0405324: Stochastic System Simulation

Lecture 7: Output analysis

Motivation

- Random input (using RNG) leads to random output (RIRO)
- Run a simulation (once) — what does it mean?
  - Was this run "typical" or not?
  - Variability from run to run (of the same model)?

- Need statistical analysis of output data
  - Evaluating the **absolute** performance of a single model configuration
  - Evaluating the **relative** performance by comparing two (Output Analyzer of Arena) or more (Process Analyzer of Arena) different configurations or scenarios of your model
  - Search for an optimal configuration (OptQuest tool of Arena)
- If the performance of a system is measured by a parameter θ, the result of a set of simulation experiments will be an estimator $\hat{\theta}$ of θ.
- The precision of $\hat{\theta}$ can be measured by: standard error of $\hat{\theta}$ or the width of the CI.
- Objective of the statistical analysis: either estimate the standard error, or the CI, or to figure out the number of observations required to achieve a standard error or a CI of a given size, for any typical output variable Y, of a simulation model.

- Statistical analysis of output is often ignored
  - This is a big mistake – no idea of precision of results → may lead to bad decision based on imprecise output results
  - Not hard or time-consuming to do this – it just takes a little planning and thought, then some (cheap) computer time

Output Analysis Activities

- **Output Analysis is the modeling stage concerned with**
  - Simulation replications
  - Computing statistics from replications
  - Presenting statistics in textual or graphical format

Types of Simulation Models

- Simulation models can be classified into two main classes, based on their time horizon:
  - ***Terminating models***
  - ***Steady-state (non-terminating) models***

Terminating Simulation Models

- **A *terminating* simulation model has a time frame with well defined and natural termination time (end) for its replications and clearly defined way to start**
  - Runs for some duration of time T_E, where E is a specified event that stops the simulation.
  - Starts at time *0* under well-specified initial conditions (for example: empty and idle).
  - Ends at the stopping time T_E (or stopping event E).
  - **Example:**
    - A bank opens at 8:30 am (time *0*) with no customers present and *8* of the *11* tellers working (initial conditions), and closes at 4:30 pm (Time $T_E = 480$ minutes).
    - The simulation analyst chooses to consider it a ***terminating*** system because the object of interest is one day's operation.
    - If the interest was the flow of money from the ATM machine of the bank, then the system would be a non-terminating one (steady state)
  - **Example:**
    - A job shop that operates for as long as it takes to produce a "run" of 500 completed assemblies specified by an order.

Steady-state or long run (non-terminating) Simulation Models

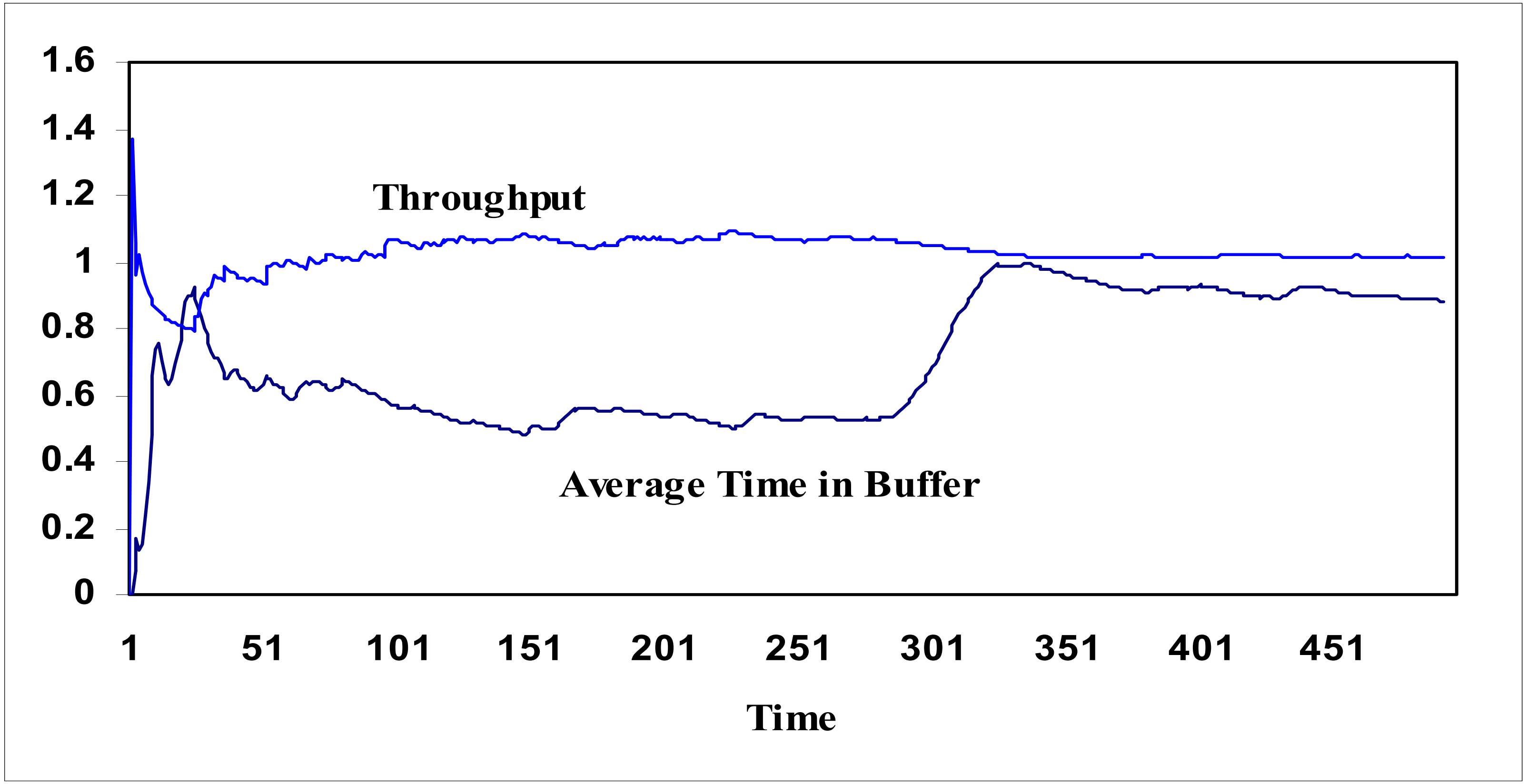
- A ***steady state*** simulation model has no natural termination time for its replications.
  - Runs continuously, or at least over a very long period of time.
  - **Examples:** assembly lines that shut down infrequently, telephone systems, hospital emergency rooms.
  - Initial conditions are defined by the analyst.
  - Runs for some analyst-specified period of time T_E (should be carefully selected: long enough).
  - Study the steady-state (long-run) properties of the system, properties that are not influenced by the initial conditions of the model.
- **Whether a simulation is considered to be terminating or non-terminating depends on both**
  - The objectives of the simulation study and
  - The nature of the system.

