Verification & Validation

Motivation

- The simulation model needs to work as intended.
- It should adequately represent the system being simulated.

Verification & Validation

- Verification –
- Validation –

Verification

- Possible methods
  -
  - …
- Use all of the prior methods in extreme cases
  - Examples
    - Examine only one part type at a time.
    - Allow only one entity into the system.

Verification

- Use all of the prior methods in extreme cases
  - Examples
    - Control the distributions used.
      - Set all processing times to constants – You should be able to predict quantities such as the entity avg. TIS.
      - Control arrival rates
        - Extremely high.
        - Extremely low.

Verification

- Need to run very large models for a long simulated time to check for gradual buildup of queues.
  - May be able to compute utilizations.
- During model development
  -

Validation

- In general this is more difficult than verification.
- Two general cases
  -

Validation

- System exists.
  - Compare results to historical results.
    - Does the data exist?
    - Are you comparing "apples to apples"?
  - Control model variability
    - Resource availability
      - If a known long maintenance period occurred for machines or if known long failures/repairs occurred, duplicate these in the model.
      - Duplicate resource schedules.
    - Duplicate the arrival pattern over which historical performance data has been collected.
      - Job types.
      - Job arrival times.

Validation

- System does not exist.
  - Change the data and modify the model so that it simulates a system that still exists. Then compare to historical data.
  - Expert opinion.
  - Run extreme cases.

Output Analysis

Introduction

- A simulation model has been constructed of a system to generate estimates of one or more performance measures.
- Random inputs/system components in the simulation imply that the outputs (performance measure estimates) will be observations from probability distributions.

Introduction

- Questions about starting and running the model.
  -

Types of Simulation "Runs"

- In general, there are two types of simulation analyses performed that dictate how the previous questions should be answered.

1. Terminating –
2. Steady state –

Example

- Simulation of a bank from open to close (9AM-5PM).
  - This is an example of a terminating simulation.
    - How do you start the simulation?
      - Empty and idle.
    - How long should the model be run (how much simulated time before stopping the run)?
      - 9AM until all customers depart after 5PM.

