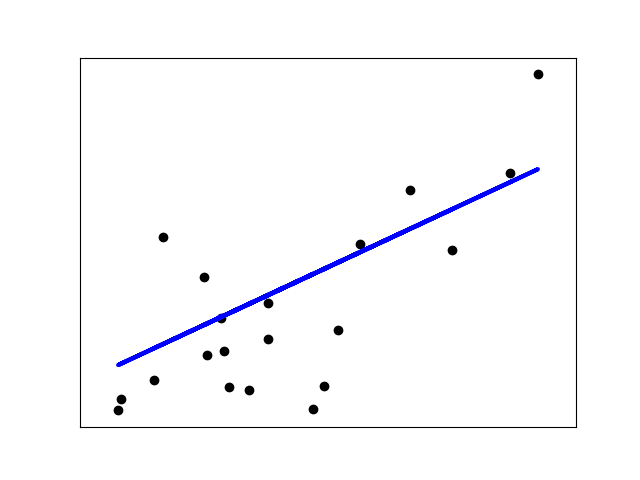
1. **Intro: Prediction and interpretation**
2. **The aim of statistical analysis in social science is mostly about interpretation.**
3. **Prediction is widely used in the field of engineering.**

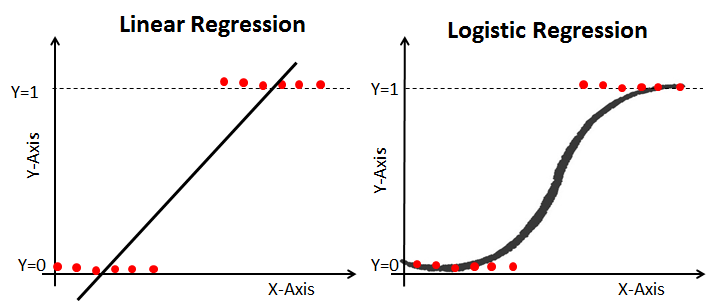


**Keywords** : interpretation, explanation, decision making

1. **Predictive Analysis**
2. **Logistic Regression model**

* What is the probability that the region A will be “응급의료취약지역”?

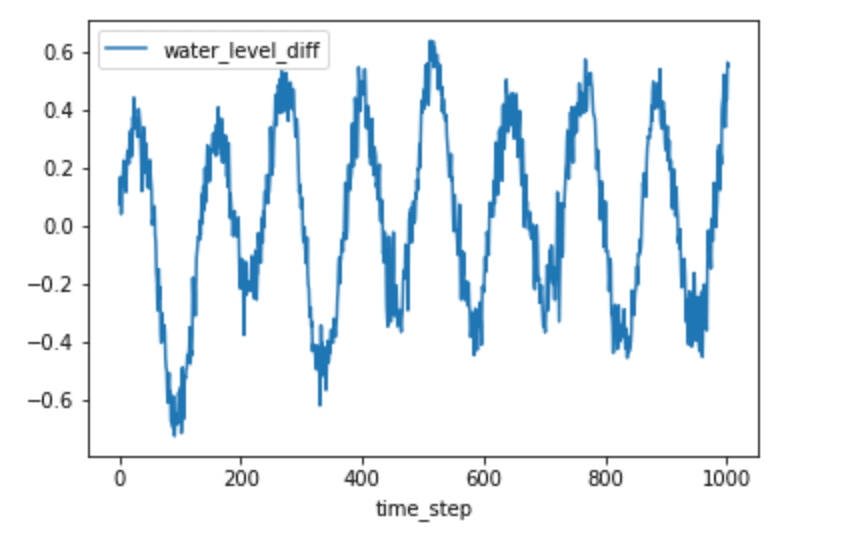
|  |  |  |  |
| --- | --- | --- | --- |
| 1일 섭취량 | 발병 | 정상 | n |
| 2 | 4 | 17 | 21 |
| 3 | 6 | 12 | 18 |
| 4 | 7 | 12 | 19 |
| 5 | 9 | 13 | 22 |
| 6 | 11 | 9 | 20 |

