Section 7.1/7.2
Discrete and Continuous Random Variables

AP Statistics
NPHS
Mrs. Skaff

Do you remember?--p.141, 143

- The annual rate of return on stock indexes is approximately Normal. Since 1945, the Standard & Poors index has had a mean yearly return of 12%, with a standard deviation of 16.5%. **In what proportion of years does the index gain 25% or more?**

- The annual rate of return on stock indexes is approximately Normal. Since 1945, the Standard & Poors index has had a mean yearly return of 12%, with a standard deviation of 16.5%. **In what proportion of years does the index gain between 15% and 22%?**

Random Variables

- A **random variable** is a variable whose value is a numerical outcome of a random phenomenon.
- For example: Flip four coins and let *X* represent the number of heads. *X* is a random variable.
- We usually use capital letters to denotes random variables.

Random Variables

- A **random variable** is a variable whose value is a numerical outcome of a random phenomenon.
- For example: Flip four coins and let *X* represent the number of heads. *X* is a random variable.
- **X = number of heads when flipping four coins.**
- S = {0, 1, 2, 3, 4}

Discrete Probability Distribution Table

◦ A **discrete random variable**, **X,** has a <u>countable number of possible values.</u>

Value of X:	x_1	x_2	x_3	…	x_n
Probability:	p_1	p_2	p_3	…	p_n

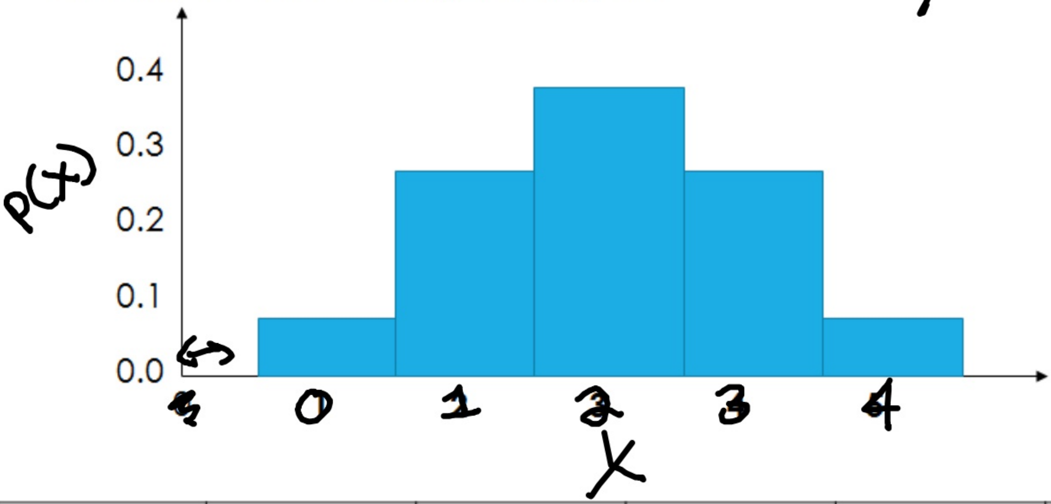
◦ The **probability distribution** of discrete random variable, X, lists the values and their probabilities.

Probability Distribution Table: X = Number of Heads Flipping 4 Coins

	TTTT	TTTH TTHT THTT HTTT	TTHH THTH HTTH HTHT THHT HHTT	THHH HTHH HHTH HHHT	HHHH
X	0	1	2	3	4
P(X)	1/16	4/16	6/16	4/16	4 16

Discrete Probability Distributions

- Can also be shown using a histogram



X	1	2	3	4	5
P(X)	.0625	.25	.375	.25	.0625

What is…

- The probability of at most 2 heads?

X	0	1	2	3	4
P(X)	0.0625	0.25	0.375	0.25	0.0625

Example: Maturation of College Students

In an article in the journal *Developmental Psychology* (March 1986), a probability distribution for the age X (in years) when male college students began to shave regularly is shown:

X	11	12	13	14	15	16	17	18	19	≥20
P(X)	0.013	0	0.027	0.067	0.213	0.267	0.240	0.093	0.067	0.013

Is this a valid probability distribution? How do you know?

