

SARIMA 모형을 이용한 태양광 발전량 예보 모형 구축

이동현* · 정아현* · 김진영** · 김창기** · 김현구*** · 이영섭****

*동국대학교 일반대학원 통계학과, 석사과정
**한국에너지기술연구원 신재생에너지자원·정책센터, 선임연구원
***한국에너지기술연구원 신재생에너지자원·정책센터, 책임연구원
****동국대학교 통계학과, 교수

Solar Power Generation Forecast Model Using Seasonal ARIMA

Lee Dong-Hyun* · Jung Ahyun * · Kim Jin-Young** · Kim Chang Ki** · Kim Hyun-Goo*** · Lee Yung-Seop****

*Department of Statistics, Dongguk University, Seoul, Korea 04620, Master Course
**Department of New and Renewable Energy Resource&Policy Center, Korea Institute of Energy Research, Daejeon, Korea 34129, Senior Researcher
***Department of New and Renewable Energy Resource&Policy Center, Korea Institute of Energy Research, Daejeon, Korea 34129, Principal Researcher
****Department of Statistics, Dongguk University, Seoul, Korea 04620, Professor

†Corresponding author: yung@dongguk.edu

Abstract

New and renewable energy forecasts are key technology to reduce the annual operating cost of new and renewable facilities, and accuracy of forecasts is paramount. In this study, we intend to build a model for the prediction of short-term solar power generation for 1 hour to 3 hours. To this end, this study applied two time series technique, ARIMA model without considering seasonality and SARIMA model with considering seasonality, comparing which technique has better predictive accuracy. Comparing predicted errors by MAE measures of solar power generation for 1 hour to 3 hours at four locations, the solar power forecast model using ARIMA was better in terms of predictive accuracy than the solar power forecast model using SARIMA. On the other hand, a comparison of predicted error by RMSE measures resulted in a solar power forecast model using SARIMA being better in terms of predictive accuracy than a solar power forecast model using ARIMA.

Keywords: 태양광 발전량(Solar power generation), 시계열 분석(Time series analysis), ARIMA (Autoregressive intergrated moving average), SARIMA(Seasonal ARIMA), MAE(Mean Absolute Error), RMSE(Root Mean Square Error)

1. 서론

최근 국내에서는 화석연료로 인한 환경문제, 원자력 에너지로 인한 위험성 때문에 정부에서는 신재생 에너지 사용을 늘리고 이에 대한 효율적인 대책을 강구하고 있다. 신재생



OPEN ACCESS

Journal of the Korean Solar Energy Society
Vol.39, No.3, pp.59-66, June 2019
<https://doi.org/10.7836/kses.2019.39.3.059>

pISSN : 1598-6411
eISSN : 2508-3562

Received: 21 May 2019

Revised: 08 June 2019

Accepted: 14 June 2019

Copyright © Korean Solar Energy Society

This is an Open-Access article distributed under the terms of the Creative Commons Attribution NonCommercial License which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

에너지 예보는 연간 신재생설비 운영비를 절감하기 위한 핵심적인 기술이며 신재생 에너지 중에서도 국내에서는 주로 태양으로 인한 태양 에너지 생산이 많이 이루어지고 있다. 태양 에너지 발전량 같은 경우에는 다양한 기상 요인으로 인해 발전량의 변동성이 매우 큰 특징을 가지고 있고 그로 인해 예보의 정확성이 무엇보다 중요하다.

기존 연구를 살펴보면 Reikard(2008)¹⁾은 ARIMA 기법을 이용하여 1시간부터 4시간까지의 일사량 예보 모형을 구축하였으며, Moreno-Muñoz(2008)²⁾은 AR기법과 ARMA 기법을 이용하여 5분 단위의 일사량 예보 모형을 구축하였다. 또한 Kim(2017)³⁾에서는 ARIMA와 ARIMAX 모형을 이용하여 1시간 단위 일사량 예보 모형을 구축하였다. 위의 연구들은 태양광 발전량을 예측을 위해 태양광 발전량에 가장 큰 영향을 미치는 일사량 예측을 하였다.

