COLLEGE OF CHARLESTON

Week 5

Math 104-03: Elementary Statistics

5.1 Discrete Random Variables

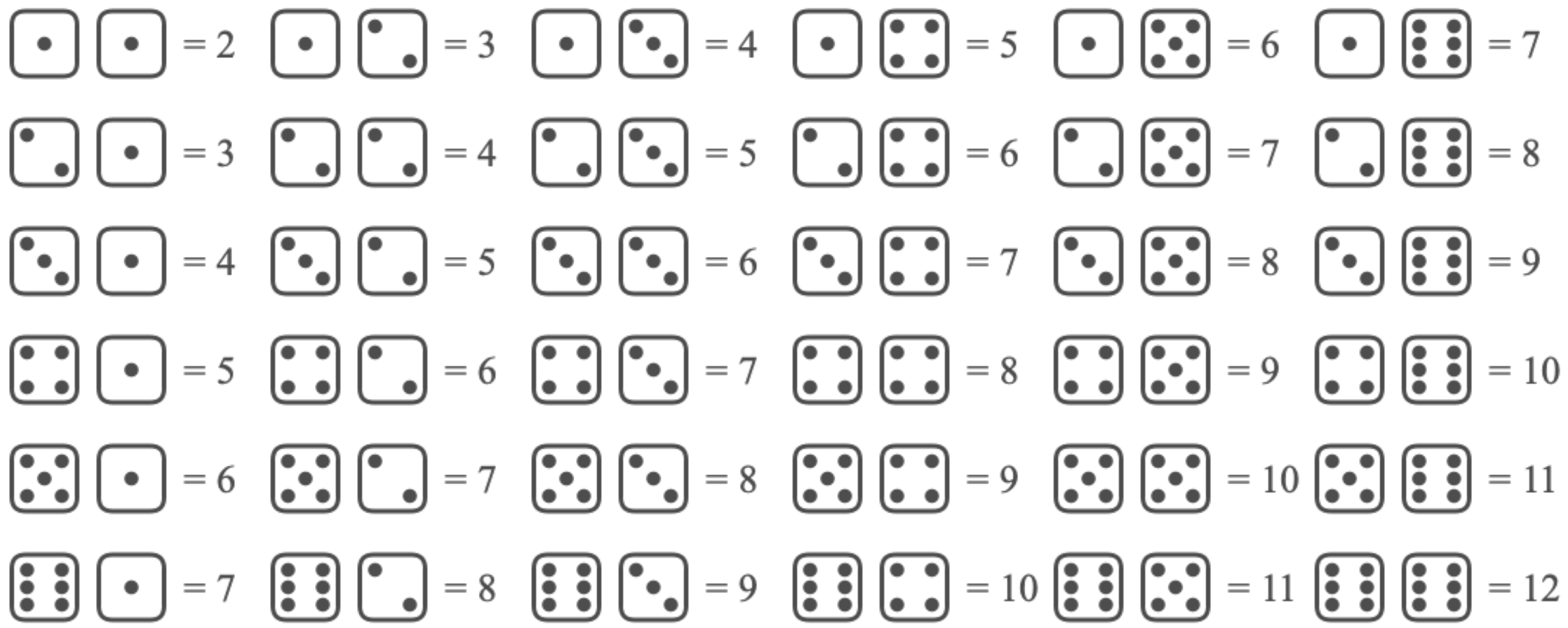
Random Variables

- A **random variable** is a numerical quantity whose value is determined by the outcome of an experiment.
- Why are we interested in random outcomes which are numerical values?
  - Example 1: In a large bag of coins 15% are pennies, 20% are nickels, 25% are dimes, and the rest are quarters. Suppose one coin is randomly selected and let X represent the value of the coin.
  - Example 2: From a forrest, select an tree at random and let X represent the height of the tree.
- A **discrete random variable** will have either finite number of possible values or a countable number
- A **continuous random variable** will have a whole interval of possible value

Probability Distributions for Random Variables

- A table or formula that gives the probabilities for every value of the random variable X is called a **probability distribution**.
  - $0 \leq \mathbb{P}(X = x) \leq 1$
  - $\sum \mathbb{P}(X = x) = 1$
- Example: we create a discrete probability distribution for the sum of two rolled dice

X	P(X=x)
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36



GAME OVERVIEW

1 The island of Catan lies before you. The isle consists of 19 terrain tiles surrounded by ocean. Your goal is to settle on Catan, and expand your territory until it becomes the largest and most glorious in Catan.

