Chapter 12
Comparison and Evaluation of Alternative System Designs

Source: Banks, Carson, Nelson & Nicol

Discrete-Event System Simulation

Purpose

- Purpose: comparison of alternative system designs.
- Approach: discuss a few of many statistical methods that can be used to compare two or more system designs.
- Statistical analysis is needed to discover whether observed differences are due to:
  - Differences in design or,
  - The random fluctuation inherent in the models.

Outline

- For two-system comparisons:
  - Independent sampling.
  - Correlated sampling (common random numbers).
- For multiple system comparisons:
  - Bonferroni approach: confidence-interval estimation, screening, and selecting the best.
- Metamodels

Comparison of Two System Designs

- Goal: compare two possible configurations of a system
  - e.g., two possible ordering policies in a supply-chain system, two possible scheduling rules in a job shop.
- Approach: the method of replications is used to analyze the output data.
- The mean performance measure for system *i* is denoted by θ_i ($i = 1,2$).
- To obtain point and interval estimates for the difference in mean performance, namely $\theta_1 - \theta_2$.

Comparison of Two System Designs

- Vehicle-safety inspection example:
  - The station performs *3* jobs: (1) brake check, (2) headlight check, and (3) steering check.
  - Vehicles arrival: Possion with rate = *9.5*/hour.
  - Present system:
    - Three stalls in parallel (one attendant makes all 3 inspections at each stall).
    - Service times for the *3* jobs: normally distributed with means *6.5, 6.0* and *5.5* minutes, respectively.
  - Alternative system:
    - Each attendant specializes in a single task, each vehicle will pass through three work stations in series
    - Mean service times for each job decreases by *10%* (*5.85, 5.4*, and *4.95* minutes).
  - Performance measure: mean response time per vehicle (total time from vehicle arrival to its departure).

Comparison of Two System Designs

- From replication r of system i, the simulation analyst obtains an estimate Y_{ir} of the mean performance measure θ_i .
- Assuming that the estimators Y_{ir} are (at least approximately) unbiased:

  $\theta_1 = E(Y_{1r}),\ r = 1, \ldots, R_1; \qquad \theta_2 = E(Y_{2r}),\ r = 1, \ldots, R_2$
- Goal: compute a confidence interval for $\theta_1 - \theta_2$ to compare the two system designs
- Confidence interval for $\theta_1 - \theta_2$ (c.i.):
  - If c.i. is totally to the left of *0*, strong evidence for the hypothesis that $\theta_1 - \theta_2 < 0$ ($\theta_1 < \theta_2$).
  - If c.i. is totally to the right of *0*, strong evidence for the hypothesis that $\theta_1 - \theta_2 > 0$ ($\theta_1 > \theta_2$).
  - If c.i. is totally contains *0*, no strong statistical evidence that one system is better than the other
    - If enough additional data were collected (i.e., R_i increased), the c.i. would most likely shift, and definitely shrink in length, until conclusion of $\theta_1 < \theta_2$ or $\theta_1 > \theta_2$ would be drawn.

Comparison of Two System Designs

- In this chapter:
  - A two-sided 100(1-α)% c.i. for $\theta_1 - \theta_2$ always takes the form of:

$$\bar{Y}_{.1} - \bar{Y}_{.2} \pm t_{\alpha/2,\upsilon} s.e.(\bar{Y}_{.1} - \bar{Y}_{.2})$$

where $\bar{Y}_{.i}$ is the sample mean performance measure for system i over all replications, and υ is the degress of freedom,

  - 3 techniques discussed assume that the basic data Y_{ir} are approximately normally distributed.

Comparison of Two System Designs

- Statistically significant versus practically significant
  - Statistical significance: is the observed difference $\bar{Y}_{.1} - \bar{Y}_{.2}$ larger than the variability in $\bar{Y}_{.1} - \bar{Y}_{.2}$?
  - Practical significance: is the true difference $\theta_1 - \theta_2$ large enough to matter for the decision we need to make?
  - Confidence intervals do not answer the question of practical significance directly, instead, they bound the true difference within a range.