본 연구에서는 위의 연구들과는 달리 태양광 발전량 데이터만을 이용한 태양광 발전량 예보 모형을 구축하고자 한다. 태양 에너지 초단기 발전예보를 위한 핵심 인자기반 스마트 예보 프레임 구축을 위한 기반연구를 진행 하고 대전지역 태양광 발전소 4개 지점의 2016년 1월 1일 00:00 ~ 2016년 12월 31일 23:00까지 1시간 단위로 측정된 태양광 시계열 자료가 계절성을 고려하지 않은 ARIMA와 계절성이나 주기적 특성을 지니고 있을 경우에 적합한 SARIMA⁴⁾를 이용해 태양광 발전량 예보 모형을 만들어서 어느 방법의 예측 정확성이 더 우수한지 비교하고자 한다.

2. 이론적 배경

2.1 시계열 분석 모형

(1) ARIMA

ARIMA(Auto-Regressive Integrated Moving Average)는 auto-regressive(AR) 모형과 moving average(MA) 모형을 동시에 포함하는 시간의 흐름에 따른 관측값들의 확률구조를 고려한 모형이다. ARIMA 모형은 자기회귀 차수 p, 차분 차수 d, 이동평균 차수 q를 가진다. 시계열 Z_t 가 ARIMA(p, d, q)과정을 따를 때 식(1) 형태의 모형으로 표현한다.

$$\begin{aligned} \phi_p(B)(1-B)^d Z_t &= \theta_0 + \theta_q(B)(1-B)^d Z_t, \\ \text{where, } \phi_p(B) &= (1 - \phi_1 B - \dots - \phi_p B^p), \\ \theta_q(B) &= (1 - \theta_1 B - \dots - \theta_q B^q), \\ \theta_0 &: \text{the deterministic trend} \end{aligned} \tag{1}$$

p : 자기회귀(AR) 차수로서, t 시점의 값에 (t-1), ..., (t-p) 시점 값이 영향을 주는 것을 의미한다.

d : 차분을 의미하며, t 시점의 값에 (t-d) 시점의 값을 빼서 비정상 시계열을 정상 시계열로 만들어준다.

q : 이동평균(MA) 차수로서, t 시점의 값이 (t-1), ..., (t-q) 시점의 연속적인 오차항들의 영향을 받는 것을 의미한다.

이때, $\phi(B)$, $\theta(B)$ 는 각각 AR, MA에 대한 다항식이며 B는 후진연산자(back shift operator)이고 d는 Z_t 의 차분차수이다. ARIMA 모형의 적용 단계는 다음과 같다. 먼저, 시계열 자료에 대한 정상성(stationarity)을 검증한다. 정상성이란 시계열을 일정한 주기로 나누었을 때, 각 주기에 해당하는 평균과 분산이 일정하다는 것이다. 만일 비정상적일 경우, 변수변환 및 차분을 통해 정상성을 만족시키는 작업이 우선되어야 하며, 이를 위해 자기상관함수(Autocorrelation Function, ACF)와 부분자기상관함수(Partial Autocorrelation Function, PACF)를 활용할 수 있다. 두 번째로, 모형의 구성요소인 AR(p), I(d), MA(q)를 조합하여 모형을 추정한 후에 AIC(Akaike's Information Criterion) 통계량 또는 BIC(Normalized Bayesian Information Criterion) 통계량 등을 산출하여 가장 적합한 모형을 결정한다. 추정된 모형이 통계적으로 적절한지에 대한 모형 진단 방법으로는 잔차의 자기상관함수와 부분자기상관함수를 관측하여 잔차들의 독립성을 확인하는 잔차 분석을 주로 이용한다. 마지막으로, 도출된 모형을 예측모형으로 이용하여 미래값을 예측한다. 이처럼 ARIMA 모형은 과거 관측값과 오차항만으로 적합한 모형을 도출할 수 있는 장점이 있으며, 최근 시점에 가까운 관측값에 더 큰 가중치를 주기 때문에 단기에측에 적합하다.

