Journal of the Korean Institute of Industrial Engineers
Vol. 17, No. 2, Dec., 1991

표준자료 산출시 작업특성치의 오차가 총작업시간의 예측에 미치는 영향평가

변재현* · 염봉진**

Evaluation of the Effect of Errors in Job Characteristics on the Predicted Total Task Time in Standard Data Systems

Jai-Hyun Byun* and Bong-Jin Yum**

Abstract

In developing a regression relationship for a standard data system in work measurement, job characteristics are frequently measured with error when measurements are made in the field under less controlled conditions or when accurate instruments are not available. This paper concerns with the prediction of the total task time when job characteristics are measured with error. Integrated mean square error of prediction(IMSE) is developed as a measure of the effect of errors in job characteristics on the predicted total task time. By evaluating how IMSE is affected by the measurement error in each job characteristic, we can determine which error should be controlled to develop a desirable standard data system.

1. 서 론

표준자료를 작성하기 위해서는 우선 작업을 몇 개의 작업요소(activity element)로 구분하여, 각 작업요소에 대하여 예측되어야 할 종속변수(즉, 작업요소의 정규시간)와 이에 영향을 미치는 독립변수(즉, 작업특성, job characteristic)간의 관계식을 추정할 필요가 있으며, 이러한 추정에 회귀분석이 종종 사용된다. 관계식이 추정되면 이것을 이용하여 새로운 작업의 작업특성치를 바탕으로 각 작업요소의 정규시간, 궁극적으로 총작업시간을 예측할 수 있게 된다. 예를 들어서, 표 1의 경우,

* 경상대학교 산업공학과
** 한국과학기술원 산업공학과

총작업은 4개의 작업요소로 이루어져 있고, 각 요소는 작업특성에 의해 영향을 받는데, 요소 "a", "b", 그리고 "c"는 작업특성 1에, 요소 "d"는 작업특성 1과 2에 영향을 받는다고 하자. 이런 상황에서 우리는 작업요소 a, b, c 그리고 d의 정규시간의 합인 총작업시간(total task time)의 예측에 관심을 가지게 된다.

표 1. 표준자료체계의 예

작업요소	a	b	c	d
작업특성	1	1	1	1, 2

한편, 정확한 측정이 이루어지지 못하는 생산현장이나 정확한 측정자체가 불가능할 경우에는 작업특성치의 측정에 오차가 수반된다. 이 때에는 미래의 새로운 작업의 정규시간을 예측하는데 있어서 각 작업특성(독립변수)에 포함되어 있는 측정오차가 예측치의 정확성에 끼치는 영향이 어느 정도인가를 파악하여 예측에 큰 영향을 미치는 작업특성의 오차를 집중적으로 통제할 필요가 있다. 예를 들어, Fein[2]은 작업측정자료가 1일 업무량을 결정하거나 장려급 설정에 이용될 때는 가능한한 정확한 수치가 필요하다고 지적하고 있다.

회귀분석시 독립변수에 측정오차가 있을 때의 예측에 관한 문제는 Lindley[4], Ganse 등[3], Yum과 Neuhardt[8], 그리고 Yum과 Byun[7]이 다루었다. 특히, Yum과 Byun은 독립변수가 2개 이상인 경우, 각 변수의 측정오차가 예측의 정확성에 미치는 영향을 평가하기 위하여 통합된 예측 평균제곱오차(Integrated Mean Square Error of Predicition, IMSE)를 개발하여 이 결과를 표준자료 산출에 적용하였다. 그들은 각 작업특성의 오차가 하나의 특정한 작업요소의 정규시간 예측치에 미치는 영향을 예를 통하여 보여 주었다.

본 연구의 목적은 표준자료 산출시 작업특성이 오차를 수반하여 측정될 때, 이러한 오차가 총작업시간의 예측에 미치는 영향을 정량적으로 파악하기 위한 방법을 마련하는 데에 있다. 이 목적을 위해, Yum과 Byun이 개발한 하나의 작업요소에 대한 IMSE의 결과를 그대로 적용할 수는 없다. 예를 들어, 표준자료체계가 표 1과 같은 경우에, 총작업시간의 IMSE가 작업요소 a, b, c 그리고 d의 IMSE의 합으로 표현될 수 없기 때문이다. 왜냐하면, 각 작업요소에 대한 회귀관계식의 추정은 각 요소에 대하여 개별적으로 이루어지지만, 총작업시간의 정규시간을 예측할 때에는 한 작업특성이 둘 이상의 작업요소에 영향을 미침으로써 예측오차의 공분산 행렬(Covariance Matrix)이 일반적으로 대각행렬이 되지 않기 때문이다.

본 논문은 다음과 같이 구성되어 있다. 2절에서는 모형의 정의와 가정을 제시하고, 3절에서는 총작업시간의 예측의 정확성 평가를 위한 IMSE를 개발하며, 4절은 3절에서 개발한 IMSE의 추정을 다루고, 5절에서는 3절에서 개발한 IMSE의 관점에서 각 작업특성의 측정오차가 예측치의 정확성에 미치는 영향에 대한 평가를 예를 통해서 살펴본다.

