

다항식 회귀분석을 이용한 전자저울의 비선형 특성 개선 연구

채규수

백석대학교 ICT학부 교수

A Study of the Nonlinear Characteristics Improvement for a Electronic Scale using Multiple Regression Analysis

Gyoo-Soo Chae

Professor, Division of ICT, Baekseok University

요 약 본 연구에서는 다항식 회귀분석(Polynomial regression analysis) 방법을 이용하여 비선형 특성을 갖는 전자저울의 질량 추정 모델 개발이 이루어 졌다. 전자저울에 사용되는 로드셀의 출력 단자 전압을 기준 질량 추를 사용하여 직접 측정하였고 이 데이터를 이용하여 MS Office 엑셀의 행렬식 계산과 데이터 추세선 분석 기능을 이용하여 다항식 회귀 모델을 구하였다. 5kg까지 측정 가능한 로드셀 전자저울을 사용하여 100g단위로 질량을 측정하였고 다항식 회귀분석(Multiple regression analysis) 모델을 구하였으며, 단순(1차), 2차, 3차 다항식 회귀분석에 대한 오차를 구하였다. 각 모델에 대한 회귀 방정식의 적합도 분석을 위해 결정계수(Coefficient of determination)를 제시하여 추정 질량과 측정 데이터와의 상관관계를 나타내었다. 본 연구에서 제안하는 3차 다항식 모델을 이용하여 추정 값의 표준편차가 10g, 결정계수 1.0으로 상당히 정확한 모델을 얻었다. 본 연구에 사용된 선형 회귀 분석 이론을 바탕으로 최근 인공지능 분야에서 많이 사용되고 있는 로지스틱 회귀 분석(Logistic regression analysis)을 활용하여 기상예측, 신약개발, 경제지표 분석 등의 분야에 대한 다양한 연구를 수행할 수 있을 것으로 생각된다.

주제어 : 회귀분석, 다항식, 비선형, 질량 추정, 전자저울

Abstract In this study, the development of a weight estimation model of electronic scale with nonlinear characteristics is presented using polynomial regression analysis. The output voltage of the load cell was measured directly using the reference mass. And a polynomial regression model was obtained using the matrix and curve fitting function of MS Office Excel. The weight was measured in 100g units using a load cell electronic scale measuring up to 5kg and the polynomial regression model was obtained. The error was calculated for simple(1st), 2nd and 3rd order polynomial regression. To analyze the suitability of the regression function for each model, the coefficient of determination was presented to indicate the correlation between the estimated mass and the measured data. Using the third order polynomial model proposed here, a very accurate model was obtained with a standard deviation of 10g and the determinant coefficient of 1.0. Based on the theory of multi regression model presented here, it can be used in various statistical researches such as weather forecast, new drug development and economic indicators analysis using logistic regression analysis, which has been widely used in artificial intelligence fields.

Key Words : Regression analysis, Polynomial, Non-linear, Weight estimation, Electronic scale

*This study was supported by 2019 Baekseok University Research Fund.

*Corresponding Author : Gyoo-Soo Chae(gschae@bu.ac.kr)

