Aachen Summer Simulation Seminar 2014

Lecture 07

Input Modelling + Experimentation + Output Analysis

Peer-Olaf Siebers

pos@cs.nott.ac.uk

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

Motivation

1. Input modelling
   – Improve the understanding about how data is collected
   – Understand how random sampling works
   – Learn more about when to use specific types of distributions

2. Experimentation
   – Understanding the need for obtaining accurate results
   – Looking at methods for supporting obtaining accurate results

3. Output analysis
   – Provide statistical tools necessary to conduct simple output analysis

Input Modelling

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

Introductory Remark

- In this lecture we will focus on DES Input Modelling
  - Robinson (2004) Chapter 2 and 7
- More about SD Input Modelling
  - Luna-Reyes and Andersen (2003)
- More about ABS Input Modelling
  - Janssen and Ostrom (2006)

Data

- Obtaining data
  - Data needs to be ***sufficiently accurate*** and ***in the right format*** for the simulation model (more later)

  - Three categories of data:
    - Category A: ***Available***
      - Data is known or has been collected before
    - Category B: ***Not available*** but ***collectable***
      - Putting a data collection exercise in place by either getting people or electronic systems to monitor the operations
    - Category C: ***Not available*** and ***not collectable***
      - Often occurs because the real world system does not yet exist or due to time limitation to collect meaningful data

Data

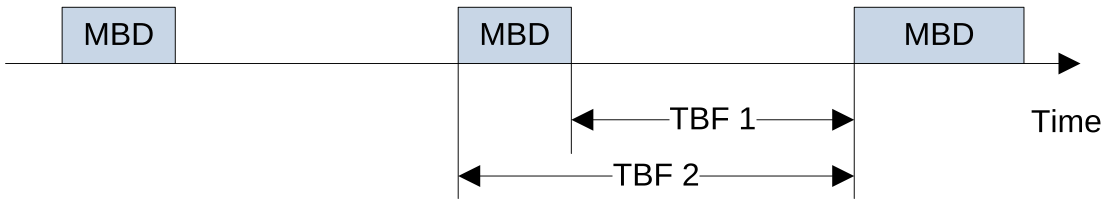
- Dealing with unobtainable Category C data
  - Estimate data
    - Use data from similar system for your estimates; using standard times; discuss with stakeholders; intelligent guess
  - Treat data as experimental factors
    - Instead of asking what the data are it is asked what do the data need to be (can only be applied when there is some control over the data in question)
  - Revise the conceptual model
  - Change the modelling objectives
  - Abandon the simulation study

Data

- Dealing with unobtainable Category C data (cont.)
  - Validity and credibility issues
    - Estimates need to be clearly identified in an assumption list
    - Sensitivity analysis should be performed on these data
    - Data might become available as project progresses

Data

- Data format
  - Information is often not in the right format for the simulation (e.g. time study data are aggregated to determine standard times for activities; in simulation the individual elements (e.g. breaks and process inefficiencies) are modelled separately
  - Important to know how the input data is interpreted by the simulation software

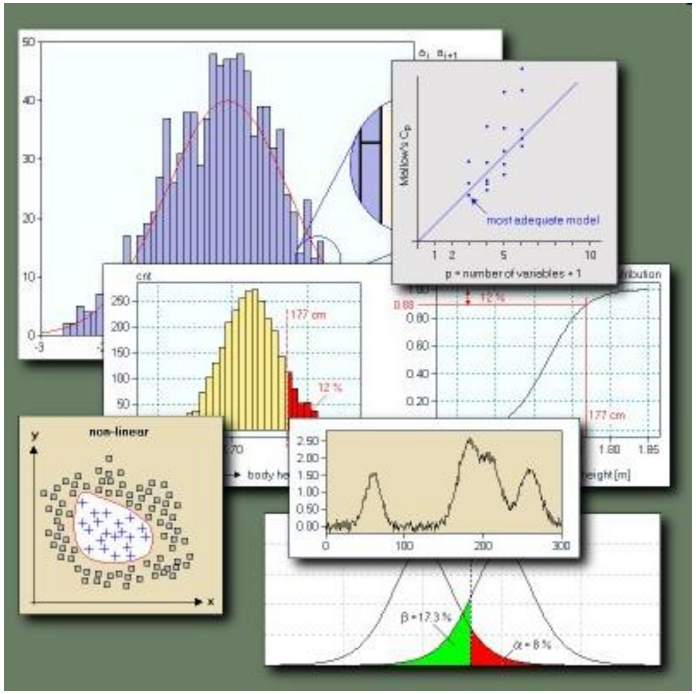


MBD = machine breakdown
TBF = time between failures

Representing Unpredictable Variability

- Modeller must determine how to present the variability that is present in each part of the model

- Three options
  - Traces
  - Empirical distributions
  - Statistical distributions



Representing Unpredictable Variability

- Traces
  - Streams of data that describe a sequence of events
    - Data about the time the events occur (e.g. call arrival times)
    - Additional data about the event (e.g. call type)
  - Trace is read by the simulation as it runs and events are recreated in the model as described by the trace
  - Automatic monitoring systems are a common source of trace data

Call arrival time (minutes)	Call type
0.07	1
0.44	3
0.97	1
1.01	1
1.20	2
...	...

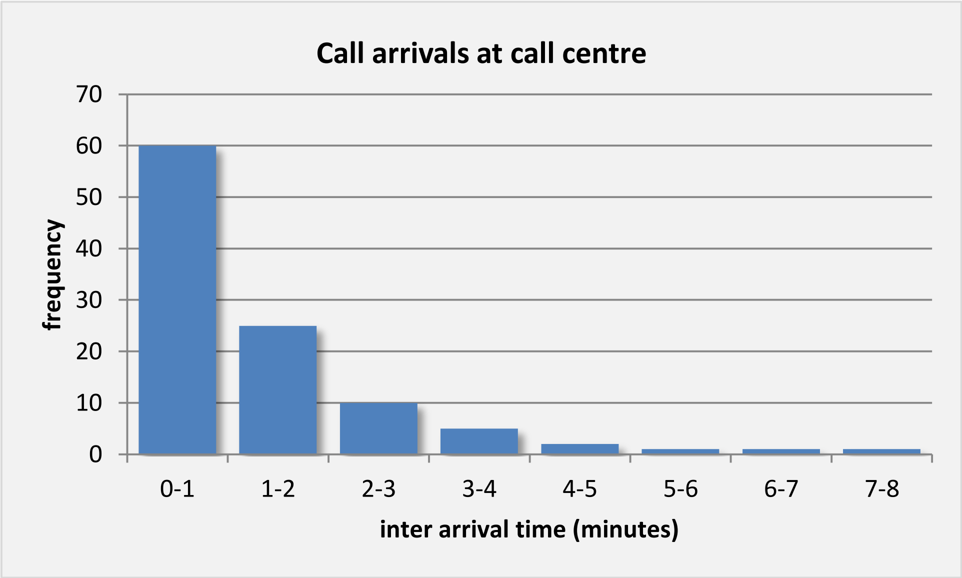
Representing Unpredictable Variability

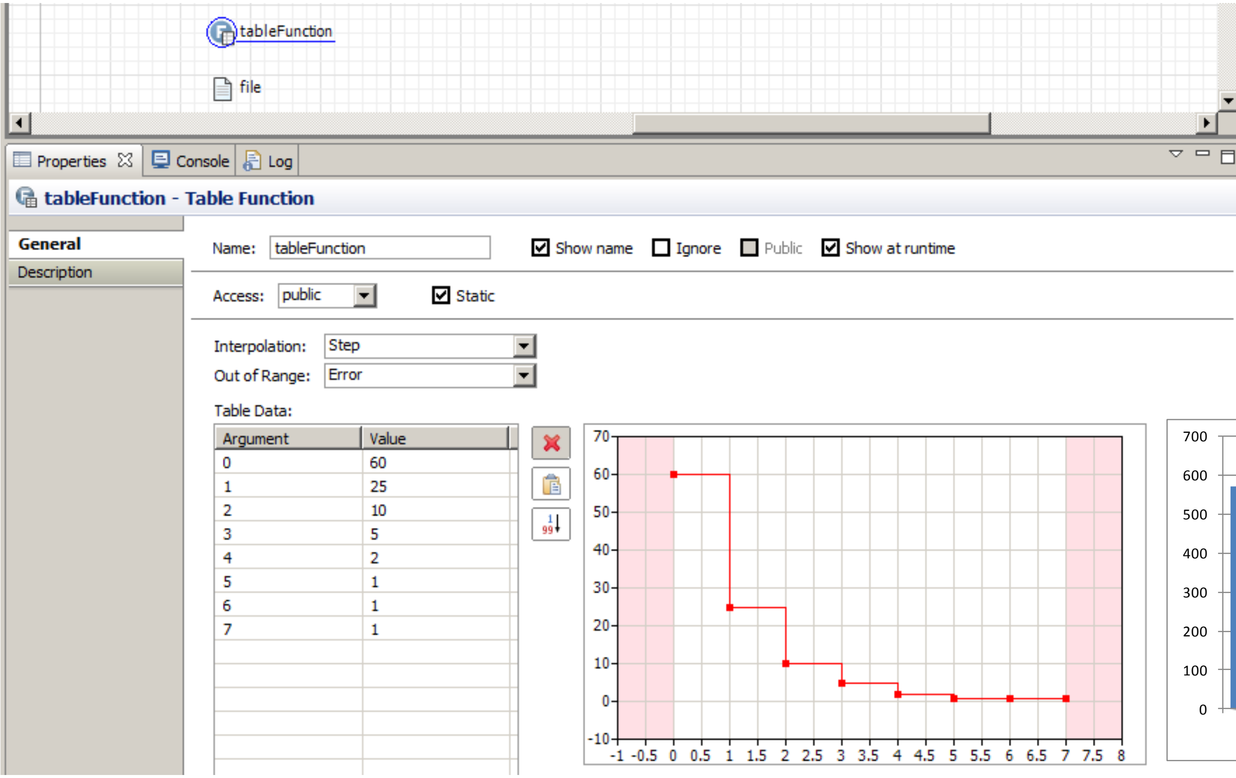
- Traces
  - Advantages?
    - Most natural approach
    - Useful for validating a model against an existing system
    - Improve credibility of the model (client can see patterns they recognise)
  - Problems?
    - Real system needs to exist
    - Data set will be quickly exhausted in long simulations or those that are replicated many times
    - Tends to be slow as it involves intensive input operations (if the simulation model involves many sources of unpredictability)
    - Simulation model results are limited to those induced by this particular data set, and are thus not as generalisable as they should be

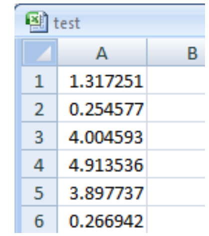
Representing Unpredictable Variability

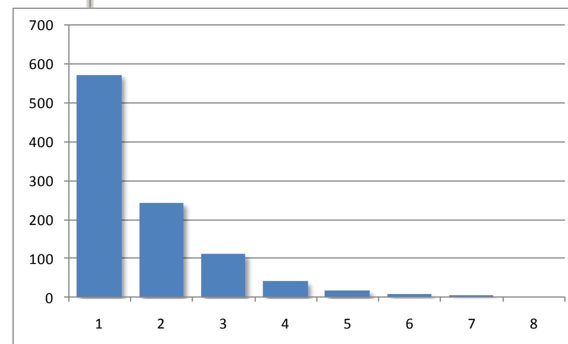
- Empirical distributions
  - Show the frequency with which data values or ranges of values occur
  - Based on historical data, often formed by summarising trace data
  - As simulation runs values are sampled from the distribution
  - Most simulation software allows the user to enter empirical distribution data directly