Independent Sampling with Equal Variances
[Comparison of 2 systems]

- Different and independent random number streams are used to simulate the two systems
  - All observations of simulated system *1* are statistically independent of all the observations of simulated system *2*.
- The variance of the sample mean, $\bar{Y}_{.i}$, is:

$$V\left(\bar{Y}_{.i}\right) = \frac{V\left(Y_{.i}\right)}{R_i} = \frac{\sigma_i^2}{R_i}, \qquad i = 1,2$$

- For independent samples:

$$V\left(\bar{Y}_{.1} - \bar{Y}_{.2}\right) = V\left(\bar{Y}_{.1}\right) + V\left(\bar{Y}_{.2}\right) = \frac{\sigma_1^2}{R_1} + \frac{\sigma_2^2}{R_2}$$

Independent Sampling with Equal Variances

[Comparison of 2 systems]

- If it is reasonable to assume that $\sigma^2_1 = \sigma^2_2$ (approximately) or if $R_1 = R_2$, a two-sample-t confidence-interval approach can be used:
  - The point estimate of the mean performance difference is:

$$\bar{Y}_{.1} - \bar{Y}_{.2}$$

  - The sample variance for system i is:

$$S_i^2 = \frac{1}{R_i - 1}\sum_{r=1}^{R_i}\left(\bar{Y}_{ri} - \bar{Y}_{.i}\right)^2 = \frac{1}{R_i - 1}\sum_{r=1}^{R_i}\bar{Y}_{ri}^{\,2} - R_i\bar{Y}_{.i}^{\,2}$$

  - The pooled estimate of σ^2 is:

$$S_p^2 = \frac{(R_1 - 1)S_1^2 + (R_2 - 1)S_2^2}{R_1 + R_2 - 2}, \qquad \text{where } \upsilon = R_1 + R_2 \text{ -2 degrees of freedom}$$

  - C.I. is given by: $\bar{Y}_{.1} - \bar{Y}_{.2} \pm t_{\alpha/2,\upsilon}\, s.e.(\bar{Y}_{.1} - \bar{Y}_{.2})$

  - Standard error: $s.e.\left(\bar{Y}_{.1} - \bar{Y}_{.2}\right) = S_p\sqrt{\frac{1}{R_1} + \frac{1}{R_2}}$

Independent Sampling with Unequal Variances

[Comparison of 2 systems]

- If the assumption of equal variances cannot safely be made, an approximate *100(1-α)%* c.i. for can be computed as:

$$s.e.\left(\overline{Y}_{.1} - \overline{Y}_{.2}\right) = \sqrt{\frac{S_1^2}{R_1} + \frac{S_2^2}{R_2}}$$

  - With degrees of freedom:

$$\upsilon = \frac{\left(S_1^2 / R_1 + S_2^2 / R_2\right)^2}{\left[\left(S_1^2 / R_1\right)^2 / \left(R_1 - 1\right)\right] + \left[\left(S_2^2 / R_2\right)^2 / \left(R_2 - 1\right)\right]}, \quad \text{round to an interger}$$

  - Minimum number of replications $R_1 > 7$ and $R_2 > 7$ is recommended.

