

품질통계에서 자유도 개념 및 유형

최성운*

*경원대학교 산업공학과

Concept and Type for Degree of Freedom in Quality Statistics

Sung Woon Choi*

*Department of Industrial Engineering, Kyungwon University

Abstract

This paper presents real examples of quality statistics for users to easily understand the concept and purpose for obtaining the degree of freedom.

Moreover degree of freedom by Satterwaite can be used for linear combinations of unbiased variance. Finally effective degree of freedom by Welch-Satterthwaite is applicable to obtain expanded uncertainty considering type A and type B uncertainty.

Keywords : Degree of Freedom, Satterthwaite, Linear Combinations of Unbiased Variance, Welch-Satterthwaite, Expanded Uncertainty

1. 서론

역학계에서 위치를 정하는 좌표(Coordinate)로서 임의로 독립적(Independent)인 변동을 할 수 있는 개수를 역학계의 자유도(Degree of Freedom : df , ν , ϕ)라 한다. 열평형에 있는 물질계에서 상(Phase)의 수를 변동시키지 않고 서로 변동시킬 수 있는 상태변수의 수를 열역학적 자유도라 한다.[4]

공간을 자유로이 운동할 수 있는 n 개의 질점계의 자유도는 $3n$ 이며 구속 또는 제약조건이 k 개 있으면 독립된 변수의 수 즉 자유도는 $3n-k$ 가 된다. 자유로운 강성체의 자유도는 6인데 X , Y , Z 축의 3차원 공간에서 직선운동, 중심회전운동을 수행하기 때문이다. 로봇의 경우 링크(Link), 관절(Joint) 운동 또는 모터(Motor)의 수가 자유도가 될 수 있으며 손목관절은 피치(Pitch), 요(Yaw), 롤(Roll)과 인쪽, 바깥쪽 운동의 X 축, 옆방향 운동의 Y 축, 아래, 위 운동의 Z 축과 함께 Shoulder, Elbow, Arm, Wrist, Gripper, End-of-Arm-Tooling 등의 Manipulator의 부품에 따라 자유도가 달라진다.

로봇의 자유도는 인간의 관절에 해당하는 것으로 크게 할수록 인간과 유사한 동작을 하게 된다. 자동차의 기구에서 핸들은 Z 축 중심회전운동의 1자유도이며 지면위의 자동차는 Y 축 직선운동과 Z 축 중심회전 운동의 2자유도를 갖는다. 기구의 자유도가 증가 할수록 다양한 기능을 수행할 수 있으나 제어하기가 힘들다.

통계학에서 자유도는 데이터의 개수에서 통계적 제약조건의 수를 뺀 것으로 정의하는데 이는 편차제곱합(Sum of Squares) 또는 변동(Variation)으로 평균제곱(Mean Square) 또는 불편분산(Unbiased Variance)을 구할 경우 적용된다. 이 경우 통계적 제약조건은 제로의 편차합을 의미하며 n 개의 데이터의 경우 자유도는 $n-1$ 이 된다. 그러나 역학계나 열역학적 자유도와 같이 통계적 자유도는 형상에 따른 개념이 구체적이지 못하고 추상적이어서 쉽게 이해할 수 없는 단점이 있다.

따라서 본 연구에서는 통계에서의 자유도의 개념 및 유형을 다음 세 가지와 같이 제시하여 품질통계 사용자에게 자유도의 이해를 높인다.

첫째, 통계학에서 자유도의 정의, 구하는 목적, 구하는 방법 등을 1차원 데이터, 중회귀 분석과 반복있는 2원배치법 등의 2차원 이상의 데이터를 중심으로 예를 들어 제시한다.

둘째, 등분산이 아닌 경우 두 모평균차의 검정, 반복 있는 2원배치법 혼합모형인 난괴법(Randomized Block Design)에서 모수인자의 모평균을 추정하는 경우 오차항의 풀링(Pooling) 전후 사용되는 Satterthwaite 자유도 정의 및 구하는 방법을 소개한다.

셋째, 측정불확도에서 A형, B형 불확도의 확장불확도를 산정할 경우 적용되는 Welch-Satterthwaite 자유도 정의 및 구하는 방법을 제시한다.

