ECON 180

Introduction to Principles of Microeconomics and Financial Project Evaluation

Lecture 27: Probabilistic Sensitivity Analysis

November 9, 2021

Version 1

Required Reading

- Marchenko, R.S. & Cherepovitsyn, A.E. (2017). Improvement of the quality of calculations using the Monte Carlo simulation method in the evaluation of mining investment projects. *IT&QM&IS 2017*, pp. 247 – 251. Retrieved from https://ieeexplore-ieee-org.ezproxy.library.uvic.ca/document/8085805
  - **Short discussion of Monte Carlo methods.**

Recommended Reading

- Shafrin, J. (2007). Bootstrapping [Web Page]. Retrieved from http://healthcare-economist.com/2007/05/24/bootstrapping/
  - A **brief summary of bootstrapping**.
- Wright, D. (n.d.). “Easy” Excel Inverse Triangular Distribution for Monte Carlo Simulations Redux [Web Page]. Retrieved from https://www.drdawnwright.com/easy-excel-inverse-triangular-distribution-for-monte-carlo-simulations/
  - **Includes a tool for allowing Excel to sample values from a triangle distribution.**
- Yang, Z.R., Zwolinkski, M. & Chalk, C.C. (1988). Bootstrap, an alternative to Monte Carlo simulation. *Electronics Letters*, 34(12), pp. 1174 – 1175. https://eprints.soton.ac.uk/251238/
  - **A short but important summary of bootstrapping.**
- European Union. (2014). *Guide to Cost-Benefit Analysis for Investment Projects*. Belgium: European Commission. Retrieved from https://ec.europa.eu/regional_policy/en/information/publications/guides/2014/guide-to-cost-benefit-analysis-of-investment-projects-for-cohesion-policy-2014-2020
  - **An excellent free textbook.**

Case Studies

- de Rooij, M. et al. (2014). Cost-effectiveness of Magnetic Resonance (MRI) Imaging and MR-Guided Targeted Biopsy Versus Systematic Transrectal Ultrasound-Guided Biopsy in Diagnosing Prostate Cancer: A Modelling Study from a Health Care Perspective. *European Urology*, 66(3), 430-436. https://doi-org.ezproxy.library.uvic.ca/10.1016/j.eururo.2013.12.012
  - **Source of a Monte Carlo scatter plot**.
- Dheskali, E., Koutinas, A. A. & Kookos, I. K. (2020). Risk assessment modeling of bio-based chemicals economics based on Monte-Carlo simulations. *Chemical Engineering Research and Design*, *163*, 273-280. https://doi-org.ezproxy.library.uvic.ca/10.1016/j.cherd.2020.09.011
  - **Interesting use of cost-acceptability curves.**
- Jinbo, S. (2009). A Decision-Making Model of Concession Period for **Municipal Waste Incineration** Build-Operate-Transfer Project. MASS '09. Retrieved from https://doi-org.ezproxy.library.uvic.ca/10.1109/ICMSS.2009.5301118
  - **Very clear Monte Carlo example, with tables and diagrams.**
- Li, C. & Sun, A. (2008). The Research of Economic Evaluation Project Risk Based on Monte Carlo Simulation. WiCOM '08. Retrieved from https://doi-org.ezproxy.library.uvic.ca/10.1109/WiCom.2008.2442
  - **Chinese construction. Good demonstration of the use of probability distributions.**

Learning Objectives

- Understand Monte Carlo simulation
- Understand Bootstrapping
- Be able to read and interpret a scatter plot of Bootstrapping or Monte Carlo cost-effectiveness simulations.
- Be able to interpret CEAC.
- Be able to calculate 95% confidence intervals for ICER from Bootstrapping or Monte Carlo simulations.

ESSENTIALS (20 slides)

Commonly…

- When working on a project, you'll have a good idea of the possible range of values for given variables, and their distribution.
- The probability distribution of the COMBINATION of these variables into NPV, IRR, BCR and so on may not be easy to determine.
- The Monte Carlo method of dealing with uncertainty (named after the casino) is to run the calculations many different times, with random variables fitting the desired distributions.
- A probability distribution for the complex value can then (hopefully) be inferred from the results.
- We'll also look at *bootstrapping*, a similar technique used when you *don't* have a distribution, but *do* have a bunch of observations of your parameters.

What do we know?

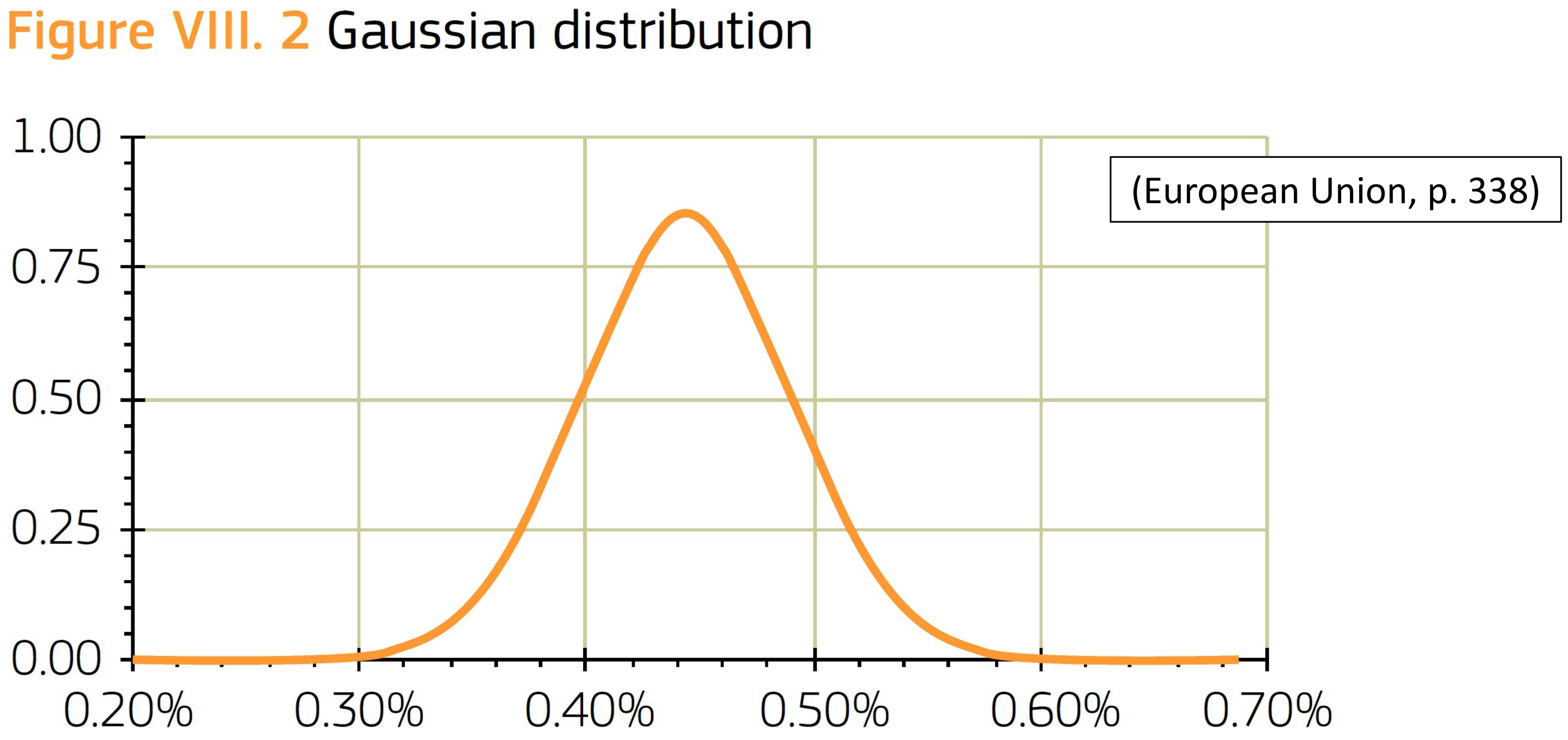
- When values are probabilistic, we often know one of two things:
- The distribution (and its mean, limits and standard deviation, if applicable):
  - Uniform distribution
  - Normal (Bell Curve) distribution
  - Etc.
  - (Use Monte Carlo for this)
- We don’t know the distribution, but we have values sampled from it:
  - History
  - Results of random trials
  - Etc.
  - Example: think of marks in a class…
  - (Use Bootstrapping for this)
- Let’s start with the first case.