2 There are five productive terrain types and one desert on Catan. Each terrain type produces a different type of resource (The desert produces nothing). Each resource you receive is represented by a card. Here's what each terrain produces:

3 You begin the game with 2 settlements and 2 roads. Each settlement is worth 1 victory point. You therefore start the game with 2 victory points! The first player to acquire 10 victory points on their turn wins the game.

4 To gain more victory points, you must build new roads and settlements and upgrade your settlements to cities. Each city is worth 2 victory points. To build or upgrade, you need to acquire resources.

5 *How do you acquire resources? It's simple.* Each turn, you roll 2 dice to determine which terrain hexes produce resources. Each terrain hex is marked with a round number token. If, for example, a "10" is rolled, all terrain hexes with a "10" number token produce resources—in the illustration on the right, those terrain hexes are a mountains hex (ore) and a hills hex (brick).

6 You only collect resources if you own a settlement or city bordering these terrain hexes. In the illustration, the red settlement [A] borders the "10" mountains and orange settlement [B] borders the "10" hills. If a "10" is rolled, the red player receives 1 ore card and the orange player receives 1 brick card.

7 Since the settlements and cities usually border on 2-3 terrain types, they can "harvest" up to 3 different resources based on the dice roll. Here, the white settlement [C] borders on forest, mountains, and pasture. A settlement at [D] would only harvest the production from 2 terrain hexes (hills and mountains). Finally, a settlement at [E] would only harvest the production from 1 terrain hex (pasture). However [E] is also at a wool harbor.

8 Since it's impossible for you to have settlements adjacent to all terrain hexes and number tokens, you may receive certain resources only at rare intervals—or never. This is tough, because building requires specific resource combinations.

9 For this reason, you can trade with other players. Make them an offer! A successful trade might yield you a big build!

10 You can only build a new settlement on an unoccupied intersection if you have a road leading to that intersection and the nearest settlement is at least two intersections away.

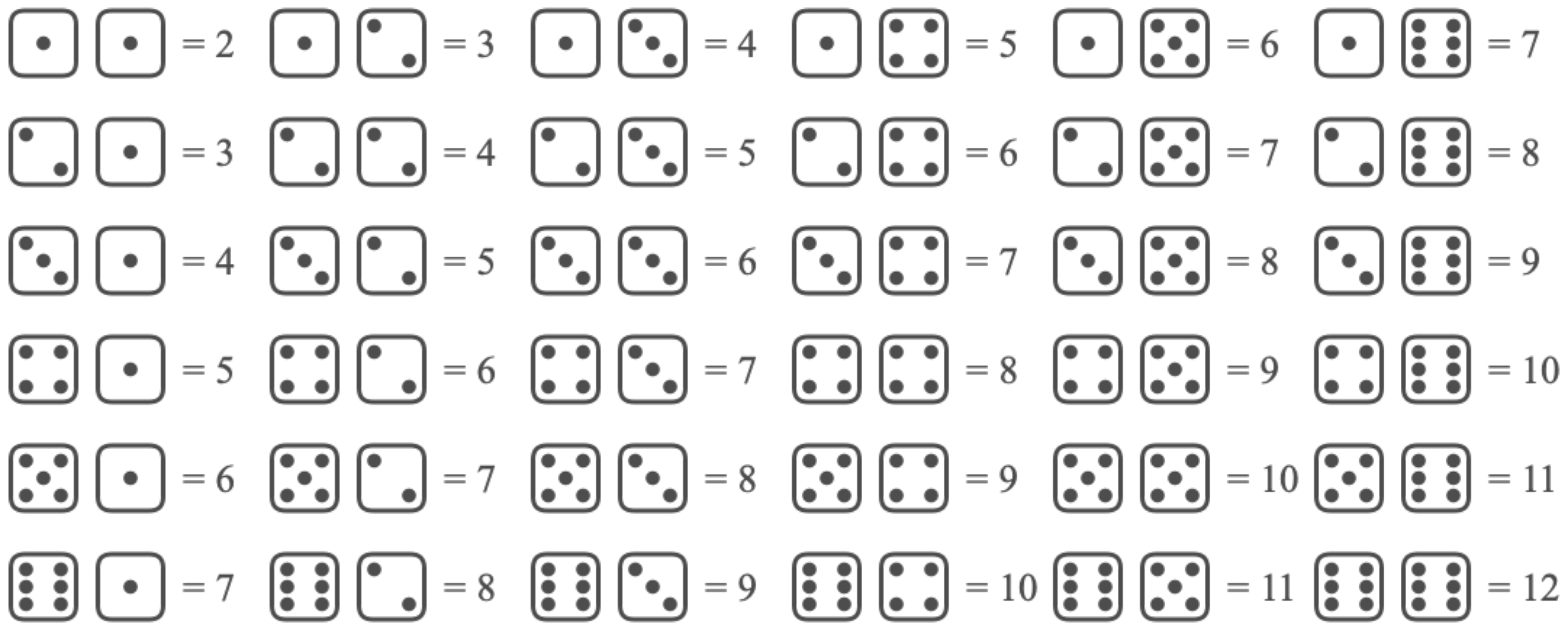
11 Carefully consider where you build settlements. The numbers on the round tokens are depicted in varying sizes. They also have dots (pips) below the numbers. The taller the depicted number, and the more pips it has, the more likely that number is to be rolled. The red numbers 6 and 8 are the tallest numbers with the most pips; they are likely to be rolled most frequently.

