

## Common Mistakes (cont)

- No analysis
- Erroneous analysis
- No sensitivity analysis
  - Measure the effect on changing a parameter
- Ignoring errors in input
- Improper treatment of outliers
  - Some should be ignored (can't happen)
  - Some should be retained (key special cases)
- Assuming no change in the future
- Ignoring variability
  - Mean is of low significance in the face of high variability
- Too complex analysis
  - Complex models are “interesting” and so get published and studied
  - Real world use is simpler
  - Decision makers prefer simpler models

## Common Mistakes (cont)

- Improper presentation of results
  - The proper metric for analyst performance is number of analyses that helped decision makers.
- Ignoring social aspects
- Omitting assumptions and limitations

# **A Systematic Approach**

1. State goals and define the system
2. List services and outcomes
3. Select metrics
4. List parameters
5. Select factors to study
6. Select evaluation technique
7. Select workload
8. Design experiments
9. Analyze and interpret data
10. Present results

# Selecting a Technique

Choices: Analytical Modeling, Simulation, Measurement

“Until validated, all evaluation results are suspect.”

Validate one of these approaches by comparing against another.

Measurement results are just as susceptible to experimental errors and biases as the other two techniques.

## Criteria

- Stage of analysis
- Time required
- Tools
- Accuracy
- Trade-off evaluation
- Cost
- Saleability

# Performance

Throughput: The rate (requests per unit of time) at which requests can be serviced by the system.

- Throughput generally increases as the load initially increases.
- Eventually it stops increasing, and might then decrease.
- Nominal capacity is maximum achievable throughput under ideal workload conditions.
- Usable capacity is maximum throughput achievable without violating a limit on response time.
- Efficiency is ratio of usable to nominal capacity.

Example: Hashing systems. Knee capacity.

Utilization: Fraction of time the resource is busy servicing requests.

Bottleneck: Component with highest utilization. Improving this component often gives highest payoff.

# Workloads

A workload is the requests made by users of the system under study.

- A test workload is any workload used in performance studies
- A real workload is one observed on a real system. It cannot be repeated.
- A synthetic workload is a reproduction of a real workload to be applied to the tested system

Examples (for CPU performance)

- Addition instruction
- Instruction mixes
- Kernels
- Synthetic programs
- Application benchmarks

What if you are an HCI researcher?