6. Sensitivity Analysis

Up to now, we have treated the parameters in the Net Value Function (NVF) as a single known value. However, there are many instances where we do not know the exact value of a parameter. Rather, we may know that the parameter falls within a range of value at a certain probability. For example, we may not know what the exact price of gasoline is going to be tomorrow, but we may estimate based on historical data that the price has a 40% chance of being $1.54/L tomorrow. How can we capture this uncertainty in our NVF?

Sensitivity analyses study how the model's (i.e. NVF) output can be affected by variation of the model's inputs (i.e. NVF parameters). To put it more simply, sensitivity analysis concerns itself with the hypothetical question "What if?". We are aiming to gain understanding of a problem. Specifically, how would a certain change in a NVF parameter affect the NVF. For example, what if the cost of production changes? What if a tax was imposed on our product? What if a competitor enters the market? Questions such as these, regarding how outputs change given variation in one or more inputs, are the basis for sensitivity analyses. This can be used to determine whether we should adjust MARR due to project risks or if there are critical measures to be taken to mitigate those risks before proceeding. This can also extend as far as having "risk factors" for the chance employees are injured on the job, for example. An oil driller has a higher chance of injury compared to an actuary, and the risk that goes into covering for injured workers, providing payouts, etc. can also be accounted for.

6.1. Applications of Sensitivity Analysis

1. Analyzing Uncertainty
    - How much of the output uncertainty is **irreducible**; in other words, how much of the output uncertainty is caused by "random" uncertainty in the input parameters (uncertainty dependent on chance)?
    - How much of the output uncertainty is **epistemic**; in other words, how much is related to a lack of knowledge that could be reduced by doing more research?
2. Optimization of Research Investment
    - Which input parameters contribute most to the model output uncertainty?
    - Which input parameters are should we spend research money on to reduce the most amount of uncertainty?
3. Model Reduction
    - Identifying and removing parameters that have little effect on the model output.

6.2. Deterministic Models

6.2.1. Definition

A **deterministic** model (as opposed to a **stochastic** one) is one whose outputs are dependent only on the inputs (and not on randomness or external factors we're not considering as inputs). That is, for a given set of inputs, deterministic models *determine* the output with 100% certainty

6.2.2. . Net Value Example, Deterministic Model

Let's perform a sensitivity analysis on a simple example. You are the owner of a lemonade stand. Your dream is to have a lemonade stand empire, but every great company has a humble beginning. For simplicity, the NVF for the lemonade stand has the following parameters: cost of ingredients and supplies (*I*), cost of labour (*L*), and revenue (*R*) (for now, we will assume the cost of space is zero since you are operating in your front yard). Given these parameters, the NVF is:

$$NV = R - I - L$$

Up to now in our course we have assumed a single value for each parameter. For example, the average cost for ingredients and supplies for 1 cup of lemonade is $0.10, the cost of labour is $15/hour (the value you give for your own time), and the average revenue per week is $400 assuming we sell 200 cups per week at $2/cup (based on your "market research" and your understanding of supply and demand). You decide to run the lemonade stand for 8 weeks in July and August. Each week you will open the stand for 2 days (only weekends) for 8 hours/day.

Using the information above we can calculate the net value of the lemonade stand:

$$R = \frac{\$400}{\text{week}} \times 8 \text{ weeks} = \$3200$$

$$I = \frac{\$0.10}{\text{cup}} \times 200 \frac{\text{cups}}{\text{week}} \times 8 \text{ weeks} = \$160$$

$$L = \frac{\$15}{\text{hour}} \times 8 \frac{\text{hours}}{\text{day}} \times 2 \frac{\text{days}}{\text{week}} \times 8 \text{ weeks} = \$1920$$

Therefore, the net value of the lemonade stand is,

$$NV = R - I - L = \$3200 - \$160 - \$1920 = \$1120$$

6.2.2.1. Scenario Analysis

The NVF above did not account for any variations in the parameters. What if it rains for one of the weekends? What if one of the weekends was very hot and sunny? What if you find a supplier that lets you buy lemons in bulk for a much cheaper price? These questions are examples of possible uncertainties affecting the NVF. One way that we can account for these uncertainties is to set values for parameters for best- and worst- case scenarios. Note, we are not considering the probabilities of how likely each case may happen, we will look at probabilities in the Stochastic Models section below. For now, we will develop deterministic models for the best- and worst- case scenarios (i.e., the *deterministic model for the best-case scenario* results from substituting the best-case parameter values into the deterministic NVF above to *determine* the resulting NV, but doesn't imply anything about how likely this best-case scenario is to happen or the expected value or risk of the overall project. Doing this for the best- and worst- case scenarios *does* determine the limits of the NV though).

For example, we want to see how revenue will affect the net value of the lemonade stand. We can recalculate NV using different values for revenue. For the worst-case scenario, we can say that this summer is wetter and cooler than usual, and we expect only 100 cups can be sold per week. Now, recalculating the NV we get:

$$NV_{WC} = R_{WC} - I_{WC} - L = -\$400$$

Similarly, for the best-case scenario where the weather is good and hot, and there are a number of community events in the neighbourhood, you may expect to sell 300 cups/week.

$$NV_{BC} = R_{BC} - I_{BC} - L = \$2640$$

By performing the analysis above, we can have an understanding of how sales volume affects the NV and we have a range of possible net values for the lemonade stand. For scenario analysis, we usually change multiple parameters at a time to give us the limits of the model.

For example, the best-case scenario could be that you sell 300 cups/week and that lemons can be bought in bulk so ingredients and supplies only costs you \$0.05/cup. We can calculate NV_{BC}:

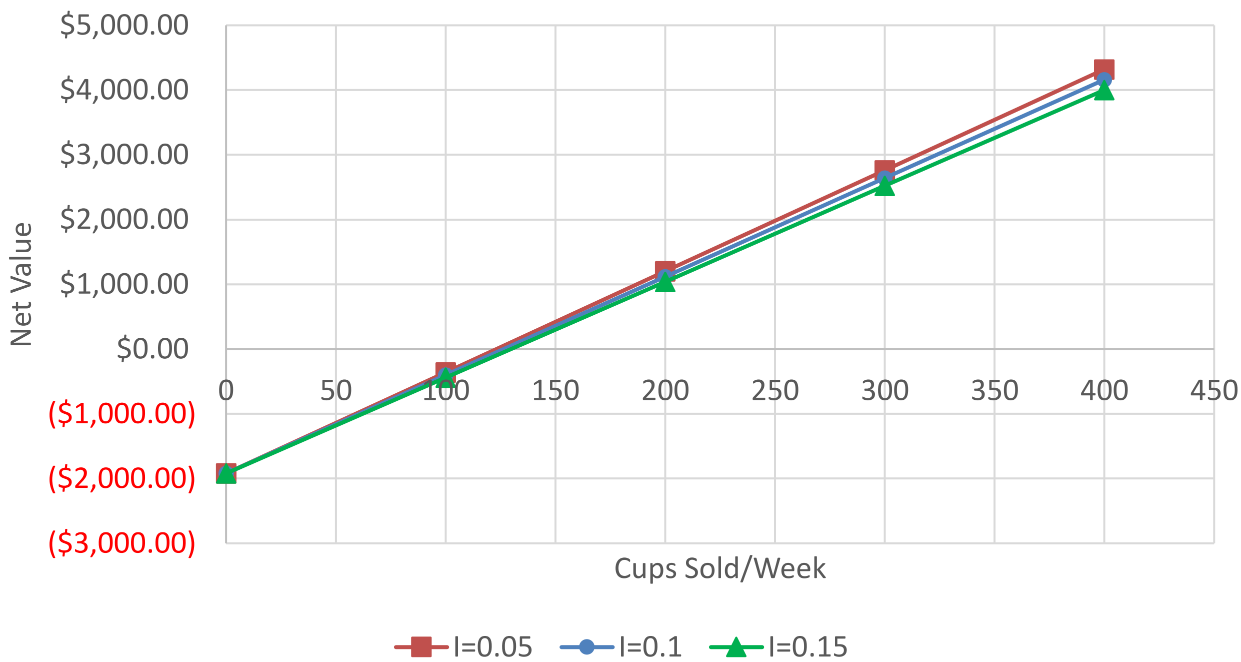
$$NV_{BC} = R_{BC} - I_{BC} - L = \$2760$$

