Section 7.1

# Discrete and Continuous Random Variables

# Brainstorming

- In small groups…
  1. Think of a random phenomenon that would produce a value of 0, 1, 2, 3, 4, 5, etc. (non-negative integers)
  2. Now, think of a random phenomenon that would produce a value of any positive real number.
  3. How would you assign probabilities to these values?

# Random Variables

- A random variable is a variable whose value is a numerical outcome of a random phenomenon.
  - X = number of dots on the upside face of a rolled die
  - X = grams of sugar in a muffin
  - X = number of free-throw shots made
- We usually use X or Y for these, but not always.
- We're going to learn about two types of random variables and the probabilities associated with their possible values.

# Discrete Random Variables

- Take on positive integer values or zero.
  - Number of free-throw shots made out of five
  - Grade in a class if only A's, B's, C's, D's, and F's are given and A=4, B=3, C=2, D=1, and F=0.

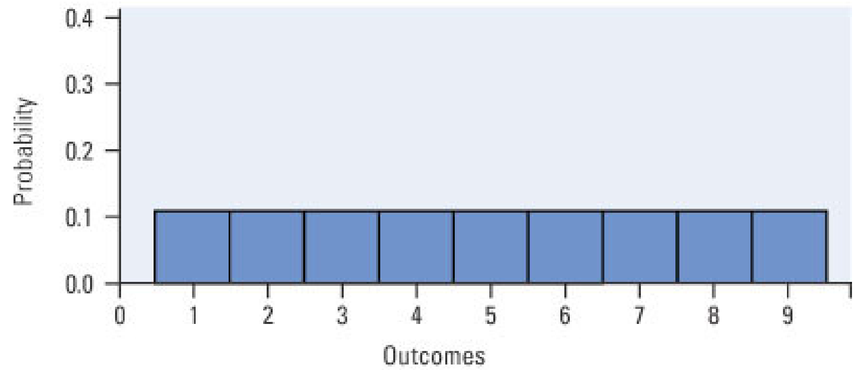
| **Value of X** | **0** | **1** | **2** | **3** | **4** |
|---|---|---|---|---|---|
| Probability | .01 | .05 | .30 | .43 | .21 |

$$P(X \geq 3) = P(X = 3) + P(X = 4)$$
$$= 0.43 + 0.21 = 0.64$$

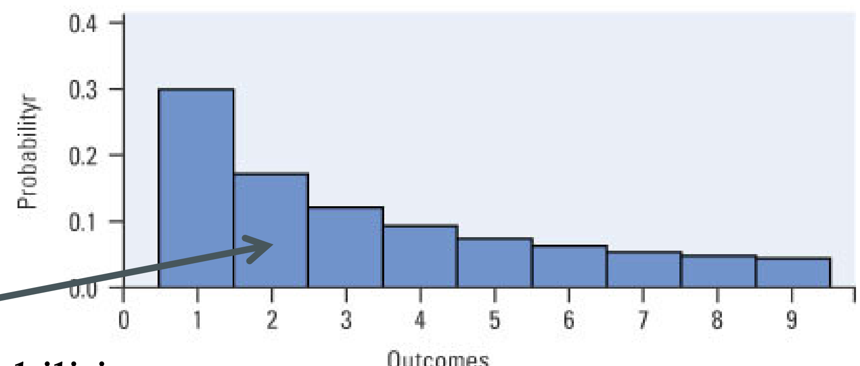
# Discrete Random Variables

- We can use histograms to display distributions of probabilities:

Benford's law, also called the first-digit law, states that in lists of numbers from many (but not all) real-life sources of data, the leading digit is distributed in a specific, non-uniform way.
http://en.wikipedia.org/wiki/Benford%27s_law



(a)



(b)

Note that the areas represent probabilities…

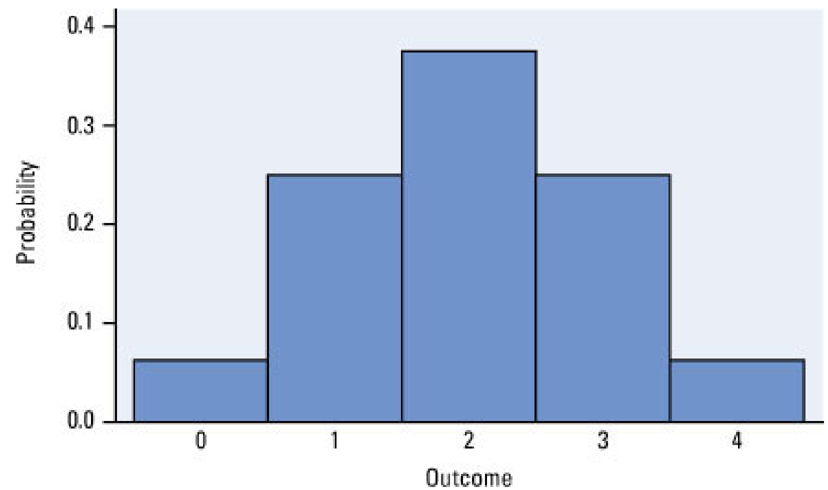
# Example 7.2 – Counting Heads

- What is the probability distribution of the discrete random variable X that counts the number of heads in four tosses of a coin?