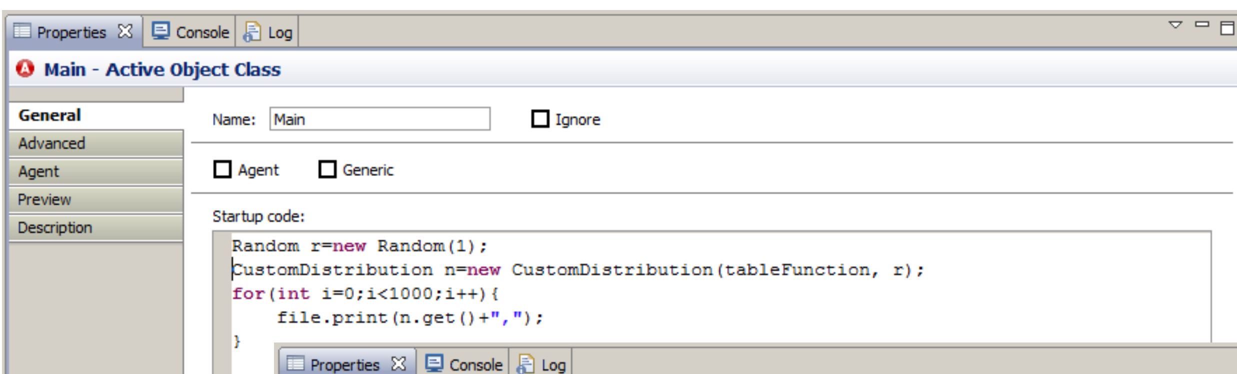
arrival time	inter-arrival time
0	-
1.5	1.5
2	0.5
4.1	2.1
4.2	0.1
7.5	3.3
7.8	0.3
8.2	0.4

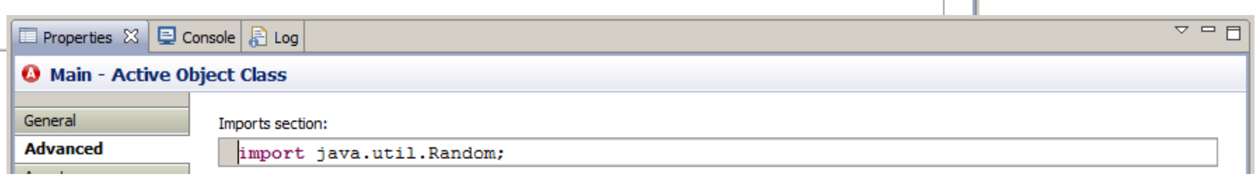



tableFunction
file
Properties
Console
Log
tableFunction - Table Function
General
Description
Name: tableFunction
Show name
Ignore
Public
Show at runtime
Access: public
Static
Interpolation: Step
Out of Range: Error
Table Data:
Argument | Value
0 | 60
1 | 25
2 | 10
3 | 5
4 | 2
5 | 1
6 | 1
7 | 1
70 60 50 40 30 20 10 0 -10
-1 -0.5 0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5 5.5 6 6.5 7 7.5 8


test
A
B
1 1.317251
2 0.254577
3 4.004593
4 4.913536
5 3.897737
6 0.266942


700
600
500
400
300
200
100
0
1 2 3 4 5 6 7 8


Properties
Console
Log
Main - Active Object Class
General
Advanced
Agent
Preview
Description
Name: Main
Ignore
Agent
Generic
Startup code:
Random r=new Random(1);
CustomDistribution n=new CustomDistribution(tableFunction, r);
for(int i=0;i<1000;i++){
file.print(n.get()+",");
}


Properties
Console
Log
Main - Active Object Class
General
Advanced
Imports section:
import java.util.Random;

Representing Unpredictable Variability

- Empirical distributions
  - Advantage?
    - There is almost no issue of poor fit, as there is little or no fitting going on
    - Generating draws from such empirical distributions is easy and fast, using linear interpolation
    - It is possible to change the stream of random numbers used for sampling and alter the pattern of variability (important when performing multiple replications)
  - Problems?
    - Range of variability is still restricted to that observed in historic events
    - With a small sample size this method tends to be high-variance, not filling in the 'holes' well at all

Representing Unpredictable Variability

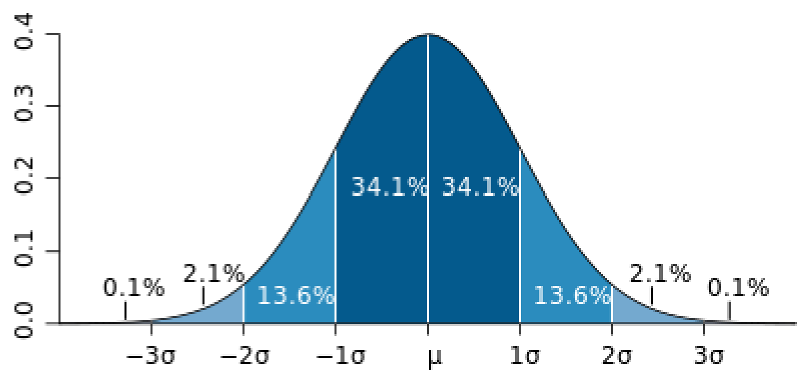
- Statistical distributions
  - Defined by some mathematical function or probability density function
  - Three different types of statistical distributions
    - ***Continuous distributions:*** For sampling data that can take any value across a range
    - ***Discrete distributions:*** For sampling data that can take only specific data across a range, for instance only integer or non-numeric values
    - ***Approximate distributions:*** Used in the absence of data
  - Process for sampling distributions is explained later

Representing Unpredictable Variability

- Statistic distributions commonly used in simulation models
  - *Continuous* distributions
    - Normal distribution
    - (Negative) exponential distribution
    - Erlang distribution
  - *Discrete* distributions
    - Binomial distribution
    - Poisson distribution
  - *Approximate* distributions
    - Discrete and continuous uniform distribution
    - Triangular distribution
  - For more details see Law and Kelton (2006) or Wikipedia

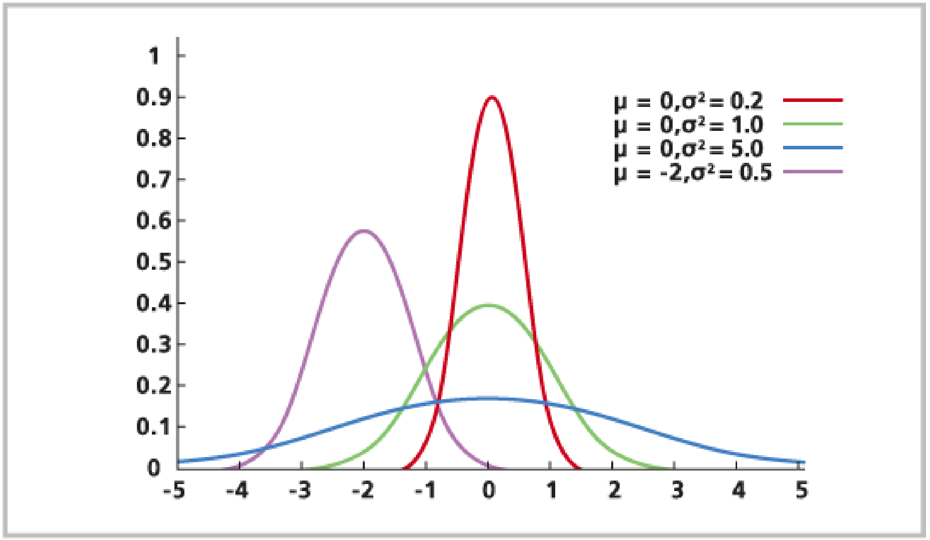
Representing Unpredictable Variability

- Continuous: Normal distribution (Gaussian distribution)
  - x axis: Values of the dependent variable (e.g. height)
  - y axis: Probability that the values of the dependent variable will occur
  - 68% of values less than one standard deviation away from the mean
  - 95% of values less than two standard deviations away from the mean
  - 99.7% of values less than three standard deviations away from the mean



Representing Unpredictable Variability

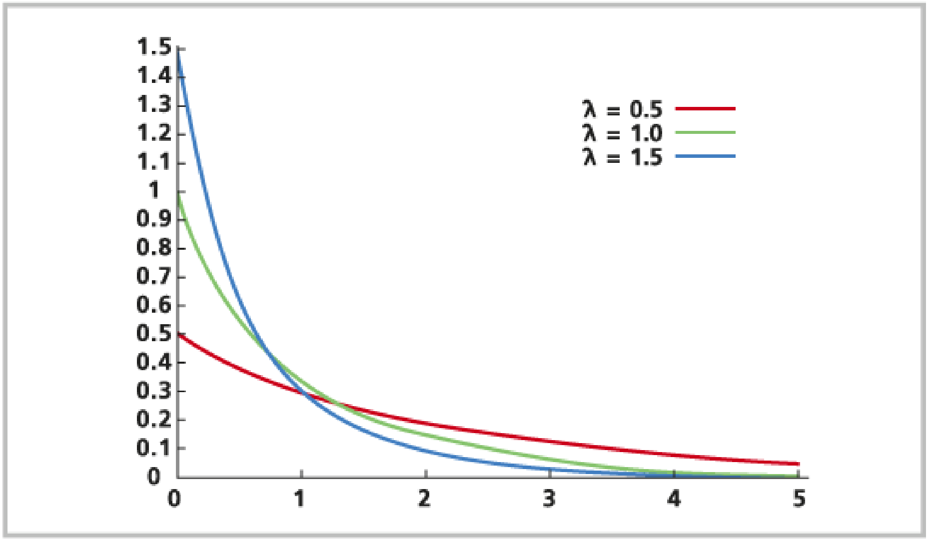
- Continuous: Normal distribution (Gaussian distribution)
  - Parameters: Mean (its location) and standard deviation (its spread)
  - For a given range of values of x, the area under the curve gives the probability of obtaining that range of values
  - Used for sampling errors; sometimes truncated normal distribution is used (e.g. when sampling for times) to avoid negative values



The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

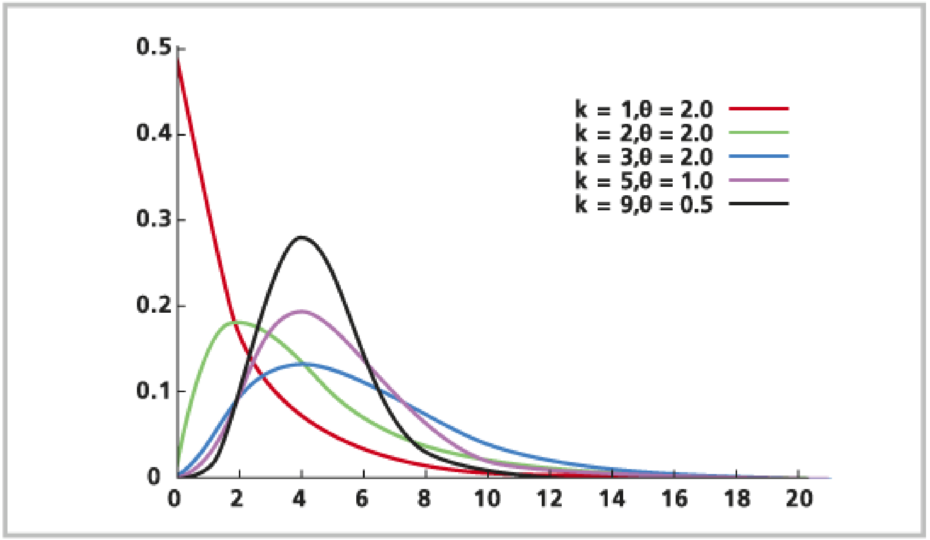
Representing Unpredictable Variability

- Continuous: (Negative) exponential distribution
  - Parameter: Mean ($1/\lambda$ = inter arrival time, λ=rate parameter)
  - Gives a high probability of sampling values close to zero and a low probability of sampling higher values
  - Used for sampling times between events or times to complete a task (problem: high probability of near zero values)



Representing Unpredictable Variability

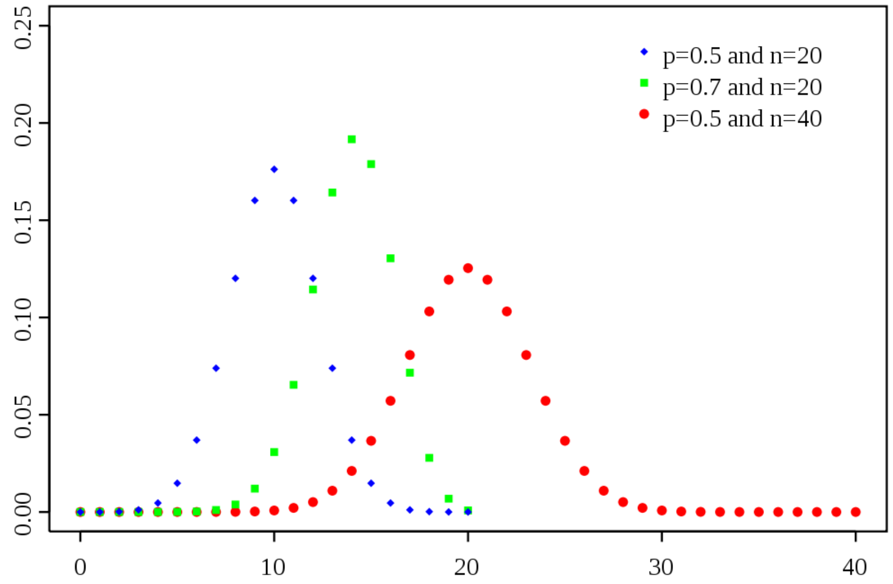
- Continuous: Erlang distribution
  - Parameters: Mean and k (determines the skew of the distribution)
  - If k equals 1 the Erlang distribution is the same as a negative exponential distribution with an equivalent mean
  - As the value of k increases, the skew reduces and the hump of the distribution moves towards the centre
  - Used for sampling the time to complete a task and inter arrival times