Terminating Simulations

- Time period of interest is defined.
-

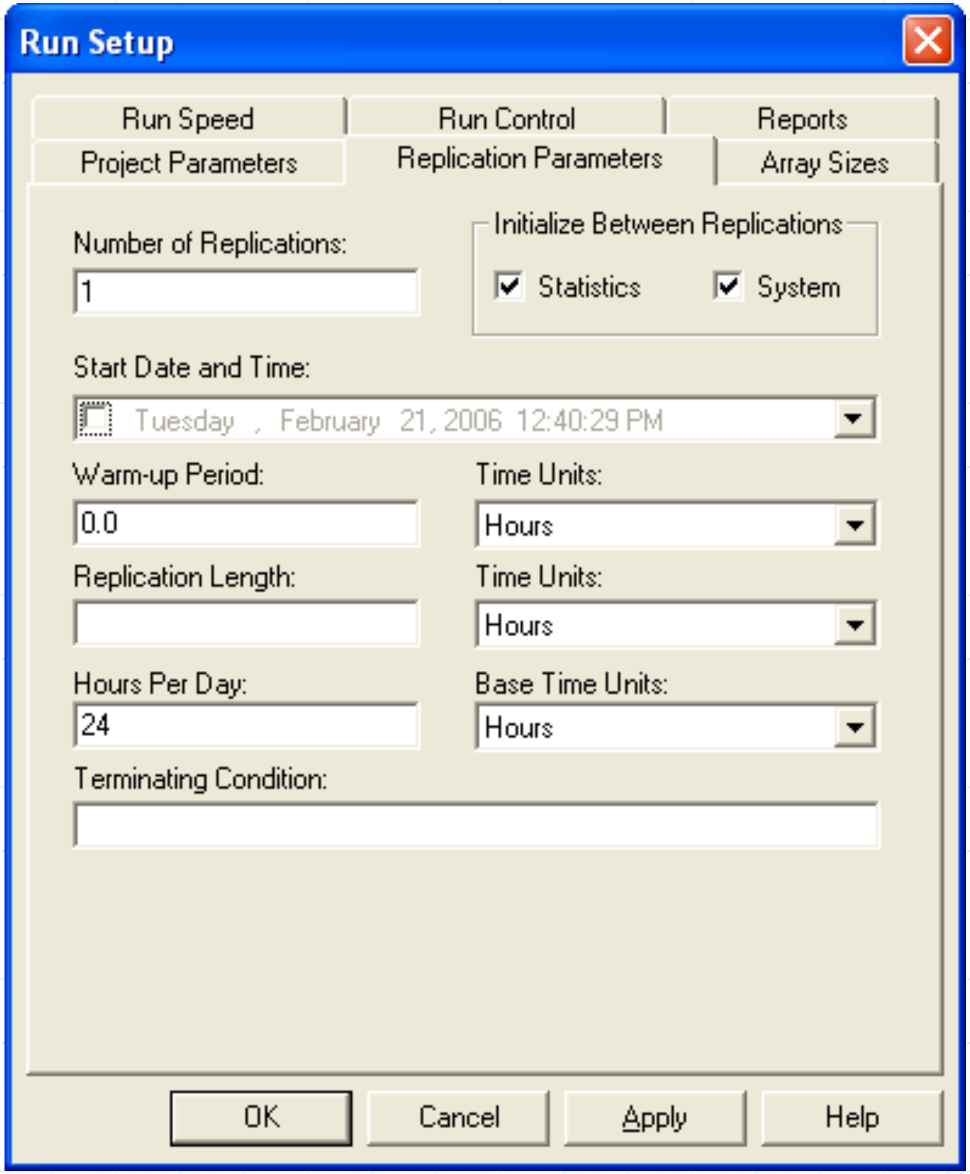
Example

- Bank simulation from 9AM-5PM.
  - Customers arriving before 5PM are all served.
  - The simulation ending time may vary from replication to replication.
  - What is the termination criteria?

Arena

- To properly end such a simulation, use
  - Run -> Setup -> Replication Parameters
    1. Set ending criteria in “Terminating Condition”.
    2. Leave “Replication Length” field blank.
- The termination criteria will use Arena “syntax”.

Arena


Run Setup
Run Speed
Run Control
Reports
Project Parameters
Replication Parameters
Array Sizes
Number of Replications:
1
Initialize Between Replications
Statistics
System
Start Date and Time:
Tuesday , February 21, 2006 12:40:29 PM
Warm-up Period:
0.0
Time Units:
Hours
Replication Length:
Time Units:
Hours
Hours Per Day:
24
Base Time Units:
Hours
Terminating Condition:
OK
Cancel
Apply
Help

Terminating Simulations

- A simulation of a service facility with “rush hour” periods is constructed. There is a specific time period of interest – 11AM -1PM. The facility is open at 9AM.
  - How long to simulate? 11AM-1PM.
  -

