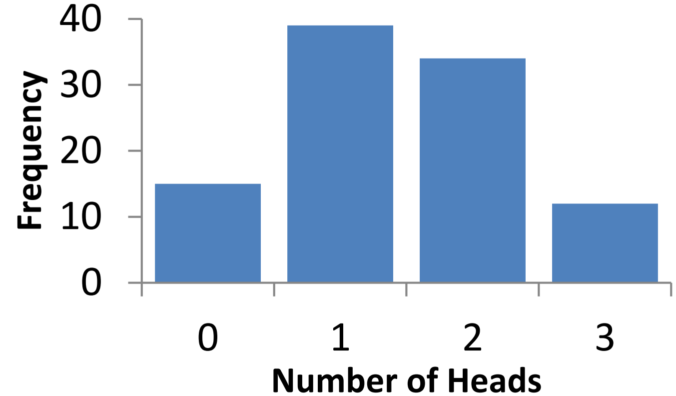
SOC 3811/5811: BASIC SOCIAL STATISTICS

Discrete Random Variables

Random Variables

Random Event
3 Coin Flips
(Repeated 100 Times)

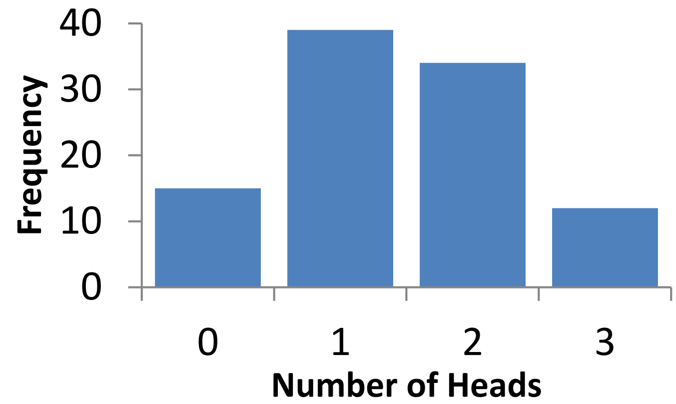
Observed Distribution
for a **Sample** of
100 Sets of 3 Coin Flips



Random Variables

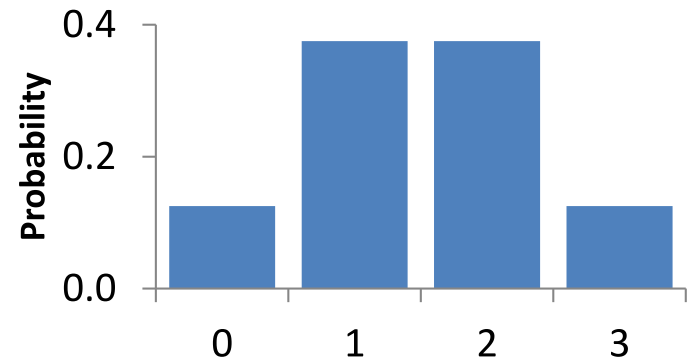
Random Event
3 Coin Flips
(Repeated 100 Times)

Observed Distribution
for a **Sample** of
100 Sets of 3 Coin Flips



Random Event
3 Coin Flips
(Repeated Infinitely)

Theoretical Distribution
for the **Population** of
All Possible 3 Coin Flips



Random Variables

Today

We <u>know</u> the actual probabilities associated with the random event that generates the theoretical distribution (e.g., coin flips)

We will learn to *describe* and *make practical use* of these theoretical distributions

Later (and in Life)

We <u>do not know</u> the actual probabilities associated with the random event (e.g., number of children per person, which candidate will win)

Our ability to *describe* and *make practical use* of theoretical distributions will allow us to <u>infer</u> those probabilities

Discrete Random Variables

For discrete random variable Y:

Y_1, Y_2, Y_3 , … , Y_k are all of the k possible values of Y

p_1, p_2, p_3, … , p_k are the probabilities of observing each value

The **expected value** of Y … the mean of Y that you would get if you repeated the random trial an infinite number of times … is:

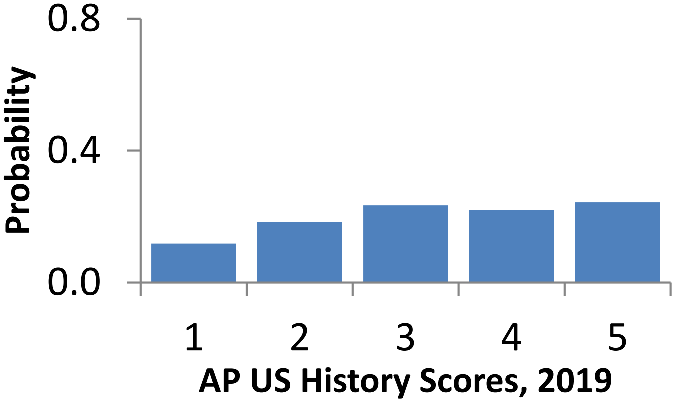
$$E(Y) = \mu_Y = \sum_{i=1}^{k} Y_i p(Y_i)$$

Discrete Random Variables

Example: The distribution of Advanced Placement (AP) United States History scores in 2019

The **probability distribution function** is:

k	1	2	3	4	5
p(Y=k)	0.118	0.184	0.234	0.220	0.243



$$E(Y) = \mu_Y = \sum_{i=1}^{k} Y_i p(Y_i)$$

$$E(Y) = (1)(0.118) + (2)(0.184) + (3)(0.234) + (4)(0.220) + (5)(0.243) = 3.283$$

Source: https://apstudents.collegeboard.org/about-ap-scores/score-distributions

Discrete Random Variables

For discrete random variable Y, the variance of Y … σ^2_Y … is

$$\sigma_Y^2 = \sum_{i=1}^{k} \left(Y_i - \mu_Y\right)^2 p(Y_i)$$

and the standard deviation of Y … σ_Y … is

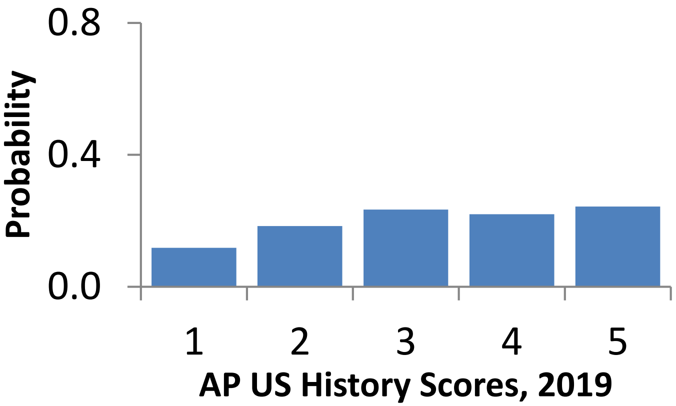
$$\sigma_Y = \sqrt{\sum_{i=1}^{k} \left(Y_i - \mu_Y\right)^2 p(Y_i)}$$

Discrete Random Variables

Example: The distribution of Advanced Placement (AP) United States History scores in 2019

The **probability distribution function** is:

k	1	2	3	4	5
p(Y=k)	0.118	0.184	0.234	0.220	0.243



$$\sigma_Y^2 = \sum_{i=1}^{k} \left(Y_i - \mu_Y\right)^2 p(Y_i)$$

$$\sigma^2_Y = (1-3.283)^2(0.118) + (2-3.283)^2(0.184) + (3-3.283)^2(0.234) + (4-3.283)^2(0.220) + (5-3.283)^2(0.243) = 1.766$$

$$\sigma_Y = \sqrt{\sum_{i=1}^{k} \left(Y_i - \mu_Y\right)^2 p(Y_i)} = \sqrt{1.766} = 1.329$$

Discrete Random Variables

For the discrete random variable Y = “2019 AP US History Scores” we found that the expected score was $E(Y)=\mu_Y=$ 3.283, with standard deviation $\sigma_Y=1.329$

If we sampled 2019 US History test takers and observed Y for each, we would expect the mean of Y would be 3.283

We wouldn’t always (or ever) see a score of exactly 3.283; σ_Y indicates how much the scores would vary from person to person

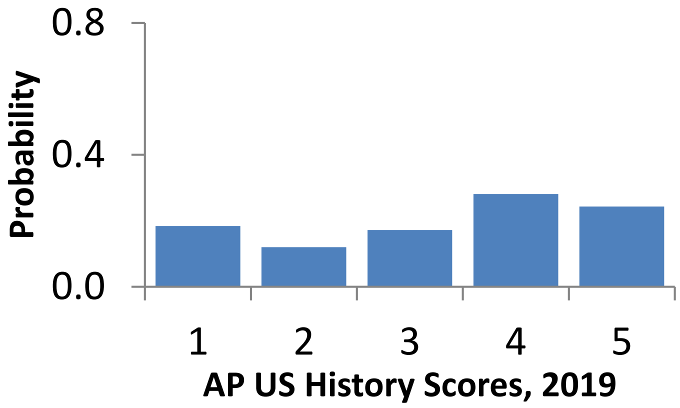
Discrete Random Variables

0.118 0.184 0.234 0.220 0.243

Example: The distribution of Advanced Placement (AP) Microeconomics scores in 2019

