Module-9:

Comparison and Evaluation of Alternative System Designs- Comparison of Two System Designs. Sampling with Equal and Unequal Variances. Common Random Numbers. Confidence Intervals with Specified Precision. Comparison of Several System Designs: Bonferroni Approaches to Multiple Comparisons and to Screening and to Selection of the Best. Meta modeling Sample Linear Regression, Testing for Significance, Multiple Linear Regressions. Random Number Assignment for Regression. Optimization via Simulation: Robust Heuristics.

Comparing Alternative Systems

Simulations are used to compare two or more alternative designs of systems. This comparison may be based on one or more decision variables such as buffer capacity, work schedule, resource availability, etc. Comparing alternative designs requires careful analysis to ensure that differences being observed are attributable to actual differences in performance and not to statistical variation. This is where running multiple replications may again be helpful.

Suppose, for example, that method A for deploying resources yields a throughput of 100 entities for a given time period while method B results in 110 entities for the same time period. Is it valid to conclude that method B is better than method A, or might additional replications actually lead the opposite conclusion?

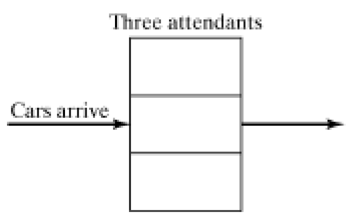
Evaluating alternative configurations or operating policies can sometimes be performed by comparing the average result of several replications. Where outcomes are close or where the decision requires greater precision, a method referred to as *hypothesis testing* should be used. In hypothesis testing, first a hypothesis is formulated (e.g., that methods A and B both result in the same throughput) and then a test is made to see whether the results of the simulation lead us to reject the hypothesis. The outcome of the simulation runs may cause us to reject the hypothesis that methods A and B both result in equal throughput capabilities and conclude that the throughput does indeed depend on which method is used.

Sometimes there may be insufficient evidence to reject the stated hypothesis and thus the analysis proves to be inconclusive. This failure to obtain sufficient evidence to reject the hypothesis may be due to the fact that there really is no difference in performance, or it may be a result of the variance in the observed outcomes being too high given the number of replications to be conclusive. At this point, either additional (perhaps time consuming) replications may be run or one of several variance reduction techniques might be employed (see Law and Kelton, 1991).

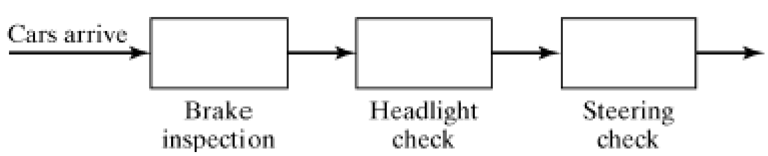
Vehicle-safety inspection example

The station performs 3 jobs:
(1) brake check,
(2) headlight check, and
(3) steering check.
• Vehicles arrival: Poisson 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

Systems design is the process of defining the architecture, modules, interfaces, and data for a system to satisfy specified requirements. Systems design could be seen as the application of systems theory to product development. There is some overlap with the disciplines of systems analysis, systems architecture and systems engineering.