Common Random Numbers (CRN)

[Comparison of 2 systems]

- For each replication, the same random numbers are used to simulate both systems.
  - For each replication *r*, the two estimates, Y_{r1} and Y_{r2}, are correlated.
  - However, independent streams of random numbers are used on different replications, so the pairs (Y_{r1}, Y_{s2}) are mutually independent.
- Purpose: induce positive correlation between $\bar{Y}_{.1}, \bar{Y}_{.2}$ (for each *r*) to reduce variance in the point estimator of $\bar{Y}_{.1} - \bar{Y}_{.2}$

$$V\left(\bar{Y}_{.1} - \bar{Y}_{.2}\right) = V\left(\bar{Y}_{.1}\right) + V\left(\bar{Y}_{.2}\right) - 2\,\mathrm{cov}\left(\bar{Y}_{.1}, \bar{Y}_{.2}\right)$$

$$= \frac{\sigma_1^2}{R} + \frac{\sigma_2^2}{R} - \frac{2\rho_{12}\sigma_1\sigma_2}{R}$$

ρ_{12} is positive

  - Variance of $\bar{Y}_{.1} - \bar{Y}_{.2}$ arising from CRN is less than that of independent sampling (with $R_1 = R_2$).

Common Random Numbers (CRN)

[Comparison of 2 systems]

- The estimator based on CRN is more precise, leading to a shorter confidence interval for the difference.
- Sample variance of the differences $\overline{D} = \overline{Y}_{.1} - \overline{Y}_{.2}$:

$$S_D^2 = \frac{1}{R-1}\sum_{r=1}^{R}\left(\overline{D}_r - \overline{D}\right)^2 = \frac{1}{R-1}\left(\sum_{r=1}^{R} D_r^2 - R\overline{D}^{\,2}\right)$$

where $D_r = Y_{r1}\text{-}Y_{r2}$ and $\overline{D} = \frac{1}{R}\sum_{r=1}^{R} D_r$, with degress of freedom $\upsilon = \mathrm{R} - 1$

- Standard error:

$$s.e.(\overline{D}) = s.e.\left(\overline{Y}_{.1} - \overline{Y}_{.2}\right) = \frac{S_D}{\sqrt{R}}$$

Common Random Numbers (CRN)

[Comparison of 2 systems]

- It is never enough to simply use the same seed for the random-number generator(s):
  - The random numbers must be synchronized: each random number used in one model for some purpose should be used for the same purpose in the other model.
  - e.g., if the i^{th} random number is used to generate a service time at work station *2* for the 5^{th} arrival in model 1, the i^{th} random number should be used for the very same purpose in model 2.

Common Random Numbers (CRN)

[Comparison of 2 systems]

- Vehicle inspection example:
  - 4 input random variables:
    - A_n, interarrival time between vehicles *n* and *n+1*,
    - $S_n^{(i)}$, inspection time for task *i* for vehicle *n* in model 1 (i=1,2,3; refers to brake, headlight and steering task, respectively).
  - When using CRN:
    - Same values should be generated for A_1, A_2, A_3, …in both models.
    - Mean service time for model 2 is *10%* less. 2 possible approaches to obtain the service times:
      - Let $S_n^{(i)}$, be the service times generated for model 1, use:

        $$S_n^{(i)} - 0.1E[S_n^{(i)}]$$

      - Let $Z_n^{(i)}$, as the standard normal variate, $\sigma = 0.5$ *minutes*, use:

        $$E[S_n^{(i)}] + \sigma Z_n^{(i)}$$

    - For synchronized runs: the service times for a vehicle were generated at the instant of arrival and stored as its attribute and used as needed.

Common Random Numbers (CRN)

[Comparison of 2 systems]

- Vehicle inspection example (cont.): compare the two systems using independent sampling and CRN where $R = R_1 = R_2 = 10$ (see Table 12.2 for results):
  - Independent sampling: $\bar{Y}_{.1} - \bar{Y}_{.2} = -5.4$ minutes

    with $\upsilon = 17,\ t_{0.05,17} = 2.11,\ S_1^2 = 118.0$ and $S_2^2 = 244.3$, c.i.: $-18.1 \le \theta_1 - \theta_2 \le 7.3$
  - CRN without synchronization: $\bar{Y}_{.1} - \bar{Y}_{.2} = -1.9$ minutes

    with $\upsilon = 9,\ t_{0.05,9} = 2.26,\ S_D^2 = 208.9$, c.i.: $-12.3 \le \theta_1 - \theta_2 \le 8.5$
  - CRN with synchronization: $\bar{Y}_{.1} - \bar{Y}_{.2} = 0.4$ minutes

    with $\upsilon = 9,\ t_{0.05,9} = 2.26,\ S_D^2 = 1.7$, c.i.: $-0.50 \le \theta_1 - \theta_2 \le 1.30$
    - The upper bound indicates that system 2 is at almost 1.30 minutes faster in expectation. Is such a difference practically significant?