The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

Representing Unpredictable Variability

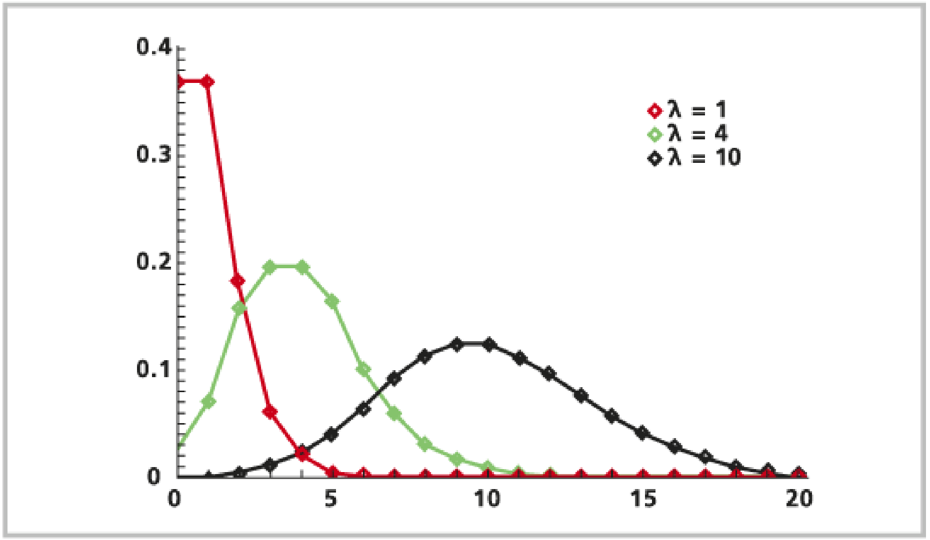
- Discrete: Binomial distribution
  - Parameters: Probability of success (p) and number of trials (n)
  - Describes the number of successes in a sequence of n independent yes/no experiments, each of which yields success with probability p



The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

Representing Unpredictable Variability

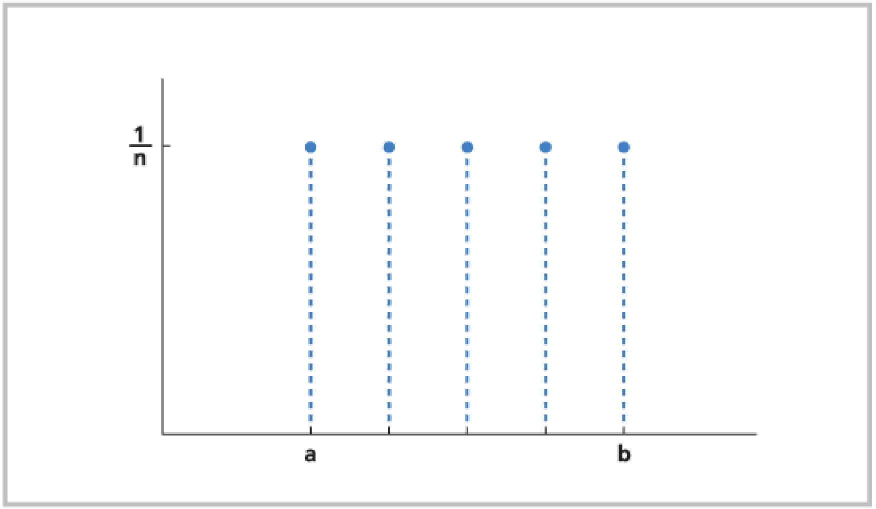
- Discrete: Poisson distribution
  - Parameter: Mean (λ = rate parameter; positive number representing the expected number of occurrences within a specified interval, e.g. if the event occurs every 10 minutes, in an hour λ would be 6)
  - Represents number of events that occur in an interval of time
  - Used for sampling arrival rates, number of items in a batch of random size (e.g. Number of boxes on a pallet)

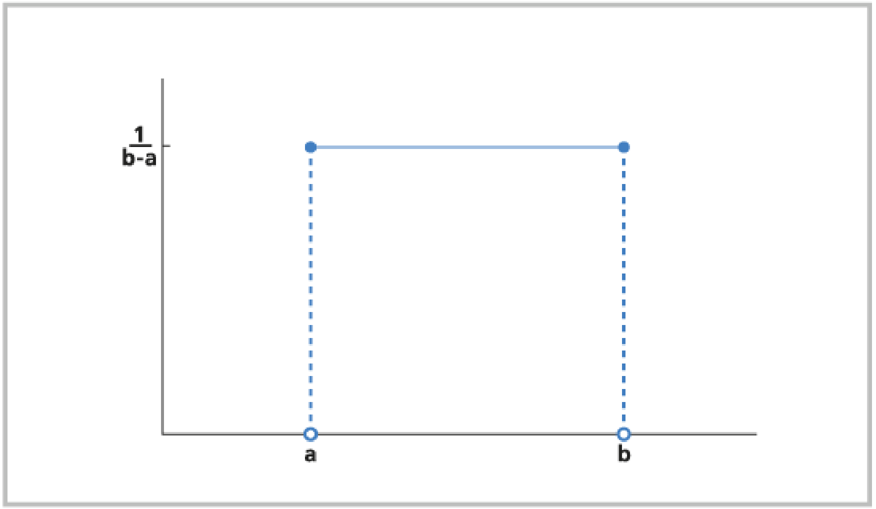


The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

Representing Unpredictable Variability

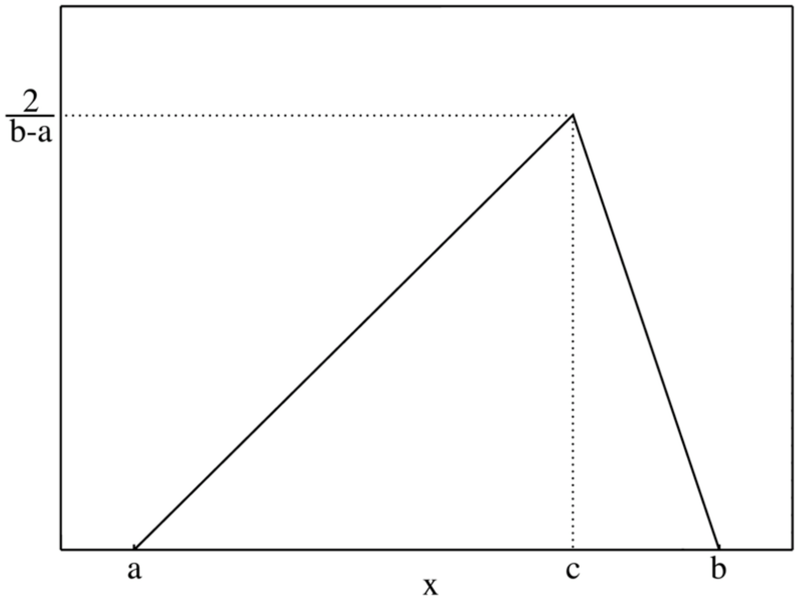
- Approximate: Uniform distribution
  - Parameter: Minimum and maximum
  - Provide a useful approximation in the absence of data
  - Frequently used when dealing with category C data
  - Discrete and continuous uniform distribution





Representing Unpredictable Variability

- Approximate: Triangular distribution
  - Parameters: Minimum, maximum, mode (value that occurs the most frequently in a data set)
  - Mean and mode can be quite different; this needs to be considered when collecting data
  - Used as an approximation for task times and inter arrival times



The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

Representing Unpredictable Variability

- Statistical distributions
  - Advantages?
    - Range of variability is not restricted to that encountered in the past
    - Limited use of computer memory
    - Not necessary to have data from the real system as parameters can be estimated; however, data from the real system is always preferred
  - Problems?
    - Many statistical distributions have long tails and therefore there might be occasions when extreme values might be sampled
    - Least transparent approach for the client
    - Possibility of direct comparison to historic results not available
    - Wanting to find a fit to our observed data from among a relatively short list of probability distributions that were developed decades ago largely for their analytic tractability.

Representing Unpredictable Variability

- Fitting empirical data to statistical distributions
  - Three step process
    1. Select a statistical distribution
    2. Determine the parameters
    3. Test of goodness of fit
  - Iterative process
    - It is not enough to select a single distribution or a single set of parameters; a series of distributions should be tried with different sets of parameter values

Representing Unpredictable Variability

- Fitting empirical data to statistical distributions
  - Select a statistical distribution
    - Inspect the data
    - Select distribution based on known properties of the process
  - Determine parameters
    - Means can be calculated from histogram data
    - Some parameters cannot be estimated: use trial and error
  - The goodness-of-fit test
    - Graphical test
      - Comparing the histogram of the empirical data with that of the proposed distribution
    - Statistical test (see Robinson, 2004, pp114-119)
  - Software: e.g. ExpertFit and StatFit

Random Sampling

- Sampling
  - Example: Booking clerk with one arrival process
    - 60% of customers: Personal enquirers (= type X)
    - 40% of customers: Phone callers (= type Y)
    - Top hat method:
      - 100 pieces of paper, 60 with X and 40 with Y
      - Every time a customer arrives draw one piece
      - Important to replace the paper to keep the ratio (60:40)
  - In computer simulation a similar principle is adopted based on random numbers

Random Sampling

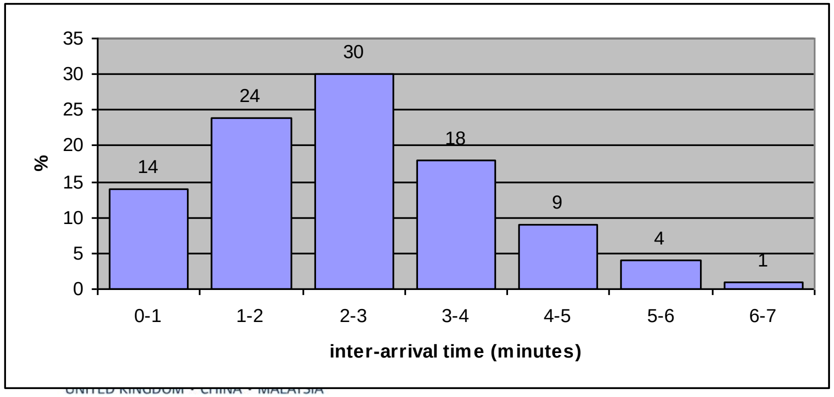
- Random numbers:
  - Sequence of numbers that appear in a random order
  - Presented as integer (e.g. [0-99]) or as real numbers (e.g. [0-1[)
  - Top head method (replacement method)
  - Properties
    - **Uniform:** same probability of any number occurring at any point in the sequence
    - **Independent:** once a number has been chosen this does not effect the probability of it being chosen again or of another number being chosen

Random Sampling

- Relating random numbers to variability in a simulation
  - Modelling proportions
    - Example: Booking clerk with one arrival process
      - For small sample sizes ratio might not be achieved (as the process is random)
      - For large sample sizes ratio will be achieved more or less

Random numbers	Customer type
00-59	X
60-99	Y

  - Modelling variability in times
    - To model continuous real variables (e.g. activity times) one could determine the range and then draw a second random number, divide it by 100 and add it to the range

