

## 1. Wilk's Lambda 란[1]

검증 통계치(test statistics)에서 많이 사용되는 검정 통계량

Wilks lambda ranges from 0 – 1 and the lower the Wilks lambda, the larger the between group

dispersion. A small (close to 0) value of Wilks' lambda means that the groups are well separated. A large (close to 1) value of Wilks' lambda means that the groups are poorly separated.

- 0 ~ 1 사이의 값을 가짐
- 1에 가까우면 그룹들이 잘 분류되지 않음
- 0에 가까우면 그룹들이 잘 분류됨

## 2. Wilk's Lambda 식 ( $\Lambda$ )

즉, Wilk's lambda 수치가 작아야 그룹들을 잘 분류해주는 중요한 변수이다.

Within Variance

$$\Lambda = \frac{|W|}{|T|} = \frac{|W|}{|W + B|}$$

Total Variance

$$T = \sum_{i=1}^g \sum_{j=1}^{n_i} (X_{ij} - \bar{X})(X_{ij} - \bar{X})'$$

$$W = \sum_{i=1}^g \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)'$$

$$B = \sum_{i=1}^g n_i (\bar{X}_i - \bar{X})(\bar{X}_i - \bar{X})'$$

## 1. 변수의 중요도

**library(klaR)**

```
data("B3") # West German Business Cycles 1955-1994
head(B3)
```

```
## PHASEN BSP91JW CP91JW DEFRATE EWAJW EXIMRATE GM1JW IAU91JW IB91JW
## 1955,4 2 10.53 9.31 0.05 5.7 3.08 11.15 23.56 14.69
## 1956,1 2 10.60 12.66 0.06 5.2 1.96 11.03 12.72 24.95
## 1956,2 3 9.21 6.55 0.05 4.8 2.82 10.04 11.52 14.90
## 1956,3 3 5.17 7.87 0.05 3.3 3.74 8.33 0.85 7.55
## 1956,4 3 4.93 8.60 0.04 2.1 4.16 7.69 -2.08 3.23
## 1957,1 3 8.39 5.62 0.04 3.2 2.90 6.62 -3.76 14.58
## LSTKJW PBSPJW PCPJW ZINSK ZINSLR
## 1955,4 3.00 2.89 1.91 6.27 3.21
## 1956,1 2.36 2.59 2.20 4.60 3.54
## 1956,2 3.39 3.01 3.09 6.19 3.22
## 1956,3 5.30 3.03 2.08 6.71 3.37
## 1956,4 6.91 3.46 1.48 7.10 3.14
## 1957,1 1.03 1.95 1.65 4.96 4.95
```

str(B3)

factor 자료구조

```
## 'data.frame': 157 obs. of 14 variables:
## $ PHASEN : Factor w/ 4 levels "1","2","3","4": 2 2 3 3 3 3 3 3 3 ...
## $ BSP91JW: num 10.53 10.6 9.21 5.17 4.93 ...
## $ CP91JW : num 9.31 12.66 6.55 7.87 8.6 ...
## $ DEFRATE: num 0.05 0.06 0.05 0.05 0.04 0.04 0.04 0.03 0 ...
## $ EWAJW : num 5.7 5.2 4.8 3.3 2.1 3.2 2.5 2.7 3 0.3 ...
## $ EXIMRATE: num 3.08 1.96 2.82 3.74 4.16 2.9 3.65 4.57 4.37 2.89 ...
## $ GM1JW : num 11.15 11.03 10.04 8.33 7.69 ...
## $ IAU91JW: num 23.56 12.72 11.52 0.85 -2.08 ...
## $ IB91JW : num 14.69 24.95 14.9 7.55 3.23 ...
## $ LSTKJW : num 3 2.36 3.39 5.3 6.91 1.03 3.73 6.2 4.12 7.94 ...
## $ PBSPJW : num 2.89 2.59 3.01 3.03 3.46 1.95 3.18 3.98 3.29 5.63 ...
## $ PCPJW : num 1.91 2.2 3.09 2.08 1.48 1.65 1.47 3.29 3.59 4.19 ...
## $ ZINSK : num 6.27 4.6 6.19 6.71 7.1 4.96 5.21 4.83 4.5 3.83 ...
## $ ZINSLR : num 3.21 3.54 3.22 3.37 3.14 4.95 3.82 3.09 3.91 1.47 ...
```

```
# AIC 이용한 변수 선택법: PHASEN이 범주형 변수이므로 사용할 수 없음.  
# step(model, direction = "both")
```

```
# Wilks.lambda: 집단내 분산/ 총분산  
# 종속변수에 미치는 영향력에 따라 변수의 중요도를 정리 (작을수록 적함)
```

```
greedy.wilks(PHASEN ~ ., data = B3, niveau = 0.1)
```

다양한 "범주형 변수"가 위치해야 함.

```
## Formula containing included variables:  
##  
## PHASEN ~ EWAJW + LSTKJW + ZINSK + CP91JW + IAU91JW + PBSPJW +  
## ZINSLR + PCPJW  
## <environment: 0x00000000142a8998>  
##  
##  
## Values calculated in each step of the selection procedure:  
##  
## vars Wilks.lambda F.statistics.overall p.value.overall  
## 1 EWAJW 0.6058201 33.18341 1.405358e-16  
## 2 LSTKJW 0.4271561 26.85606 1.218146e-25  
## 3 ZINSK 0.3614525 21.20584 7.607587e-29  
## 4 CP91JW 0.3002868 19.05337 1.153881e-32  
## 5 IAU91JW 0.2624925 17.11094 6.597858e-35  
## 6 PBSPJW 0.2451025 14.99388 3.695840e-35  
## 7 ZINSLR 0.2205325 13.94619 1.442943e-36  
## 8 PCPJW 0.1999847 13.10739 9.454573e-38  
## F.statistics.diff p.value.diff  
## 1 33.18341 1.405358e-16  
## 2 21.19203 1.554268e-11  
## 3 9.14942 1.326989e-05  
## 4 10.18453 3.783582e-06  
## 5 7.15112 1.604993e-04  
## 6 3.50019 1.708972e-02  
## 7 5.45920 1.379166e-03  
## 8 5.00033 2.486333e-03
```