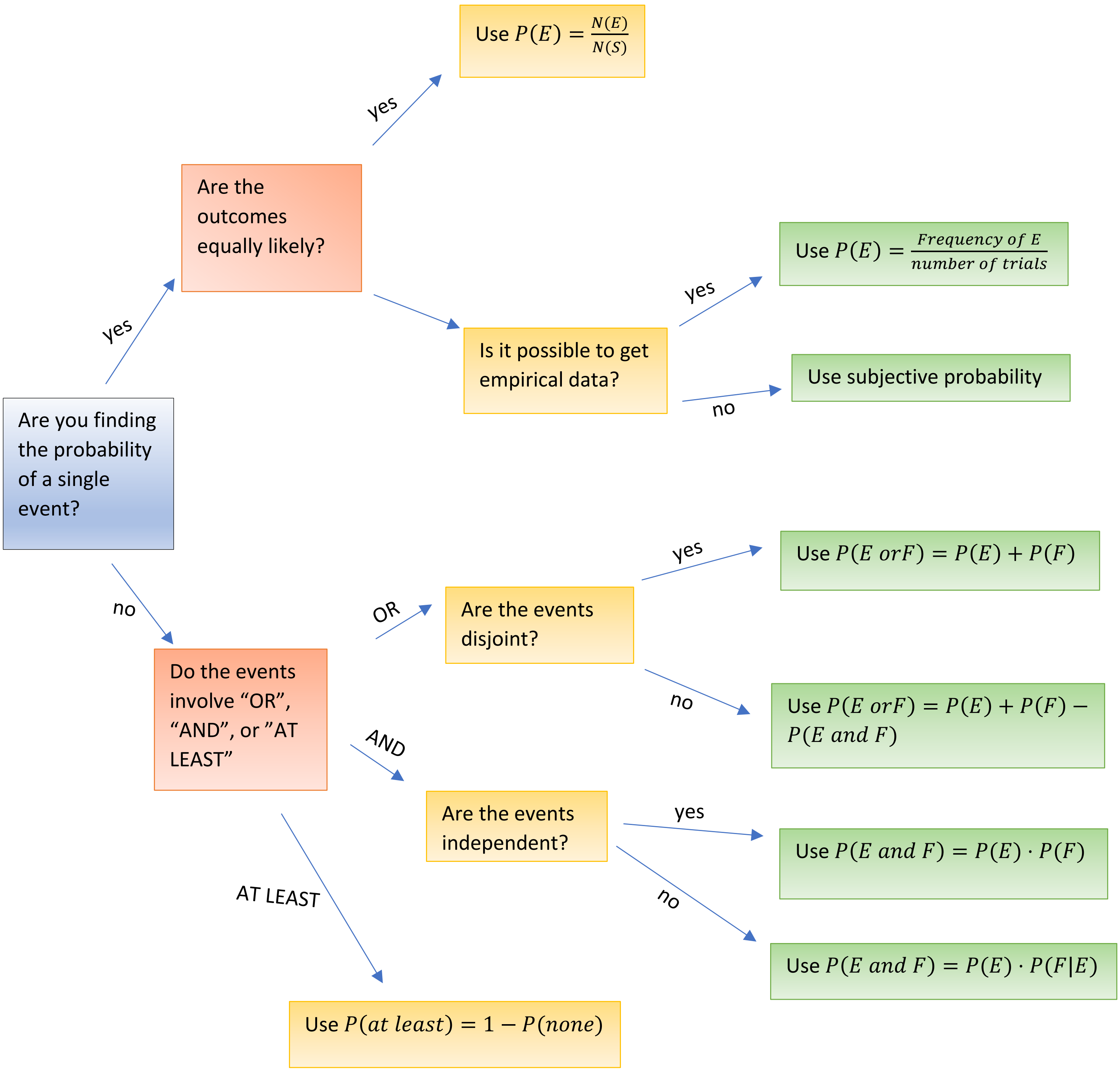
(C) What is the probability that 80 or more flights are on time?