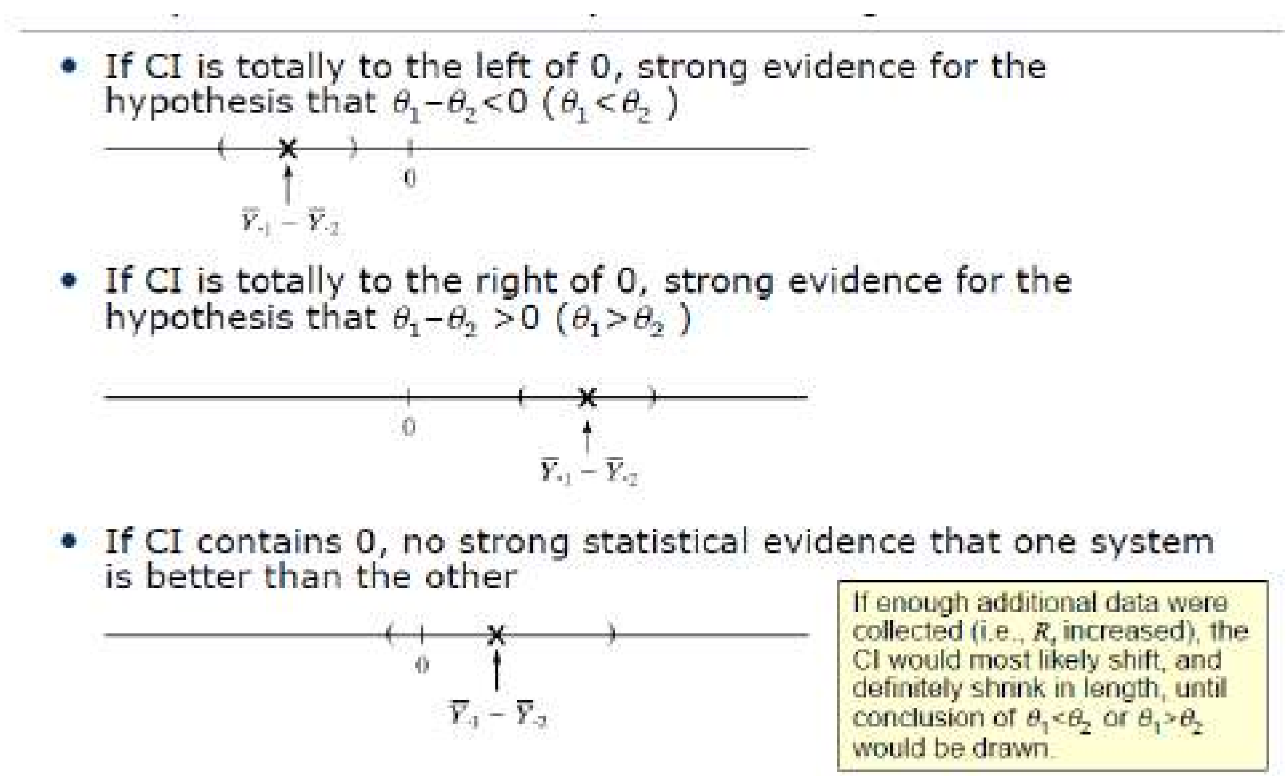
- ✓ From replication *r* of system *i*, the analyst obtains an estimate *Yir* of the mean performance measure θ
- ✓ Assuming that the estimators *Yir* are (at least approx.) unbiased:

θ1 = *E*(*Y1r*), *r* = 1, … , *R*1

θ2 = *E*(*Y*2*r*), *r* = 1, … , *R*2

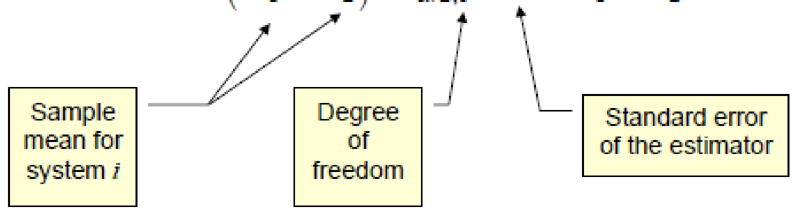
• Goal:

Compute a confidence interval for θ1 – θ2 to compare the two system designs



- A two-sided 100(1-α)% CI for $\theta_1 - \theta_2$ always takes the form of:

$$\left(\overline{Y}_{\bullet 1} - \overline{Y}_{\bullet 2}\right) \pm t_{\alpha/2,v} \cdot \text{s.e.}(\overline{Y}_{\bullet 1} - \overline{Y}_{\bullet 2})$$



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

- Statistically significant versus practically significant
  - Statistical significance: is the observed difference $\overline{Y}_{\bullet 1} - \overline{Y}_{\bullet 2}$ larger than the variability in $\overline{Y}_{\bullet 1} - \overline{Y}_{\bullet 2}$?
  - 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 the range:

$$\left(\overline{Y}_{\bullet 1} - \overline{Y}_{\bullet 2}\right) - t_{\frac{\alpha}{2},v} s.e.(\overline{Y}_{\bullet 1} - \overline{Y}_{\bullet 2}) \le \theta_1 - \theta_2 \le \left(\overline{Y}_{\bullet 1} - \overline{Y}_{\bullet 2}\right) + t_{\frac{\alpha}{2},v} s.e.(\overline{Y}_{\bullet 1} - \overline{Y}_{\bullet 2})$$

  - Whether a difference within these bounds is practically significant depends on the particular problem.

