SECTION 5

Chapter 16, 17 and 18

CHAPTER 16

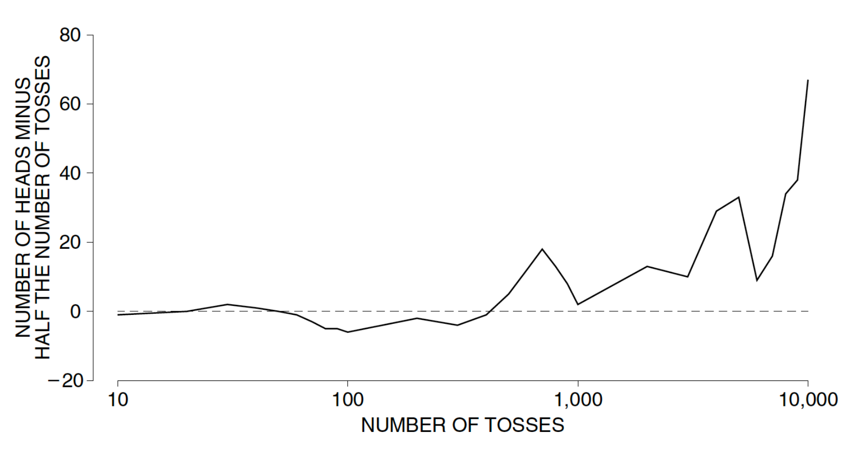
The Law of Averages

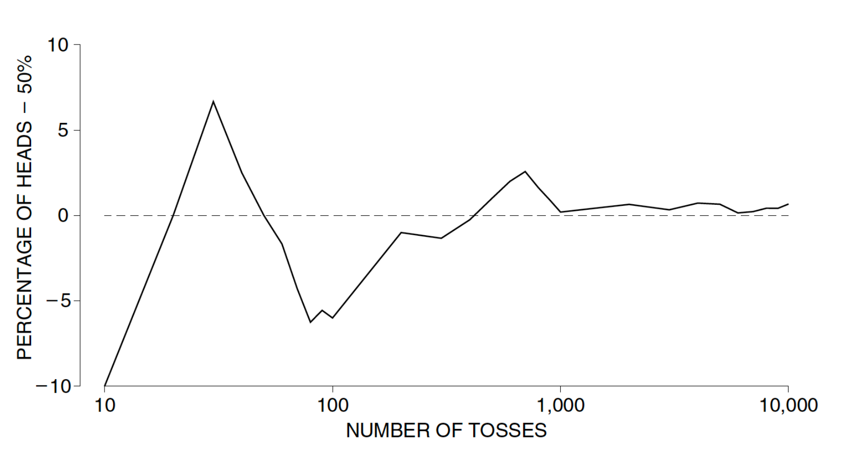
- The Law of Averages
- Chance Processes
- The Box Model

The Law of Averages

Coin Tossing Example:

- As the number of tosses ↑, the difference between the number of heads and half the number of tosses gets ↑
- As the number of tosses ↑, the difference between the percentage of heads and 50% gets ↓





number of heads = half the number of tosses + chance error.

4. (a) A coin is tossed, and you win a dollar if there are more than 60% heads. Which is better: 10 tosses or 100? Explain.
   (b) As in (a), but you win the dollar if there are more than 40% heads.
   (c) As in (a), but you win the dollar if there are between 40% and 60% heads.
   (d) As in (a), but you win the dollar if there are exactly 50% heads.

Chance Processes

Motivation: when a coin is tossed a large number of times, the actual number of heads is likely to differ from the expected number.

So, want to find a way to measure the variability of *the sum of the draws*

Sum of the draws

- Draw tickets at random from a box.
- Add up the numbers on the tickets.

One hundred draws are made at random with replacement from the box $\boxed{1}\ \boxed{2}$.

