Chapter 11

Understanding Randomness

At the end of this chapter, you should be able to

- Identify a random event.
- Describe the properties of random events.
- Use random number table to study random events.
- Do simulation.

We know what it means to be random—right?

- Rolling dice, spinning spinners, shuffling cards… all select at random ⇨ rely on chance outcomes
- What's the most important aspect of randomness in these games?
  - **Must be *FAIR***
  - **Nobody can guess the outcome before it happens**
  - **Some set of outcomes will be equally likely**

Random

An event is random if we know what outcomes could happen, but not which particular values did or will happen.

Randomness

Lets simulate 100 tosses with a random number table.

Line								
101	19223	95034	05756	28713	96409	12531	42544	82853
102	73676	47150	99400	01927	27754	42648	82425	36290
103	45467	71709	77558	00095	32863	29485	82226	90056
104	52711	38889	93074	60227	40011	85848	48767	52573

`randInt(0,1)` Simulates coin toss (0 = tails, 1 = heads)

`randInt(0,1,5)` Simulates 5 coin tosses

`sum(randint(0,1,100))` Sum of 100 coin tosses

randInt(

- The command **randInt(** provides a random integer between the first argument and the second. randInt(can be used to provide a number for guessing games.
- **randInt(lowNum, highNum, numTrials)**
  - This creates a *numTrials* length list that is composed of random integers between *lowNum* and *highNum*.

`randInt(0,1)`

`randInt(0,1,5)`

`sum(randint(0,1,100))`

Command Syntax
randInt(*min*,*max*[,*# of numbers*])

Menu Location
Press:
1) MATH to access the mathmenu.
2) LEFT to access the PRB submenu.
3) 5 to select randInt(, or use arrows.

Randomness

Do a calculator simulation.

randInt(0,1) simulates coin toss(0 = tails, 1 = heads)
randInt (0, 1, 5) simulates 5 coin tosses

Any case of all heads or all tails?

List Math 5 Math PRB 5 (0, 1, 100) : Simulates 100 tosses of a fair coin
sum(randInt(0, 1, 100)) : each coin is 0 or 1, therefore the sum of these 100 values = counting # of heads

Simulate 100 tosses 5 or 6 times. Is the sum > 65?

Make a good decision rule for testing coins: any coin landing Heads>65 times in 100 tosses is suspect. Even after tossing the mystery coin 100 times and declaring it OK or biased

Randomness

Suppose a basketball player has an **<u>80%</u>** free throw success rate. How can we use random numbers to simulate whether or not he makes a foul shot? How many shots might he be able to make in a row without missing?

Identify the component, trial, response variable and the statistic.

Line								
101	19223	95034	05756	28713	96409	12531	42544	82853
102	73676	47150	99400	01927	27754	42648	82425	36290
103	45467	71709	77558	00095	32863	29485	82226	90056
104	52711	38889	93074	60227	40011	85848	48767	52573

Randomness

80%

Component: *the most basic event we're simulating*
- Here that's **taking one shot**.
- Get a random digit 0-9; let 0-7 be good, 8-9 missed.

Trial: *The sequence of events we want to investigate*
- here that's **shooting shot after shot until he misses**.
- Look at the random digits until we find an 8 or 9.

Response variable: *What happened in the trial? What are we interested in?*
- *Count the number of shots made before the miss.*

Statistic: *Find the mean number of shots made.*

Randomness

72%

How might the simulation change if the percentage was only **<u>72%</u>** free throw success rate?

Component:
- Get random digits 00-99; let 00-71 be good, 72-99 missed.

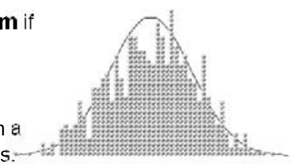
Trial:
- Look at the random digits until we find 72-99.

The gambling industry relies on probability distributions to calculate the odds of winning. The rewards are then fixed precisely so that, *on average,* players lose and the house wins.

The industry is very tough on so called "cheaters" because their probability to win exceeds that of the house. Remember that it is a business, and therefore it is supposed to be profitable.

Randomness and probability

A phenomenon is **random** if individual outcomes are uncertain, but there is nonetheless a regular distribution of outcomes in a large number of repetitions.



Randomness is a description of a kind of order that emerges only in the long run.

Probability is empirical, i.e., it is based on observation rather than theorizing.