2. 자유도의 개념

2.1 정의 및 구하는 방법

품질통계에서는 모집단에서 랜덤샘플링한 샘플은 효율성을 추구하기 위해 가능한한 최소로 구성되면서 서로 영향을 받지 않는 독립적인 랜덤 샘플 즉 iid(Identically and Independently Distributed) 샘플로 가정한다. 자유도는 순수한 독립인 개수로 정의되며 품질통계에서 자유도를 구하는 목적은 샘플 편차제공합의 평균 즉 샘플 분산을 구할 경우 불편성(Unbiasedness)을 만족해 주기 위해 자유도를 사용한다. 계량형 데이터가 3개 이상일 경우 산포(흩어짐)를 나타내는 가장 좋은 측도는 각 데이터와 평균값의 차인 편차제공합이며 이 경우 데이터의 기준이 되는 평균값을 구할 경우 데이터의 독립성을 잃고 종속적인 관계가 된다.

따라서 편차제공합을 자유도를 나누어 불편분산을 구할 경우 자유도를 구하는 방법은 편차제공합의 형태에 따라 달라진다. 자유도는 편차제공합의 데이터 총 개수에서 데이터에 의해 결정되어 지는 평균의 수 또는 계수의 수만큼 증가, 감소하는 개수로 자유도=총 데이터 수 ± 데이터에 의해 계산되는 평균 또는 계수의 수로 표현된다.

2.2 유형

2.2.1 1차원 데이터

모집단에서 랜덤샘플링한 샘플의 계량치 데이터가 x_1, x_2, \dots, x_n 인 랜덤 샘플 즉 iid 샘플일 경우 평균의 자유도는 n 개다. 편차제공합 $S = \sum_{i=1}^n (x_i - \bar{x})^2$ 의 자

유도는 x_i 의 총 데이터 수 n 개에서 데이터에 의해 계산된 평균 \bar{x} , 1 개를 빼 $n-1$ 개다. 이 경우 불편분산은

$$V = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \frac{S}{\nu} \text{ 가 되며 } E(s^2) = \sigma^2 \text{ 을 만족한다.}$$

2.2.2 2차원 이상 데이터

2.2.2.1 중회귀분석의 자유도

$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \dots + \hat{\beta}_k x_{ki}$ 의 중회귀분석은 x 의 원인과 y 의 결과 즉 인과관계를 예측회귀식으로 구할 수 있다. 중회귀식이 유의적인가를 알아보기 위해 분산분석에 의한 검정을 실시할 경우의 편차제공합에 대한 자유도를 구해 보기로 한다. $S_T = \sum_{i=1}^n (y_i - \bar{y})^2$ 의 자유도는 y_i 의 총 데이터의 수 n 개에서 데이터에 의해 계산된 평균 \bar{y} , 1 개를 빼 $\nu_T = n-1$ 이다.

$S_E = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \dots + \hat{\beta}_k x_{ki})]^2$ 의 자유도는 y_i 의 총 데이터의 수 n 개에서 데이터에 의해 계산된 계수의 수 $k+1$ 개를 빼 $\nu_E = n-k-1$ 이다. $S_R = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ 의 자유도는 \hat{y}_i 의 계수의 수 $k+1$ 에서 \bar{y} 평균의 수 1 개를 빼 $\nu_R = k$ 이다.

2.2.2.2 반복있는 2원배치법의 자유도

반복있는 2원배치법의 데이터 구조식은 $x_{ijk} = \mu + a_i + b_j + (ab)_{ij} + e_{ijk}$, $i=1, 2, \dots, l$, $j=1, 2, \dots, m$, $k=1, 2, \dots, r$ 이다. 분산분석을 실시하기 위한 편차제공합의 자유도를 구해 보기로 한다.

