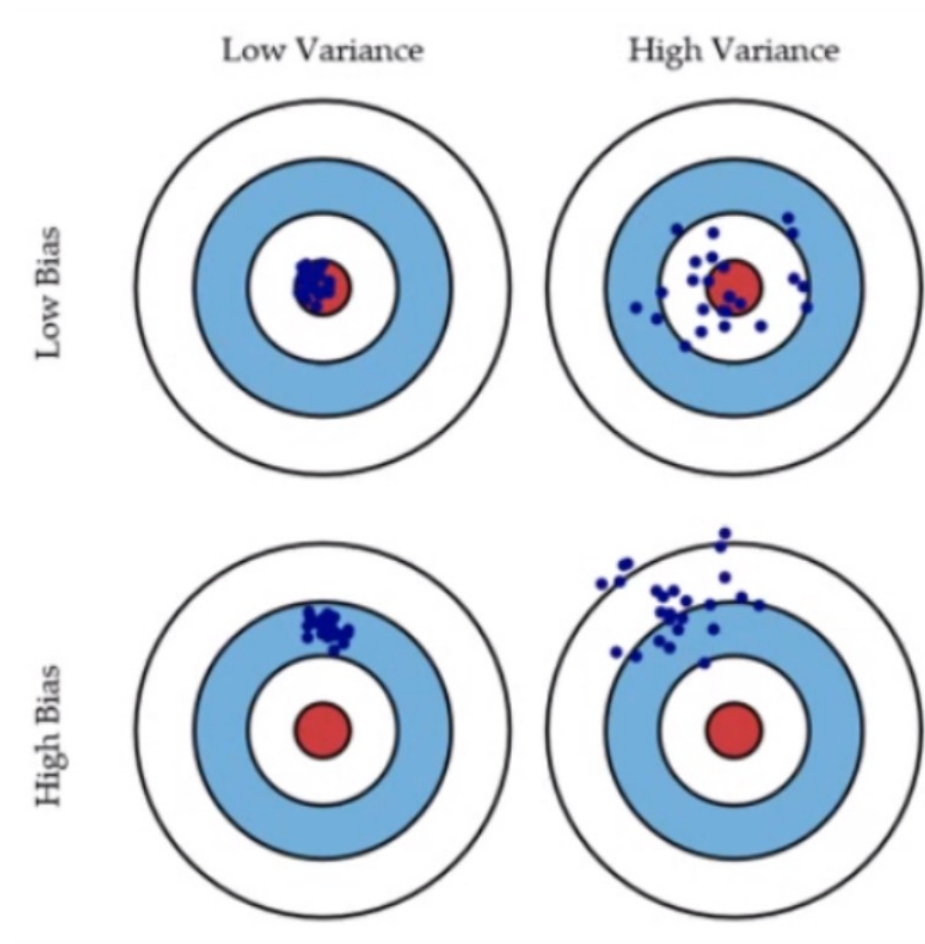


# 네트워크 분석

산업경영공학과 박혜원

# 1. Lasso



Bias와 Variance 모두 줄여주면 좋겠지만  
둘 중 하나를 포기해야 될 수도 있음

Overfitting을 해결함으로써 적당한 지점을 찾을 수 있음

## Overfitting을 해결하기 위한 방법

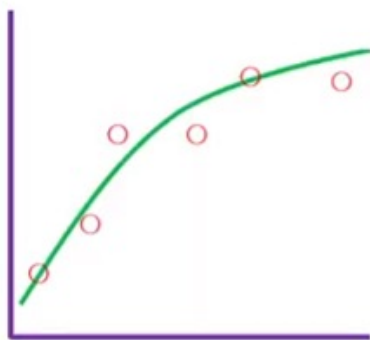
1. 특성의 수 줄이기 (주요 특징 직접 선택)
2. 정규화 수행 (모든 특성 사용, 베타값 조정하기)

# 1. Lasso



$$\beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 x^4$$

설명변수의 수가 많을 수록 Overfitting



$$\beta_0 + \beta_1 x + \beta_2 x^2$$

일반성을 띄고  
해석력을 높임

$$\min_{\beta} \sum_{i=1} (y_i - \hat{y}_i)^2 + 5000\beta_3^2 + 5000\beta_4^2$$

상단의 식을 가장 최소화 하는 베타값은 0

베타값에 제약을 줌으로써 모델 정돈해주는 것이 **정규화**

# 1. Lasso

$$\sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p |\beta_j| = \text{RSS} + \lambda \sum_{j=1}^p |\beta_j|.$$

일반화 정확도  
(Generalization Accuracy)  
: 새로운 Data에 대해  
잘 작동하는가?

$$\begin{aligned} \text{RSS} &= \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \\ &= \sum_{i=1}^n (Y_i - \hat{b}_0 - \hat{b}_1 X_i)^2 \end{aligned}$$

잔차 제곱합  
(Training Accuracy)