(a) How small can the sum be? How large?
(b) How many times do you expect the ticket $\boxed{1}$ to turn up? The ticket $\boxed{2}$?
(c) About how much do you expect the sum to be?

(a) small if all the draws are with the smallest number, 1: $1 \times 100 = 100$
large if all the draws are with the largest number, 1: $2 \times 100 = 200$
(b) $P(1) = 0.5 = P(2)$, $Exp(1) = N \times P(1) = 50$; $Exp(2) = N \times P(2) = 50$
(c) $Avg = 4$; $Exp = N \times Avg = 4 \times 100 = 400$

The Box Model

Goal: analyze the chance variability

- The tickets in the box show the various amounts that can be won or lost on a single play.

| 1 ticket +$35 | 37 tickets −$1 |

Play 100 times

1 chance in 38 of winning
⇔ 37 chances in 38 of losing

a single number pays 35 to 1

number of draws = 100

- The chance of drawing any particular number from the box must equal the chance of winning that amount on a single play.
- The number of draws equals the number of plays

A gambler will play roulette 50 times, betting a dollar on four joining numbers each time (like 23, 24, 26, 27 in figure 3, p. 282). If one of these four numbers comes up, she gets the dollar back, together with winnings of $8. If any other number comes up, she loses the dollar. So this bet pays 8 to 1, and there are 4 chances in 38 of winning. Her net gain in 50 plays is like the sum of ______ draws from the box ______. Fill in the blanks; explain.

| __4[$8]____34[$=1]___|

CHAPTER 17

The Expected Value and Standard Error

- The Expected Value
- The Standard Error
- Using the Normal Curve
- A Short-cut for SD
- Classifying and Counting

The Expected Value

The expected value for the sum of draws made at random with replacement from a box:

number of draws × average of box

Average of box: the average of the numbers in the box
on average, each draw adds around **average of box** to the sum

A game is *fair* if the expected value for the net gain equals 0: on the average, players neither win nor lose. A generous casino would offer a bit more than $1 in winnings if a player staked $1 on red-and-black in roulette and won. How much should they pay to make it a fair game? (Hint: Let x stand for what they should pay. The box has 18 tickets $\boxed{x}$ and 20 tickets $\boxed{-\$1}$. Write down the formula for the expected value in terms of x and set it equal to 0.)

Box model: |___18[x]__20[-$1]___| **avg:** $\frac{1}{38}\big(18x+20(-1)\big)=0$ since it is a fair game => x≈1.11

The Standard Error

Sum = Expected Value + Chance Error

⇔ Chance Error = Sum – Expected Value

The sum of the draws is likely to be around the expected value, give or take SE or so.

To calculate the Standard Error (SE):

Square Root Law: $\sqrt{\boldsymbol{number\ of\ draws}} \times (\boldsymbol{SD\ of\ box})$

SD: measure spread of the lists of numbers

SE: measure the chance variability

NOTE:

- Each draw adds some extra variability to the sum. Number of draws ↑, the sum gets harder to predict, the chance errors ↑, and the SE ↑.
- The spread of the in the box ↑ (SD ↑), so does SE ↑

Tickets are drawn at random with replacement from a box of numbered tickets. The sum of 25 draws has expected value equal to 50, and the SE is 10. If possible, find the expected value and SE for the sum of 100 draws. Or do you need more information?