Terminating Simulations

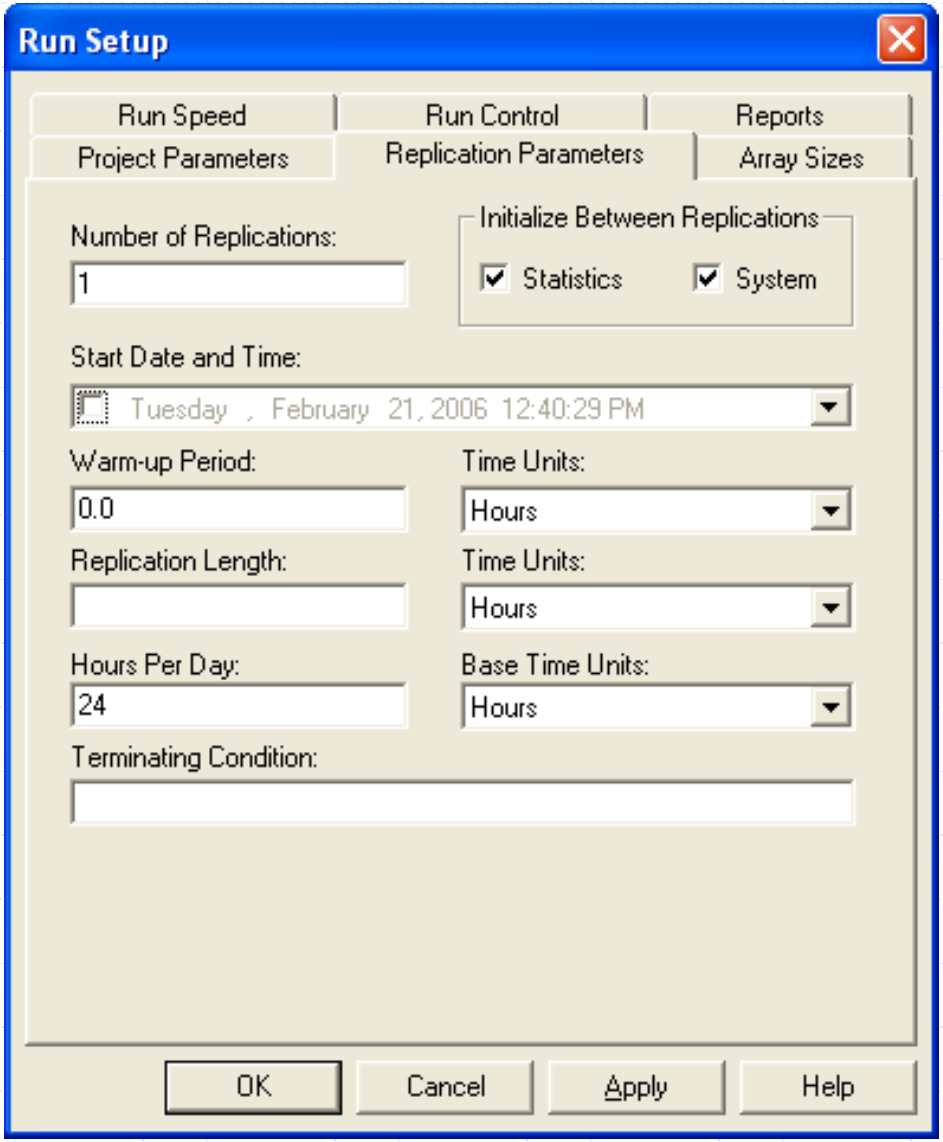
- Approaches for starting a simulation with unknown initial conditions.
  1. Collect data on the system state at the start of the rush hour period.
     - Initialize the simulation with the “average” system state.
     - Initialize the simulation with a random system state based on the collected data.
       - Both hard to do in Arena,
       - Straightforward in an “event driven” model (e.g., the manual simulation).

Terminating Simulations

- Approaches for starting a simulation with unknown initial conditions.

2.

Arena


Run Setup
Run Speed
Run Control
Reports
Project Parameters
Replication Parameters
Array Sizes
Number of Replications:
1
Initialize Between Replications
Statistics
System
Start Date and Time:
Tuesday , February 21, 2006 12:40:29 PM
Warm-up Period:
0.0
Time Units:
Hours
Replication Length:
Time Units:
Hours
Hours Per Day:
24
Base Time Units:
Hours
Terminating Condition:
OK
Cancel
Apply
Help