Steady-State Simulation Models

- While long-term statistics are of interest, initial system conditions tend to bias its long-term statistics
  - it, therefore, makes sense to start statistics collection after an initial period of system **warm-up**, namely, after the biasing effect of the initial conditions decays to insignificance:

    Arena: *Run>Setup>Replication Parameters>Warmup period*
  - the transient-state regime is characterized by statistics that vary as function of time, while the steady-state regime prevails when statistics stabilize and do not vary over time

Sample Steady-State Simulation Output



Workstation throughput and average time in buffer (queue) as functions of time

Steady-State Simulation Issues

- Since steady-state models have no natural termination time, how does one select a replication length?
  - the replication can be stopped when statistics at the end of several successive increments are sufficiently close (e.g., within some difference, to be determined by the analyst)
- Since a warm-up period is needed to eliminate statistical bias, how does one select the warm-up length?
  - in a similar vein, the length of a warm-up period is determined by observing experimentally when the time variability of the statistics of interest largely disappears

Stochastic Nature of Output Data

- **Model output consists of one or more random variables (r.v.) because the model is an input-output transformation and the input variables are r.v.'s.**
- **M/G/1 queueing example:**
  - Poisson arrival rate = *0.1* per minute;
    service time ~ $N(\mu = 9.5, \sigma = 1.75)$.
  - System performance: long-run mean queue length, $L_q(t)$.
  - Suppose that we run a single simulation for a total of 5,000 minutes
    - Divide the time interval [*0, 5000*) into *5* equal subintervals of *1000* minutes.
    - Average number of customers in queue from time *(j-1)1000* to *j(1000)* is Y_j.

- **M-Markovian, G-General, 1-one server**

Stochastic Nature of Output Data

- **M/G/1 queueing example (cont.):**
  - *Batched* average queue length for 3 independent replications:

Distinguish the *within-replication* data from *across-replication* data

Batching Interval (minutes)	Batch, j	Replication, R (i)		
		1, Y_{1j}	2, Y_{2j}	3, Y_{3j}
[0, 1000)	1	3.61	2.91	7.67
[1000, 2000)	2	3.21	9.00	19.53
[2000, 3000)	3	2.18	16.15	20.36
[3000, 4000)	4	6.92	24.53	8.11
[4000, 5000)	5	2.82	25.19	12.62
[0, 5000)		3.75 ($\bar{Y}_{1.}$)	15.56 ($\bar{Y}_{2.}$)	13.66 ($\bar{Y}_{3.}$)

Yij: average number of customers in the queue in batch j of replication i.

  - Inherent variability in stochastic simulation both within a single replication and across different replications.
  - The average of the 3 replications, $\bar{Y}_{1.}, \bar{Y}_{2.}, \bar{Y}_{3.}$, can be regarded as independent observations, but averages within a replication, $Y_{11}, \ldots, Y_{15}$, are not.

Note: The *Batch Means* method is a practical way of collecting multiple estimates from a single replication (terminating simulation)

Measures of performance

- **Consider the estimation of a performance parameter, θ (or ϕ), of a simulated system.**
  - Simulation output data: discrete time "tally" data $[Y_1, Y_2, \ldots, Y_R]$, with ordinary mean: θ
    - Average System Time
    - Average Waiting Time
  - Simulation output data: continuous-time "time-persistent" data $\{Y(t), 0 \leq t \leq T_E\}$ with time-weighted mean: ϕ
    - Average Queue Length
    - Average Utilization
  - It does not matter whether we use θ (or ϕ): this is only to distinguish between discrete time and time-persistent measure of performance.

Point Estimator [Performance Measures]

- **Point estimation for discrete-time data.**
  - The point estimator of θ based on simulation data [$Y_1, Y_2, \ldots, Y_R$],

$$\hat{\theta} = \frac{1}{R}\sum_{i=1}^{R} Y_i \qquad \hat{\theta} : \text{is a sample (size } R\text{) mean}$$

  - Is unbiased if $E(\hat{\theta}) = \theta$ ⇐ ***Desired***

$$E(\hat{\theta}) - \theta : \text{bias in the point estimator}$$

- **Point estimation for continuous-time data.**
  - The point estimator:

$$\hat{\phi} = \frac{1}{T_E}\int_0^{T_E} Y(t)\,dt$$

    - Is biased if $E(\hat{\phi}) \neq \phi$
    - An unbiased or low-bias estimator is desired.

Confidence-Interval Estimation

[Performance Measures]

- **CI defines how precise is the point estimator assuming that it is described with a probability model (α and the width of the CI define the precision)**
- **Suppose the probability model is a normal distribution with mean θ, variance σ^2 (both unknown).**
  - Let $\overline{Y}_{i\cdot}$ be the average cycle time for parts produced on the i^{th} replication of the simulation (its mathematical expectation is θ): average of the ***within-replication*** data.
  - Average cycle time will vary from day to day (from replication to replication), but over the long-run the average of the averages $\overline{Y}_{\cdot\cdot}$ (***cross-replication*** average) will be close to θ.
  - Sample variance across R replications:

$$S^2 = \frac{1}{R-1}\sum_{i=1}^{R}(\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot})^2$$

Where, $\overline{Y}_{\cdot\cdot} = \frac{1}{R}\sum_{i=1}^{R}\overline{Y}_{i\cdot}$

Note:

But $\overline{Y}_{\cdot\cdot}$ is estimated based on a sample and it has error. A CI is a measure of that error.

Confidence-Interval Estimation

[Performance Measures]

- **Confidence Interval (CI):**
  - A measure of error.
  - Where $Y_{i.}$ are normally distributed.

$$\bar{Y}_{..} \pm t_{\alpha/2, R-1} \frac{s}{\sqrt{R}}$$

  - We cannot know with certainty how far $\bar{Y}_{..}$ is from θ but CI attempts to bound that error.
  - A CI, such as 95%, tells us how much we can trust the interval to actually bound the error between $\bar{Y}_{..}$ and θ.
  - The more replications we make, the less error there is in $\bar{Y}_{..}$ (converging to 0 as R goes to infinity).

Output Analysis for Terminating Simulations

- **A terminating simulation: runs over a simulated time interval [$0, T_E$].**
- **A common goal is to estimate:**

$$\theta = E\left(\frac{1}{R}\sum_{i=1}^{R} Y_i\right), \qquad \text{for discrete output}$$

$$\phi = E\left(\frac{1}{T_E}\int_0^{T_E} Y(t)\,dt\right), \quad \text{for continuous output } Y(t), 0 \le t \le T_E$$

- **In general, independent replications are used: each run uses a different stream of random numbers and independently chosen initial conditions.**

Statistical Background

[Terminating Simulations]

- **Important to distinguish within-replication data from across-replication data.**
- **For example, simulation of a manufacturing system**
  - Two performance measures of that system: cycle time for parts and work in process (WIP).
  - Let Y_{ij} be the cycle time for the j^{th} part produced in the i^{th} replication.
  - Across-replication data are formed by summarizing within-replication data $\overline{Y}_{i.}$.