람다 커질수록 더 적은 베타 사용

$\lambda=1$

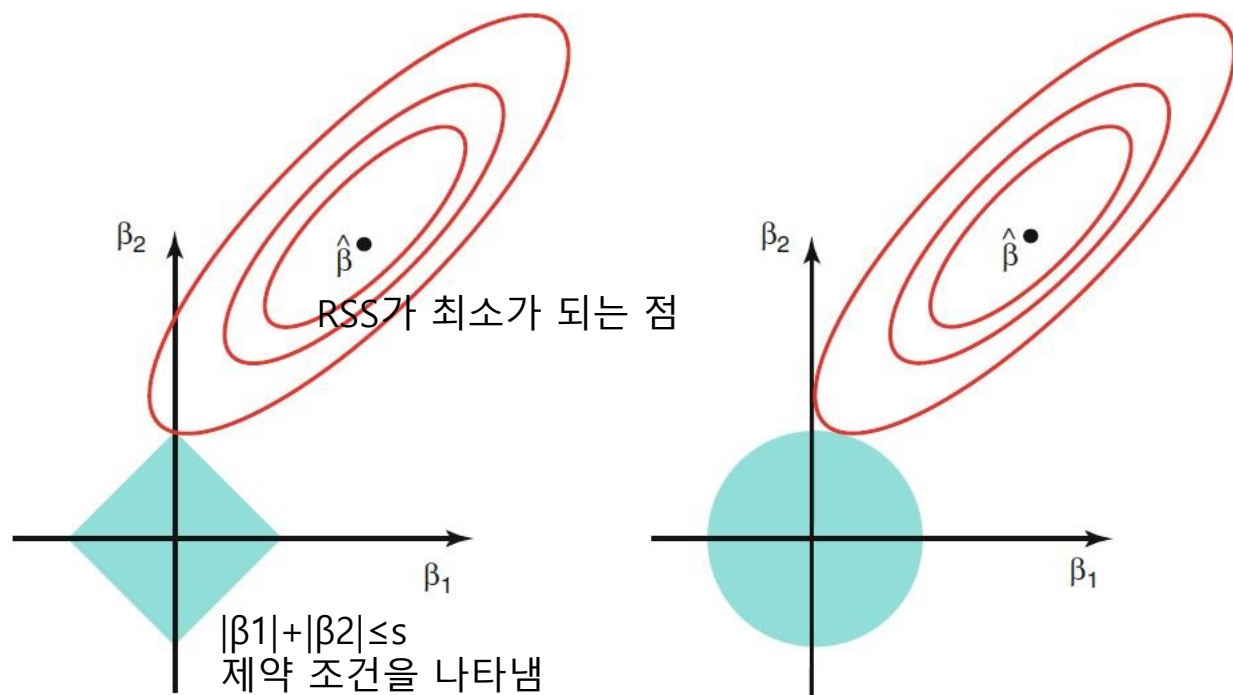
100개의  $\beta$  중 96개가 0

$\lambda=0.0001$

100개의  $\beta$  중 6개가 0이고 94개 변수 사용

$\lambda=0.1$   $\beta$  33개 사용 (적정)

# 1. Lasso



$\beta_1, \beta_2$  두 개의 베타값을 가지는  
간단한 선형 모델을 푼다고 가정할 때,

계수가 적당히 작은 마름모 내 베타 값을 찾기 위해서는,  
빨간선(MSE)과 맞닿아 있는 값 중에 골라야 됨

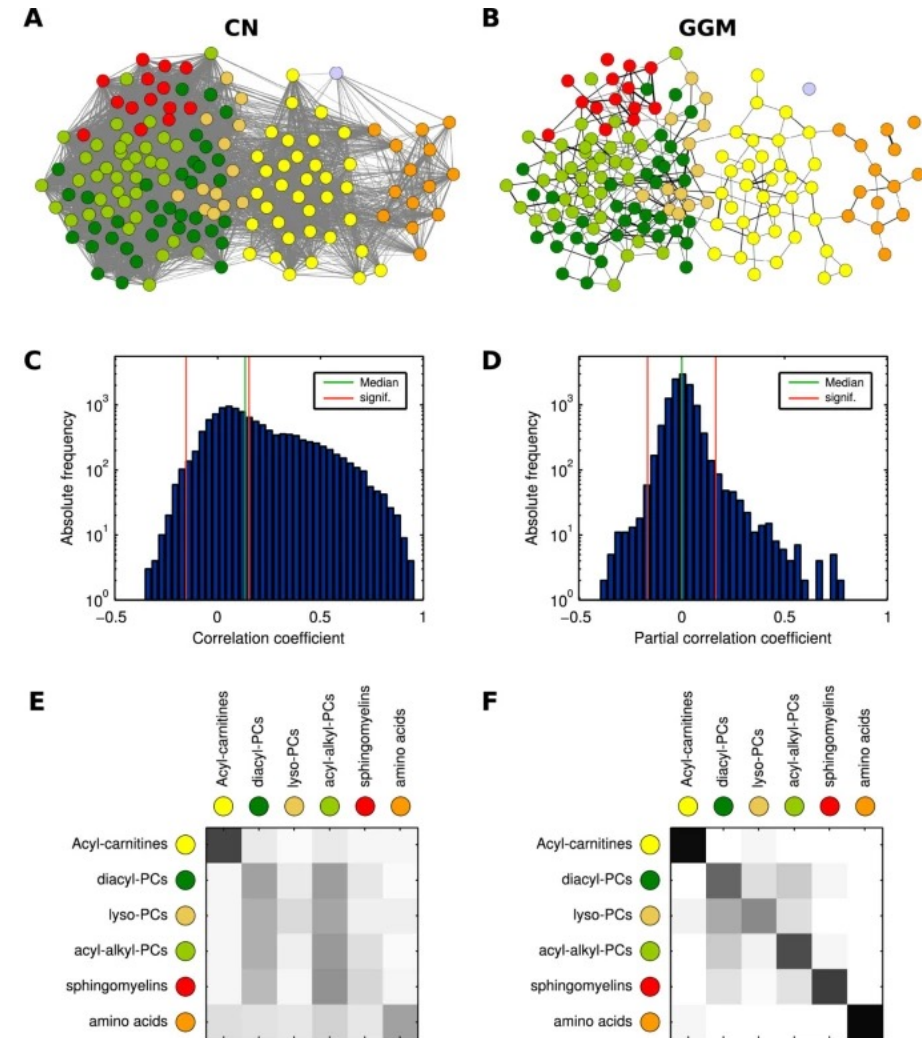
Lasso는 일부 베타값이 0에 수렴하는 특징 잘 드러남  
따라서 Lasso는 Ridge와 다르게 변수 선택이 가능하지만  
변수 간 상관 관계가 높은 상황에서는  
Ridge에 비해 상대적으로 예측 성능이 떨어짐