Sampling with Equal and Unequal Variances

Sampling is a general technique for estimating properties of a particular distribution, while only having samples generated from a different distribution than the distribution of interest. **Variance reduction** is a procedure used to increase the precision of the estimates that can be obtained for a given simulation. Every output random variable from the simulation is associated with a variance which limits the precision of the simulation results. In order to make a simulation statistically efficient, i.e., to obtain a greater precision and smaller confidence intervals for the output random variable of interest, variance reduction techniques can be used.

Independent Sampling with Equal Variances

- 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 $\bar{Y}_{\bullet i}$ is:

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

- For independent samples:

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

- If it is reasonable to assume that $\sigma^2_1 = \sigma^2_2$ (approx.) 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: $\bar{Y}_{\bullet 1} - \bar{Y}_{\bullet 2}$
  - The sample variance for system i is:

$$S_i^2 = \frac{1}{R_i - 1}\sum_{r=1}^{R_i}(Y_{ri} - \bar{Y}_{\bullet i})^2 = \frac{1}{R_i - 1}\sum_{r=1}^{R_i} Y_{ri}^{\ 2} - R_i \bar{Y}_{\bullet i}^{\ 2}$$

  - The pooled estimate of σ^2 is:

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

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

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

Independent Sampling with Unequal Variances

- If the assumption of equal variances cannot safely be made, an approximate $100(1-\alpha)\%$ CI can be computed as:

$$s.e.(\bar{Y}_1 - \bar{Y}_2) = \sqrt{\frac{S_1^2}{R_1} + \frac{S_2^2}{R_2}}$$

- With degrees of freedom:

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

- In this case, the minimum number of replications $R_1 > 7$ and $R_2 > 7$ is recommended.

Common Random Numbers (CRN)

The common random numbers variance reduction technique is a popular and useful variance reduction technique which applies when we are comparing two or more alternative configurations (of a system) instead of investigating a single configuration. CRN has also been called *correlated sampling*, *matched streams* or *matched pairs*.

CRN requires synchronization of the random number streams, which ensures that in addition to using the same random numbers to simulate all configurations, a specific random number used for a specific purpose in one configuration is used for exactly the same purpose in all other configurations. For example, in queueing theory, if we are comparing two different configurations of tellers in a bank, we would want the (random) time of arrival of the *N*th customer to be generated using the same draw from a random number stream for both configurations.

Confidence Intervals with Specified Precision

A confidence interval (CI) is a type of interval estimate, computed from the statistics of the observed data, that might contain the true value of an unknown population parameter. The confidence interval in the frequentist school is by far the most widely used statistical interval and the Layman's definition would be the probability that you will have the true value for a parameter such as the mean or the mean difference or the odds ratio under repeated sampling.

Sample size determination is targeting the interval width, that's quite closely tied to the idea of the standard error of the estimate which is tied to your law of large number theory.

- Sometimes we need to do the inverse, given a level of error and confidence, how many replications are needed?
- The half-length (h.i.) of a $100(1-\alpha)\%$ confidence interval for a mean θ, based on the *t* distribution, is

$$h.i. = t_{\alpha/2,R-1} * \hat{\sigma}(\hat{\theta})(*)$$

where $\hat{\sigma}(\hat{\theta}) = S/\sqrt{R}$, *S* is the sample standard deviation, *R* is the number of replications.

- Assume an error criterion ϵ is specified with a confidence level $1-\alpha$, it is desired that a sufficiently large sample size *R* be taken such that

$$P(|\hat{\theta} - \theta| < \epsilon) \geq 1 - \alpha$$

Since we have the relation (*), the desired the error control condition can be written as

$$h.i. = \frac{t_{\alpha/2,R-1} S_0}{\sqrt{R}} \leq \epsilon$$

Solve the above relation, we have

$$R \geq \left(\frac{t_{\alpha/2,R-1} S_0}{\epsilon}\right)^2$$

since $t_{\alpha/2,R-1} \geq z_{\alpha/2}$ the above relation can be written

$$R \geq \left(\frac{z_{\alpha/2} S_0}{\epsilon} \right)^2$$

For $R \geq 50, t_{\alpha/2,R-1} \approx z_{\alpha/2}$ the inequality with standard normal distribution holds. This says we need to run that many (*R*) replications to satisfy the error requirement.