CRN with Specified Precision

[Comparison of 2 systems]

- Goal: The error in our estimate of $\theta_1 - \theta_2$ to be less than ε.
- Approach: determine the number of replications *R* such that the half-width of c.i.: $H = t_{\alpha/2,\upsilon} s.e.\left(\bar{Y}_{.1} - \bar{Y}_{.2}\right) \le \varepsilon$
- Vehicle inspection example (cont.):
  - $R_0 = 10$, complete synchronization of random numbers yield 95% c.i.: 0.4 ± 0.9 minutes
  - Suppose $\varepsilon = 0.5$ minutes for practical significance, we know *R* is the smallest integer satisfying $R \ge R_0$ and: $R \ge \left(\frac{t_{\alpha/2,R-1} S_D}{\varepsilon}\right)^2$
  - Since $t_{\alpha/2,R-1} \le t_{\alpha/2,R_0-1}$, a conservation estimate of *R* is:

$$R \ge \left(\frac{t_{\alpha/2,R_0-1} S_D}{\varepsilon}\right)^2$$

  - Hence, 35 replications are needed (25 additional).

Comparison of Several System Designs

- To compare K alternative system designs
  - Based on some specific performance measure, θ_i, of system *i* , for *i = 1, 2, …, K.*
- Procedures are classified as:
  - Fixed-sample-size procedures: predetermined sample size is used to draw inferences via hypothesis tests of confidence intervals.
  - Sequential sampling (multistage): more and more data are collected until an estimator with a prespecified precision is achieved or until one of several alternative hypotheses is selected.
- Some goals/approaches of system comparison:
  - Estimation of each parameter θ_i.
  - Comparison of each performance measure θ_i, to control θ_1.
  - All pairwise comparisons, θ_i, - θ_j, for all i not equal to j
  - Selection of the best θ_i.

Bonferroni Approach [Multiple Comparisons]

- To make statements about several parameters simultaneously, (where all statements are true simultaneously).
- Bonferroni inequality:

$$P(\text{all statements } S_i \text{ are true}, i = 1, \ldots, C) \geq 1 - \sum_{j=1}^{C} \alpha_j = 1 - \alpha_E$$

Overall error probability, provides an upper bound on the probability of a false conclusion

  - The smaller α_j is, the wider the j^{th} confidence interval will be.
- Major advantage: inequality holds whether models are run with independent sampling or CRN
- Major disadvantage: width of each individual interval increases as the number of comparisons increases.

Bonferroni Approach [Multiple Comparisons]

- Should be used only when a small number of comparisons are made
  - Practical upper limit: about 20 comparisons
- 3 possible applications:
  - Individual c.i.'s: Construct a $100(1-\alpha_j)\%$ c.i. for parameter θ_i, where # of comparisons = K.
  - Comparison to an existing system: Construct a $100(1-\alpha_j)\%$ c.i. for parameter $\theta_i - \theta_1$ (i = 2,3, …K), where # of comparisons = $K - 1$.
  - All pairwise: For any 2 different system designs, construct a $100(1-\alpha_j)\%$ c.i. for parameter $\theta_i - \theta_j$. Hence, total # of comparisons $= K(K-1)/2$.