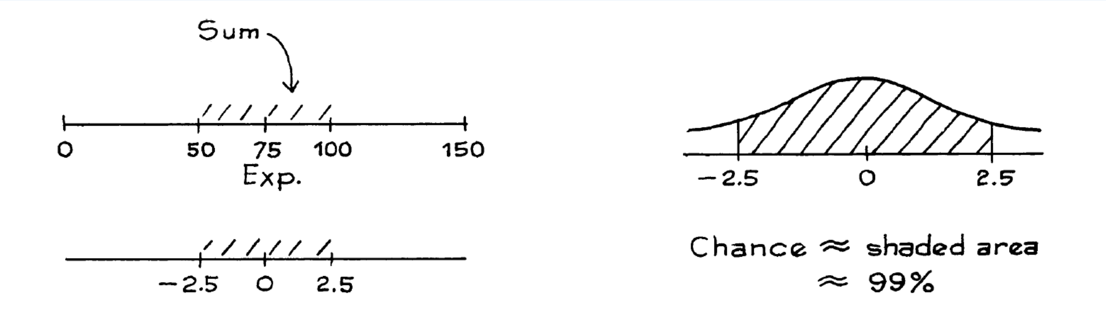
Find the **Avg and SD** of the list in the box. They will not change with the sum of draws.

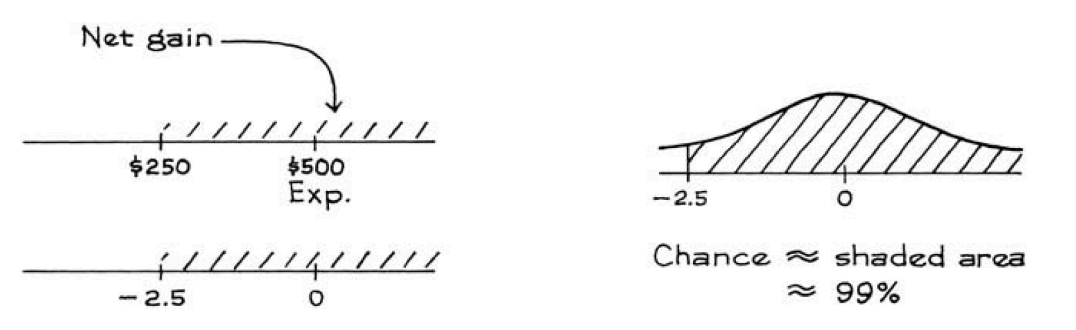
Exp = 50 -> Avg = Exp/25 = 2 -> Exp_new = Avg × 100 = 200;

SE = 10 -> SD = $\frac{SE}{\sqrt{25}}$ =2 -> SE_new = $\sqrt{100}$ × SD = 10 × 2 = 10

Using the Normal Curve

Questions like: find the chance that the sum of the draws will be in a given range (between/below/above..)





Steps:

1. Find out the box model
2. Calculate the expected value of the sum
3. Calculate the SE of the sum
4. Tansfer the bonds of the interval into the units
5. Check the normal table to get the shaded area -> chance

Using the Normal Curve

Exercises:

One hundred draws will be made at random with replacement from the box $\boxed{1}\ \boxed{3}\ \boxed{3}\ \boxed{9}$.

(a) How large can the sum be? How small?
(b) How likely is the sum to be in the range from 370 to 430?

There are two options:

(i) One hundred draws will be made at random with replacement from the box $\boxed{1}\ \boxed{1}\ \boxed{5}\ \boxed{7}\ \boxed{8}\ \boxed{8}$.
(ii) Twenty-five draws will be made at random with replacement from the box $\boxed{14}\ \boxed{17}\ \boxed{21}\ \boxed{23}\ \boxed{25}$.

Which is better, if the payoff is—

(a) \$1 when the sum is 550 or more, and nothing otherwise?
(b) \$1 when the sum is 450 or less, and nothing otherwise?
(c) \$1 when the sum is between 450 and 550, and nothing otherwise?

A Short-cut for SD

When a list has only 2 different numbers ("big" and "small")

$$SD = (\text{big number} - \text{small number}) \times \sqrt{\frac{\textit{fraction with}}{\textit{big number}} \times \frac{\textit{fraction with}}{\textit{small number}}}$$

Exercises:

Does the formula give the SD of the list? Explain.