Statistical Background

[Terminating Simulations]

- **Within replication:**
  - For example: the WIP (a continuous time data)
    - The average: $\bar{Y}_{i.} = \frac{1}{T_{Ei}} \int_0^{T_{Ei}} Y_i(t)\,dt$
    - The sample variance: $S_i^2 = \frac{1}{T_{Ei}} \int_0^{T_{Ei}} \left(Y_i(t) - \bar{Y}_{i.}\right)^2 dt$
- **Across Replication:**
  - For example: the daily cycle time averages (discrete time data)
    - The average: $\bar{Y}_{..} = \frac{1}{R} \sum_{i=1}^{R} \bar{Y}_{i.}$
    - The sample variance: $S^2 = \frac{1}{R-1} \sum_{i=1}^{R} (\bar{Y}_{i.} - \bar{Y}_{..})^2$

Statistical Background

[Terminating Simulations]

- **Overall sample average, $\bar{Y}_{..}$, and the interval replication sample averages, $\bar{Y}_{i.}$, are always unbiased estimators of the expected daily average cycle time or daily average WIP.**

- **Across-replication data are independent (based on different random numbers) and identically distributed (same probability model), but within-replication data do not have these properties.**

C.I. with Specified Precision

[Terminating Simulations]

- **Confidence interval:** $\bar{Y}_{..} \pm t_{\alpha/2,R-1} \frac{S}{\sqrt{R}}$

- **The half-width *H* of a *100(1 – α)%* confidence interval for a mean *θ*, based on the *t* distribution, is given by:**

$$H = t_{\alpha/2,R-1} \frac{S}{\sqrt{R}}$$

R is the # of replications

*s*² is the sample variance

- **Prefer smaller confidence intervals — *precision***

- **Half-width,** $H = t_{\alpha/2,R-1} \frac{S}{\sqrt{R}}$

Want this to be "small," say ≤ h where h is pre-specified half-width or error

Called the standard error (interpreted as the average error in Y.. in estimating θ)

- **Can't control *t* or *s***

- **Must increase *R* — how much?**

C.I. with Specified Precision
[Terminating Simulations]

- **If a predetermined half width interval h is required, and we want to determine the value of R that provides such interval, then solving** $H = t_{\alpha/2,R-1}\frac{s}{\sqrt{R}} = h$ **for *R* gives:**

$$R = \left(t_{\alpha/2,R-1}\frac{s}{h}\right)^2$$

Depends on R through $t_{\alpha/2,R-1}$ and s

- **Assume that an initial sample of size R_0 (independent) replications have been observed.**
- **Using R_0, obtain an initial estimate s_0^2 of the population variance σ^2.**
- **Substitute $t_{\alpha/2,R-1}$ by $z_{\alpha/2}$ since for large R, t-distribution → z-distribution**
- **Then, choose sample size *R* such that $R \geq R_0$:**
  - Since $t_{\alpha/2,R-1} \geq z_{\alpha/2}$, an initial estimate of *R*:

$$R \geq \left(\frac{z_{\alpha/2}s_0}{h}\right)^2, \quad z_{\alpha/2} \text{ is the standard normal distribution.}$$

  - *R* is the smallest integer satisfying $R \geq R_0$ and $R \geq \left(\frac{z_{\alpha/2}s_0}{h}\right)^2$
- **Collect $R - R_0$ additional observations (or start over and make R new replication).**
- **The *100(1-α)%* C.I. for θ:** $\bar{Y}_{..} \pm t_{\alpha/2,R-1}\frac{S}{\sqrt{R}}$

C.I. with Specified Precision

[Terminating Simulations]

- **Example-The cross-replication summary outputs from M/G/1 simulation model based on 10 replications is as follows:**

	Total Cost ($)	Percent Rejected
Sample Mean	21,618.33	11.12
Sample Standard Deviation	1,136.24	1.30
95% Confidence Interval Half Width	812.82	0.93
Minimum Summary Output Value	20,056.75	9.43
Maximum Summary Output Value	23,837.38	13.55

  - **What should be the sample size to get the half-width down to ± 250 or less for total cost?**

Try yourself: Consider percent rejected and repeat the same for half-width of 0.50.

C.I. with Specified Precision

[Terminating Simulations]

- **Solution:**
  - **Given that from initial 10 replications, 95% half-width on Total Cost was ± 812.82.**
  - **Using approximation:**

$$\bar{Y}_{..} = 21618.33,\ s_0 = 1136.24,\ z_{0.025} = 1.96,\ h_0 = 812.76,\ R_0 = 10,\ h = 250$$

$$R \geq \left(z_{0.025} \frac{s_0}{h} \right)^2 = 1.96^2 \cdot \frac{1136.24^2}{250^2} = 79.4 \quad \rightarrow R=80$$

  - Finalize *R*

R	**80**	**81**	**82**
$t_{0.025,\ R-1}$	**1.95996**	**1.95996**	**1.95996**
$H = t_{0.025,R-1} \frac{s_0}{\sqrt{R}}$	**248.98**	**247.44**	**?**
$\left(\frac{t_{\alpha/2,R-1} s_0}{h} \right)^2$	**79.35**	**?**	**?**

Since new H < 250 and new R=79.35→80 then
Final *R* = 80

C.I. with Specified Precision

[Terminating Simulations]

- **Example**—The data for four replications of a technical-support call center simulation model are as follows:

Run	Utilization	Avge time in system (min)
1	0.808	3.74
2	0.875	4.53
3	0.708	3.84
4	0.842	3.98

(i) Estimate 95% CI half-width and the CI for the server utilization.

(ii) What do you suggest to do if we want to estimate the utilization in (i) with a precision of ± 0.04 with a probability 0.95?