Received May 2, 2019

Revised May 27, 2019

Accepted June 20, 2019

Published June 28, 2019

1. 서론

일상생활에서 무게를 정확히 측정하기 위하여 여러 가지 방법들이 사용되는데, 기준 질량 추를 이용하는 양팔 저울, 코일 형태의 스프링의 탄성을 이용한 저울, 디지털 전자저울 등이 주로 사용되고 있다. 현재 사용하고 있는 대부분의 전자저울은 일종의 압력 센서를 이용하여 압력 값을 질량 값으로 환산하여 디지털 형태로 보여주는 방식이다. 센서를 사용하는 전자저울의 특성상 외부 압력 데이터를 질량으로 변환하는 단계가 필요하다. 일반적으로 압력을 받는 전기/기계 센서의 변화에 따른 출력 신호를 전기적인 회로를 이용하여 질량 값으로 변환하여 보여주는 방법을 사용한다. 이 때 문제가 되는 것은 압력을 받는 금속 물질이 일정 압력 이상 한계에 이르면 비선형적인 특성을 갖게 되므로 이에 대한 정확한 보정 과정이 필요하게 된다. 이러한 압력 데이터를 질량 데이터로 변환하기 위한 다양한 방법들이 소개 되었다[1,2]. 선형 회귀분석(Regression analysis)법을 이용하여 측정된 데이터의 비선형 특성을 고려하여 좀 더 정확한 질량 표시 방법에 대한 연구들이 소개되었다[3,4]. 본 연구에서는 외부의 힘 또는 압력에 의해 전기저항이 변화하는 스트레인지지를 로드셀(Load cell)에 부착한 전자저울이 사용되었다 [5]. 본 연구에서는 5kg까지 측정이 가능한 소형 전자저울 키트를 사용하여 100g 단위로 50개의 측정 데이터를 구하고, 이 데이터들에 다항식 회귀분석(Multi regression analysis)을 적용하여 전 측정 구간에서 정확도가 향상된 질량 추정 모델을 제시하였다.

2. 본 론

2.1 회귀분석

회귀분석은 기상예측, 경제지표, 시스템 제어, 생물학, 화학 등의 다양한 분야에서 어떤 변화요인들이 또 다른 변화요인에 의해 결정된다고 보고 그 변화요인들 사이의 함수관계를 분석하는 통계적 해석 기법이다[6,7]. 기본적으로 하나 이상의 독립변인(x)이 한 단위 변할 때, 종속변인(y)이 얼마나 변할 것인지를 예측하는 데 주로 사용하는 통계분석 기법이다. 이렇게 예측된 함수가 원래 데이터와 어느 정도 일치하는가를 나타내는 적합도를 예측하는 분석 방법이다. 하나의 독립변수만을 고려하여 종속변수와의 관계를 분석하는 방법을 단순(1차) 회귀 분석(Simple(1st order) linear regression model), 두 개 이상의 독립변수들을 고려하는 방법을 다중 회귀분석(Multiple regression analysis)이라고 한다. 또한 독립변수와 종속변수

간의 관계에 따라 선형회귀분석(Linear regression analysis)과 비선형회귀분석(Non-linear regression analysis)으로 구분된다. 본 연구에서 제시된 전자저울에 사용되는 로드셀이 외부에서 받는 압력과 이 때 발생하는 내부 인장 관계와 같이, 일정 압력 이상이 되면 인장 값이 한계에 이르러 같은 비율로 지속적으로 증가하지 않는 비선형 특성을 갖는다. 본 논문에서는 선형 및 비선형 회귀분석 방법을 사용하여 전자저울의 측정 데이터를 사용하여 질량 추정 모델을 제시하고자 한다.

2.1.1 단순(1차) 선형 회귀분석

단순(1차) 선형 회귀분석의 경우에는 상관성이 있는 독립 변수와 종속변수의 관계를 나타내는 1차 함수를 추정할 수 있는데 식 (1)과 같이 표현된다[8].

$$y = \alpha_0 + \alpha_1 x \tag{1}$$

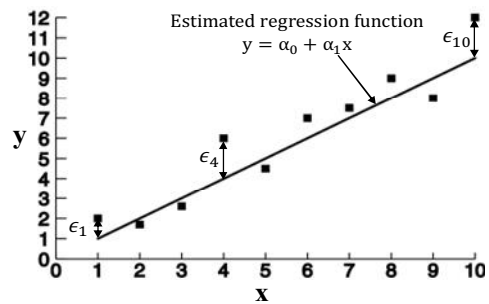


Fig. 1. Simple(1st) Regression Analysis

식 (1)에서 α_0 는 절편이고 기울기가 α_1 인 독립변수 x 와 종속변수 y 의 관계식을 회귀방정식이라고 한다. Fig. 1에서는 단순 선형 회귀 방정식의 예를 보여주는 그래프이다. 회귀직선과 실제 데이터 사이에 오차(ϵ)들이 존재하며 이를 다음과 같이 표기한다.