	List	*Formula*
(a)	7, 7, 7, −2, −2	$5 \times \sqrt{3/5 \times 2/5}$
(b)	0, 0, 0, 0, 5	$5 \times \sqrt{1/5 \times 4/5}$
(c)	0, 0, 1	$\sqrt{2/3 \times 1/3}$
(d)	2, 2, 3, 4, 4, 4	$2 \times \sqrt{1/6 \times 2/6 \times 3/6}$

At Nevada roulette tables, the "house special" is a bet on the numbers 0, 00, 1, 2, 3. The bet pays 6 to 1, and there are 5 chances in 38 to win.

(a) For all other bets at Nevada roulette tables, the house expects to make about 5 cents out of every dollar put on the table. How much does it expect to make per dollar on the house special?

(b) Someone plays roulette 100 times, betting a dollar on the house special each time. Estimate the chance that this person comes out ahead.

Classifying and Counting

Example 4. A die is rolled 60 times.

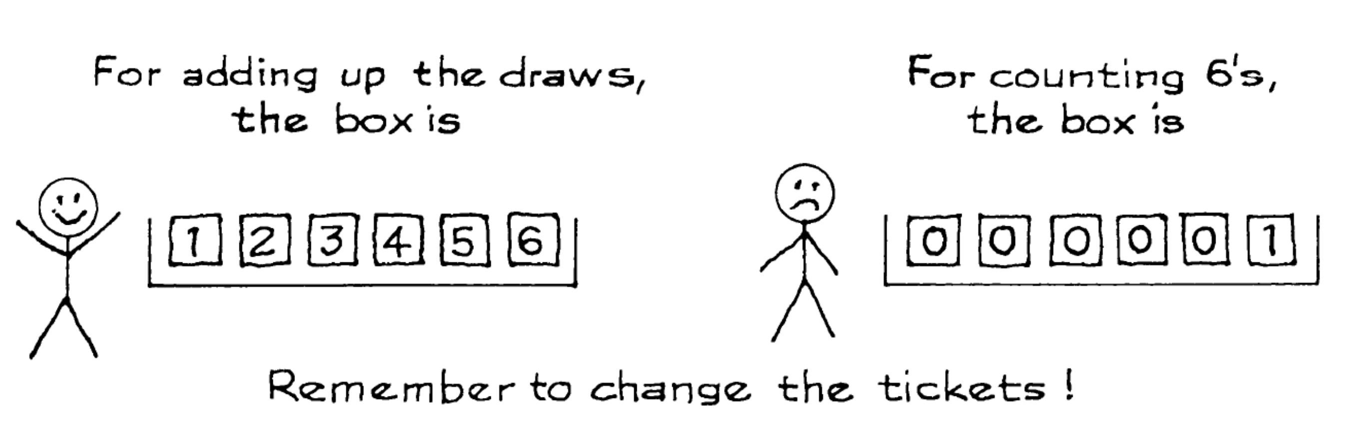
(a) The total number of spots should be around ______, give or take ______ or so.

(b) The number of 6's should be around ______, give or take ______ or so.

Adding

Classifying and Counting

If you have to classify and count the draws, put 0's and 1's on the tickets. Mark 1 on the tickets that count for you, 0 on the others.



One hundred draws are made at random with replacement from | 1 2 3 4 5 |. What is the chance of getting between 8 and 32 tickets marked "5"?

- New box: |________________|
- Avg:
- SD:
- Exp:
- SE:
- Chance:

CHAPTER 18

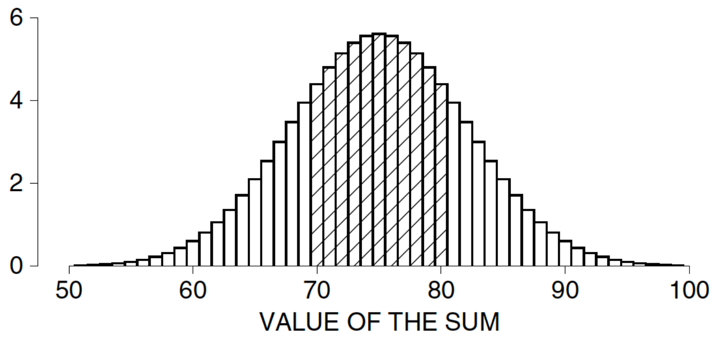
The Normal Approximation for Probability Histograms