Try yourself: Consider average time in system and repeat (i) and (ii)

C.I. with Specified Precision

[Terminating Simulations]

- **Solution—**
  here, R_0 = 4, $\bar{Y}_{..} = \frac{0.808 + ... + 0.842}{4} = 0.808$,

$$s_0^2 = \frac{(0.808 - 0.808)^2 + ... + (0.842 - 0.808)^2}{4-1} = 5.18 \times 10^{-3}$$

or, $s_0 = 0.072$,

$t_{0.025,3} = 3.18$

(i) 95% CI half-width, $H = t_{\alpha/2,R-1}\frac{s}{\sqrt{R}} = 3.18 \times \frac{0.072}{\sqrt{4}} = 0.114$

95% CI on utilization (ρ) is

$$\bar{Y}_{..} - t_{\alpha/2,R-1}\frac{s}{\sqrt{R}} \le \rho \le \bar{Y}_{..} + t_{\alpha/2,R-1}\frac{s}{\sqrt{R}}$$

or, $0.808 - 0.114 \le \rho \le 0.808 + 0.114$

or, $0.694 \le \rho \le 0.922$

C.I. with Specified Precision

[Terminating Simulations]

(ii) 95% CI half-width, $h = 0.04$

$$R \geq \left(z_{0.025} \frac{s_0}{h} \right)^2 = 1.96^2 \cdot \frac{0.072^2}{0.04^2} = 12.44 \longrightarrow \text{try } R=13$$

Inacceptable >0.04

R	13	14	15
$t_{0.025,\ R-1}$	2.18	2.16	2.14
$H = t_{0.025,R-1} \dfrac{S_0}{\sqrt{R}}$	0.0435	0.0415	0.0397
$\left(\dfrac{t_{\alpha/2,R-1} S_0}{h} \right)^2$	15.39	15.10	14.83

Final $R = 15$

Suggestion– Use replicate size of at least 15 to estimate the utilization in (i) with a precision of ±0.04.

Output Analysis for Steady-State (long run) Simulation

- **Consider a single run of a simulation model to estimate a steady-state or long-run characteristics of the system.**
  - The single run produces observations $Y_1, Y_2, \ldots$ (generally the samples of an autocorrelated time series).
  - Performance measure:

$$\theta = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} Y_i, \qquad \text{for discrete measure}$$

(with probability 1: because it is defined as a limit)

$$\phi = \lim_{T_E \to \infty} \frac{1}{T_E} \int_0^{T_E} Y(t)\,dt, \qquad \text{for continuous measure}$$

(with probability 1: because it is defined as a limit)

    – Independent of the initial conditions.

Output Analysis for Steady-State Simulation

- **The simulation analyst could decide to stop the simulation after some number of observations–say, *n*–have been collected, or**
- **Could decide to simulate for length of time T_E that determines *n* (may vary from run to run)**

Output Analysis for Steady-State Simulation

- **The sample size n (or T_E) is a design choice (i.e. it is not inherently determined by the nature of the problem), with several considerations in mind:**
  - Any bias in the point estimator that is due to artificial or arbitrary initial conditions (bias can be severe if run length is too short).
  - The desired precision of the point estimator (for instance the CI half width).
  - Budget constraints on the time available to execute the simulation or the computer resources.
- **Notation: the estimation of θ from a discrete-time output process (easier than continuous time output).**
  - One replication (or run), the notation of output data: $Y_1, Y_2, Y_3, \ldots$
  - With several replications, the notation of output data for replication r: $Y_{r1}, Y_{r2}, Y_{r3}, \ldots$

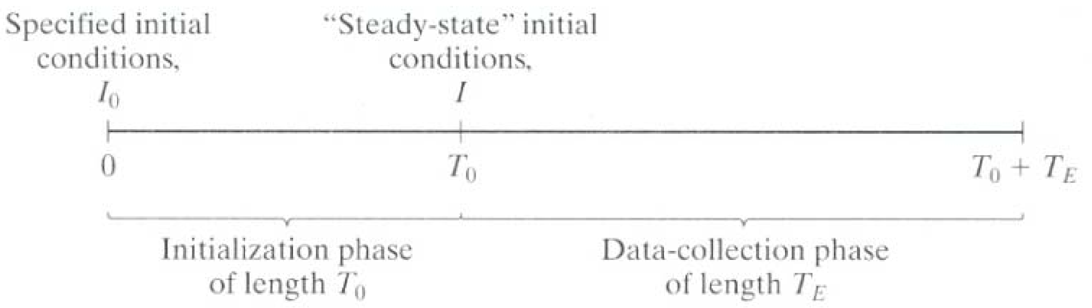
Initialization Bias [Steady-State Simulations]

- **Methods to reduce the point-estimator bias caused by using artificial and unrealistic initial conditions in a steady state simulation:**
  - **Intelligent initialization.**
  - **Divide each simulation into two phases (initialization phase and data-collection phase).**

- **Intelligent initialization**
  - Initialize the system in a state that is more representative of long-run conditions (e.g. by setting the number in the queue, instead of starting with an idle system):
    - If the system exists, collect data on it and use these data to specify more nearly typical initial conditions.
    - If the system can be simplified enough to make it mathematically solvable, e.g. queueing models, solve the simplified model to find long-run expected or most likely conditions–such as the expected number of customers in the queue–and these conditions can be used to initialize the simulation.

Initialization Bias [Steady-State Simulations]

- Divide each simulation into two phases
  - This method can be used in conjunction with the previous one.
  - An initialization phase, from time 0 to time T_0.
  - A data-collection phase, from T_0 to the stopping time T_0+T_E.
  - The choice of T_0 is important:
    - After T_0, the system state should be more nearly representative of steady-state behavior.



Initialization and data collection phases of a steady-state simulation run.

  - I: is a random variable → saying that the system has reached a steady state means: the probability distribution of the system state is close to the steady-state probability distribution (bias of response variable is negligible).

Initialization Bias [Steady-State Simulations]

- **M/G/1 queueing example: A total of 10 independent replications were made (table in the next slide).**
  - Each replication beginning in the empty and idle state.
  - Simulation run length on each replication was $T_0+T_E = 15{,}000$ minutes.
  - Response variable: queue length, $L_q(t,r)$ (at time t of the rth replication).
  - Batching intervals of 1,000 minutes, batch means
- **Ensemble averages: average corresponding batch means *across* replication**
  - Goal is to identify trend in the data due to initialization bias
  - The average corresponding batch means *across* replications:

$$\bar{Y}_{.j} = \frac{1}{R}\sum_{r=1}^{R} Y_{rj}$$

R replications

  - **The preferred method to determine deletion point.**

Initialization Bias [Steady-State Simulations]

