Honors Class02 Activity: Data to Die For...

Team Members: ____

Learning Objectives

- Plotting data by hand
- Probability Distributions
- Sample distributions and sample size

Activity Introduction [1]

This activity introduces several concepts we will revisit throughout the semester, including probability, and sample size. Just for this one activity, you'll be plotting the data by hand because, hey, everyone needs to remember what it was like before computers!

If you roll a single die, there are six possible outcomes: S = {1, 2, 3, 4, 5, 6}. If the die is not loaded these outcomes are equally likely, so each has a one in six chance of occurring or a probability of 1/6, where 0 means no chance and 1 means 100% certainty. Because each outcome is equally likely, the distribution of probabilities is "uniform."

Sample Space: $\boldsymbol{S} = \{(1, 1), \ldots, (6, 6)\}$ $\#(\boldsymbol{S}) = 6^2 = 36$

(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

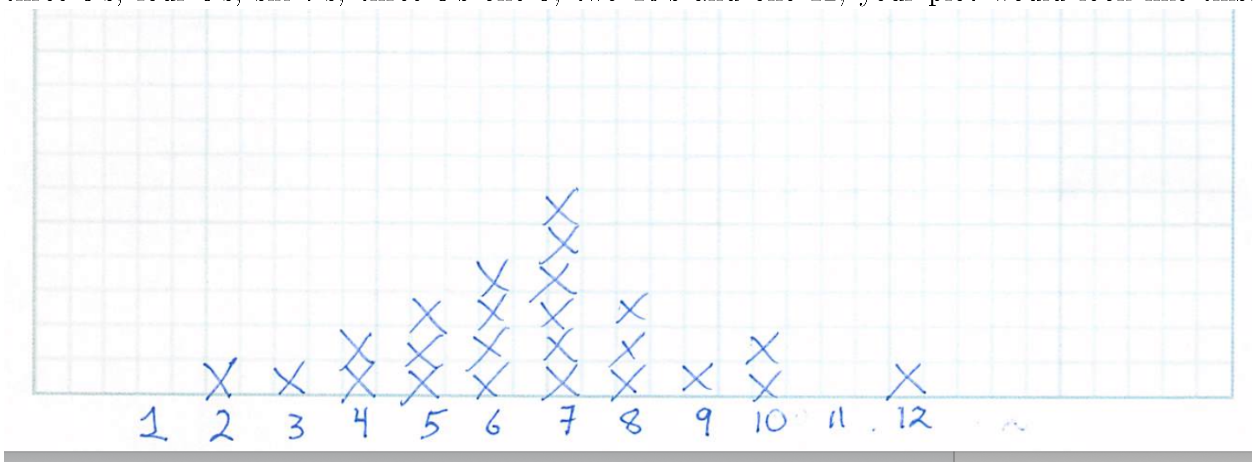
Figure 1: Sample space

[1]This activity and the associated graphic were adapted in large part from the excellent introduction: http://pages.stat.wisc.edu/~ifischer/Intro_Stat/Lecture_Notes/4_-_Classical_Probability_Distributions/4.1-_Discrete_Models.pdf

Experiment

The sum of the two dice is any integer between 2 and 12, but the probability distribution is no longer uniform. Let's find it experimentally.

On the graph paper provided, you will plot the frequency distribution, which is the number of times each value occurred in the data sample. For example, if you had one 2, one 3, two 4's, three 5's, four 6's, six 7's, three 8's one 9, two 10's and one 12, your plot would look like this:



Data from 50 rolls of a pair of dice

7	11	5	3	8	6	6	9	5	4
9	7	10	5	7	6	6	7	10	8
7	4	6	8	11	4	6	5	7	9
10	8	5	8	12	9	7	5	6	10
9	4	6	3	8	7	9	8	7	7
3	6	5	11	7	5	8	8	7	6
4	10	7	9	6	5	8	3	7	7
6	9	10	4	9	7	11	8	8	7
6	7	9	6	7	8	6	5	8	5
7	9	9	9	5	9	9	10	6	9

Figure 2: Graph your data here

What you have just created is a data histogram, a way to visualize the distribution of a data sample. Once you learn some Python coding, you'll no longer have to plot these by hand. (And there was much rejoicing.)