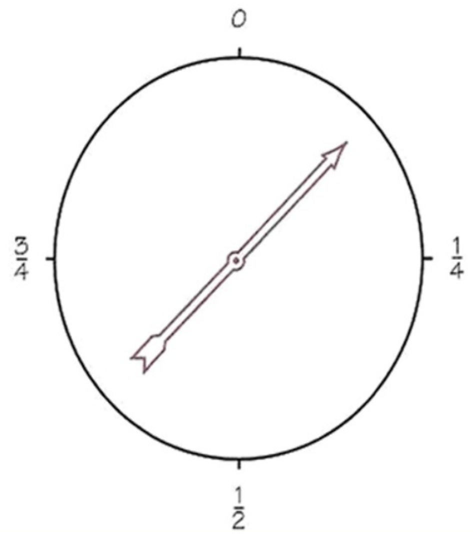
What is the random variable of interest?

Is the random variable discrete?

Continuous Random Variable

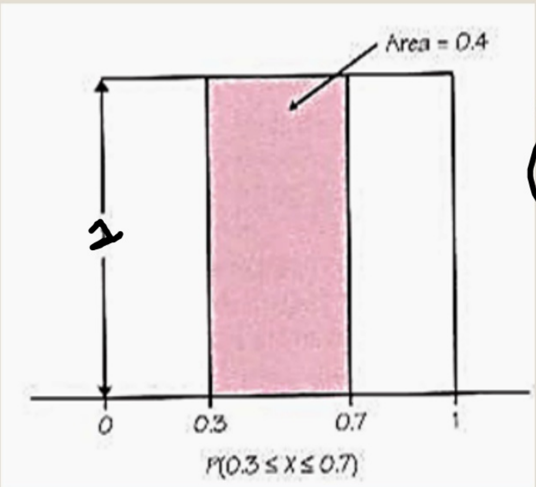
- A **continuous random variable** *X* takes **all** values in an interval of numbers.

$S = \{\text{all numbers } x \text{ such that } 0 \leq x \leq 1\}$



Distribution of Continuous Random Variable

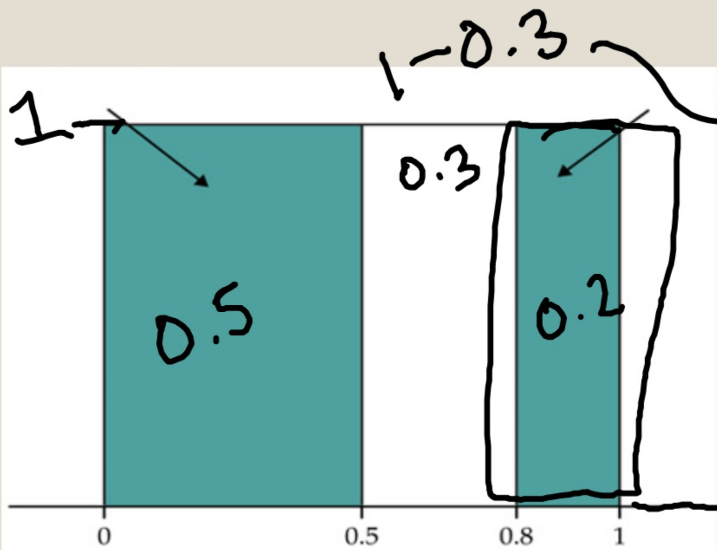
- The **probability distribution** of X is described by a density curve.
- The probability of any event is the **area** under the density curve and above the values of X that make up that event.