$$P(E) =$$

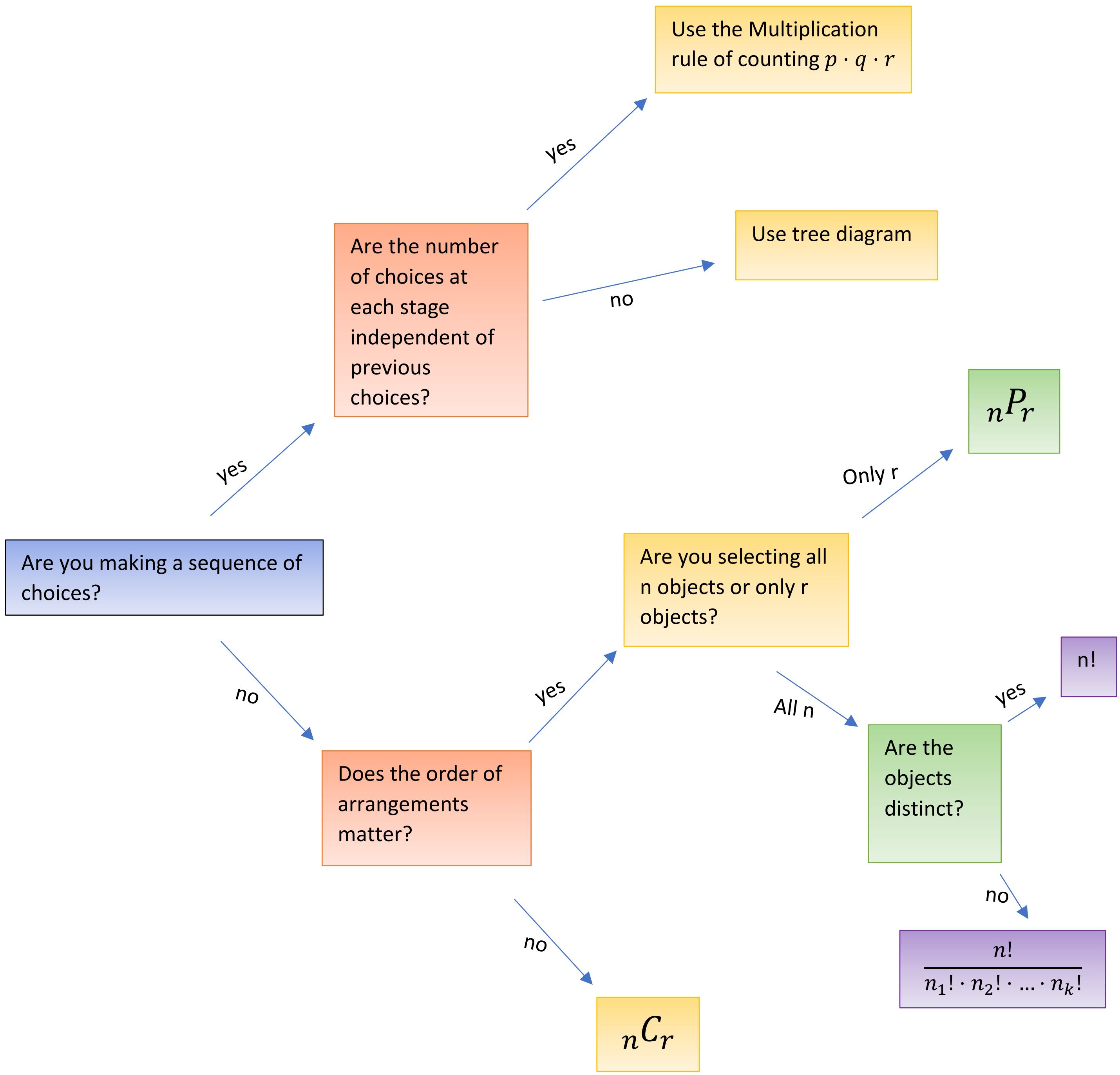
(D) Compute the mean and standard deviation of the random variable X, the number of flights of on-time flights in 100 trials of the probability experiment.

**Which method do I use?**

Determining the appropriate probability rule to use



Determining the appropriate counting technique to use



# Answer Key

1) (A) 0.35 (B) 0.20

2) 0.3154

3) (A) 0.55 (B) 0.80

4) (A) 0.1124 (B) 0.9178

5) (A) 0.1225 (B) 0.0909 (C) 0.0909 (D) 0.1818

6) (A) 0.2125 (B) 0.28 (C) Yes

7) (A) 0.3193 (B) 0.0925 (C) 0.4286

8) (A) $10^9$ (B) $\frac{1}{10^9}$

9) (A) 0.0025 (B) 0.0539 (C) 0.2247

10) (A) 0.0369 (B) 0.9342(C) 0.0658 (D) $\mu_x = 74$ and $\sigma_x = 4.3863$