MTH245 Unit 3 Module 4 Random Variables and Probability Distributions

The last Module was all "how to", now we had better talk about "why" we would want to create those pivot tables and histograms. In the video of examples from Unit 3 Module 3 we created this table:

Row Labels	Count of Score
15	34.66%
20	29.30%
25	8.76%
30	15.04%
35	8.38%
40	1.02%
45	2.10%
50	0.58%
60	0.16%
Grand Total	**100.00%**

The first column is a list of all possible score totals when three darts are thrown with possible individual scores of 5, 10 and 20. The first column is our **sample space** for the experiment, a list of all possible outcomes. The second column is the percentage of times we got each of those scores in 5000 experiments. The second column is the (estimated) probability of each of those events. Together the two columns are an **Empirical Probability Distribution** for the experiment of throwing the three darts. These are also called **probability tables**.

A **random variable** is a function that assigns a **numerical** value to each event in a sample space. (Not relevant in all situations.)

Experiment	Define x	Sample Space
Roll a die twice	x = Sum of the 2 #s	x = 2, 3, …, 11, 12
Toss a coin 3 times	x = number of heads	x = 0, 1, 2, 3

If we have an experiment where the sample space is {red, green, blue} it would not make sense to be talking about random variables, this is only used if we have numerical results.

A **probability table** for a random variable lists all possible values of x and the probability of each of those values occurring. The sum of those probabilities must be 100% (or 1)! Why? Because we are listing all possible outcomes for an event with disjoint outcomes.

Example: Construct a probability table for x, the number of heads when a coin is tossed 3 times.
To find the theoretical probability table we would first list the sample space:
S = HHH, HHT, HTH, THH, TTH, THT, HTT, TTT
Are these equally likely? Yes
How many outcomes are there? 8
How many of the outcomes result in 0 heads? 1 (TTT)
1 Head? 3 (TTH, THT, HTT)
2 Heads? 3 (HHT, HTH, THH)
3 Heads? 1 (HHH)

From that we can construct the theoretical probability table:

x	P(x)
0	1/8 = .125
1	3/8 = .375
2	3/8 = .375
3	1/8 = .125

Note that the probabilities add to 1. Because P(0 heads) + P(1 head) + P(2 heads) + P(3 heads) = P(0 heads or 1 head or 2 heads or 3 heads) and if we flip 3 coins it is guaranteed that one of those outcomes will occur.

Alternatively, we can set up an experiment in Excel and construct an empirical probability table:

This was based on a sample of 1000, and you can see that the probability of each outcome is within 1% of the theoretical results.

Row Labels	Count of # of heads
0	11.90%
1	37.20%
2	38.90%
3	12.00%
Grand Total	**100.00%**

Using a sample size of 5000 the results were within 0.6% of the theoretical results:

Row Labels	Count of # of heads
0	12.22%
1	37.88%
2	36.86%
3	13.04%
Grand Total	**100.00%**

Once we have created a probability table for an experiment we can use it to answer questions about the experiment:

If three coins are tossed what is the probability of more than one head?
P(more than one head) = P(1, 2 or 3 heads) + P(1) + P(2) + P(3) = .3788 + .3686 + .1304 =.8778
Alternatively, P(x >=1) = 1 – P(0) = 1 – .1222 = .8778

A girl scout is going door to door selling cookies. When she knocks on a door, there is a 30% chance that someone will answer and buy a box of cookies. If she knocks on 2 doors, what is the probability that she sells cookies at both houses?
P(sells at house 1 and sells at house 2) = P(sell H1) * P(sell H2) = (0.3)(0.3) = 0.09