## 2. Gaussian Graphical Model

CN(Correlation Network)보다 훨씬 안정적이고  
편상관관계 (Partial Correlation Coefficient)에 기반함

CN: 직/간접적 연관성 구분이 없고  
대규모 데이터에서 연관성이 과다하게 나타남

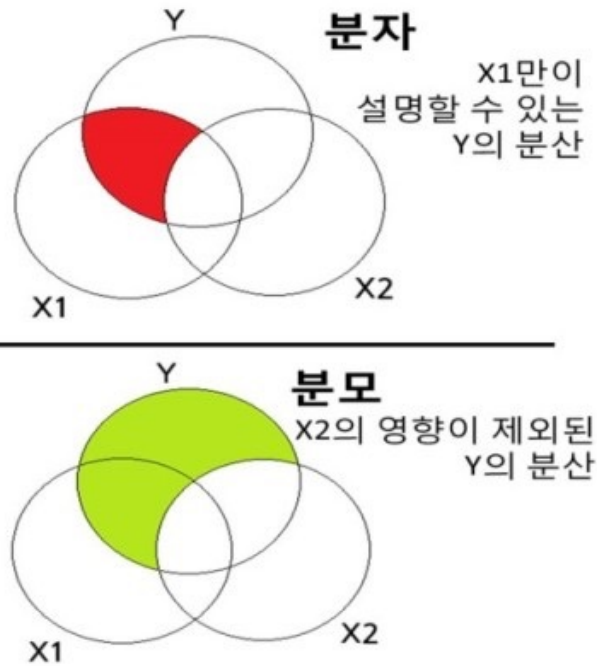
GGM: 직/간접적 연관성 정확히 구분함  
조건부 의존성을 평가함으로써 간접 연관 효과를 회피함



## 2. Gaussian Graphical Model

편상관관계  
Squared **partial**  
**correlation**  
(부분상관의 제곱 값)