(2) SARIMA (Seasonal ARIMA)

시계열이 계절성이나 주기적 특성을 지니고 있을 경우에 사용하는 SARIMA 모형은 시계열의 구성성분들이 확률적이거나 다른 성분들과 상관이 있는 경우에 사용할 수 있다. 시계열 Z_t 가 계절주기가 s인 SARIMA $(p, d, q)(P, D, Q)_s$ 과정을 따를 때 식(2) 형태의 모형으로 표현한다. 여기서, B는 후행연산자(Backshift Operator), $\phi(B)$ 와 $\Phi(B_s)$ 는 각각 비계절적 및 계절적 AR연산자, $\theta(B)$ 와 $\Theta(B_s)$ 는 각각 비계절적 및 계절적 MA연산자, d와 D는 각각 비계절적 및 계절적 차분 차수, p와 P는 각각 비계절적 및 계절적 AR차수, q와 Q는 각각 비계절적 및 계절적 MA차수이다.

$$\begin{aligned} \phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D Z_t &= \theta_q(B)\Theta_Q(B^s)\epsilon_t & (2) \\ \text{where, } \phi_p(B) &= 1 - \phi_1 B - \dots - \phi_p(B^p) \\ \theta_q(B) &= 1 - \theta_1 B - \dots - \theta_q(B^q) \\ \Phi_P(B^s) &= 1 - \Phi_1 B^s - \dots - \Phi_P(B^{Ps}) \\ \Theta_Q(B^s) &= 1 - \Theta_1 B^s - \dots - \Theta_Q(B^{Qs}), \epsilon_t \sim iid N(0, \sigma_\epsilon^2) \end{aligned}$$

SARIMA 모형의 전반적인 구축절차는 다음과 같다. 첫 번째로 데이터의 비정상성 및 계절성을 확인한다. 비정상성과 계절성이 존재한다면, 비계절적 및 계절 차분을 실시한다. 두 번째는 모형 식별단계이다. 이 단계에서는 비계절적 p, q와 계절적 P, Q의 계수를 결정한다. ACF와 PACF그래프의 관측을 통해 차수를 판단할 수 있다. 세 번째는 모형 추정단계이다. 식별단계에서 선택한 계수들을 정확히 추정하며 통계적으로 유의한지 여부를 확인한다. 네 번째는 모형 진단이다. 추정된 모델이 통계적으로 적합한지 결정하는 단계이다. 모형이 기각된

다면 식별단계부터 다시 되풀이한다. 최종적으로 확정된 모형을 이용하여 예측한다.

이하 논문 내용에서는 모형을 구분하기 위해 ARIMA 모형은 계절성을 고려하지 않은 모형으로, SARIMA 모형은 계절성을 고려한 모형으로 정의하도록 한다.

3. 태양광 발전량 예보모형 구축 실험

대전지역 4개 지점에 대한 태양광 발전량의 시계열 예보모형을 구축 할 때 1월부터 12월까지 모든 달의 특징을 반영되어야 한다. 따라서 1월부터 12월까지 모든 달의 특징을 모두 반영하기 위해서 태양광 발전량 데이터 중 초순과 중순(1일~20일)을 모형 구축(training) 용도로 하순(21일~말일)을 모형 평가(test) 용도로 구분하여 사용하였다. 데이터를 나눈 개요 그림은 Fig. 1과 같다.

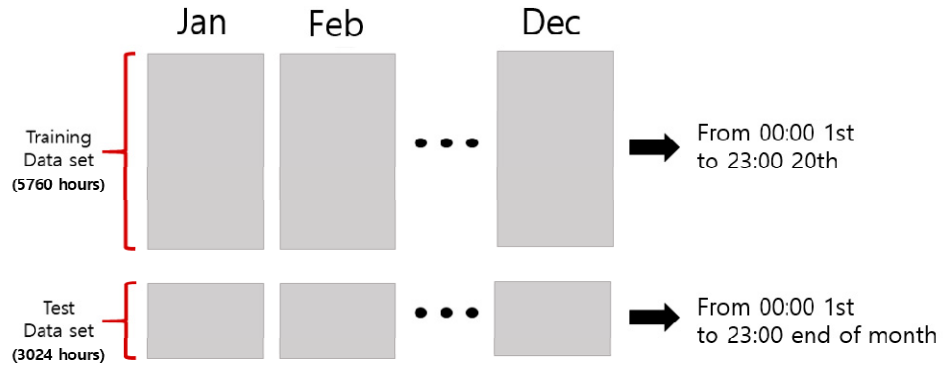


Fig. 1 Overview of training data set and test data set for solar power data

3.1 시계열 분석 기법을 이용한 태양광 발전량 예보모형

(1) ARIMA 기법을 이용한 예보모형

본 연구에서는 ARIMA(p,d,q) 모형에서의 p,d,q 차수를 결정하기 위해 통계적 모형 적합 판별 척도인 AIC와 SBC를 사용하였다. AIC(Akaike's information criterion)와 SBC(Schwartz's bayesian criterion) 모두 낮을수록 모형의 적합도가 우수함을 의미한다(Alkaikie, 1974, Schwarz, 1978)^{5,6,7}. Table 1은 AIC와 SBC 척도를 이용하여 4개 태양광 발전소 각 지점의 ARIMA 모형 차수를 결정한 결과를 나타낸 것이다.