The probability that X = a particular value is 0

$P(X = 0.7) = 0$

Distribution of a Continuous Random Variable



$P(X \leq 0.5 \text{ or } X > 0.8)$

(b)

P(X≤0.5 or X>0.8)

= 0.7

P(X<0.5 or X≥0.8)

= 0.7

P(0.8<X<1.5)

= 0.2

Normal distributions as probability distributions

- Suppose X has N(μ,σ) then we can use our tools to calculate **probabilities.**
- One tool we may need is our formula for standardizing variables:

$$z = \frac{X - \mu}{\sigma}$$

Cheating in School

- A sample survey puts this question to an SRS of 400 undergraduates: "You witness two students cheating on a quiz. Do you go to the professor?" Suppose *if* we could ask *all* undergraduates, **12%** would answer "Yes"

- We will learn in Chapter 9 that the proportion **p=0.12** is a ***population parameter*** and that the proportion $\hat{p}$ of the *sample* who answer "yes" is a ***statistic*** used to estimate p.
- We will see in Chapter 9 that $\hat{p}$ is a random variable that has approximately the **N(0.12, 0.016) distribution.**
  - The mean 0.12 of the distribution is the same as the population parameter. The standard deviation is controlled mainly by the sample size.

Continuous Random Variable

- $\hat{p}$ (proportion of the sample who answered yes) is a random variable that has approximately the **N(0.12, 0.016) distribution.**
- What is the probability that the poll result differs from the truth about the population by more than two percentage points?

$P(\hat{p} < 0.10 \text{ or } \hat{p} > 0.14)$

$1 - P(0.10 < \hat{p} < 0.14)$

$1 - \text{ncdf}(.1, .14, .12, .016)$

$\sim 21\%$

Check Point

- $\hat{p}$ (proportion of the sample who answered drugs) is a random variable that has approximately the **N(0.12, 0.016) distribution.**
- What is the probability that the poll result is greater than 13%?
- What is the probability that the poll result is less than 10%?

Random Variables: MEAN

- The Michigan Daily Game you pick a 3 digit number and win $500 if your number matches the number drawn.
- There are 1000 three-digit numbers, so you have a probability of 1/1000 of winning
- Taking X to be the amount of money your ticket pays you, the probability distribution is:

Payoff X:	$0	$500
Probability:	0.999	0.001

Random Variables: MEAN

- The Michigan Daily Game you pick a 3 digit number and win $500 if your number matches the number drawn.
- There are 1000 three-digit numbers, so you have a probability of 1/1000 of winning
- Taking X to be the amount of money your ticket pays you, the probability distribution is:

Payoff X:	$0	$500
Probability:	0.999	0.001

We want to know your average payoff if you were to buy many tickets.

Why can't we just find the average of the two outcomes (0+500/2) = $250?

Random Variables: Mean

So…what is the average winnings? (Expected long-run payoff)

expected = μ

Payoff X:	\$0	\$500
Probability:	0.999	0.001

$$\mu_x = 0(0.999) + 500(0.001)$$

$$\mu_x = \$0.50$$

Random Variables: Mean

$$\mu_X = p_1x_1 + p_2x_2 + p_3x_3 + \cdots + p_nx_n$$

$$\mu_X = \sum p_i x_i$$

Random Variables: Example

- The Michigan Daily Game you pick a 3 digit number and win \$500 if your number matches the number drawn.
- You have to pay \$1 to play
- What is the average **PROFIT**? = Y
- Mean = Expected Value

Payoff X:	0	500
Probability:	.999	0.001

Y	-1	499
Prob.	.999	.001

$\mu_Y = 0.999(-1) + 0.001(499)$

$\mu_Y = -\$0.50$

payoff = \$0.50

profit = \$0.50 − 1 = −\$0.50

Random Variables: **Variance**

(the average of the squared deviation from the mean)

$$\sigma_X^2 = p_1\left(x_1-\mu_x\right)^2 + p_2\left(x_2-\mu_x\right)^2 + \cdots + p_n\left(x_n-\mu_x\right)^2$$