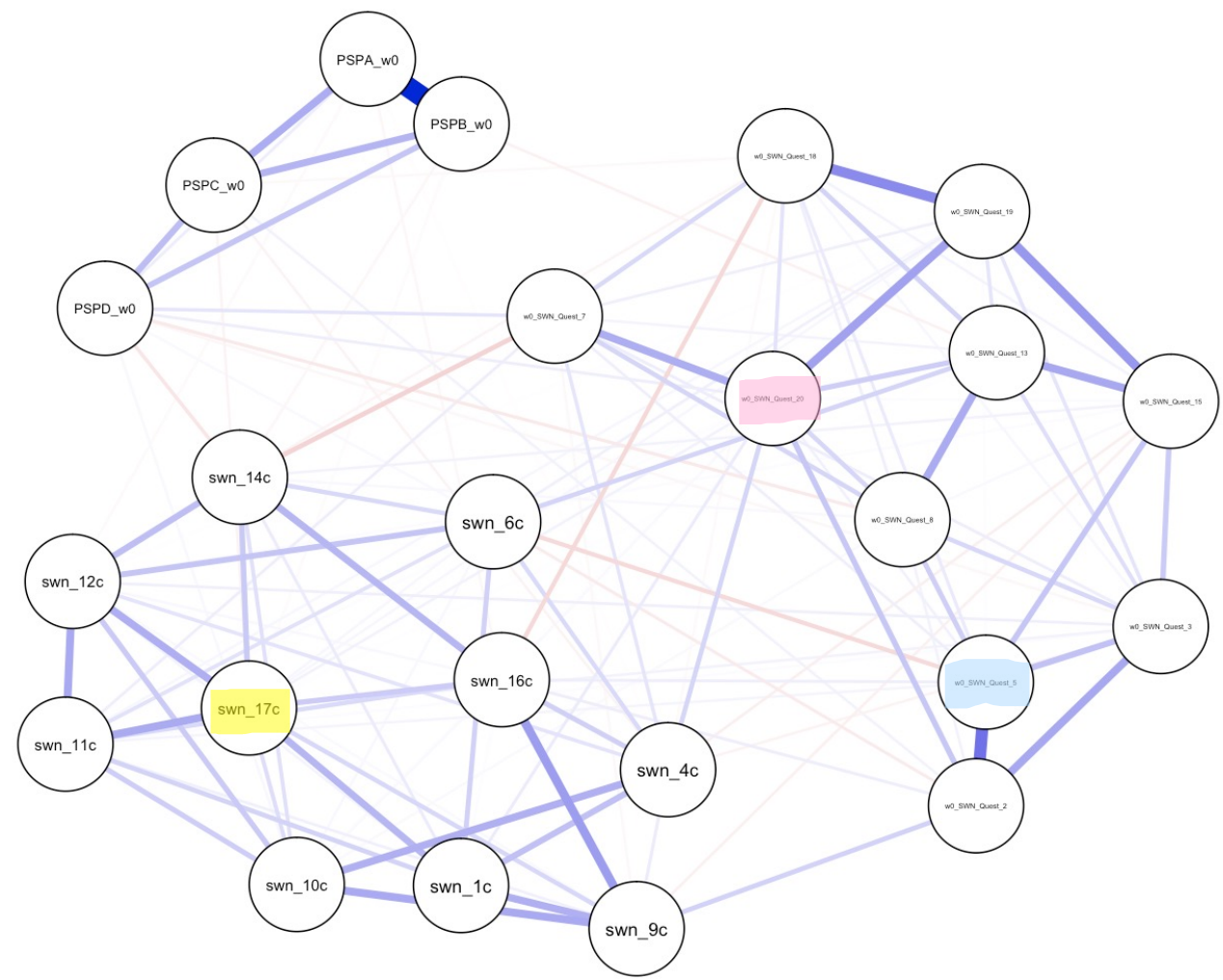
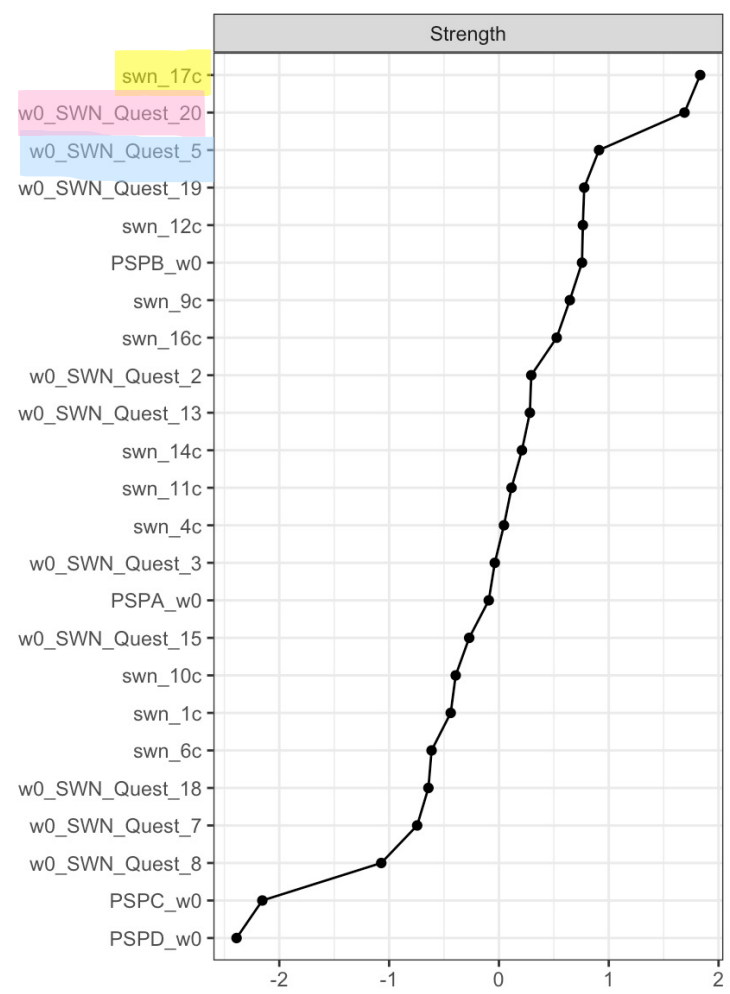
=