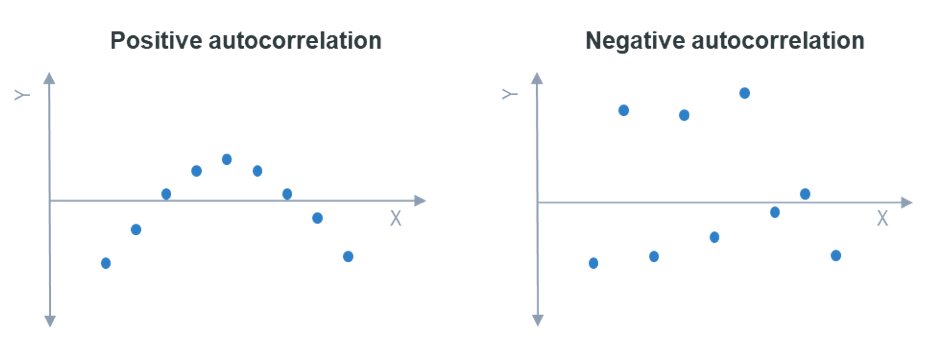


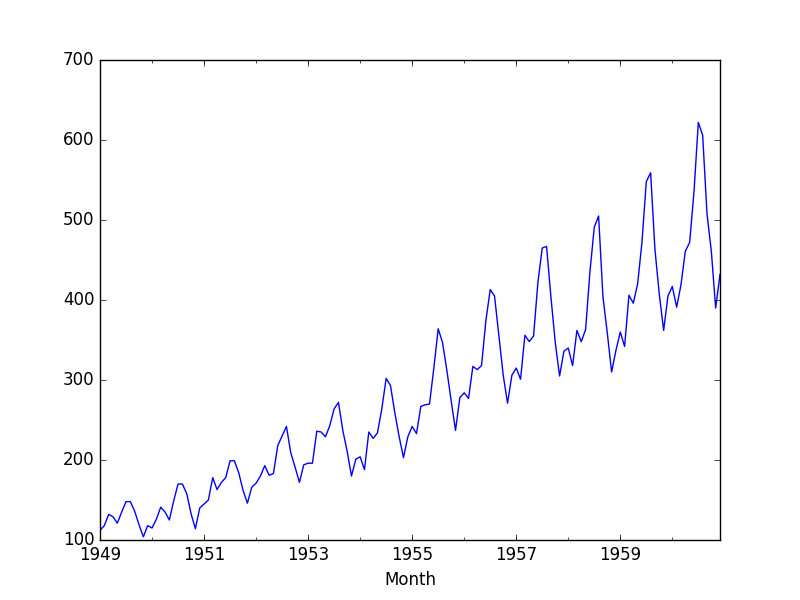
1. **Time Series Analysis**

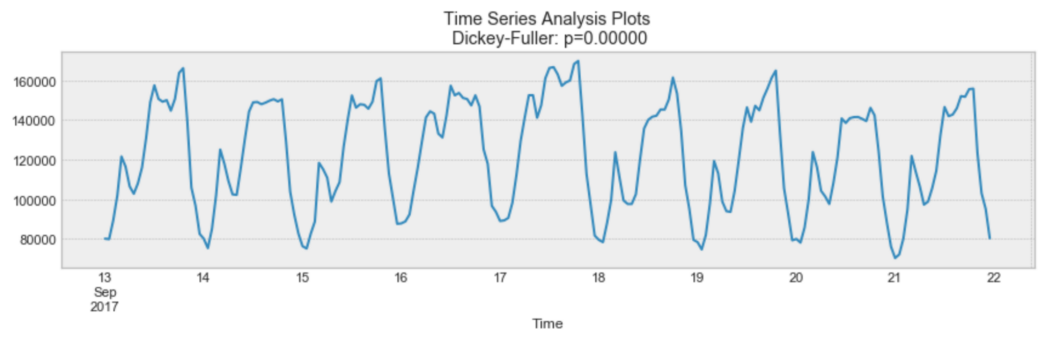
* How many people will be in the region A in the next month?
* Autocorrelation

In common cases, it is assumed that the observations are independent each other e.g., iid(independently and identically distributed). That is, . However in time series analysis, it is assumed that there is covariance between observations, .