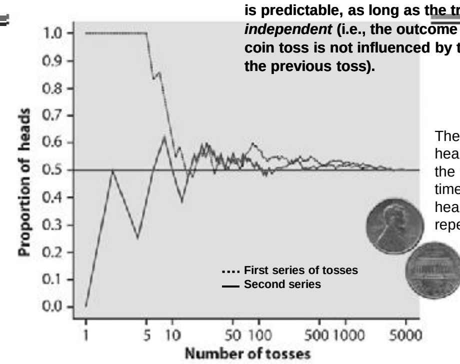
The **probability** of any outcome of a random phenomenon can be defined as the proportion of times the outcome would occur in a very long series of repetitions.

Randomness

- Is not the same as haphazard. Is not always what we might think of as "***at random***"
- We call a phenomenon **<u>random</u>** if individual outcomes are uncertain but there is nonetheless a regular distribution of outcomes in a large number of repetitions.
- Random outcomes have a lot of structure, especially in long run
  - Think of flipping a coin—can't predict how it'll land on 1 toss, but can be confident it will land on heads 50% of the time if you flip it 1,000 times
- As statisticians we use randomness as a tool
- It's harder to pick a random number than you think!

Coin Toss

The result of any single coin toss is random. But the result over many tosses is predictable, as long as the trials are *independent* (i.e., the outcome of a new coin toss is not influenced by the result of the previous toss).



The probability of heads is 0.5 = the proportion of times you get heads in many repeated trials.

Why Be Random?

- What is it about chance outcomes being random that makes random selection seem fair?
  1) Nobody can guess the outcome before it happens.
  2) Randomness has structure in the long run.
  3) When we want things to be fair, usually some underlying set of outcomes will be equally likely (although in many games some combinations of outcomes are more likely than others).

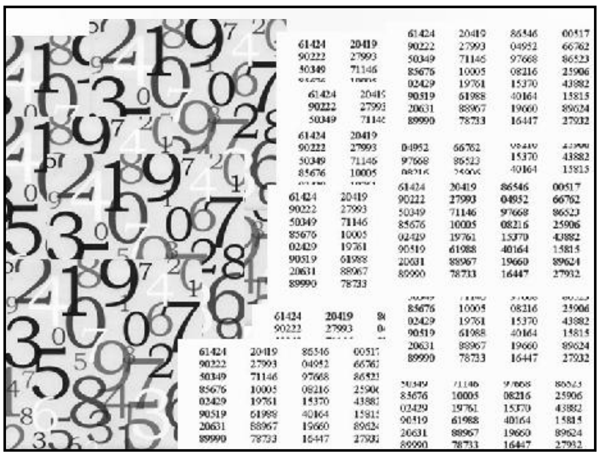
Ready for a TEST?

When I change the slide, look at the numbers quickly

Pick a NUMBER

Write it DOWN

READY???



Randomly Pick a Number…

1 2 3 4

Did you pick 3?

- About 75% of people pick the number 3
- 20% pick either 2 or 4
- Only about 5% choose 1

Randomly Pick a Number…

1 2 3 4

Examples:
- Pick "heads" or "tails."
- Flip a fair coin. Does the outcome match your choice? Did you know before flipping the coin whether or not it would match?

How do we get random numbers?

- Computers/graphing calculator
  - However, these are generated from a program, and thus are not truly random
  - These are *pseudorandom*
  - Because there are finite possibilities, the sequence *must* repeat at some point
- In the past, books of random numbers were published.
- Table of Random Numbers.

Practical Randomness

- We need an imitation of a real process so we can manipulate and control it.
- In short, we are going to **simulate** reality.

Using Random Numbers

Ex) How many times do we need to roll a die before we get a 6?

- Simulation
  - 1,2,3,4,5,6: Roll on die
  - 0,7,8,9: Throw out
- Look at random numbers

31087 42430 60322 34765 15757 53300 97392 98035

05228 68970 84432 04916 52949 78533 316

Using Random Numbers

31087 42430 60322 34765 15757 53300

97392 98035 05228 68970 84432 04916

52949 78533 316

1, 2, 3, 4, 5, 6: Roll on die
0, 7, 8, 9: Throw out

Trial Number	Component Outcomes	Trial Outcomes; y = Number of times to roll a die
1	31087 42430 6	7
2	0322 3476	6
3	5 15757 53300 97392 98035 05228 6	15
4	8970 84432 04916	7
5	52949 78533 316	44/5 = 8.8 9

Using Random Numbers

- Throw out 0,7,8,9
  3142436322346515553332
  3552264432416524533316

- Trial 1: 3142436 __7__ rolls
- Trial 2: 322346 __6__ rolls
- Trial 3: 515553332355226 __15__ rolls
- Trial 4: 4432416 __7__ rolls
- Trial 5: 524533316 __9__ rolls

44/5 = 8.8

Using Random Numbers

- Out of 5 trials
  - mean number of rolls = ____
- More trials = more accurate result.
  - Out of 200 trials
    - mean number of rolls = ______ .