The worst-case scenario could be that you sell 100 cups/week and there is a supply chain issue with lemons so it costs you \$0.15/cup. We can calculate NV_{WC}:

$$NV_{WC} = R_{WC} - I_{WC} - L = -\$440$$

6.2.2.2. Scenario Analysis Visualization

We can visually represent the effect of the parameters on the net value by plotting the net value against the range of possible parameter values as shown below. The graph below also shows how two parameters interact and affect the net value. The top and bottom lines are the limits (best- and worst-case) of the model based on our assumptions.



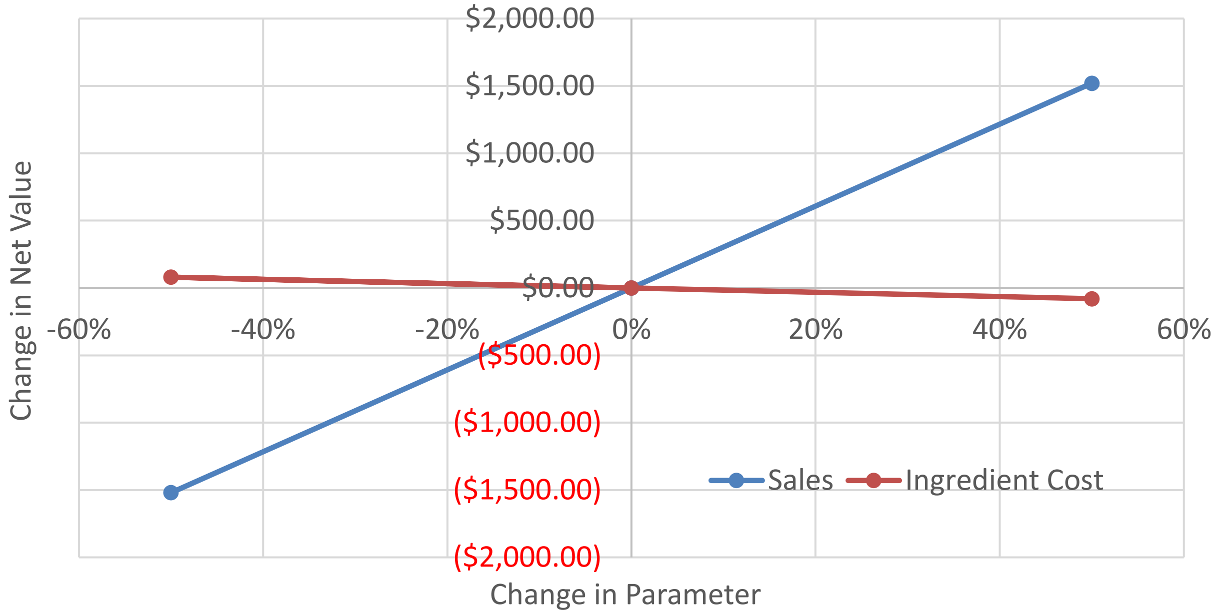
6.2.2.3. Sensitivity Analysis

We can also perform a sensitivity analysis on the model to understand how one specific parameter affect the model outcome. Unlike scenario analysis, sensitivity analysis changes one parameter at a time. This treatment of NVF is sometimes called *Ceteris Paribus*, which means "all other things being equal". To perform Ceteris Paribus analysis, we follow these general steps:

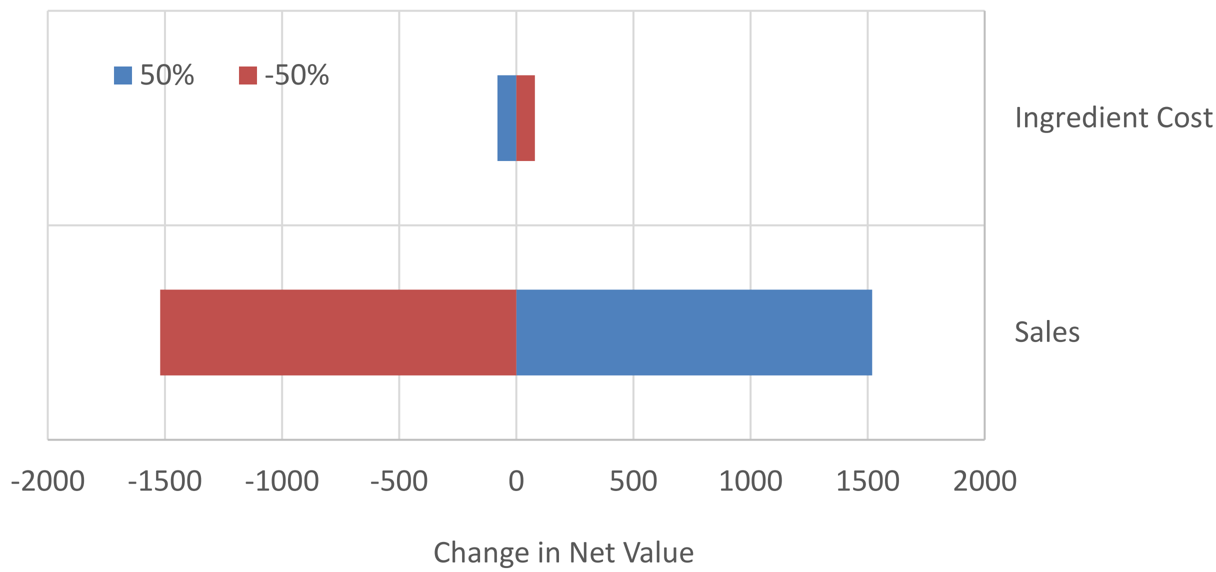
1. Determine the possible range of variation of one parameter,
2. Evaluate the model outcome over this range (choose a mesh of as many points as you like),
3. Determine how (and whether) key conclusions change (profitability, payback period, etc.), and
4. Select decisions that give the lowest and highest expected net value considering probabilities (if conclusions change).