- Probability Histograms
- Probability Histograms and the Normal Curve
- The Normal Approximation (coin)
- The Scope of the Normal Approximation (Box)

Probability Histograms

A probability histogram represents **chance** by area

The figure below is a probability histogram for the sum of 25 draws from the box $\boxed{\boxed{1}\ \boxed{2}\ \boxed{3}\ \boxed{4}\ \boxed{5}}$. The shaded area represents the chance that the sum will be between ______ and ______ (inclusive).



Empirical Histogram VS Probability Histogram

Empirical Histogram:

- Based on observations, percentage of the times that something comes up
- Can be different due to the draws are different, chance variation
- Converges to the probability histogram

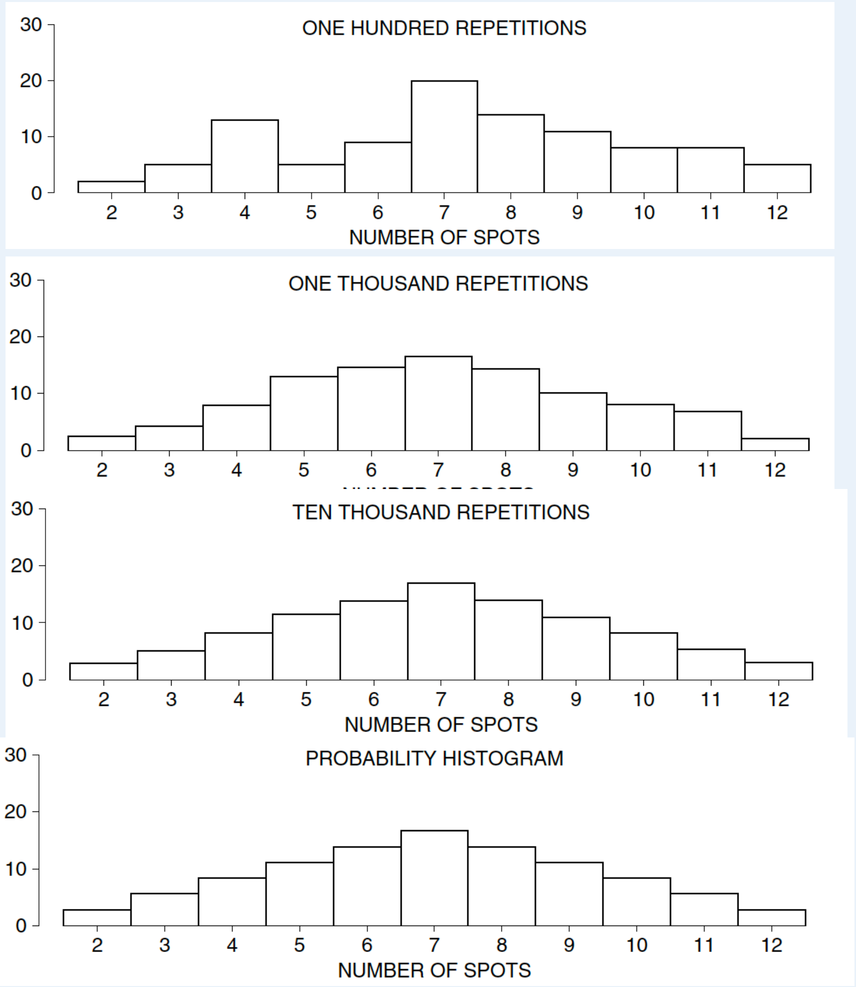
Probability Histograms

Exercises:

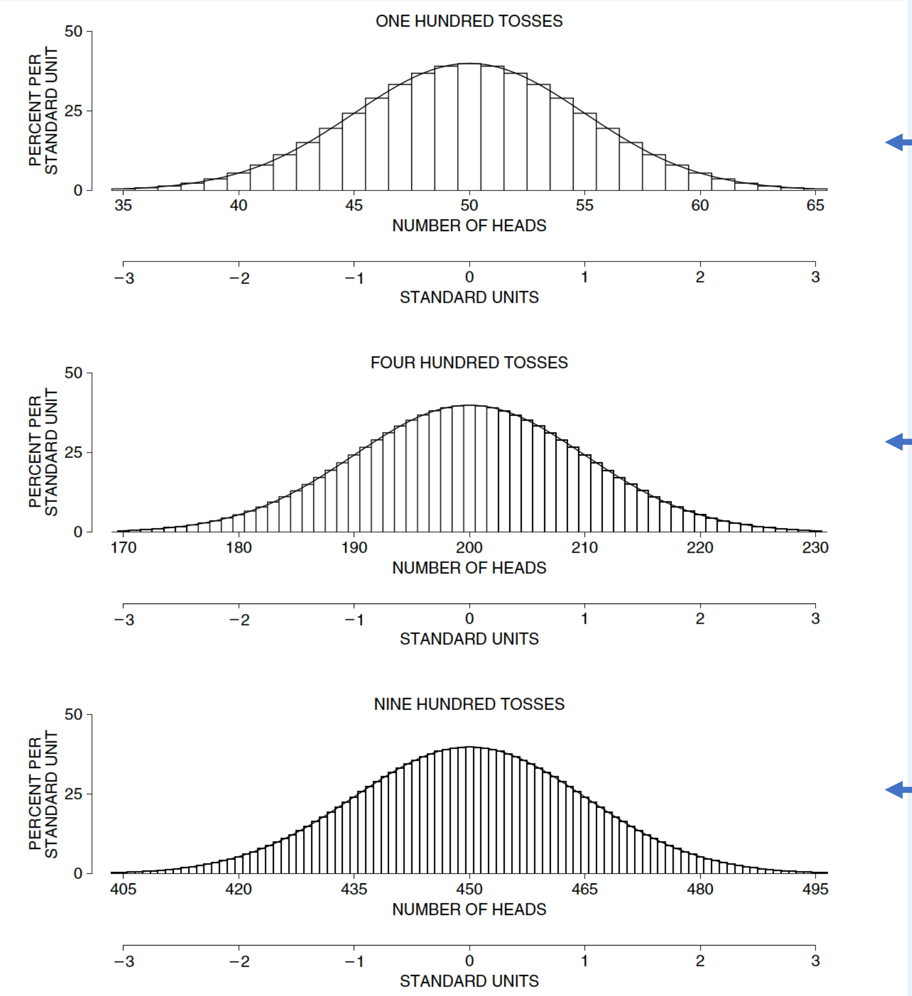
2. The bottom panel of figure 1 (p. 311) shows the probability histogram for the total number of spots when a pair of dice are rolled.
   (a) The chance that the total number of spots will be between 7 and 10 (inclusive) equals the area under the histogram between 6.5 and 10.5.
   (b) The chance that the total number of spots will be 7 equals the area under the histogram between 6.5 and 7.5.

3. This exercise—like exercise 2—refers to figure 1 on p. 311.
   (a) If a pair of dice are rolled, the total number of spots is most likely to be 7.
   (b) In 1,000 rolls of the pair of dice, which total came up most often? 7, the tallest bar
   (c) In the top panel of figure 1, the rectangle over 4 is bigger than the rectangle over 5. Is this because 4 is more likely than 5? Explain. No. chance variation
   (d) Look at the top panel of the figure. The area of the rectangle above 8 represents—
      (i) the chance of getting a total of 8 spots when a pair of dice are rolled.
      (ii) the chance of getting a total of 8 spots when 100 dice are rolled.
      (iii) the percentage of times the total of 8 comes up in table 2.