Bottom line: The more frequently a number is rolled, the more often the hexes with those numbers produce resources. You should consider settling on hexes that have good potential for production (*i.e. 6 and 8 vs. 2 and 12*). However, these same high-producing hexes are often the primary target for the robber.

Expectation

- The **expected value** for a discrete random variable X is equal to the mean of the probability distribution of X and is given by
  - $\mathbb{E}(X) = \mu = \sum x_i \cdot \mathbb{P}(X = x_i)$
- Example: what is the expected value of X = sum of two 6-sided dice? Answer: $\mathbb{E}(X) = 7$

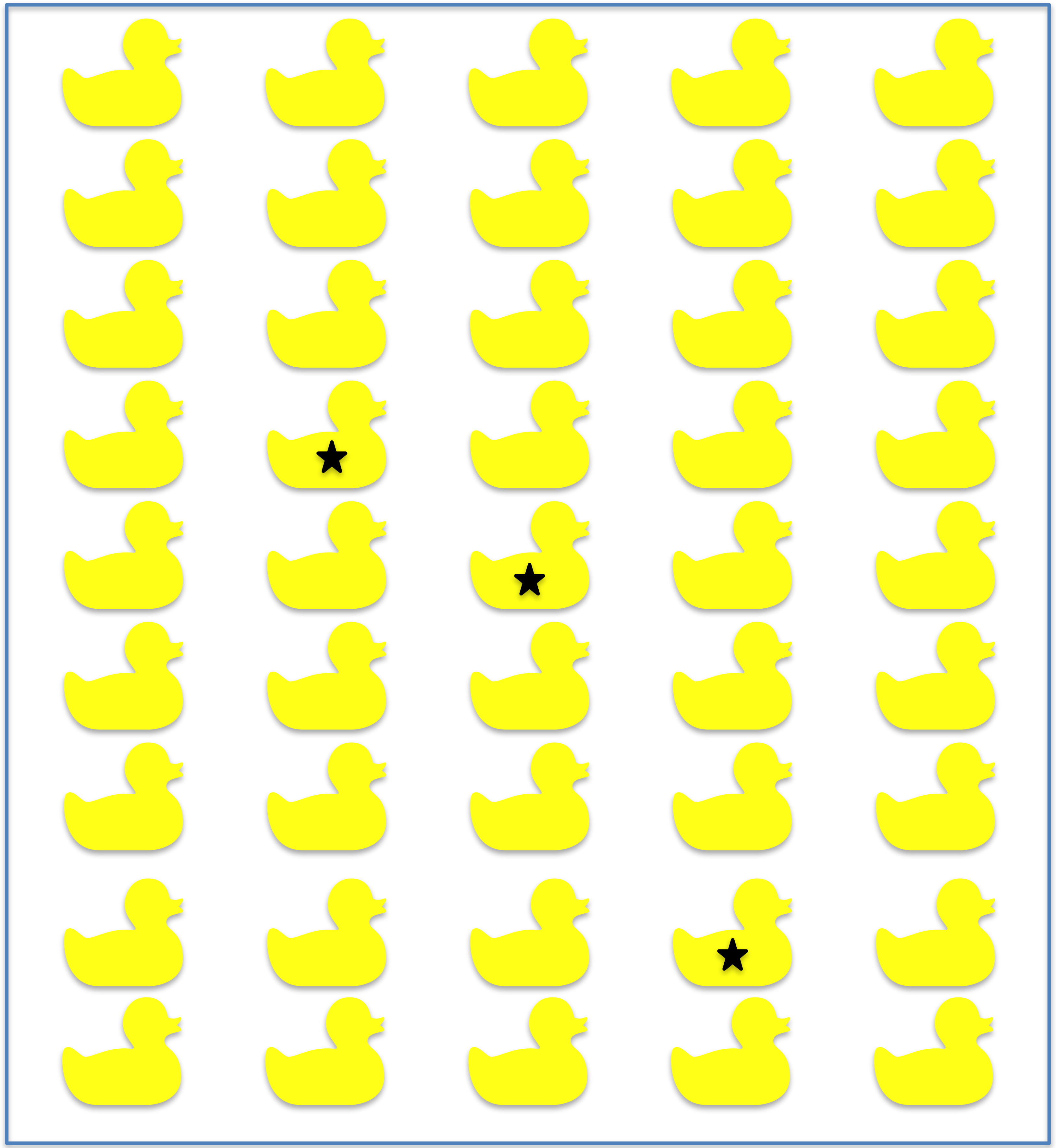
x	P(X=x)	x P(X=x)
2	1/36	2/36
3	2/36	6/36
4	3/36	12/36
5	4/36	20/36
6	5/36	30/36
7	6/36	42/36
8	5/36	40/36
9	4/36	36/36
10	3/36	30/36
11	2/36	22/36
12	1/36	12/36
		7