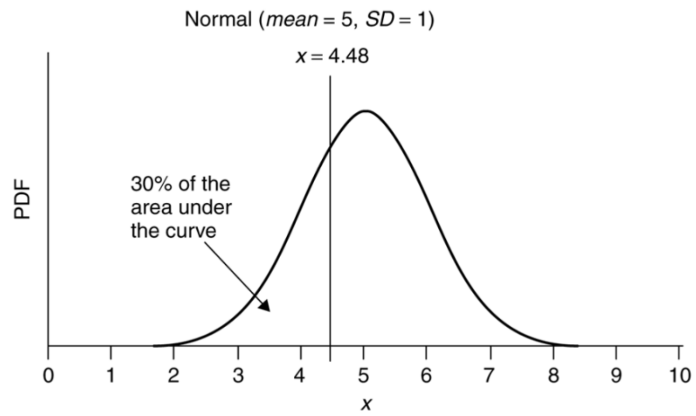


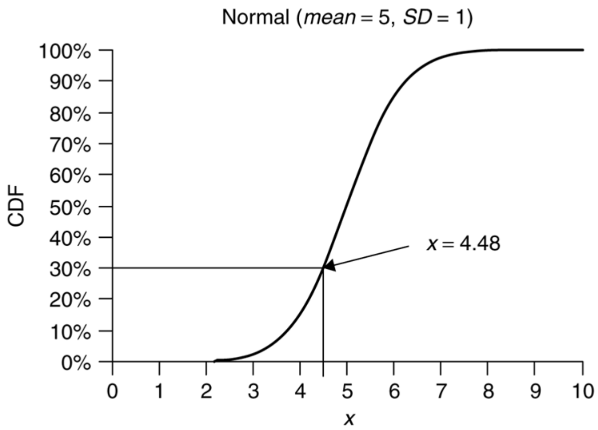
Random numbers	inter arrival time
00-13	0-1
14-37	1-2
38-67	2-3
68-85	3-4
86-94	4-5
95-98	5-6
99	6-7

Customer	rnd 1	range	rnd 2	inter arrival time
1	41	2-3	27	2.27
2	44	2-3	24	2.24
3	90	4-5	5	4.05

Random Sampling

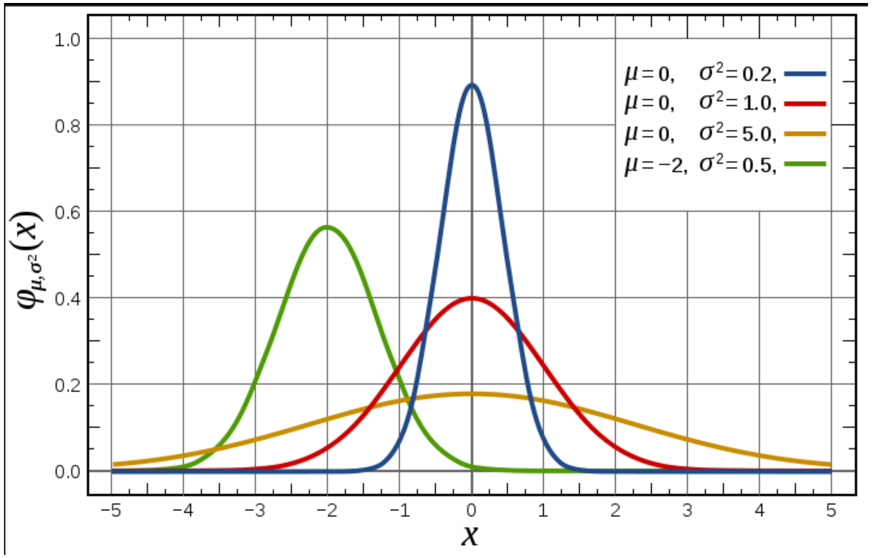
- Random sampling (generating variates) from standard statistical distributions
  - To sample a value from the distribution the random number is taken to be the percentage of the area under the curve
    - Difficult to think in terms of identifying area under the curve
    - Instead of the probability density function (pdf) we use the cumulative distribution function (cdf)

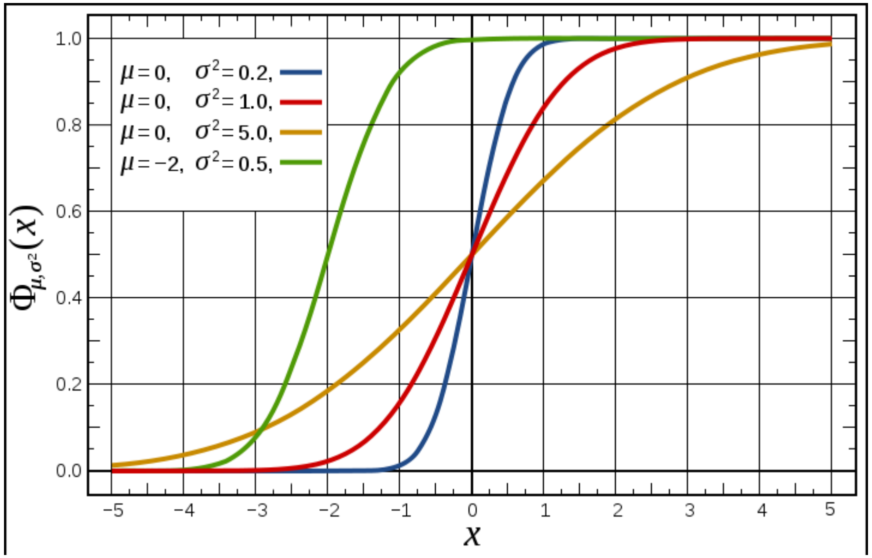




Random Sampling

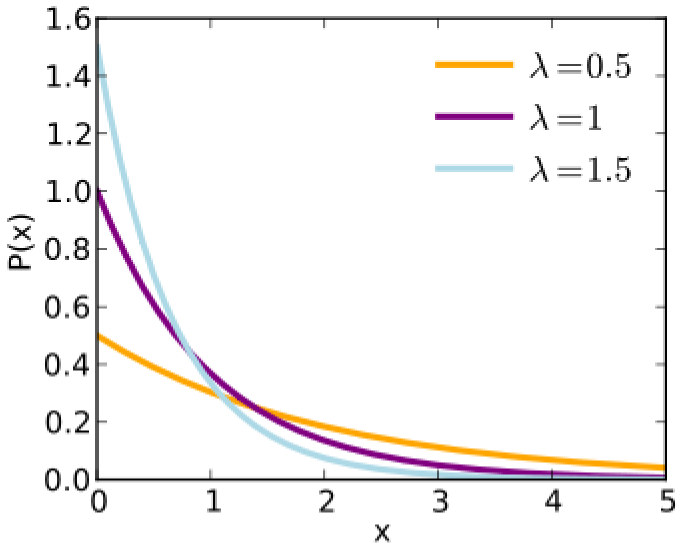
- Normal distribution (pdf and cdf)

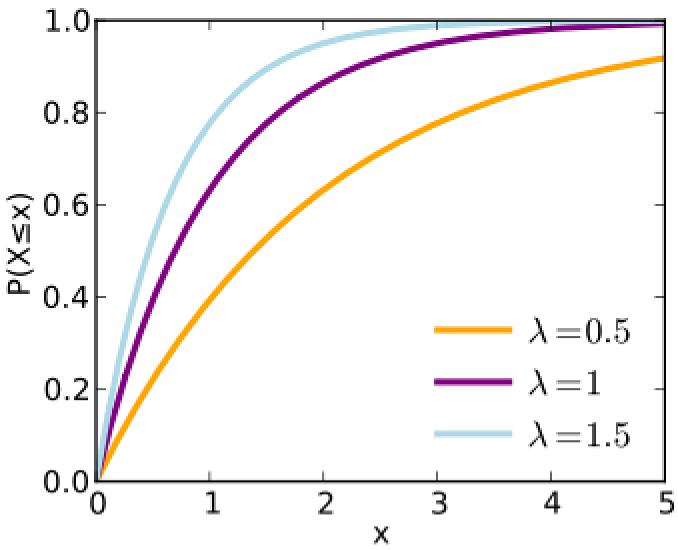




Random Sampling

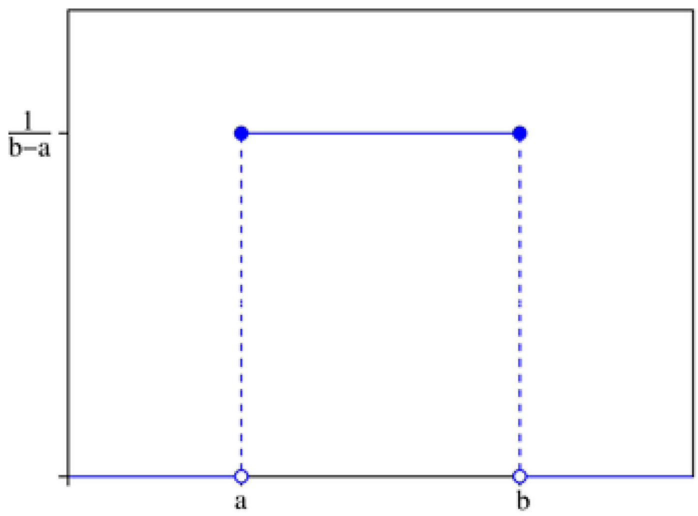
- Negative exponential distribution (pdf and cdf)

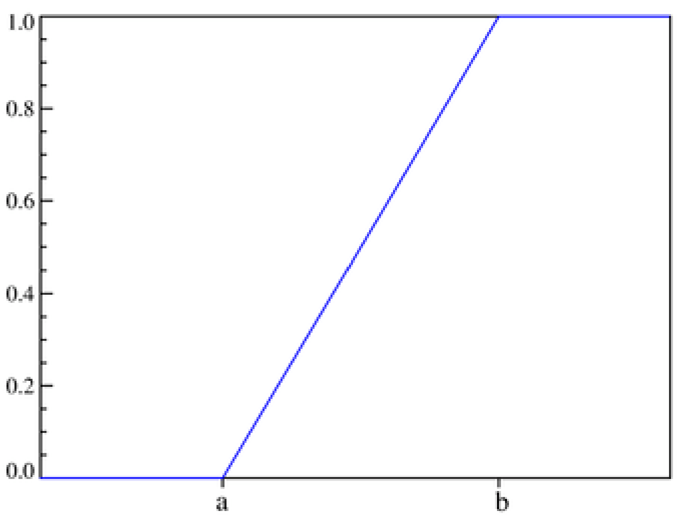




Random Sampling

- Continuous uniform distribution (pdf and cdf)





Random Sampling

- Computer generated random numbers
  - By nature computers do not behave in a random fashion
  - There are algorithms that give the appearance of producing random numbers although the results are completely predictable
    - Numbers produced by these algorithms (called pseudo random numbers) have the properties of uniformity and independence
  - Commonly used algorithm for generating random numbers:
    - $X_{i+1} = aX_i + c \pmod{m}$
      - X_i: Stream of random numbers (integer) on the interval (0, m-1)
      - a: Multiplier constant
      - c: Additive constant
      - m: modulus
      - X_0 = starting value for X = seed

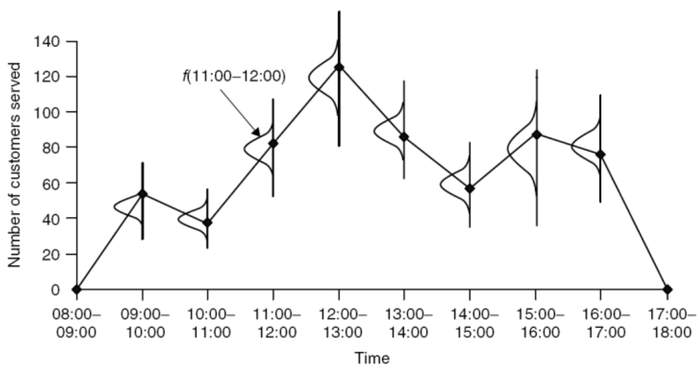
Experimentation

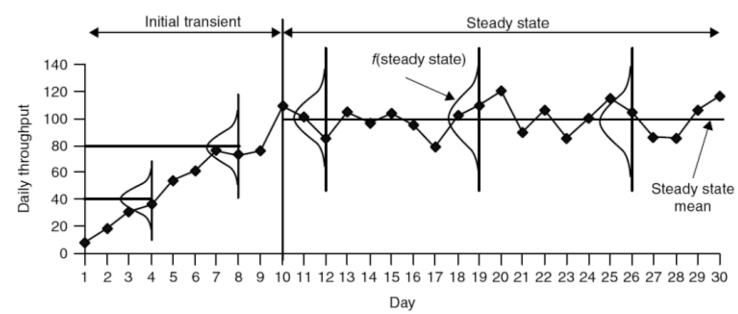
Experimentation Preparation

- What is experimental preparation about?
  - Obtaining accurate results (for stochastic simulation)
    - Obtaining accurate data on the performance of the model (estimate of average performance and its variability).
    - This says nothing about how accurately the model predicts the performance of the real system
    - To answer the latter you would use black box validation