- Two possible outcomes each time it's tossed, so there are $2^4$ total possible outcomes in 4 consecutive tosses.

| | | | | |
|---|---|---|---|---|
| | | HHTT | | |
| | | HTHT | | |
| | HTTT | HTTH | HHHT | |
| | THTT | THHT | HHTH | |
| | TTHT | THTH | HTHH | |
| TTTT | TTTH | TTHH | THHH | HHHH |
| $X = 0$ | $X = 1$ | $X = 2$ | $X = 3$ | $X = 4$ |

| | | | | | |
|---|---|---|---|---|---|
| Number of heads: | 0 | 1 | 2 | 3 | 4 |
| Probability: | 0.0625 | 0.25 | 0.375 | 0.25 | 0.0625 |



# Continuous Random Variables

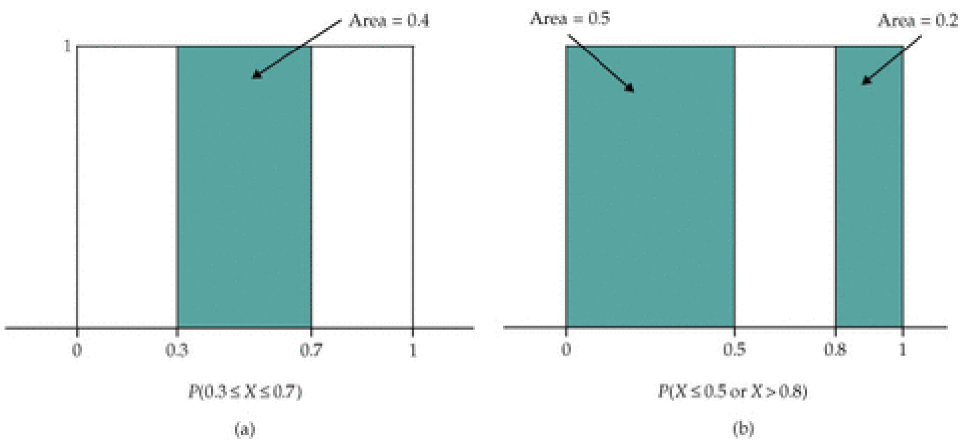
- Many random variables don't take on integer values.
  - X = grams of sugar in a muffin (if we're being really precise…)
  - X = any value between 0 and 1
  - X = amount of time required to understand this concept, in a decimal value
- How do we assign probabilities to these infinitely precise values?

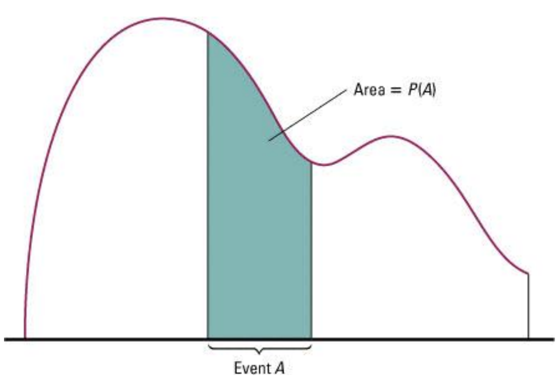
# Continuous Random Variables

- The probability of any <u>one</u> of these values would be zero.
- So, instead, we look at the probability of a range of values…

# Continuous Random Variables

- We rely on something we mentioned earlier – the area in a histogram.
- Or, since we like to approximate these distributions with smooth curves, we use the area under a density curve.



# Our Favorite Density Curves – Coming back like an old friend that just found you on Facebook

- Normal Curves!
- Normal distributions *are* probability distributions…you already knew this.
- We used this fact when we analyzed the performance of Carly Patterson in the 2004 Olympics.

# A Quick Review

- Remember to convert values to standardized values, or "z-scores."

$$Z = \frac{X - m}{s}$$

- Also, *N*(0,1) means "a Normal density curve with mean of 0 and standard deviation of 1."

# Example – Tattle-tales

- "You witness two students cheating on a quiz. Do you go to the professor?" was asked to an SRS of 400 undergrads.
- Suppose that if we asked *all* students, 12% would answer "Yes."
- The proportion, $p = 0.12$, would be a parameter describing all undergraduates (the population).
- The proportion $\hat{p}$ of the sample who answer "Yes" is a *statistic* used to estimate $p$.
- $\hat{p}$ is a random variable because repeating the SRS would give different samples of students and different values of $\hat{p}$.