- The true value of θ is in the following range with probability of $100(1-\alpha)\%$

$$\hat{\theta} - \frac{t_{\alpha/2,R-1} S}{\sqrt{R}} \leq \theta \leq \hat{\theta} + \frac{t_{\alpha/2,R-1} S}{\sqrt{R}}$$

In statistics, the **Bonferroni correction** is one of several methods used to counteract the problem of multiple comparisons

This method answers the question:

"Which treatment means are significantly different from each other?"

Bonferroni's method provides a pairwise comparison of the means. To determine which means are significantly different, we must compare all pairs. There are k = (a) (a-1)/2 possible pairs where a = the number of treatments. In this example, a= 4, so there are 4(4-1)/2 = 6 pairwise differences to consider.

To start, we must select a value for alpha (α), the confidence level. We will select α = 0.05. In Bonferroni's method, the idea is to divide this familywise error rate (0.05) among the k tests. So each test is done at the α/k level.

We will use the t distribution to help determine the pairwise confidence interval. To start, we need to calculate the pooled variance. This is an estimate of the variance based on the four treatment means. The following equation is used to determine the pooled variance:

$$s_p^2 = \frac{\sum (n_i - 1) s_i^2}{\sum (n_i - 1)}$$

where n_i is the sample size and s_i is the standard deviation for the i^{th} sample. The standard deviations are given in the table above. The pooled variance for this data is then:

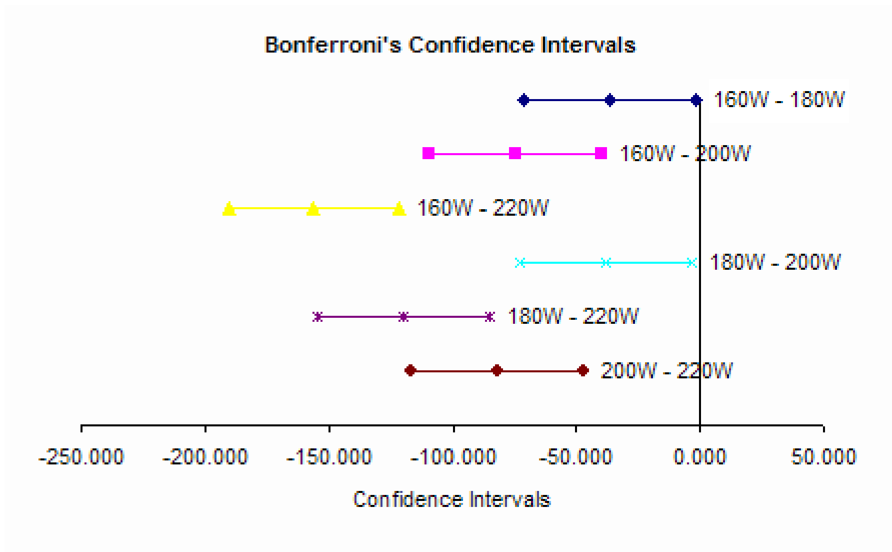
$$s_p^2 = \frac{\sum (n_i - 1)s_i^2}{\sum (n_i - 1)} = \frac{4(20.017^2) + 4(16.742^2) + 4(20.526^2) + 4(15.248^2)}{16} = 333.7$$

If you look at the ANOVA table from last month, you will see that this value corresponds to the mean square error.