2. 추정실험과 예측에 관한 모형

작업요소의 수가 m개일 때, i번째 작업요소에 대하여 종속변수(작업요소의 정규시간)와 이에 영향을 주는 독립변수(작업특성)간의 관계식을 다음과 같이 표현할 수 있다고 하자.

$$y_i = \beta_{i0} + \beta_{i1}\zeta_{i1} + \beta_{i2}\zeta_{i2} + \cdots + \beta_{ip_i}\zeta_{ip_i} + v_i \quad \cdots\cdots (1)$$

여기서 y_i는 종속변수, p_i는 독립변수의 수, β_{ij}는 미지의 모수, ζ_{ij}는 독립변수($j=1, 2, \cdots, p_i$), 그리고 v_i는 i번째 관계식에 고유한 산포이다. i번째 관계식 (1)을 추정하기 위해 i번째 작업요소의 모집단으로 부터 n_i개의 표본을 취하여 각각에 대해 ζ_{ijk}, y_{ik}($k=1, 2, \cdots, n_i$)를 측정한다고 하자. 이

때, ζ_{ijk}의 측정에는 측정오차 u_{ijk}가 수반되어 x_{ijk}로 관측된다고 하면 다음과 같은 모형을 생각할 수 있다.

$$y_{ik}=\beta_{i0}+\beta_{i1}\zeta_{i1k}+\beta_{i2}\zeta_{i2k}+\cdots+\beta_{ip_i}\zeta_{ip_ik}+v_{ik}$$
$$x_{ijk}=\zeta_{ijk}+u_{ijk} \quad \cdots\cdots (2)$$

단, $i=1, 2, \cdots, m$; $j=1, 2, \cdots, p_i$;

$k=1, 2, \cdots, n_i$.

주어진 i, k에 대해 독립변수의 오차백터를 다음과 같이 정의하자.

$$u_{ik}=(0, u_{i1k}, u_{i2k}, \cdots, u_{ip_ik})'.$$

첫번째 요소를 0으로 한 것은 식(2)의 우변의 첫째항 β_{i0}를 $\beta_{i0}\zeta_{i0k}$, $\zeta_{i0k}=1$로 보고 ζ_{i0k}의 오차 u_{i0k}를 0으로 간주했기 때문이다. 그리고 백터$(v_{ik}, u_{ik}')'$는 평균이 0백터이고 다음과 같은 공분산 행렬(covariance matrix)을 갖는다고 가정한다.

$$\mathrm{Cov}(v_{ik}, u_{ik}')'=\begin{bmatrix}\delta_i^2 & 0\\ 0 & \Sigma_i\end{bmatrix}$$

단,

$$\delta_i^2=\mathrm{Var}(v_{ik})$$
$$\Sigma_i=\mathrm{Cov}(u_{ik})=\mathrm{diag}(0, \sigma_{i1}^2, \sigma_{i2}^2, \cdots, \sigma_{ip_i}^2).$$
$$\sigma_{ij}^2=\mathrm{Var}(u_{ijk}),\ j=1, 2, \cdots, p_i.$$

추정실험에서 i번째 관계식은 다른 관계식에 대해 독립적인 실험을 통해 추정된다고 가정한다. i번째 관계식의 추정을 위해 독립변수와 종속변수의 관측치에 대한 행렬과 백터를 다음과 같이 정의한다.

$$X_i=\begin{bmatrix}1 & x_{i11} & x_{i21} & \cdots & x_{ip_i1}\\ 1 & x_{i12} & x_{i22} & \cdots & x_{ip_i2}\\ \cdot & \cdot & \cdot & \cdots & \cdot\\ \cdot & \cdot & \cdot & \cdots & \cdot\\ \cdot & \cdot & \cdot & \cdots & \cdot\\ 1 & x_{i1n_i} & x_{i2n_i} & \cdots & x_{ip_in_i}\end{bmatrix}$$

$$y_i=(y_{i1}, y_{i2}, \cdots, y_{in_i})'$$

아울러 관계식 (1) 또는 (2)에서 미지의 모수들로 이루어진 백터 β를 다음과 같이 정의한다.

$$\beta=(\beta_1', \beta_2', \cdots, \beta_m')'$$
$$\beta=(\beta_{i0}, \beta_{i1}, \beta_{i2}, \cdots, \beta_{ip_i})'.$$

그러면, β_i에 대한 보통최소제곱(Ordinary Least Squares, OLS) 추정치 b_i는 다음과 같이 주어진다.

$$b_i=(X_i'X_i)^{-1}X_iy_i,\ i=1, 2, \cdots, m.$$

따라서, β에 대한 OLS 추정치 b는

$$b=(b_1', b_2' \cdots, bm')'$$

가 된다.