Bonferroni Approach to Selecting the Best

[Multiple Comparisons]

- Among K system designs, to find the best system
  - "Best" - the maximum expected performance, where the i^{th} design has expected performance θ_i.
- Focus on parameters: $\theta_i - \max_{j \neq i} \theta_j \quad \text{for } i = 1,2,\ldots,K$
  - If system design *i* is the best, it is the difference in performance between the best and the second best.
  - If system design *i* is not the best, it is the difference between system *i* and the best.
- Goal: the probability of selecting the best system is at least *1* – α, whenever $\theta_i - \max_{j \neq i} \theta_j \geq \varepsilon$.
  - Hence, both the probability of correct selection *1*-α, and the practically significant difference ε, are under our control.
- A two-stage procedure.

Bonferroni Approach to Selecting the Best
[Multiple Comparisons]

- Vehicle inspection example (cont.): Consider K = 4 different designs for the inspection station.
  - Goal: 95% confidence of selecting the best (with smallest expected response time) where ε = *2 minutes.*
  - A minimization problem: focus on $\theta_i - \min_{j \neq i} \theta_j \quad \text{for } i = 1,2,\ldots,K$
  - $\varepsilon = 2$, $1-\alpha = 0.95$, $R_0=10$ and $t = t_{0.0167,9} = 2.508$
  - From Table 12.5, we know:
    $S_{12}^2 = 4.498,\ S_{13}^2 = 28.498,\ S_{14}^2 = 5.498,\ S_{23}^2 = 11.857,\ S_{24}^2 = 0.119,\ S_{34}^2 = 9.849$
  - The largest sample variance $\hat{S}^2 = \max_{i \neq j} S_{ij}^2 = S_{13}^2 = 28.498$, hence,
$$R = \max\left\{10, \left\lceil \frac{2.508^2 * 28.498}{2^2} \right\rceil \right\} = \max\{10, \lceil 44.8 \rceil\} = 45$$
  - Make *45 - 10 = 35* additional replications of each system.

Bonferroni Approach to Selecting the Best

[Multiple Comparisons]

- Vehicle inspection example (cont.):
  - Calculate the overall sample means:

$$\overline{\overline{Y}}_i = \frac{1}{45}\sum_{r=1}^{45} Y_{ri}$$

  - Select the system with smallest $\overline{\overline{Y}}_i$ is the best.
  - Form the confidence intervals:

$$\min\left\{0, \overline{\overline{Y}}_i - \min_{j\neq i} \overline{\overline{Y}}_j - 2\right\} \leq \theta_{\text{i}} - \min_{j\neq i} \theta_j \leq \max\left\{0, \overline{\overline{Y}}_i - \min_{j\neq i} \overline{\overline{Y}}_j - 2\right\}$$

  - Note, for maximization problem:
    - The difference for comparison is: $\theta_i - \max_{j\neq i} \theta_j$
    - The c.i. is: $\min\left\{0, \overline{\overline{Y}}_i - \max_{j\neq i} \overline{\overline{Y}}_j - 2\right\} \leq \theta_{\text{i}} - \max_{j\neq i} \theta_j \leq \max\left\{0, \overline{\overline{Y}}_i - \max_{j\neq i} \overline{\overline{Y}}_j - 2\right\}$

Bonferroni Approach for Screening

[Multiple Comparisons]

- A screening (subset selection) procedure is useful when a two-stage procedure isn't possible or when too many systems.
- Screening procedure: The retained subset contains the true best system with *probability* $\geq 1-\alpha$ when the data are normally distributed (independent sampling or CRN).
  - Specify $1-\alpha$, common sample size from each system and R≥2.
  - Make R replications of system i to obtain Y_{1i}, Y_{2i}, … < Y_{Ri} for system i = *1,2, …, K.*
  - Calculate the sample means for all systems $\bar{Y}_{.i}$.
  - Calculate sample variance of the difference for every system pair S_{ij}^2.
  - If bigger is better, then retain system i in the selected subset if:

$$\bar{Y}_{.i} \geq \bar{Y}_{.j} - t_{\alpha/(K-1),R-1} \frac{S_{ij}}{\sqrt{R}} \text{ for all } j \neq i$$

