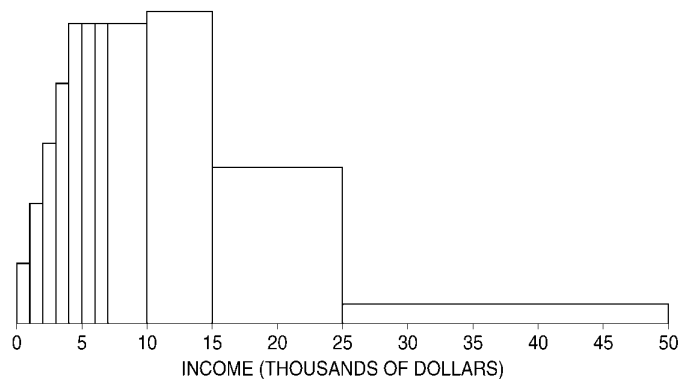


- **Chapter 18: The Normal Approximation for Probability Histograms**

- The histograms we saw in Chapter 3 are called “empirical” histograms because they represent **data**. The area above a class interval represents the **percentage of data values** falling into that class interval. For example:

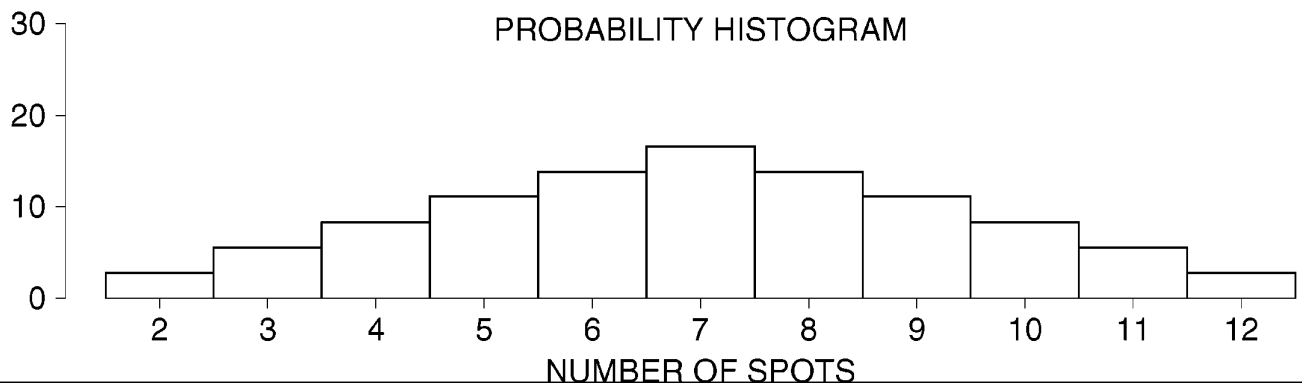
Figure 1. A histogram. This graph shows the distribution of families by income in the U.S. in 1973.



Source: Current Population Survey.

- **Probability Histograms**

- **Probability histograms** represent *chance*. The area above a class interval represents the *chance* for that class interval.
- Example: roll two dice and look at the total number of spots. A **probability histogram** for this chance process is given by:



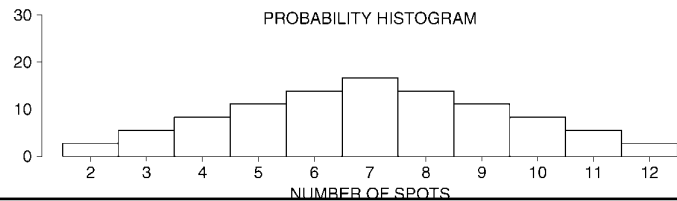
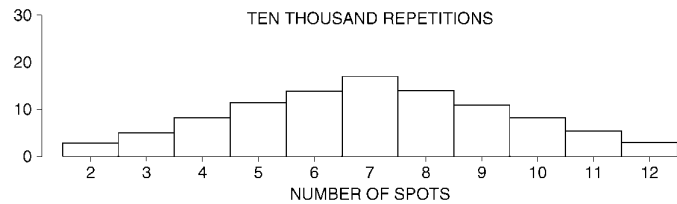
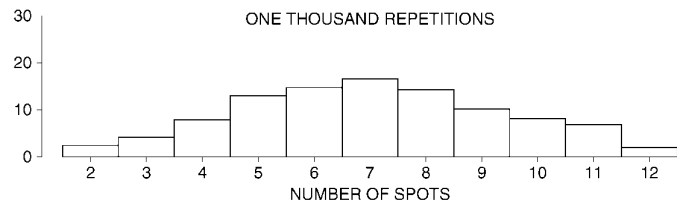
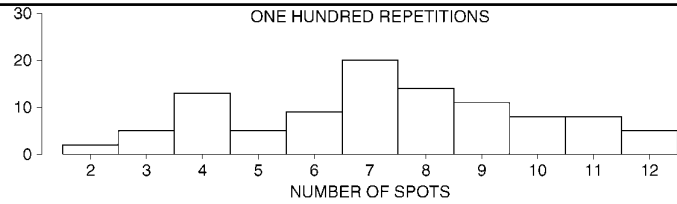
- We can get a ***probability histogram*** for a chance process by finding the chances and doing a histogram.
- We can get an ***empirical histogram*** for a chance process by repeating the process many times and doing a histogram.
- The more repetitions we do, the closer the ***empirical histogram*** will get to the ***probability histogram***.