Expectation

- The **expected value** for a discrete random variable X is equal to the mean of the probability distribution of X and is given by
  - $\mathbb{E}(X) = \mu = \sum x_i \cdot \mathbb{P}(X = x_i)$
- Example: Game at state fair with rubber ducks floating in a pond. For \$1, you get two chances to win a small stuffed animal worth \$5. If you choose two ducks both with stars on the bottom, you win the extra large stuffed animal worth \$15. After you first pick, the rubber duck is placed back in the water for your second pick. If there are 50 ducks and 3 with stars on them, what is expected value of this carnival game?

  Answer: $-\$0.38$



Variance

- The **variance** for a discrete random variable X is given by

$$\sigma^2 = \sum \left(x_i^2 \cdot \mathbb{P}(X = x_i)\right) - \mu^2$$
$$= \sum \left(x_i - \mu)^2 \cdot \mathbb{P}(X = x_i)\right)$$

- What does this mean?

$$\sigma^2 = \mathbb{E}(X^2) - \mu^2$$
$$= \mathbb{E}\left(((X - \mathbb{E}(X))^2 \right)$$

- Standard deviation $\sigma = \sqrt{\sigma^2}$
- Example: what is the variance for the random variable X which equals the sum of two dice?

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \mu^2 = \left(2^2 \cdot \mathbb{P}(X = 2) + 3^2 \cdot \mathbb{P}(X = 3) + \cdots + 12^2 \cdot \mathbb{P}(X = 12)\right) - 7^2 \approx 5.83$$

Answer: 5.83

Example: Coin Toss Game

- Toss three coins.
  - If all the same (HHH or TTT), win \$2
  - Otherwise, win -\$1
- X = money you win
  - Calculate $\mathbb{E}(X)$

    $$E(X) = 2 \cdot \mathbb{P}(\text{HHH or TTT}) + -1 \cdot \mathbb{P}(\text{not HHH or TTT}) = 2 \cdot \frac{1}{4} + -1 \cdot \frac{3}{4} = -\ \$0.25$$

  - Calculate $\mathrm{Var}(X)$

    $$\mathrm{Var}(X) = (2 + 0.25)^2 \cdot \frac{1}{4} + (-1 + 0.25)^2 \cdot \frac{3}{4} = \$1.12$$

Case Study: Poker

- You're playing Texas Hold'em, and you're heads-up (just two players). You have a pocket pair 10♦ 10♣, and your opponent has A♠ K♠ The board so far shows 8♥ 6♦ 2♣, so you have an overpair, and your opponent currently has just two overcards with a potential to hit an Ace, King, or backdoor straight or flush. You're trying to decide whether to **go all-in**, which would cost you $100, with the pot already at $150.
- At this stage, your pocket tens are ahead. But your opponent still has outs (cards that could make them a better hand), including:
  - 3 Aces
  - 3 Kings
  - Z Some chance to backdoor a flush or straight.
- The chances of winning with your pocket 10s after the flop (with two cards still to come) is roughly 75%, while your opponent has a 25% chance of improving to a better hand.
- If you win, the pot will be $250 ($150 from the pot + $100 from your opponent). The probability of this is 75%, so the expected value of winning is:
  - $\mathbb{E}(\text{winning}) = 0.75 \times 250 + 0.25 \times (-100) = \$162.50$
- In this scenario, the expected value of going all-in is $162.50. Since this value is positive and greater than your $100 investment, it indicates that going all-in is a profitable decision in the long run.