. CONFIDENCE INTERVALS **Part of the output from the SPC for Excel program using Bonferroni's method to analyze this data is shown below. This table includes the confidence intervals for each difference in means. These are determined simply by adding and subtracting the critical value from the difference in the treatment means. As long as the confidence interval does not contain 0, there is significant difference in the two means.**

Bonferroni's Method					
Family Conf. Int.=95%, Individual Conf. Int.=98.75%					
Comparisons	Diff. in Means	Critical Value	LCon	UCon	Sig Diff?
160W - 180W	-36.2	34.756	-70.956	-1.444	Yes
160W - 200W	-74.2	34.756	-108.956	-39.444	Yes
160W - 220W	-155.8	34.756	-190.556	-121.044	Yes
180W - 200W	-38	34.756	-72.756	-3.244	Yes
180W - 220W	-119.6	34.756	-154.356	-84.844	Yes
200W - 220W	-81.6	34.756	-116.356	-46.844	Yes
There is evidence that some pairs of means are different.					

The confidence intervals can also be plotted as shown below. This shows that none of the intervals includes 0. All the treatment means are significantly different from one another.



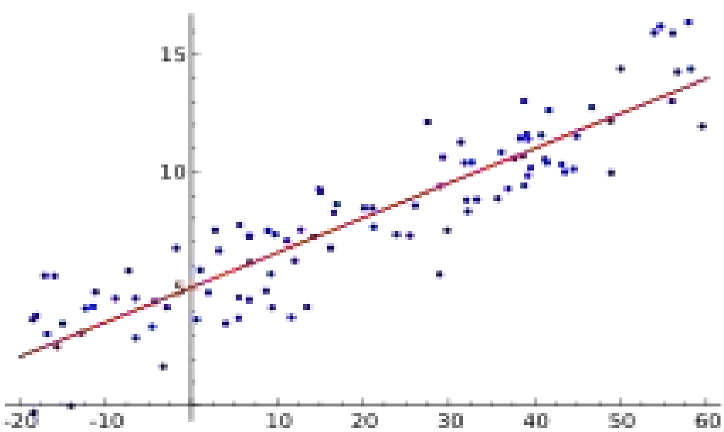
Meta modeling Sample Linear Regression

Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. One variable is considered to be an explanatory variable, and the other is considered to be a dependent variable. For example, a modeler might want to relate the weights of individuals to their heights using a linear regression model.

Linear regression is a basic and commonly used type of predictive analysis. The overall idea of regression is to examine two things:

(1) does a set of predictor variables do a good job in predicting an outcome (dependent) variable?

(2) Which variables in particular are significant predictors of the outcome variable, and in what way do they–indicated by the magnitude and sign of the beta estimates–impact the outcome variable?



These regression estimates are used to explain the relationship between one dependent variable and one or more independent variables. The simplest form of the regression equation with one dependent and one independent variable is defined by the formula y = c + b*x, where y = estimated dependent variable score, c = constant, b = regression coefficient, and x = score on the independent variable.

Naming the Variables. There are many names for a regression's dependent variable. It may be called an outcome variable, criterion variable, endogenous variable, or regressand. The independent variables can be called exogenous variables, predictor variables, or regressors.

Three major uses for regression analysis are

(1) determining the strength of predictors,

(2) forecasting an effect, and

(3) trend forecasting.

Types of Linear Regression

Simple linear regression

1 dependent variable (interval or ratio), 1 independent variable (interval or ratio or dichotomous)

Multiple linear regression

1 dependent variable (interval or ratio) , 2+ independent variables (interval or ratio or dichotomous)

Logistic regression

1 dependent variable (dichotomous), 2+ independent variable(s) (interval or ratio or dichotomous)

Ordinal regression

1 dependent variable (ordinal), 1+ independent variable(s) (nominal or dichotomous)

Multinomial regression

1 dependent variable (nominal), 1+ independent variable(s) (interval or ratio or dichotomous)

Discriminant analysis

1 dependent variable (nominal), 1+ independent variable(s) (interval or ratio)