Simulation
- consists of a sequence of random outcomes that model a situation

Component
- The basic building block of simulation

Outcomes
- possible results, occur at random

Trial
- the sequence of events/components we want to investigate

Response Variable
- what happened in the trial

Running a Simulation

1) Identify the component to be repeated.
2) Explain how you will model the component's outcome—assign digits to outcomes.
3) Explain how you simulate (how you will combine the components to model a trial—how does a trial stop, repeat?)
4) State clearly what the response variable is.
5) Run several trials—more is better.
6) Analyze the response variable, i.e., collect and summarize the results of all the trials.
7) State your conclusion (in context).

Suppose a cereal manufacturer puts pictures of famous athletes on cards in boxes of cereal in the hope of boosting sales. The manufacturer announces that 20% of the boxes contain a picture of Tiger Woods, 30% a picture of David Beckham, and the rest a picture of Serena Williams. You want all three pictures. How many boxes of cereal do you expect to have to buy in order to get the complete set? **How could we simulate our purchases using this information if we're trying to get all 3?**

Let's Use a Random Model!

- Why random?
- When we pick a box off the shelf, we don't know what picture is inside.
- We assume: pictures are randomly placed in the boxes and that the boxes are distributed randomly to stores around the country.
- Why a model?
- Because we don't want to actually buy hundreds of cereal boxes.
- We need an imitation of the real process that we can manipulate and control.
- We are going to **simulate** reality.

A Simulation

- We are asking how many boxes do you expect to buy to get a complete card collection.
- We can't answer this question by completing our collection only once!
- We want to understand the *typical* number of boxes to open, how that number varies, and, often, the shape of the distribution.
- We will have to do this over and over, and each time we attain a simulated answer to our question we will call this a ***trial.***

Building Our Simulation

- We know how to find equally likely random digits
- How do we get from there to simulating the trial outcomes?
- We know the relative frequencies of the cards:
  20% Tiger
  30% Beckham
  50% Serena
- Here are our random digits:

0 1 2 3 4 5 6 7 8 9

Out of these ten digits each one has a 10% chance of being generated at random

So…

Building Our Simulation

0 1 2 3 4 5 6 7 8 9

20% Tiger – 0 and 1
30% Beckham – 2, 3, 4
50% Serena -- 5, 6, 7, 8, 9

We can interpret the digits 0 and 1 as finding Tiger; 2, 3, and 4 as finding Beckham; and 5 through 9 as finding Serena

- Generating one random number between 0 and 9 now simulates opening one box
- Opening the box is the basic building block of our simulation, called a **component** of our simulation

Building Our Simulation

- The component is opening the box.
- However, the component's outcome isn't the result we want.
- We need to observe a sequence of components until our card collection is complete.
- The *trial's* outcome is called the **response variable**, for this simulation that is the ***number* of components (boxes) in the sequence**
- Let's look at the steps for making a simulation:

Running a Simulation

1) **Identify a component**
   Component (basic act that is being repeated) "*opening a box of cereal.*" The trial's outcome is "*which picture is in the box*"
2) **Model a component's outcome**
   (decide how to use random numbers to model its outcome) Probability for each outcome must be the same for the random numbers "0,1 (Tiger Woods) and 2,3,4 (David Beckham) and 5-9 (Serena Williams) " percents the same as given for the pictures

Running a Simulation

3) **Explain how will you simulate the trial**
   "pretend to open cereal boxes"
4) **State clearly what the response variable is** (*what are we interested in?)*
   "how many boxes opened until you get all three pictures"
5) **Model a trial –** combine components until you get what you want, that's **one trial** "look at a series of random digits until the trial is complete" (Use line 3 on page 257)

Running a Simulation

89064 2730 8645681 41219
822665388587328580 169902
78431 1038 042006 7664

0,1 Tiger Woods
2,3,4 David Beckham
5-9 Serena Williams

Trial Number	Component Outcomes	Trial Outcomes; y = Number of Boxes
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		

Running a Simulation

5) **Model a trial** (Use line 3 on page 257)

89064 2730 8645681 41219 822665388587328580 169902 78431 1038 042006 7664.