  - If smaller is better, then retain system i in the selected subset if:

$$\bar{Y}_{.i} \leq \bar{Y}_{.j} - t_{\alpha/(K-1),R-1} \frac{S_{ij}}{\sqrt{R}} \text{ for all } j \neq i$$

Metamodeling

- Goal: describe the relationship between variables and the output response.
- The simulation output response variable, Y, is related to k independent variables $x_1, x_2, \ldots, x_k$ (the design variables).
- The true relationship between variables Y and x is represented by a (complex) simulation model.
- Approximate the relationship by a simpler mathematical function called a metamodel, some metamodel forms:
  - Linear regression.
  - Multiple linear regression.

Simple Linear Regression [Metamodeling]

- Suppose the true relationship between Y and x is suspected to be linear, the expected value of Y for a given x is: $E(Y|x) = \beta_0 + \beta_1 x$

where β_0 is the intercept on the Y axis, and β_1 is the slope.

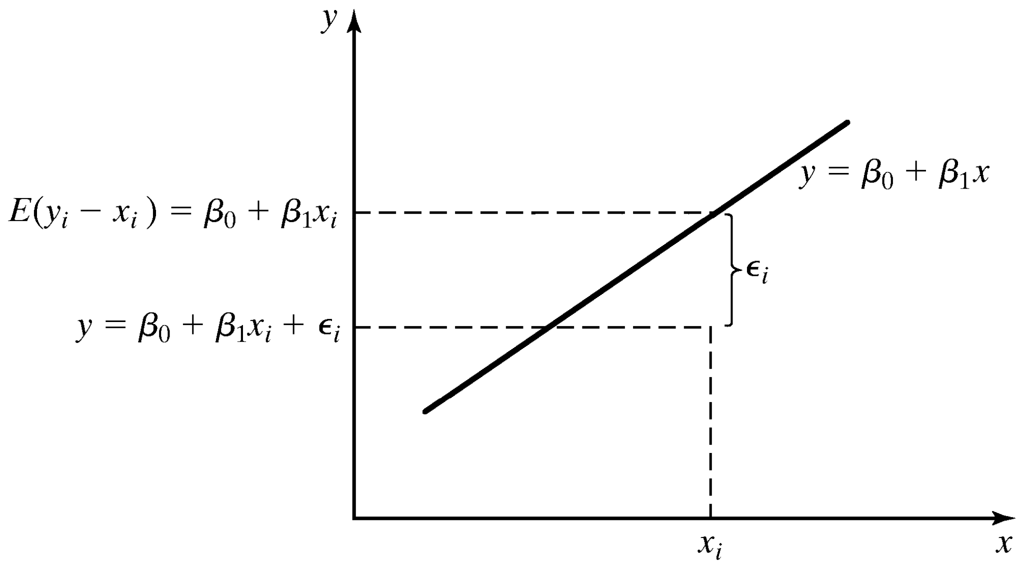
- Each observation of Y can be described by the model:

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

where ε is the random error with mean zero and constant variance σ^2

Simple Linear Regression [Metamodeling]

- Suppose there are *n* pairs observations, the method of least squares is commonly used to estimate β_0 and β_1.
  - The sum of squares of the deviation between the observations and the regression line is minimized.



Simple Linear Regression [Metamodeling]

- The individual observation can be written as:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

where $\varepsilon_1, \varepsilon_2 \ldots$ are assumed to be uncorrelated r.v.