지금까지 i번째 관계식에 대한 모형과 추정방법에 대해 살펴 보았는데, 이제 각 작업요소의 정규시간의 합인 총작업시간의 예측을 위한 모형에 관하여 살펴보자.

미래의 어떤 작업 f의 i번째 작업요소에 대하여 작업특성치 ζ_{ijf}와 y_{if} 사이에 다음과 같은 관계가 성립한다고 하자(식 (2) 참조).

$$y_{if}=\beta_{i0}+\beta_{i1}\zeta_{i1f}+\beta_{i2}\zeta_{i2f}+\cdots+\beta_{ip_i}\zeta_{ip_if}+v_{if}.$$
$$i=1, 2, \cdots, m.$$

그러면, 총작업시간 t_f는

$$t_f=y_{1f}+y_{2f}+\cdots+y_{mf}$$
$$=\beta'\zeta_f+\varepsilon_f \quad \cdots\cdots (3)$$

로 주어진다. 단,

$$\zeta_f=(\zeta_{1f}', \zeta_{2f}', \cdots, \zeta_{mf}')'$$
$$\zeta_{if}=(1, \zeta_{i1f}, \zeta_{i2f}, \cdots, \zeta_{ip_if})'$$
$$\varepsilon_f=v_{1f}+v_{2f}+\cdots+v_{mf}$$

이다. t_f를 추정하기 위해, $\zeta_{ijf}(i=1, 2, \cdots, m$; $j=1, 2, \cdots, p_i)$를 측정함에 있어 추정실험시와 마찬가지로 다음과 같이 측정오차 u_{ijf}가 수반된다고 가정한다.

$$x_{ijf}=\zeta_{ijf}+u_{ijf}.$$

아울러, v_{if}는 평균 0, 분산 δ_i^2, u_{ijf}는 평균 0, 분산

σ_{ij}^2, 그리고 v_{it}와 u_{ijt}는 서로 상관관계가 없다고 가정한다. 독립변수의 오차에 대해 다음 백터를 정의하자.

$u_f=(u_{1f}', u_{2f}', \cdots, u_{mf}')'$
$u_{if}=(0, u_{i1f}, u_{i2f}, \cdots, u_{ip_if})'.$

u_f는 평균이 0백터이고 공분산 행렬이 Σ_f인 것으로 가정한다. 여기서 Σ_f는 일반적으로 대각행렬이 되지 않음에 주의할 필요가 있다. 왜냐하면, 어떤 독립변수(작업특성)가 두 개 이상의 종속변수(작업요소의 정규시간)에 관련되어 있을 때 그 독립변수를 모든 관련 관계식마다 따로 따로 독립적으로 측정하는 것이 아니라 한 번 측정한 값을 공통적으로 사용하므로 한 관계식의 독립변수의 측정오차가 다른 관계식의 해당 독립변수의 측정오차와 상관되어 있기 때문이다. 예를 들어, m=2일 때 독립변수 ζ_1이 두 관계식에 영향을 미친다고 하자. 즉,

$y_{1f}=\beta_{10}+\beta_{11}\zeta_{11f}+v_{1f},$
$y_{2f}=\beta_{20}+\beta_{21}\zeta_{21f}+v_{2f},$

그리고, 독립변수의 관측치는 수리적으로는

$$x_{11f}=\zeta_{11f}+u_{11f} \quad \cdots\cdots (4)$$
$$x_{21f}=\zeta_{21f}+u_{21f} \quad \cdots\cdots (5)$$

이다. 그러나, 식 (4)와 (5)는 동일한 ζ_f에 대한 측정상황을 나타내는 것으로서 사실은 $\zeta_{11f}=\zeta_{21f}$, $u_{11f}=u_{21f}$, $x_{11f}=x_{21f}$이다. 따라서, $u_f=(0, u_{11f}, 0, u_{21f})'$의 공분산 행렬 Σ_f는

$$\Sigma_f=\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \mathrm{Var}(u_{11f}) & 0 & \mathrm{Cov}(u_{11f}, u_{21f}) \\ 0 & 0 & 0 & 0 \\ 0 & \mathrm{Cov}(u_{11f}, u_{21f}) & 0 & \mathrm{Var}(u_{21f}) \end{bmatrix}$$

$$\Sigma_f=\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \sigma_1^2 & 0 & \sigma_1^2 \\ 0 & 0 & 0 & 0 \\ 0 & \sigma_1^2 & 0 & \sigma_1^2 \end{bmatrix}$$

이 되어, 대각행렬이 되지 않는다.

