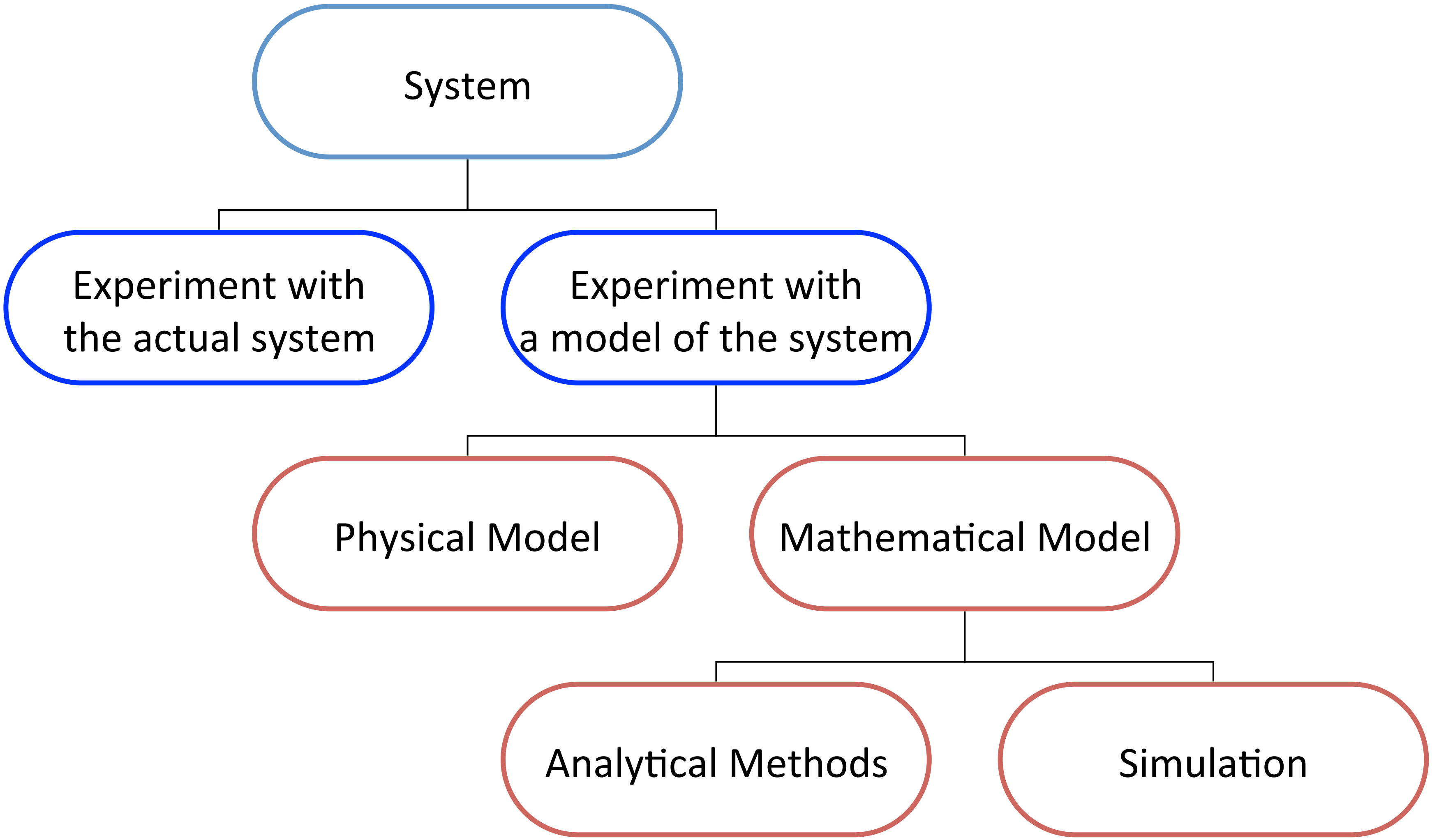
Building Descriptive Models: Simulation

Ways of Studying a System



Selection of Different Methods

- Optimization (LP, IP, MILP, NLP)
  - Finds "best" solution or recommendation (Prescriptive)
  - Best suited for deterministic problems with no uncertainty
- Regression
  - Measures impact of independent variables on a dependent variable (Predictive)
  - Best suited for establishing relationships between variables
- Simulation
  - Captures outcomes of different policies within an uncertain or stochastic environment (Descriptive)
  - Best suited for situations where the exact relationship between decisions and outcome are not known