For sensitivity analysis, we relate the change in model output to the change in a specific input parameter. For example, increasing the number of cups sold by 50% increases NV by $1520 or 136%. Increasing the ingredient cost by 50% decreases NV by $80 or 7.1%. Therefore we can say that the model (i.e. NV) is more sensitive to changes in cups sold versus changes in ingredient cost. Based on this analysis, we may want to spend more effort in reducing the uncertainty in project sales. Furthermore, if the model gets more complicated (i.e. more input variables), it may be acceptable to remove ingredient cost as an input parameter if we deem it to have an insignificant effect on the model.

We can also visually show the sensitivity of the net value to each parameter. The spider plot below shows that the net value has a much greater sensitivity to changes in sales when compared with changes in ingredient and supplies cost.



Another common way of visualizing the sensitivity of the model is by a tornado plot as shown below. The color shows the percent change in the parameter value and the length of the bars show the change in net value.



6.3. Stochastic Models

6.3.1. Definition

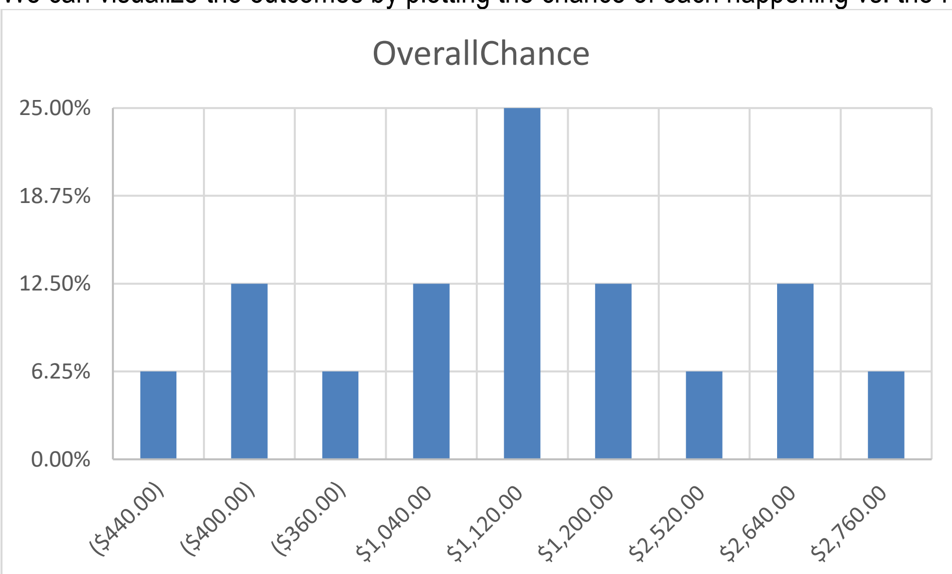
A stochastic model is one which uses random variables under different conditions as inputs and forecasts the probability of different outcomes. That is, the input variables/parameters are not given a fixed value but rather they vary randomly given a probability distribution. Therefore, stochastic models are inherently random as they have uncertain factors built directly into them. Stochastic models produce many different answers, estimators, and outcomes to observe the different effects on the solution, repeatedly, under different scenarios.

6.3.2. Stochastic Model Example

Returning to the example of the lemonade stand from the Deterministic Models section above, suppose that there are three possible temperatures for the summer: cool, average, and hot. There's a 25% chance we have a cool summer in which we sell only 100 cups of lemonade per week, a 50% chance of an average summer in which we sell 200 cups of lemonade per week, and a 25% chance of a hot summer in which we sell 300 cups of lemonade per week. Furthermore, there's a 25% chance we can buy lemons in bulk for $0.05 ingredient cost per cup, a 50% chance we can't but don't have other supply chain problems and stick with $0.10 ingredient cost per cup, and a 25% chance of supply chain problems resulting in $0.15 ingredient cost per cup. Suppose further that the outcome for our ability to buy in bulk is independent from the hotness or coolness of the summer – this means the NV our stand ends up with depends on two independent random events. Using this information, we can use the same NVF to determine the NVF of each possible outcome and the likelihood of that outcome happening:

Temperature	Chance	CupsPerWeek	Supply	Chance	IngredientCostPerCup	NV	OverallChance
Cool	25%	100	Bulk	25%	$0.05	-$360.00	6.25%
Cool	25%	100	Normal	50%	$0.10	-$400.00	12.50%
Cool	25%	100	Bad	25%	$0.15	-$440.00	6.25%
Avg	50%	200	Bulk	25%	$0.05	$1,200.00	12.50%
Avg	50%	200	Normal	50%	$0.10	$1,120.00	25.00%
Avg	50%	200	Bad	25%	$0.15	$1,040.00	12.50%
Hot	25%	300	Bulk	25%	$0.05	$2,760.00	6.25%
Hot	25%	300	Normal	50%	$0.10	$2,640.00	12.50%
Hot	25%	300	Bad	25%	$0.15	$2,520.00	6.25%

We can visualize the outcomes by plotting the chance of each happening vs. the NV of that outcome:



(note that the horizontal axis is not linear in this plot). This shows that the single most likely outcome is a revenue of $1120 which has a 25% chance of occurring. Further, there is a cluster of three "bad" outcomes around -$400, three "OK" outcomes around $1120, and three "good" outcomes around $2640 (corresponding to the chances of the cool, avg, and hot summer, since as-before the number of cups sold per week impacts the resulting NV much more strongly than the ingredient cost).

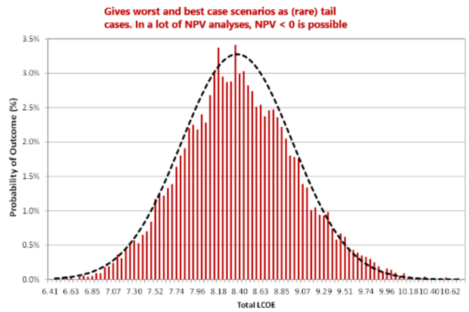
The **expected NV** results from multiplying the NV of each outcome by the chance of it happening and summing the results:

$$E[NV] = \sum_i NV_i \cdot \Pr(i) = -\$360 \times 6.25\% - \$400 \times 12.5\% + \cdots + \$2520 \times 6.25\%$$

$$= \$1120$$

(In this example, this is the same as the NV resulting from the average temperature and normal supply chain because of the symmetry of the outcomes and their likelihood about these averages). The expected value doesn't necessarily equal the value that we actually expect to occur. Even though in this example it *does* happen to correspond to the most likely outcome, there's still a 75% chance of some *other* outcome happening, so you should "expect" the NV *won't* be exactly $1120. Rather, the expected value is the average value that would occur if you ran this experiment many times and averaged the outcome (this suggests a way to determine expected values for complicated scenarios which we'll explore in the next section, Monte-Carlo simulations). Whether we can accept a project depends not just on its expected value, but whether we can accept the risk of dealing with all of the possible outcomes.