$$y = \alpha_0 + \alpha_1 x + \epsilon \tag{2}$$

$y = f(x, \alpha_0, \alpha_1, \dots, \alpha_n)$ 대하여 오차(ϵ) 제곱의 합이 최소가 되는 조건을 찾는 것이 최소자승법(LMS: Least Mean Square)이다. 이것은 식 (3), (4), (5)와 같이 회귀모델 $y = \alpha_0 + \alpha_1 x$ 에 대하여 계수 α_0, α_1 에 대한 S_r 의 미분이 0이 될 때를 의미한다.

$$S_r = \sum_{i=1}^n (y_i - \alpha_0 - \alpha_1 x_i)^2 \tag{3}$$

$$\frac{\partial S_r}{\partial \alpha_0} = \sum_{i=1}^n 2(y_i - \alpha_0 - \alpha_1 x_i)(-1) \quad (4)$$

$$\frac{\partial S_r}{\partial \alpha_1} = \sum_{i=1}^n 2(y_i - \alpha_0 - \alpha_1 x_i)(-x_i) \quad (5)$$

2.1.2 다항식 회귀 분석

전자저울에 사용되는 로드셀에 부착된 스트레인 게이지의 변형에 따른 저항 값이 전 구간에서 선형적으로 증가하지 않는다. 이러한 비선형 특성으로 인해 단순(1차) 선형 회귀 분석에 한계가 있어 식 (6)과 같은 다항식 회귀분석을 적용하였다. Fig. 2에서 비선형 특성의 데이터에 대한 다항식 회귀 모델이 제시되었다.

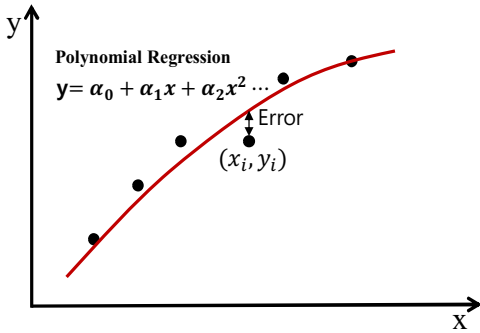


Fig. 2. Polynomial Regression Analysis

본 연구에서는 여기 제시된 2차 다항식 모델인 식 (6)을 기본으로 하여 오차를 줄이기 위해 3차 다항식 모델인 식 (7)을 적용하였다[9,10].

$$y = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \epsilon \quad (6)$$

$$y = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3 + \epsilon \quad (7)$$

다항식을 행렬식으로 변환하여 계산하면 많은 데이터를 사용 할 때 효과적으로 사용할 수 있다. 식 (6), (7)의 방정식을 행렬식으로 표현하면 식 (8)과 같이 되고, 구하고자 하는 방정식에서 행렬 [X]는 식 (9)로 표현된다.

$$[Y] = [X] \{A\} + \{E\} \quad (8)$$

$$[X] = \begin{bmatrix} x_{01} & \dots & x_{m1} \\ \vdots & \ddots & \vdots \\ x_{0m} & \dots & x_{mm} \end{bmatrix} \quad (9)$$

계수 $\alpha_0, \alpha_1, \alpha_2, \alpha_3$ 를 각각 구하는 방법은 모든 데이터들에서 오차가 최소가 되는 새로운 추정 값을 식 (3), (4), (5)에 나타난 것과 같이 각 계수의 미분을 이용하여 구할 수 있다.