$$\sigma_X^2 = \sum p_i\left(x_i-\mu_x\right)^2$$

The **standard deviation** σ of X is the square root of the variance

$$\sigma_X = \sqrt{\sigma_X^2}$$

Random Variables: Example

- The Michigan Daily Game you pick a 3 digit number and win $500 if your number matches the number drawn.
- The probability of winning is .001
- **What is the variance and standard deviation of X?**

$$\sigma_X^2 = p_1(x_1-\mu_X)^2 + p_2(x_2-\mu_X)^2 + \cdots + p_n(x_n-\mu$$

$$\sigma_X^2 = \sum p_i(x_i-\mu_X)^2$$

$$\sigma_X^2 = 0.999(0-0.5)^2 + 0.001(500-0.5)^2$$

$$\sigma_X^2 = \$249.64$$

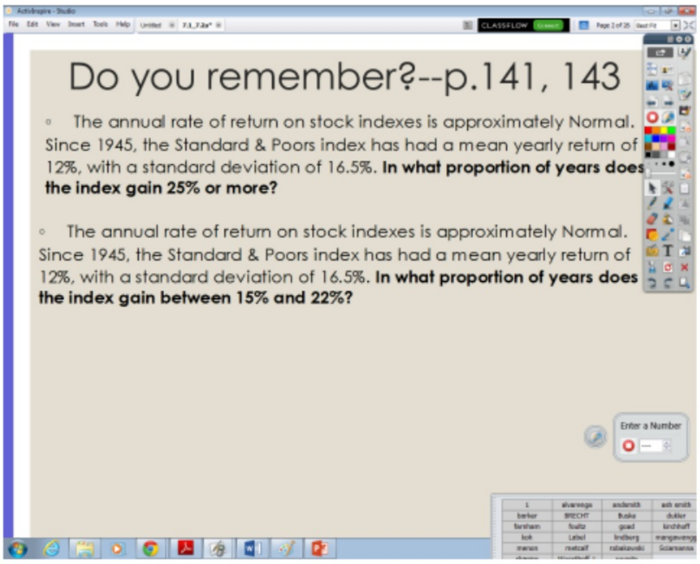
$$\sigma_X = \$15.80$$