Y와 X1의 상관을 알고 싶을 때  
X1과 Y 외 다른 모든 변수를 통제

상관관계는 X1, Y의 연관성에 의해 직접 결정됨

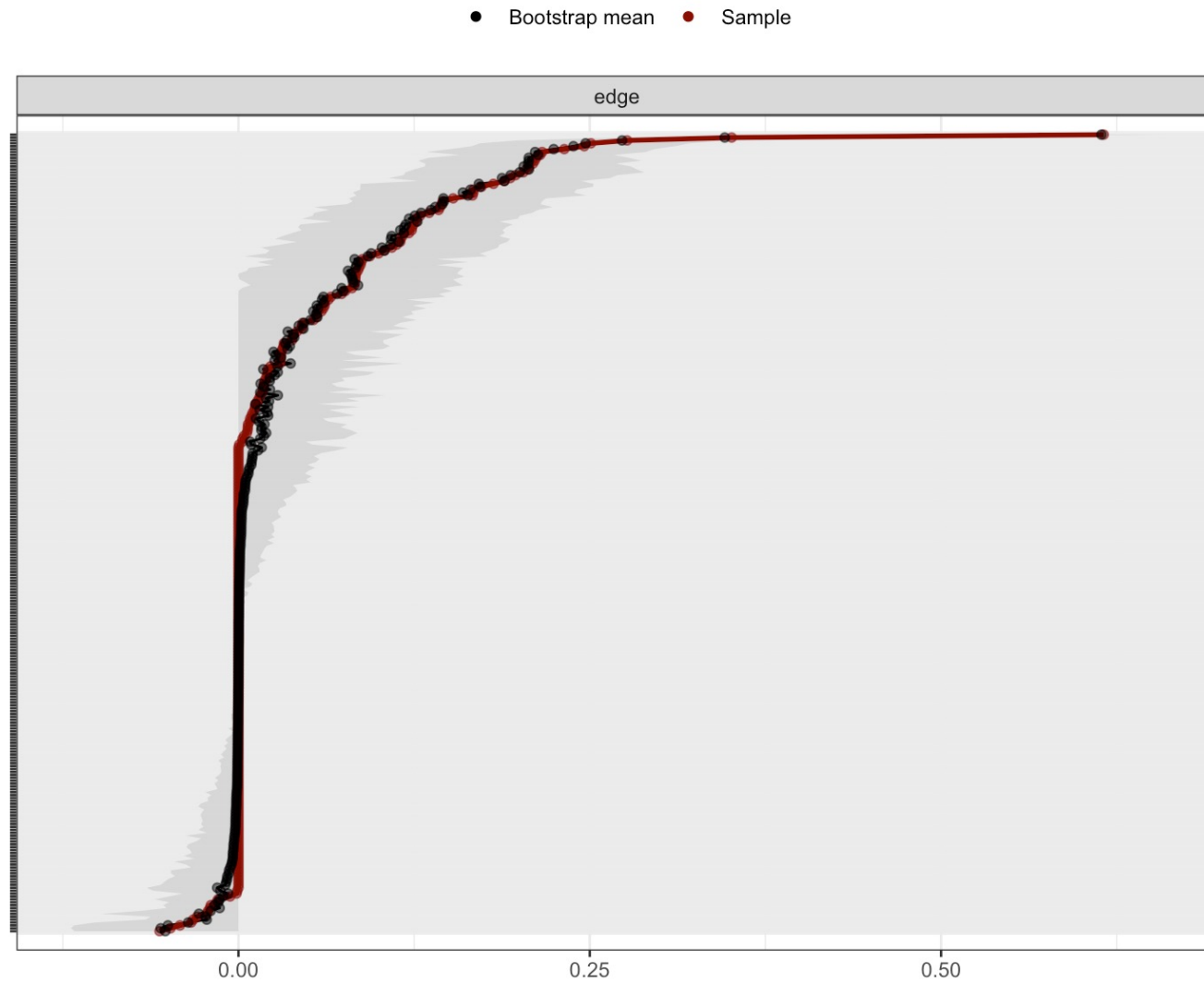