식 (3)으로 부터 t_f의 최선의 예측치는 $\beta'\zeta_f$임을 알 수 있다. 그러나, β와 ζ_f는 알 수 없으므로 t_f의 예측치로서

$$\hat{t}_f=b'x_f \quad \cdots\cdots (6)$$

로 삼는다. 식 (6)에서 b는 추정실험을 통해 구한 β의 OLS 추정 백터이며,

$x_f=(x_{1f}', x_{2f}', \cdots, x_{mf}')'$
$x_{if}=(1, x_{i1f}, x_{i2f}, \cdots, x_{ip_if})'$

이다. 따라서 식 (6)은 다시 다음과 같이 쓸 수 있다.

$$\hat{t}_f=b_1'x_{1f}+b_2'x_{2f}+\cdots+b_m'x_{mf}$$
$$=\hat{y}_{1f}+\hat{y}_{2f}+\cdots+\hat{y}_{mf}.$$

3. 예측치의 통합평균제곱오차

식 (6)의 $\hat{t}_f$가 t_f와 얼마나 가까운가를 나타내는 척도로서 평균제곱오차(mean square error, MSE)를 고려하고자 한다. 또한, 미래의 여러 작업에 대해 예측이 이루어지므로 예측치들의 평균적 형태를 나타내는 척도로서 소위 통합평균제곱오차 IMSE를 채택하고자 한다.

IMSE를 구하기 전에 우선 x_f에 대한 $\hat{t}_f$의 조건부 MSE를 구해보면

$$\begin{aligned} &\mathrm{MSE}(\hat{t}_f \mid x_f) \\ &=\mathrm{MSE}\Big(\sum_{i=1}^{m}\hat{y}_{if} \mid x_f\Big) \\ &=E\Big\{\Big(\sum_{i=1}^{m}\hat{y}_{if}-\sum_{i=1}^{m}y_{if}\Big)^2 \mid x_f\Big\} \\ &=E\Big[\Big\{\sum_{i=1}^{m}(\hat{y}_{if}-y_{if})\Big\}^2 \mid x_f\Big\} \end{aligned}$$

가 된다. x_f에 대하여 조건부 평균제곱오차의 기대값을 취하면 $\hat{t}_{if}$의 평균제곱오차를 얻을 수 있는데, 그 전에 다음을 정의한다.

$V_i=\mathrm{Cov}(b_i),$
$\phi_i=E(b_i)-\beta_i, \; i=1, 2, \cdots, m.$

$V = \mathrm{diag}(V_1, V_2, \cdots, V_m)$,

$\phi = (\phi_1', \phi_2', \cdots, \phi_m')'$.

Seber[5]의 Theorem 1. 7을 이용하고 수식을 정리하면, $\hat{t}_f$의 MSE는 다음과 같이 된다.

$$\begin{aligned} MSE(\hat{t}_f) &= E[MSE(\hat{t}_f \mid x_f)] \\ &= tr[\{V + (\beta+\phi)(\beta+\phi)'\} \Sigma_f] \\ &\quad + \zeta_f'(V + \phi\phi')\zeta_f + \sum_{i=1}^{m} \delta_i^2 \quad \cdots\cdots\cdots\cdots (7) \end{aligned}$$

예측이 어떤 R이라는 흥미영역(Region of Interest)에서 이루어진다면, 우리는 예측치들의 "평균적" 행태에 관심을 갖게 된다. 독립변수의 측정오차가 예측에 미치는 영향에 대한 평가기준으로 IMSE를 다음과 같이 정의한다.

$$IMSE = \int_R MSE(\hat{t}_f) \quad w(\zeta_f) d\zeta_f \quad \cdots\cdots\cdots\cdots\cdots\cdots (8)$$

여기서 가중함수(weight function) $w(\zeta_f)$는 ζ_f값들의 상대적 중요성(예를 들어, 각 ζ_f값을 갖는 작업들이 출현할 빈도)을 나타내며 다음을 만족한다.

$$\int_R w(\zeta_f) d\zeta_f = 1 \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (9)$$

그리고 다음과 같이 ζ_f의 2차 모멘트가 존재한다고 가정한다.

$$\int_R \zeta_f \zeta_f' \; w(\zeta_f) d\zeta_f = M \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (10)$$

식 (7)－(10)을 조합하면 다음과 같이 IMSE를 구할 수 있다.

$$\begin{aligned} IMSE &= tr\{(V + \phi\phi')(\Sigma_f + M)\} + \beta'\Sigma_f\beta \\ &\quad + 2\beta'\Sigma_f\phi + \sum_{i=1}^{m} \delta_i^2 \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots (11) \end{aligned}$$

만약에 독립변수에 측정오차가 없었다면, OLS 추정을 사용한다고 할 때 식 (11)에서 $\phi = 0$, $\Sigma_f = 0$이므로 IMSE는 다음과 같이 표현된다.

$$IMSE_0 = \sum_{i=1}^{m} \delta_i^2 + tr(V_0 M)$$

여기서

$V_0 = \mathrm{diag}(V_{10}, V_{20}, \cdots, V_{m0})$,

$V_{i0} = \delta_i^2 \cdot (\Xi_i' \Xi_i)^{-1}$, $i = 1, 2, \cdots, m$.

$$\Xi_i = \begin{bmatrix} 1 & \zeta_{i11} & \zeta_{i21} & \cdots & \zeta_{ip_i1} \\ 1 & \zeta_{i12} & \zeta_{i22} & \cdots & \zeta_{ip_i2} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 1 & \zeta_{i1n_i} & \zeta_{i2n_i} & \cdots & \zeta_{ip_in_i} \end{bmatrix}$$

4. IMSE의 추정

IMSE의 추정을 위해서는 식(11)의 미지의 값들을 추정해야 하는데, OLS 추정법을 이용했을 때 공분산 행렬 V와 편의(Bias)벡터 ϕ의 근사값은 Davies and Hutton[1]으로 부터 다음과 같이 구할 수 있다.