$S_A = mr \sum_{i=1}^l (\bar{x}_{i.} - \bar{x})^2$ 의 자유도는 $\bar{x}_{i.}$ 의 총 데이터 수 l 개에서 데이터에 의해 계산된 평균 \bar{x} , 1 개를 빼 $\nu_A = l-1$ 이다. $S_B = lr \sum_{j=1}^m (\bar{x}_{.j} - \bar{x})^2$ 의 자유도는 $\bar{x}_{.j}$ 의 총 데이터 수 m 개에서 데이터에 의해 계산된 평균 \bar{x} , 1 개를 빼 $\nu_B = m-1$ 이다. $S_{AB} = r \sum_{i=1}^l \sum_{j=1}^m (\bar{x}_{ij.} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x})^2$ 의 자유도는 $\bar{x}_{ij.}$ 의 총 데이터 수 lm 개에서 데이터에 의해 계산된 평균 $\bar{x}_{i.}$, l 개를 빼고 데이터에 의해 계산된 평균 $\bar{x}_{.j}$, m 개를 빼고 데이터에 의해 계산된 평균 \bar{x} , 1 개를 더해서 $\nu_{AB} = lm - l - m + 1$ 이 된다.

$S_E = \sum_{i=1}^l \sum_{j=1}^m \sum_{k=1}^r (x_{ijk} - \bar{x}_{ij.})^2$ 의 자유도는 x_{ijk} 의 총 데이터 수 lmr 개에서 데이터에 의해 계산된 평균 $\bar{x}_{ij.}$, lm 개를

편 $\nu_E = lmr - lm$ 이다. $S_T = \sum_{i=1}^l \sum_{j=1}^m \sum_{k=1}^r (x_{ijk} - \bar{x})^2$ 의 자유도는 x_{ijk} 의 총 데이터수 lmr 개에서 데이터에 의해 계산된 평균 \bar{x} , 1개를 뺀 $\nu_T = lmr - 1$ 이다.

3. Satterthwaite 자유도

3.1 정의

Satterthwaite 공식[2]은 불편분산 또는 평균제곱 V_i 들의 선형결합식에 대한 자유도 ν^* 을 구하는 공식을 다음과 같이 정의하였다.

$$V = b_1 V_1 + b_2 V_2 + \dots + b_k V_k$$

$$\nu^* = \frac{b_1 V_1 + b_2 V_2 + \dots + b_k V_k}{\frac{(b_1 V_1)^2}{\nu_1} + \frac{(b_2 V_2)^2}{\nu_2} + \dots + \frac{(b_k V_k)^2}{\nu_k}}$$

3.2 용도

3.2.1 등분산이 아닌 경우 두 모평균차의 검정에서의 자유도

두 모분산비의 검정 즉 F검정이 유의적일 경우 등분산이 성립되지 않으므로 Satterthwaite 자유도에 의한 t검정을 실시해야 한다.

검정통계량 $t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{V_1}{n_1} + \frac{V_2}{n_2}}}$ 의 Satterthwaite

자유도 $\nu^* = \frac{(\frac{V_1}{n_1} + \frac{V_2}{n_2})^2}{\frac{(\frac{V_1}{n_1})^2}{\nu_1} + \frac{(\frac{V_2}{n_2})^2}{\nu_2}}$ 이다.

3.2.2 반복있는 2원배치 혼합모형의 모수인자 모평균 추정에서의 자유도

3.2.2.1 오차항을 풀링하지 않는 경우

혼합모형인 경우 모수인자 A의 모평균 구간추정식은 $\mu(A_i) = \bar{x}_{i..} \pm t_{1-\frac{\alpha}{2}}(\nu^*) \sqrt{\frac{V_B + lV_{A \times B} - V_E}{lmr}}$ 이며 Satterthwaite

자유도 $\nu^* = \frac{[(V_B + lV_{A \times B} - V_E)/lmr]^2}{\frac{(V_B)^2}{lmr} + \frac{(lV_{A \times B})^2}{lmr} + \frac{(-V_E)^2}{lmr}} = \frac{(V_B + lV_{A \times B} - V_E)^2}{\frac{(V_B)^2}{\nu_B} + \frac{(lV_{A \times B})^2}{\nu_{A \times B}} + \frac{(-V_E)^2}{\nu_E}}$

가 된다.