Published example (Jinbo 2009)

TABLE I. STATISTICAL DISTRIBUTION OF OPERATION COST

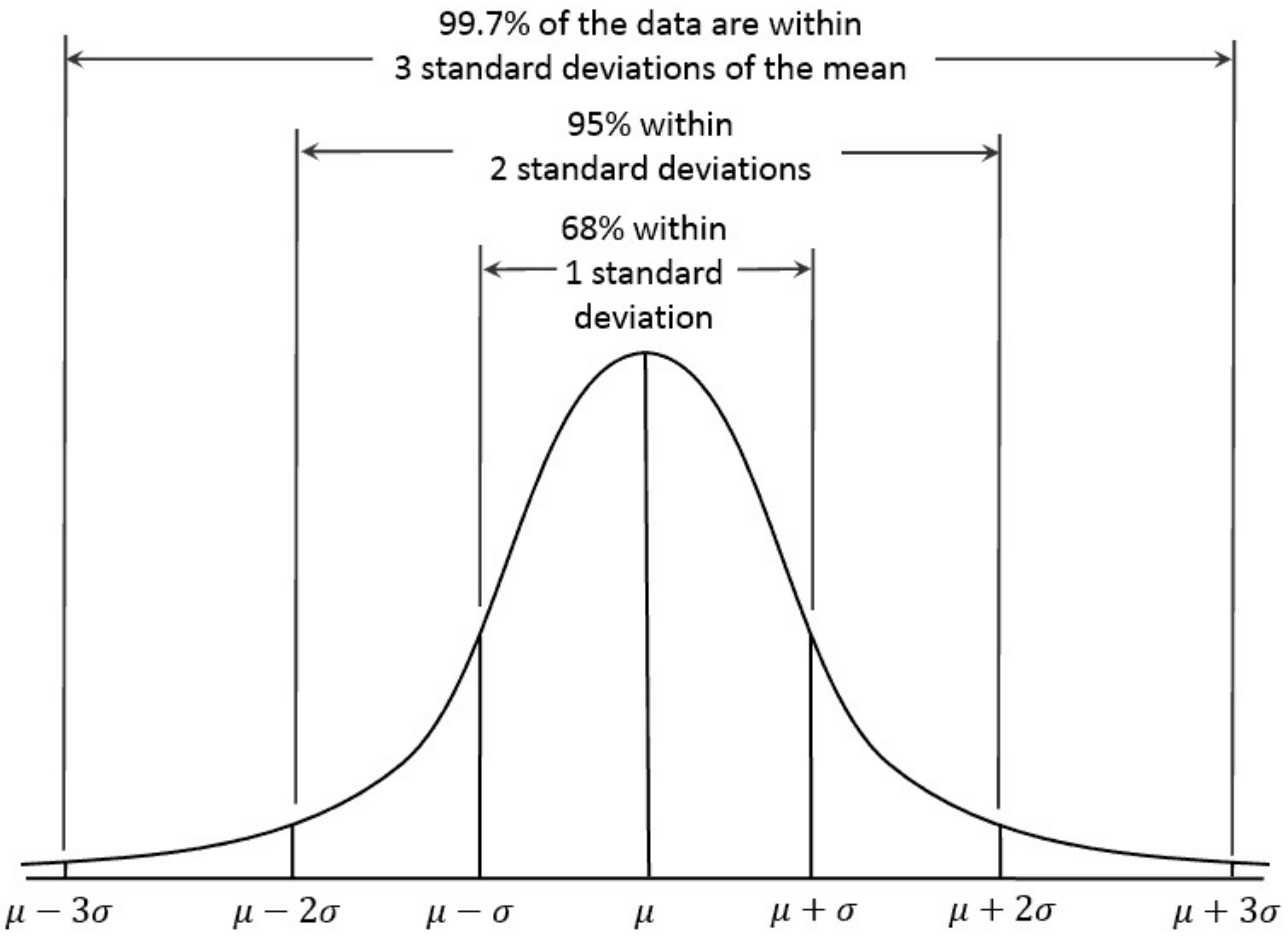
Random Variable	Name	Distribution
C_1	Raw materials	Normal distribution, with the mean of 20 Yuan/ton and the standard deviation of 2 Yuan/ton.
C_2	Fuel and power	Normal distribution, with the mean of 39.5 Yuan/ton and the standard deviation of 4.0 Yuan/ton.
C_3	Ash transportation	Normal distribution, with the mean of 3 Yuan/ton and the standard deviation of 0.3 Yuan/ton
C_4	Management fees and wages	Normal distribution, with the mean of 2800000 Yuan/year and the standard deviation of 280000 Yuan/year

Normal, ‘Bell Curve’, or ‘Gaussian’ distribution



- Student grades tend to fall into this distribution.
- To describe it, we need at least a MEAN and a STANDARD DEVIATION.
- Excel has built-in functions to draw numbers from a normal distribution.
- Specifically, NORM.INV(RAND(),MEAN,STD)

From (MIN,BASELINE,MAX) to Bell Curve



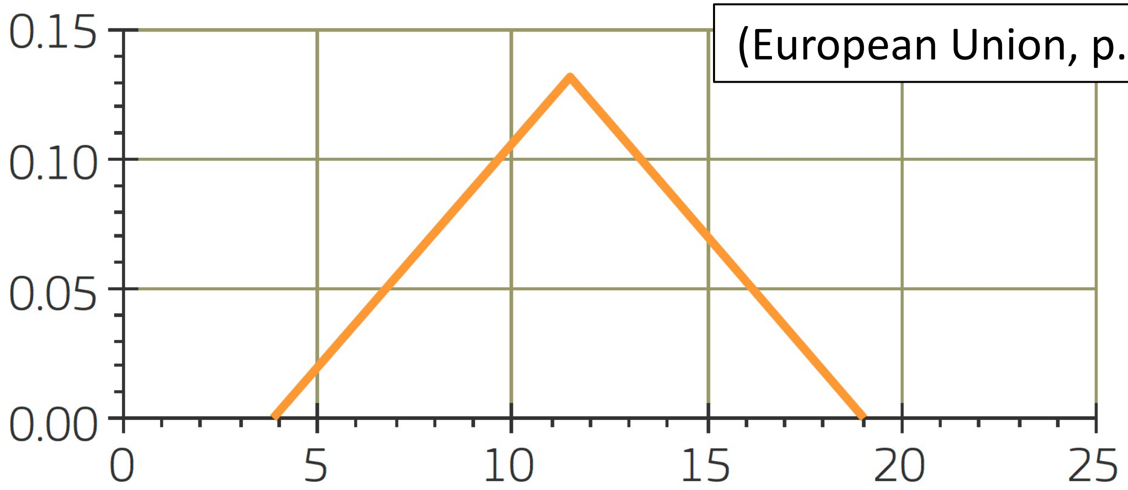
- μ = MEAN (Average)
- σ = Standard Deviation
- Since for a Normal Distribution, 99.7% of values are within 3 σ of the mean…
- (MIN – MAX) should cover approximately 6 standard deviations
- →A common shortcut is σ = (MAX – MIN)/6

Image Source: [Wikipedia]

A slightly richer variety (Li & Sun, 2008)

The first year		
Risk factor	*Distribution*	*parameter*
Residential sales	*Triangle*	*None*
Commercial housing sales	*Triangle*	*None*
Land costs	*Uniform*	*a=11182,b=12105*
Early costs	*Normal*	$\varepsilon = 911, \sigma = 50$
Development and construction costs	*Triangle*	*A=3112,m=3276,b=3374*
Sales costs	*Triangle*	*a=235,m=313,b=329*
Other costs	*Normal*	$\varepsilon = 249, \sigma = 15$

The Triangle Distribution

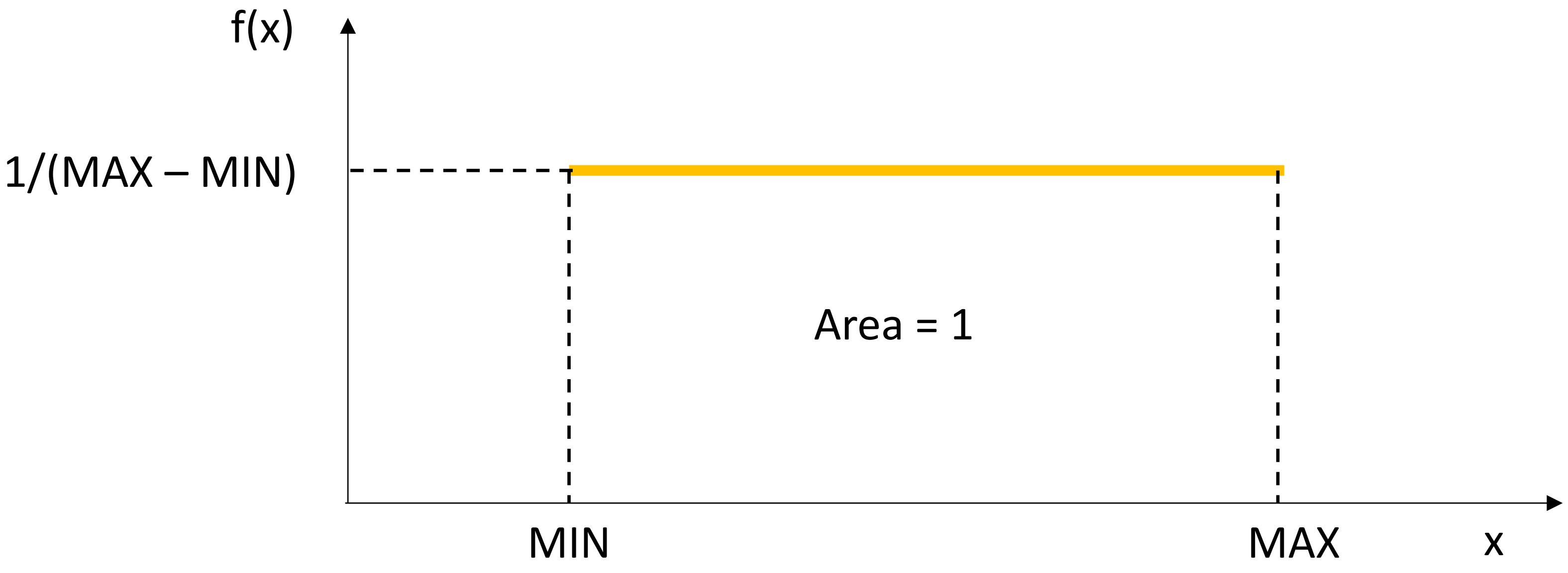


(European Union, p. 339)