# 3. Bootnet





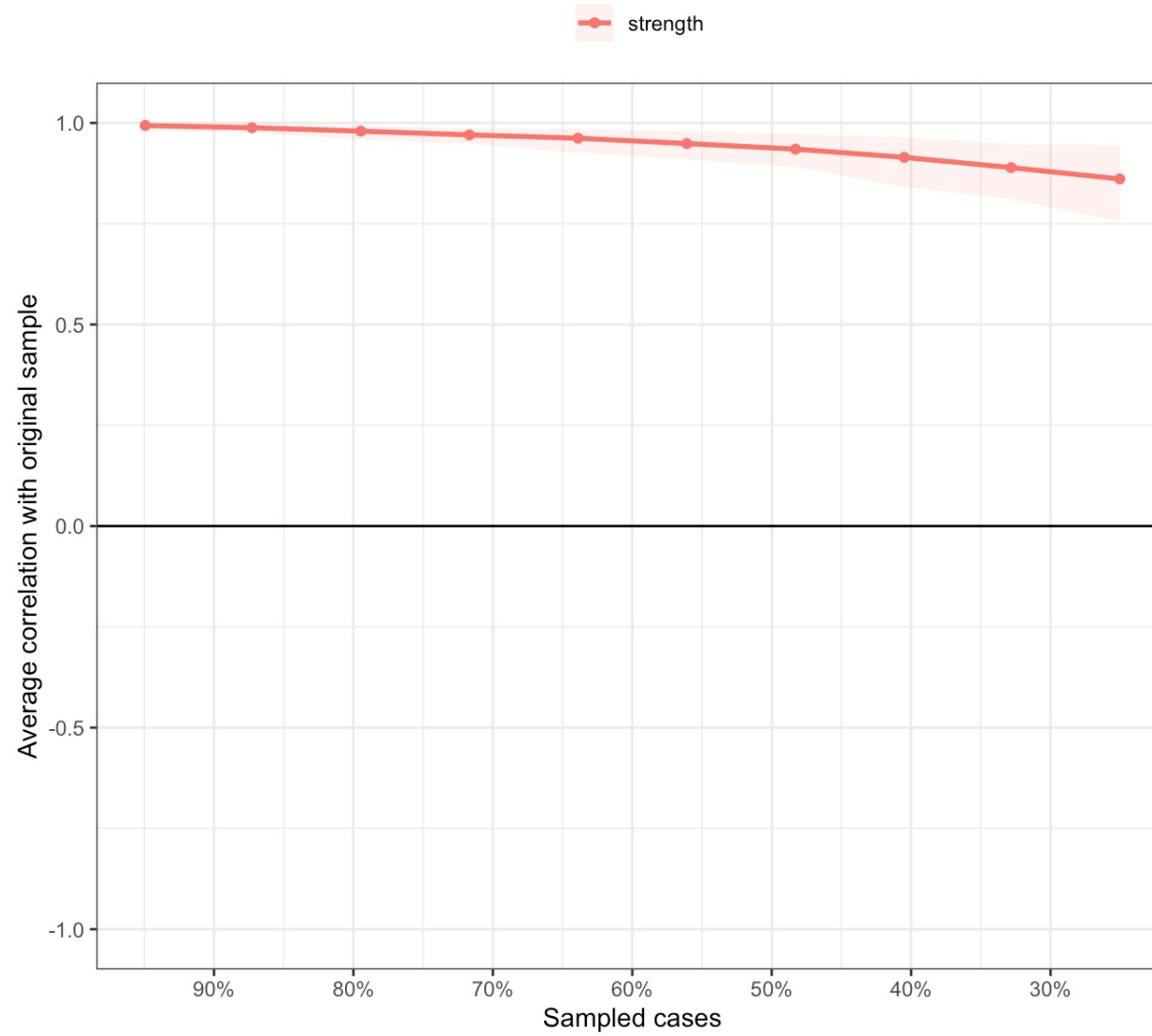
### 3. Bootnet

Edge Weight  