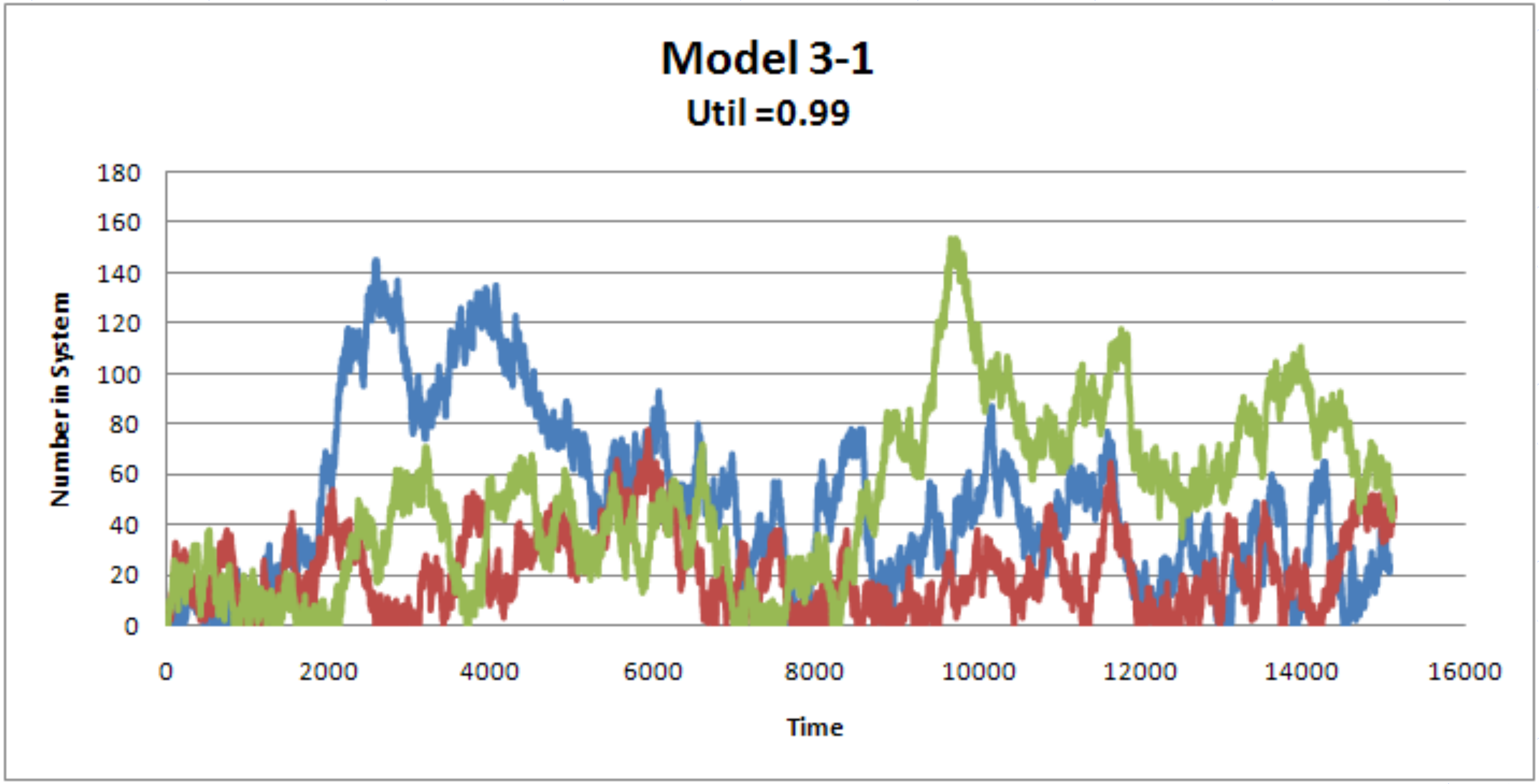
Steady State Simulations

- Steady state simulations are used to understand how a system performs after being in operations for a long time. The system has reached "steady state" where the performance is independent of the initial conditions.
- Examples
  - Production line simulations – the line starts where it left off at the end of prior shifts.
  - Emergency rooms.
  - Worst case analysis.

Steady State Simulations

- How to start the simulation?
  - Typically a warm-up period is used to minimize any impact of initial conditions.
  - How long should the warm-up period be?
    - Can determine from some sample runs of the simulation.
- How long to run the simulation?
  - Can determine from some sample runs of the simulation.

Steady State Simulations



Steady State Simulations

- What are you looking for in the plots?
  - Impact of initial conditions.
    - Does the plot from time zero look different than other portions of the graph?
    - Does the performance measure seem to be growing without bound?

The Number of Replications and Analysis of Output

- We will focus on the analysis of terminating simulations.
  - They are more common in practice.
  - The analysis is more straightforward.
- Simulation models are used for experimentation.
  - One simulation replication $\rightarrow$ a single realization of each system performance measure.
  - n independent replications $\rightarrow$ n independent samples from the same distribution.