- Example on the left:
  - MIN = 4, BASE = 12, MAX = 19
- Height at 4 = 0
- Height at 19 = 0
- Height at 12 : 2/(19 – 4) = 0.1333…

- If you don’t think a normal distribution is right for your variable, but don’t have much more than (MIN,BASELINE,MAX), you can use a triangle distribution (‘three point estimation’).
- A triangle distribution has a height of 0 at MIN and MAX, and a height of 2/(MAX – MIN) at the baseline.
- (Need help? See this site for more details, including formulas and an Excel implementation: https://www.drdawnwright.com/easy-excel-inverse-triangular-distribution-for-monte-carlo-simulations/)

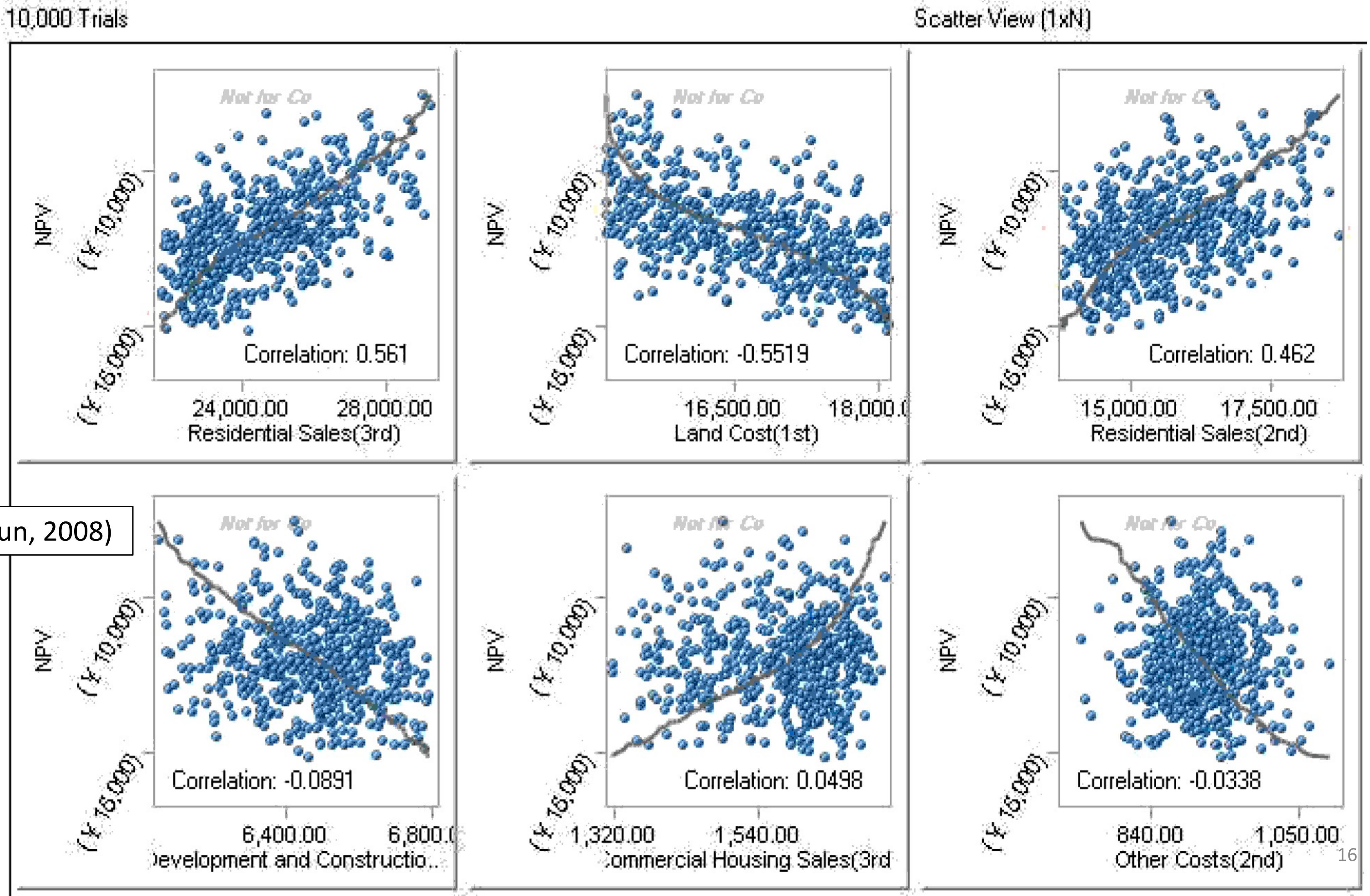
The uniform distribution



- Used when any value between the minimum and maximum is equally likely.
- It just looks like a horizontal line from the minimum to maximum value, with height 1/(MAX – MIN). You only need MIN and MAX to completely describe it.
- Excel's random number generator can be used to approximate drawing values from a uniform distribution (for example, by using RANDBETWEEN(MIN,MAX)).

Suppose we know the distribution

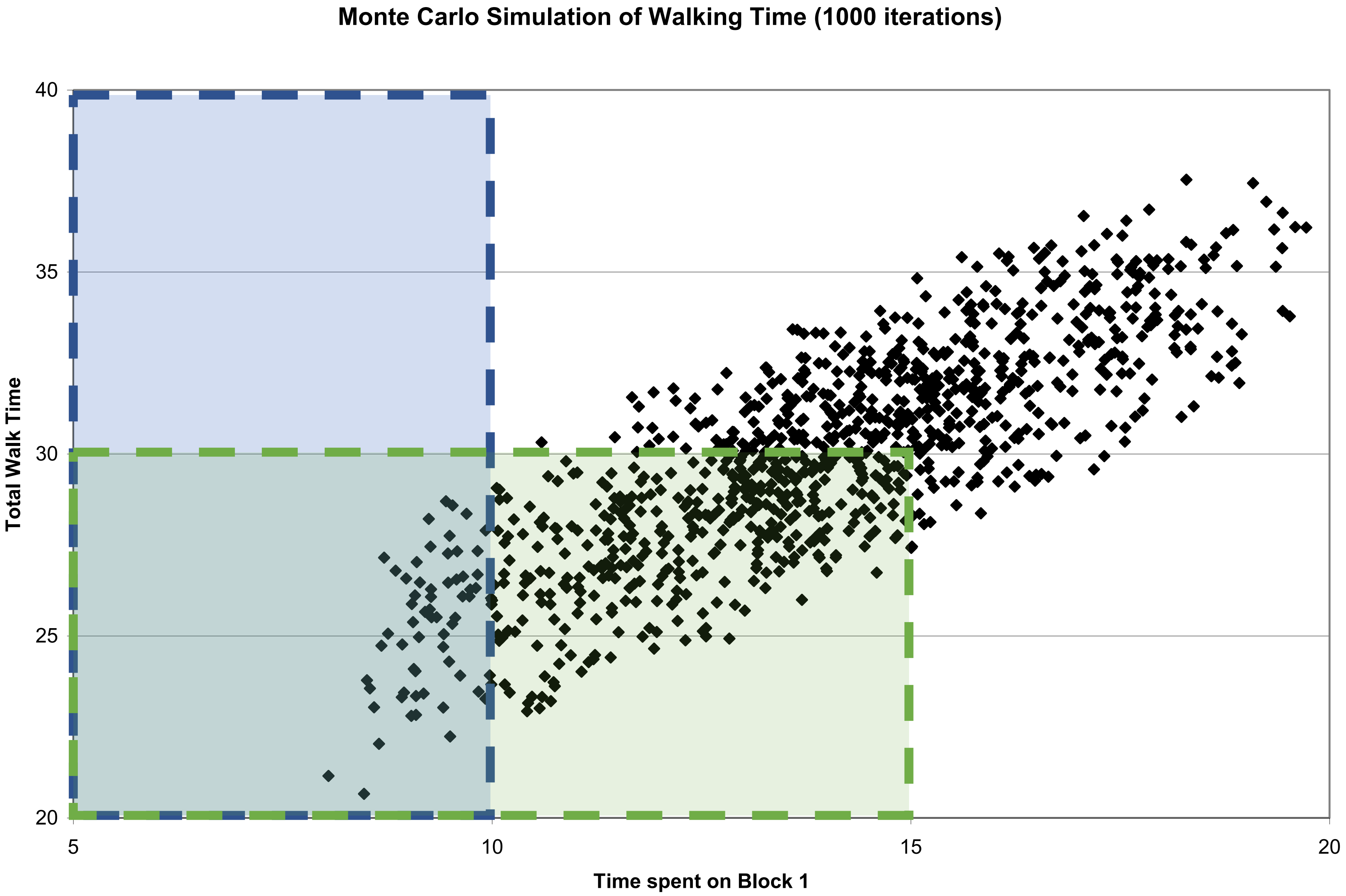
- Suppose you know the distributions for Cost and Benefits for a project.
- What’s the mean NPV? The median?
- How likely is it that the NPV will be negative?
- We can use a computer to generate variables that fit the distributions we require, and then plug them into our NPV equation.
- Doing this 1,000 (or more) times will give us a good idea of the result.
- This sort of analysis is called ‘Monte Carlo’, after the casino.