단순선형회귀, 다항식회귀 등 모두 동일하게 적용이 가능한 일반화된 행렬식은 다음과 같이 표현된다.

$$[[X]^T [X]] [A] = \{[X]^T [Y]\} \quad (10)$$

위 식에서 회귀방정식의 계수 벡터 [A]를 구하기 위해 식 (10)을 이용하여 다음과 같이 구할 수 있다.

$$\{A\} = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = [[X]^T [X]]^{-1} \{[X]^T [Y]\} \quad (11)$$

2.1.3 회귀모델의 적합도 분석

본 연구에서 사용된 각 회귀모델이 원래의 측정 데이터와 얼마나 일치하는가를 판별하는 적합도 평가를 위하여 결정 계수(R^2)를 사용하였다[8]. 결정계수는 다음 식 (12)로 표현된다.

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (12)$$

여기서, SST(total sum of square)는 각 데이터 값과 데이터의 평균(\bar{y})과의 차이의 제곱이며, SSE(explained sum of square)는 회귀 모델로 구한 추정 값(\hat{y})과 측정 데이터 차의 제곱이며, SSR(residual sum of square)은 추정 값과 데이터 평균의 차이 제곱이다.

2.2 로드셀 전자저울

금속 저항체가 응력을 받을 때 전기저항이 변하는 스트레인게이지를 로드셀에 부착하고 하중을 가했을 때 일어나는 저항변화를 검출하여 하중을 측정하는 것이 로드셀 전자저울이다. 로드셀은 스트레인게이지를 휘스톤브릿지(Wheatstone bridge) 형태로 구성하여 사용한다[11]. Fig. 3에서 로드셀에 부착된 4개의 저항 역할을 하는 스트레인게이지(R1, R2, R3, R4)와 이 회로의 등가회로인 휘스톤브릿지 회로의 연결 상태가 제시되었다.

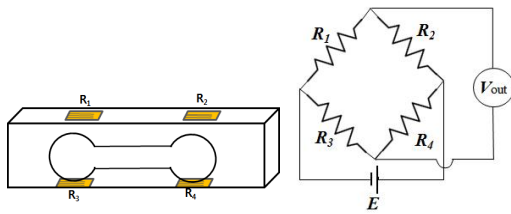


Fig. 3. Load cell with strain gauge and Wheatstone bridge circuit

로드셀의 원리를 이용한 스트레인게이지의 전기적인 특성의 변화를 측정하여 가해지는 압력을 구하는 방법이 식 (13)에 제시되었다[11].

$$V_{out} = k \times \text{변형율}(\text{strain}) \times E = \frac{\text{저항의 변화}(dR)}{\text{기준저항}(R)} \quad (13)$$

여기서, V_{out} 은 로드셀의 출력 전압, E 는 외부 인가 전압, k 는 탄성체가 외력에 의해 변형되었을 때 원래 상태로 돌아오려고 하는 복원력의 크기를 나타내는 상수이다. 로드셀에 부착된 4개의 스트레인게이지는 저항으로 동작하게 되고 외부에서 전압(E)을 인가했을 때 로드셀 출력 전압(V_{out})은 다음과 같이 표현된다.

$$V_{out} = \left(\frac{R_1}{R_1 + R_2} - \frac{R_4}{R_3 + R_4} \right) E = \frac{R_1 R_3 - R_2 R_4}{(R_1 + R_2)(R_3 + R_4)} \quad (14)$$

여기서, 스트레인게이지가 외부 응력을 받지 않은 상태, $R_1 R_3 = R_2 R_4$ 이면 $V_{out} = 0$ 이 된다.

3. 실험 결과 및 분석

본 연구의 질량 기준 데이터를 얻기 위하여 아두이노 HX711 로드셀 5kg 전자저울 키트가 사용되었다. 실험에 사용된 로드셀, HX711 증폭기, 아두이노 모듈의 연결 구성도와 실제 실험을 위한 준비 상태가 Fig. 4에 제시되었다.