6.3.3. Monte Carlo Simulations



The calculation approach we used in the previous section works well enough if the probabilities of various events that impact NV are simple and independent. But what if our ability to get volume discounts depends not just on random chance but also on the amount we order? This inter-dependence may change the best strategy depending on independent random events (temperature and general availability of volume discounts at different order amounts), which now complicates the NV calculation in our stochastic model to the point that we could use a better approach.

Rather than calculating it, another way to determine the probability of various outcomes in a stochastic process is to run a simulation of the process many times and *see* how often each outcome happens. If you run enough experiments (e.g., 10000), then the fraction of times each outcome happens will limit to the probability of that outcome, so this can be a way to determine outcome probabilities and expected values for the process. This simulation approach is called a Monte Carlo simulation.

6.3.3.1. Flipping a Coin

Rather than complicated examples that actually require a Monte Carlo simulation, we'll first do the process for some simple examples where we can calculate the true probability of the outcome so we'll be able to see whether it's working.
Given an evenly-weighted coin, the chance of flipping a coin to its “heads” side is equal to the chance it flips to its “tails” side – 50/50. Why not test that? Trusting that Excel’s RAND() function is truly random, I’ve flipped a coin 20 times, with these results:

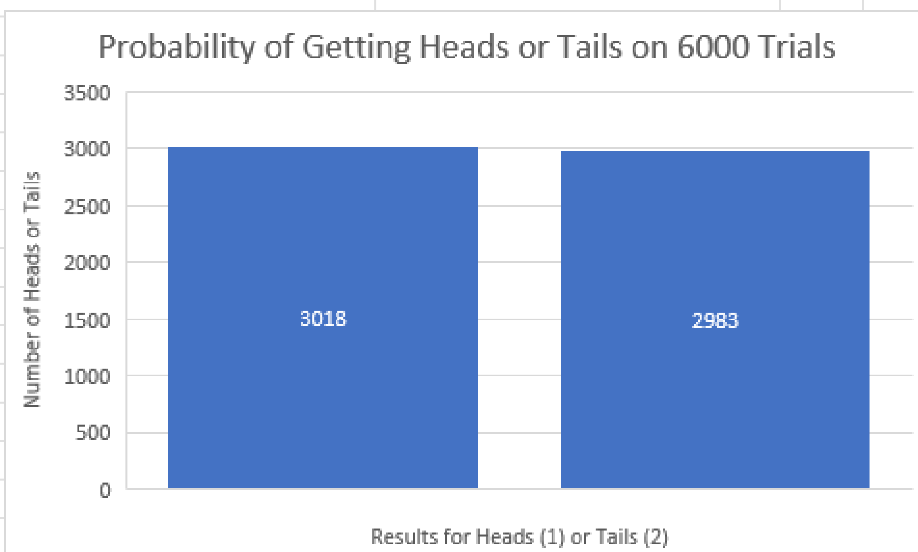
Trial	Result		Translated	
1	0	<-- =ROUND(RAND(),0)	Tails	<-- =IF(D14=1,"Heads","Tails")
2	0		Tails	
3	0		Tails	
4	0		Tails	
5	1		Heads	
6	1		Heads	
7	1		Heads	
8	1		Heads	
9	0		Tails	
10	1		Heads	
11	0		Tails	
12	1		Heads	
13	0		Tails	
14	1		Heads	
15	0		Tails	
16	0		Tails	
17	1		Heads	
18	0		Tails	
19	0		Tails	
20	0		Tails	

Distribution	
Number of Data Points	20
Number of Times Flipped Heads	8
Chance of Flipping Heads	40.0%

Behold! Undeniable proof that there is a 40% chance – not 50% – of flipping a coin resulting in “heads” side up. Oddly, repeating this experiment resulted in a 55% rate of “heads.” Repeating it again gave 65%. Why is the deviation from experiment to experiment so large? It has to do with the number of attempts – the coin isn’t flipped enough times per experiment to return useful data. Instead of 20 flips, let’s try 6,000. This is the resulting distribution:

Trial	Result	Translated	
1	0	Tails	<-- =IF(L14=1,"Heads","Tails")
2	0	Tails	
3	0	Tails	
4	0	Tails	
5	1	Heads	
6	0	Tails	
7	1	Heads	
8	1	Heads	
9	0	Tails	
10	1	Heads	
11	0	Tails	
12	1	Heads	
13	1	Heads	
14	1	Heads	
15	0	Tails	
16	1	Heads	
17	0	Tails	
18	0	Tails	
19	0	Tails	
20	0	Tails	

Distribution	
Number of Data Points	6000
Number of Times Flipped Heads	2983
Chance of Flipping Heads	49.7%



The distribution is much closer to 50% in this simulation. In fact, redoing the 6,000 flips experiment four more times yields the chance of heads as: 49.6%, 50.2%, 51.0%, 50.0%. It stands to reason that adding even *more* trials to each simulation would yield a simulated chance closer to the true probability of 50%.

Remember, a Monte Carlo method refers to random simulations used to determine values of parameters. By that definition, we've done a Monte Carlo simulation here: we had random simulations (coin flip results) used to determine the value of our one parameter (chance of flipping heads).

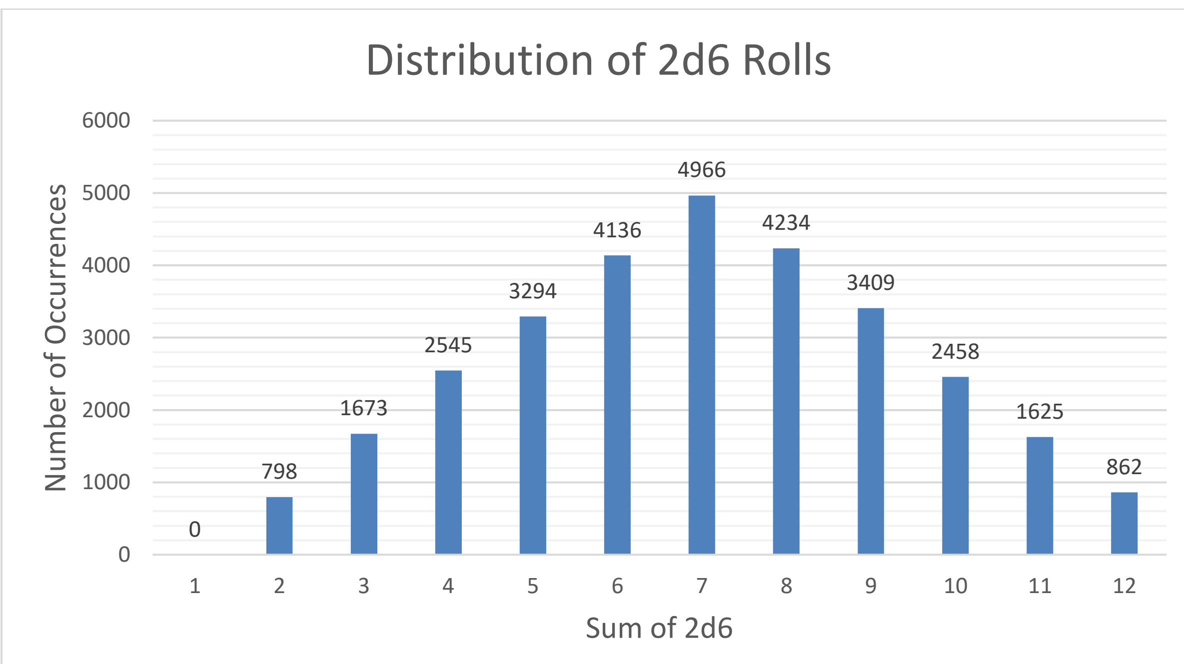
Monte Carlo has a typical use-case of determining model outputsfor which there are no analytical solutions, but we were able to use it here to test the probability of a coin flip landing heads.

