Chapter03

통계분석 I

I 기술통계

FASTCAMPUS ONLINE

금융공학/퀀트 I

강사. 장순용

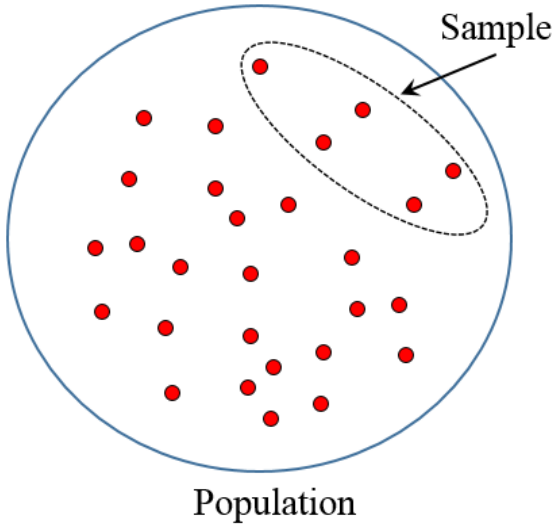
I 키포인트

- 모집단과 표본.
- 기술통계.
- 분위수.
- 상자그림.

I 모집단과 표본

- 모집단 (population): 통계 분석 대상 전체.
- 표본 (sample): 모집단에서 추출한 일부.

예). 대한민국 20세이상 남성의 체질량 지수 BMI 평균을 구하기 위해서 500명을 표본으로 뽑는다.



I 기술 통계와 통계적 추론

- 기술 통계: 통계적 특성을 있는 그대로 묘사한다.

  ⇒ 표본을 요약한다. 통계량 계산.

- 통계적 추론: 표본의 특성을 가지고 모집단의 특성 즉 모수를 알아낸다.

  ⇒ 일반화를 의미한다.

I 기술 통계와 통계적 추론

- 기술 통계: 통계적 특성을 있는 그대로 묘사한다.

  ⇒ 표본을 요약한다. 통계량 계산.

- 통계적 추론: 표본의 특성을 가지고 모집단의 특성 즉 모수를 알아낸다.

  ⇒ 일반화를 의미한다.