Pros and Cons of Using Simulation

- Use simulation models to . . .
  - capture complex system interactions.
  - model system uncertainty.
  - systematically experiment with new system designs or decision rules.
  - train or experiment without disrupting real operations.
  - generate data to analyze, describe, and visualize interactions, outcomes, and sensitivities.

- Do not use simulation models if . . .
  - the problem can be solved analytically (simulation is expensive!)
  - direct experimentation with the actual physical system is feasible and cost effective.
  - cost and/or time requirements are prohibitively expensive.

Simulation Models in Supply Chains

- Manufacturing processes
  - How well do sequential processes interact with each other?
- Flow of goods in a warehouse
  - How do different picking rules perform?
- Vehicle routing from a distribution center
  - Which routing heuristics work best?
- Service operations in a call center
  - How does the system operate with more staff?
- Inventory replenishment policies
  - How well do different re-order points and order quantities perform?

Simulation Example:

Suzie's Sushi Shack

Suzie's Sushi Shack

- Suzie sells sushi in the Seychelles from her shack at the seashore . Each day she buys certain fish from the local fishermen and creates a number of her special rolls. If she has extra unsold rolls at the end of the day, she has to throw them out (no one eats day old sushi). If she sells out what she has during a day, she is done and cannot make any more rolls that day. Each roll costs 25 Seychelles Rupees (SCR) to make and she sells them for 80 SCR each.
- The demand for her sushi rolls is highly variable. She has collected two months of demand, to include the number of times customers requested rolls when she was already sold out.
- Her rule of thumb has been to make 1.5 times the average number of rolls typically demanded each day. She is reporting an average profit of about 800 SCR a day.
- Suzie wants to know if this policy makes sense, and if not, how many rolls she should make each day to maximize her profit.

Simple Steps in a Simulation Study

1. Formulate & plan the study
2. Collect data & define a model
3. Construct model & validate

    Go to Step 2 as many times as needed!

4. Make experimental runs
5. Analyze output

Step 1. Study Plan

- Develop a simulation model where daily demand varies.
- A "production policy" will be applied.
- Based on the demand and the policy, we will calculate profitability.
- We will assess profitability and other performance metrics of different policies

Collect Data and Define Model

Step 2. Collect Data and Define Model

a) Determine random variables

X = number of sushi rolls demanded each day

b) Determine relationships between various variables

s = sales price of each sushi roll (SCR/roll)
c = cost of each sushi roll (SCR/roll)
Q = quantity of sushi rolls made each day
R = daily revenue = s*MIN(Q,X)
C = daily cost = c*Q

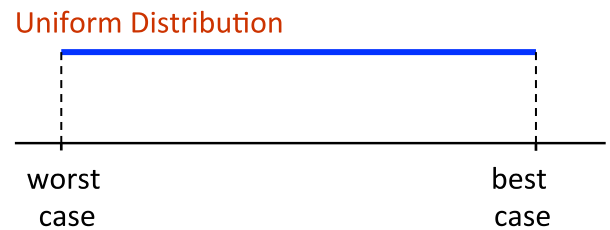
c) Determine performance metrics

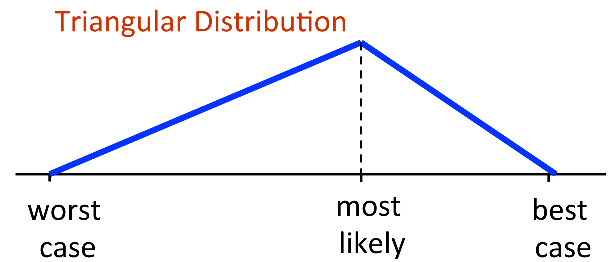
Daily Profit = R – C
Daily Unfilled Demand = MAX(X-Q, 0)
Daily Extra Rolls = MAX(Q-X, 0)

