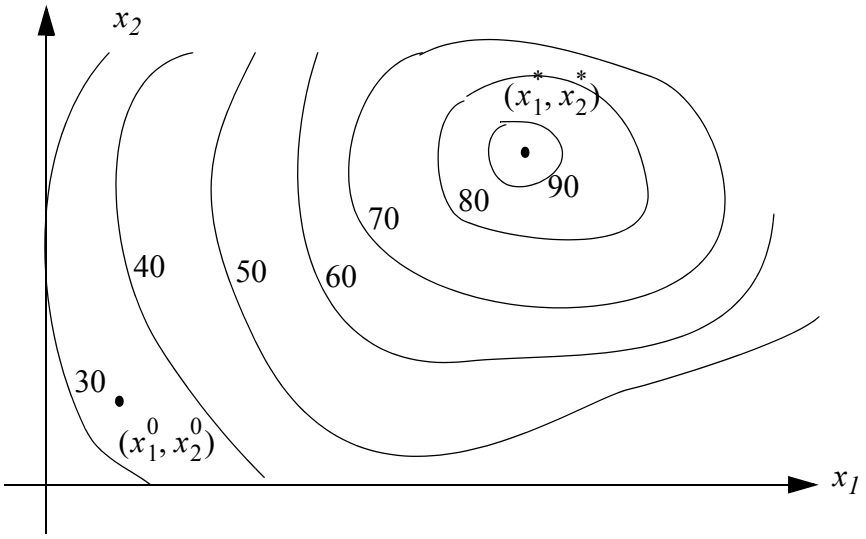
14.0 RESPONSE SURFACE METHODOLOGY (RSM)

(Updated Spring 2001)

So far, we've focused on experiments that:

- Identify a few important variables from a large set of candidate variables, i.e., a screening experiment.
- Ascertain how a few variables impact the response

Now we want to answer the question: "What specific levels of the important variables produce an optimum response?"



How do we get from starting position (x_1^0, x_2^0) to optimum position (x_1^*, x_2^*), "top of the hill"?

Our approach will be based on characterizing the true response surface,

$$y = \eta + \varepsilon$$

where η is the mean response and ε is the error, with a model, $\hat{y}$.

RSM: An experimental optimization procedure:

1. Plan and run a factorial (or fractional factorial) design near/at our starting point. $(x_1^0, x_2^0, \ldots)$
2. Fit a linear model (no interaction or quadratic terms) to the data.
3. Determine path of steepest ascent (PSA) - quick way to move to the optimum - gradient based
4. Run tests on the PSA until response no longer improves.

5. If curvature of surface large go to step 6, else go to step 1
6. Neighborhood of optimum - design, run, and fit (using least squares) a 2nd order model.
7. Based on 2nd order model - pick optimal settings of independent variables.

Example:

Froth Flotation Process - Desire to whiten a wood pulp mixture.

2 independent variables:

(1) Concentration of environmentally safe bleach
(2) Mixture temperature.

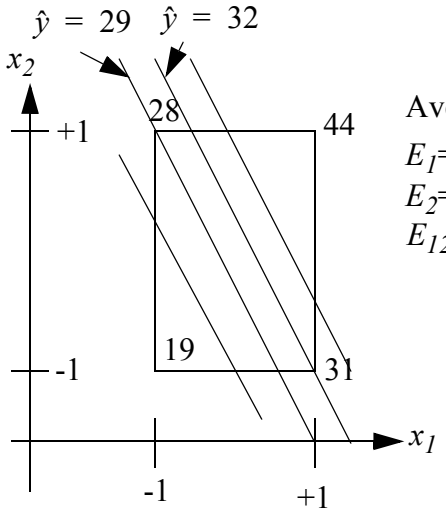
Current operating point:

%Bleach = 4%, Temp = 80^{o}F. we believe we are well away from the optimum.

Allowable variable ranges:
0-20% Bleach 60-130 Temp.

An initial 2^2 factorial design

Variable	Low Level	Midpoint	High Level
(1)% Bleach	2	4	6
(2) Temp	75	80	85



Averge=30.5
E_1=14
E_2=11
E_{12}=2 (It is small relative to E_1 and E_2, so the response surface is fairly linear)

Linear model: $y = b_0 + b_1x_1 + b_2x_2 + \varepsilon$

Fitted model is: $\hat{y} = 30.5 + 7x_1 + 5.5x_2$

A one unit change in x_1 is a change of 2% in Bleach.
A one unit change in x_2 is a change of 5 deg. in Temperature.

$$x_1 = \frac{\%\text{Bleach}-4}{2}, \qquad \text{Bleach}\% = 2x_1+4$$

$$x_2 = \frac{\text{Temp} - 80}{5}, \qquad \text{Temp} = 5x_2 + 80$$

We want to move from our starting point ($x_1 = 0$, $x_2 = 0$ or % Bleach = 4, Temp = 80) in the direction that will increase the response the fastest. Thus, we want to move in a direction normal to the contours.