# Example – Tattle-tales

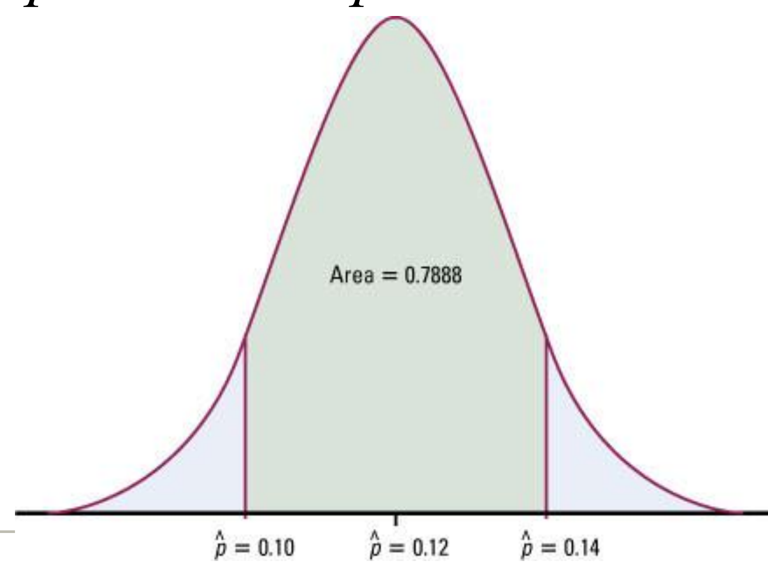
- What is the probability that the survey result differs from the truth about the population by more than 2 percentage points, given that $N(0.12, 0.016)$?
- In other words… what is $P(\hat{p} < 0.10$ or $\hat{p} > 0.14)$?

$$Z = \frac{X - 0.12}{0.016}$$

$$= P(-1.25 \text{ £ } Z \text{ £ } 1.25)$$

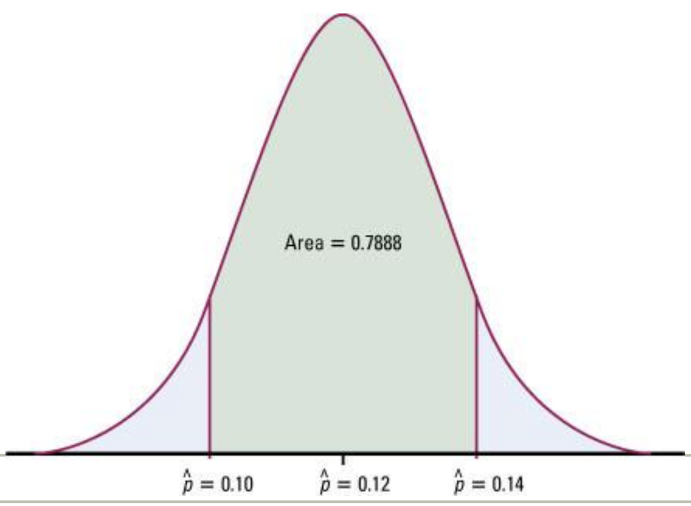
$$= 0.8944 - 0.1056$$

$$= 0.7888$$



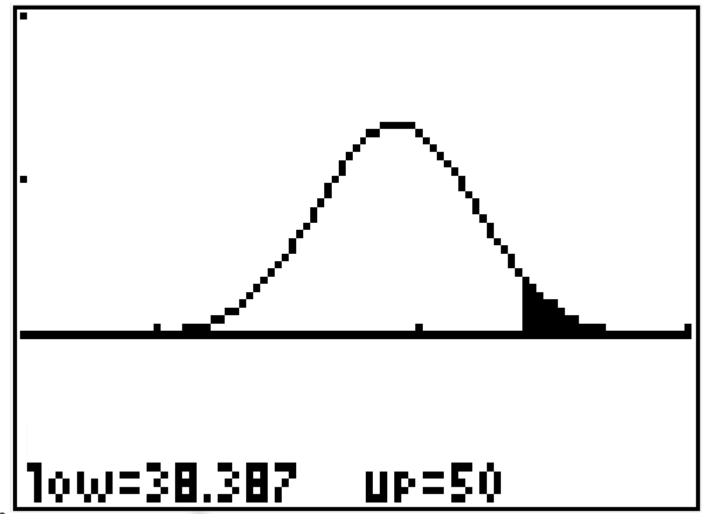
# Example – Tattle-tales

- What is the probability that the survey result differs from the truth about the population by more than 2 percentage points, given that *N*(0.12,0.016)?
- Answer: 1 - 0.7888, or .2112. About 21% of sample results will be off by more than two percentage points.



# Example

- What is the probability of Carly getting a total score of 38.387 or above?
- $P(X \geq 38.387) = ?$
- Find z if her mean total was 37.9 and her standard deviation was 0.30.
- What is the answer to our original question?



# In Summary

- Discrete Random Variables –

| Value of X | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| Probability | .01 | .05 | .30 | .43 | .21 |

- Continuous Random Variables –