Types of Systems

- Types of Systems
  - **Terminating** (natural end point that determines the length of a run) vs. **non terminating** (there is no specific reason why the simulation experiment should terminate)
  - **Transient** (the distribution of the output is constantly changing; true for most terminating simulations) vs. **steady state** (the output is varying to some fixed (steady-state) distribution; true for most non-terminating simulations)





Robinson (2004)

Experimentation Preparation

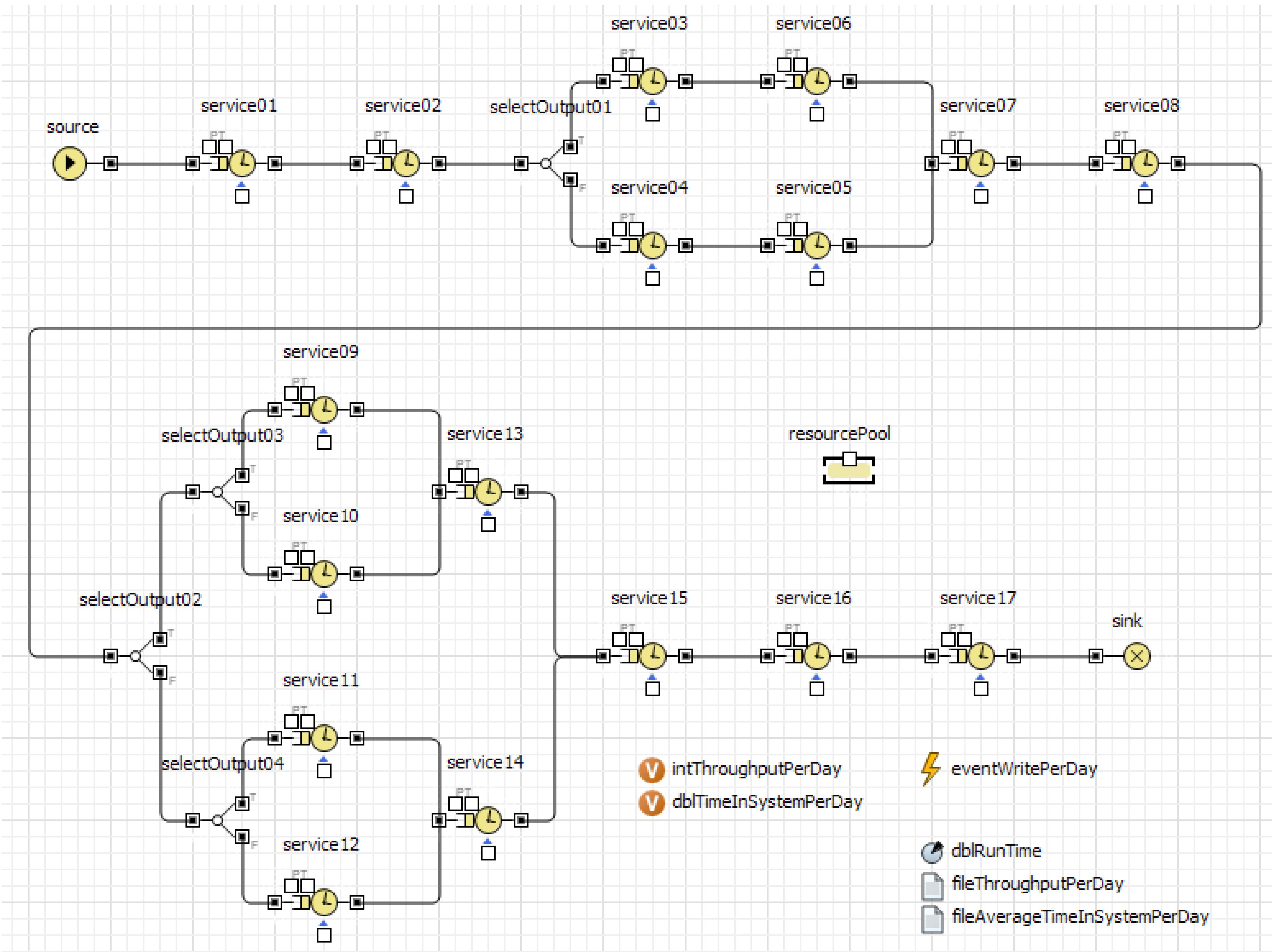
1. Dealing with initialisation bias
   - Solutions:
     a) Run model for a warm-up period: Running the model until it reaches a realistic condition and only collect results from the model after this point
     b) Set initial conditions in the model: Place the model in realistic conditions at the start of the simulation run [not practical]

2. Obtaining sufficient output data
   - Solutions:
     a) Multiple replications (terminating or non-terminating simulations): Equivalent to taking multiple samples in statistics; multiple runs of the simulation model with different random number streams
     b) Single long run (non-terminating simulations): Equivalent to taking one large sample in statistics [not practical; some statistical concerns]

Dealing with Initialisation Bias

1a. Running the model for a warm-up period

- Needs to be long enough to ensure that the model is in a realistic condition (the difficulty lies in determining whether the model is in a realistic condition)
- Method categories for determining the warm-up period length (Robinson, 2002)
  - Graphical methods; heuristic approaches; statistical methods; initialisation bias tests; hybrid methods
- Most commonly used methods for estimating the warm-up period
  - **Time series inspection**
  - **Welch's method**



source
service01
service02
selectOutput01
service03
service06
service04
service05
service07
service08
service09
selectOutput03
service13
resourcePool
service10
selectOutput02
service15
service16
service17
sink
service11
selectOutput04
service14
intThroughputPerDay
dblTimeInSystemPerDay
eventWritePerDay
service12
dblRunTime
fileThroughputPerDay
fileAverageTimeInSystemPerDay

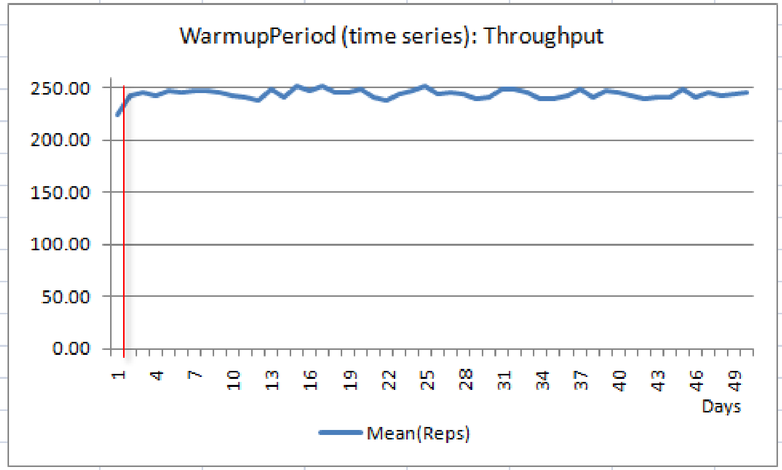
ThroughputPerDay														
Day	Rep 1	Rep 2	Rep 3	Rep 4	Rep 5	Rep 6	Rep 7	Rep 8	Rep 9	Rep 10	Rep 11	Rep 20	Mean(Reps)	StDev(Reps)
1	254	243	228	225	214	220	226	205	193	244	231	250	223.65	19.62
2	261	228	255	262	260	237	232	237	230	257	247	231	242.35	14.06
3	230	263	240	252	239	247	246	241	223	208	251	246	245.45	20.44
4	241	235	244	242	267	239	246	242	269	246	248	217	243.00	16.09
5	252	261	235	231	253	245	271	238	245	244	236	251	247.30	11.65
6	220	218	244	266	244	276	246	255	230	258	237	244	246.30	17.99
7	251	263	255	238	234	237	260	274	235	250	255	221	247.05	12.75
8	257	241	256	236	281	257	238	241	249	241	242	258	247.05	12.10
9	250	256	246	257	248	223	228	232	264	254	264	242	245.55	13.16
10	255	242	250	260	218	245	239	222	269	250	237	233	242.60	14.47
11	207	266	231	247	254	241	246	250	232	243	249	262	240.90	17.23
12	265	215	239	235	282	254	256	225	222	247	229	245	238.70	20.81
13	235	262	236	251	253	253	262	231	269	269	223	225	249.30	14.26
100	234	250	234	240	233	236	257	235	225	286	236			
Mean(Days)	246.02	243.51	244.79	243.99	245.65	243.66	243.52	241.05	242.98	245.94	245.24			
StDev(Days)	16.03	17.59	13.00	14.82	14.91	15.12	16.76	14.48	15.49	15.70	14.63			

AverageTimeInSystemPerDay														
Day	Rep 1	Rep 2	Rep 3	Rep 4	Rep 5	Rep 6	Rep 7	Rep 8	Rep 9	Rep 10	Rep 11	Rep 20	Mean(Reps)	StDev(Reps)
1	134.73	94.10	94.13	96.27	92.23	93.07	91.30	85.83	94.98	115.86	92.93	138.76	98.05	14.68
2	98.81	90.92	100.21	105.72	115.85	94.37	107.07	89.87	100.44	89.84	96.70	99.22	98.23	7.90
3	87.67	112.62	92.11	115.71	92.60	96.68	88.49	91.93	90.69	86.49	92.98	92.96	103.90	29.50
4	92.21	88.46	105.28	91.99	108.48	90.01	97.76	93.40	113.16	98.45	103.44	98.60	99.27	9.30
5	100.70	107.43	95.88	88.65	96.72	130.39	139.82	90.52	92.04	91.25	88.74	99.30	101.78	15.35
6	89.07	92.50	102.44	101.06	95.62	137.93	90.06	126.58	87.30	94.91	95.71	91.60	101.10	13.59
7	99.45	96.56	101.53	88.83	91.92	88.74	108.22	122.74	94.15	120.29	110.42	87.29	100.34	12.67
8	98.88	102.23	94.27	89.76	123.28	97.78	96.01	95.10	94.56	102.31	91.87	99.32	97.13	8.10
9	95.21	94.99	113.02	96.21	95.53	85.89	96.73	89.09	155.85	123.31	99.98	90.66	99.70	15.96
10	93.84	90.74	95.84	103.43	88.66	102.15	91.69	87.60	121.74	134.60	94.71	90.39	97.19	11.97
11	86.88	111.63	91.06	95.33	120.87	102.23	98.13	102.78	91.09	103.32	91.93	128.31	97.58	11.44
12	94.73	89.80	102.91	93.86	129.82	97.87	119.70	89.07	98.30	103.04	94.18	101.25	98.46	10.53
13	90.22	100.24	96.45	103.32	96.72	98.48	109.35	87.10	103.86	126.29	93.32	93.48	100.10	9.82
100	92.95	98.20	94.62	91.02	88.80	97.61	97.57	88.19	86.64	207.19	90.65			
Mean(Days)	100.66	97.68	99.38	98.89	99.41	97.81	99.88	95.88	98.50	101.50	99.23			
StDev(Days)	13.49	10.43	11.29	10.86	11.59	11.39	16.19	8.60	11.81	17.30	11.80			

WarmupPeriod

ThroughputPerDay

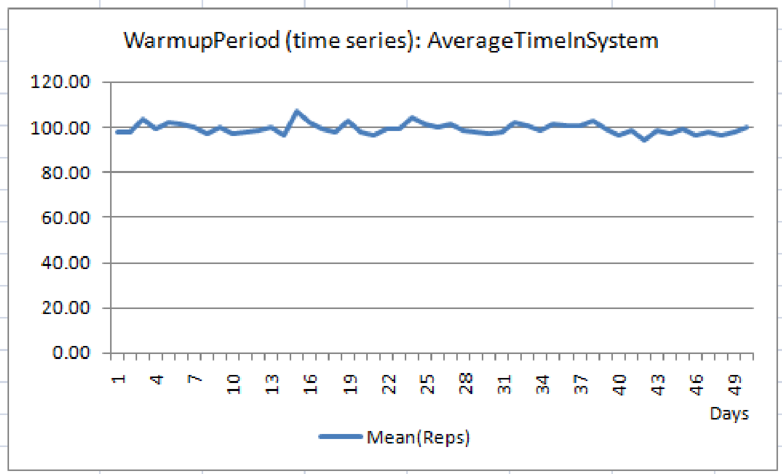
Day	Mean(Reps)
1	223.65
2	242.35
3	245.45
4	243.00
5	247.30
6	246.30
7	247.05
8	247.05
9	245.55
10	242.60
11	240.90
12	238.70