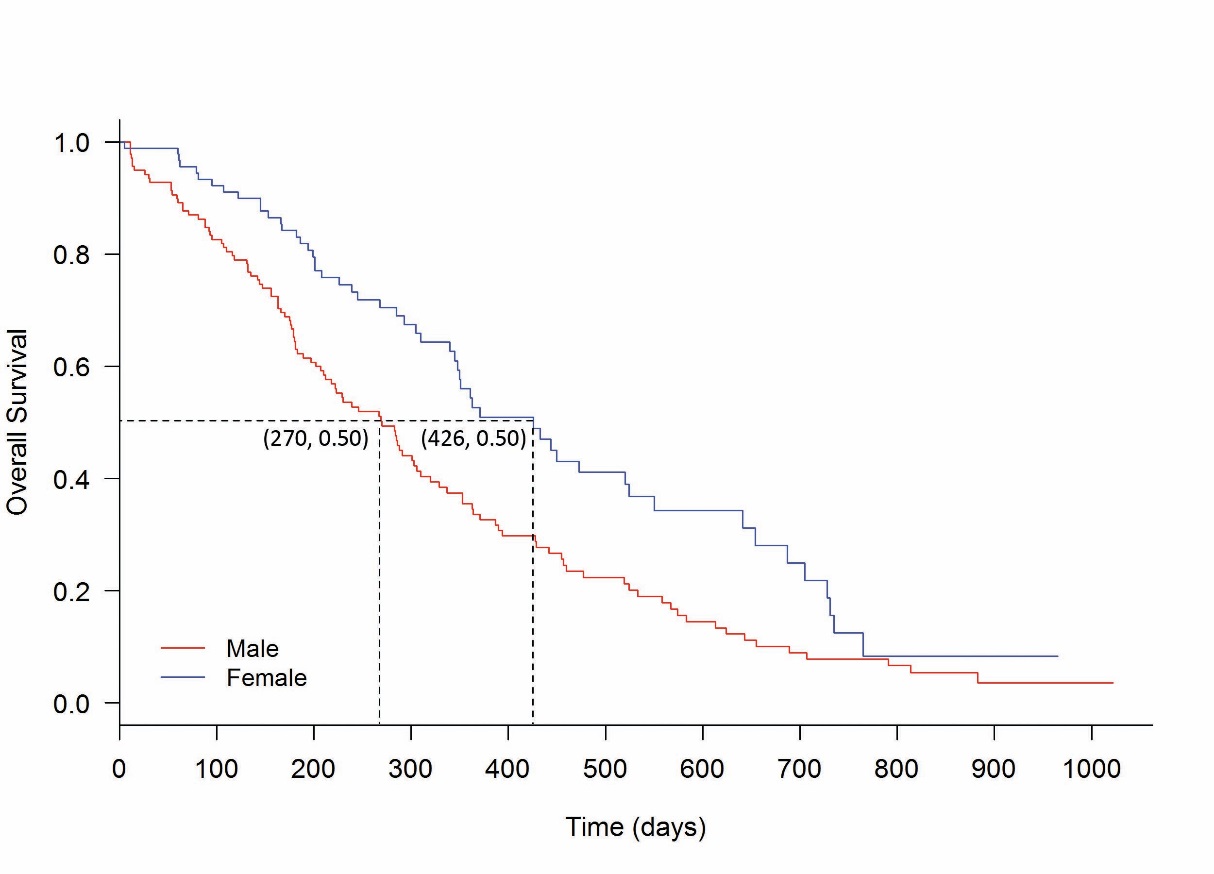
* Stationary
* Seasonality



* AR model
* MA model
* ARMA, ARIMA model
* SARIMA model

1. **Survival Analysis**

* When will the region A be “응급의료취약지역”?
* Survival function :
* Lifetime Distribution Function :



* Hazard Function

**Example)** Let’s say we have 500 graduate students in our sample and (amazingly), 15 of them (3%) manage to finish their dissertation in the first year after advancing.