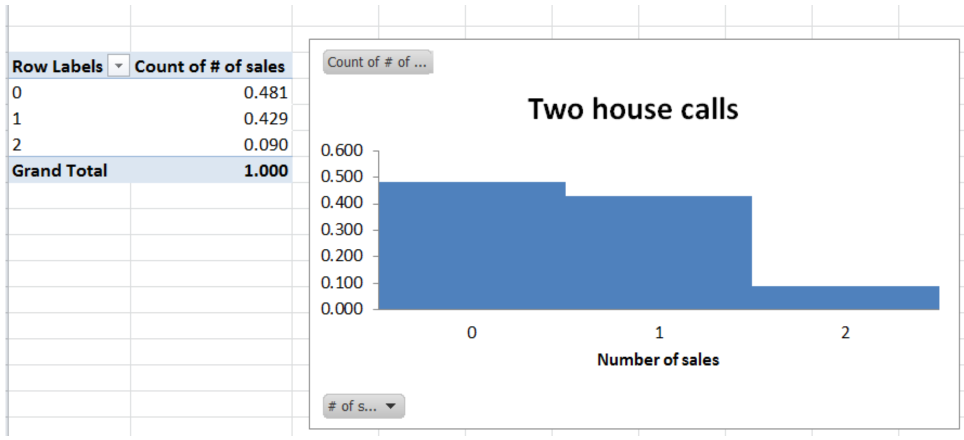
What is the probability she sells at neither house?
P(sells at neither house) = P(no sell H1) * P(no sell H2) = (0.7)(0.7) = 0.49

The third option is that she makes exactly 1 sale, either at the first but not the second or at the second but not at the first. This probability can be calculated two ways.
The short cut would be to note that:
P(2 sales) + P(1 sale) + P(0 sales) = 1, so P(1 sale) = 1 – 0.09 - .49 = 0.42.
Alternatively:
P(1 sale) = P(sell H1)*P(no sell H2) + P(no sell H1)*P(sell H2) = (.3)(.7)+(.7)(.3) = .42

Then the theoretical probability distribution for the random variable x = number of sales made will be:

x = number of sales	P(x)
0	0.49
1	0.42
2	0.09

Running the same problem as an experiment in Excel with 2500 iterations resulted in an empirical probability distribution of:

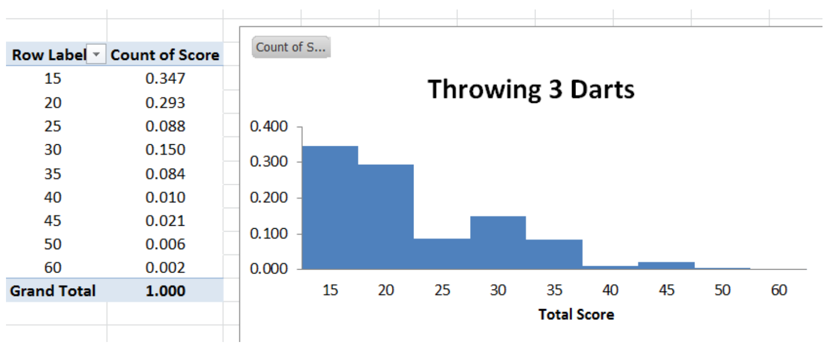


Row Labels	Count of # of sales
0	0.481
1	0.429
2	0.090
Grand Total	**1.000**

With empirical distributions the results are not exact, but with a large number of iterations they will be very close to the theoretical distributions.

What does the histogram tell us about the probability distribution? Across the horizontal (x) axis we have a list of all possible outcomes of the experiment–the sample space. The height of each bar, or the area under the bar, is the probability of each event occurring. The total shaded area under the bars totals 1 (or 100%).

From the dart throwing example in the video for Unit 3 Module 3 we have the distribution:



Row Label	Count of Score
15	0.347
20	0.293
25	0.088
30	0.150
35	0.084
40	0.010
45	0.021
50	0.006
60	0.002
Grand Total	**1.000**

The histogram is a visual display of the information in the table. Had you noticed before that the probability of scoring 30 points is higher than that of scoring 25 points? The histogram can make some information pop out that might be overlooked.

If you are interested, the theoretical distribution of this experiment is given. It was calculated using formulas beyond the scope of this class.

x = total score on 3 darts	P(x)
15	.343
20	.294
25	.084
30	.155
35	.084
40	.012
45	.021
50	.006
60	.001