--Cumulative means

Individual Batch Means (Y_{rj}) for *M/G/1* Simulation with Empty and Idle Initial State

	j → Batch														
Replication (r ↓)	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
1	3.61	3.21	2.18	6.92	2.82	1.59	3.55	5.60	3.04	2.57	1.41	3.07	4.03	2.70	2.71
2	2.91	9.00	16.15	24.53	25.19	21.63	24.47	8.45	8.53	14.84	23.65	27.58	24.19	8.58	4.06
3	7.67	19.53	20.36	8.11	12.62	22.15	14.10	9.87	23.96	24.50	14.56	6.08	4.83	16.04	23.41
4	6.62	1.75	12.87	8.77	1.25	1.16	1.92	6.29	4.74	17.43	18.24	18.59	4.62	2.76	1.57
5	2.18	1.32	2.14	2.18	2.59	1.20	4.11	6.21	7.31	1.58	2.16	3.08	2.32	2.21	3.32
6	0.93	3.54	4.80	0.72	2.95	5.56	1.96	2.07	2.74	3.45	14.24	13.39	7.87	0.94	3.19
7	1.12	2.59	5.05	1.16	2.72	5.12	5.03	4.14	4.98	15.81	9.29	2.14	8.72	29.80	28.94
8	1.54	5.94	5.33	2.91	2.69	1.91	3.27	3.61	10.35	9.66	4.13	6.14	7.90	2.61	7.95
9	8.93	4.78	0.74	2.56	9.43	18.63	8.14	1.49	4.51	1.69	12.62	11.28	3.32	3.42	3.35
10	4.78	2.84	10.39	5.87	1.01	2.59	16.77	27.25	26.81	20.96	7.26	2.32	5.04	8.50	9.11

$$\bar{Y}_{.j} = \frac{1}{R}\sum_{r=1}^{R} Y_{rj}$$

Initialization Bias [Steady-State Simulations]

--Cumulative means

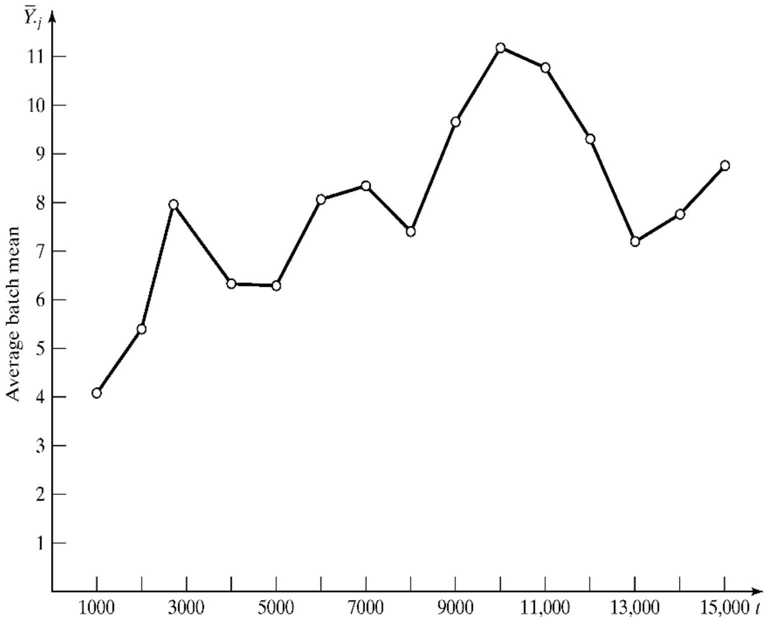
M/G/1 output: Ensemble batch means and cumulative means

Run Length T	Batch j	Average Batch Mean, $\bar{Y}_{\cdot j}$	Cumulative Average (No Deletion), $\bar{Y}_{\cdot\cdot}(j, 0)$	Cumulative Average (Delete 1), $\bar{Y}_{\cdot\cdot}(j, 1)$	Cumulative Average (Delete 2), $\bar{Y}_{\cdot\cdot}(j, 2)$
1,000	1	4.03	4.03	—	—
2,000	2	5.45	4.74	5.45	—
3,000	3	8.00	5.83	6.72	8.00
4,000	4	6.37	5.96	6.61	7.18
5,000	5	6.33	6.04	6.54	6.90
6,000	6	8.15	6.39	6.86	7.21
7,000	7	8.33	6.67	7.11	7.44
8,000	8	7.50	6.77	7.16	7.45
9,000	9	9.70	7.10	7.48	7.77
10,000	10	11.25	7.51	7.90	8.20
11,000	11	10.76	7.81	8.18	8.49
12,000	12	9.37	7.94	8.29	8.58
13,000	13	7.28	7.89	8.21	8.46
14,000	14	7.76	7.88	8.17	8.40
15,000	15	8.76	7.94	8.21	8.43

Averaged to (batch means 4.03 and 5.45 → 4.74; 4.03, 5.45 and 8.00 → 5.83)

Initialization Bias [Steady-State Simulations]

- A plot of the ensemble averages, $\bar{Y}_{\cdot j}$, versus *1000j*, for *j = 1,2, …,15*.



Illustrates the downward bias of the initial observations. As time becomes larger, the effect of the initial conditions on later observations lessens and the observation appear to vary around a common mean. When the analyst feels that this point has been reached, then the data collection phase starts.

It is apparent that downward bias is present and this bias can be reduced by deleting one or more observations.

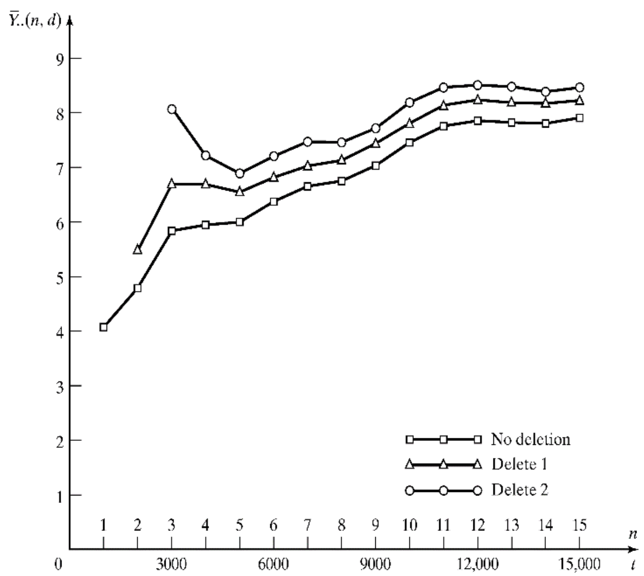
Initialization Bias [Steady-State Simulations]

--Cumulative means

- **After deleting *d* observations out of *n* observations, compute**

$$\bar{Y}_{..}(n,d) = \frac{1}{n-d}\sum_{j=d+1}^{n}\bar{Y}_{.j}$$

- **The results in the table (slide # 34) are for d = 0, 1, and 2.These cumulative averages with deletion, namely, are plotted for comparison purposes in the following figure:**



However, it is not convenient to use cumulative averages to determine the initialization phase.

Initialization Bias [Steady-State Simulations]

- **No widely accepted, objective and proven technique to guide how much data to delete to reduce initialization bias to a negligible level.**
- **Plots can, at times, be misleading but they are still recommended.**
  - Ensemble averages reveal a smoother and more precise trend as the # of replications, R, increases.
  - Ensemble averages can be smoothed further by plotting a moving average.
  - Cumulative average becomes less variable as more data are averaged.
  - The more correlation present, the longer it takes for $\bar{Y}_{.j}$ to approach steady state.
  - Different performance measures could approach steady state at different rates.