10,000 Trials
Scatter View (1xN)
NPV
(¥ 10,000)
(¥ 15,000)
Correlation: 0.561
24,000.00
28,000.00
Residential Sales(3rd)
Correlation: -0.5519
16,500.00
18,000.0
Land Cost(1st)
Correlation: 0.462
15,000.00
17,500.00
Residential Sales(2nd)
Correlation: -0.0891
6,400.00
6,800.0
Development and Constructio...
Correlation: 0.0498
1,320.00
1,540.00
Commercial Housing Sales(3rd
Correlation: -0.0338
840.00
1,050.00
Other Costs(2nd)

(Li & Sun, 2008)

Example: use and interpretation of MC graphs

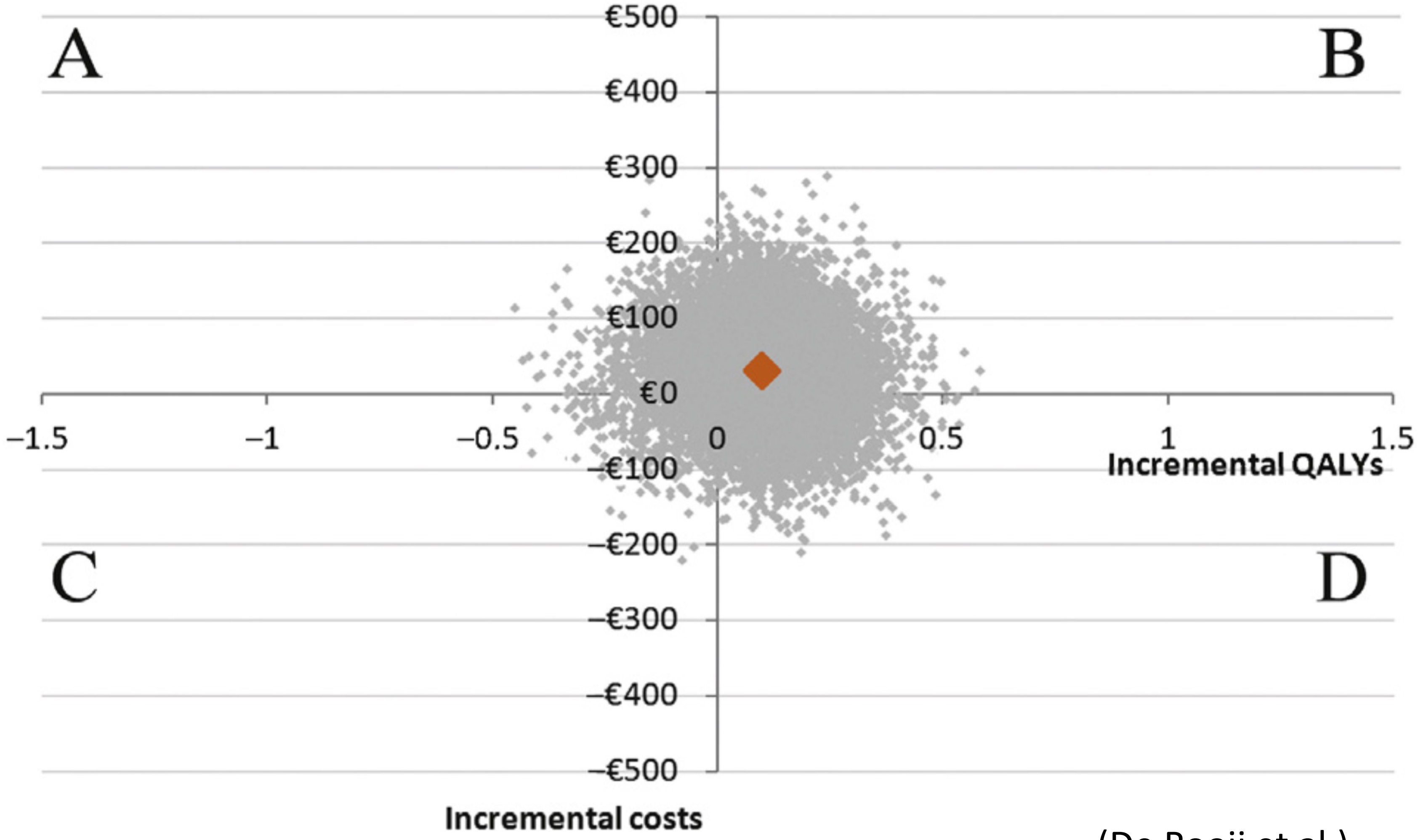
- You have a part-time job walking dogs up and down two city blocks.
- Exactly how long the walk takes varies, and depends on such factors as how interested dogs are in sniffing around on a particular day.
- Block 1 is in a dangerous neighbourhood, and your boss wants to minimize the amount of time you spend there.
- She gives you two choices: i) your walk must not last longer than 30 minutes, and at most 15 minutes can be spent on Block 1,
- or ii) you can take as long as you like on Block 2, but you must not spend more than 10 minutes on Block 1.
- Based on the Monte Carlo simulation graphed below, which **rule are you more likely to be able to follow?**



Since the 30-minute option (green) contains more dots than the other option (blue), accepting the 30-minute limitation is the better choice.

Note that for our simulation the blue option is a subset of a green option, but we CANNOT assume this will ALWAYS be true from a Monte Carlo graph alone.

Your turn: what about a third option – 30 minutes total time, with half that time or less spent on Block 1?



(De Rooij et al.)

Fig. 1 – Scatter plot of probabilistic sensitivity analysis for magnetic resonance imaging (MRI) versus transrectal ultrasound–guided biopsy (TRUSGB) strategies, including the total costs. (A) The MRI strategy is less effective and more expensive; (B) the MRI strategy is more effective and more expensive; (C) the MRI strategy is less effective and less expensive; (D) the MRI strategy is more effective and less expensive.

Five steps to a Monte Carlo Simulation

1. Deterministic Model
   - Analytical, as if values were known.
2. Probability distribution
   - Establish a reasonable one for each variable.
   - Limits, basic shape. (e.g. Normal distribution, mean 4, standard deviation 1.2)
3. Random sampling
   - Generate a set of variable values that fits the distributions and constraints.
4. Repeat Sampling
   - Calculate the number of repetitions needed for required confidence.
   - Generate that many sets of variable values, and evaluate the model for those values.
5. Summarize results
   - Histograms, scatter plots, tables, etc. Choose a summary tool to fit the task.

Soup and Sandwich go to Monte Carlo

	Minimum	Maximum	Baseline
Soup	\$2	\$5	\$4
Sandwich	\$3	\$9	\$7

- The cost of lunch is equal to Soup + Sandwich
- I want to know when lunch is affordable with a given budget.
- If my budget is \$8, how often will soup & sandwich be affordable? What if my budget is \$10?
- Since I have minimum, maximum & baseline values, I decided to use a **triangle distribution** defined by the minimum, maximum & baseline costs for Soup and Sandwich.
- I had Excel draw numbers from the appropriate distribution using the Excel formula developed by (Wright, n.d.).

Sample trials (Lunch = Soup + Sandwich)

<u>Step 1</u>: Draw parameter values from the appropriate distributions.
<u>Step 2</u>: Plug the numbers from Step 1 into your Value function & record the result.
<u>Step 3</u>: Repeat as needed & analyze.

Define the distribution for each parameter

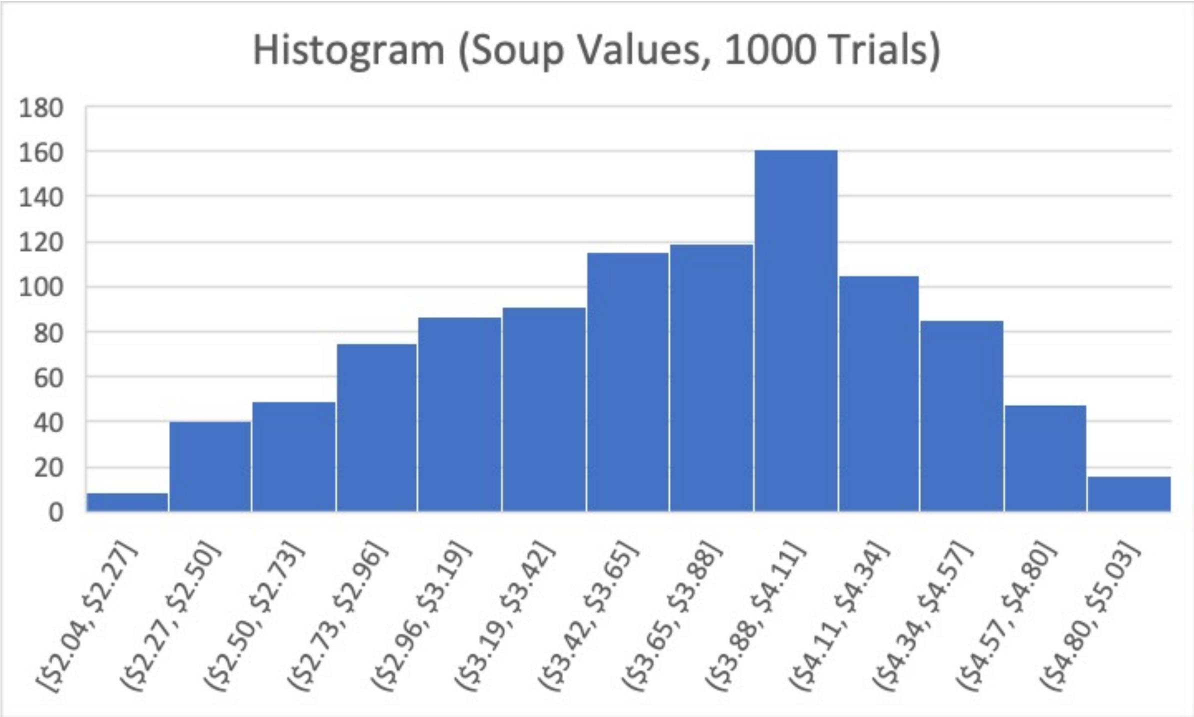
Variable	HELPER	Soup	HELPER	Sandwich	Lunch
Minimum	[redacted]	2	[redacted]	3	
Baseline	[redacted]	4	[redacted]	7	
Maximum	[redacted]	5	[redacted]	9	
Distribution	[redacted]	Triangle	[redacted]	Triangle	
Trial	P(X)	Value	P(X)	Value	Value
1	[redacted]	$4.17	[redacted]	$7.55	$11.72
2	[redacted]	$2.81	[redacted]	$6.77	$9.58
3	[redacted]	$4.33	[redacted]	$4.49	$8.82
4	[redacted]	$3.72	[redacted]	$4.25	$7.97
5	[redacted]	$4.56	[redacted]	$7.70	$12.26