$\hat{V} = \mathrm{diag}(\hat{V}_1, \hat{V}_2, \cdots, \hat{V}_m)$.

여기서

$$\hat{V}_i \simeq n_i^{-1}\{\delta_i^2(M_{\xi i} + \Sigma_i)^{-1} + (M_{\xi i} + \Sigma_i)^{-1} \; T \; (M_{\xi i} + \Sigma_i^{-1}\}$$

이며

$$\begin{aligned} T = {} & (M_{\xi i} + \Sigma_i) \; \beta_i' M_{\xi i} \; (M_{\xi i} + \Sigma_i)^{-1} \; \Sigma_i (M_{\xi i} + \Sigma_i)^{-1} \\ & M_{\xi i}\beta_i + \Sigma_i \beta_i' M_{\xi i} \; (M_{\xi i} + \Sigma_i)^{-1} \; \Sigma_i M_{\xi i}^{-1} \Sigma_i (M_{\xi i} + \Sigma_i)^{-1} \\ & M_{\xi i}\beta_i - \Sigma_i (M_{\xi i} + \Sigma_i)^{-1} \; M_{\xi i}\beta_i\beta_i' M_{\xi i} \; (M_{\xi i} + \Sigma_i)^{-1} \\ & \Sigma_i, \end{aligned}$$

$$M_{\xi i} = \lim_{n_i \to \infty} n_i^{-1}(\Xi_{ni}'\Xi_{ni}) \; ; \; \Xi_{ni} \equiv \Xi_i \quad \cdots\cdots\cdots\cdots\cdots (12)$$

한편,

$\hat{\phi} = (\hat{\phi}_1', \hat{\phi}_2', \cdots, \hat{\phi}_m')'$

여기서

$\hat{\phi}_i \simeq -(M_{\xi i} + \Sigma_i)^{-1} \Sigma_i \beta_i$, $i = 1, 2, \cdots, m$.

미지의 모수를 추정하기 위해 Seber[5]로 부터 다음의 결과를 얻을 수 있다.

$$E(X_i'X_i/n_i) = (\Xi_i'\Xi_i)/n_i + \Sigma_i, \quad \cdots\cdots\cdots\cdots\cdots\cdots (13)$$

$$E(S_i^2) = \delta_i^2 + \beta_i'\Sigma_i\beta_i, \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (14)$$

여기서

$S_i^2=(y_i-X_ib_i)'(y_i-X_ib_i)/(n_i-p_i-1)$.

식 (12)로 부터 M_{ij}는 $(\Xi_i'\Xi_i)/n_i$로 근사될 수 있으며, 이는 다시 식 (13)를 이용하면 $(X_i'X_i)/n_i-\Sigma_i$로 근사화된다. β_i는 b_i로 추정되며, 식 (14)로 부터 δ_i^2은 $S_i^2-b_i'\Sigma_ib_i$로 근사화된다. 만일 Σ_i 또는 Σ_i의 σ_{ij}^2 $(i=1, 2, \cdots, m; j=1, 2, \cdots, p_i)$이 알려지지 않은 경우, 이들은 과거의 자료나 실험을 통하여 추정되어야 하며, 2차 모멘트 행렬 M은 미래에 관심 있는 영역에서 식(10)을 이용하여 추정되어야 한다.

5. 예 제

실험데이터가 주어지면 4절의 결과를 이용하여 IMSE를 추정할 수 있다. 본 예제에서는 IMSE의 추정보다는, 독립변수의 측정오차에 따른 IMSE의 민감도를 분석하고자 한다. 이를 통하여 어떤 독립변수의 측정오차가 IMSE의 크기에 큰 영향을 미치는가를 파악하여 이들을 집중적으로 통제함으로써 보다 바람직한 예측치를 얻을 수 있는 표준자료시스템을 구축할 수 있게 될 것이다.

독립변수의 측정오차가 총작업시간의 예측에 미치는 영향을 Smith[6]의 수평보오링 밀링작업(Horizontal Boring Mill Operation)의 예를 통해서 알아보기로 한다. 수평 보오링밀링이 끝난 작업물을 하역하는 활동(unloading activity)은 6개의 작업요소로 구성되어 있는데, 2개의 작업요소("크레인 부름, "크레인 정위치")는 어떤 작업에 대해서든지 그 시간이 일정하고, 다른 4개의 작업요소("체인 걸기," "체인길이 조정," "들어서 옆으로 옮김," 그리고 "체인 제거")는 여러가지 작업에 대하여 그 시간이 다르게 나타났다. 작업특성에 따라 변하는 4개의 작업요소중, "들어서 옆으로 옮김"은 작업 대상물의 "무게"와 밀링머시인에서 저장장소까지의 "이동 거리"에 의존하는데 반하여, 나머지 3개의 요소는 "무게"에만 의존하였다.