The **probability distribution function** is:

k	1	2	3	4	5
p(Y=k)	0.184	0.120	0.172	0.281	0.243



What is the expected value of random variable Y, "AP Microeconomics Scores in 2019"? How do you interpret that number?

What is the standard deviation of that random variable?

Source: https://apstudents.collegeboard.org/about-ap-scores/score-distributions

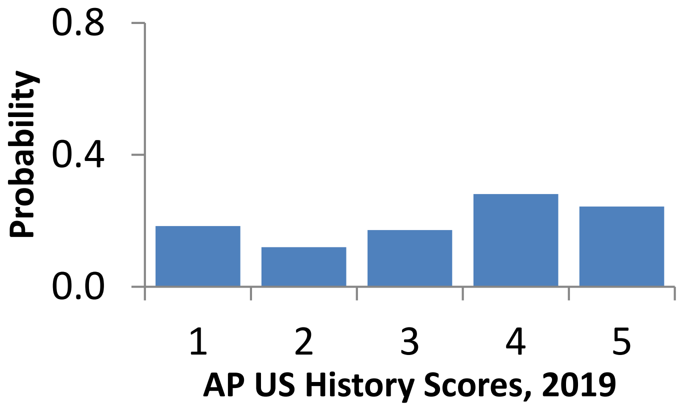
Discrete Random Variables

0.118 0.184 0.234 0.220 0.243

Example: The distribution of Advanced Placement (AP) Microeconomics scores in 2019

The **probability distribution function** is:

k	1	2	3	4	5
p(Y=k)	0.184	0.120	0.172	0.281	0.243



$$E(Y) = \mu_Y = \sum_{i=1}^{k} Y_i p(Y_i)$$

$$E(Y) = (1)(0.184) + (2)(0.120) + (3)(0.172) + (4)(0.281) + (5)(0.243) = 3.279$$

Source: https://apstudents.collegeboard.org/about-ap-scores/score-distributions

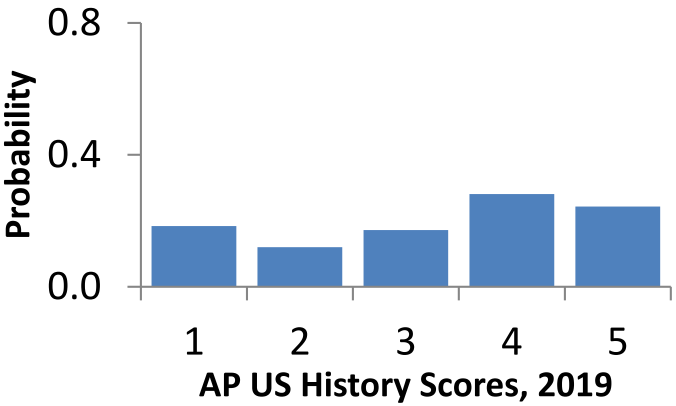
Discrete Random Variables

0.118 0.184 0.234 0.220 0.243

Example: The distribution of Advanced Placement (AP) Microeconomics scores in 2019

The **probability distribution function** is:

k	1	2	3	4	5
p(Y=k)	0.184	0.120	0.172	0.281	0.243



$$\sigma_Y^2 = \sum_{i=1}^{k} \left(Y_i - \mu_Y\right)^2 p(Y_i)$$

$$\sigma^2_Y = (1\text{-}3.279)^2(0.184) + (2\text{-}3.279)^2(0.120) + (3\text{-}3.279)^2(0.172) + (4\text{-}3.281)^2(0.220) + (5\text{-}3.279)^2(0.243) = 2.031$$

$$\sigma_Y = \sqrt{\sum_{i=1}^{k} \left(Y_i - \mu_Y\right)^2 p(Y_i)} = \sqrt{2.031} = 1.425$$

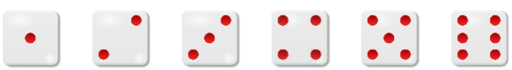
Source: https://apstudents.collegeboard.org/about-ap-scores/score-distributions

Binomial Random Variables

A special kind of discrete random variable … called a binomial random variable … is one in which the random event generating it has only two possible outcomes

Discrete Random Variable

Random event has a finite number of distinct values; the random variable is a count of how often each value occurs.

