**7.SP Waiting Times**

**Alignments to Content Standards:** 7.SP.C.8

**Task**

Suppose each box of a popular brand of cereal contains a pen as a prize. The pens come in four colors, blue, red, green and yellow. Each color of pen is equally likely to appear in any box of cereal. Design and carry out a simulation to help you answer each of the following questions.

a. What is the probability of having to buy at least five boxes of cereal to get a blue pen? What is the mean (average) number of boxes you would have to buy to get a blue pen if you repeated the process many times?

b. What is the probability of having to buy at least ten boxes of cereal to get a full set of pens (all four colors)? What is the mean (average) number of boxes you would have to buy to get a full set of pens if you repeated the process many times?

**IM Commentary**

As the standards in statistics and probability unfold, students will not yet know the rules of probability for compound events. Thus, simulation is used to find an approximate answer to these questions. In fact, part b would be a challenge to students who do know the rules of probability, further illustrating the power of simulation to provide relatively easy approximate answers to wide-ranging problems. Modeling with simulation follows four steps: state assumptions about how the real process works; describe a model that generates similar random outcomes; run the model over many repetitions and record the relevant results; write a conclusion that reflects the fact that the simulation is an approximation to the theory.
**Solution**

a. If each color of pen is equally likely to appear in any box, the chance of getting a blue pen in any one box is $\frac{1}{4}$ or 0.25. Simulation is then used to find an approximate answer to the question posed. Students select a device, or devices, that generate a specified outcome with probability 0.25 to model the process of buying boxes of cereal until a blue pen is found. Random integers 1, 2, 3, 4, with, say, 1 denoting blue, will work (as will using four sides of a six-sided die, etc.). They then generate many outcomes for the simulated event and collect the data to produce a distribution of waiting times.

Here is a string of random integers that produces 9 trials of a simulation, the respective waiting times to get a 1 (blue) being 2, 5, 4, 1, 4, 3, 4, 3, 3.

4 1] 4 3 3 2 1] 4 2 2 1] 1] 3 2 2 1] 2 4 1] 3 2 4 1] 2 2 1] 2 3 1] ...

The plot below is based on 100 simulated waiting times (Tb) to get a blue pen. The probability of having to purchase at least 5 boxes is approximated by the proportion of simulated waiting times greater than or equal to 5, which is $(33/100) = 0.33$. The mean of the 100 simulated waiting times is 3.8 or approximately 4. It should seem intuitively reasonable that an event with probability $\frac{1}{4}$ would happen, on average, about every four trials.

![Histogram of Waiting Times](image.png)

Mean = 3.8
b. Modeling the outcome of getting a full set of pens (all four colors) works in a similar way. Using the same sequence of random integers as above, the waiting times are 6, 7, 7, 8.

\[4 \ 1 \ 4 \ 3 \ 3 \ 2 \ 1 \ 4 \ 2 \ 2 \ 1 \ 1 \ 3 \ 2 \ 2 \ 1 \ 2 \ 4 \ 1 \ 3 \ 2 \ 4 \ 1 \ 2 \ 2 \ 1 \ 2 \ 3 \ 1 \ ...\]

The plot below is based on 100 simulated outcomes (T4) resulting in a full set of pens. The probability of having to purchase at least 10 boxes to get a full set is approximated by the proportion of waiting times greater than or equal to 10, which is \(\frac{31}{100} = 0.31\) for this simulation. The mean of the 100 simulated waiting times is 8.2, which is not so intuitive.

![Histogram of Waiting Times](image)

Mean = 8.2

Even though students will not yet have the tools to figure this out, it is worth noting that the theoretical solution is \(\frac{1}{4} + \frac{1}{3} + \frac{1}{2} + \frac{1}{1} = 8\frac{1}{3}\). The results of the simulation agree well with this.