Path of steepest ascent (PSA): line passing through point ($x_1 = 0$, $x_2 = 0$) that is perpendicular to the contours.

- Define contour passing through (0,0) $\hat{y} = 30.5 + 7x_1 + 5.5x_2 = 30.5$
- Relation between x_1 & x_2 for this contour $7x_1 + 5.5x_2 = 0$ or

$$x_2 = \frac{-7x_1}{5.5}$$

- Line normal to this one (slope of PSA = $\frac{-1}{\text{contour slope}}$)

$$x_2 = \frac{5.5x_1}{7}$$

- Interpretation: for a one unit change in x_1, x_2 must change by 5.5/7.0 units to remain on the PSA. A 7 unit change in x_1 -> a 5.5 unit change in x_2.

- Gradient of $\hat{y}$ may also be used to determine PSA.

Let's define some points (candidate tests) on the path of steepest ascent (PSA)

Candidate Test	x_1	$x_2 = \frac{5.5x_1}{7}$	%Bleach $= 2x_1+4$	Temp $=5x_2+80$	y
1	0	0	4	80	
2	0.5	0.39	5	82	
3	1.0	0.79	6	84	43
4	1.5	1.18	7	86	49
5	2.0	1.57	8	88	52
6	2.5	1.96	9	90	59
7	3.0	2.36	10	92	66
8	3.5	2.75	11	94	75
9	4.0	3.14	12	96	81
10	**4.5**	**3.54**	**13**	**98**	**86**
11	5.0	3.93	14	100	80
12	5.5	4.32	15	102	69
13	6.0	4.71	16	104	

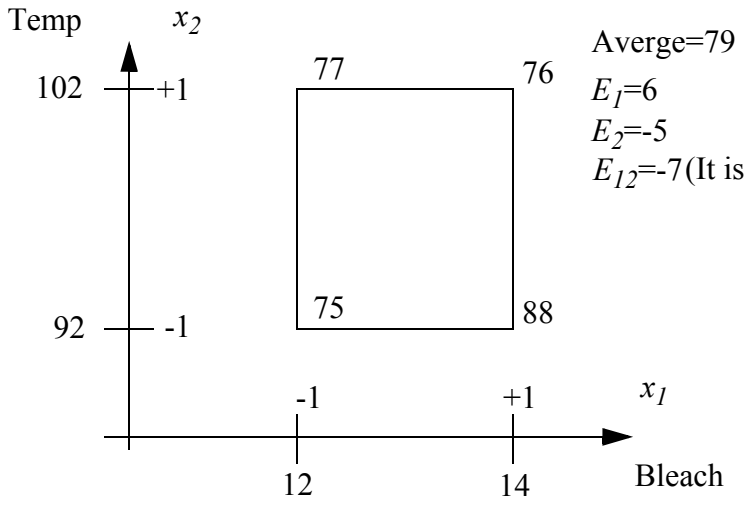
Since it appears that the response surface is fairly linear (E_{12} is small) where we conducted tests, no reason to examine/conduct test 1 or 2.
Run 1st test on or beyond boundary defined by $x_1, x_2 = \pm 1$.

Continue running tests on the PSA until response no longer increases.
Highest response occurs at (Bleach = 13%, Temp = 98^o). Run new factorial at this point.

Ran initial factorial design - defined path of steepest ascent - moved on PSA to

Bleach=13%, Temperature = 98^o, Whiteness = 86%.

Conduct another factorial design near this point



$\hat{y} = 79 + 3x_1 - 2.5x_2$ describes response surface with a plane.

At $x_1 = 0$, $x_2 = 0$, $\hat{y} = 79$

The combinations of (x_1, x_2) that give $\hat{y} = 79$ are specified by:

$3x_1 - 2.5\, x_2 = 0$

$x_2 = \frac{3}{2.5}x_1$ <------ (x_1, x_2) relationship along $\hat{y} = 79$ contour

PSA: $x_2 = \frac{-2.5}{3}x_1 = -0.833x_1$ or $x_1 = -1.2x_2$

Let's follow the PSA but take steps in the x_1 direction since E_1 is bigger than E_2.

Candidate Test	x_1	$x_2 = -.833x_1$	%Bleach $= 13+x_1$	Temp $=97+5x_2$	y
1	0	0	13	97	
2	0.5	-0.4165	13.5	94.9	
3	1.0	-0.833	14	92.8	88
4	1.25	-1.04125	14.25	91.8	90
5	**1.5**	**-1.2495**	**14.50**	**90.8**	**92**
6	1.75	-1.4578	14.75	89.7	89
7	2.00	-1.666	15.00	88.7	84
8	2.25	-1.87425	15.25	87.6	
9	2.50	-2.0825	15.50	86.6	
10	2.75	-2.29075	15.75	85.5	
11	3.00	-2.50	16.00	84.5	

Highest response at Bleach 14.50%, Temp 90.8^o -----> 92% Whiteness.