3.2.2.2 오차항을 풀링하는 경우

교호작용이 무시되어 오차항을 풀링한 경우 모수인자 A의 모평균 구간추정식은

$\mu(A_i) = \bar{x}_{i..} \pm t_{1-\frac{\alpha}{2}}(\nu^*) \sqrt{\frac{V_B + (l-1)V_E}{lmr}}$ 이며 Satterthwaite

자유도 $\nu^* = \frac{[(V_B + (l-1)V_E)/lmr]^2}{\frac{(V_B)^2}{lmr} + \frac{((l-1)V_E)^2}{lmr}} = \frac{(V_B + (l-1)V_E)^2}{\frac{(V_B)^2}{\nu_B} + \frac{[(l-1)V_E]^2}{\nu_E}}$

이다.

4. Welch-Satterthwaite 자유도

4.1 정의

Welch-Satterthwaite 공식[3]은 합성표준불확도의 A형, B형 평가방법에 따라 확장불확도를 구하려고 하는 경우 사용되는 유효 자유도

$$\nu_{eff} = \frac{\sum_{i=1}^y [C_i U(x_i)]^4}{\sum_{i=1}^y \frac{[C_i U(x_i)]^4}{\nu_i}} = \frac{U_C^4(y)}{\sum_{i=1}^y \frac{[U_C^4(y)]^4}{\nu_i}}$$
 이다.

여기서 C_i 는 감도계수, $U(x_i)$ 는 x_i 의 표준불확도, $U_C(y)$ 는 합성표준불확도, ν_{eff} 는 $U(x_i)$ 의 A형, B형 평가방법에 따른 유효자유도이다.

4.2 확장불확도에서의 자유도 적용

확장불확도를 구하는 절차는 다음과 같다. [1]

단계 1 : 합성표준불확도 $U_C(y) = \sum_{i=1}^y C_i U(x_i)$ 를 구한다.

단계 2 : 유효자유도 ν_{eff} 를 4.1절과 같이 구한다.

단계 3 : 신뢰수준 $1-\alpha$ 와 유효자유도 ν_{eff} 로 포함인자

$t_{1-\frac{\alpha}{2}}(\nu_{eff})$ 구한다.

단계 4 : 확장불확도 $y = t_{1-\frac{\alpha}{2}}(\nu_{eff}) U_C(y)$ 를 구한다.

5. 결론

본 연구에서는 역학계와 열역학적 자유도의 개념에 비해 추상적인 통계에서의 자유도의 개념 및 구하는 방법을 품질통계 사용자가 이해하기 쉬도록 1차원 데이터, 중회귀분석, 반복있는 2원배치법등의 편차제곱합에 대한 예를 들어 제시하였다. 또한 불편분산 또는 평균제곱의 선형결합식으로 표현되는 품질통계 모형인 경우 적용되는 Satterwaite 자유도를 제시하였으며 합성표준불확도의 A형, B형 평가에 의해 확장불확도를 구하고자 할 경우 적용되는 Welch-Satterthwaite 의 유효자유도를 제안하였다.

향후 연구로는 다변량분석, 시계열분석 등을 이용한 품질통계에서 자유도의 개념 및 유형을 연구해 보고자 한다.

6. 참고 문헌

- [1] KSA 3000 : 2005, 측정결과의 불확도 추정 및 표현을 위한 지침, 한국표준협회, 2005.
- [2] Satterthwaite F.E., "An Approximate Distribution of Estimates of Variance Components", Biometric Bulletin, 2(6) (1946) : 110 - 114.
- [3] Welch B.L., "The Generalization of Students Problem When Several Different Population Variances Are Involved", Biometrika, 34(112) (1947) : 28 - 35.
- [4] <http://en.wikipedia.org>

저자 소개

최성운



현 경원대학교 산업공학과 교수. 한양 대학교 산업공학과에서 공학사, 공학석사, 공학박사 학위를 취득하고, 1994년 한국과학재단 지원으로 University of Minnesota에서 1년간 Post-Doc을 수행했으며, 2002년부터 1년 반동안 University of Washington에서 Visiting Professor를 역임하였음. 주요 관심분야는 자동화 생산 및 장치 산업에서의 품질관리이며, 컴퓨터, 정보시스템의 신뢰성 설계 및 분석, 서비스 사이언스, RFID시스템에서도 관심을 가지고 있음.

주소: 경기도 성남시 수정구 복정동 산65번지 경원대학교
산업공학과 ☎031)750-5366