- Rewrite: $Y_i = \beta_0^{'} + \beta_1 (x_i - \bar{x}) + \varepsilon_i$

where $\beta_0^{'} = \beta_0 + \beta_1 \bar{x}$ and $\bar{x} = \sum_{i=1}^{n} x_i / n$

- The least-square function (the sum of squares of the deviations):

$$L = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left(Y_i - \beta_0 + \beta_1 x_i\right)^2 = \sum_{i=1}^{n} \left[Y_i - \beta_0^{'} + \beta_1 (x_i - x)\right]^2$$

- To minimize L, find $\partial L / \partial \beta_0^{'}$ and $\partial L / \partial \beta_1$, set each to zero, and solve for:

$$\hat{\beta}_0^{'} = \bar{Y} = \sum_{i=1}^{n} \frac{Y_i}{n} \quad \text{and} \quad \hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum_{i=1}^{n} Y_i (x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

S_{xy} – corrected sum of cross products of x and Y

S_{xx} – corrected sum of squares of x

Test for Significance of Regression

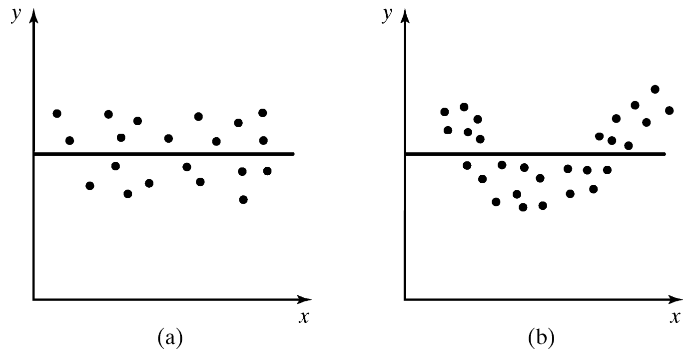
[Metamodeling]

- The adequacy of a simple linear relationship should be tested prior to using the model.
  - Testing whether the order of the model tentatively assumed is correct, commonly called the “lack-of-fit” test.
  - The adequacy of the assumptions that errors are *NID*$(0,\sigma^2)$ can and should be checked by residual analysis.

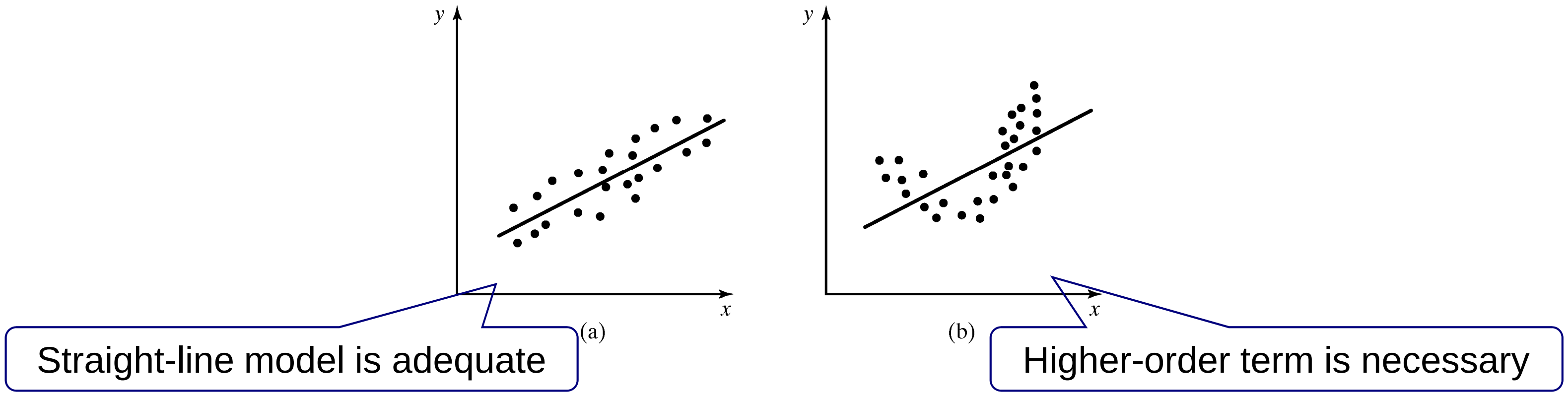
Test for Significance of Regression

[Metamodeling]

- Hypothesis testing: $H_0 : \beta_1 = 0$ and $H_1 : \beta_1 \neq 0$
  - Failure to reject H_0 indicates no linear relationship between x and Y.