6.3.3.2. Rolling Two Dice

In a slightly more complex version of a Monte Carlo simulation, let's approximate the chance two six-sided dice roll a combined value of 10, 11, or 12. While this is completely possible to calculate via combinatorics, a Monte Carlo simulation would work just as well. This slightly complicates things now that we need two parameters – the first die, and the second.

	First Die	Second Die	
Trial	x1	x2	Sum
1	2	5	7
2	4	3	7
3	6	2	8
4	2	3	5
5	4	5	9

Distribution		
Number of Data Points	30000	
Rolls of a Combined Value 10-12	4885	<-- =COUNTIF(F:F,">=10")
Chance of Rolling 10-12	16.28%	<-- =I7/I6



6.3.3.2.1. Review of Confidence Intervals

The confidence interval for an estimated parameter is a quantitative measurement of its accuracy. A confidence interval defines a range within which the "true" value of the parameter lies.

Review from statistics class:
μ = population mean, or "true mean"
$\bar{x}$ = sample mean
k = critical value, or the "z-score" value
σ = standard deviation
n = number of observations in sample
α = significance level, or 1-(the desired confidence level)

The equation for calculating the confidence interval is below:

$$\mu_{1-\alpha} = (\bar{x} + k_{\alpha/2}\frac{\sigma}{\sqrt{n}}; \bar{x} + k_{1-\alpha/2}\frac{\sigma}{\sqrt{n}})$$

If you have all the inputs for this formula, it will return a "lower" and a "higher" value: the true mean will have a (confidence level) probability of being between these two values.
For example, you test 25 rebars to see if they meet the minimum strength requirements. The sample mean of these 25 rebars is 300 MPa. The rebar supplier quotes a 25 MPa standard deviation. With a 95% confidence level, what is the true mean of the rebar population?

$$\mu_{1-\alpha} = (\bar{x} + k_{\alpha/2}\frac{\sigma}{\sqrt{n}}; \bar{x} + k_{1-\alpha/2}\frac{\sigma}{\sqrt{n}})$$
$$\mu_{95\%} = \left(300\ \mathrm{MPa} + k_{0.025}\frac{25\ \mathrm{MPa}}{\sqrt{25}}; 300\ \mathrm{MPa} + k_{0.975}\frac{25\ \mathrm{MPa}}{\sqrt{25}}\right)$$
$$\mu_{95\%} = (300\ \mathrm{MPa} - 1.96 * 5\ \mathrm{MPa}; 300\ \mathrm{MPa} + 1.96 * 5\ \mathrm{MPa})$$
$$\mu_{95\%} = (290.2\ \mathrm{MPa}; 309.8\ \mathrm{MPa})$$

There is a 95% probability that the true mean is somewhere between 290.2 MPa and 309.8 MPa.

We can also calculate this with Excel. The product $k_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ is also called the **confidence radius** or **margin of error**, and we can calculate it using the formula **confidence.norm(alpha, sigma, n)**
For example,

Confidence level	95%	
alpha	5%	
standard_dev	25	
sample size	25	
mean	300	
Confidence Radius (Margin of Error)	9.7998	<-- =CONFIDENCE.NORM(B2, B3, B4)
Low Value	290.2002	<-- =B5 - B6
High Value	309.7998	<-- =B5 + B6

6.3.3.2.2. Rolling Two Dice, Confidence Interval

Here's the Excel method for calculating the confidence interval for the 30,000 2d6 rolls shown above:

Confidence level	95%	
alpha	5%	<-- =1-M40
standard_dev	2.416	<-- =STDEV.S($F$6:$F$30006)
sample size	30000	
mean	7.011	<-- =AVERAGE($F$6:$F$30006)
Confidence Interval	0.0273	<-- =CONFIDENCE.NORM(M41,M42,M43)

This means that based on this Monte Carlo simulation we are 95% sure that the true mean of rolling 2d6 is 7.011 ± 0.0273, or (6.9837, 7.0383)..
Because the two dice rolls are independent events each with an expected value of 3.5, the true expected value of the sum is 7. This was indeed within the range we specified for our Monte Carlo simulation's estimation of the mean (as you'd expect 19 times out of 20 when working with a 95% confidence level).

What about the confidence interval for the chance of rolling 10-12 with 2d6? Using a data table to simulate 30 trials of the 30,000 dice rolls, we can generate a confidence interval through Excel as well:

Data Table	
Trial	Chance of Rolling 10-12
Base	16.72%
1	16.50%
2	16.67%
3	16.72%
4	16.92%
5	16.97%
6	16.88%
7	16.52%
8	16.71%
9	16.67%
10	16.59%
11	16.56%
12	16.65%
13	16.70%
14	16.76%
15	16.76%
16	16.96%
17	16.62%
18	16.25%
19	17.13%
20	16.53%
21	16.59%
22	16.57%
23	16.97%
24	16.62%
25	16.82%
26	16.60%
27	16.53%
28	16.55%
29	16.54%
30	16.67%

Confidence level	95%	
alpha	5%	<-- =1-Y38
standard_dev	0.00182	<-- =STDEV.S(V41:V70)
sample size	30	<-- =COUNT(V41:V70)
mean	0.1668	<-- =AVERAGE(V41:V70)
Confidence Interval	0.000651	<-- =CONFIDENCE.NORM(Y39,Y40,Y41)

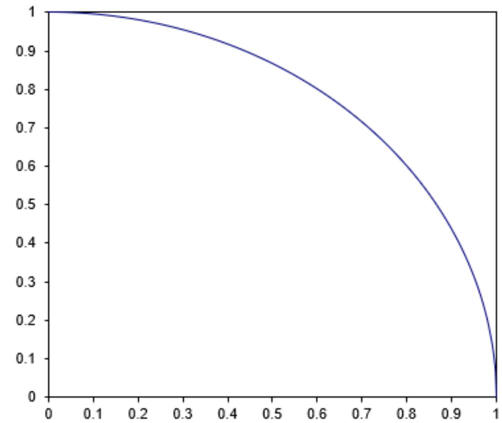
A little confusing: the mean of this data table (0.1671) is the average of 30 iterations of the "how often a 2d6 roll summed to 10-12, simulated 30,000 times," experiment.

From this, we've calculated that there is a 95% probability the true probability of [rolling 2d6 and the result being 10-12] as 0.1668 ± 0.000651, or (0.166149, 0.167451).