   Choose one option, and explain. iii, the top panel is an empirical histogram



Probability Histograms and the Normal Curve



N = 100

N = 400

N = 900

As the number of tosses goes up, the probability histogram follows the normal curve better

The Normal Approximation

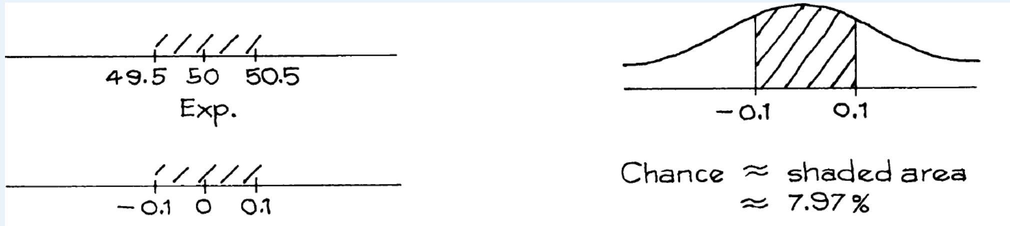
Idea: Since the histogram follows the normal curve so closely, we can use normal curve to **approximate** the chance

Note: it is an approximation, not giving you the exact chance.
In order to obtaining the exact chance, look for the probability histogram

Eg: get the area under the histogram (chance) between a and b, can be replaced by area under the normal curve between the corresponding values (in standard units).

Example 1. A coin will be tossed 100 times. Estimate the chance of getting—

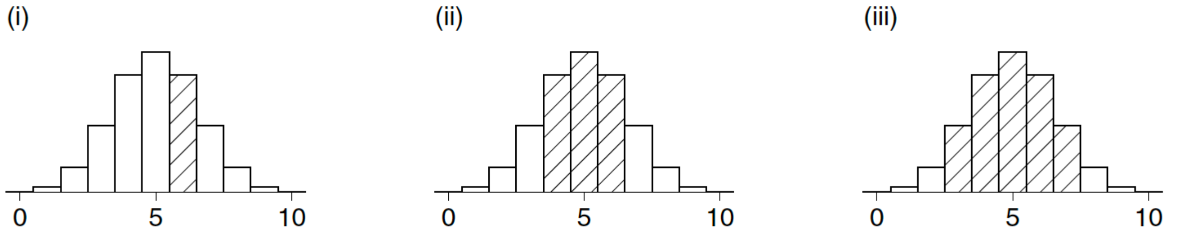
(a) exactly 50 heads.
(b) between 45 and 55 heads inclusive.
(c) between 45 and 55 heads exclusive.



The Normal Approximation

Exercises:

1. A coin is tossed 10 times. The probability histogram for the number of heads is shown at the top of the next page, with three different shaded areas. One corresponds to the chance of getting 3 to 7 heads inclusive. One corresponds to the chance of getting 3 to 7 heads exclusive. And one corresponds to the chance of getting exactly 6 heads. Which is which, and why?



3. A coin is tossed 100 times. Estimate the chance of getting 60 heads.

5. A coin is tossed 10,000 times. Estimate the chance of getting—
   (a) 4,900 to 5,050 heads
   (b) 4,900 heads or fewer
   (c) 5,050 heads or more

6. (a) Suppose you were going to estimate the chance of getting 50 heads or fewer in 100 tosses of a coin. Should you keep track of the edges of the rectangles?
   (b) Same, for the chance of getting 450 heads or fewer in 900 tosses.

The Scope of the Normal Approximation

The normal approximation works perfectly well for the box model as well

Note:
The more the histogram of the numbers in the box differs from the normal curve, **the more draws** are needed before the approximation takes hold.

Eg: the sum of draws from the box