d) Collect data & estimate probability distributions

Determining Random Variable Distribution

- If sample data is **not** available . . .
  - Determine the "range" of the variable
    - Talk to stakeholders or experts
    - Get worst/best/most likely values
  - Use known distributions
    - Number of customers arriving ~ Poisson
    - No information ~ Uniform or Triangular





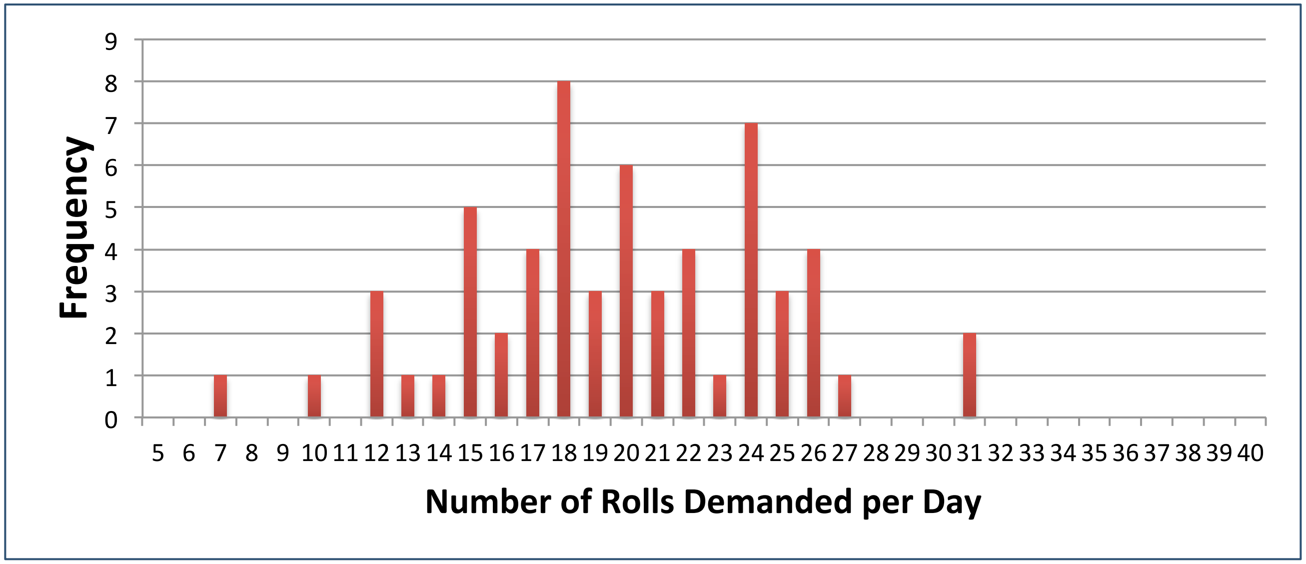
- If sample data is available . . .
  - Examine histograms
  - Calculate summary sample statistics
  - Conduct Chi-Square tests to fit sample to "traditional" distributions, (i.e., Normal, Poisson, Exponential etc.)
  - Use a "custom" empirical distribution (be careful of over-fitting!!)
    - Discrete Empirical – use % of observation as probabilities
    - Continuous - Use histogram to compute probabilities of each range and then "uniform" within the range

Testing the Fit for Sample Data

Chi-Square Test

- Recall the general approach
  - Create buckets or categories, c, that make sense
  - Count the expected and observed (actual) values in each category
  - Calculate the Chi-square statistic and find the p-value
  - If the p-value is greater than level of significance, reject the null hypothesis that the data follows the proposed (expected) distribution

$$\chi^2 = \sum \left(\frac{(Observed - Expected)^2}{Expected} \right) \qquad df = c - 1$$



Min	7
25th Pct	17
Mode	18
Median	20.00
Mean	19.77
75th Pct	24
Max	31
StDev.S	4.91
IQ Range	7

Chi-Square Test

Looking at the data, we could set c=1, but let's use c=2 in order to illustrate the process.