The Number of Replications and Analysis of Output

- Consider a single performance measure. Let X_i be the random variable that represents the value of the performance measure for the ith simulation replication.
  - x_i = outcome/realization of X_i from the ith simulation replication.
- Since the X_i are independent and identically distributed random variables the performance can be characterized using the "typical" confidence interval.

Analysis of Output

- The approximate $(1-\alpha)*100\%$ confidence interval

- Assumes the observations are from a normal distribution.

The Number of Replications

- How to estimate number of independent replications required for a desired precision expressed as a confidence interval half-width.
- The half-width h of this confidence interval is
- This cannot be used to precisely calculate n to get the desired precision h since the t-value is a function of n.

The Number of Replications

- Substitute

- Use this formula to approximate the number of replications needed to get a desired half-width (precision) for some performance measure.

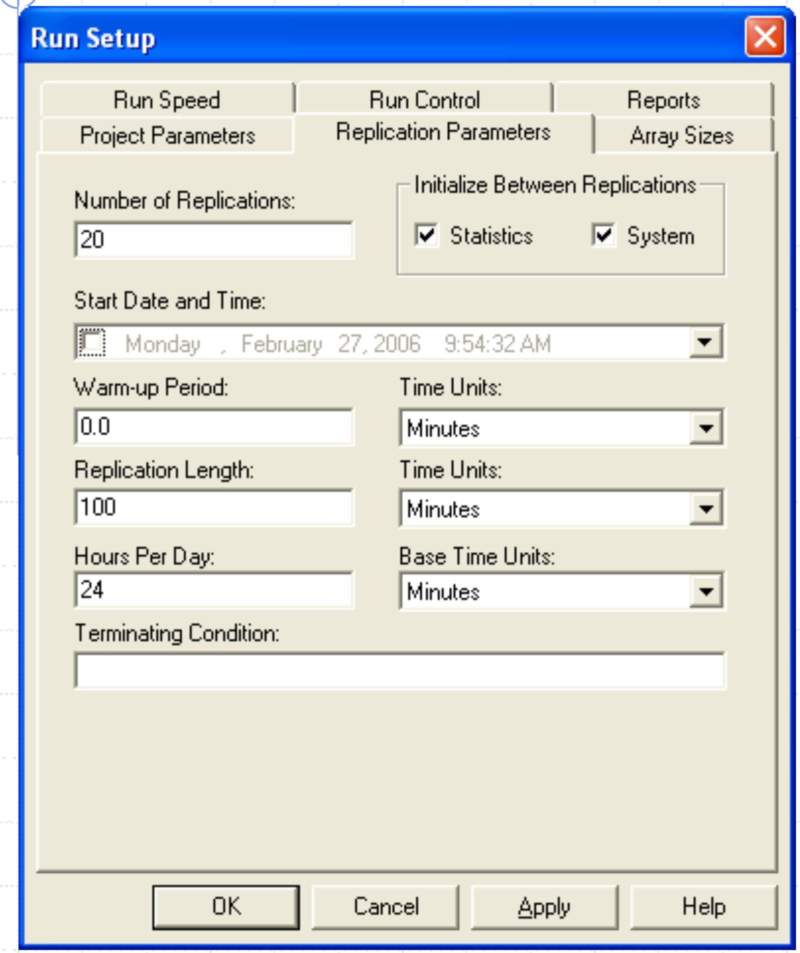
The Number of Replications

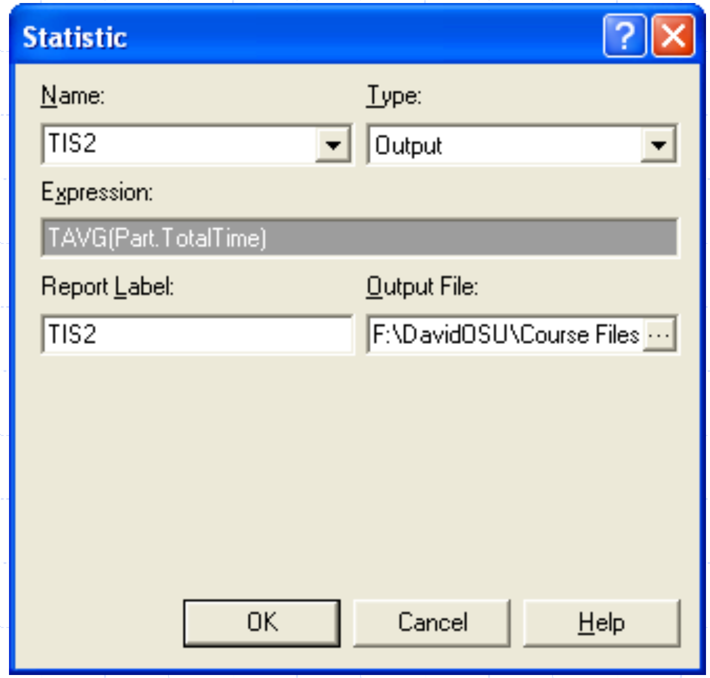
- Example – Average Time In System (TIS) for entities being processed.

The Number of Replications

- Use Arena to send TIS average results from independent replications to a text file.
- Use this data to estimate n – the number of independent replications needed for a desired precision.

The Number of Replications





The Number of Replications

Project:	Exercise 3.1	
User:	Kelton	
Data item:	TIS2	
Run date:	2/27/2006	
Options:	YDT	20

Time	Observation
1	8.341297
-1	0
2	7.39797
-2	0
3	5.40304
-3	0
4	9.393243
-4	0
5	3.859661
-5	0
6	14.76574
-6	0
7	26.15339
-7	0

