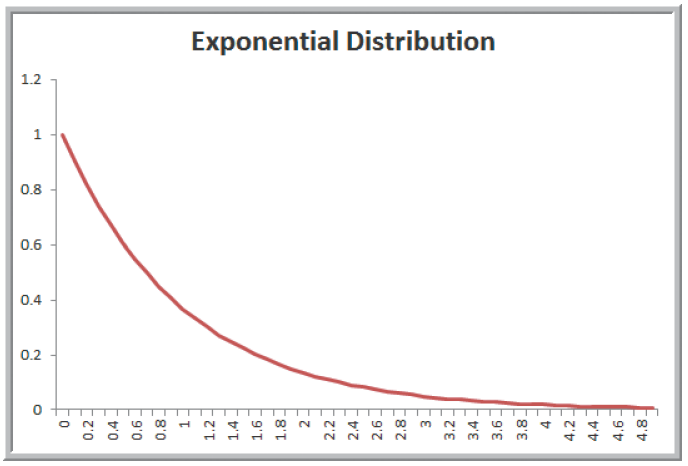
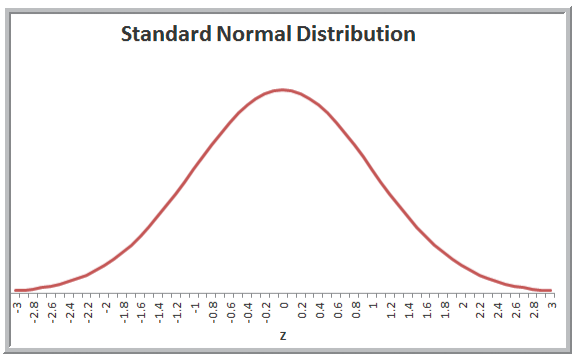
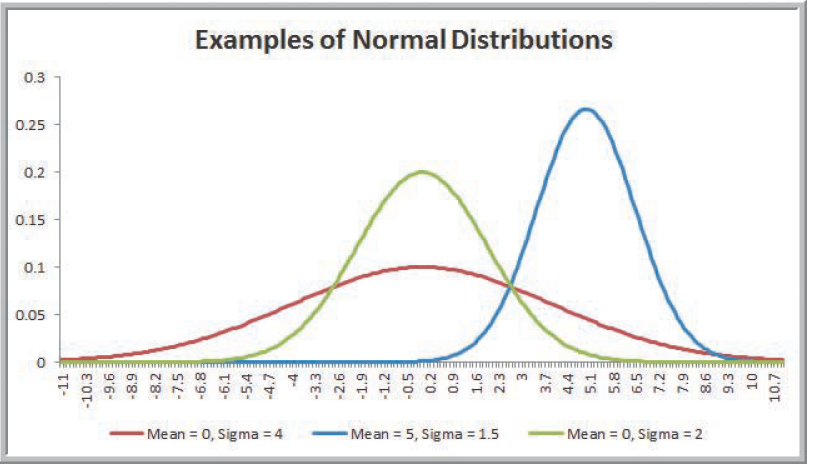
# Class 7 – Simulation and Risk Analysis – 2017-10-03

Reading 7 p156-172; 379-402

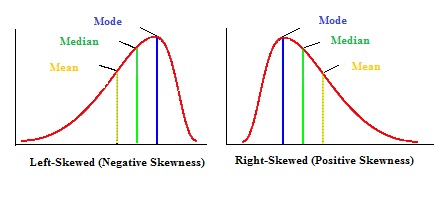
**Probability Distributions (Recap)**

|  |
| --- |
| **Potential Distributions for the Exams (re: ICI case)**   * GIVEN: All costs, revenues are normally distributed **=NORM.INV(RAND(),MEAN,STDEV)**   + STDEV = MEAN\*VARIABILITY/#STDEV (eg $25M\*50%/3) = variance of 3 stdev * GIVEN: All probabilities are uniform **=RAND()\*(A-B)+B** |