신뢰구간 추정

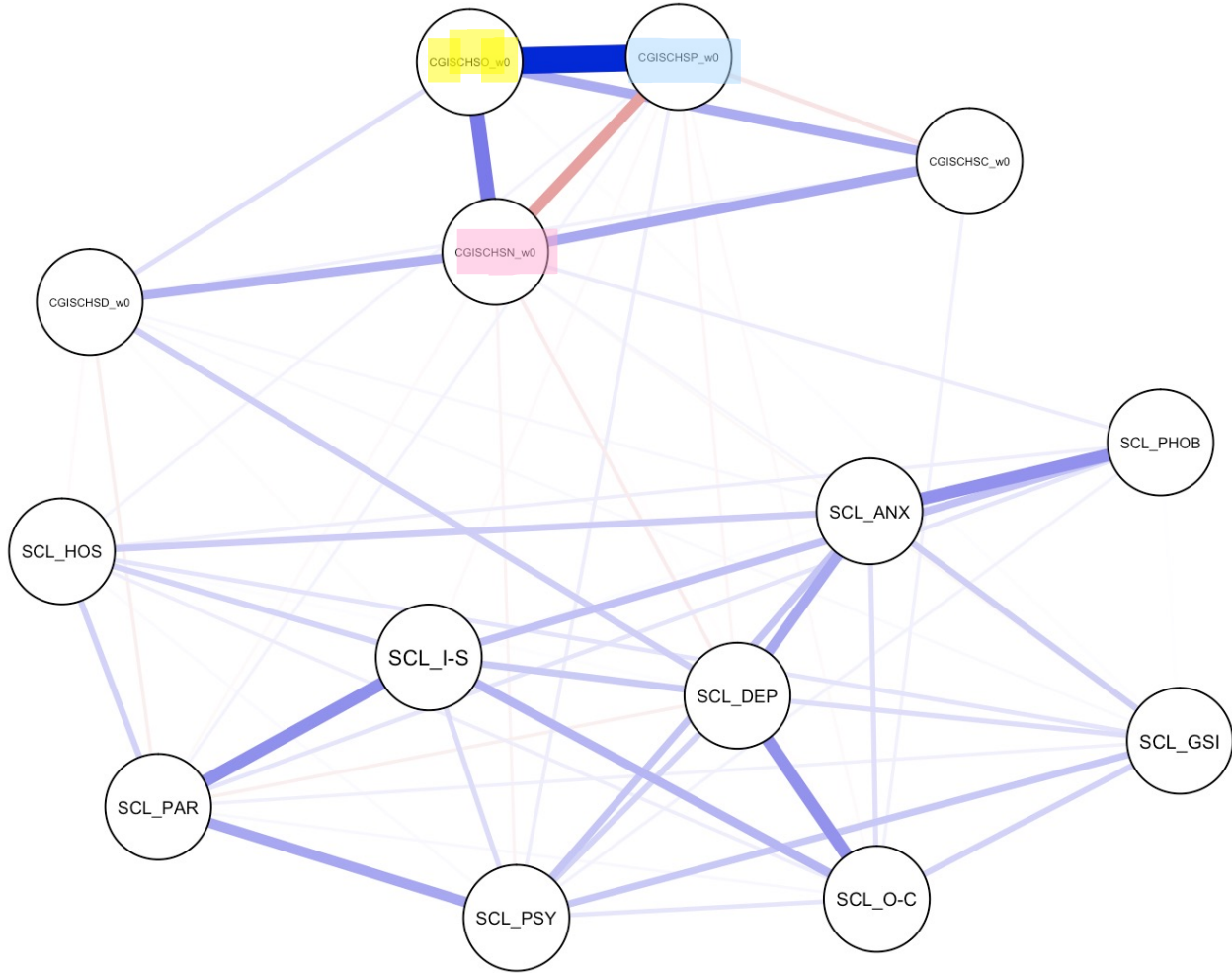
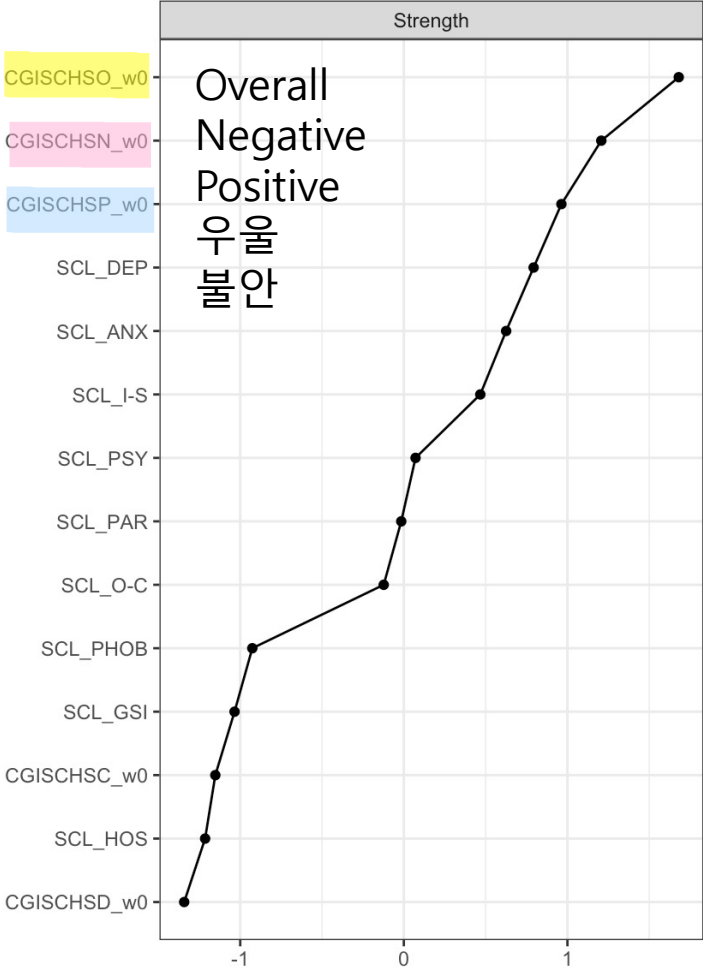