Trial Number	Component Outcomes	Trial Outcomes: y = Number of boxes
1	89064 = Serena, Serena, Tiger, Serena, Beckham	5
2	2730 = Beckham, Serena, Beckham, Tiger	4
3	8645681 = Serena, Serena, Beckham, . . . , Tiger	7
4	41219 = Beckham, Tiger, Beckham, Tiger, Serena	5
5	822665388587328580 = Serena, Beckham, . . . , Tiger	18
6	169902 = Tiger, Serena, Serena, Serena, Tiger, Beckham	6
7	78431 = Serena, Serena, Beckham, Beckham, Tiger	5
8	1038 = Tiger, Tiger, Beckham, Serena	4
9	042006 = Tiger, Beckham, Beckham, Tiger, Tiger, Serena	6
10	7664 . . . = Serena, Serena, Serena, Beckham . . .	?

Running a Simulation

5) Model a trial (Use line 3 on page 257)

Wait! Only 10 trials?

If you fear that these may not be accurate estimates because we ran only nine trials, you are absolutely correct. The more trials the better and nine is woefully inadequate. Twenty trials is probably a reasonable minimum if you are doing this by hand. Even better, use a computer and run a few hundred trials!

Running a Simulation

6) Analyze the response variable

"number of boxes / number of trials"

What is the mean and stddev of the number of boxes from each trial?

N(______ , ______)

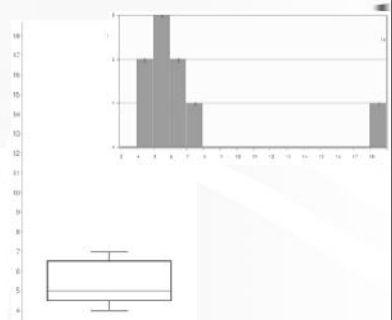
- Out of 10 trials
  mean = ________ boxes
- More trials = more accurate result.
  Out of 200 trials
  mean = ________ boxes.

Running a Simulation

7) State your conclusion (in context)

"Based on our simulation, we expect to open about _5__ boxes on the average to get all 3 pictures."

Population size: 9
Median: 5
Minimum: 4
Maximum: 18
First quartile: 4.5
Third quartile: 6.5
Interquartile Range: 2
Outlier: 18



Another Example!

Simulating a Dice Game

The game of 21 can be played with an ordinary 6-sided die. Competitors each roll the die repeatedly, trying to get the highest total less than or equal to 21. If your total exceeds 21, you lose.

Suppose your opponent has rolled an 18. Your task is to try to beat him by getting more than 18 points without going over 21. How many rolls do you expect to make, and what are your chances of winning?

Question:
- How will you simulate the components?

Question: How will you simulate the components?

- A component is **one roll of the die**. A roll will be simulated by looking at a random digit from a table or an internet site. The **digits 1 through 6 will represent the results on the die** and we shall **ignore digits 7-9 and 0**.

~~0~~ 1 2 3 4 5 6 ~~7 8 9~~

Question: How will you combine components to model a trial? What's the response variable?

- Add components until the total is greater than 18, counting the number of "rolls".
- If my total is greater than 21, it is a loss. If not, it is a win.
- These two components are variables. I'll count the number of times I roll the die and I'll keep track of whether I win or lose.

Question: How would you use those random digits to run trials? Show your method clearly for two trials

91129 58757 69274 92380 82464 33089

Trial 1: ~~9~~ 1 1 2 ~~9~~ 5 ~~8~~ ~~7~~ 5 ~~7~~ 6
Total: 1 2 4 9 14 **20**
Outcome: 6 rolls, won

Trial 2: ~~9~~ 2 ~~7~~ 4 ~~9~~ 2 3 ~~8~~ ~~0~~ ~~8~~ 2 4 6
Total: 2 6 8 11 13 17 **23**
Outcome: 7 rolls, lost

Question: Suppose you run 30 trials, getting the outcomes tallied below. What is your conclusion?

Number of Rolls		Result	
4	III	Won:	21
5	卌 卌	Lost	9
6	卌 卌 I		
7	卌		
8	I		

- Based on my simulation, competing against an opponent who has a score of 18, I expect my turn to usually last 5 or 6 rolls and I should win about 70% of the time.

Just Checking

World Series

Just Checking, World Series

- The baseball World Series consists of up to seven games.
- The first team to win four games wins the series.
- The first two are played at one team's home ballpark, the next three at the other team's park, and the final two (if needed) are played back at the first ballpark.

Home Field Advantage

- Records show that over the past century there is a home field advantage – the home team has about a 55% chance of winning.
- Does the current system of alternating ballparks even out the home field advantage? How often will the team that begins at home win the series?
- 1) What is the component to be repeated?
- Answer: The component is one game.
- 2) How will you model each component from equally likely random digits?
- Answer: Generate random numbers and assign numbers from 00 to 54 to the home team's winning and from 55 to 99 to the visitors winning.

Home Field Advantage (Cont.)