Day	Demand (X)
1	25
2	30
3	17
4	15
5	16
6	19
7	31
8	13
9	17
10	25
11	16
12	20
13	13
14	24
15	15
16	33
17	18
18	20
19	6
20	24
.	.
.	.
.	.
60	23

=COUNTIF(*DemandArray*,"<=”&*ValueCell*)

X≤	CumObs	Obs	ExpCumProb	Expected
6	1	1	0.01	0.56
8	1	0	0.02	0.75
10	2	1	0.05	1.47
12	4	2	0.09	2.58
14	9	5	0.16	4.05
16	18	9	0.25	5.66
18	26	8	0.37	7.08
20	36	10	0.50	7.91
22	42	6	0.63	7.91
24	48	6	0.75	7.06
26	53	5	0.84	5.63
28	54	1	0.91	4.02
30	56	2	0.95	2.56
32	58	2	0.98	1.46
34	59	1	0.99	0.74
36	59	0	1.00	0.34
38	60	1	1.00	0.14
40	60	0	1.00	0.05

=*CumObs$_i$* – *CumObs$_{i-1}$*

=NORM.DIST(*ValueCell*,*Mean*,*StdDev*,1)

=N*(*ExpCumProb$_i$* – *ExpCumProb$_{i-1}$*)

=CHISQ.TEST(*ObsArray*, *ExpectedArray*)

0.71

H_0: X~N(20,5)

H_1: X not ~N(20,5)

alpha = 0.01

p = 0.71

Do NOT Reject H_0

Assume X~N(20,5)

Define and Validate the Model

Generating Random Variables

- Underlying principle
  - Generate a random (or usually pseudo-random) number
  - Transform that RV to fit the desired distribution
- Manual Techniques:
  - Rolling a die ~U(1, 2, 3, 4, 5, 6)
  - Turning a roulette wheel ~U(38 options) ala Monte Carlo
  - Flipping open a book and picking last digit
  - Random number tables
- In Spreadsheets – there are other functions in add-ins:
  - RAND() = Returns a continuous variable between 0 and 1
  - Uniform Distribution ~U(a,b)
    - ~U(a, b) = a + (b – a) * *RAND*()
    - ~U(500, 900) =500 + 400 * *RAND*()
  - Normal Distribution ~N(μ, σ)
    - ~N(μ, σ) =*NORMINV*(*RAND*(), μ, σ)
    - ~N(650, 75) =*NORMINV*(*RAND*(), 650, 75)
  - Rand() refreshes upon hitting enter, F9, or Ctrl-R

The Sushi Simulation Model

Data:

Sales Price	80	SCR
Cost	25	SCR
Mean μ	20	#/day
StdDev σ	5	#/day
Production Quantity	30	#/day

=INT(NORM.INV(RAND(),*μ*,*σ*))

=MAX(*Demand-ProductionQuantity*,0)

=MAX(*ProductionQuantity - Demand*,0)

=MIN(*Demand, ProductionQuantity*)**Price*

=*ProductionQuantity*Cost*

=*Revenue - Cost*

Day	Demand	#Short	#Extra	Revenue	Cost	Profit
1	17	0	13	1360	750	610
2	18	0	12	1440	750	690
3	26	0	4	2080	750	1330
4	33	3	0	2400	750	1650
5	28	0	2	2240	750	1490
6	25	0	5	2000	750	1250
7	18	0	12	1440	750	690
8	18	0	12	1440	750	690
9	15	0	15	1200	750	450
. . .	. . .	. . .	. . .	. . .	. . .	. . .