Note that we weren't able to move as far along the path before we "fell off".

We now believe we are fairly close to the optimum, so we must characterize the response surface with more than a plane, perhaps a model of the form:

$$y = b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2 + b_{11}x_1^2 + b_{22}x_2^2 + \varepsilon$$

To estimate the b's in this model, we need more than a 2 - level factorial design. A Central Composite Design (CCD) is more efficient than a 3-level factorial design.

Given the results of a 2nd order design - must use least squares (linear regression) to estimate the b's. We will examine the topic of regression after we finish up the RSM procedure for our example.

Fit the following model to the response surface near the point:
%Bleach=14.50, Temp = 90.8, Whiteness=92%

Note a small change in nomenclature:
use B's as the model parameter to be estimated,
use b's as the estimate of the parameter.

So, we are assuming the response surface can be characterized by the model:

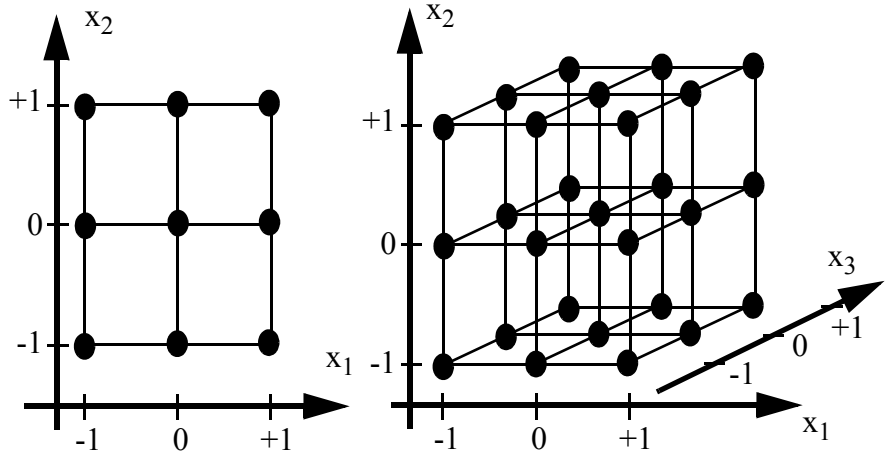
$$y = B_0 + B_1 x_1 + B_2 x_2 + B_{12} x_1 x_2 + B_{11} x_1^2 + B_{22} x_2^2 + \varepsilon$$

The fitted model will be of the form:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_{12} x_1 x_2 + b_{11} x_1^2 + b_{22} x_2^2$$

$$\underset{\sim}{b} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_{12} \\ b_{11} \\ b_{22} \end{bmatrix} = (\underset{\sim}{x}^T \underset{\sim}{x})^{-1} \underset{\sim}{x}^T \underset{\sim}{y}$$

Three-level Factorial Design

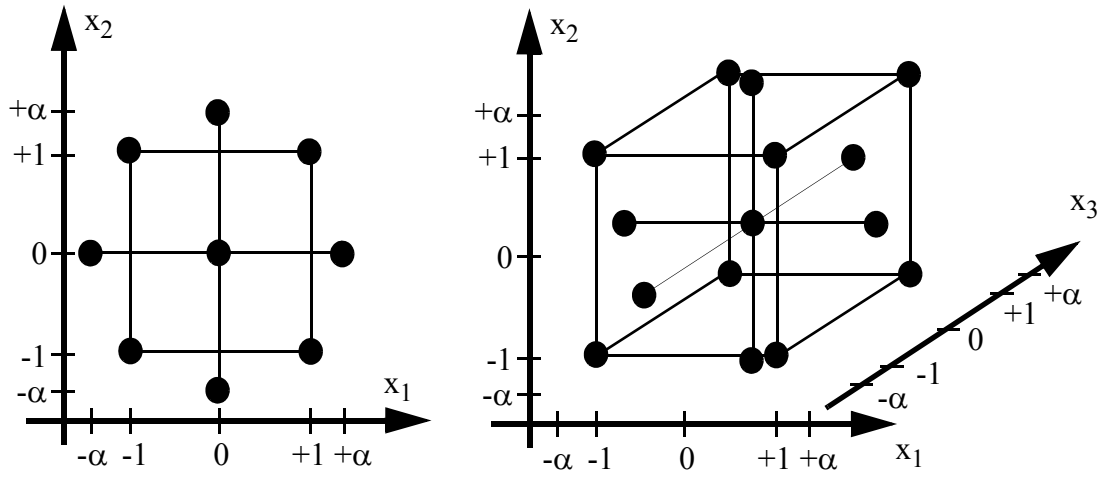


For three-level factorial designs, or 3^k designs, the number of tests in the

design is 3^k, where k is the number of variables being examined in the experiment.