Replication Method [Steady-State Simulations]

- **Used to estimate a point-estimator variability and to construct a confidence interval.**
- **Approach: make R replications, initializing and deleting from each one the same way.**
- **Important to do a thorough job of investigating the initial-condition bias:**
  - Bias is not affected by the number of replications, instead, it is affected only by deleting more data (i.e., increasing T_0) or extending the length of each run (i.e. increasing T_E).
  - Increasing R could produce a better CI around the "wrong point".
- **Basic raw output data $\{Y_{rj}, r = 1, \ldots, R; j = 1, \ldots, n\}$ is derived by:**
  - Individual observation from within replication *r* (n observations).
  - Batch mean within replication *r* of some number of discrete-time observations.
  - Batch mean within replication *r* of a continuous-time process over time interval *j*.

Replication Method [Steady-State Simulations]

- **Each replication is regarded as a single sample for estimating θ. For replication r, define the sample mean of all (undeleted) observations within replication :**

$$\bar{Y}_{r.}(n,d) = \frac{1}{n-d}\sum_{j=d+1}^{n} Y_{rj}$$

  - n: total # of observations for each replication
  - d: # of deleted observations (initialization)

- **All replications use different random number streams, and initialized with the same initial conditions → $\bar{Y}_{1.}(n,d)$, $\bar{Y}_{2.}(n,d)$,…,$\bar{Y}_{R.}(n,d)$ are iid r.v. → a random sample from a population with unknown mean $\theta_{n,d}$**

- **The overall point estimator:**

$$\bar{Y}_{..}(n,d) = \frac{1}{R}\sum_{r=1}^{R}\bar{Y}_{r.}(n,d) = \frac{1}{R}\sum_{r=1}^{R}\left(\frac{1}{n-d}\sum_{j=d+1}^{n} Y_{rj}\right) \quad \text{and} \quad \mathrm{E}[\bar{Y}_{..}(n,d)] = \theta_{n,d}$$

- **If d and n are chosen sufficiently large:**
  - $\theta_{n,d} \sim \theta$.
  - $\bar{Y}_{..}(n,d)$ is an approximately unbiased estimator of θ.

Replication Method [Steady-State Simulations]

- **The sample variance and standard error of $\bar{Y}_{..}$:**

$$S^2 = \frac{1}{R-1}\sum_{r=1}^{R}(\bar{Y}_{r.} - \bar{Y}_{..})^2 = \frac{1}{R-1}\left(\sum_{r=1}^{R}\bar{Y}_{r.}^{\,2} - R\bar{Y}_{..}^{\,2}\right) \quad \text{and} \quad s.e.(\bar{Y}_{..}) = \frac{S}{\sqrt{R}}$$

- **Then a 100(1 - α)% CI for θ, based on the t distribution is:**

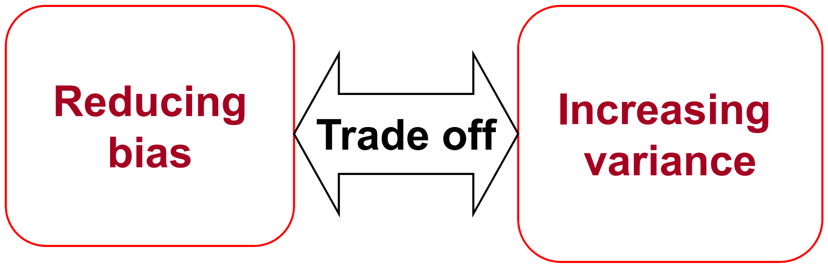
$$\bar{Y}_{..} - t_{\alpha/2,R-1} S/\sqrt{R} \le \theta \le \bar{Y}_{..} + t_{\alpha/2,R-1} S/\sqrt{R}$$

Replication Method [Steady-State Simulations]

- ***Rough rule***: **length of each replication (*n*) beyond deletion point (*d*):**

$$(n - d) > 10d$$

- ***Rough rule***: **number of replications (*R*) should be as many as time permits, up to about *25* replications.**
- **For a fixed total sample size (*R*), as fewer data are deleted (*d*):**
  - C.I. shifts: greater bias (more effect of initial biased data).
  - Standard error of $\bar{Y}_{..}(n,d)$ decreases: decrease variance.



Replication Method [Steady-State Simulations]

- **Example: suppose that the simulation analyst of a M/G/1 queuing model decides to make 10 replications, each of length 15000 minutes, each starting at time 0 in the empty and idle state, and each initialized for 2000 minutes before data collection begins. The output data on queue length are given on slide (# 33). The first two batch means are deleted.**

(i) Estimate by 95% CI, the long-run time-average queue length and comment on it.

(i) How many additional replications would be needed to estimate the long-run mean queue length, *Lq*, within ±2 with *90%* confidence? The replication averages are shown in the table (slide # 43).

Replication Method [Steady-State Simulations]

Data Summary for *M/G/1* Simulation by Replication

Replication, r	Sample Mean for Replication r (No Deletion) $\bar{Y}_{r\cdot}(15,0)$	(Delete 1) $\bar{Y}_{r\cdot}(15,1)$	(Delete 2) $\bar{Y}_{r\cdot}(15,2)$
1	3.27	3.24	3.25
2	16.25	17.20	17.83
3	15.19	15.72	15.43
4	7.24	7.28	7.71
5	2.93	2.98	3.11
6	4.56	4.82	4.91
7	8.44	8.96	9.45
8	5.06	5.32	5.27
9	6.33	6.14	6.24
10	10.10	10.48	11.07
$\bar{Y}_{\cdot\cdot}(15,d)$	7.94	8.21	8.43
$\sum_{r=1}^{R} \bar{Y}_{r\cdot}^2$	826.20	894.68	938.34
S^2	21.75	24.52	25.30
S	4.66	4.95	5.03
$S/\sqrt{10} = \text{s.e.}(\bar{Y}_{\cdot\cdot})$	1.47	1.57	1.59

$$\bar{Y}_{r\cdot}(n,d) = \frac{1}{n-d}\sum_{j=d+1}^{n} Y_{rj}$$

$$\bar{Y}_{\cdot\cdot}(n,d) = \frac{1}{R}\sum_{r=1}^{R} \bar{Y}_{r\cdot}(n,d)$$

May not be given in the exam

Needed for the example

Replication Method [Steady-State Simulations]

Solution.