Table 1 ARIMA model for 4 stations

Station	ARIMA (p,d,q)
Station 1	ARIMA (2,1,2)
Station 2	ARIMA (4,1,2)
Station 3	ARIMA (4,1,2)
Station 4	ARIMA (4,1,2)

Table 1의 결과를 바탕으로 각 지점에 해당하는 ARIMA 모형을 이용하여 1시간부터 3시간까지의 태양광 발전량을 예측하였다.

Fig. 2는 4개 지점 중 하나인 지점4의 모형 평가 기간 중 9월의 모형 평가 기간(2016년 09월 21일 00:00~2016년 09월 30일 23:00)의 실제 태양광 발전량 관측값과 1시간부터 3시간까지의 태양광 발전량 예측값을 나타낸 그림이다.

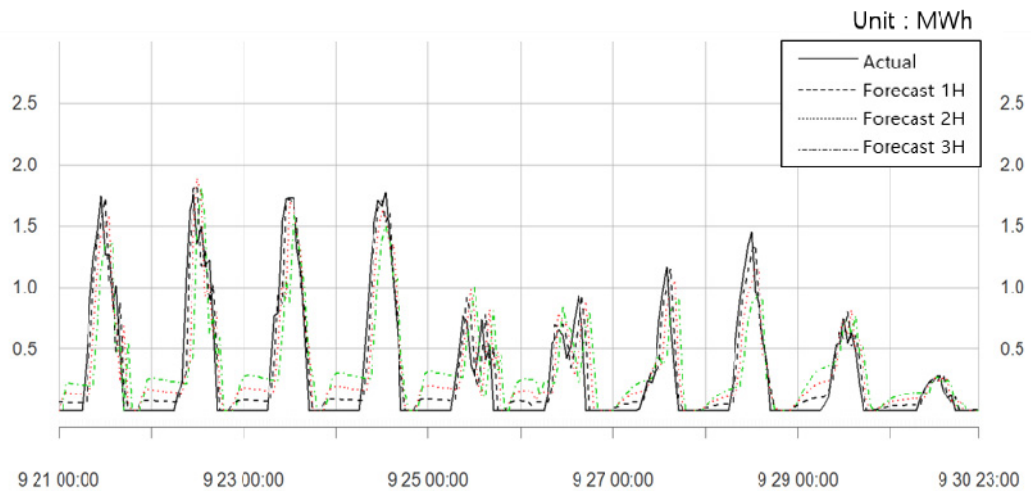


Fig. 2 The actual value and the predicted value(from 1 hour to 3 hour) of solar power generation at station 4 using ARIMA model

(2) SARIMA 기법을 이용한 예보 모형

SARIMA 모형도 위에서 한 ARIMA 방식과 마찬가지로 AIC와 SBC를 이용하여 모형의 차수를 결정하였다. Table 2는 AIC와 SBC 척도를 이용하여 4개 태양광 발전소 각 지점의 SARIMA 모형 차수를 결정한 결과를 나타낸 것이다.

Table 2 SARIMA model for 4 stations

Station	SARIMA (p,d,q) (P,D,Q) ₂₄
Station 1	SARIMA (4,0,1) (2,1,0) ₂₄
Station 2	SARIMA (2,0,1) (2,1,0) ₂₄
Station 3	SARIMA (2,0,3) (2,1,0) ₂₄
Station 4	SARIMA (2,0,1) (2,1,0) ₂₄

Table 2의 결과를 바탕으로 각 지점에 해당하는 SARIMA 모형을 이용하여 1시간부터 3시간까지의 태양광 발전량을 예측하였다.