The Number of Replications

$z_{1-\alpha/2}$ for $t_{n-1,1-\alpha/2}$

$$\Rightarrow n \approx z_{1-\alpha/2}^{2} * \frac{s^{2}(AvgTIS)}{h^{2}}$$

Avg.	8.187022
Stdev	5.428285

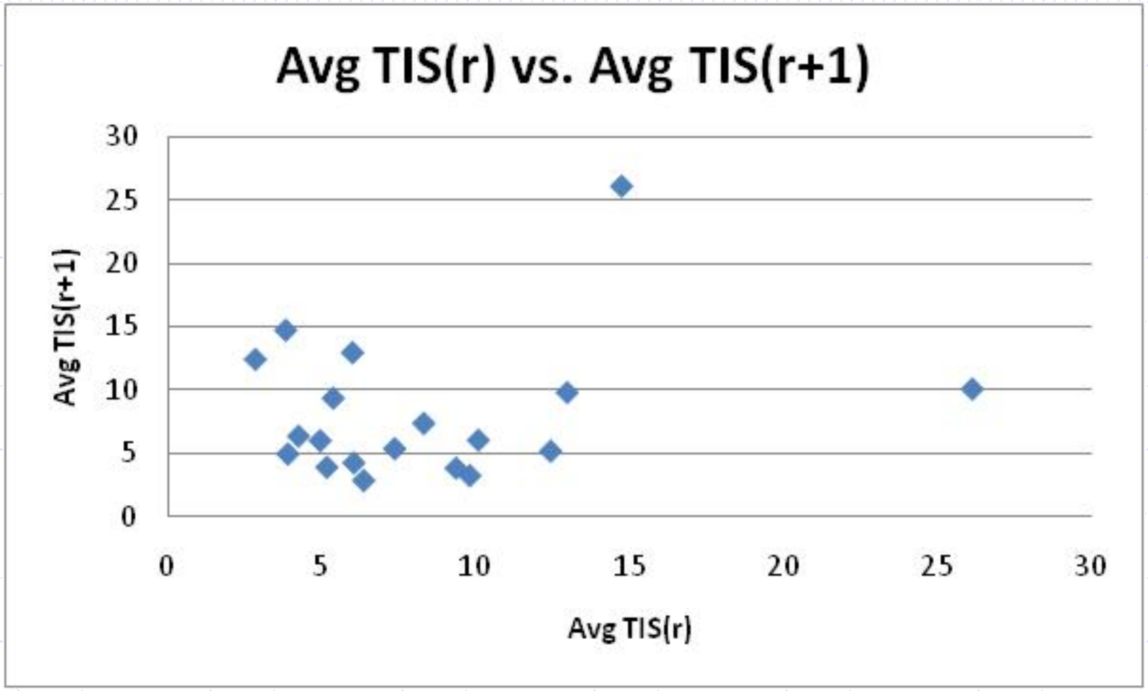
$$z_{1-0.025} = 1.96$$

$$n \approx 1.96^{2} \frac{5.43^{2}}{.5^{2}} = 453$$

The Number of Replications

Project: Exercise 3.1
User: Kelton
Data item: TIS2
Run date: 2/27/2006
Options: YDT

Time	Observation
1	8.341296692
2	7.397970199
3	5.403039681
4	9.393242671
5	3.859660527
6	14.7657369
7	26.15339254
8	10.11390026
9	6.071090565
10	4.278436904
11	6.390677328
12	2.874782431
13	12.45966732
14	5.197794785
15	3.92990998
16	4.978496969
17	6.029672064
18	12.99574679
19	9.837051072
20	3.268866729
Avg.	8.18702162
Stdev	5.428284629



Comparing Two Alternatives

- In many situations where simulation is used, changes to an existing or "base" system are explored.
- There is a need to do a comparison of two systems.
- Simulation has been used to generate estimates of performance measure for the base system and a changed system.
- Analysis
  - Two sample t-confidence interval.
  - Paired t-confidence interval.

Comparing Two Alternatives

Let

$X_{i1}, X_{i2}, \ldots, X_{in_i}$ be a sample of n_i independent and identically distributed observations of some system performance measure (e.g., X_{i1} is the average TIS from replication 1 of the simulation model) for $i = 1,2$.

$\mu_i = E(X_{ij})$

$\delta = \mu_1 - \mu_2$

The independence of X_{1j} and X_{2j} depends on how the simulation experiments were conducted (over which there is control).

Comparing Two Alternatives

- "Standard" two-sample t-confidence interval.