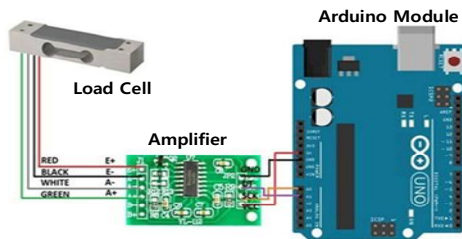


Fig. 4. Arduino module and load cell with Amplifier

본 연구에서 제시하고자 하는 질량 추정 모델 개발을 위해 실제 전자저울의 질량 데이터를 사용하여 다항식 회귀 방정식을 구하고자 한다. 질량 추정 회귀방정식을 구하기 위해 Fig. 4에 제시된 구성과 아두이노 연결 프로그램을 이용하여 기준 추의 질량에 대한 전자저울의 측정 질량을 Table 1과 같이 구하였다. Table 1에 나타난 측정 질량은 아두이노 측정 과정에 100g 단위의 기준 추를 사용한 질량 측정 결과가 순간적으로 변화가 심하기 때문에 5회의 평균값을 표시하여 얻은 결과이다.

Table 1. Measured results using standard weight

Reference (g)	Measured (g)	Reference(g)	Measured (g)
0	0	2,600	2,622.3
100	99.7	2,700	2,724.8
200	199.9	2,800	2,804.5
300	302.2	2,900	2,925.3
400	397.4	3,000	3,026.4
500	497.6	3,100	3,105.2
600	602.1	3,200	3,206.6
700	698.5	3,300	3,297.2
800	796.8	3,400	3,406.4
900	901.2	3,500	3,504.8
1,000	1,002.4	3,600	3,606.5
1,100	1,099.2	3,700	3,706.3
1,200	1,198.9	3,800	3,807.5
1,300	1,299.5	3,900	3,900.4
1,400	1,400.5	4,000	3,992.0
1,500	1,498.5	4,100	4,090.0
1,600	1,603.6	4,200	4,185.0
1,700	1,697.5	4,300	4,284.0
1,800	1,803.4	4,400	4,367.0
1,900	1,895.5	4,500	4,440.0
2,000	1,996.0	4,600	4,547.5
2,100	2,095.0	4,700	4,642.4
2,200	2,196.2	4,800	4,720.2
2,300	2,300.1	4,900	4,800.4
2,400	2,398.5	5,000	4,850.7
2,500	2,510.3		

Fig. 5에서는 Table 1에서 제시된 측정 질량 값들을 이용하여 식 (11)의 행렬식 계산으로 단순 선형 회귀 방정식을 구하는 과정을 보여주고 있고 회귀 방정식은 $y = 98.903x + 18.66$ 이다. Fig. 6에서는 측정된 질량 데이터를 이용하여 엑셀의 추세선(curve fitting) 기능을 이용하여 구한 회귀 방정식 $y = 98.904x - 80.243$ 과 결정 계수가 0.9997로 계산되었다. 그래프에서 5kg 근처의 비선형 구간에서 여전히 약간의 오차가 있는 것을 볼 수 있다. Fig. 7에서는 단순 선형 회귀 방정식을 구하는 과정과 동일한 방법으로 3차 다항식 선형 회귀 방정식을 구하는 과정을 보여주고 있고 회귀 방정식은 $y = -0.006x^3 + 0.35x^2 + 95.017x + 12.811$ 로 나타났다. Fig.

8에서는 엑셀을 이용하여 구한 회귀 방정식 $y = -0.006x^3 + 0.3683x^2 + 94.298x - 81.849$ 을 보여주고 있다. 결정계수는 1.0으로 나타나 측정 데이터와 거의 일치하는 것으로 판단되며, Fig. 8에서는 추세선이 5kg 근처의 비선형 데이터 구간에서도 거의 유사하게 따라가는 것을 볼 수 있다.