Variables	2	3	4	5
Tests	9	27	81	243

Central Composite Design



In general, a central composite design (CCD) consists of 2^k points that form the base design, 2k star points, and several center points.

Central Composite Designs (adapted from Montgomery)

k	2	3	4	5	6
Factorial Portion = F	4	8	16	32	64
Axial Points	4	6	8	10	12
α	1.414	1.682	2.000	2.378	2.828
Center Points (unif. prec.) = n	5	6	7	10	15
Center Points (orthog.) = n	8	9	12	17	24
Total Tests (unif. prec.) = N	13	20	31	52	91
Total Tests (orthog.) = N	16	23	36	59	100

Number of Factorial Points, $F = 2^k$, $\alpha = (F)^{1/4}$: rotatability property - variance contours of $\hat{y}$ are concentric circles. Unif. Prec.: variance at origin same as variance a unit away from origin.

Variable Levels

	$-\alpha = -\sqrt{2}$	-1	0	+1	$+\alpha = +\sqrt{2}$
Bleach (%)	13.1	13.5	14.5	15.5	15.9
Temp (deg)	84	86	91	96	98

Test Plan

Test	x_1	x_2	%Bleach	Temp
1	-1	-1	13.5	86
2	+1	-1	15.5	86
3	-1	+1	13.5	96
4	+1	+1	15.5	96
5	$-\sqrt{2}$	0	13.1	91
6	$+\sqrt{2}$	0	15.9	91
7	0	$-\sqrt{2}$	14.5	84
8	0	$+\sqrt{2}$	14.5	98
9	0	0	14.5	91

Calculating the Parameter Estimates

$$\underset{\sim}{x} = \begin{bmatrix} 1 & -1 & -1 & +1 & +1 & +1 \\ 1 & +1 & -1 & -1 & +1 & +1 \\ 1 & -1 & +1 & -1 & +1 & +1 \\ 1 & +1 & +1 & +1 & +1 & +1 \\ 1 & -\sqrt{2} & 0 & 0 & 2 & 0 \\ 1 & +\sqrt{2} & 0 & 0 & 2 & 0 \\ 1 & 0 & -\sqrt{2} & 0 & 0 & 2 \\ 1 & 0 & +\sqrt{2} & 0 & 0 & 2 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad \underset{\sim}{y} = \begin{bmatrix} 87 \\ 85 \\ 89 \\ 83 \\ 86 \\ 82 \\ 98 \\ 87 \\ 92 \end{bmatrix}$$

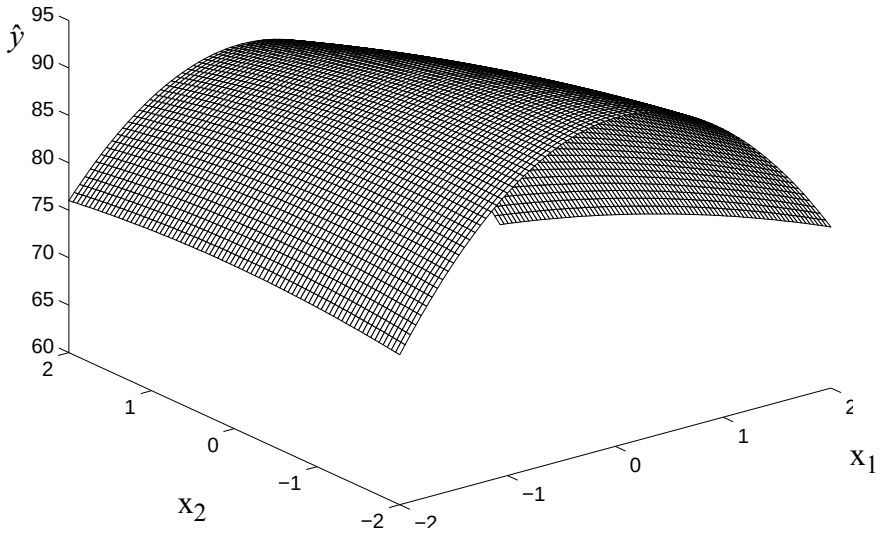
$$(\underset{\sim}{x}^T\underset{\sim}{x}) = \begin{bmatrix} 9 & 0 & 0 & 0 & 8 & 8 \\ 0 & 8 & 0 & 0 & 0 & 0 \\ 0 & 0 & 8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 & 0 \\ 8 & 0 & 0 & 0 & 12 & 4 \\ 8 & 0 & 0 & 0 & 4 & 12 \end{bmatrix}$$

$$(\underset{\sim}{x}^T\underset{\sim}{x})^{-1} = \begin{bmatrix} 1 & \dot{0} & 0 & 0 & -0.5 & -0.5 \\ 0 & 0.125 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.125 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.25 & 0 & 0 \\ -0.5 & 0 & 0 & 0 & 0.344 & 0.219 \\ -0.5 & 0 & 0 & 0 & 0.219 & 0.344 \end{bmatrix}$$