The theoretical answer to this example is 16.667% - which is inside our confidence interval! This works! (Or at least, we didn't hit the 1/20 chance our simulation would have a confidence interval for the mean outside the true mean)

6.3.3.3. Using Monte Carlo Simulation to Calculate Pi

If the area of the unit circle is π, it follows that the area of one quarter of the unit circle is $\pi/4$. If we plot the curve of a quarter-unit circle inside a unit square – of area 1 – it should look like this:



If we were to plot a random point with x-value between 0 and 1, and a y-value between 0 and 1 (in other words, within the unit square), the chance of the point ***also*** being inside the quarter-unit circle is:

$$\frac{Area\ Quarter\ Circle}{Area\ Unit\ Square} = \frac{\pi/4}{1} = \pi/4$$

If you plot *many* random points inside the unit square, should therefore approximately be equal to $\pi/4$, and so multiplying this approximate value by 4 should give you an approximate value of π.

Iteration	Random X	Random Y	Inside Unit Circle?
1	0.234717943	0.431540191	TRUE
2	0.791687041	0.609845122	TRUE
3	0.518362657	0.79680997	TRUE
4	0.627755539	0.303957941	TRUE
5	0.464243367	0.706159894	TRUE
6	0.812937618	0.92061859	FALSE
7	0.406585304	0.886413792	TRUE
8	0.701687605	0.126584372	TRUE
9	0.223821405	0.063891356	TRUE
10	0.408692904	0.856905529	TRUE
11	0.010608921	0.959851006	TRUE
12	0.798018698	0.744488639	FALSE
13	0.222729019	0.936242631	TRUE
14	0.791073082	0.666258176	FALSE
15	0.870713399	0.221578871	TRUE
16	0.927568278	0.257138825	TRUE
17	0.16373457	0.197024427	TRUE
18	0.475767288	0.266443378	TRUE
19	0.316216398	0.667812694	TRUE
20	0.611655458	0.90088867	FALSE

Calculation of π		
Number of Data Points	20	<-- =COUNT(B:B)
Number of Data Points Inside Curve	16	<-- =COUNTIF(D:D,TRUE)
Pi?	3.200	<-- =H6/H5 * 4

With 20 datapoints, our estimation of π is 3.200; not particularly accurate. So, how about with 60 datapoints?

Iteration	Random X	Random Y	Inside Unit Circle?
1	0.369987824	0.270985223	TRUE
2	0.781443584	0.422638753	TRUE
3	0.633738004	0.789260599	FALSE
4	0.264538659	0.746651258	TRUE
5	0.719804557	0.150428502	TRUE

Calculation of π		
Number of Data Points	60	<-- =COUNT(B:B)
Number of Data Points Inside Curve	50	<-- =COUNTIF(D:D,TRUE)
Pi?	3.333	<-- =H6/H5 * 4

With 60 datapoints, our estimation is even farther away than before! With the huge level of variability in our two parameters (x and y), this is possible in a stochastic model. How many iterations do we need to run through to find a ***consistently*** accurate estimation? While one of the 20 or 60 datapoint models may calculate 3.1415… if the model is re-run enough times, it (as shown) does not calculate that every time.

Behold! 35,000 iterations.

Iteration	Random X	Random Y	Inside Unit Circle?
1	0.4271779	0.995217772	FALSE
2	0.133015941	0.788383998	TRUE
3	0.296077202	0.469228586	TRUE

Calculation of π		
Number of Data Points	35000	<-- =COUNT(B:B)
Number of Data Points Inside Curve	27499	<-- =COUNTIF(D:D,TRUE)
Pi?	3.143	<-- =H6/H5 * 4

That's a bit better. Exponentially more iterations than this would reasonably lead to a more accurate calculation to double-digit decimal point accuracy.

There's one more thing to think about, though: how can I be certain my model didn't just fluke an accurate answer? For this, we can check a confidence interval.

Data Table	
Base	3.140
1	3.152343
2	3.137943
3	3.135429
4	3.152457
5	3.150629
6	3.161714
7	3.136229
8	3.132
9	3.147771
10	3.149829
11	3.136914
12	3.150286
13	3.132686
14	3.141829
15	3.1368
16	3.134286
17	3.1536
18	3.148
19	3.133257
20	3.1432
21	3.126057
22	3.122171
23	3.145714
24	3.129714
25	3.131314
26	3.14
27	3.150629
28	3.124571
29	3.148343
30	3.139886

Confidence level	95%	
alpha	5%	<-- =1-Q4
standard_dev	0.00982	<-- =STDEV.S(N6:N35)
sample size	30	<-- =COUNT(N6:N35)
mean	3.1409	<-- =AVERAGE(N6:N35)
Confidence Interval	0.003514	<-- =CONFIDENCE.NORM(Q5,Q6,Q7)

With this, we can predict with 95% certainty that the true value of π is between (3.137386, 3.144414).

6.3.3.4. Using Monte Carlo Simulation to Plan Your Retirement

Picture being 65 years old, ready to retire. You have amassed to this point in your life a wealth sum of $1,000,000. Ideally, you would like to live the next ten years with the following holding true: that you can drawdown $150,000 from your retirement savings each year, and that when you turn 75 there is still *some* money leftover.

You decide to split your money between risky stock investments and safe, risk-free bond investments. For the stock investments, assume the expected annual return (μ) is 8%, and the standard deviation (σ) of returns is 20%. For the bonds, assume the risk-free rate is 4%. These interests are continuously compounded.

(Meaning: $100 invested in the “risk-free” grows to $FV = PV * e^{rt} = \$100 * e^{4\%}$ at the end of the year, and the “risky” grows by a factor of $e^{\mu+\sigma * Z} = e^{8\%+(20\% * Z)}$ (where Z is a random number normally distributed with mean 0 and standard deviation 1.))

You would like to know if this is a feasible plan considering your wealth and the expected returns on investments. As a starting point, you consider investing 70% of your wealth in stocks, and 30% in bonds. For simplicity’s sake, assume you take all your money out of investments at the end of each year, drawdown the $150,000, then reinvest the remaining money with the same 70/30 proportion split.

To explore this situation, we can use a Monte Carlo simulation. We will be simulating the return on the stock investments, as everything else can be calculated from the given information.