작업특성에 따라 변하는 작업요소들의 정규시간치들, 즉, "체인걸기"의 y_1, "체인길이 조정"의 y_2, "체인제거"의 y_3, 그리고 "들어서 옆으로 옮김"의 y_4가 작업특성 "무게"의 참값 ζ_1 및 "이동 거리"의 참값 ζ_2와 다음과 같은 관계를 갖는다고 가정하자.

$$y_i=\beta_{i0}+\beta_{i1}\zeta_{i1}+v_i,\ i=1, 2, 3$$
$$y_4=\beta_{40}+\beta_{41}\zeta_{41}+\beta_{42}\zeta_{42}+v_4$$

i번째 작업요소에 대하여, n_i개의 작업을 선택하고 각 작업의 정규시간치를 r_i번 반복하여 관측한다. 관측된 데이터에 대한 표현을 위하여, 주어진 i에 대해서 다음을 정의한다.

ζ_{i1k}=k번째 작업의 "무게"의 참값, $i=1, 2, 3, 4$

ζ_{i2k}=k번째 작업의 "이동 거리"의 참값, $i=4$

u_{i1k}=k번째 작업의 "무게"의 측정오차, $i=1, 2, 3, 4$

u_{i2k}=k번째 작업의 "이동 거리"의 측정오차, $i=4$

v_{ikq}=k번째 작업의 q번째 반복에 대한 정규시간 추정실험에 고유한 변동

그러면 $i=1, 2, 3, 4$, $k=1, 2, \cdots, n_i$, 그리고 $q=1, 2, \cdots, r_i$에 대하여, 우리는 다음과 같은 관측치를 얻게 된다.

$$y_{ikq}=\beta_{i0}+\beta_{i1}\zeta_{i1k}+v_{ikq},\ i=1, 2, 3$$
$$y_{4kq}=\beta_{40}+\beta_{41}\zeta_{41k}+\beta_{42}\zeta_{42k}+v_{4kq} \quad \cdots\cdots\cdots\cdots (15)$$

$$x_{i1k}=\zeta_{i1k}+u_{i1k},\ i=1, 2, 3, 4$$
$$x_{42k}=\zeta_{42k}+u_{42k}$$

분석의 편의상, 각 k에 대해서 y_{ikq}, $q=1, 2, \cdots, r_i$의 표본평균을 취하여 그 값을 $\bar{y}_{ik}$라 하면 식 (15)는 다음과 같이 쓸 수 있다.

$$\bar{y}_{ik}=\beta_{i0}+\beta_{i1}\zeta_{i1k}+\bar{v}_{ik},\ i=1, 2, 3$$
$$\bar{y}_{4k}=\beta_{40}+\beta_{41}\zeta_{41k}+\beta_{42}\zeta_{42k}+\bar{v}_{4k}$$

단,

$$\bar{y}_{ik} = \sum_{q=1}^{r_i} y_{ikq}/r_i,$$
$$\bar{v}_{ik} = \sum_{q=1}^{r_i} v_{ikq}/r_i.$$

제 2절의 추정실험 모형과 연관지어 보면, 식 (2)의 y_{ik}와 v_{ik}가 본 예제에서는 각각 $\bar{y}_{ik}$와 $\bar{v}_{ik}$로 대치되었음을 알 수 있다. 아울러 독립변수의 오차에 대해서는

$$u_{ik}=(0,\ u_{i1k})',\quad i=1,\ 2,\ 3,$$
$$u_{4k}=(0,\ u_{41k},\ u_{42k})',$$
$$\Sigma_i=\mathrm{Cov}(u_{ik})=\mathrm{diag}(0,\ \sigma_{i1}^2),\quad i=1,\ 2,\ 3,$$
$$\Sigma_4=\mathrm{Cov}(u_{4k})=\mathrm{diag}(0,\ \sigma_{41}^2,\ \sigma_{42}^2)$$

이 된다. $\sigma_{i1}^2(i=1,\ 2,\ 3,\ 4)$은 무게의 측정에 따른 오차분산이므로 모든 i에 대해 같으며, 이를 σ_1^2으로 하고, 마찬가지로 σ_{42}^2을 간단히 σ_2^2이라고 하자.

y_{ikq}대신에 $\bar{y}_{ik}$를 종속변수로 하여 β_{i0}, $\beta_{i1}=(i=1,\ 2,\ 3,\ 4)$, β_{42}의 OLS 추정량 b_{i0}, $b_{i1}(i=1,\ 2,\ 3,\ 4)$, b_{42}를 구할 수 있다.