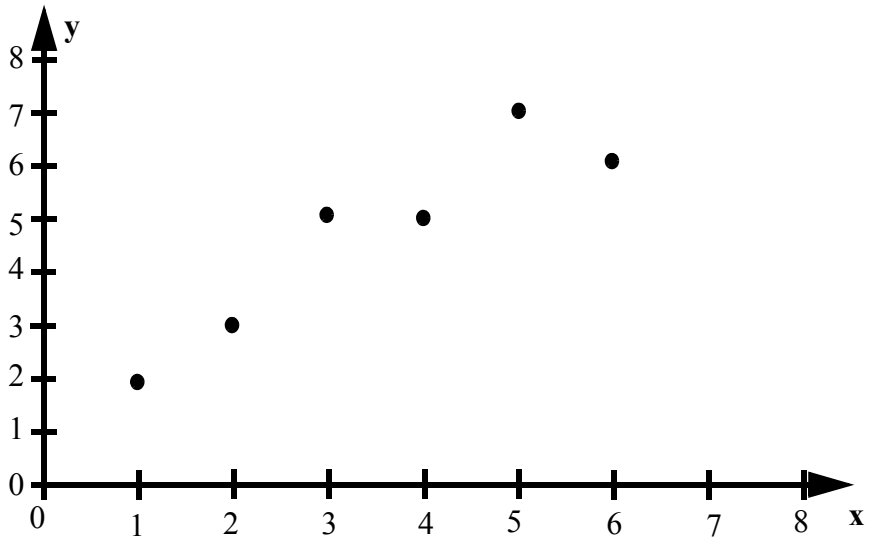
$$(\underset{\sim}{x}^T\underset{\sim}{y}) = \begin{bmatrix} 789 \\ -13.657 \\ -15.556 \\ -4 \\ 680 \\ 714 \end{bmatrix}$$

$$\underset{\sim}{b} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_{12} \\ b_{11} \\ b_{22} \end{bmatrix} = (\underset{\sim}{x}^T\underset{\sim}{x})^{-1}\underset{\sim}{x}^T\underset{\sim}{y} = \begin{bmatrix} 92 \\ -1.707 \\ -1.945 \\ -1.000 \\ -4.563 \\ -0.313 \end{bmatrix}$$

The fitted model is shown in the graph below



Linear Regression



To describe the data above, propose the model:

$$y = B_0 + B_1 x + \varepsilon$$

Fitted model will then be $\hat{y} = b_0 + b_1 x$

Want to select values for $b_0 \& b_1$ that minimize $\sum_{i=1}^{n=6} (y_i - \hat{y}_i)^2$.

Define $S(b_0, b_1) = \sum_{i=1}^{n=6} (y_i - \hat{y}_i)^2$: the model residual Sum of Squares.

Minimize $S(b_0, b_1) = \sum_{i=1}^{n=6} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n=6} (y_i - b_0 - b_1 x_i)^2$

To find minimum S, take partial derivatives of S with respect to $b_0 \& b_1$, set these equal to zero, and solve for $b_0 \& b_1$.

$$\frac{\partial}{\partial b_0} S(b_0, b_1) = 2\Sigma(y_i - b_0 - b_1 x_i)(-1) = 0$$

$$\frac{\partial}{\partial b_1} S(b_0, b_1) = 2\Sigma(y_i - b_0 - b_1 x_i)(-x_i) = 0$$

$$-\Sigma y_i + \Sigma b_o + \Sigma b_1 x_i = 0$$

$$-\Sigma x_i y_i + \Sigma b_o x_i + \Sigma b_1 x_i^2 = 0$$

Simplifying, we obtain:

$$n b_0 + b_1 \Sigma x_i = \Sigma y_i$$

$$b_0 \Sigma x_i + b_1 \Sigma x_i^2 = \Sigma x_i y_i$$

These two equations are known as "Normal Equations".
The values of $b_0 \& b_1$ that satisfy the Normal Equations are the least squares estimates -- they are the values that give a minimum S.