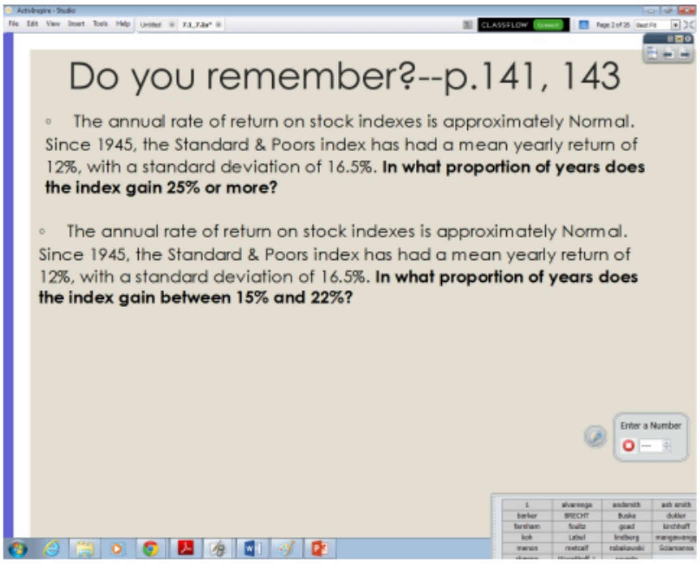
Technology

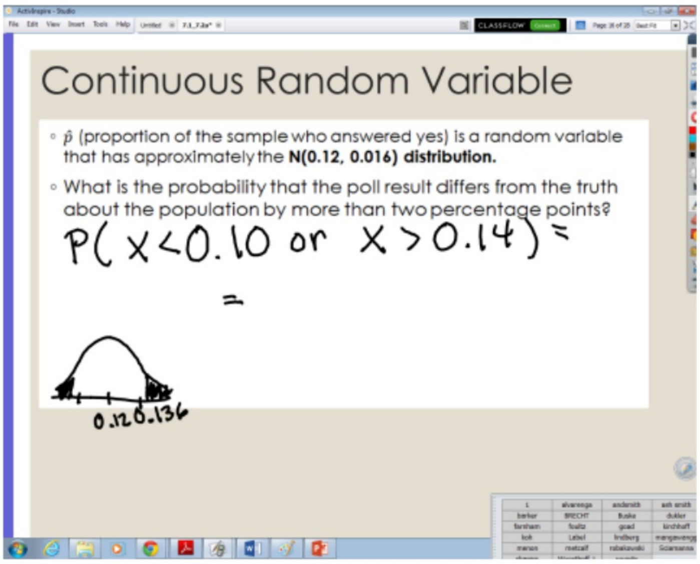
- When you work with a larger data set, it may be a good idea to use your calculator to calculate the standard deviation and mean.
- Enter the X values into List1 and the probabilities into List 2. Then 1-Var Stats L1, L2 will give you μ_x (as x-bar) and σ_x (to find the variance, you will have to square σ_x)
- EX: find μ_x and σ^2_x for the data in example 7.7 (p.485)

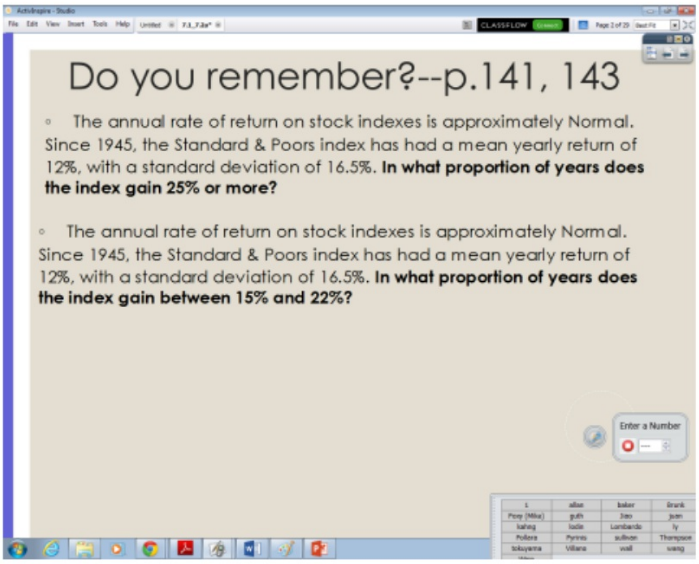
Assignment:

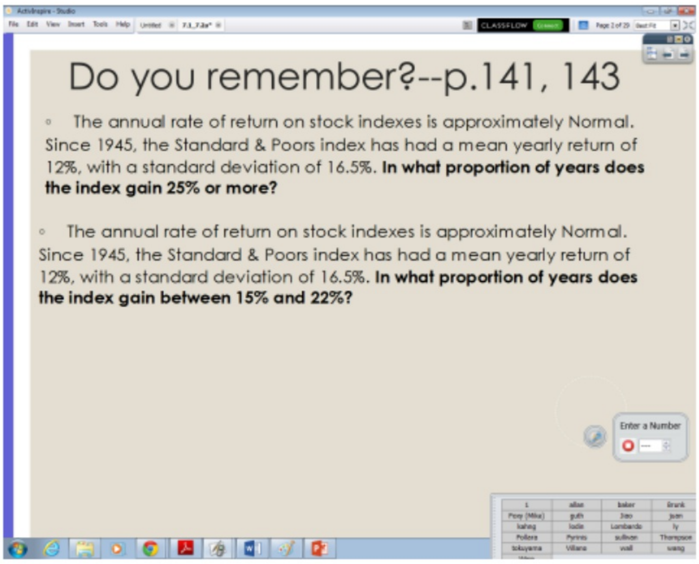
Exercises: 7.3, 7.4, 7.7, 7.9, 7.13-7.15, 7.20, 7.24, 7.25, 7.27, 7.32, 7.34