Fig. 3은 4개 지점 중 하나인 지점4의 모형 평가 기간 중 9월(2016년 09월 21일 00:00~2016년 09월 30일 23:00)의 실제 태양광 발전량 관측값과 1시간부터 3시간까지의 태양광 발전량 예측값을 나타낸 그림이다.

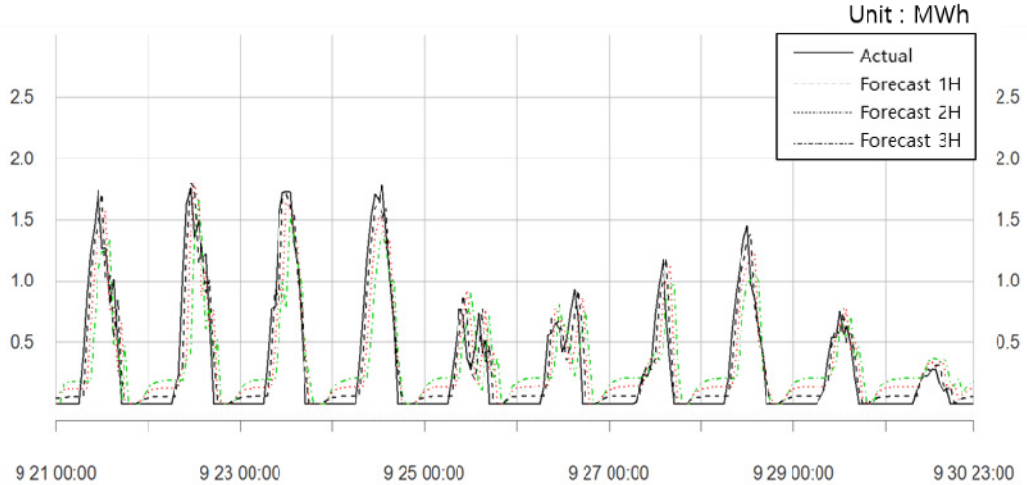


Fig. 3 The actual value and the predicted value (from 1 hour to 3 hour) of solar power generation at station 4 using SARIMA model

3.2 ARIMA 기법과 SARIMA 기법을 이용한 태양광 발전량 예보모형 예측 정확도 비교

본 연구에서는 시계열 분석 방법인 ARIMA와 SARIMA를 이용한 태양광 발전량 예보 모형의 예측 정확도를 비교하기 위해 MAE 측도와 RMSE측도를 이용하였다.

MAE와 RMSE은 추정값 또는 모델이 예측한 값과 실제 환경에서 관찰되는 값의 차이인 잔차를 비교할 때 흔히 사용하는 측도이다. MAE(Mean Absolute Error)는 평균 절대 오차로, (예측값-실제값)의 절대값의 평균을 구한 값이다. RMSE(Root Mean Square Error)는 평균 제곱근 오차로, (예측값-실제값)의 제곱의 평균을 구하고 그것의 제곱근을 구한 값이다. MAE와 RMSE는 낮은 값을 가질수록 보다 정확한 예측을 한 모형이다.

Table 3은 전체 모형 평가(Test) 구간(매월 21일~말일)의 태양광 발전량 실제 관측값과 1시간부터 3시간까지의 예측값 간의 MAE를 나타낸 결과이다.

Table 3 Comparison of MAE between ARIMA model and SARIMA model

[Unit : MWh]

Station	ARIMA model			SARIMA model		
	Forecast 1 H	Forecast 2 H	Forecast 3 H	Forecast 1 H	Forecast 2 H	Forecast 3 H
Station 1	0.0032	0.0061	0.0090	0.0037	0.0066	0.0090
Station 2	0.0936	0.1638	0.2224	0.0966	0.1720	0.2330
Station 3	0.0871	0.1527	0.2068	0.0899	0.1591	0.2153
Station 4	0.1006	0.1769	0.2424	0.1054	0.1852	0.2509

Table 4는 전체 모형 평가(Test) 구간(매월 21일~말일)의 태양광 발전량 실제 관측값과 1시간부터 3시간까지의 예측값 간의 RMSE를 나타낸 결과이다.