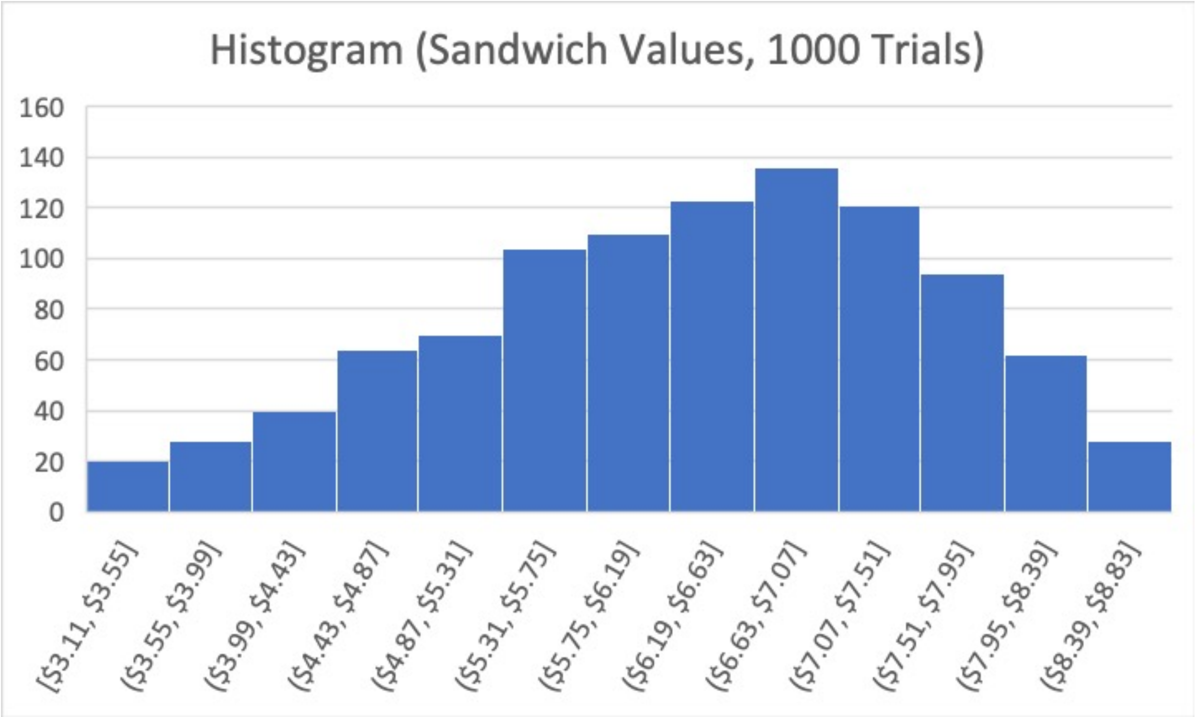
Plug into the value function.

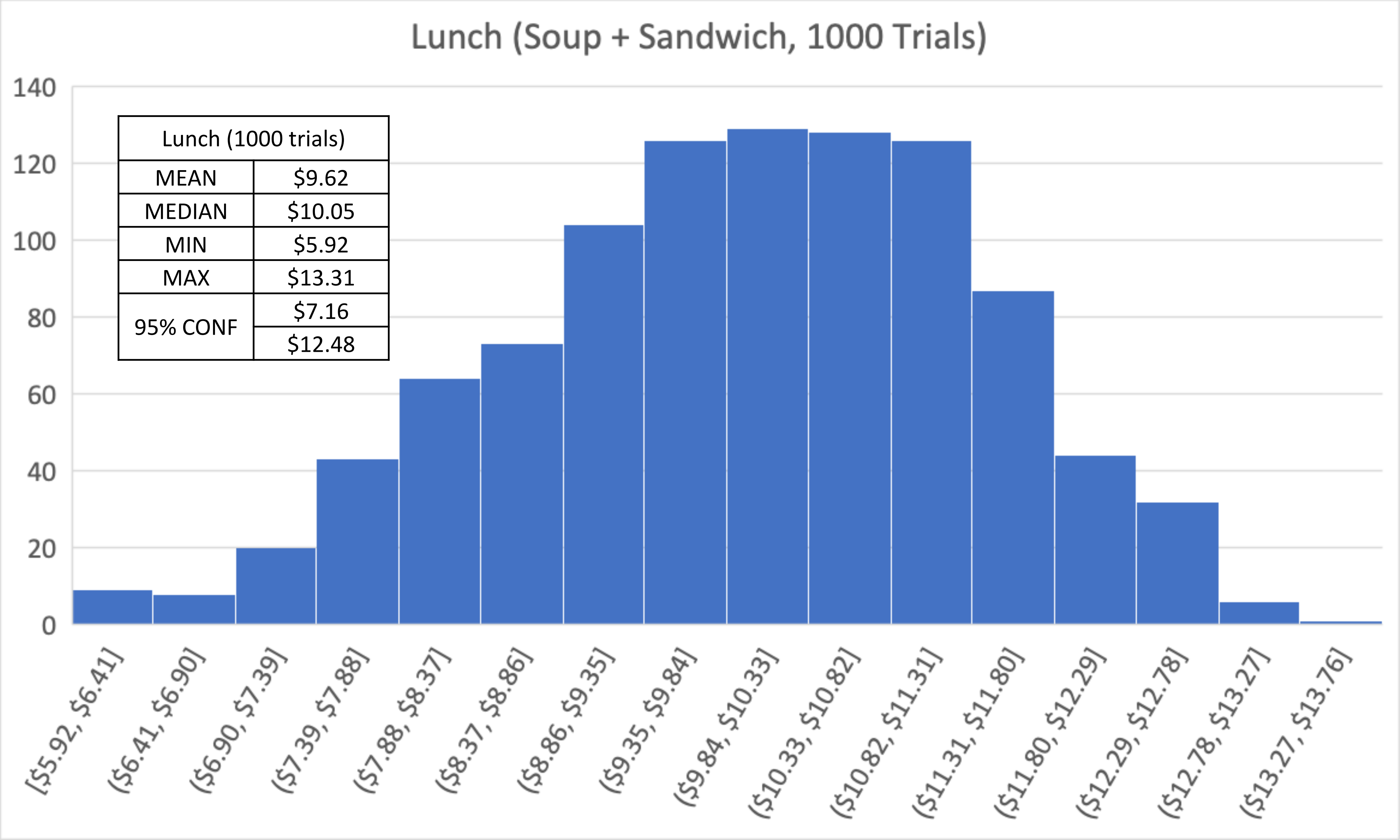
Draw from the distributions.

Don't mind these. They're 'under the hood' calculations used in my specific implementation of a triangle distribution.