In matrix form

$$\begin{bmatrix} n & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{bmatrix} \begin{Bmatrix} b_0 \\ b_1 \end{Bmatrix} = \begin{Bmatrix} \Sigma & y_i \\ \Sigma & x_i y_i \end{Bmatrix}$$

$$\begin{Bmatrix} b_0^* \\ b_1^* \end{Bmatrix} = \text{Least Squares Estimates} = \begin{bmatrix} n & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^{\,2} \end{bmatrix}^{-1} \begin{Bmatrix} \Sigma & y_i \\ \Sigma & x_i y_i \end{Bmatrix}$$

$$\begin{Bmatrix} b_0^* \\ b_1^* \end{Bmatrix} = \frac{1}{n\Sigma x_i^{\,2} - (\Sigma x_i)^2} \begin{bmatrix} \Sigma x_i^{\,2} & -\Sigma x_i \\ -\Sigma x_i & n \end{bmatrix} \begin{Bmatrix} \Sigma & y_i \\ \Sigma & x_i y_i \end{Bmatrix}$$

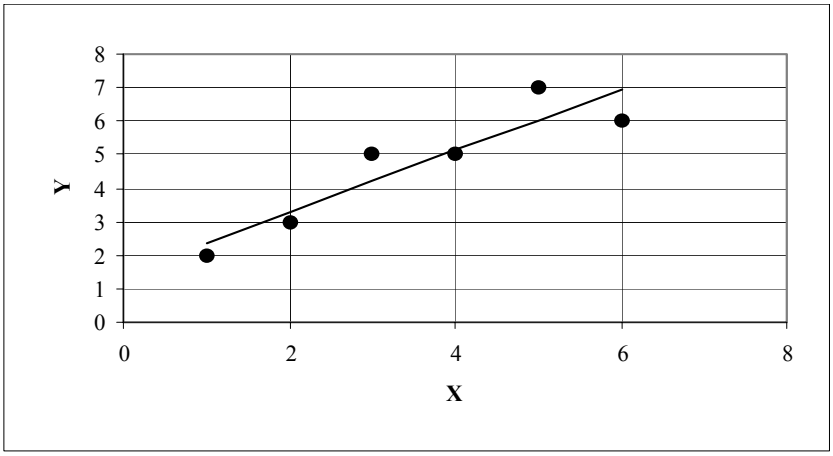
$$= \frac{1}{n\Sigma x_i^{\,2} - (\Sigma x_i)^2} \begin{bmatrix} \Sigma x_i^{\,2} \Sigma y_i - \Sigma x_i \Sigma x_i y_i \\ -\Sigma x_i \Sigma y_i + n\Sigma x_i y_i \end{bmatrix}$$

$$\begin{Bmatrix} b_0^* \\ b_1^* \end{Bmatrix} = \begin{bmatrix} 1.4667 \\ 0.9143 \end{bmatrix}$$

b_0^*, b_1^* are the values of b_0 & b_1 that minimize S, the Residual Sum of Squares.

$b_0^* = \hat{B}_0$ = an estimate of B_0

$b_1^* = \hat{B}_1$ = an estimate of B_1



Matrix Approach

$\underset{\sim}{y}$: Vector of Observations = $\begin{bmatrix} 2 \\ 3 \\ 5 \\ 5 \\ 7 \\ 6 \end{bmatrix}$,

$\underset{\sim}{x}$: Matrix of Independent Variables (Design Matrix) = $\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \\ 1 & 6 \end{bmatrix}$

$\underset{\sim}{\hat{y}}$ = Vector of Observations = $\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix} = \underset{\sim}{x}\underset{\sim}{b}$

$\underset{\sim}{b}$ coefficients = $\begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$

$\underset{\sim}{e}$ = Vector of Prediction Errors = $\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$

$\underset{\sim}{e} = \underset{\sim}{y} - \underset{\sim}{\hat{y}}$

Want to Min $\underset{\sim}{e}^T \underset{\sim}{e}$ or Min $(\underset{\sim}{y} - \underset{\sim}{\underset{\sim}{x}}\underset{\sim}{b})^T(\underset{\sim}{y} - \underset{\sim}{\underset{\sim}{x}}\underset{\sim}{b})$

Take derivative with respect to b's and set = 0

$$-\underset{\sim}{\underset{\sim}{x}}^T(\underset{\sim}{y} - \underset{\sim}{\underset{\sim}{x}}\underset{\sim}{b}) = 0 = -\underset{\sim}{\underset{\sim}{x}}^T\underset{\sim}{y} + (\underset{\sim}{\underset{\sim}{x}}^T\underset{\sim}{\underset{\sim}{x}})\underset{\sim}{b}$$

$$(\underset{\sim}{\underset{\sim}{x}}^T\underset{\sim}{\underset{\sim}{x}})\underset{\sim}{b} = \underset{\sim}{\underset{\sim}{x}}^T\underset{\sim}{y}$$

Therefore,

$$\underset{\sim}{b} = (\underset{\sim}{\underset{\sim}{x}}^T\underset{\sim}{\underset{\sim}{x}})^{-1}\underset{\sim}{\underset{\sim}{x}}^T\underset{\sim}{y}$$