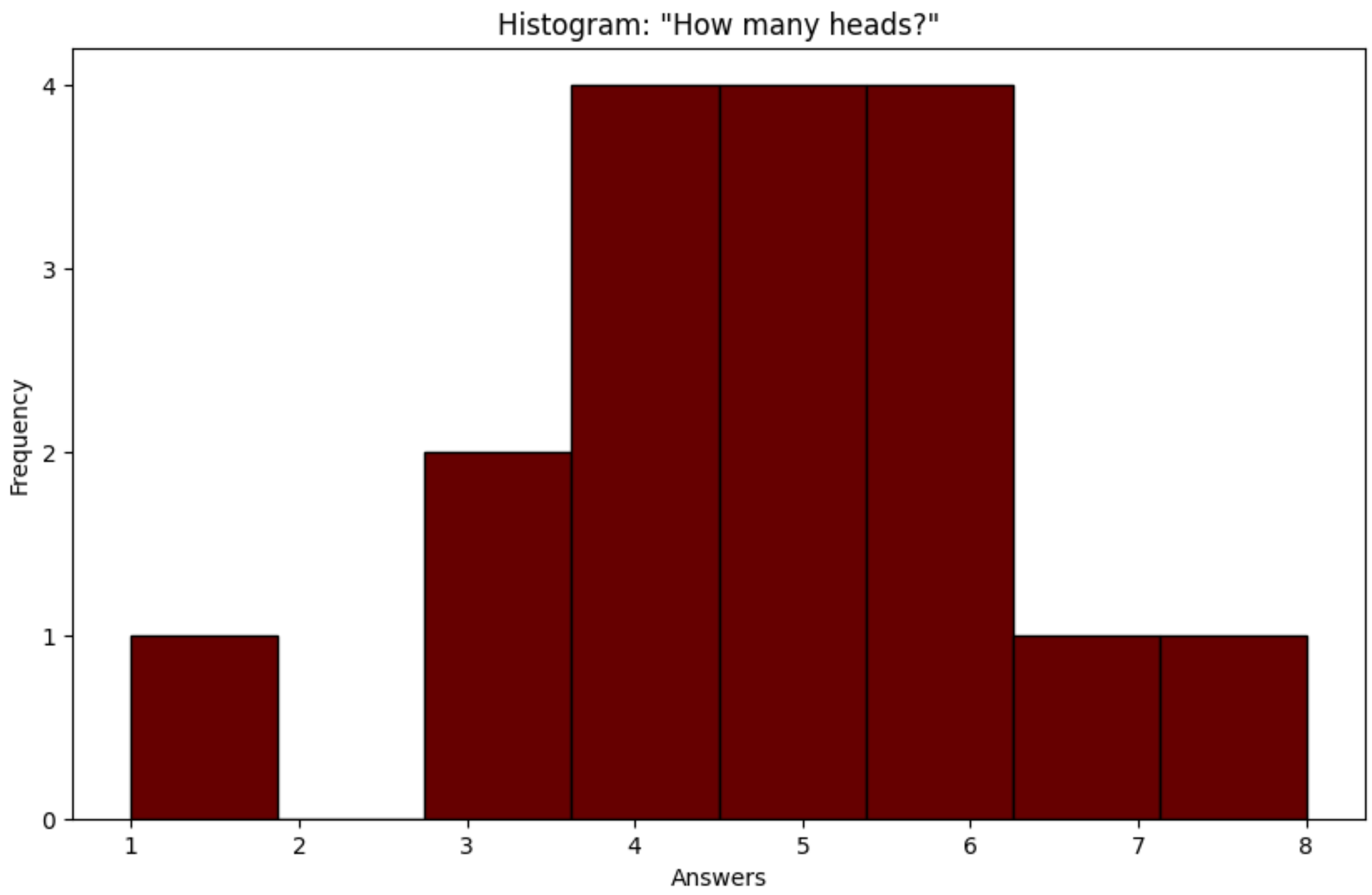
5.2 Binomial Distribution

Coin Toss Experiment

- With a partner, toss a coin a total of n = 10 times and record the number of successes
  - One person flips the coin
  - Other writes down number of heads
- Every do the experiment at least once then submit each result to elementary-stats.com

- Questions:
  - Describe the distribution of the number of successes.
  - What would the distribution look like if the coin was "bent" i.e. "biased."