Results of sampling from triangle distribution

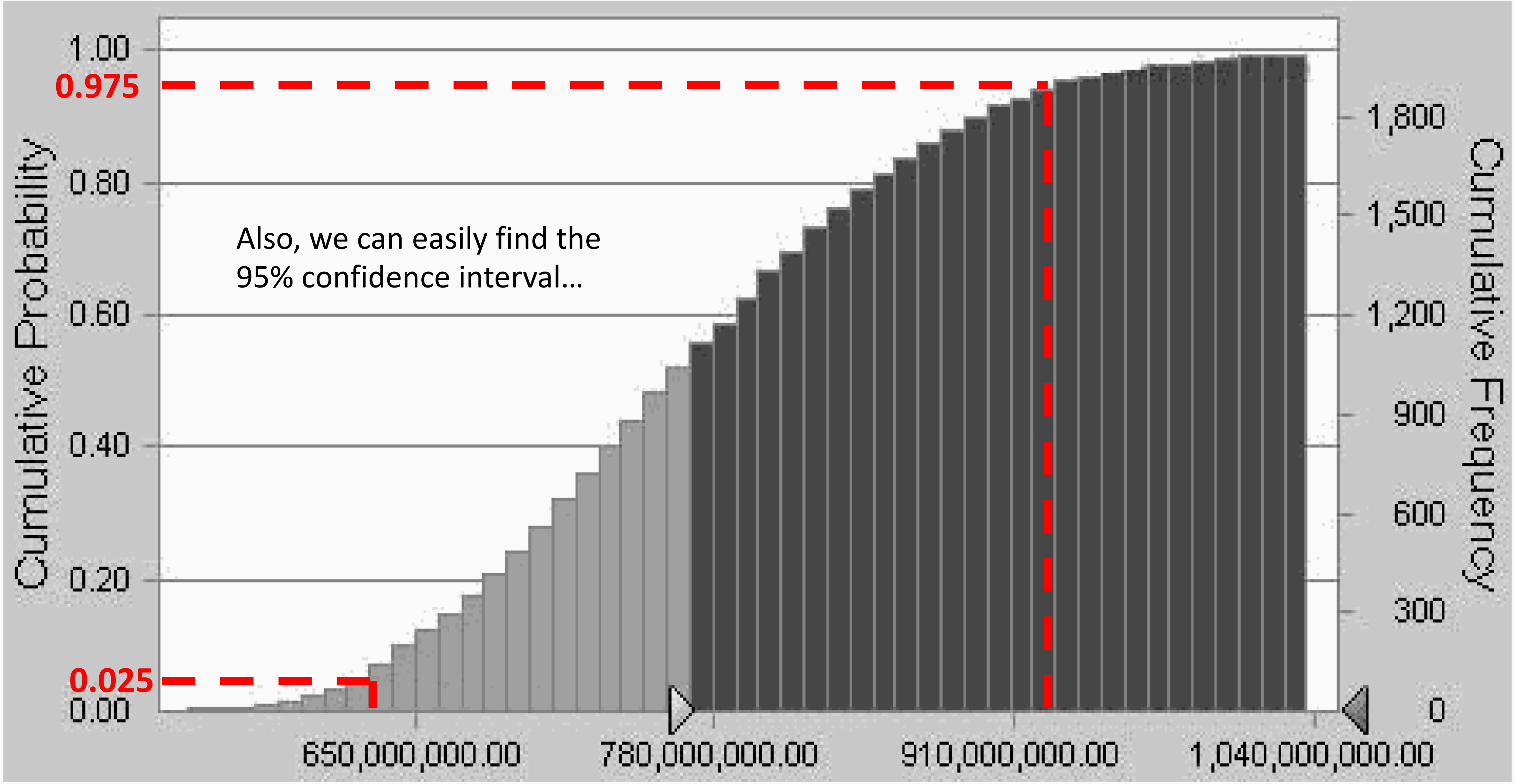





Lunch (Soup + Sandwich, 1000 Trials)
Lunch (1000 trials)
MEAN $9.62
MEDIAN $10.05
MIN $5.92
MAX $13.31
95% CONF $7.16
$12.48
0
20
40
60
80
100
120
140
[$5.92, $6.41]
($6.41, $6.90]
($6.90, $7.39]
($7.39, $7.88]
($7.88, $8.37]
($8.37, $8.86]
($8.86, $9.35]
($9.35, $9.84]
($9.84, $10.33]
($10.33, $10.82]
($10.82, $11.31]
($11.31, $11.80]
($11.80, $12.29]
($12.29, $12.78]
($12.78, $13.27]
($13.27, $13.76]

Presenting the results

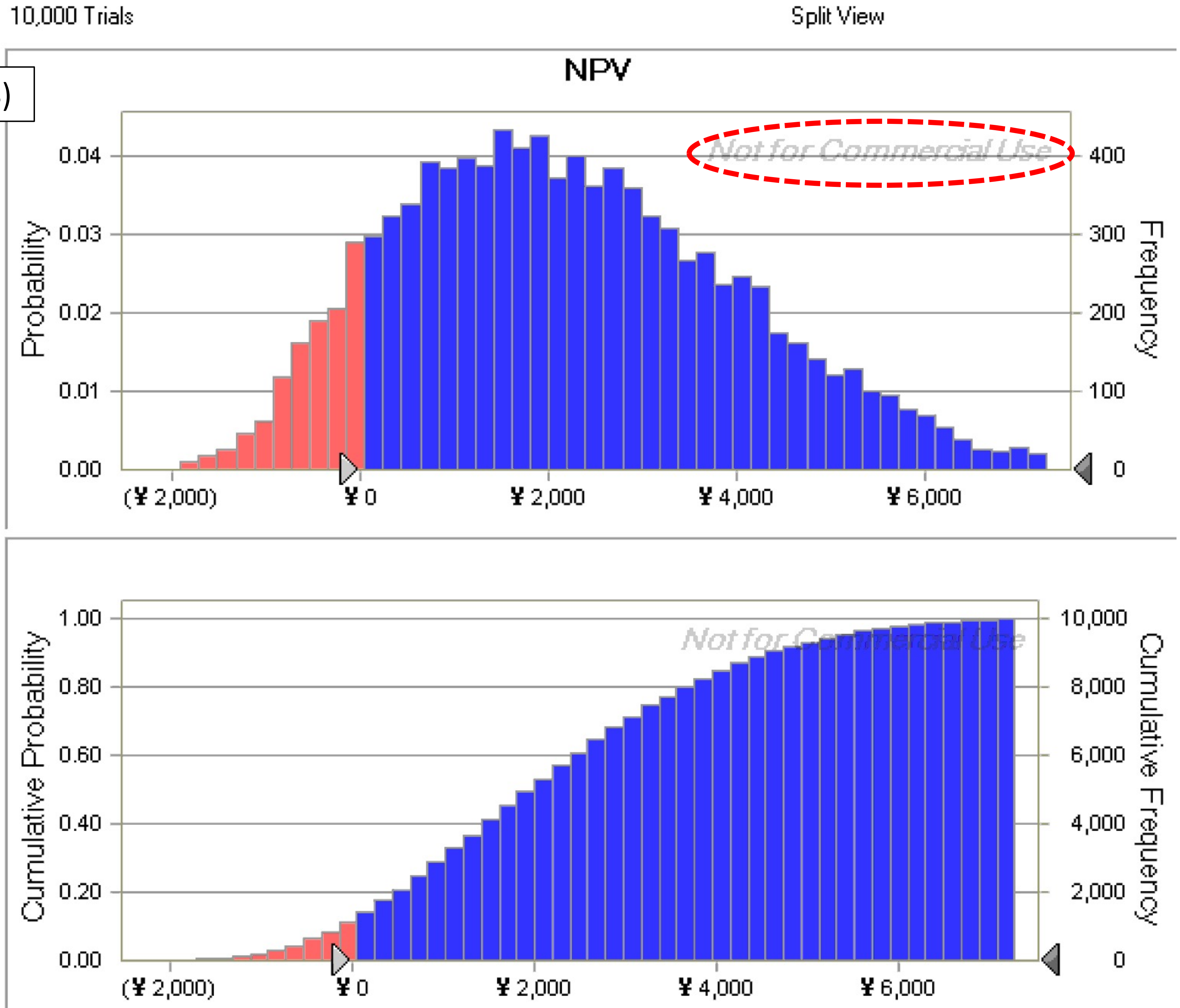
- For project evaluation, we're most often concerned with answering the question, "How likely is it that alternative 1 is better than alternative 2?"
- Another common question: "How likely is it that NPV<0, or BCR<1?"
- The standard way to answer this is via a cost acceptability curve, CAC.
- After we've created the simulated values…
- ...create a cumulative probability density function, showing what proportion of simulations were equal to or less than the value on the horizontal axis.
- This diagram makes it very easy to see how feasible a target is.



(Jinbo, 2009)

Figure 2. Cumulative frequency of $NPV_{min} \mid t{=}17$.

(Li & Sun, 2008)



10,000 Trials
Split View
NPV
Not for Commercial Use
Probability
Frequency
0.04
0.03
0.02
0.01
0.00
400
300
200
100
0
(¥ 2,000)
¥ 0
¥ 2,000
¥ 4,000
¥ 6,000
Cumulative Probability
Cumulative Frequency
1.00
0.80
0.60
0.40
0.20
0.00
10,000
8,000
6,000
4,000
2,000
0
Not for Commercial Use
(¥ 2,000)
¥ 0
¥ 2,000
¥ 4,000
¥ 6,000

AFTER HOURS

- Creating a CAC in Excel (2 slides)
- Bootstrapping (6 slides)

How to create a CAC in Excel

- Calculate the value of interest (e.g. NPV or BCR) for each trial.
- Sort the values in ascending order.
- Create an adjacent column, and have it be a counter starting at 1/N and going up to N/N (=1), where N is the number of trials.
- Plot the two columns as a smooth-line scatter plot with the counter column on the horizontal axis.
- Tidy things up (axis limits and number type, etc.)
- The 95% confidence interval is equal to the values between the row with counter column index 0.025, and the row with counter column index 0.95 (corresponding to 2.5% and 97.5%).