- Roll two dice 100 times and look at the number of spots:

<i>Repe-</i> <i>tion</i>	<i>Total</i>	<i>Repe-</i> <i>tion</i>	<i>Total</i>	<i>Repe-</i> <i>tion</i>	<i>Total</i>	<i>Repe-</i> <i>tion</i>	<i>Total</i>	<i>Repe-</i> <i>tion</i>	<i>Total</i>
1	8	21	10	41	8	61	8	81	11
2	9	22	4	42	10	62	5	82	9
3	7	23	8	43	6	63	3	83	7
4	10	24	7	44	3	64	11	84	4
5	9	25	7	45	4	65	9	85	7
6	5	26	3	46	8	66	4	86	4
7	5	27	8	47	4	67	12	87	7
8	4	28	8	48	4	68	7	88	6
9	4	29	12	49	5	69	10	89	7
10	4	30	2	50	4	70	4	90	11
11	10	31	11	51	11	71	7	91	6
12	8	32	12	52	8	72	4	92	11
13	3	33	12	53	10	73	7	93	8
14	11	34	7	54	9	74	9	94	8
15	7	35	7	55	10	75	9	95	7
16	8	36	6	56	12	76	11	96	9
17	9	37	6	57	7	77	6	97	10
18	8	38	2	58	6	78	9	98	5
19	6	39	6	59	7	79	9	99	7
20	8	40	3	60	7	80	7	100	7

The sum of 2 rolls.

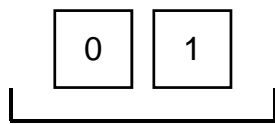
The more repetitions we do, the closer the *empirical histogram* will get to the *probability histogram*.



- **The Normal Curve**

- The *probability histogram* for the
 - **sum of the draws**
- will approximately follow the normal curve if the number of draws is large enough, even if the tickets in the box do not follow the normal curve!