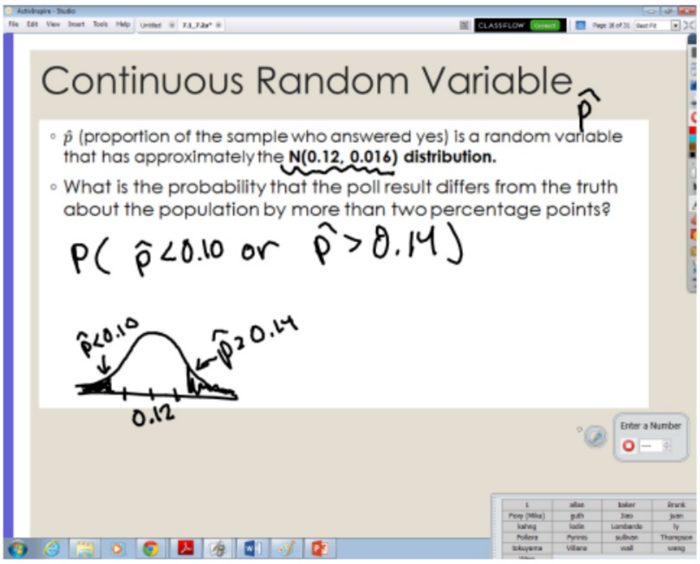

Do you remember?--p.141, 143
The annual rate of return on stock indexes is approximately Normal. Since 1945, the Standard & Poors index has had a mean yearly return of 12%, with a standard deviation of 16.5%. In what proportion of years does the index gain 25% or more?
The annual rate of return on stock indexes is approximately Normal. Since 1945, the Standard & Poors index has had a mean yearly return of 12%, with a standard deviation of 16.5%. In what proportion of years does the index gain between 15% and 22%?
Enter a Number
CLASSFLOW


Do you remember?--p.141, 143
The annual rate of return on stock indexes is approximately Normal. Since 1945, the Standard & Poors index has had a mean yearly return of 12%, with a standard deviation of 16.5%. In what proportion of years does the index gain 25% or more?
The annual rate of return on stock indexes is approximately Normal. Since 1945, the Standard & Poors index has had a mean yearly return of 12%, with a standard deviation of 16.5%. In what proportion of years does the index gain between 15% and 22%?
Enter a Number


Continuous Random Variable
p̂ (proportion of the sample who answered yes) is a random variable that has approximately the N(0.12, 0.016) distribution.
What is the probability that the poll result differs from the truth about the population by more than two percentage points?
P(X<0.10 or X>0.14) =
=
0.12 0.136
CLASSFLOW


Do you remember?--p.141, 143
The annual rate of return on stock indexes is approximately Normal. Since 1945, the Standard & Poors index has had a mean yearly return of 12%, with a standard deviation of 16.5%. In what proportion of years does the index gain 25% or more?
The annual rate of return on stock indexes is approximately Normal. Since 1945, the Standard & Poors index has had a mean yearly return of 12%, with a standard deviation of 16.5%. In what proportion of years does the index gain between 15% and 22%?
Enter a Number
CLASSFLOW


Do you remember?--p.141, 143
The annual rate of return on stock indexes is approximately Normal. Since 1945, the Standard & Poors index has had a mean yearly return of 12%, with a standard deviation of 16.5%. In what proportion of years does the index gain 25% or more?
The annual rate of return on stock indexes is approximately Normal. Since 1945, the Standard & Poors index has had a mean yearly return of 12%, with a standard deviation of 16.5%. In what proportion of years does the index gain between 15% and 22%?
Enter a Number



Continuous Random Variable
p̂
◦ p̂ (proportion of the sample who answered yes) is a random variable that has approximately the N(0.12, 0.016) distribution.
◦ What is the probability that the poll result differs from the truth about the population by more than two percentage points?
P(p̂ < 0.10 or p̂ > 0.14)
p̂ < 0.10
p̂ > 0.14
0.12
Enter a Number