Simple Linear Regression

- Suppose the true relationship between Y and x is assumed to be linear, the expected value of Y for a given x is:

$$E(Y \mid x) = \beta_0 + \beta_1 x$$

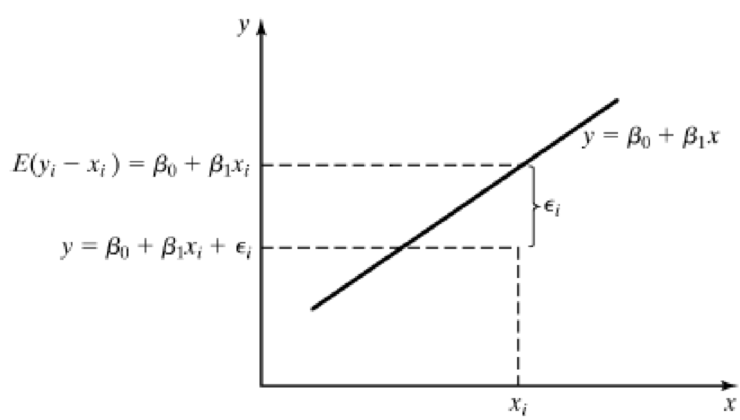
where β_0 is the intercept on the Y axis, and β_1 is the slope.

- 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

- Suppose there are n pairs of 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.



- 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 random variables

- Rewrite:

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

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

- 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

Multiple Linear Regression

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

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

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

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

Meta-regression

Meta-regression is a tool used in meta-analysis to examine the impact of moderator variables on study effect size using regression-based techniques. Meta-regression is more effective at this task than are standard meta-analytic techniques. Meta-regression analysis (MRA) is a quantitative method of conducting literature surveys. Meta-regression has gained popularity in social, behavioral and economic sciences. Important applications have focused on qualifying estimates of policy-relevant parameters, testing economic theories, explaining heterogeneity, and qualifying potential biases. Generally, three types of models can be distinguished in the literature on meta-analysis: simple regression, fixed effect meta-regression and random effects meta-regression.

The model can be specified as

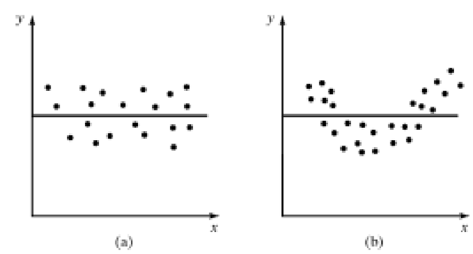
$$y_j = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \cdots + \varepsilon$$

Where y_j is the effect size in study j and β_0 (intercept) the estimated overall effect size. The variables x_i $(i = 1 \ldots k)$ specify different characteristics of the study, ε specifies the between study variation. Note that this model does not allow specification of within study variation.

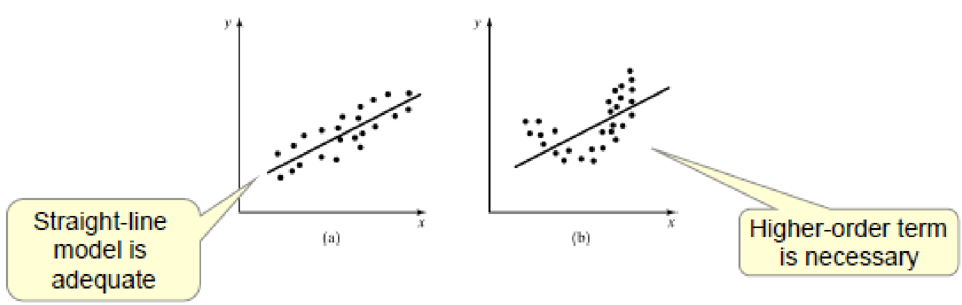
Test for Significance of Regression

- 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 (normally and independent) $NID(0,\sigma^2)$ can and should be checked by residual analysis.

- 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

- 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 $|t_0| > t_{\alpha/2,\, n-2}$

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

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

Robust Regression with Optimisation Heuristics

Linear regression is widely-used in finance. While the standard method to obtain parameter estimates, Least Squares, has very appealing theoretical and numerical properties, obtained estimates are often unstable in the presence of extreme observations which are rather common in financial time series. One approach to deal with such extreme observations is the application of robust or resistant estimators, like Least Quantile of Squares estimators. Unfortunately, for many such alternative approaches, the estimation is much more difficult than in the Least Squares case, as the objective function is not convex and often has many local optima. We apply different heuristic methods like Differential Evolution, Particle Swarm and Threshold Accepting to obtain parameter estimates. Particular emphasis is put on the convergence properties of these techniques for fixed computational resources, and the techniques' sensitivity for different parameter settings.