- Assumptions
  - X_{1j}'s and X_{2j}'s are independent.
  - Var(X_{1j}) = Var(X_{2j}) .
  - n_1 and n_2 need not be equal.
- Not recommended for simulation output analysis since the equal variance assumption is often not met (Kelton and Law 2000).
  - However, the test is robust if $n_1 = n_2$.

Comparing Two Alternatives

- "Welch" two-sample t-confidence interval.
- Assumptions
  - X_{1j}'s and X_{2j}'s are independent and normally distributed.
  - n_1 and n_2 need not be equal.

Comparing Two Alternatives

Compute sample means and sample variances for the output of the two simulations,

$$\bar{X}_1(n_1), \bar{X}_2(n_2), s_1^2(n_1), s_2^2(n_2).$$

Compute the estimated degrees of freedom $\hat{f}$

$$\hat{f} = \frac{[s_1^2(n_1)/n_1 + s_2^2(n_2)/n_2]^2}{[s_1^2(n_1)/n_1]^2/(n_1-1) + [s_2^2(n_2)/n_2]^2/(n_2-1)}$$

Form the confidence interval as

$$\bar{X}_1(n_1) - \bar{X}_2(n_2) \pm t_{\hat{f},1-\alpha/2}\sqrt{s_1^2(n_1)/n_1 + s_2^2(n_2)/n_2}$$

Comparing Two Alternatives

- Paired t-confidence interval.
- Assumptions
  - n_1 and n_2 are equal.

Comparing Two Alternatives

Let

$$D_j = (X_{1j} - X_{2j})$$

$$\overline{D}(n) = \frac{\sum_{j=1}^{n} D_j}{n}, \quad s^2(D_j) = \text{sample variance.}$$

The confidence interval is :

$$\overline{D}(n) \pm t_{n-1,1-\alpha/2}\sqrt{\frac{s^2(D_j)}{n}}$$

Comparing Two Alternatives

- The paired t-confidence interval forms a new sample as the difference between corresponding outputs from the same replication number for system 1 and system 2.
  - X_{1i} and X_{2i} may be dependent.
  - The D_i's must be independent.
  - Var(X_{1j}) $=$ Var(X_{2j}) is not necessary.