### 3. Bootnet

중심성 지수  
안정성 평가

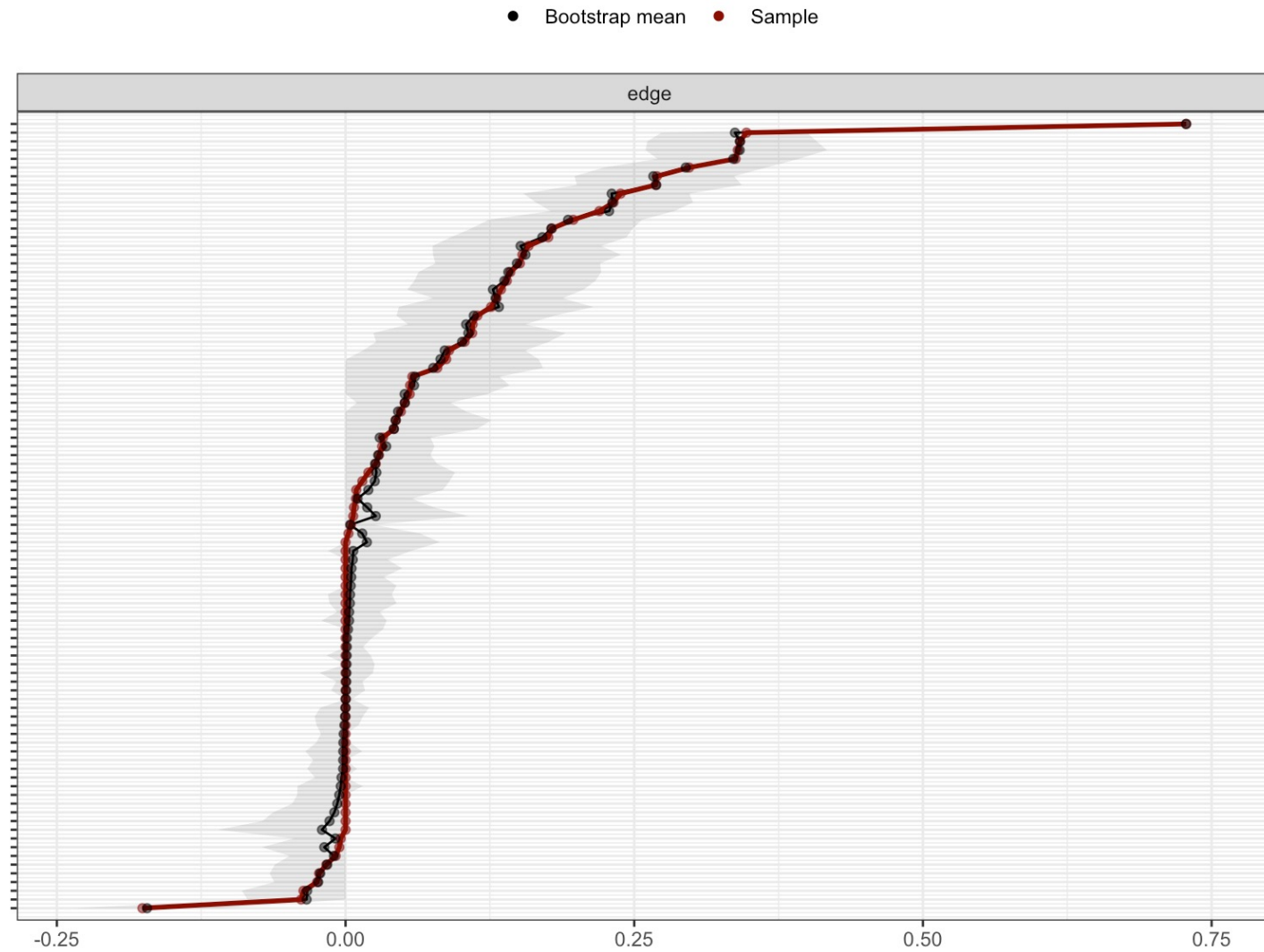


# 3. Bootnet



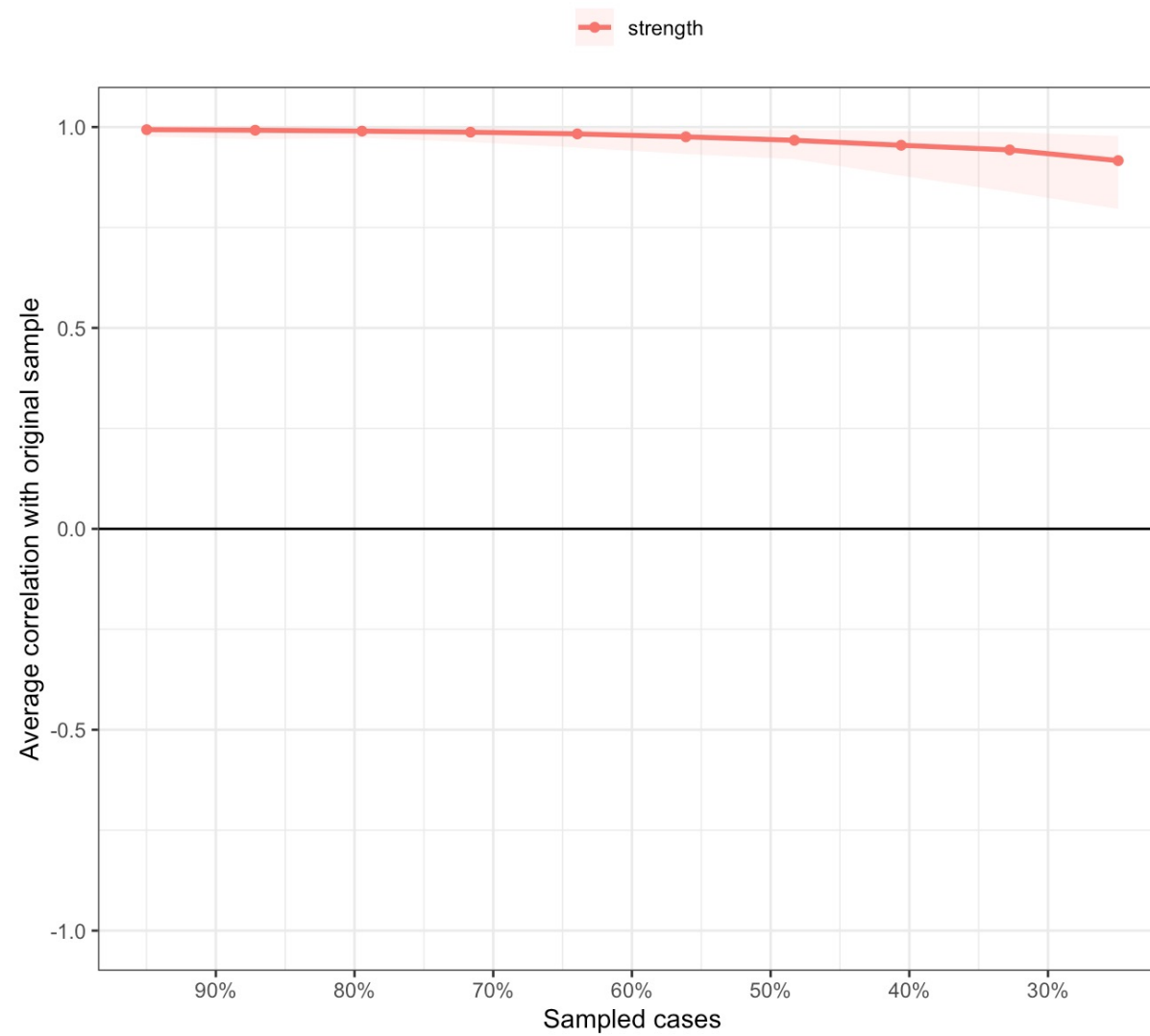
### 3. Bootnet

Edge Weight  
신뢰구간 추정



### 3. Bootnet

중심성 지수  
안정성 평가



[Todolist]