미래의 어떤 작업 f에 대하여 우리는 다음의 총 작업시간을 예측하는 데에 관심을 가지게 된다.

$$t_f=y_{1f}+y_{2f}+y_{3f}+y_{4f}.$$

여기서

$$y_{if}=\beta_{i0}+\beta_{i1}\zeta_{i1f}+v_{if},\quad i=1,\ 2,\ 3,$$
$$y_{4f}=\beta_{40}+\beta_{41}\zeta_{41f}+\beta_{42}\zeta_{42f}+v_{4f}.$$

t_f의 예측을 위해 작업특성인 무게와 이동거리를 측정하게 되는데, 수리적으로는

$$x_{i1f}=\zeta_{i1f}+u_{i1f},\quad i=1,\ 2,\ 3,\ 4,$$
$$x_{42f}=\zeta_{42f}+u_{42f} \quad\cdots\cdots\cdots\cdots\cdots\cdots (16)$$

로 쓸 수 있다. 실제로 작업 f의 무게는 한 번 측정하여 4개의 작업요소에 공통으로 사용하고 이동거리는 4번째 작업요소에만 관계되므로 식 (16)에서 i를 생략하면

$$x_{1f}=\zeta_{1f}+u_{1f},$$
$$x_{2f}=\zeta_{2f}+u_{2f}$$

로 간단히 쓸 수 있다.

$\varepsilon_f=\sum_{i=1}^{m} v_{if}$는 평균 0, 분산 $\sum_{i=1}^{m}\delta_i^2$을 갖고, $u_f=(u_{1f}',\ u_{2f}',\ u_{3f}',\ u_{4f}')'=(0,\ u_{1f},\ 0,\ u_{1f},\ 0,\ u_{1f},\ 0,\ u_{1f},\ u_{2f})'$는 ε_f와 독립이며, 평균 0과 다음과 같은 공분산 행렬을 갖는다고 가정한다.

$$\Sigma_f=\begin{bmatrix} \Sigma_{1f} & & & & \\ \Sigma_{1f} & \Sigma_{1f} & & & \\ \Sigma_{1f} & \Sigma_{1f} & \Sigma_{1f} & & \\ \Sigma_{1f} & \Sigma_{1f} & \Sigma_{1f} & \Sigma_{1f} & \\ 0 & 0 & 0 & 0 & \sigma_2^2 \end{bmatrix}$$

여기서

$$\Sigma_{1f}=\begin{bmatrix} 0 & \\ 0 & \sigma_1^2 \end{bmatrix}$$

그러면 4개의 작업요소의 정규시간은 다음과 같이 예측된다.

$$\hat{y}_{if}=b_{i0}+b_{i1}x_{1f},\quad i=1,\ 2,\ 3$$
$$\hat{y}_{4f}=b_{40}+b_{41}x_{1f}+b_{42}x_{2f}.$$

총작업시간의 예측치 $\hat{t}_f$는

$$\hat{t}_f=\hat{y}_{1f}+\hat{y}_{2f}+\hat{y}_{3f}+\hat{y}_{4f}$$

이 된다. 본 예에서는 IMSE의 추정보다는, 모수의 값이 주어진 상태에서 IMSE가 σ_1^2과 σ_2^2에 어느 만큼 의존하는가를 보이는데 관심이 있다. 추정을 위한 실험에 선택된 작업특성의 참값은 다음의 표 2와 같다고 하자.

표 2. 작업특성의 참값

작업특성 / 작업요소(i)	무게(kg) ξ_1	이동거리(m) ξ_2
1	350 1600 2700	
2	350 1600 2700	
3	350 1600 2700	
4	350 1600 1600 2700 350 2700	10 10 20 20 30 30

선택된 작업의 수는 $n_i=3(i=1,\ 2,\ 3)$, 그리고 $n_4=6$, 반복횟수는 $r_1=15$, $r_2=13$, $r_3=14$, 그리고 $r_4=15$이고, 관계된 모수(parameters)는 다음의 표 3과 같다(Smith[6], pp.73-81 참조).

표 3. 수평보오링밀작업 예제에 대한 모수의 값

모 수 / 작업요소(i)	β_{i0}	β_{i1}	β_{42}	δ_i^2
1	.391	.000132		$.4\times10^{-3}$
2	.063	.000029		$.4\times10^{-7}$
3	.163	.000024		$.36\times10^{-6}$
4	.237	.000097	.025466	$.9\times10^{-3}$

OLS 추정법을 이용했을 때의 IMSE를 계산하기 위해 $M_{\xi i}$는 표 2의 데이터를 이용하여 계산한 $(\Xi_i'\Xi_i)/n_i$ $(i=1,\ 2,\ 3,\ 4)$로 근사치를 구할 수 있는데, 그 값은 다음과 같다.

$$M_{\xi i}=\begin{bmatrix} 1 & \\ & \\ .155\times10^4 & .3324\times10^7 \end{bmatrix}$$

$$M_{\xi 4}=\begin{bmatrix} 1 & & \\ .155\times10^4 & .3324\times10^7 & \\ .2\times10^2 & .3283\times10^5 & .4667\times10^3 \end{bmatrix}$$