Starting wealth	1,000,000
Risk-free rate	4%
Parameters of risky investment	
Expected annual return	8%
Standard deviation of return	20%
Proportion invested in "risky investments"	70%
Annual drawdown	150,000

Year	Wealth at beginning of year	Invested in risky	Invested in bonds	Random number, normally distributed (Z-score)	1+return on risky investment	Wealth at end of year	Drawdown	Left at end of 10 years
1	1,000,000	700,000	300,000	-1.1697	0.8573	912,366	150,000	
2	762,366	533,656	228,710	0.4097	1.1758	865,506	150,000	
3	715,506	500,854	214,652	0.1894	1.1251	786,933	150,000	
4	636,933	445,853	191,080	1.6421	1.5044	869,638	150,000	
5	719,638	503,747	215,891	0.4386	1.1826	820,441	150,000	
6	670,441	469,309	201,132	-0.2422	1.0321	693,693	150,000	
7	543,693	380,585	163,108	-0.4862	0.9829	543,844	150,000	
8	393,844	275,691	118,153	-0.3132	1.0175	403,494	150,000	
9	253,494	177,446	76,048	1.2333	1.3863	325,150	150,000	
10	175,150	122,605	52,545	-1.0346	0.8808	162,679	150,000	$ 12,679.42

Normally distributed random numbers generated by <-- =NORM.S.INV(RAND())

<-- =D21*G21+E21*EXP($C$4)

Wealth at beginning of year <-- =H20-I20

1+return on risky investment <-- =EXP($C$6+$C$7*F21)

Investment in risky asset <-- =C21*$C$8

In this instance, we reach the end of 10 years with $12,679.42 leftover after drawing-down $150,000 per year and investing in stocks with uncertain returns – it worked! Using a data table, let’s repeat the simulation 30 times to see if we consistently have a positive outcome.

Simulation	$ 12,679.42		How Well Did We Do?		
1	$ 704,988.89		Positive Outcome Percentage		
2	$ 80,605.57				
3	$ 893,502.52	<-- {=TABLE(,J6)}	Runs	30	
4	$ (183,554.23)		Successful Runs	53.3%	<-- =(COUNTIF(INDIRECT("O14:O"&ROWS(O:O)),">0")/S15)
5	$ (389,307.42)				
6	$ 118,647.42				
7	$ 230,747.69				
8	$ 92,567.37				
9	$ 188,459.27				
10	$ 133,948.74				
11	$ (178,782.70)				
12	$ (534,501.68)				
13	$ 482,739.28				
14	$ 896,276.55				
15	$ 481,153.33				
16	$ 347,796.03				
17	$ (243,286.36)				
18	$ (564,706.63)				
19	$ (51,282.60)				
20	$ (95,191.54)				
21	$ (33,435.52)				
22	$ (640,632.95)				
23	$ (139,489.94)				
24	$ 5,229.97				
25	$ (1,120,061.01)				
26	$ 936,000.52				
27	$ 190,564.42				
28	$ 268,270.48				
29	$ (437,089.20)				
30	$ (520,681.58)				

Based on 30 simulated runs, 53.3% of the time we can meet the initial requirements set out at the beginning of the question. However, we have a huge amount of variance – notice in simulation #25 that we end the 10-year period in debt by $1.1 million. There are many factors to modify in an attempt to raise the successful run percentage: living in retirement on a smaller budget than $150,000 or making any number of potential changes to the investment strategy.

6.3.3.5. Net Value Example, Monte Carlo

Remember your lemonade stand? Well, it was a great success and after a few years, you have made enough money to consider renting a store front in the neighbourhood to sell lemonade and other delicious drinks and treats. Below are the details of your potential lemonade business (empire):

- The lease you are considering is a 5-year lease, $3,000/month + 3% increase per year
- Expected monthly revenue from the business is $12,000/month, with a standard deviation of $1,000/month normally distributed.
- The revenue is expected to grow yearly (i.e. prices of food/drink will change yearly) with inflation which is estimated to be 3%/year, with a standard deviation of 0.5% normally distributed.
- Labour cost is $4,000/month and will increase $200/month each year.
- Utilities and maintenance cost $500/month, with a standard deviation of $50 normally distributed.
- Material costs are $1500/month with a standard deviation of $100 normally distributed and is expected to increase with inflation each year.
- Being a wise and risk-averse entrepreneur, you save all the profit you make from the business in a savings account that accrues 5% yearly interest compounded monthly.

- For the next phase of your lemonade empire, you need $200,000 to start multiple stores in the city. What is the probability that you will have $200,000 in your bank account after 5 years?
- Perform a Monte Carlo simulation of 300 runs.

We can do this on Excel. The Excel file is provided on Avenue. First, we set up the values for the lease, inflation, revenue, labour cost, and material cost as shown below. The lease is increasing by a fixed about (3%) each year. Inflation is randomized using the formula “=NORM.INV(RAND(),3,0.5)”. This randomizes the inflation rate using a normal distribution with an average of 3% and a standard deviation of 0.5%. For the first year, inflation is not applied, therefore the inflation rate is set to 0 for year 1. The expected monthly revenue for year 1 is $12,000. The expected revenue in subsequent year is increased by the inflation rate calculated in the row above. Similarly, the expected material cost is increased by the inflation each year. Lastly, the labour cost increases by $200 each year.

L	M	N	O	P	Q
Year	1	2	3	4	5
Lease	$ 3,000.00	$ 3,090.00	$ 3,182.70	$ 3,278.18	$ 3,376.53
Inflation	0	2.93925804	3.477367928	3.277088715	2.391675594
Expected Monthly Revenue	$ 12,000.00	$ 12,352.71	$ 12,782.26	$ 13,201.15	$ 13,516.87
Labour Cost	$ 4,000.00	$ 4,200.00	$ 4,400.00	$ 4,600.00	$ 4,800.00
Expected Material Cost	$ 1,500.00	$ 1,544.09	$ 1,597.78	$ 1,650.14	$ 1,689.61

We can then calculate the monthly revenue, lease, labour cost, utilities cost, materials cost, monthly net value, and bank account balance as shown below. For revenue, utilities, and materials we use the NORM.INV function to randomize the values. The monthly net value is calculated as:

$$NV = Revenue - Lease - Labour - Utilities - Materials$$

Lastly, the monthly bank account balance is calculated by multiplying the previous month’s account balance by the monthly interest rate plus the month’s NV.

	A	B	C	D	E	F	G	H	I	J
1	**Year**	**Month**	**Revenue**	**Lease**	**Labour Cost**	**Utilities**	**Materials**	**Monthly NV**	**Bank Account Balance**	
2	1	1	$11,816.44	$3,000.00	$ 4,000.00	$ 480.82	$1,474.90	$ 2,860.72	$ 2,860.72	
3	1	2	$13,135.77	$3,000.00	$ 4,000.00	$ 552.58	$1,542.63	$ 4,040.56	$ 6,913.20	
4	1	3	$12,779.32	$3,000.00	$ 4,000.00	$ 407.50	$1,534.73	$ 3,837.09	$ 10,779.10	
5	1	4	$10,605.56	$3,000.00	$ 4,000.00	$ 461.92	$1,389.01	$ 1,754.63	$ 12,578.65	
6	1	5	$11,514.61	$3,000.00	$ 4,000.00	$ 498.62	$1,577.46	$ 2,438.53	$ 15,069.58	
7	1	6	$13,082.83	$3,000.00	$ 4,000.00	$ 546.68	$1,328.28	$ 4,207.87	$ 19,340.24	
8	1	7	$10,011.08	$3,000.00	$ 4,000.00	$ 453.14	$1,355.36	$ 1,202.58	$ 20,623.41	
9	1	8	$10,482.42	$3,000.00	$ 4,000.00	$ 499.76	$1,467.29	$ 1,515.37	$ 22,224.70	
10	1	9	$12,615.82	$3,000.00	$ 4,000.00	$ 430.09	$1,496.44	$ 3,689.29	$ 26,006.60	
11	1	10	$12,087.79	$3,000.00	$ 4,000.00	$ 520.24	$1,463.05	$ 3,104.50	$ 29,219.45	
12	1	11	$13,327.91	$3,000.00	$ 4,000.00	$ 541.51	$1,484.61	$ 4,301.79	$ 33,642.99	
13	1	12	$12,279.85	$3,000.00	$ 4,000.00	$ 531.39	$1,451.38	$ 3,297.08	$ 37,080.26	
14	2	13	$10,789.09	$3,090.00	$ 4,200.00	$ 499.47	$1,728.06	$ 1,271.56	$ 38,506.32	
15	2	14	$12,479.93	$3,090.00	$ 4,200.00	$ 447.24	$1,315.71	$ 3,426.98	$ 42,093.74	
16	2	15	$13,033.87	$3,090.00	$ 4,200.00	$ 501.95	$1,503.14	$ 3,738.78	$ 46,007.91	