(i)

- Here, $R = 10$, each of length $T_E = 15,000$ minutes, starting at time 0 in the empty and idle state, initialized for $T_0 = 2,000$ minutes before data collection begins.
- Each batch means is the average number of customers in queue for a *1,000*-minute interval.
- The 1st two batch means are deleted ($d = 2$).
- The point estimator and standard error are:

$$\bar{Y}_{..}(15,2) = 8.43 \qquad \text{and} \qquad s.e.\left(\bar{Y}_{..}(15,2)\right) = 1.59$$

- The *95%* C.I. for long-run mean queue length is:

$$\bar{Y}_{..} - t_{\alpha/2,R-1} S / \sqrt{R} \le \theta \le \bar{Y}_{..} + t_{\alpha/2,R-1} S / \sqrt{R}$$

$$8.43 - 2.26(1.59) \le L_Q \le 8.42 + 2.26(1.59)$$

- That is, the long-run mean queue length is between 4.84 and 12.02 (if d and n are "large" enough).

Replication number [Steady-State Simulations]

(ii)

- We know: $R_0 = 10, d = 2$ and $s_0^2 = 25.30$.
- To estimate the long-run mean queue length, Lq, within h $= 2$ customers with *90%* confidence ($\alpha = 10\%$).
- Initial estimate:

$$R \geq \left(\frac{z_{0.05} s_0}{h} \right)^2 = \frac{1.645^2 (25.30)}{2^2} = 17.1$$

- Hence, at least 18 replications are needed, next try $R = 18, 19, \ldots$ using $R \geq \left(t_{0.05,R-1} s_0 / h \right)^2$. We found that:

$$R = 19 \geq \left(t_{0.05,19-1} s_0 / h \right)^2 = (1.74 \times 25.3 / 2)^2 = 18.93$$

- Additional replications needed is $R - R_0 = 19\text{-}10 = 9$.

Interpretation of Confidence Intervals

- **Interval with random (data-dependent) endpoints that's supposed to have stated probability of containing, or covering, expected value**
  - "Target" expected value is a fixed, but unknown number
  - Expected value = average of infinite number of replications
- **Usual formulas assume normally-distributed data**
  - Never true in simulation
  - Might be approximately true if output is an average, rather than an extreme
  - Central limit theorem
  - Robustness, coverage, precision

Interpretation of Confidence Intervals [Comparing Two Scenarios]

- **Usually compare alternative system scenarios, configurations, layouts, sensitivity analysis**
  - lets compare two scenarios (for example: two queuing disciplines, two inventory control policies)
  - Make confidence intervals on expected outputs from each scenario, see if they overlap.
  - Reasonable but not-quite-right idea (given the meaning of CI, that depend on data)
  - Example-
    - Base case:
      22175.19 ± 369.54, or [21805.65, 22544.73]
    - More-resources case:
      24542.82 ± 329.11, or [24213.71, 24871.93]

    No overlap
  - No overlap → this means that the two scenarios are different.
  - If there was overlap → no significant difference
  - But this doesn't allow for a precise, efficient statistical conclusion. Hypothesis testing could be used (next slides).

Interpretation of Confidence Intervals [Comparing Two Scenarios]—Hypothesis testing (t-test)

- H_0: $\theta_1 - \theta_2 = \Delta_0$
- Normality assumption
- $\bar{X}_1$ and $\bar{X}_2$ are the sample means, and θ_1 and θ_2 are the population means
- **Case-1: populations have equal standard deviations.**

- **Finding the value of the test statistic requires two steps:**

Step 1: Pool (or combine) the sample standard deviations.

$$s_p^2 = \frac{(R_1 - 1)s_1^2 + (R_2 - 1)s_2^2}{R_1 + R_2 - 2}$$

Step 2: Determine the value of *t* from the following formula.

$$t = \frac{\bar{X}_1 - \bar{X}_2 - \Delta_0}{\sqrt{s_p^2\left(\frac{1}{R_1} + \frac{1}{R_2}\right)}}$$

has (R_1 + R_2 - 2) degrees of freedom

Interpretation of Confidence Intervals [Comparing Two Scenarios]—Hypothesis testing (t-test)

- H_0: $\theta_1 - \theta_2 = \Delta_0$
- Normality assumption
- **Case-2: populations have unequal standard deviations.**

- **Finding the value of the test statistic requires two steps:**

Step 1: Determine the value of degrees of freedom ν from following formula.

$$\nu = \frac{(s_1^2 / R_1 + s_2^2 / R_2)^2}{[(s_1^2 / R_1)^2 / (R_1 - 1)] + [(s_2^2 / R_2)^2 / (R_2 - 1)]}$$

Step 2: Determine the value of t from the following formula.

$$t = \frac{\bar{X}_1 - \bar{X}_2 - \Delta_0}{\sqrt{\left(\frac{s_1^2}{R_1} + \frac{s_2^2}{R_2} \right)}}$$

has ν degrees of freedom

Interpretation of Confidence Intervals [Comparing Two Scenarios] – Confidence Interval on the difference in the means

- A CI on the difference between θ_1-θ_2 can be:

$$(\overline{Y_{.1}} - \overline{Y_{.2}}) \pm t_{\frac{\alpha}{2},\upsilon} \times s.e(\overline{Y_{.1}} - \overline{Y_{.2}})$$

υ: degrees of freedom associated with the variance estimator,

$t_{\frac{\alpha}{2},\upsilon}$: 100(1-α/2) percentage of a t-distribution with υ degrees of freedom

$s.e(\overline{Y_{.1}} - \overline{Y_{.2}})$: standard error of the specified point estimator.

$\overline{Y_{.1}}$: sample mean (cross replications) of system 1

$\overline{Y_{.2}}$: sample mean (cross replications) of system 2

Interpretation of Confidence Intervals [Comparing Two Scenarios] – Confidence Interval on the difference in the means

- **Case-2: populations have unequal standard deviations. (we will not discuss case 1 with equal standard deviations: because it is rarely to be known to be true in practice))**

Step 1: Determine the value of degrees of freedom ν from following formula (rounded up).

$$\nu = \frac{(s_1^2/R_1 + s_2^2/R_2)^2}{[(s_1^2/R_1)^2/(R_1-1)] + [(s_2^2/R_2)^2/(R_2-1)]}$$

Step 2: Determine the value of $t_{\frac{\alpha}{2},\nu}$ with ν degrees of freedom

use $s.e.(\bar{Y}_{.1} - \bar{Y}_{.2}) = \sqrt{\frac{s_1^2}{R_1} + \frac{s_2^2}{R_2}}$ to find the CI

$$(\overline{Y}_{.1} - \overline{Y}_{.2}) \pm t_{\frac{\alpha}{2},\nu} \times s.e(\overline{Y}_{.1} - \overline{Y}_{.2})$$