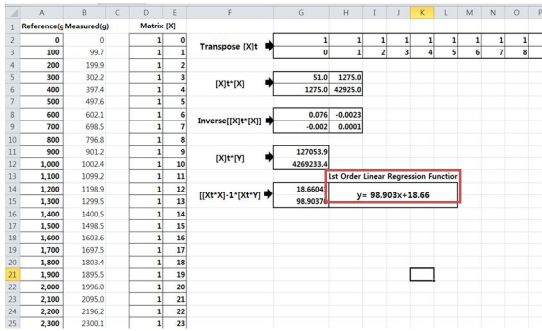


Fig. 5. Matrix coefficient Calculation for Simple(1st order) linear regression model using measured mass data

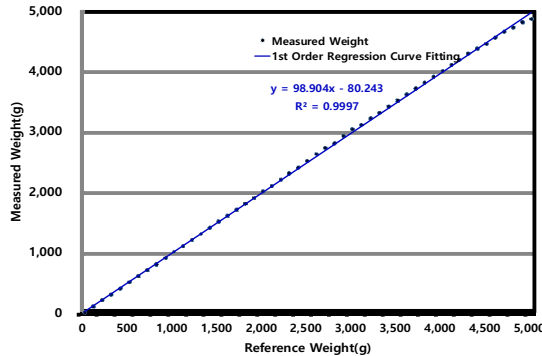


Fig. 6. Calculation for Simple(1st order) linear regression model using excel curve fitting

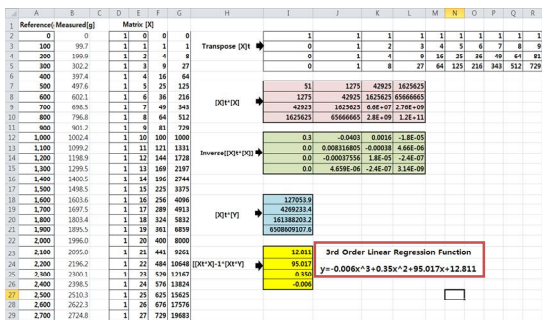


Fig. 7. Matrix coefficient Calculation for 3rd order polynomial regression model using measured

mass data

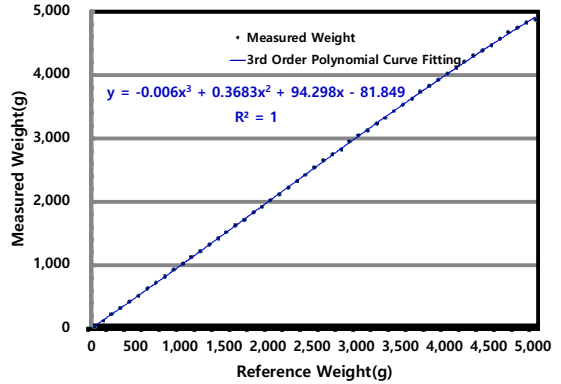


Fig. 8. Calculation for 3rd order polynomial regression model using excel curve fitting

Fig. 9에서는 전자저울의 질량 데이터를 이용하여 개발된 3가지 질량 추정 모델의 오차가 나타나 있다. 단순 선형 회귀 모델, 2차 다항 식 모델, 3차 다항식 모델에서 각각 27g, 18g, 10g의 표준편차가 나타나는 것으로 예측되었다. 추정 오차는 고차 다항식을 사용하는 경우에 더 개선되는 것을 알 수 있으나 모든 경우에 저울의 한계 측정 범위 근처인 5kg 근처에서 다른 구간에 비해 오차가 다소 크게 나타나는 것을 볼 수 있다.