Example: 

Binomial Random Variable

Random event has exactly two possible values ("success" and "failure"); the random variable is a count of how often "success" occurs over *n* repeated trials. *Example*: 

Binomial Random Variables

For example: Your probability of winning *anything* when you play Powerball is 0.0402. (That is, you win *nothing* 95.98% of the time)

Each play is a binomial random “trial” in which winning is a “success.” If you play 3 times, you might observe 0, 1, 2, or 3 successes.

Binomial Random Variables

For binomial random variable Y, the probability that Y = k (where k ranges from 0 to n) is:

$$P(Y=k)=\frac{n!}{k!(n-k)!}p^{k}(1-p)^{n-k}$$

where n is the number of binomial events, k is some possible value of Y, and the probability of “success” in any one event is p

For example, what is the probability of getting k=2 heads out of n=3 flips of a fair coin where P(Heads)=0.5?

$$P(Y=2) = \frac{3!}{2!(3-2)!}0.5^{2}(1-0.5)^{3-2} = \frac{6}{2 \text{ x } 1}(0.25)(0.5) = 0.375$$

Binomial Random Variables

For example, what is the probability of winning anything in the Powerball all k=3 times you play if you play n=3 times and the probability of winning anything each time you play is 0.0402?

$$= \frac{3!}{3!(3-3)!} 0.0.0402^3 (1 - 0.0402)^{3-3}$$

$$= \frac{6}{6 \text{ x } 1} (0.000065)(1) \ = 0.000065$$

Binomial Random Variables

For binomial random variable Y based on n trials with probability of "success" on any given trial equal to p, the expected value of Y (or μ_Y) equals

$$E(Y) = \mu_Y = np$$

the variance σ^2_Y equals

$$\sigma_Y^2 = np(1-p)$$

and the standard deviation σ_Y equals

$$\sigma_Y = \sqrt{\sigma_Y^2} = \sqrt{np(1-p)}$$

Binomial Random Variables

For the Powerball example the expected value of Y (or μ_Y) equals

$$E(Y) = \mu_Y = np = (3)(0.0402) = 0.1206$$

the variance σ^2_Y equals

$$\sigma_Y^2 = np(1-p) = (3)(0.0402)(0.9598) = 0.1158$$

and the standard deviation σ_Y equals

$$\sigma_Y = \sqrt{\sigma_Y^2} = \sqrt{np(1-p)} = \sqrt{0.1158} = 0.3402$$

WORKSHEET

The probability that your personal income taxes get audited by the IRS is 0.04 in any given tax year.

a) Across 5 tax years, what is the probability that you get audited exactly twice?

b) What is the expected value of the number of times you will be audited over those 5 years?

c) What is the standard deviation of the number of times you will be audited over those five years?

WORKSHEET

The probability that your personal income taxes get audited by the IRS is 0.04 in any given tax year.

Across 5 tax years, what is the probability that you get audited exactly twice?

$$P(Y = k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

$$= \frac{5!}{2!(5-2)!} 0.04^2 (1 - 0.04)^{5-2} = (10)(0.0016)(0.884736)$$

$$= 0.014$$

WORKSHEET

The probability that your personal income taxes get audited by the IRS is 0.04 in any given tax year.

What is the expected value of the number of times you will be audited over those 5 years?

$$E(Y) = \mu_Y = np = (5)(0.04) = 0.2$$

WORKSHEET

The probability that your personal income taxes get audited by the IRS is 0.04 in any given tax year.

What is the standard deviation of the number of times you will be audited over those five years?

$$\sigma_Y = \sqrt{\sigma_Y^2} = \sqrt{np(1-p)} = \sqrt{(5)(0.04)(0.96)} = 0.438$$

Binomial Random Variables

Because a binomial random variable is just a special case of discrete random variables, the formula used to compute the mean and variance of a discrete random variable…

$$E(Y) = \mu_Y = \sum_{i=1}^{k} Y_i p(Y_i)$$

$$\sigma_Y^2 = \sum_{i=1}^{k} (Y_i - \mu_Y)^2 p(Y_i)$$

…will also work for binomial random variables (if you are inclined to do extra math)