Compare Means via ARENA Output Analyzer

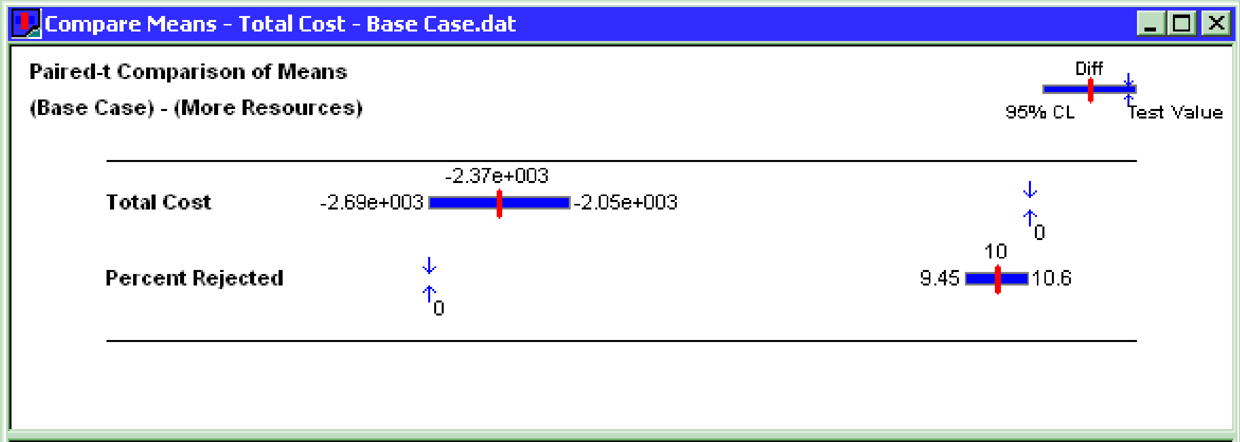
- **Output Analyzer is a separate application that operates on .dat files produced by Arena (by the statistics module)**
  - Launch separately from Windows, not from Arena
- **To save output values (Expressions) of entries in Statistic data module (Type = Output) – enter filename.dat in Output File column**
  - Did for both Total Cost and Percent Rejected
  - Will overwrite these file names next time
    - Either change names in Arena model, or out in operating system before next run
  - .dat files are binary … can only be read by Output Analyzer

Compare Means via Output Analyzer
(cont'd.)

- **Start Output Analyzer, open a new data group**
  - Basically, a list of .dat files of current interest
  - Can save data group for later use – .dgr file extension
  - Add button to select (Open) .dat files for data group
- ***Analyze > Compare Means* menu option**
  - Add data files … "A" and "B" for two scenarios
  - Select "Lumped" for Replications field
  - Title, confidence level, accept Paired-t Test (more general and to be used unless we made action to make the two scenarios statistically independent), do not Scale Display since two output performance measures have different units

Compare Means via Output Analyzer
(cont'd.)

- **Results:**



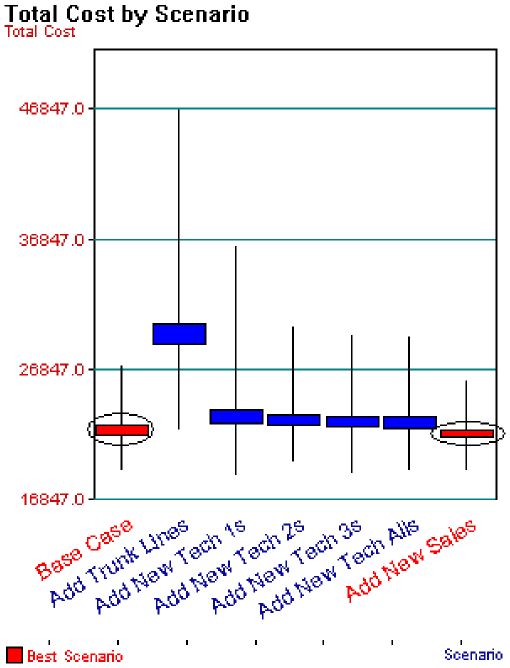
IDENTIFIER	ESTD. MEAN DIFFERENCE	STANDARD DEVIATION	0.950 C.I. HALF-WIDTH	MINIMUM VALUE	MAXIMUM VALUE	NUMBER OF OBS
Total Cost	-2.37e+003	1.7e+003	321	1.9e+004	2.71e+004	110
				2.16e+004	3.15e+004	110
REJECT H0 => MEANS ARE NOT EQUAL AT 0.05 LEVEL						
Percent Rejected	10	2.96	0.559	5.52	18.8	110
				0	9.31	110
REJECT H0 => MEANS ARE NOT EQUAL AT 0.05 LEVEL						

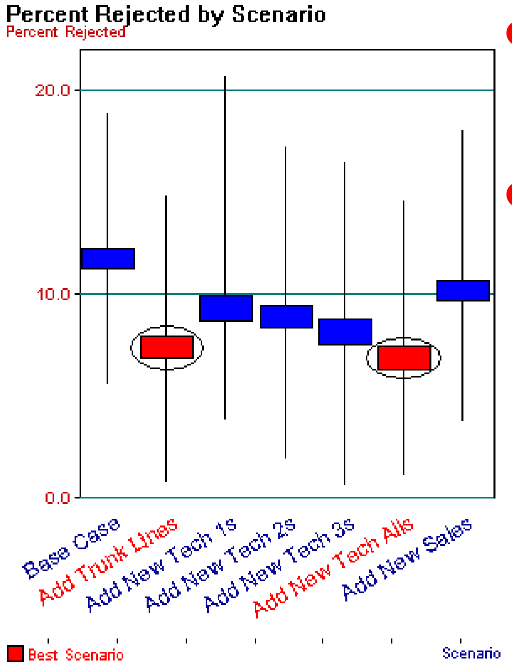
- **Confidence intervals on differences both miss 0**
  - Conclude that there *is* a (statistically) significant difference on both output performance measures

Evaluating Many Scenarios with Process Analyzer (PAN)

- **With (many) more than two scenarios to compare, two problems are**
  - Simple mechanics of making many parameter changes, making many runs, keeping track of many output files
  - Statistical methods for drawing reliable, useful conclusions
- **Process Analyzer (PAN) addresses these**
- **PAN operates on ARENA program (.p) files – produced when .doe file (ARENA model) is run (or just checked:** *Run>Check Model***)**
- **Start PAN from Arena (*Tools* > *Process Analyzer*) or via Windows**
- **PAN runs on its own, separate from Arena**

Statistical Comparisons with PAN (cont'd.)





- **Vertical boxes: 95% confidence intervals**
- **Red scenarios statistically significantly better than blues**
  - More precisely, red scenarios are 95% sure to contain best one
  - Narrow down red set – more replications, or Error Tolerance > 0
  - More details in text

Numerical values (including c.i. half widths) in chart – right click on chart, Chart Options, Data

So which scenario is "best"? Criteria disagree. Combine them somehow?

See Tutorial-4 (Lecture-7)

Continued in Lecture 8