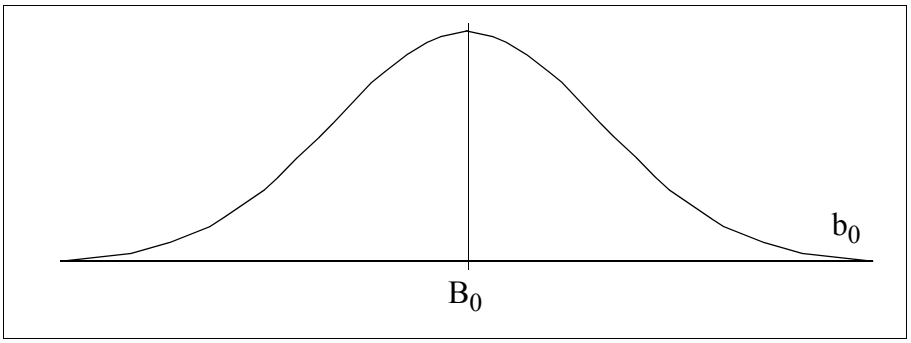
It is analogous to $\begin{Bmatrix} b_0 \\ b_1 \end{Bmatrix} = \begin{bmatrix} n & \Sigma x \\ \Sigma x & \Sigma x^2 \end{bmatrix}^{-1} \begin{Bmatrix} \Sigma y \\ \Sigma xy \end{Bmatrix}$

Re-run experiments several times

$$\begin{Bmatrix} b_0 \\ b_1 \end{Bmatrix} = \begin{Bmatrix} 1.4667 \\ 0.9143 \end{Bmatrix}, \begin{Bmatrix} b_0 \\ b_1 \end{Bmatrix} = \begin{Bmatrix} 1.5309 \\ 0.9741 \end{Bmatrix}, \begin{Bmatrix} b_0 \\ b_1 \end{Bmatrix} = \begin{Bmatrix} 1.5512 \\ 1.0134 \end{Bmatrix}$$

If true model is $y = B_0 + B_1 x + \varepsilon$

Then $E(b_0) = B_0$, $E(b_1) = B_1$, $E[\underset{\sim}{b}] = \underset{\sim}{B}$



$$Var(\underset{\sim}{b}) = (\underset{\sim}{x}^T\underset{\sim}{x})^{-1}\sigma_y^2$$

where σ_y^2 describes the experimental error variation in the y's(σ_ε^2).

For our example, $Var(\underset{\sim}{b}) = \begin{bmatrix} \text{Var}(b_0) & \text{Cov}(b_0, b_1) \\ \text{Cov}(b_0, b_1) & \text{Var}(b_1) \end{bmatrix}$.

If σ_y^2 (or σ_ε^2) is unknown, we can estimate it with

$$s^2 = \frac{(y-\hat{y})^T(y-\hat{y})}{(\text{n - \# of parameters})} = \frac{e^Te}{n-p} = \frac{S_{res}}{\nu}$$

and for the example,

$$s_y^2 = \frac{\sum_{i=1}^{n=6}(y_i-\hat{y}_i)^2}{n-2} = 0.67619$$

$$\widehat{\text{Var}}(\underset{\sim}{b}) = (\underset{\sim}{x}^T\underset{\sim}{x})^{-1}s_y^2 = \begin{bmatrix} 0.586 & -0.135 \\ -0.135 & 0.039 \end{bmatrix}$$

$s_{b_0}^2 = 0.586$, $s_{b_0} = 0.767$ standard error of b_0

$s_{b_1}^2 = 0.039$, $s_{b_1} = 0.197$ standard error of b_1

Marginal Confidence Interval for Parameter Values

Given our parameter estimates, we can develop a 100(1-α)% confidence intervals for the unknown parameter values. The confidence intervals take

the form:

$$b_i \pm t_{\nu,\, 1-\frac{\alpha}{2}}\ s_{b_i}$$

For our example ($t_{4,0.975} = 2.776$),

For B_0: $1.4667 \pm (2.776)(0.767)$, 1.4667 ± 2.125

For B_1: $0.9143 \pm (2.776)(0.197)$, 0.9143 ± 0.546

Note that these confidence intervals do not consider the fact that the parameter estimates may be correlated.

Joint Confidence Interval for Parameter Values

Joint Confidence Interval (approximately) is bounded by a Sum of Squares Contour, S_0. Consider all coefficients simultaneously,

$$S_0 = S_R\left(1 + \frac{p}{n-p} F_{p,\, n-p,\, 1-\alpha}\right)$$

where

$$S_R = S(\underset{\sim}{b}) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Sum of Squares contours are only circles if $b_0 \& b_1$, are uncorrelated - $(x^T x)^{-1}$ is diagonal.

Sum of Squares often contours appear as ellipses.

Confidence Interval for Mean Response

The predicted response at an arbitrary point, $\underset{\sim}{x}_o$, is $\hat{y}_0 = \underset{\sim}{x}_o \underset{\sim}{b}$. Due to the uncertainty in the b's, there is uncertainty in $\hat{y}_o$ (remember $\hat{y}_o$ is the best estimate available for the mean response at $\underset{\sim}{x}_o$).