Standard units: use EV for sum and SE for sum



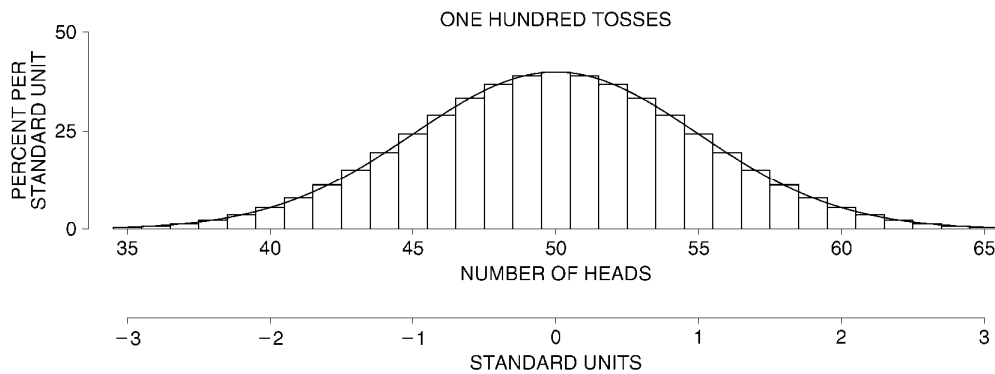
Box AV = 0.5

Box SD = 0.5

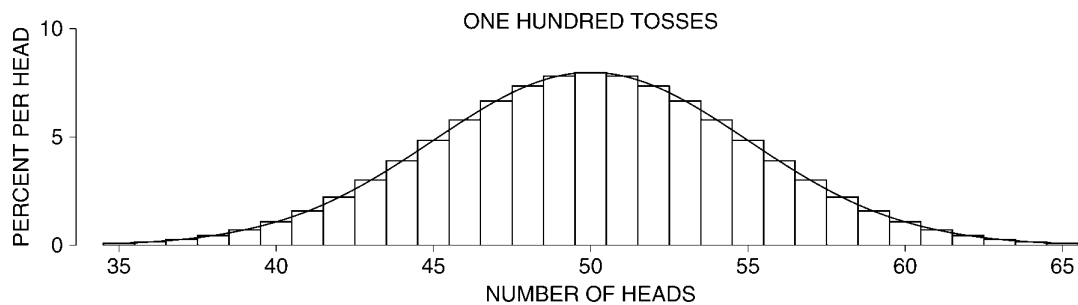
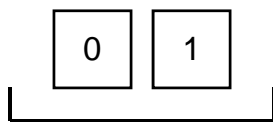
EV for sum = $(0.5)(100) = 50$

SE for sum = $(0.5)(\sqrt{100}) = 5$

$$z = \frac{x - EV \text{ for sum}}{SE \text{ for sum}}$$



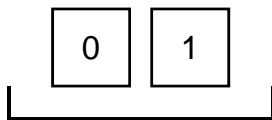
- The probability histogram and the normal curve for the number of heads when we toss a coin 100 times. This is the sum of 100 draws from the box:



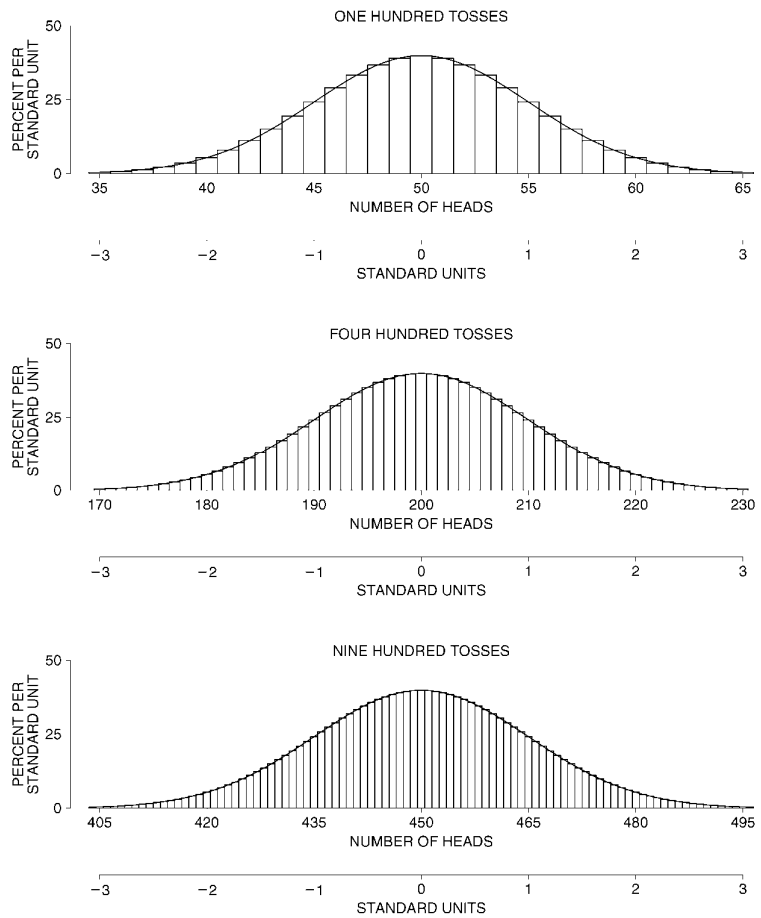
The sum of the draws.

The more draws we do, the closer the *probability histogram* will get to the *normal curve*.

The box:



histogram for the box:



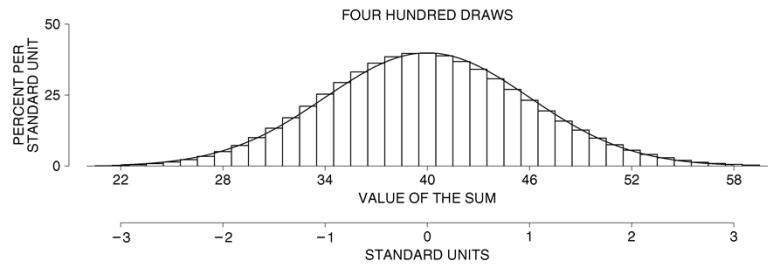
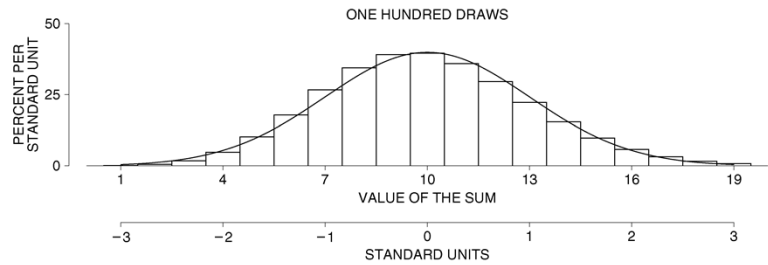
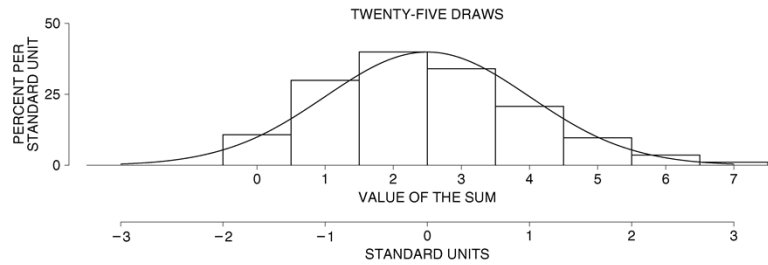
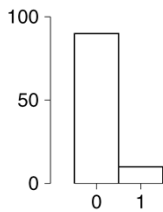
The sum of the draws.

The more draws we do, the closer the *probability histogram* will get to the *normal curve*.

The box:

[0,0,0,0,0,0,0,0,0,0,1]

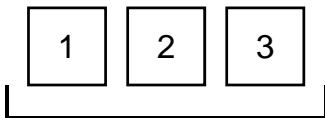
histogram for the box:



The sum of the draws.

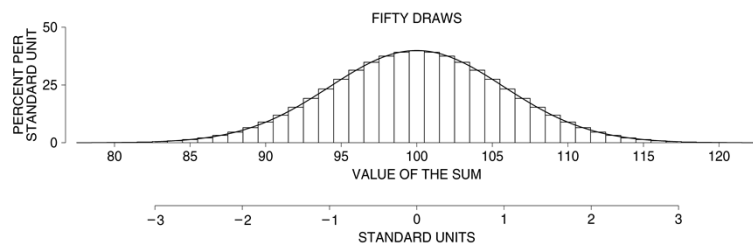
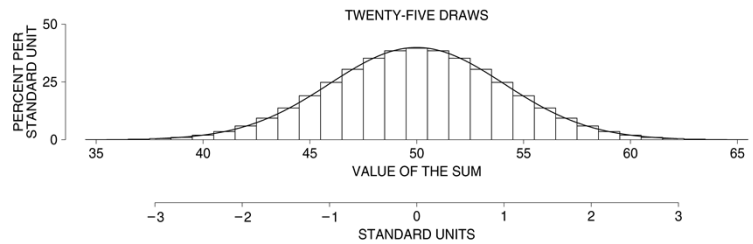
The more draws we do, the closer the **probability histogram** will get to the **normal curve**.

The box:



histogram for the box:

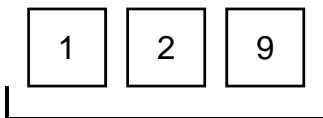
Figure 7. Probability histograms for the sum of 25 or 50 draws from the box [1 2 3]. These histograms follow the normal curve very well.