Coin Toss Experiment: Results



- $\bar{x} = 4.92$
- $s = 1.62$

Binomial Distribution

- **Binomial distribution** is a special discrete probability distribution for independent trials.
- Random variable measuring the number of **successes** out of a total of n trials
  - Each trial is independent of the others
  - Each trial is either a success or failure
  - Probability of success is p; probability of failure $1 - p$
  - Random variable X counts the number of successes in n trials
- Mean and variance of the binomial distribution:
  - Mean: $\mu = np$. Let's prove it!
  - Variance: $\sigma^2 = np(1 - p)$. Let's prove it!
- We will use two facts: $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ and $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$

Coin Toss Experiment

- With a partner, toss a coin a total of n = 10 times and record the number of successes
  - One person flips the coin
  - Other writes down number of heads
- Every do the experiment at least once then submit each result to elementary-stats.com

- Questions:
  - Describe the distribution of the number of successes.
  - What would the distribution look like if the coin was "bent" i.e. "biased."
  - How does the sample mean $\bar{x}$ and sample standard distribution s compare to the
    - mean $\mu = n \cdot p = 10 \cdot 0.5 = 5$?
    - standard deviation $\sigma = \sqrt{np(1-p)} = \sqrt{10 \times 0.5 \times 0.5} \approx 1.581$?

4.4 Combinations and Permutations

Combinations

- A **combination** is a selection of objects with a group without regard to their order
- A **permutation** is a selection of objects from a group where the arrangement is specific
- When order is not important, the number of combinations of r objects from a total of n is the following: $\binom{n}{r} = \frac{n!}{r!(n-r)!}$ also denoted $n^{C}r$
- The symbol $n! = n \cdot (n-1) \cdot (n-2) \cdot \cdots \cdot 2 \cdot 1$ is the **factorial**
- Example: how many hands can I get in 5-card poker?
  Answer: $\binom{52}{5} = \frac{52!}{47!5!} = \frac{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 2{,}598{,}960$
- Example: in a tournament with 8 players how many possible matchups are there?
  Answer: $\binom{8}{2} = \frac{8!}{(8-2)!2!} = \frac{8 \times 7}{2 \times 1} = 28$

5.2 Binomial Distribution

Binomial Distribution

- Formula for binomial distribution
  - $\mathbb{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$
  - p, probability of success
  - n, number of trails
  - x, number of successes
- Example: what's the probability of getting exactly six heads in 10 coin tosses?

  Apply the formula: $\mathbb{P}(X = 6) = \binom{10}{6} 0.5^6 (1-0.5)^{10-6} = 210 \cdot 0.5^{10} \approx 0.2050$

Binomial Distribution

- Formula for binomial distribution
  - $\mathbb{P}(X = x) = \binom{n}{x} p^x (1 - p)^{1-x}$
  - p, probability of success
  - n, number of trails
  - x, number of successes
- Example: You are a QA expert at a factory which estimates 10% of all the batteries are defective. If you take a random sample of fifteen batteries, what is the probability that exactly two are defective?

  Use the formula! $\mathbb{P}(X = 2) = \binom{15}{2} \cdot (0.1)^2 \cdot (0.9)^{15-2} \approx 0.2669$

Binomial Distribution

- Formula for binomial distribution
  - $\mathbb{P}(X = x) = \binom{n}{x} p^x (1-p)^{1-x}$
  - p, probability of success
  - n, number of trails
  - x, number of successes
- Also can calculate the probabilities $\mathbb{P}(X \leq x) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \cdots + \mathbb{P}(X = x)$
- Example: you are a QA expert at a factory which estimates 10% of all the batteries are defective. If you take a random sample of fifteen batteries, what is the probability that no more than two are defective? Answer: $\mathbb{P}(X \leq 2) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2)$ which is equal to $\binom{15}{0}(1-0.1)^{10} + \binom{15}{1}(0.1)^1(1-0.1)^{15-1} + \binom{15}{2}(0.1)^2(1-0.1)^{15-3} \approx 0.8159$