  - If H_0 is rejected, implies that x can explain the variability in Y, but there may be in higher-order terms.



Test for Significance of Regression

[Metamodeling]

- The appropriate test statistics:

$$t_0 = \frac{\hat{\beta}_1}{\sqrt{MS_E / S_{xx}}}$$

  - The mean squared error is:

$$MS_E = \sum_{i=1}^{n} \frac{e_i^2}{n-2} = \frac{S_{yy} - \hat{\beta}_1 S_{xy}}{n-2}$$

    which is an unbiased estimator of $\sigma^2 = V(\varepsilon_i)$.

  - t_0 has the *t*-distribution with *n-2* degrees of freedom.
  - Reject H_0 if $|t_0| > t_{\alpha/2, n-2}$.

Multiple Linear Regression [Metamodeling]

- Suppose simulation output Y has several independent variables (decision variables). The possible relationship forms are:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_m x_m + \varepsilon$$

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x^2 + \varepsilon$$

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon$$

Random-Number Assignment for Regression [Metamodeling]

- Independent sampling:
  - Assign a different seed or stream to different design points.
  - Guarantees that the responses Y from different design points will be significantly independent.
- CRN:
  - Use the same random number seeds or streams for all of the design points.
  - A fairer comparison among design points (subjected to the same experimental conditions)
  - Typically reduces variance of estimators of slope parameters, but increases variance of intercept parameter

Optimization via Simulation

- Optimization usually deals with problems with certainty, but in stochastic discrete-event simulation, the result of any simulation run is a random variable.
- Let $x_1, x_2, \ldots, x_m$ be the m controllable design variables & $Y(x_1, x_2, \ldots, x_m)$ be the observed simulation output performance on one run:
  - To optimize $Y(x_1, x_2, \ldots, x_m)$ with respect to $x_1, x_2, \ldots, x_m$ is to maximize or minimize the mathematical expectation (long-run average) of performance, $E[Y(x_1, x_2, \ldots, x_m)]$.
- Example: select the material handling system that has the best chance of costing less than *$D* to purchase and operate.
  - Objective: maximize $Pr(Y(x_1, x_2, \ldots, x_m) \le D)$.
  - Define a new performance measure:

$$Y'(x_1, x_2, \ldots x_m) = \begin{cases} 1, & \text{if } Y(x_1, x_2, \ldots x_m) \le D \\ 0, & \text{otherwise} \end{cases}$$

  - Maximize $E(Y'(x_1, x_2, \ldots, x_m))$ *instead.*

Robust Heuristics

- The most common algorithms found in commercial optimization via simulation software.
- Effective on difficult, practical problems.
- However, do not guarantee finding the optimal solution.
- Example: genetic algorithms and tabu search.
- It is important to control the sampling variability.

Control sampling variability

[Optimization via Simulation]

- To determine how much sampling (replications or run length) to undertaken at each potential solution.
  - Ideally, sampling should increase as heuristic closes in on the better solutions.
  - If specific and fixed number of replications per solution is required, analyst should:
    - Conduct preliminary experiment.
    - Simulate several designs (some at extremes of the solution space and some nearer the center).
    - Compare the apparent best and apparent worst of these designs.
    - Find the minimum for the number of replications required to declare these designs to be statistically significantly different.
    - After completion of optimization run, perform a *2^nd^* set of experiments on the top *5* to *10* designs identified by the heuristic, rigorously evaluate which are the best or near-best of these designs.

Summary

- Basic introduction to comparative evaluation of alternative system design:
  - Emphasized comparisons based on confidence intervals.
  - Discussed the differences and implementation of independent sampling and common random numbers.
  - Introduced concept of metamodels.