Ch. 18: Normal Approximation for Probability Histograms

A _____ is a histogram of chances and not of data.

For a sum of draws, each possible value for the sum has a rectangle (of width 1) centered on the value.

The area of the rectangle is the chance of that value.

If we repeat a chance process several times, we can construct an _____ of the observed outcomes.

The more repetitions we do, the more the empirical histogram will look like the probability histogram.

Ex: We toss 3 coins and count the number of heads.

What is the corresponding box model?

What is the expected number of heads?

What is the standard error?

We repeat this experiment 10 times.

How does the empirical histogram look like?

How does the probability histogram look like?

The Central Limit Theorem

The probability histogram of a _____ of a large number of independent draws from a box follows the normal curve even if the contents of the box do not look like a normal curve, provided that the number of draws is large.

To convert to standard units, we use EV_{sum} and SE_{sum}.

How large is "large"?

The answer depends on the box.

- If the box is symmetric, or nearly so, we can get a pretty good approximation by 30 or even less draws.

- If the box is somewhat asymmetric, the number of draws will have to be larger to get a good approximation.

- 100 draws will give a pretty good approximation to all but *extremely* asymmetric boxes.

Note:
The normal curve is associated with the _____.

Other functions of many draws (such as _____) will usually *not* follow a normal curve.

Summary

- If we draw many times from a box, the probability histogram of the sum will follow a normal curve, even if the box does not.
- The EV fixes the center of the distribution, and the SE fixes the spread. To calculate probabilities, we should put our results in standard units.
- To calculate the EV and SE requires: the number of draws, the average of the box, and the SD of the box.

Ex: Suppose we are interested in rolls of a standard, six-sided die. Are the following well represented by the normal curve?

1. The empirical histogram of 1000 rolls?

2. The probability histogram of the sum of 50 rolls?

3. The empirical histogram of 1000 sums, each of 50 rolls?

4. The probability histogram of the product of 50 rolls?

5. The empirical histogram of 1000 products, each of 50 rolls?