* **Normal Distribution:** Continuous distribution escribed by bell curve (symmetric; mode, media, mean are equal; tails are unbounded; 1 stdev = 68.3% under curve, 2 stdev = 95.4%; 3 stdev = 99.7%)
  + **=NORM.DIST(**x,mean,stdev,True**) Function:** Calculate probability to the left of X
  + **=NORM.INV(**prob,mean,stdev**) Function:**  Calculate value of X based on given cumulative probability
* **Standard Normal Distribution, Z:** Normal distribution centered at 0 (μ=0, σ=1)
  + **=NORM.S. DIST(**z**) Function:** Probability to the left of z
  + Convert any value X into equivalent standard normal value:
* **Exponential Distribution:** Continuous distribution models time **between** random occurring events (eg: customer arrivals, failure of machines etc) where λ is a rate (#/time) and μ is in units of time (bounded by 0, greatest density at 0, and declines as X increases)
  + Density Function: Cumulative Distribution Function:
  + In Excel 🡪 Expressed in terms of μ rather than λ; therefore expected value is 1/λ and variance is
  + **=EXPON.DIST(**x,1/μ (or λ), cumulative**):** Probability of arrival before x



* **Other Distributions:** 
  + Triangular (distribution is estimated judgmentally)
  + Lognormal (low probabilities of large values but non zero and non-negative)
    - =LOGNORM.INV(probability,mean,stdev)
  + Beta (modeling variation over a fixed interval from 0 to beta)
    - =BETA.INV(probability,alpha,beta,A,B)
* **Plug in random number for x to provide random outcome of distribution (see next page)**



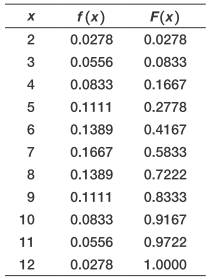
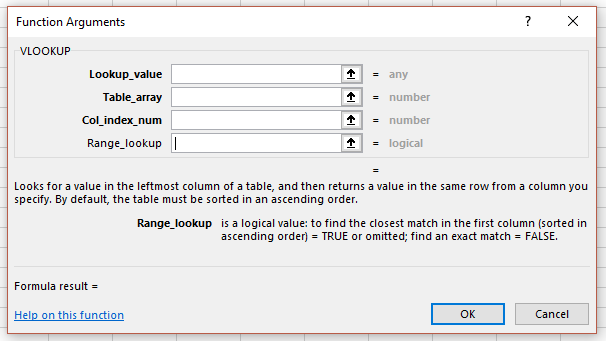
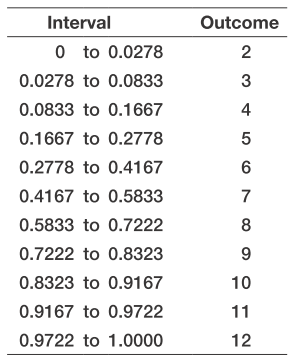
|  |
| --- |
| * **Freezing Random Number Calculations:**    + Paste as Values   + “File” 🡪 “Options” 🡪 “Formulas” 🡪 “Manual” & Use F9 to refresh |

**Random Sampling from Probability Distributions**

Applications in business analytics require random samples from specific probability distributions. Example: in a financial model, interested in distribution of cumulative discounted cash flow over several years – where all factors are uncertain and described by probability distributions. **Outcome variables are functions of random input variables** and the probability distribution of the variables can be accomplished by sampling procedures such as Monte Carlo Simulation. Input variables are generated using **random numbers** or **RAND()**.

**1. Sampling from Discrete Probability Distributions**

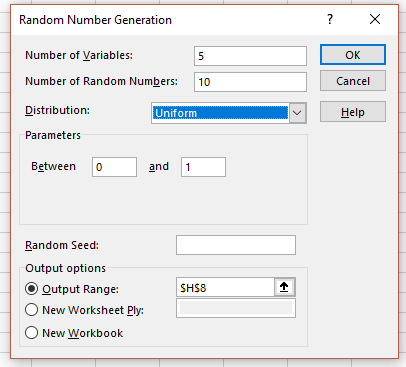
**=VLOOKUP()** - Range goes up to, but does not include, upper limit

**2. Sampling from Common Probability Distributions**

* Transform a random number into a variate from a uniform distribution between A-B:

**3. Sampling from Other Distributions (normal, exponential)**

* Use **Random Number Generation** in the **Analysis Toolpak**
  + Data 🡪 Data Analysis 🡪 Random Number Generation
  + # of Variables: # of columns
  + # of Random Numbers: #s per column
  + Distribution & Parameters
  + Output Range
  + **Random Seed**: value from which a stream of random numbers if generated (keep constant to reproduce “identical” random events across policies or decision variables under the same circumstance)
* **Use “Analytic Solver Platform” (not in this course)**

**Data Modeling and Distribution Fitting**

In applications of BA, we collect sample data (failures, demand, service time) to gain an understanding of the distribution of the variables. We can construct distributions, histograms, descriptive stats; however, these are still samples. Using sample data limits ability to predict uncertainty because of potential values outside sample; a better approach: identify underlying probability distribution from which the sample data come by “fitting” a theoretical distribution to the data and verifying goodness of fit.

* Start with “Eyeball Test”: plot data on scatter plot/histogram & observe shapes of distribution; summary statistics
* **Better approach is** **Goodness of Fit:** Draw conclusion of nature of the distribution using statistical methods (Chi2, Kolmogorov-Smirnof, Anderson-Darling) or software (eg **“Analytic Solver Platform” (not in this course)**)

|  |  |
| --- | --- |
| Focus: | * Understand Monte Carlo Simulation and use in Risk Analysis * Simulate with Excel and MCSim * Understand how to test data to see if it fits known probability distributions * Understand Statistical Issues that need to be addressed when using simulation * Awareness of dynamic simulation |

**Introduction - Simulation**

* **Risk:** likelihood of an undesirable outcome – assessed by evaluating probability that the outcome will occur along with severity of the outcome
* **Types of Simulation**
  + **Monte Carlo:** Process of generating random values for uncertain inputs (based on probability distributions) in a model, compute the out put variables of interest, and repeating process for many trials to understand distribution of the output results. (ie should 2 companies merge?)
  + **Systems (Time/Event Base Monte Carlo):** Models dynamics & behaviour of a system over time, simulated clock, event driven (ie arrival time of customers, failure time of a machine)
* **Simulation Uses**
  + **Explanatory Devices:** Understand system or problem 🡪 gain insights
  + **Analysis Vehicles:** Determine critical elements, assess uncertainty, find good solutions, what-if comparative and sensitivity analysis
  + **Examples:** *Investment choice (riskiest, probability of NPV, probability of minimum yield); Corporate applications (M&A, Chrysler concept to showroom); Vancouver Airport preboard screening*
* **Benefits of Simulation**
  + **Descriptive:** provides the output, up to you to find the answer(not prescriptive – prescribe best answer)
  + **Understand systems** without building them; disturbing them; or destroying them
  + **Model any assumption** and easier to explain then other approaches
* **Disadvantages/Issues with Simulation**
  + **Cost:** Expensive for custom models
  + **Programming Errors –** Verify Model (take historic data and stick it in to check!)
  + **“Mushy” Answers**: no guarantee of optimality
  + **Important (critical) Inputs can be Missed –** Validate Model
  + **Time**
  + **Sampling Error** – We can deal with this
  + **Flaw of Averages:** Only time averages inputs are guaranteed to result in average outputs when spreadsheet model is linear in all uncertain variables 🡪 **Jensen’s Inequality:** Outputs ≠ Inputs ???
* **Simulation Languages**
  + General Purpose: Basic, Fortran, Visual Basic, C, C++, Java
  + Special Purpose: SLAM, GPSS, SIMSCRIPT, GASP, DYNAMO
  + Simulators: Simfactory, XCELL+, Extend, Micro Saint, Arena
  + Excel Add-ins/**Replication Software**: Crystal Ball, @Risk, MCSim