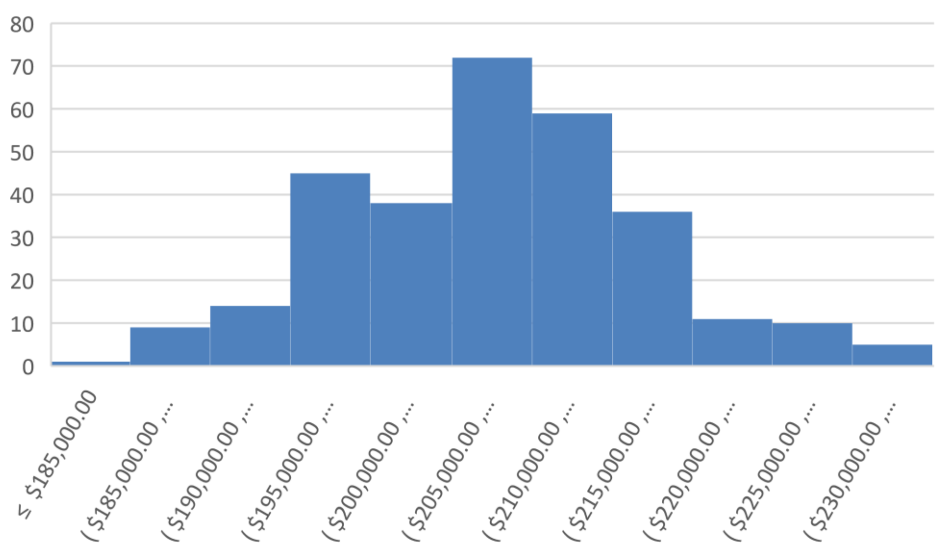
We then perform 300 simulations, taking the final account balance at the end of the 5 years.

	A	B
62		
63	Simulations	
64	1	$ 208,650.44
65	2	$ 192,742.06
66	3	$ 214,603.84
67	4	$ 210,732.81
68	5	$ 193,891.32
69	6	$ 197,223.00
70	7	$ 198,361.88
71	8	$ 227,507.27
72	9	$ 221,203.89
73	10	$ 205,084.32
74	11	$ 207,365.02
75	12	$ 206,493.06
76	13	$ 215,265.20
77	14	$ 211,129.06

To calculate the chance of having more than $200,000 in the bank account at the end of the 5 years, we can count the number of simulations that have a final bank account balance of greater than $200,000 and divide it by the total number of simulation runs. For this Monte Carlo analysis, it is estimated that you will have an 80% chance of having more than $200,000 in your bank account at the end of 5 years.

	J	K	L	M	N	O
10						
11						
12			**Monte Carlo Analysis**			
13			Average Account Balance	$ 207,843.40		
14			Std Dev	$ 9,475.00		
15			# of Runs	300		
16			Probability of > 200000	80%		
17						

Lastly, we can plot a histogram of our simulation outcomes to visualize the probability of different outcomes.



.

6.3.3.6. Monte Carlo, Midterm-Style Example

With the large amount of random number generation and analysis needed in a Monte Carlo simulation (that provides useful data), it is not often done by hand. However, you *can* perform a small-scale Monte Carlo simulation given a small handful of "random numbers" to use.

A distribution of wait times at Service Ontario is seen in the table below.

Wait Time (min)	Probability of Wait Time
30	5%
45	15%
60	20%
75	30%
90	10%
105	10%
120	10%

What is the average wait time?

Mathematically, we would multiply the wait times by their probability "weights" and compute the mean. This would look something like:

$\mu = \sum x * P(x)$

$$\mu = (30\text{ minutes} * 5\%) + (45\text{ minutes} * 15\%) + (60\text{ minutes} * 20\%) + (75\text{ minutes} * 30\%) + (90\text{ minutes} * 10\%) + (105\text{ minutes} * 10\%) + (120\text{ minutes} * 10\%)$$

$\mu = 74.25$ minutes, or $1\text{h}14\text{m}15\text{s}$.

What we have here is the mean (or expected value) of a probability distribution. But this isn't exactly a simulation; there is no information here about the range of the probability distribution that a Monte Carlo method can easily provide. So, let's adjust the question so that a Monte Carlo method can be emulated by hand. We know that the probabilities in the distribution must sum to 1, so let's assign ranges of random numbers that correspond to the probabilities. Assigning a random number interval can be a good way of handling obtuse outcomes.

The italicized red text denotes changes from the previous iteration of the example.

A distribution of wait times at Service Ontario is seen in the table below.

Wait Time (min)	Probability of Wait Time	*Random Number Range*
30	5%	*01-05 (or 5% of the range; equal to the probability)*
45	15%	*06-20 (15% of the range)*
60	20%	*21-40 (20%, and so on)*
75	30%	*41-70*
90	10%	*71-80*

105	10%	*81-90*
120	10%	*91-100*

In a Monte Carlo simulation, the following sample random numbers were drawn: 05, 60, 75, 72, 29, 30, 02, 88. What is the average wait time?

By having the random numbers given, now we can complete a rough Monte Carlo simulation by hand.

These random numbers correspond to the following wait times, respectively: 30 min, 75 min, 90 min, 90 min, 60 min, 60 min, 30 min, 105 min. The mean of these times is:

$\bar{x} = \frac{\sum x}{n} = \frac{(30+75+90+90+60+60+30+105)}{8} = 67.5$ minutes.

Just as in the previous examples, having a larger random number sample would likely result in a higher degree of accuracy for the mean of the probability distribution (eg., closer to 74.5 minutes).

6.4. Conclusions

We now know the concept of sensitivity analysis, as well as its applications and some examples of the kinds of sensitivity analyses which can be conducted. We understand how inputs can influence the outputs of a model and can describe this qualitatively or quantitatively. In addition, we understand the difference between deterministic and stochastic models – certainty and a lack of randomness in deterministic models, while stochastic models consider input data to follow one of multiple known distributions. Moreover, we have seen both ceteris paribus analyses (where we only note a change in one input at a time), as well as Monte-Carlo simulations, in which multiple or all inputs can change at a time.