- 3) How will you model a trial by combining components?
- Answer: Generate components until one team wins 4 games. Record which team wins the series.
- 4) What is the response variable?
- Answer: The response variable is who wins the series.
- 5) How will you analyze the response variable?
- Answer: Calculate the proportion of wins by the team that starts at home.

Set up a simulation

57 students participated in a lottery for a particularly desirable dorm room—a triple with a fireplace and private bath in the tower. Twenty of the participants were members of the same varsity team. When all three winners were members of the team, the other students cried foul. Use a simulation to determine whether an all-team outcome could reasonably be expected to happen if everyone had a fair shot at the room.

Question:

- Could an all-football-team outcome reasonably be expected to happen if everyone had a fair shot at the room?

Running a Simulation

1) Identify a component

"*selection of a student*"

2) Model a component's outcome

"*To model the outcome we will let two digit random numbers represent the students. The digits 01 through 20 represent the twenty varsity team applicants. The digits 21 through 57 represent the other thirty-seven applicants. Ignore the rest.*"

e

Running a Simulation

3) Explain how will you simulate the trial

"*Each trial consists of identifying pairs of digits as V (varsity) or N (non-varsity) until 3 are chosen, ignoring out of range or repeated numbers (can't put the same person in the room more than once).*"

4) State clearly what the response variable is

"*3 students selected, how many trials have all 3 from the team. The response variable is whether or not all selected students are on the varsity team.*"

Running a Simulation

5) Run many trials.

"*Repeat at least 20 times.*"

6) Analyze the response variable

"*The response variable is whether or not all three roommates are varsity team members. The response can only be either yes or no. The response is yes only when all three roommates are varsity team members. After each trial a determination is made on yes or no.*"

Running a Simulation

7) State your conclusion (in context)

" *Is it likely to have randomly happened that the 3 were from the football team? After 10 trials were run by random selection and each trial's response was determined the number of yes responses were counted. If three students are selected from a pool of fifty-seven students, twenty of which were varsity team member, then all varsity room occurred only 10% of the time. While this situation could happen, this result is not very likely to happen and thus the "fair" selection would be suspicious.*"

Set up a simulation

A basketball player with a 65% shooting percentage has just made 6 shots in a row. The announcer says this player “is hot tonight! He’s in the zone!” Assume the player takes about 20 shots per game. Is it unusual for his to make 6 or more shots in a row during a game?

Running a Simulation

1) Identify a component
“*one shot*”

2) Model a component’s outcome
“Generate pairs of random digits 00-99.
Let *01-65 represent a made shot*
66-99, 00 represent a missed shot”

Running a Simulation

3) Explain how will you simulate the trial
A trial consists of 20 simulated shots.

4) State clearly what the response variable is
“*whether or not 20 simulated shots contain a run of 6 or more baskets.*”

Running a Simulation

5) Run many trials.

6) Analyze the response variable
Number of successes / total number of trials

Running a Simulation

7) State your conclusion (in context)
According to the simulation, the player is expected to make 6 or more shots in a row in about 40% of games. This isn’t unusual. The announcer was wrong to characterize his performance as extraordinary.

BAD SIMULATIONS

#11, p. 265

a) The outcomes are not equally likely. For example, the probability of getting 5 heads in 9 tosses is not the same as the probability of getting 0 heads, but the simulation assumes they are equally likely.
b) The even-odd assignment assumes that the player is equally likely to score or miss the shot. In reality, the likelihood of making the shot depends on the player’s skill.

BAD SIMULATIONS

#11, p. 265

c) Suppose a hand has four aces. This might be represented by 1,1,1,1, and any other number. The likelihood of the first ace in the hand is not the same as for the second or third or fourth. But with this simulation, the likelihood is the same for each.

WRONG CONCLUSION

#13, p. 265

The conclusion should indicate that the simulation **suggests** that the average length of the line would be 3.2 people. Future results might not match the simulated results exactly.

Remember!

- Whenever we make a simulation in some sense it is always wrong. After all, its not the real thing. We never did roll the dice in front of the board and found the average of the rolls need to land exactly on the last space. **Remember** your simulation is only predicting what *might* happen, however it is up to you to make the simulation as accurate as possible.

What Can Go Wrong?

- The biggest mistake you can make is not running enough tests. I only ran 11 tests due to space, however you should always run **at least** 20 tests to get a good simulation of the randomness occurring.

What Can Go Wrong?

- Don’t overstate your case
  - Beware of confusing what ***really*** happens with what a simulation suggests ***might happen***
- Model the outcome chances accurately
  - Oftentimes a simulation is set up to show the desired result rather than accurately model what happens.
- Run enough trials
  - Simulation is cheap and fairly easy to do.