Based on your histogram, what is the most likely sum when you roll two dice?

Least likely?

Here's the histogram from a 10,000-roll simulation, which now closely matches the expected probability distribution for the sum of two fair dice.

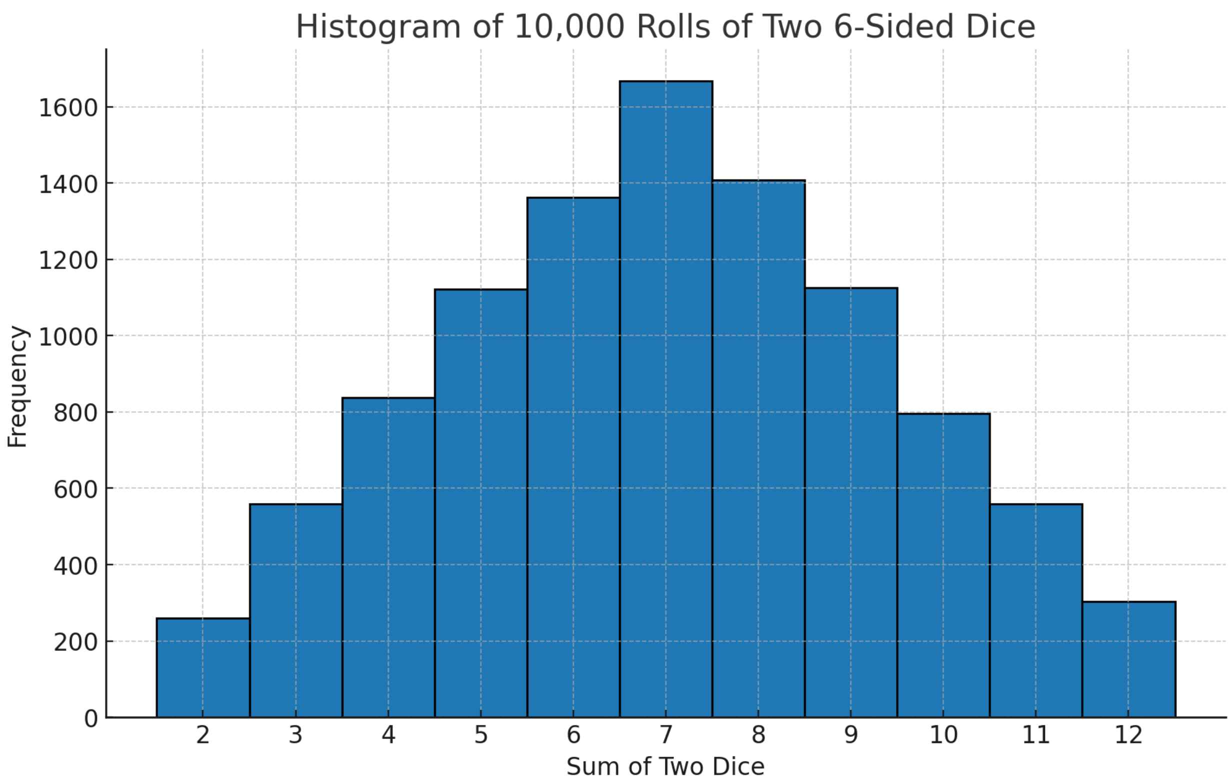


Figure 3: Histogram for 10,000 rolls

Based on a simulation of 10,000 rolls, the most likely sum is 7.

Discussion Questions

When you rolled two dice, you were no longer working with a uniform distribution. **What changed between rolling one die and rolling two dice?**

Look at your hand-drawn frequency histogram. **Are the peaks more reliable than the tails, or vice versa?**

Your histogram might not show 7 as the most frequent outcome, even though the 10,000-roll simulation does. **Does that mean your data contradicts probability theory?**

What would it mean, statistically, to say your result is "wrong"?

Suppose you repeated this activity with:

- 30 rolls
- 300 rolls
- 3,000 rolls

At which point would you feel comfortable predicting the most likely sum for future rolls? Why?

This activity is about dice — but what other kinds of real data behave similarly? Consider, for example, medical trials, or polling data.