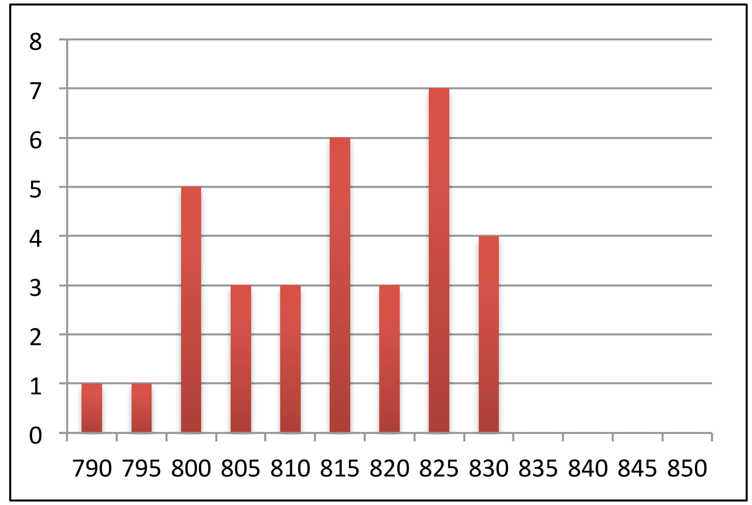
Notes:
- Ran 300 simulated days
- Refreshing the RAND() function changes these numbers!

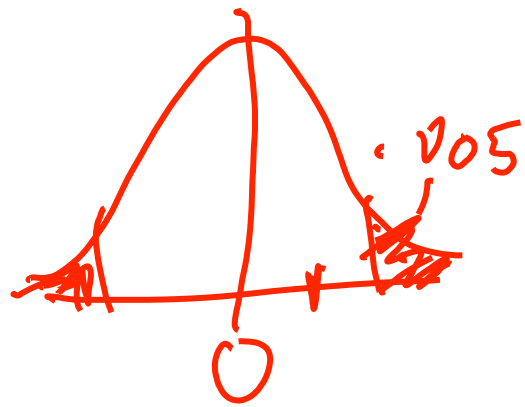
	Demand	#Short	#Extra	Revenue	Cost	Profit	
Min	1	0	0	80	750	-670	Min
25th Pct	17	0	7	1360	750	610	25th Pct
Mode	21	0	9	1680	750	930	Mode
Median	20	0	10	1600	750	850	Median
Mean	19.45	0.02	10.57	1554	750	804	Mean
75th Pct	23	0	13	1840	750	1090	75th Pct
Max	32	2	29	2400	750	1650	Max
StDev.S	5.11	0.17	5.07	406	0	406	StDev.S
IQ Range	6	0	6	480	0	480	IQ Range

Validation:
- Does the "as-is" simulation match reality?

Validation of Model

- Validation Runs
  - Made 40 runs of 300 days each
  - Sample mean ranged from 763 to 850
    - Mean of Sample Means ~N(806, 20.9)
    - 99% Confidence Interval [797, 816]
  - Hypothesis Test:
    - H_0: $\bar{X}$ = 800
    - H_1: $\bar{X}$ ≠ 800
  - This is a two tailed test, ~N since 40 obs.
  - For alpha = 0.01
    - Critical Value = NORM.S.INV(0.01/2) = +/- 2.57
  - Test statistic (using the sample standard deviation)
    - $z = (\bar{X} - \mu_{\bar{X}})/\sigma_{\bar{X}} = = (\bar{X} - \mu_{\bar{X}})/(\sigma_x/\sqrt{n}) = (806 - 800)/(20.9/\sqrt{40}) = 1.82$
  - So, we reject H_0 if z < -2.57 or if z > +2.57
  - Since z = 1.82, we cannot reject the Null Hypothesis and assume the model is valid in that it replicates the current situation.