**Random Numbers & Variables drive Simulations**

* **Random Number Characteristics -** Excel: =RAND()
  + **Uniformly Distributed** (btwn 0.000 and 0.999)
    - Can use normal, beta, binomial, chisq, f, t, lognorm distributions (see above)
  + Pseudo random vs truly random
* **Random Variates: Outcomes** generated from a probability distribution
  + Correspondence between RV and actual outcome is key to valid & credible models
  + RV are assigned proportional relative frequency we want to generate in simulation – key is to build a cumulative distribution
* **Defining Uncertain Model Inputs:** Inputs need to be characterized by a probability distribution; can be empirical/discrete or common probability distribution mentioned above (typically uniform or triangular distributions are used in absence of data)
  + Use Excel Functions
  + Use Data Analysis – Random Number Generator
* **Simulating with Excel:** 
  + Formulate problem & build spreadsheet model
  + Specify probabilistic assumptions; assign probability distribution to input variables
  + Implement model: sample input variables from probability distribution; compute output variables and record result; repeat until sufficient # of trials generate useful distribution of outputs (ie MCSim)

|  |
| --- |
| * **Using MCSim**: Excel Add-in – Makes Monte Carlo simulation easy – **Refer to MCSim.doc**  1. Develop spreadsheet model 2. Determine probability distributions and define cells which represent random information 3. Define “output” cells to be predicted 4. Set number of replications 5. Run simulation 6. Interpret results |

* **Criteria for Decision Making: How do we select between various options based on simulation results? - ???**
  + Expected Value (largest mean)
  + Optimistic (best scenario max/min)
  + Pessimistic (worst scenario max/min)
  + Minimum Payoff (select option with greatest probability max p(x>xo) or p(x<xo))
  + Maximum Variation
  + Management Preferences etc…
* **Statistical Issues**
  + Simulation experience is a **sample** from an unknown population
  + Purpose to **estimate population parameters** (mean, variance…) & confidence internals and gain knowledge about distribution of outcome variables **to assess risk**
    - * α is CI, x is mean, z is z-value at α/2 (=NORM.S.INV(1-α/2)), s is STDEV, n is trials
  + Initially need to determine input probability distributions
* **Inputs: Historical Data** (see “Data Modeling and Distribution Fitting” above)
  + **Assumption:** Past represents future; requires large amounts of data
  + **Approaches:**
    - **Eyeball Test:** Plot & examine histogram – does it look like a standard distribution?
    - **Goodness of Fit** (Computer automates this)
* **Model Verification and Validation:** Building an **accurate** model and **convince** end users (GIGO)
  + **Verification:** Does model perform as intended? Try it on something!!!!
    - Performed by expert review & computer code
    - Test with dummy data (extremes and averages)
  + **Validation**: Does conceptual model accurately depict real system?
    - Conduct a “walk thru”; if possible run simulation with past data & compare to historic results
    - 3 Step Process: (1) develop model with face validate (2) validate model assumptions (3) validate model output
* **Experimental Design:** to be addressed before collecting and analyzing output – usually interested in steady state (long run) operation of system
  + Input data (use same **“seed” data** for policy/alternatives under consideration)
  + Length of time of simulation
  + Treatment of initial data outputs (remove transient or startup conditions?)
  + Sample results or analyze all?
  + Performance measures and output statistics to collect
* **# Of Trials/Replications:** Different random numbers yield different results
  + # depends on system & objectives 🡪 longer the better; but could be a time issue or resource issue
  + Higher # provides more accurate estimate of population
  + When comparing alternatives – use the same set of random numbers and # of trials
* **Simulation Types**
  + **Static Simulation:** Independent of time
  + **Event-Based MonteCarlo Simulation:** Time dependent, simulation clock
    - **Time Driven:** move clock forward a fixed amount (**arrival times**)
    - **Event Driven:** move clock forward to time next event occurs (usual approach) (**failure times**)

**Questions:**

* **Do we need any control for random numbers with negative values? Use If statements, control it, use a different distribution**

**Queuing Theory**

* **Littles Law: I = F x T**
* Queues are a mismatch between supply and demand
* Queuing models and simulations are heavily tied together 🡪 critical in answering how to improve service
  + Sample Responses: increase servers, more lines, pre-process customers, additional training, limit services at peak hours etc
* Psychology of waiting: unoccupied time feels longer than occupied; pre-process wait feels longer than in-process wait; anxiety makes wait seem longer; uncertain waits are longer than known; unexplained waits are longer than explained; unfair waits are longer than equitable waits; more valuable to service, the longer the wait; solo waiting feels longer than group waiting