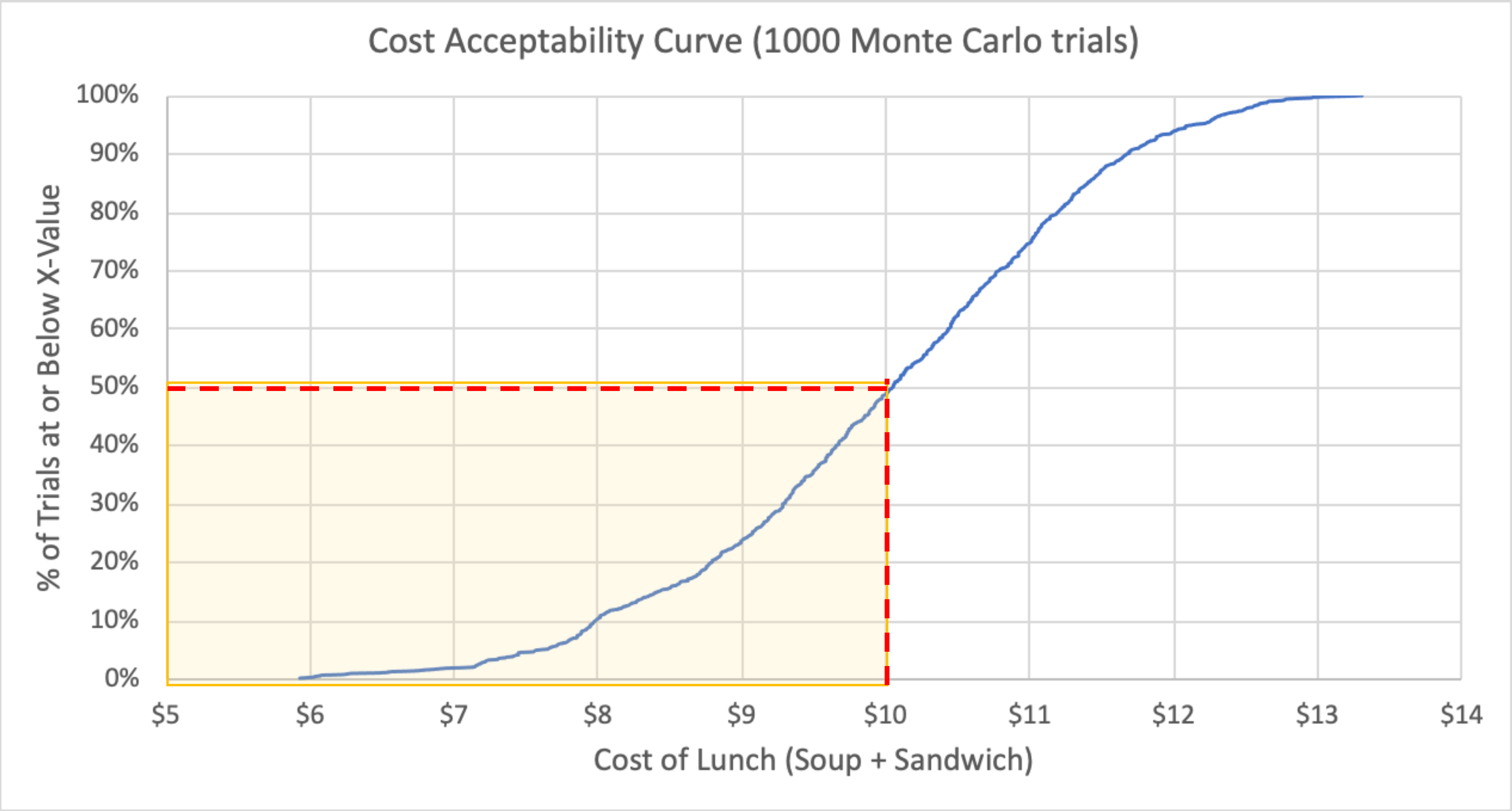
Lunch (1000 trials)	
MEAN	$9.62
MEDIAN	$10.05
MIN	$5.92
MAX	$13.31
95% CONF	$7.16
	$12.48

Lunch	%
$5.92	0.1%
$5.98	0.2%
$6.02	0.3%

$7.16	2.4%
$7.16	2.5%
$7.18	2.6%

$12.48	97.4%
$12.48	97.5%
$12.49	97.6%

$12.97	99.7%
$13.07	99.8%
$13.22	99.9%
$13.31	100.0%



Y-axis: Proportion of simulations *below* the budget: e.g. 50% for a budget of $10, or about 10% for a budget of $8..

95% confidence interval: [$7.16,$12.48] (since 95% of trials were in this range)

What if you don’t know the distribution?

- Often, we conduct studies because we DON’T know the distribution of the values of interest.
- Or, it may be very expensive to sample the values, so you have a limited number of observations to work with.
- Suppose all you have are three (but hopefully many more!) values for each of the variables. What can we do with this?
- As long as our sample is representative and free of systematic bias…
- ...we can use a technique called ‘bootstrapping’ to get a pretty good approximation of means, standard deviations, confidence intervals, etc. for the population.

Bootstrapping: If it DOES happen, it CAN happen.

- Bootstrapping works a lot like Monte Carlo simulation…
- ...only instead of asking the computer to generate a random number fitting a distribution...
- ...we ask it to randomly pick an appropriate number from the data we have.
- (This is easy to do in Excel with VLOOKUP.)
- It’s as if you wrote each entry on a piece of paper, and put them in a hat.
- For each trial, the computer draws a ballot from the hat, makes a note of the number, then replaces the piece of paper in the hat (’draw with replacement’).
- For our bootstrapping example, I took the first 100 trials from the Monte Carlo simulation and used them as our “observations“.

Setting things up in Excel...

Caution! The ‘lock’ I have on the other random number generators in the companion spreadsheet doesn’t work with VLOOKUP.

Index	Soup	Sandwich
1	$3.99	$8.31
2	$2.82	$7.90
3	$2.81	$3.55
4	$3.56	$4.51
5	$4.05	$7.05
95	$3.77	$4.19
96	$3.58	$6.86
97	$4.40	$4.96
98	$4.15	$7.65
99	$3.88	$6.95
100	$4.33	$4.53
COLUMN	2	3

Careful! Excel re-rolls random numbers whenever *anything* changes in the spreadsheet.

Copy the values you want, then Edit → Paste Special → Values to keep them from changing.

Formula that tells Excel to pick a number between 1 and 100, and retrieve the value found in (COLUMN, Index):

=VLOOKUP(ROUND(RANDBETWEEN(1,100),0),[Table Range],COLUMN,FALSE)

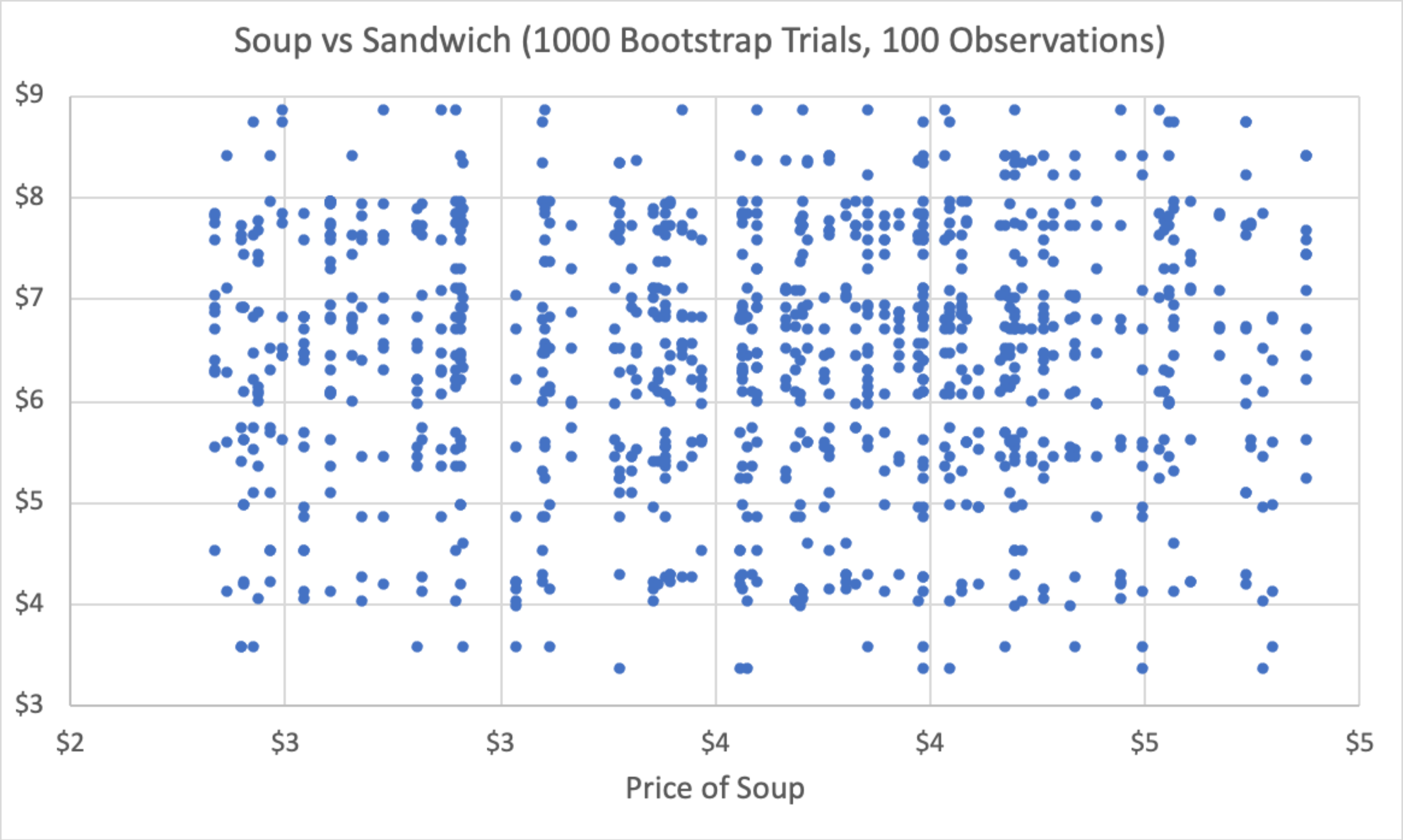
We need the rounding because RAND() returns decimals, and all we want is an integer from 1 to 100. In one of my spreadsheets, from which the tables are taken, the exact code to obtain a Soup value was:

=VLOOKUP(ROUND(RANDBETWEEN(1,100),0),$A$1:$C$102,2,FALSE)

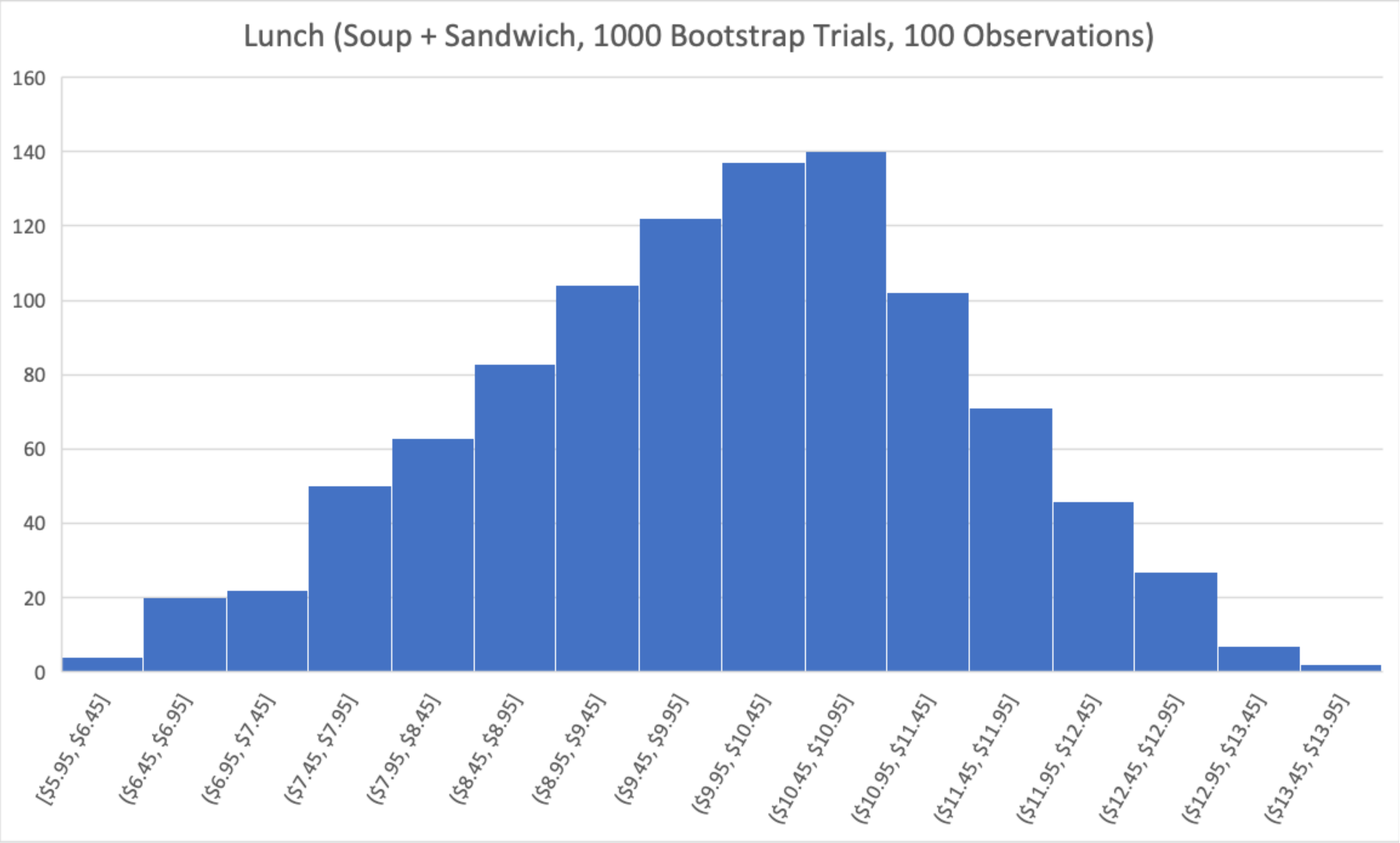
For a Sandwich value, the COLUMN value would be 3 instead of 2. The dollar signs are there so the range doesn’t change when I ‘Fill’ a column.

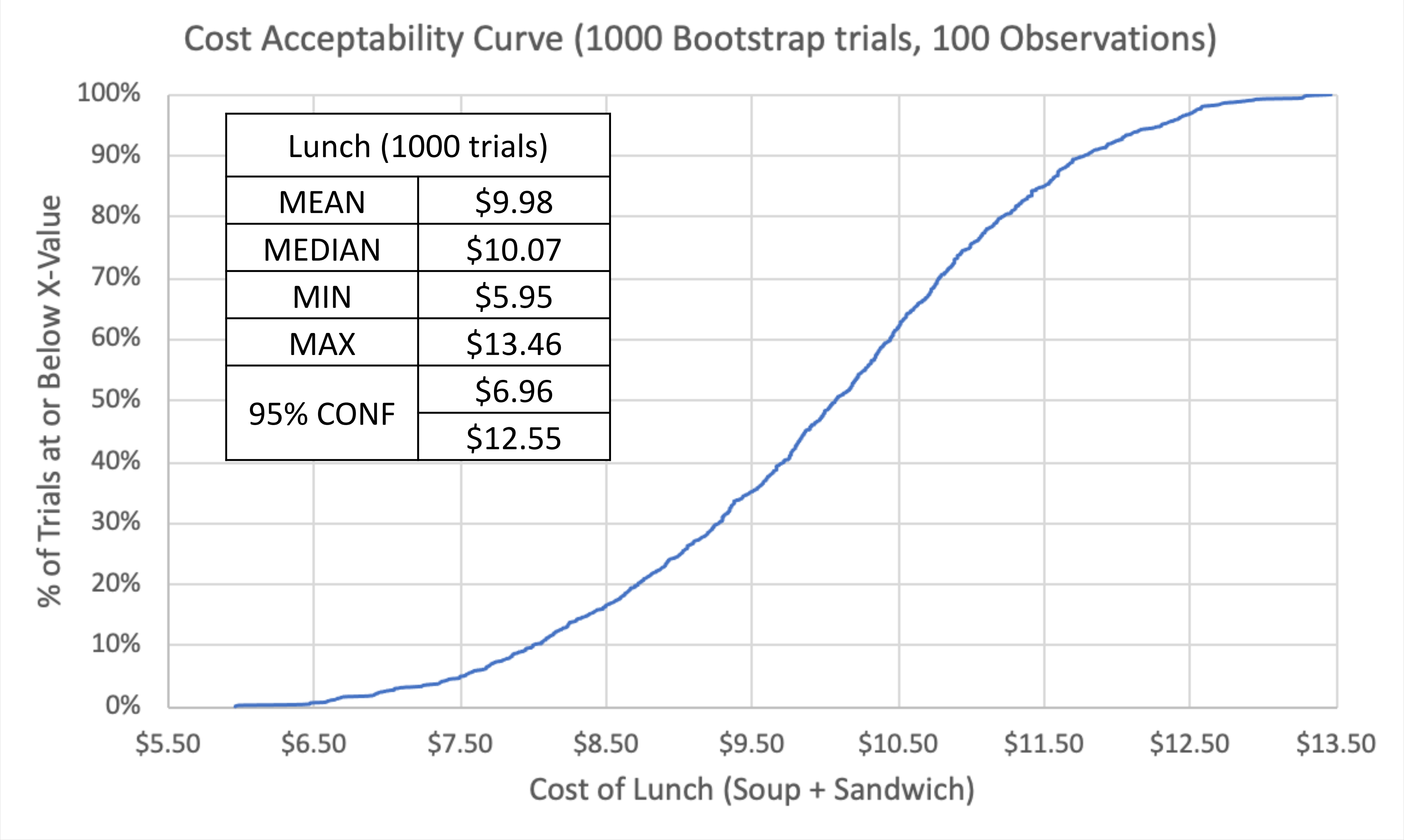
Trial	Soup	Sandwich	Lunch
1	$3.99	$8.31	$12.30
2	$2.82	$7.90	$10.72
3	$2.81	$3.55	$6.36
4	$3.56	$4.51	$8.07
5	$4.05	$7.05	$11.10
6	$3.57	$7.42	$10.99

Excel rolls a virtual one-hundred-sided die, and writes down either the Soup or Sandwich value corresponding to that Index, as appropriate. Add to get Lunch cost.



The plot is 'boxier' and has white areas because we are limited by our observations.


Lunch (Soup + Sandwich, 1000 Bootstrap Trials, 100 Observations)
160
140
120
100
80
60
40
20
0
[$5.95, $6.45]
($6.45, $6.95]
($6.95, $7.45]
($7.45, $7.95]
($7.95, $8.45]
($8.45, $8.95]
($8.95, $9.45]
($9.45, $9.95]
($9.95, $10.45]
($10.45, $10.95]
($10.95, $11.45]
($11.45, $11.95]
($11.95, $12.45]
($12.45, $12.95]
($12.95, $13.45]
($13.45, $13.95]



Lunch (1000 trials)	
MEAN	$9.98
MEDIAN	$10.07
MIN	$5.95
MAX	$13.46
95% CONF	$6.96
	$12.55