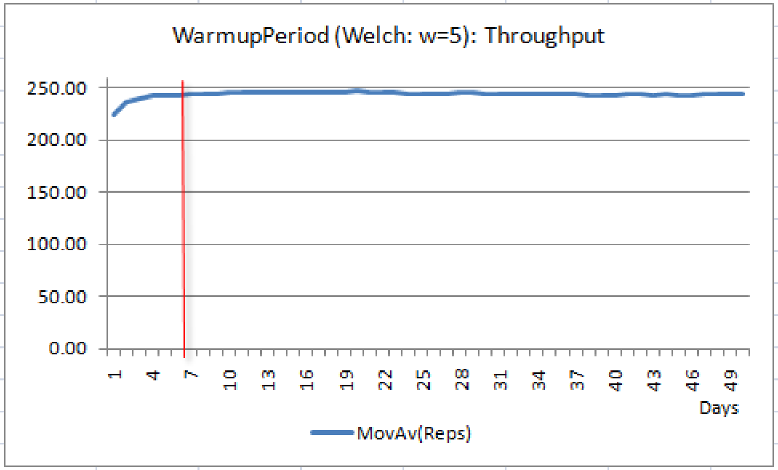
WarmupPeriod

AverageTimeInSystem

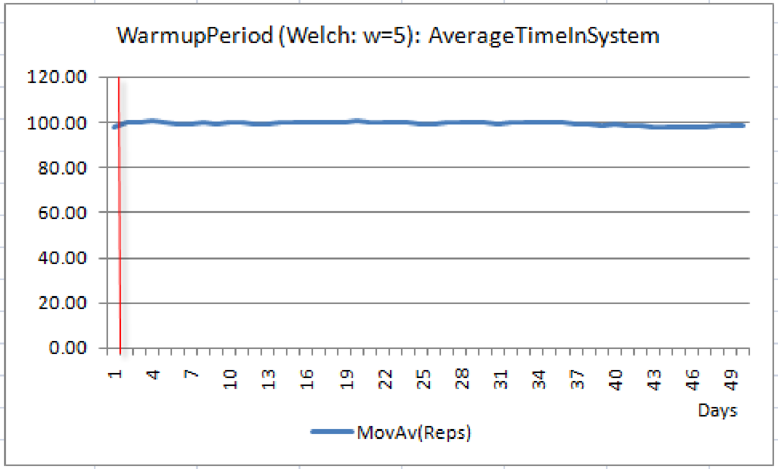
Day	Mean(Reps)
1	98.05
2	98.23
3	103.90
4	99.27
5	101.78
6	101.10
7	100.34
8	97.13
9	99.70
10	97.19
11	97.58
12	98.46



WarmupPeriod		
ThroughputPerDay: w=5		
Day	Mean(Reps)	MovAv(Rep)
1	223.65	223.65
2	242.35	237.15
3	245.45	240.35
4	243.00	242.16
5	247.30	243.08
6	246.30	242.84
7	247.05	244.20
8	247.05	244.84
9	245.55	244.50
10	242.60	245.27
11	240.90	245.22
12	238.70	245.71



WarmupPeriod		
AverageTimeInSystem: w=5		
Day	Mean(Reps)	MovAv(Rep)
1	98.05	98.05
2	98.23	100.06
3	103.90	100.24
4	99.27	100.38
5	101.78	99.94
6	101.10	99.48
7	100.34	99.52
8	97.13	99.69
9	99.70	98.99
10	97.19	99.72
11	97.58	99.72
12	98.46	99.56



Dealing with Initialisation Bias

1a. Running the model for a warm-up period: Important

- If the model has more than one key response the initial transient should be investigated for each one
- In theory, the warm-up period should be determined separately for every experimental scenario; in practice it is only done for the base scenario!

Run length

- Warm-up period generally takes less than 10 percent of the run length
  - Rule of thumb:
    - The experiment run length should be 10 x warm-up period
    - This implies a total run length of 11 x warm-up period (including 1 x for the warm up period that is removed)

Experimentation Preparation (Part 1)

1. Dealing with initialisation bias
   - Solutions:
     a) Run model for a warm-up period: Running the model until it reaches a realistic condition and only collect results from the model after this point
     b) Set initial conditions in the model: Place the model in realistic conditions at the start of the simulation run [not practical]

2. Obtaining sufficient output data
   - Solutions:
     a) Multiple replications (terminating or non-terminating simulations): Equivalent to taking multiple samples in statistics; multiple runs of the simulation model with different random number streams
     b) Single long run (non-terminating simulations): Equivalent to taking one large sample in statistics [not practical; some statistical concerns]

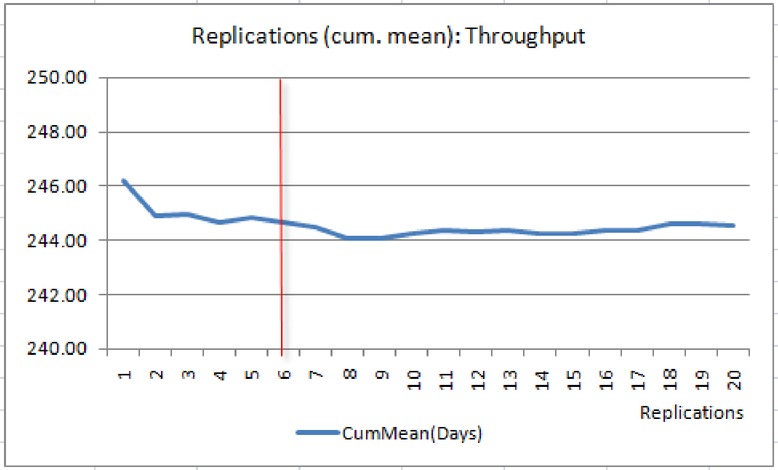
Obtaining Sufficient Output Data

2a. Multiple replications

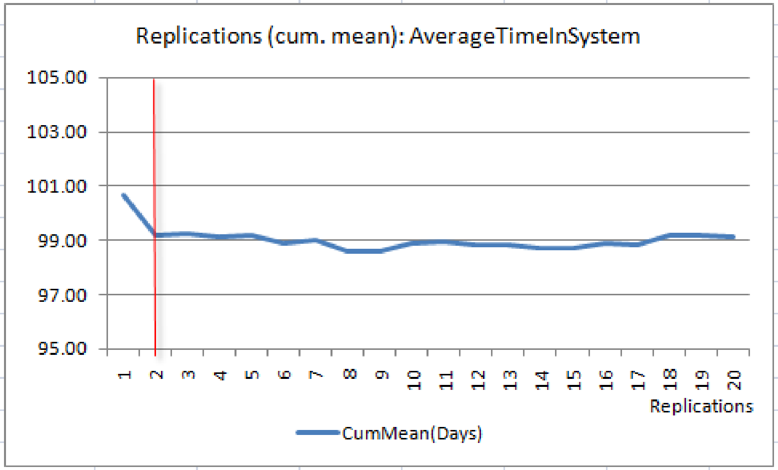
- Multiple replications are performed by changing the random number stream
- Producing multiple samples in order to get better estimates for the mean performance

- Most commonly used methods for estimating the number of required replications:
  - **Plotting cumulative means**
  - **Confidence interval method**

Number of replications		
Throughput		
Rep	Mean(Days)	CumMean(Days)
1	246.21	246.21
2	243.65	244.93
3	245.03	244.96
4	243.84	244.68
5	245.62	244.87
6	243.64	244.66
7	243.46	244.49
8	241.35	244.10
9	243.70	244.06
10	246.14	244.26
11	245.47	244.37
12	243.36	244.29



Number of replications		
AverageTimeInSystem		
Rep	Mean(Days)	CumMean(Days)
1	100.67	100.67
2	97.68	99.18
3	99.45	99.27
4	98.82	99.16
5	99.36	99.20
6	97.22	98.87
7	99.72	98.99
8	95.85	98.60
9	98.63	98.60
10	101.84	98.92
11	99.50	98.98
12	97.39	98.84



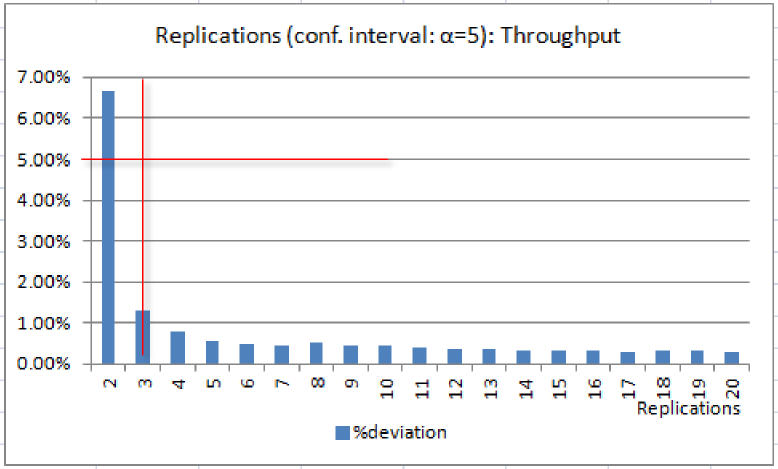
Obtaining Sufficient Output Data

2a. Confidence intervals (see www.wileyeurope.com/go/robinson - replications.xls)

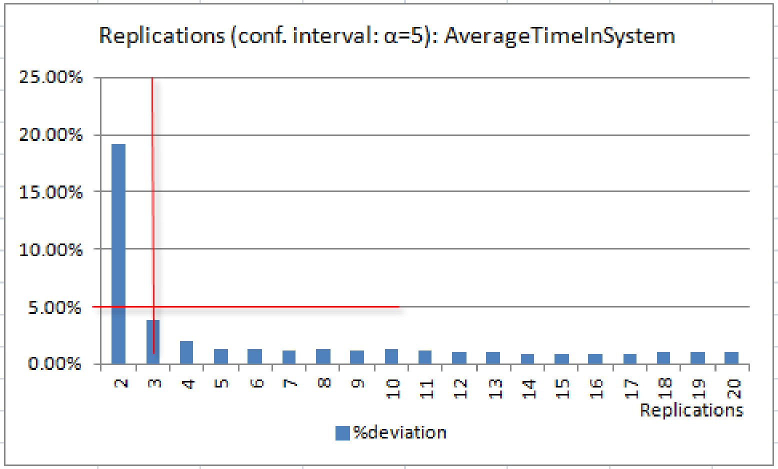
- Statistical means for showing how accurately the mean average of a value is estimated
- The narrower the interval the more accurate the estimate (i.e. the smaller the deviation between the upper and lower limit)
- Standard applications: Use 95% confidence interval (sign. level α=5%)
  - This gives a 95% probability that the value of the true mean (obtained if the model is run for an infinite period) lies within the confidence interval
- Critical application: Use 99% confidence interval (sign. level α=1%)

- The modeller needs to decide which %deviation between the upper and the lower limit is acceptable and choose the required number of replications to stay below this %deviation from the statistics

Number of replications			alpha	5%		
Throughput						
Rep	Mean(Days)	CumMean(Days)	StdDev(Days)	Lower interval	Upper interval	%deviation
1	246.21	246.21	n/a	n/a	n/a	n/a
2	243.65	244.93	1.81	228.64	261.22	6.65%
3	245.03	244.96	1.28	241.78	248.15	1.30%
4	243.84	244.68	1.19	242.79	246.58	0.77%
5	245.62	244.87	1.11	243.49	246.25	0.56%
6	243.64	244.66	1.11	243.50	245.83	0.48%
7	243.46	244.49	1.11	243.46	245.52	0.42%
8	241.35	244.10	1.52	242.83	245.37	0.52%
9	243.70	244.06	1.42	242.96	245.15	0.45%
10	246.14	244.26	1.50	243.19	245.33	0.44%
11	245.47	244.37	1.46	243.39	245.36	0.40%
12	243.36	244.29	1.43	243.38	245.20	0.37%