Example

- Experiment with the single server model arrival process.
  - $t_{9,.975} = 2.26$, Performance measure is avg. TIS.

Time	Expo(5)	U(1,9)	D
1	8.341	8.328	0.013
2	7.398	6.228	1.170
3	5.403	5.943	-0.540
4	9.393	7.924	1.469
5	3.860	2.475	1.385
6	14.766	8.086	6.679
7	26.153	12.673	13.480
8	10.114	15.218	-5.104
9	6.071	5.893	0.178
10	4.278	4.161	0.117
		Avg.	1.885
		Stdev.	4.974

In-class Exercise

- For the following data.
  - Form 95% CIs for system 1 and system 2 output separately. Check for overlap.
  - Form a 95% paired t-confidence interval. Check for the inclusion of zero.
  - Construct a Welch two-sample t-confidence interval.

Let

$$D_j = (X_{1j} - X_{2j})$$

$$\bar{D}(n) = \frac{\sum_{j=1}^{n} D_j}{n}, \quad s^2(D_j) = \text{sample variance.}$$

The confidence interval is:

$$\bar{D}(n) \pm t_{n-1,1-\alpha/2}\sqrt{\frac{s^2(D_j)}{n}}$$

$$\hat{f} = \frac{[s_1^2(n_1)/n_1 + s_2^2(n_2)/n_2]^2}{[s_1^2(n_1)/n_1]^2/(n_1-1) + [s_2^2(n_2)/n_2]^2/(n_2-1)}$$

Form the confidence interval as

$$\bar{X}_1(n_1) - \bar{X}_2(n_2) \pm t_{\hat{f},1-\alpha/2}\sqrt{s_1^2(n_1)/n_1 + s_2^2(n_2)/n_2}$$

Data Pt.	System 1 Output	System 2 Output	Difference Sys 1- Sys 2
1	0.66	3.68	-3.02
2	0.53	3.93	-3.40
3	3.42	1.37	2.05
4	4.79	5.89	-1.10
5	1.11	5.30	-4.20
6	4.16	1.24	2.92
7	0.85	3.39	-2.54
8	3.22	1.58	1.64
9	4.66	4.48	0.18
10	1.08	5.00	-3.92
11	2.97	1.68	1.28
12	2.97	1.47	1.50
13	2.82	2.58	0.23
14	0.57	5.97	-5.40
15	2.48	4.76	-2.29
16	4.03	1.71	2.33
17	4.98	5.50	-0.52
18	3.30	3.36	-0.06
19	0.53	2.87	-2.34
20	1.12	3.65	-2.53
21	2.01	3.80	-1.79
22	4.63	1.83	2.80
23	4.52	5.03	-0.51
24	0.70	2.84	-2.14
25	1.48	1.49	-0.01
26	0.21	1.18	-0.97
27	2.07	4.93	-2.86
28	4.89	1.86	3.03
29	3.12	1.59	1.52
30	1.92	1.63	0.29
Average =	2.53	3.19	-0.66
Std Dev =	1.58	1.59	2.32
		t(29,0.975) =	2.05

T(58,.975) = 2.0

In-class Exercise

Comparing >2 Systems

- Comparisons with a base system.
- All pairwise comparisons.
- Simple procedure is to apply what is called the Bonferroni inequality.

$$\text{Prob}(\text{True value} \in CI_j, \text{ for all } j = 1,2,\ldots,k) \geq 1 - \sum_{j=1}^{k} \alpha_j$$

Comparing >2 Systems

- Example – One base system and four alternatives => four confidence intervals when comparing each system with the base system.
- To get 90% overall confidence level each individual confidence interval should be a 97.5% CI.

$$\alpha_i = 1 - \alpha / c$$

Comparing >2 Systems

- Experimental design methods are helpful for exploring a "factor space".
- Care must be taken when conducting the experiments and analyzing experimental results due to the non-random nature of generating "random" values in a simulation.