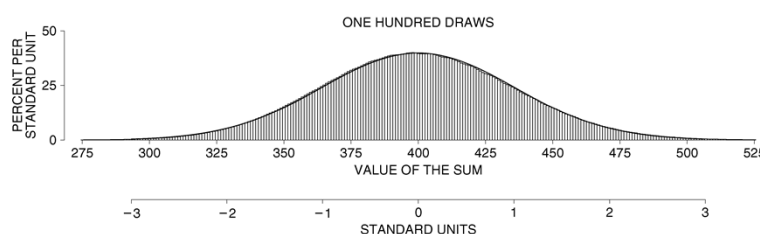
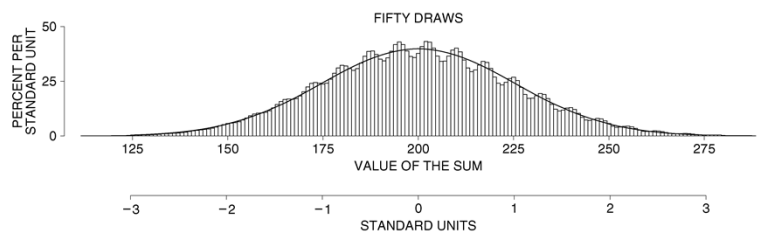
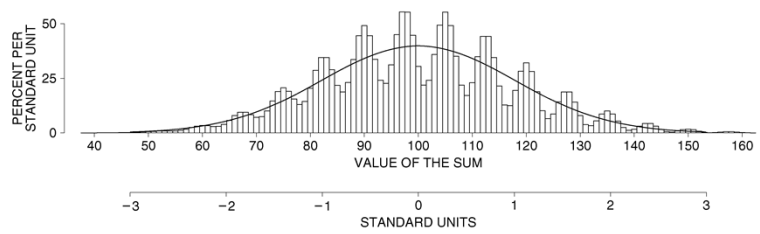
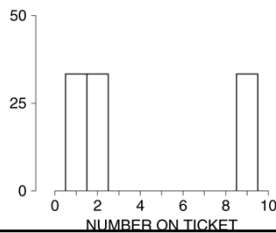
The sum of the draws.

The more draws we do, the closer the **probability histogram** will get to the **normal curve**.

The box:

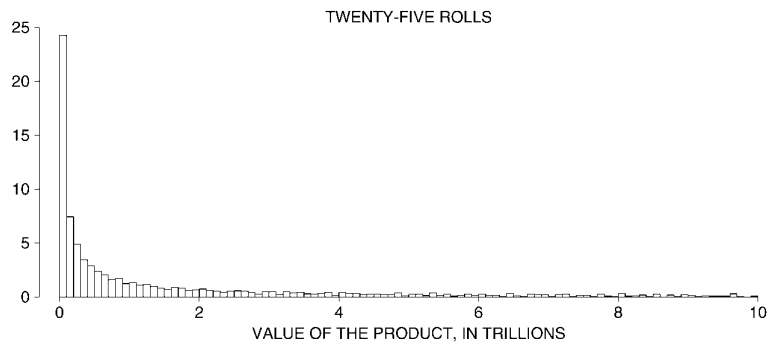
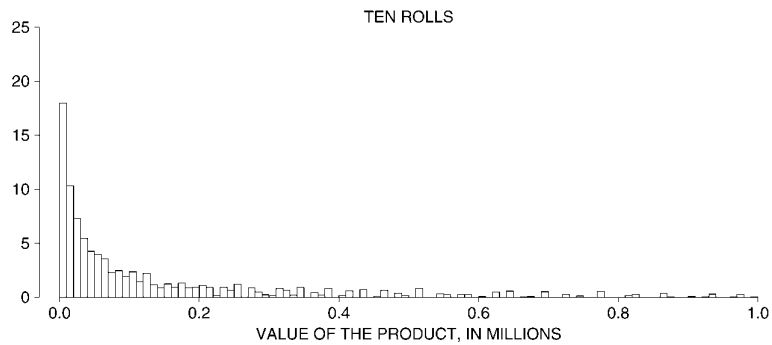


histogram for the box:



Not everything is normal!

Sums, averages, percentages are normal. Products are not!



"Life is like drawing from a box of chocolates."

And sometimes you have a pretty good idea of what you are going to get.

Normal Approximation Theorem or Central Limit Theorem:

[..... π , $\sqrt{2}$, 1943, e , 3,]

Draw n times with replacement and consider the sum (or average) of the draws.

When drawing at random with replacement from a box, the probability histogram for the sum will follow the normal curve, even if the contents of the box do not. The histogram must be put into standard units, and the number of draws must be reasonably large.

Stat 1040, Normal Approximation

1. Given the box $[1 , 1 , 0 , 0 , 0 , 0 , 0 , 0 , 0 , 0]$, draw 100 times with replacement and consider the sum of the draws.

a) What is the EV for the sum of the draws?

b) Find the SE for the sum of the draws.

c) Find the probability that the sum of the draws is between 16 and 22. What is the chance that it is greater than 28 ?

2. Current registration in a local voting district shows that 80 percent of the voters are Republican with 20 percent Democrat. If you randomly sample (with replacement) 100 voters, find the chance of getting more than 28 Democrats in your sample.

3. Toss a coin 100 times. Find the probability of getting more than 60 heads.

4. A die is rolled 180 times.

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

b) The number of aces should be around _____, give or take _____ or so.

c) Find the chance of getting less than 40 aces.