Number of replications			alpha	5%		
AverageTimeInSystem						
Rep	Mean(Days)	CumMean(Days)	StdDev(Days)	Lower interval	Upper interval	%deviation
1	100.67	100.67	n/a	n/a	n/a	n/a
2	97.68	99.18	2.12	80.17	118.18	19.16%
3	99.45	99.27	1.50	95.53	103.00	3.76%
4	98.82	99.16	1.25	97.17	101.14	2.00%
5	99.36	99.20	1.08	97.85	100.54	1.36%
6	97.22	98.87	1.26	97.54	100.19	1.34%
7	99.72	98.99	1.20	97.88	100.09	1.12%
8	95.85	98.60	1.57	97.28	99.91	1.33%
9	98.63	98.60	1.47	97.47	99.73	1.14%
10	101.84	98.92	1.72	97.69	100.15	1.24%
11	99.50	98.98	1.64	97.87	100.08	1.11%
12	97.39	98.84	1.63	97.81	99.88	1.05%



Obtaining Sufficient Output Data

2a. Multiple replications: Important

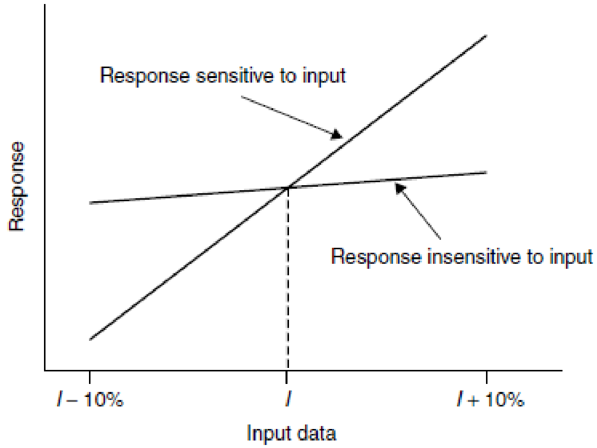
- Run length:
  - Transient models have a defined run length (e.g. one day)
  - Steady state models should be run at least 10x the warm-up period
- Remember to delete the warm-up period data before conducting any further analysis if you have a non-terminating simulation!
- In theory, the number of replications should be determined separately for every experimental scenario; in practice it is only done for the base scenario!
- If the model has more than one key response the number of replications should be chosen on the basis of the response that requires the most replications

Obtaining Sufficient Output Data

- Best solution:
  - Multiple replications of long runs!
    - The more output data can be obtained, the larger the sample, and the more certainty there can be in the accuracy of the results

Experiment Preparation

- Sensitivity Analysis
  - Improving the understanding of the model
  - Assessing the consequences of changes in model inputs
  - Model inputs: experimental factors and model data
  - Process:
    - Vary input (I)
    - Run the simulation
    - Measure the effect on the response
  - Result:
    - Significant shift in response?
      - Response sensitive to input change
    - No significant shift in response?
      - Response insensitive to input change



Robinson (2004)

Experiment Preparation

- Sensitivity Analysis
  - Useful for improving the understanding of the model:
    - Assessing the effect of uncertainties in the data
    - Understanding how changes to the experimental factors affect responses
    - Assessing the robustness of a solution
  - Sensitivity analysis can be time consuming and should be restricted to key inputs
    - Inputs about which there is greatest uncertainty
    - Inputs which are believed to have the greatest impact on the response

Experiment Preparation

- Developing an understanding of the solution space
  – Simulating a limited number of scenarios often allows you already to form an opinion as to the likely outcomes of other scenarios
  – How does it work? What methods can you apply?
    - Thinking about the likelihood of a scenario to give the desired output and only simulating the ones likely to lead to success
    - Using linear interpolation (assumes that the solution space is linear)
    - Using unconstrained models (e.g. removing queue limits) – in this way maximum requirements can be established [this is very useful!]
    - Perform some experiments with factor levels that are far apart

Experiment Preparation

- Experiment design: 2^k factorial designs
  - k = number of experimental factors
  - Each factor is set to two levels (+ and -)
  - Example:
    - k = 3 → 2^3 = 8 scenarios are simulated and the responses recorded
    - This allows to calculate the mean average effect on the response of changing a factor from its – to its + level → **Main Effect**
    - Main effect

Scenario	Factor 1	Factor 2	Factor 3	Response
1	−	−	−	R_1
2	+	−	−	R_2
3	−	+	−	R_3
4	+	+	−	R_4
5	−	−	+	R_5
6	+	−	+	R_6
7	−	+	+	R_7
8	+	+	+	R_8

$$e_1 = \frac{(R_2 - R_1) + (R_4 - R_3) + (R_6 - R_5) + (R_8 - R_7)}{4}$$

$$e_1 = \frac{-R_1 + R_2 - R_3 + R_4 - R_5 + R_6 - R_7 + R_8}{4}$$

Experiment Preparation

- Experiment design - 2^k factorial designs
  - Example (cont.)
    - Main effect
      - If main effect of factor is positive, changing the factor from – to + level increases the response by the value of the main effect (at average)
        - » Indicates the direction of change required to achieve a certain effect
        - » Identifies the most important factors
    - Interaction effect
      - Interaction effect between factor 1 and 2 (half the difference between the mean average effect of changing factor 1 (- to +) when factor 2 is at its + level and half the difference between the mean average effect of changing factor 1 (- to +) when factor 2 is at its - level)

$$e_{12} = \frac{1}{2}\left(\frac{(R_4 - R_3) + (R_8 - R_7)}{2} - \frac{(R_2 - R_1) + (R_6 - R_5)}{2}\right)$$

$$e_{12} = \frac{R_1 - R_2 - R_3 + R_4 + R_5 - R_6 - R_7 + R_8}{4}$$

Experiment Preparation

- Other approaches to Experimental Design
  - Fractional factorial design
    - When you have too many factors
  - ANOVA
    - Involves a series of hypothesis tests in which it is determined whether changes to the experimental factors have an effect on the response

Experimentation

- Interactive experimentation
  - Involves watching the simulation and making changes to the model to see the effects
  - Aim:
    - Develop an understanding of the model
    - Identify key problem areas
    - Identify potential solutions

- Batch experimentation
  - Setting experimental factors and leaving the models to run for a pre-defined run length
  - Display is normally switched off to improve run speed

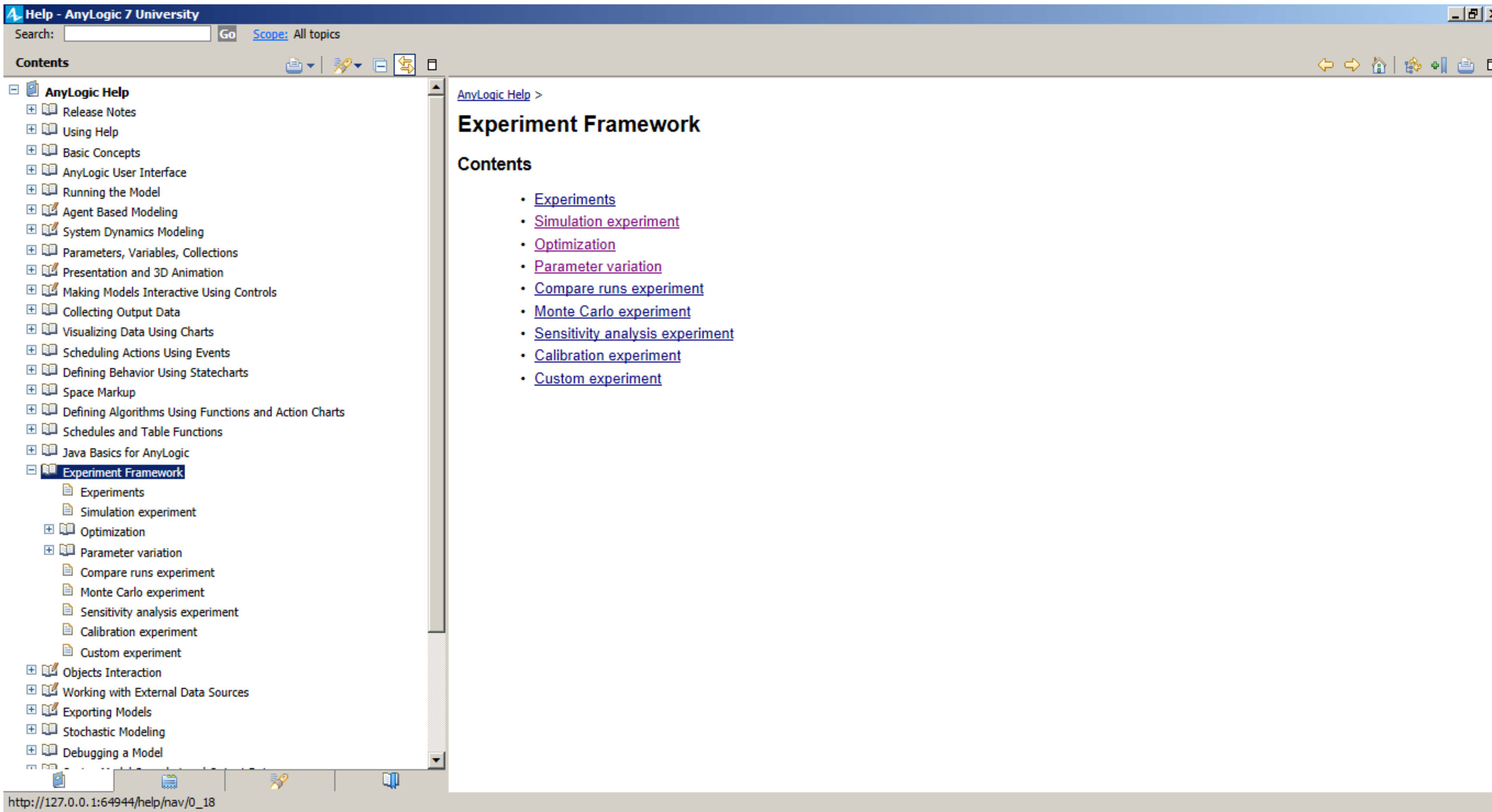
Experimentation

- Comparing alternatives
  - There is a limited number of scenarios to be compared
  - Scenarios emerge as the simulation study progresses

- Search experimentation
  - There are no predefined scenarios
  - One or more experimental factors are varied until a target or optimum level is reached

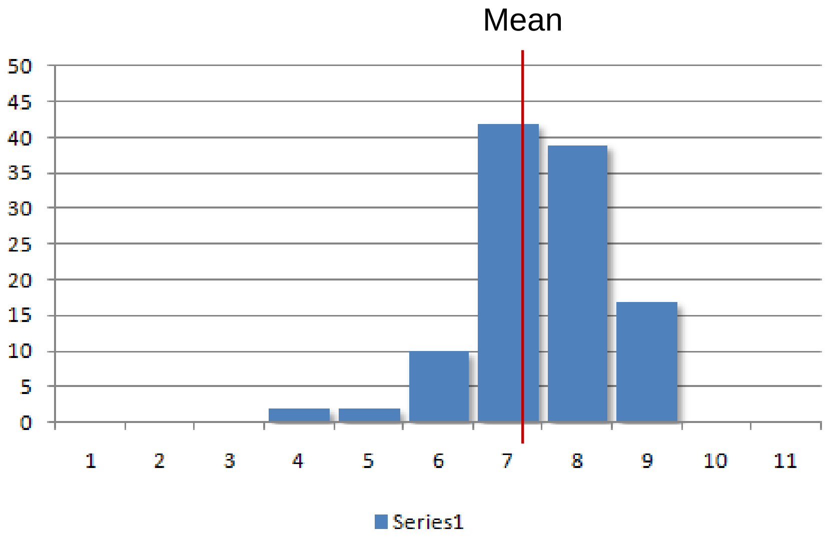
Both approaches are not mutually exclusive!