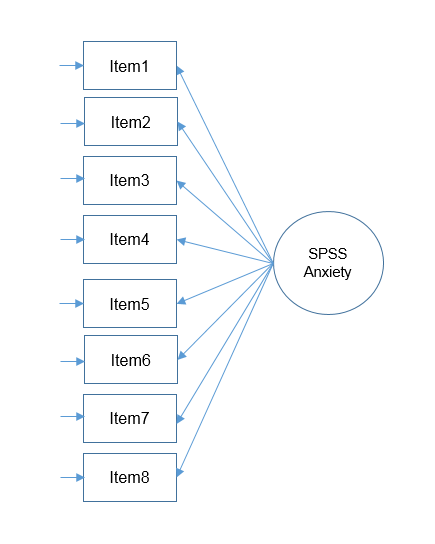
Our first year hazard, the probability of finishing within one year of advancement, is .03. That is the number who finished (the event occurred)/the number who were eligible to finish (the number at risk).

In the first year, that’s 15/500. 15 finished out of the 500 who were eligible.

Now let’s say that in the second year 23 more students manage to finish. The second year hazard is 23/485 = .048. You’ll notice this denominator is smaller than the first, since the 15 people who finished in year 1 are no longer in the group who is “at risk.”

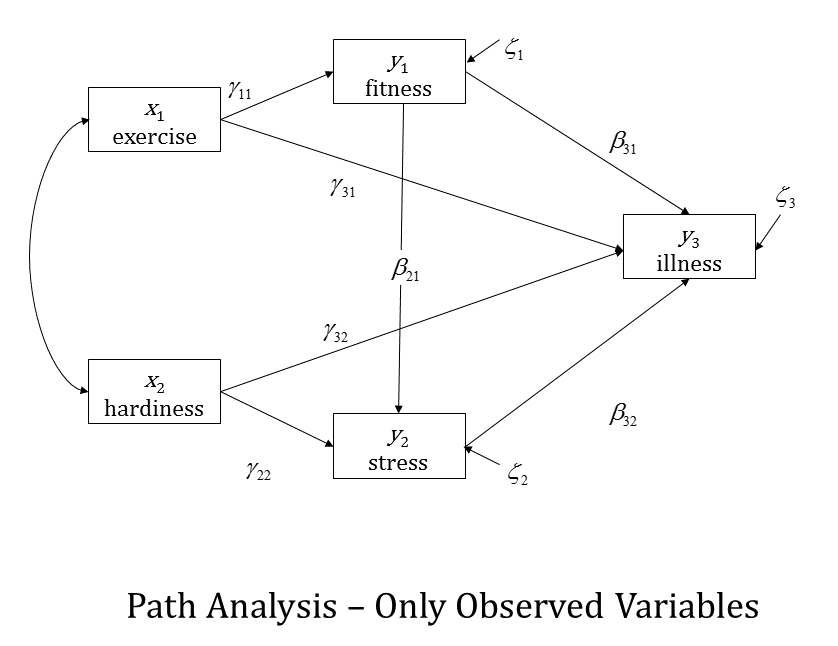
1. **Prescriptive Analysis**
2. **Factor Analysis**

* What are the latent traits which lies in this dataset?



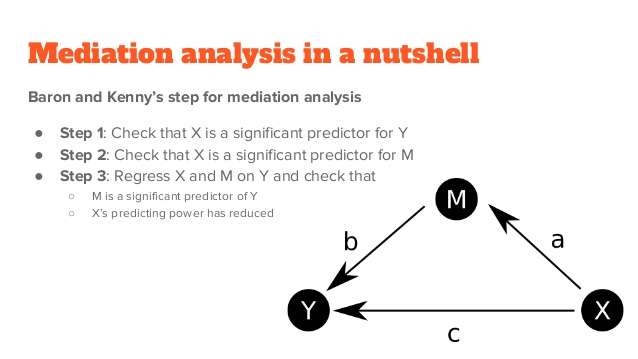
1. **Path Analysis**

* What is the relationship between the traits?



1. **Causal Analysis**
2. **Mediation Analysis**

* How does the “population size” affect “응급의료취약지역”?



3. Insert 2) to 1)

Total effect :

Indirect effect :

Direct effect :