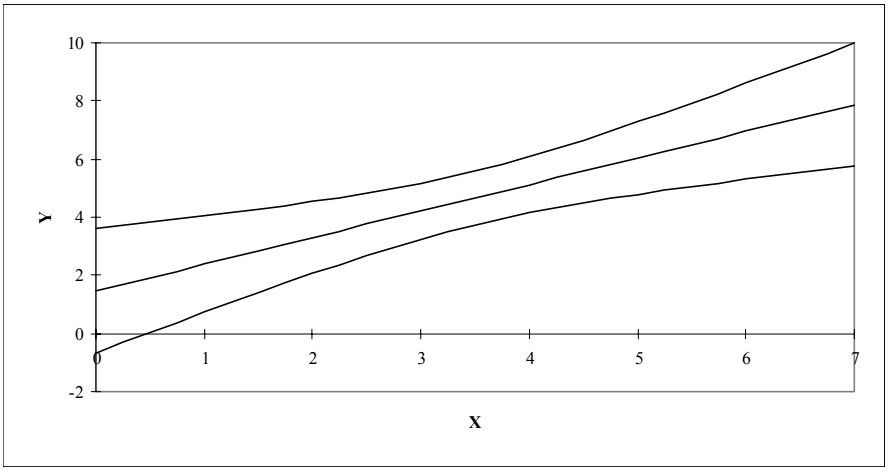
$$\mathrm{Var}(\hat{y}_0) = x_0 (x^T x)^{-1} x_0^T \sigma_y^2$$

of course, if the variance must be estimated from the data, we obtain:

$$\widehat{\mathrm{Var}}(\hat{y}_0) = x_0(x^T x)^{-1} x_0^T s_y^2$$

We can develop a 100(1 – α)% C.I. for the true mean response at x_0 (μ_0):

$$\hat{y} \pm t_{\alpha,\, 1-\frac{\alpha}{2}} [\widehat{Var}(\hat{y}_0)]^{\frac{1}{2}}$$



Prediction Interval

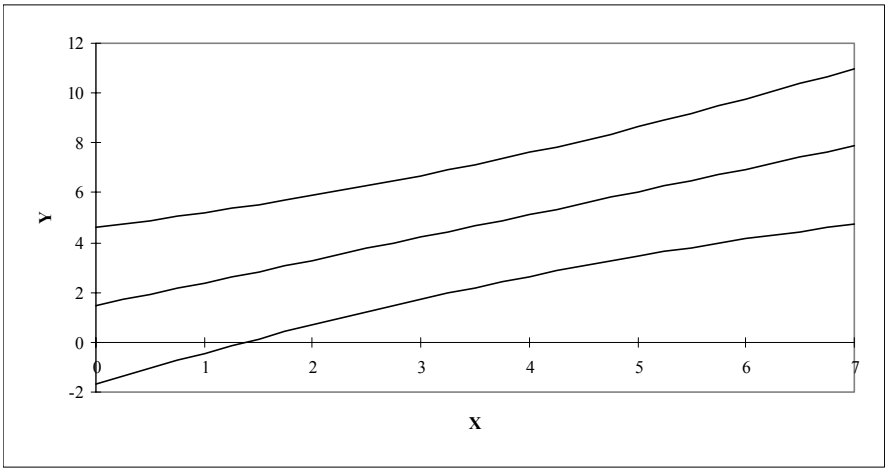
Even if we know with certainty the $\hat{y}_o$ for a given x_0, there would still be variation in the responses due to system noise(ε).

The variance of the mean of g future trials at $\underset{\sim}{x}_0$ is:

$$\widehat{Var}(\overline{y_0}) = \frac{s^2}{g} + \widehat{Var}(\hat{y}_0)$$

We can define limits within which the mean of g future observations will fall. These limits for a future outcome are known as a 100 (1-α)% prediction interval.

$$\hat{y}_0 \pm t_{\nu,\,1-\frac{\alpha}{2}}\left[\frac{s^2}{g} + \widehat{Var}(\hat{y}_0)\right]^{\frac{1}{2}}$$



Model Adequacy Checking

1. Residual Analysis

The model adequacy can be checked by looking at the residuals. The normal probability plot should not show violations of the normality assumption. If there are items not included in the model that are of potential interest, then the residuals should be plotted against these omitted factors. Any structure in such a plot would indicate that the model could be improved by the addition of that factor.

2. F test

If the sum of square of the residuals of the model is S_0. The sum of squares of the residuals of a new model with s factors added is S_1, then

$$\frac{S_0/\nu_0}{S_1/\nu_1} \sim F(\nu_0, \nu_1)$$

and

$$\nu_0 = n - p$$

$$\nu_1 = n - p - s$$

where n is the number of total observations, and p is the number of parameters in the original model.