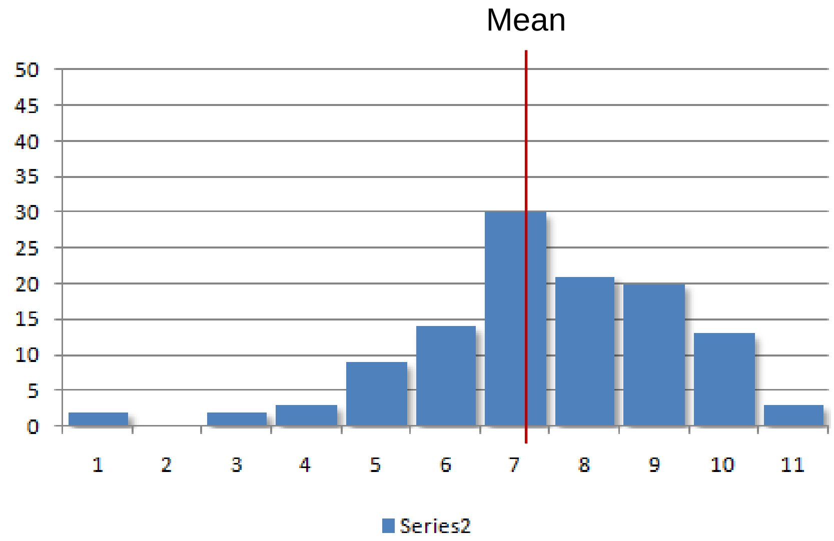
AnyLogic Experiment Framework



Output Analysis – Single Scenario

- For each response two measures are generally of interest
  - Mean
  - Variability





Output Analysis – Single Scenario

- Point estimate: Mean
  - A point estimate is a single value given as the estimate of a population parameter that is of interest, for example the mean of some quantity

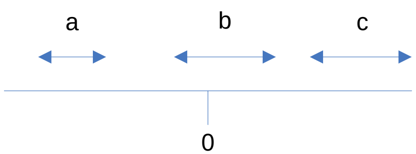
- Interval estimate: Variability
  - Because simulation experiments provide only a sample of output data it is important that a confidence interval for each mean is reported
  - The confidence interval provides information about the range within which the population mean is expected to lie

Output Analysis - Comparing Alternatives

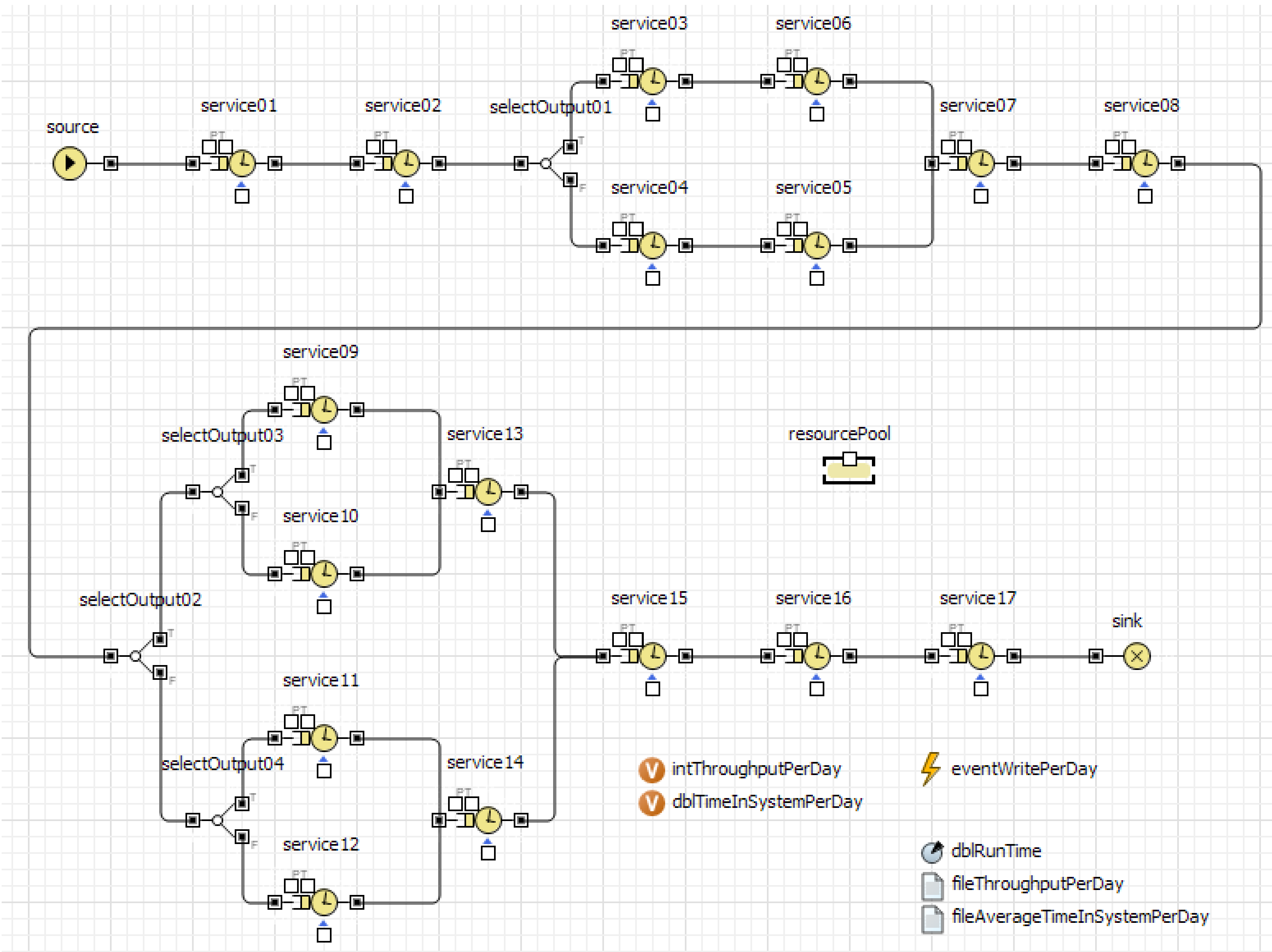
- Is the difference in a results significant?
  - Not simply a case of comparing mean values of key responses
  - Example:
    - Key response: daily throughput
      - Scenario A: Mean = 1050 units per day
      - Scenario B: Mean = 1080 units per day
    - Is scenario B the better alternative?
    - We need to consider two more factors:
      - Standard deviation of each mean daily throughput
      - Number of replications (or batches)
    - A small number of replications and a lot of variation in the results gives little confidence that the difference is significant!

Output Analysis - Comparing Alternatives

- A paired-t confidence interval helps to identify the statistical significance of a difference in the result of two scenarios.
- Resulting confidence interval can have three outcomes
  - It can be concluded **with the specified level of confidence** (usually 95%) that the result of scenario 1 is **less than (<)** the result for scenario 2
  - It can be concluded **with the specified level of confidence** (usually 95%) that the result of scenario 1 is **not significantly different** from the result of scenario 2
  - It can be concluded **with the specified level of confidence** (usually 95%) that the result of scenario 1 is **greater than (>)** the result for scenario 2



Potential outcomes from a paired-t confidence interval


source
service01
service02
selectOutput01
service03
service06
service04
service05
service07
service08
service09
selectOutput03
service13
resourcePool
service10
selectOutput02
service15
service16
service17
sink
service11
selectOutput04
service14
service12
intThroughputPerDay
dblTimeInSystemPerDay
eventWritePerDay
dblRunTime
fileThroughputPerDay
fileAverageTimeInSystemPerDay

Output Analysis - Comparing Alternatives

- Experimental set up:
  - Warm-up period: 6 days
  - Number of replications: 6
  - Run length: 10x warm-up period = 60 days

Output Analysis - Comparing Alternatives

- Throughput per day

Comparison of Two Scenarios									
						Sign. level	5.0%		
	randomTrue(0.5)	randomTrue(0.7)				Confidence interval			
Replication	Scenario 1 result	Scenario 2 result	Difference	Cum. mean difference	SD of cum. mean difference	Lower interval	Upper interval	Conclusion
1	246.02	239.88	6.14	6.14	n/a	n/a	n/a	n/a
2	243.51	241.64	1.87	4.01	3.019	-23.12	31.13	No difference
3	244.79	244.68	0.11	2.71	3.101	-5.00	10.41	No difference
4	243.99	244.59	-0.60	1.88	3.024	-2.93	6.69	No difference
5	245.65	244.24	1.41	1.79	2.627	-1.48	5.05	No difference
6	243.66	243.2	0.46	1.57	2.411	-0.97	4.10	No difference

- Average time in system per day

Comparison of Two Scenarios								
						Sign. level	5.0%	
	randomTrue(0.5)	randomTrue(0.7)				Confidence interval		
Replication	Scenario 1 result	Scenario 2 result	Difference	Cum. mean difference	SD of cum. mean difference	Lower interval	Upper interval	Conclusion
1	100.66	97.36	3.30	3.30	n/a	n/a	n/a	n/a
2	97.68	97.35	0.33	1.81	2.102	-17.07	20.70	No difference
3	99.38	98.40	0.98	1.54	1.562	-2.35	5.42	No difference
4	98.89	98.37	0.52	1.28	1.374	-0.91	3.47	No difference
5	99.41	100.30	-0.89	0.85	1.535	-1.06	2.75	No difference
6	97.81	97.39	0.42	0.78	1.384	-0.68	2.23	No difference

Output Analysis - Comparing Many Scenarios

- Use paired-t confidence interval + Bonferroni inequality
  - If we want to make ***c*** confidence interval statements, the confidence interval should be formed with a significance level of **α/c**
  - If we want to compare each scenario to the current set-up (base scenario) then **c=s-1** where **s** is the number of scenarios
    - Example: Comparing 4 scenarios to the base scenario
      - c = 5-1 = 4
      - If overall confidence required = 95% (α = 5%) then the individual significance levels α = 5/4 = 1.25%
  - If we want to compare each scenario to each scenario then **c = s(s-1)/2**
    - Example: Comparing 5 scenarios
      - c = 5*(4/2) = 10
      - If overall confidence required = 95% (α = 5%) then the individual significance levels α = 5/10 = 0.5%

Output Analysis - Comparing Many Scenarios

- The examples in tables:

Comparing Scenario 2-5 to Base Scenario 1: Overall confidence 95%			
	98.75% confidence intervals for differences		
Comparison	Lower interval	Upper interval	Conclusion
Scenario 1 to 2	...	...	Scen. 1 > Scen. 2
Scenario 1 to 3	...	...	No difference
Scenario 1 to 4	...	...	...
Scenario 1 to 5	...	...	...

Comparing between all scenarios: Overall confidence 95%				
Calculations	99.5% confidence intervals for differences			
Scenario	2	3	4	5
1	lower int., upper int.	lower int., upper int.	lower int., upper int.	lower int., upper int.
2		lower int., upper int.	lower int., upper int.	lower int., upper int.
3			lower int., upper int.	lower int., upper int.
4				lower int., upper int.

Comparing between all scenarios: Overall confidence 95%				
Conclusions	99.5% confidence intervals for differences			
Scenario	2	3	4	5
1	Scen. 1 > Scen. 2	No difference	...	...
2		...	...	...
3			...	...
4				...

Output Analysis - Comparing Many Scenario

- Conclusion:
  - Use of Bonferroni inequality is quite effective as long as the number of scenarios stays small
  - As the amount of scenarios increases the number of confidence intervals can quickly become unmanageable (in particular for full comparisons)
  - What do you think about the following statement?
    - For finding the best result out of a group of scenarios we can simply look at the mean values of the different scenarios
      - Use confidence intervals
      - Choose other statistical methods (ranking and selection methods)
      - For more see Law and Kelton (2006)

Questions / Comments

References

- Biller B and Nelson BL (2002). Answers to the Top Ten Input Modeling Questions. In: Proceedings of the 2002 Winter Simulation Conference, San Diego, California, USA
- Janssen MA and Ostrom E (2006). Empirically based, agent-based models. Ecology and Society 11(2): 37
- Luna-Reyes LF and Andersen DL (2003). Collecting and Analyzing Qualitative Data for System Dynamics: Methods and Models. System Dynamics Review 19(4) pp.271–296
- Robinson S (2004). Simulation: The Practice of Model Development and Use. John Wiley & Sons: Chichester, UK
- Law A and Kelton WD (2006). Simulation Modeling and Analysis, 4th Ed, McGraw Hill: New York, USA
- Robinson (2002). A statistical process control approach for estimating the warm-up period. In: Proceeding of the 2002 Winter Simulation Conference, San Diego, CA