ME430 Statistics Review

Last Updated October 16 2003

RSS Experiment Design

Model: $y=f(x_1,x_2,..., x_i,..., x_n)$. y is dependent variable, x_i, i=1,..n are independent variables.

Independent variables specified by one-standard deviation tolerances ε_{xi}.

Nominal measurement of independent variables $\bar{x}_i$, i=1,...n

Propagated uncertainty

$$\varepsilon_y = \left[\left(\frac{\partial f}{\partial x_1}\Big|_{x=\bar{x}}\ \varepsilon_{x_1}\right)^2 + \left(\frac{\partial f}{\partial x_2}\Big|_{x=\bar{x}}\ \varepsilon_{x_2}\right)^2 + ...+\left(\frac{\partial f}{\partial x_n}\Big|_{x=\bar{x}}\ \varepsilon_{x_n}\right)^2\right]^{1/2}$$

Percent Contribution of Uncertainty as Design Tool

$$\%\text{ Contribution from measurement } x_i \quad = \frac{\left(\frac{\partial f}{\partial x_i}\Big|_{x=\bar{x}}\ \varepsilon_{x_i}\right)^2}{\varepsilon_y^2}\bullet 100$$

Structural Sensitivity

$$\left|\frac{\partial f}{\partial x_i}\Big|_{x=\bar{x}}\right|$$

When Used: RSS is used when a mathematical model is available to relate dependent variables to independent variables.

Confidence Intervals

Confidence Interval on One Mean:

$$\Pr\left\{\bar{y} - t_{\alpha,\upsilon}\frac{s}{\sqrt{N}} < \mu < \bar{y} + t_{\alpha,\upsilon}\frac{s}{\sqrt{N}}\right\} = (1-2\alpha)100$$

Degrees of Freedom $\nu=N$-1

Sample Average

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$$

Sample Standard Deviation

$$s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(y_i - \overline{y})^2}$$

Confidence Interval for Difference Between Two Means

$$\Pr\left\{(\overline{y}_1 - \overline{y}_2) - t_{\alpha,\upsilon}s_p\sqrt{\frac{1}{N_1}+\frac{1}{N_2}} < \mu_1 - \mu_2 < (\overline{y}_1 - \overline{y}_2) + t_{\alpha,\upsilon}s_p\sqrt{\frac{1}{N_1}+\frac{1}{N_2}}\right\} = (1-2\alpha)100$$

Pooled Variance

$$s_p^2 = \frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2}$$

Degrees of Freedom $\nu = N_1 + N_2 - 2$

When Used: Used when cause-and-effect is only determined by experimental data.

Calibration/Regression

Model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_{k+1} x_k$

k Independent variables $x_1, x_2, ..., x_k$. Considered known with infinite precision

Dependent variable y, subject to measurement uncertainty.

n Data "Pairs" Assumed Measured. Data "Pair" = $(y, x_1, x_1, ..., x_k)$. Data placed into matrices

$$\{y\} = \begin{Bmatrix} y_1 \\ y_2 \\ .. \\ y_n \end{Bmatrix} \quad [X] = \begin{bmatrix} 1 & x_{11} & x_{21} & .. & x_{k1} \\ 1 & x_{12} & x_{22} & .. & x_{k2} \\ .. & .. & .. & .. & .. \\ 1 & x_{1n} & x_{2n} & .. & x_{kn} \end{bmatrix}$$

Estimate for Fit Coefficients

$$\hat{\beta} = \left[X^T X\right]^{-1} X^T y$$

Sum-Square Residuals

$$SS_E = y^T y - \hat{\beta}^T X^T y$$

Sample Estimate of Residual Standard Deviation

$$\hat{\sigma}^2 = \frac{SS_E}{n-p} \qquad p = k+1$$

Covariance Matrix* $[C]=[X^TX]^{-1}$

Confidence Interval for Fit Coefficients

$$\Pr\left\{\hat{\beta}_j - t_{\alpha,\upsilon}\sqrt{\hat{\sigma}^2 C_{jj}} < \beta_j < \hat{\beta}_j + t_{\alpha,\upsilon}\sqrt{\hat{\sigma}^2 C_{jj}}\right\} = (1-2\alpha)100$$

Degrees of Freedom $\nu = n - p$

Prediction Estimate

$$\hat{y} = x_o{}^T \hat{\beta}$$

where $x_o{}^T=\{1 \;\; x_{o1} \;\; x_{o2} \;\; ..... \;\; x_{ok}\}$.

Prediction Interval

$$\hat{y} = x_o{}^T \hat{\beta}, \quad \Pr\left\{\hat{y} - t_{\alpha,\upsilon}\sqrt{\hat{\sigma}^2\left(1 + x_o{}^T C x_o\right)} < y < \hat{y} + t_{\alpha,\upsilon}\sqrt{\hat{\sigma}^2\left(1 + x_o{}^T C x_o\right)}\right\} = (1-2\alpha)100$$

Calibration Problem, Uncertainty on x given y

$$\varepsilon_x = \left[\left(\frac{1}{\hat{\beta}_1^2}\left(y - \hat{\beta}_o\right)\varepsilon_{\hat{\beta}_1}\right)^2 + \left(\frac{1}{\hat{\beta}_1}\varepsilon_{\hat{\beta}_0}\right)^2 + \left(\frac{1}{\hat{\beta}_1}\varepsilon_y\right)^2\right]^{1/2}$$

For tolerances, use

$$\varepsilon_{\beta_0} = \frac{1}{2} t_{\alpha,\upsilon}\sqrt{\hat{\sigma}^2 C_{00}}\,, \quad \varepsilon_{\beta_1} = \frac{1}{2} t_{\alpha,\upsilon}\sqrt{\hat{\sigma}^2 C_{11}}$$

Regression used when a detailed numerical relationship between independent and dependent variables is desired.

* Our text denotes the quantity $\hat{\sigma}^2[C]$ as the covariance matrix.

Factorial Experiment Design and ANOVA

Design Grid; Factors (A,B) , Levels (a=2, b=2) and Replications (n). Total number of experiments=$n2^k$

Main Effects Factors

$$\text{Main Effect A} = \left[\frac{ab-b}{n} + \frac{a-(1)}{n}\right]\frac{1}{2}, \quad \text{Main Effect } B = \left[\frac{ab-a}{n} + \frac{b-(1)}{n}\right]\frac{1}{2}$$

Interactions Effect

$$\text{Interaction Effect AB} = \left[\frac{ab-b}{n} - \frac{a-(1)}{n}\right]\frac{1}{2}$$

ANOVA Table and Interpretation

Source of Variation	*Degrees of Freedom*	*Mean Square*	*F*
A	a-1	MS_A	$F=MS_A/MS_E$
B	b-1	MS_B	$F=MS_B/MS_E$
AB	(a-1)(b-1)	MS_{AB}	$F=MS_{AB}/MS_E$
Error	$ab(n-1)$	MS_E	

If $F>F_{\alpha,\nu1,\nu2}$, then probability is less than $\alpha\bullet 100$ % that the observed F is due to statistical anomaly.

Factorial experiments are used when it is desired to learn of cause-and-effect relationships between independent variables and dependent variables using a minimum amount of data.