Fig. 9. Errors for each polynomial regression model

4. 결론

본 연구에서는 회귀분석 방법에 대한 연구를 위해 측정 데이터로 로드셀을 사용한 전자저울의 질량 데이터가 사용되었다. 측정된 질량 데이터를 이용하여 Excel의 행렬식 계산과 데이터 추세선 분석 기능을 이용하여 다항식 회귀 모델을 구하였다. 전자저울의 질량 예측을 모델 개발을 위해 로드셀 전자저울과 표준 질량 추를 사용하여 100g단위로 질량을 직접 측정하였다. 측정된 데이터를 이용하여 단순(1

차), 2차, 3차 다항식 회귀분석 방법을 사용하여 전자저울의 질량 추정을 위한 함수를 구하였다. 각 모델에 대해 질량 추정 방정식의 적합도 분석을 위해 결정계수를 구하여 추정 질량과 측정 데이터와의 상관관계를 나타내었다. 2차 이상의 고차 다항식을 사용할 때 더 정확한 회귀 방정식을 얻을 수 있는 것을 알게 되었다. 본 연구에 사용된 선형 회귀 분석 이론을 바탕으로 최근 인공지능 분야에서 많이 사용되고 있는 로지스틱 회귀분석을 활용하여 기상예측, 신약개발, 경제지표 분석 등의 분야에 대한 다양한 연구를 수행 할 수 있을 것으로 생각된다.

REFERENCES

- [1] J. P. Seo et al. (2018). Evaluation of the Applicability of Sediment Discharge Measurement in Mountain Stream using the Load-cell Sensor. *Journal of the Korea Academia-Industrial cooperation Society*, 19(1), 644-653.
- [2] František Trebuňal et al. (2016). Application of Polynomial Regression Models in Prediction of Residual Stresses of a Transversal Beam. *American Journal of Mechanical Engineering*, 4(7), 247-251.
- [3] Y. U. Park & J. W. Joo. (2014). Study on Measurement of Sustained Load Using Loadcells. *Journal of Industrial Science and Technology Institute*, 28(2), 77-82.
- [4] H. J. Seo, H. S. Jung, G. J. Ryu & T. W. Cho. (2012). High Accurate Creep Compensation of the Loadcell using the Strain Gauge. *Journal of IKEEE*, 16(1), 34-44.
- [5] S. P. Fang, T. Schumann, L. Garcia & Y. K. Yoon. (2017). Emerging Nanotechnology for Strain Gauge Sensor. In *Semiconductor-Based Sensors* (pp. 435-472).
- [6] G. G. Vining, E. A. Peck & D. C. Montgomery. (2012) *Introduction to Linear Regression Analysis*, 5th Edition, Wiley.
- [7] D. C. Montgomery & G. C. Runger. (2003). *Applied Statistics and Probability for Engineers*, John Wiley & Sons.
- [8] S. H. Seo, M. I. Roh & H. K. Shin. (2014). A Study on the Weight Estimation Model of Floating Offshore Structures using the Non-linear Regression Analysis. *Journal of the Society of Naval Architects of Korea*, 51(6), 530-538.
- [9] K. W. Lee, J. S. Ha & S. S. Kang. (2012). Study on Torque precision measuring System using Curve Fitting Algorithm. *Journal of the Korea Society of Digital Industry and Information Management*, 8(4), 1-11.
- [10] K. Y. Cho et al. (2017). A Study on the Decision Making for the Inland Transportation of Shippers by Logistic Regression Analysis. *Journal of Digital Convergence*, 15(11), 187-197.
- [11] Tech application. (1999). Signal Conditioning Wheatstone Resistive Bridge Sensors. *SLOA034*, 1-5. Texas instruments co., Ltd.

채 규 수(Gyoo-Soo Chae)

[종신회원]



- 2000년 12월 : Virginia Tech. 전기 공학과(공학박사)
- 2001년 1월 ~ 2003년 2월 : Amphenol Mobile, RF manager
- 2003년 3월 ~ 현재 : 백석대학교 ICT 학부 교수

- 관심분야 : 안테나 설계, 초고주파 이론
- Email : gschae@bu.ac.kr