Make Experimental Runs

Making Experimental Runs

- Need to make multiple runs for each policy
- Use Hypothesis Tests to evaluate results
- Remember, simulation will never recommend a solution – only evaluate the quality of proposed solutions

Production Quantity	E[Profit]	StdDev [Profit]
30	806	20.9
25	906	10.13
20	921	8.51
15	784	3.01

- These few runs suggest that Suzie should consider making fewer Sushi rolls each day. Too much is being thrown out.
- In later courses we will show how to set the optimal Q!

Final Notes on Simulation

Simulation Modeling Tips for Spreadsheets

- Start simple. Add complexity as needed
  - Helps model validation
  - Keep track of model versions
- Keep input parameters in separate sheets
- Keep performance metrics in separate sheets
  - Macros will be useful to compile multiple run results
- Random numbers change every time a cell changes.
  - Consider creating a separate table with all random numbers in a separate tab.
  - Consider a different random stream for each random variable
  - consider "variance-reduction" when comparing multiple scenarios
- Remember output results are random variables
  - Need multiple runs for each scenario
  - If dynamic system, may need a warm-up period and run it for a long time. Initial conditions matter on dynamic systems.
  - Think carefully about number of "runs" needed for each "scenario".
  - When comparing scenarios you need to use statistical hypothesis testing to give final conclusions

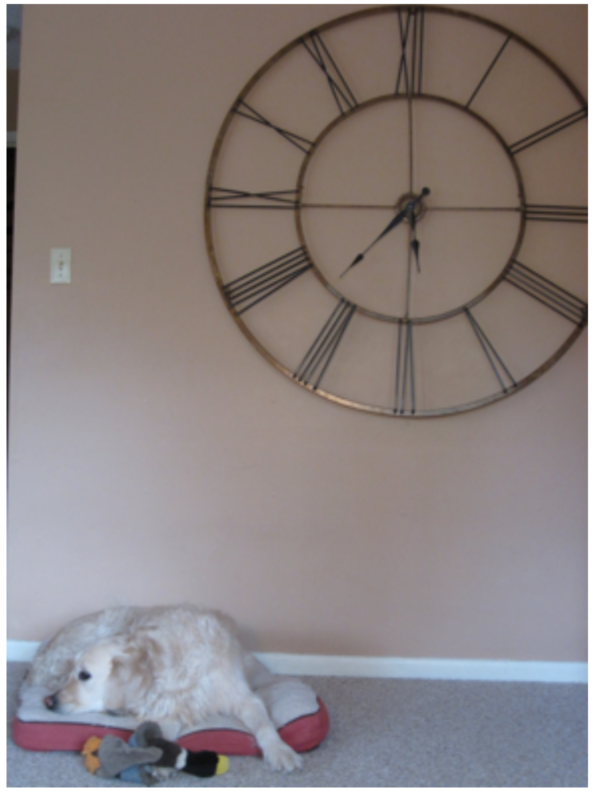
Types of Simulation Models

- Classes of Models
  - System Dynamics
    - Models the non-linear behavior of complex systems
  - Monte Carlo Simulation
    - Uses random sampling to mimic specified distribution
  - Discrete Time Simulation
    - Models the operation of a system as a discrete sequence of events in time. Each event occurs at a particular instant in time and marks a change of state in the system.
  - Agent Based Simulations
    - Models behavior and interaction of individual agents within a system using defined rules
- Suzie's Sushi Simulation Model
  - Static Monte Carlo Simulation with Stochastic Demand
  - Each day did not depend on the "state of the system" from the previous day

Common pitfalls in simulation studies

- Not well defined objectives
- Lack of communication with stakeholders
- Inappropriate level of detail
- Treating study as a primarily mathematical or programming exercise
- Failure to take into account all sources of randomness in the model
- Making a single "run" to obtain conclusions
- Using output statistics as "true" answer
- Using wrong measures to evaluate the system

Questions, Comments, Suggestions?
Use the Discussion Forum!



"Dexter – simulating his way to dinner time"
Yankee Golden Retriever Rescued Dogs
(www.ygrr.org)

MIT Center for
Transportation & Logistics

caplice@mit.edu
ctl.mit.edu