Table 4 Comparison of RMSE between ARIMA model and SARIMA model

[Unit : MWh]

Station	ARIMA model			SARIMA model		
	Forecast 1 H	Forecast 2 H	Forecast 3 H	Forecast 1 H	Forecast 2 H	Forecast 3 H
Station 1	0.0060	0.0105	0.1490	0.0057	0.0091	0.0119
Station 2	0.1563	0.2431	0.3190	0.1493	0.2414	0.3102
Station 3	0.1418	0.2262	0.2919	0.1396	0.2244	0.2865
Station 4	0.1623	0.2615	0.3461	0.1612	0.2592	0.3361

4. 결론

위의 Tables 3과 4를 통해 ARIMA 기법을 이용한 태양광 발전량 예보모형과 SARIMA 기법을 이용한 태양광 발전량 예보모형을 비교한 결과를 정리하면 다음과 같다.

- (1) 4개 지점의 1시간부터 3시간까지의 태양광 발전량의 예측 오차를 MAE 측도에 의한 예측 오차를 비교해 본 결과 ARIMA를 이용하여 구축한 태양광 발전량 예보모형이 SARIMA를 이용한 태양광 발전량 예보모형보다 예측 정확도 측면에서 더 우수하였다. 반면, RMSE 측도에 의한 예측 오차를 비교해 본 결과 SARIMA를 이용하여 구축한 태양광 발전량 예보모형이 ARIMA를 이용한 태양광 발전량 예보모형보다 예측 정확도 측면에서 더 우수하였다.
- (2) 본 논문에서는 2016년 1월 1일 00시부터 2016년 12월 31일까지 1년의 데이터를 이용하여 ARIMA 기법과 SARIMA 기법을 이용한 태양광 발전량 예보모형의 예측 정확도를 비교하였다. 추후 연구에서는 2016년 만의 데이터가 아니라 더 긴 기간의 데이터를 이용하여 예보모형을 구축한다면 ARIMA 기법을 이용한 예보 모형과 SARIMA 기법을 이용한 예보모형 사이의 예측 정확도 차이를 더 명확하게 알 수 있을 것으로 기대된다.
- (3) 추후 예보 지점을 확대하여 전국 단위 태양광 발전량 예보모형 구축을 위해 다양한 시계열 모형 적용 방안을 연구해보고자 한다. 하지만 모든 지점에 대한 발전량을 예측하기 위해서는 발전소의 규모 및 용량이 다양하고, 발전소의 개수도 많기때문에 발전소마다 개별 예측을 한다면 많은 시간과 비용이 든다. 따라서 여러 개의 지점을 군집화하여 군집으로 묶인 지점들의 연관성을 바탕으로 군집별 태양광 발전량 예보모형 구축을 하고자 한다⁸⁾. 이는 전국 단위 태양광 발전량 예보모형 구축에 있어 효율성을 증대시키는 효과가 있을 것으로 기대된다.

후기

본 연구는 한국에너지기술연구원의 주요사업으로 수행한 결과입니다 (과제번호: B9-2414).

REFERENCES

1. Reikard, G., Predicting Solar Radiation at High Resolutions: A Comparison of Time Series Forecasts, *Solar Energy*, 83 (2009) 342-349, 2008.
2. Moreno-Muñoz, A., Short Term Forecasting of Solar Radiation, *IEEE International Symposium on Industrial Electronics*, 2008.
3. Kim, S., A Study on Solar Irradiance Forecasting with Weather Variables, *The Korean Journal of Applied Statistics*, 30(6), 1005-1013, 2017.
4. Akaike, H., A New Look at Statistical Model Identification, *IEEE Transactions on Automatic Control*, AU-19, 716-722, 1974.
5. Schwarz, G., Estimating the Dimension of a Model, *Annals of Statistics*, 6, 461-464, 1978.
6. Min, K. C., Jun, Y. I., and Ha, Y. I., Forecasting the Air Cargo Demand With Seasonal ARIMA Model: Focusing on ICN to EU Route, *Journal of Korean society of Transportation*, Vol. 31 , No. 3, pp. 3-18, 2013.
7. Kim, H. G., Lee, Y. S., and Jang, M. S., Cluster Analysis and Meteor-Statistical Model Test to Develop a Daily Forecasting Model for Jejudo Wind Power Generation, *Journal of Environmental Science International*, Vol. 19, No. 10, pp. 1229-1235, 2010.
8. Wei, W. S., *Time Series Analysis: Univariate and Multivariate Methods*, 2nd Edition, pp. 394-398, 2005.