미래에 있어서 독립변수의 참값의 2차 모멘트행렬 M이 다음과 같다고 가정한다.

$$M=\begin{bmatrix} M_e & & & & \\ M_e & M_e & & & \\ M_e & M_e & M_e & & \\ M_e & M_e & M_e & M_e & \\ m_e & m_e & m_e & m_e & .484\times10^3 \end{bmatrix}$$

여기서

$$M_e=\begin{bmatrix} 1 & \\ & \\ .15\times10^4 & .324\times10^7 \end{bmatrix}$$

$m_e=(.2\times10^2,\ .324\times10^5)$

위와 같이 모수의 값들이 주어지면 IMSE를 구할 수 있다. 다음의 표 4는 여러가지 σ_1과 σ_2의 수준 조합에 대하여

$$\delta=100\cdot(\mathrm{IMSE}-\mathrm{IMSE}_0)/\mathrm{IMSE}_0$$

의 값을 나타내고 있다. IMSE_0는 독립변수에 오차가 없을 때의 IMSE 값이다. 표 4로 부터 IMSE의 증가분은 σ_1의 변화보다 σ_2의 변화에 더 민감하다는 것을 뚜렷이 알 수 있다. 즉, σ_2(이동거리를 측정하는데 있어서의 오차의 표준편차)가 σ_1(무게를 측정하는데 있어서의 오차의 표준편차) 보다 IMSE를 줄이는 데에 보다 결정적인 역할을 한다는 것을 의미한다. 예를 들어, 현재의 측정체계로는 σ_1과 σ_2가 각각 80과 1이라고 하자. 그러면, σ_1 보다는 σ_2를 줄이는 것이 IMSE의 감소를 위해 더욱 효과적이다. 이때 물론 각 측정오차의 크기를 줄이는데 필요한 경비, 기술적 문제 등이 함께 고려

표 4. 수평보오일밀작업 예제에 대한 IMSE의 증가 퍼센트

σ_2 \ σ_1	0	8	25	40	80
0	0	.4628	4.516	11.54	45.88
0.1	.7449	1.208	5.261	12.29	46.63
0.3	6.693	7.156	11.21	18.24	52.60
0.5	18.53	18.99	23.05	30.09	64.49
1.0	72.94	73.40	77.48	84.56	110.13

되어야 할 것이다.

6. 결　　론

독립변수(작업특성)에 측정오차가 있을 때, 총 작업시간의 예측치의 평균적 행태를 나타내는 척도로서 IMSE를 제시하였다. 각 독립변수의 측정 오차에 따른 IMSE의 상대적 크기를 비교함으로써 어떤 독립변수의 측정오차가 IMSE의 증가에 더 큰 역할을 하는가를 밝혀내고, 궁극적으로 어떤 측정 오차를 중점적으로 통제해야 하는지를 결정할 수 있다. 그 외에 IMSE는 여러 추정방법의 우열을 비교할 수 있는 기준으로도 사용될 수 있다. 끝으로, 각 독립변수의 측정오차가 IMSE에 미치는 영향도를 평가할 수 있는 척도를 마련하는 것이 매우 중요하다고 여겨진다. 앞으로 이에 대한 연구가 필요하리라고 보며, 그와 같은 척도가 마련되기 까지는 관련모수들을 데이터로 부터 추정하여 예제에서와 같은 수치적 방법을 사용할 것을 추천한다.

참고문헌

[1] Davies, R. B. and Hutton, B., "The Effect of Errors in the Independent Variables in Linear Regression," *Biometrika*, Vol. 11, pp. 383-392, 1975.

[2] Fein, M., "How 'Reliability,' 'Precision,' and 'Accuracy' Refer to Use of Work Measurement Data," *Ind. Eng.*, Vol. 13, pp. 26-33, 1981.

[3] Ganse, R. A., Amemiya, Y., and Fuller, W. A., "Prediction When Both Variables Are Subject to Error, with Application to Earthquake Magnitudes," *J. Amer, Statist, Assoc*, Vol. 78, pp. 761-765, 1983.

[4] Lindley, D. V., "Regression Lines and the Linear Functional Relationship," *J. R. Statist. Soc., Supp.*, Vol. 9, pp. 219-244, 1947.

[5] Seber, G. A. F., *Linear Regression*, Wiley, New York, 1977.

[6] Smith, G. L., Jr., *Work Measurement : A Systems Approach*, Grid Publishing, Inc., Columbus, 1978.

[7] Yum, B. J. and Byun, J. H., "Analysis of the Prediction Problem with Errors in the Variables," *IIE Trans*, Vol. 22, pp. 73-83, 1990.

[8] Yum, B. J. and Neuhardt, J. B., "Analysis of the Prediction Problem in a Simple Functional Relationship Model," *IIE Trans.*, Vol. 16, 177-184, 1984.