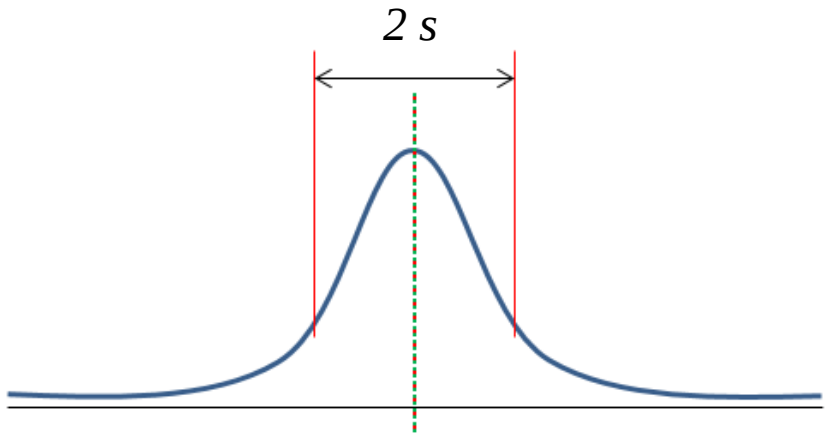
I 표본의 특성: 통계량

- 평균 (mean value): $\bar{x}$
- 중앙값 (median): m
- 분산 (variance): s^2

  ⇒ 표준편차 (standard deviation): $s = \sqrt{s^2}$
- 공분산 (covariance): s_{XY}

  ⇒ 상관계수 (correlation): r
- 왜도 (skewness): ς
- 첨도 (Kurtosis): κ

I 확률분포의 형상

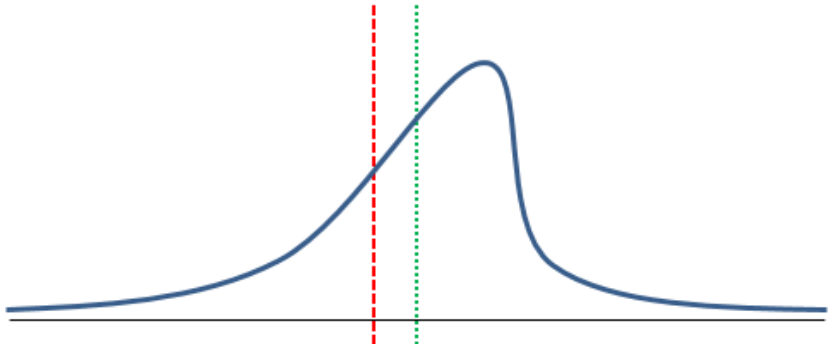


$\bar{x} = m$

$\zeta = 0$

좌우 대칭

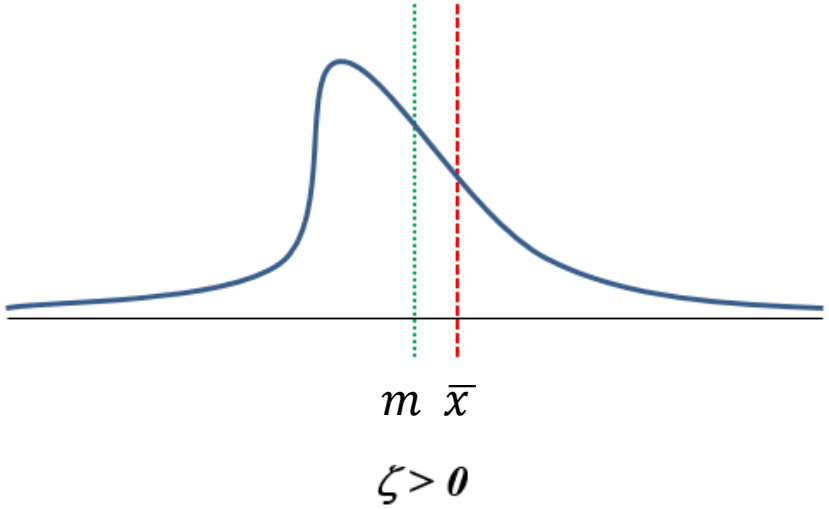
I 확률분포의 형상



$\zeta < 0$

왼 쪽으로 길게 뻗음

I 확률분포의 형상



오른 쪽으로 길게 뻗음

I 표본의 분산(Var), 공분산(Cov), 상관계수(Cor)

- $s^2 = Var(X) = Cov(X, X)$
- $s_{XY} = Cov(X, Y)$
- $r = Cor(X, Y)$

I 표본의 분산(Var), 공분산(Cov), 상관계수(Cor)

- $s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$
- $s_{XY} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}$
- $r = \frac{s_{XY}}{s_X s_Y}$이며 -1과 1 사이의 수치이다.

I 분위수

- 분위수 (quantile) : α 분위수 x_α는 누적확률이 α와 같은 지점을 일컫는다. (α 는 0과 1사이의 수치).

$$CDF(x_\alpha) = \alpha$$

$$x_\alpha = CDF^{-1}(\alpha)$$

I 분위수

- 분위수 (quantile) : α 분위수 x_α는 누적확률이 α와 같은 지점을 일컫는다. (α 는 0과 1사이의 수치).

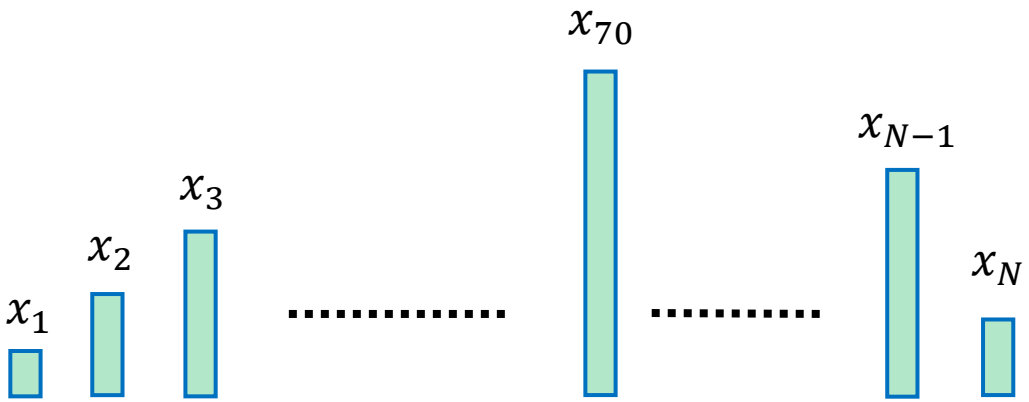
$$CDF(x_\alpha) = \alpha$$

$$x_\alpha = CDF^{-1}(\alpha)$$

그런데, 조금 난해하죠?

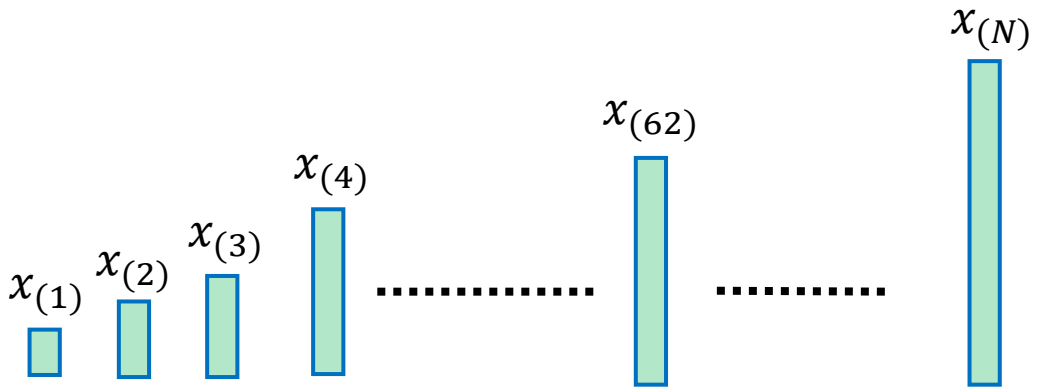
I 분위수

- $x_1, x_2, x_3, \cdots, x_N$ 와 같으로 이루어진 표본이 있다. 이들의 값은 각 양각색이다.



I 분위수

- 데이터를 **소→대** 순서대로 정렬한다.
- 정렬된 데이터를 $x_{(1)}, x_{(2)}, x_{(3)}, \cdots, x_{(N)}$와 같이 표기한다.



- 그러면, α 분위수는 α × 100 % 위치의 값이다. (α 는 0과 1사이의 수치).

I 백분위수, 사분위수

- 백분위수 (percentile) : α 분위수와 같은데 α를 백분율 (0% ~ 100%)로 나타낸 경우.
- 사분위수 (quartile) : α 를 4개의 구간으로 나눈 분위수.

  → 제1사분위수 (Q1) : α = 25%에 해당하는 분위수.

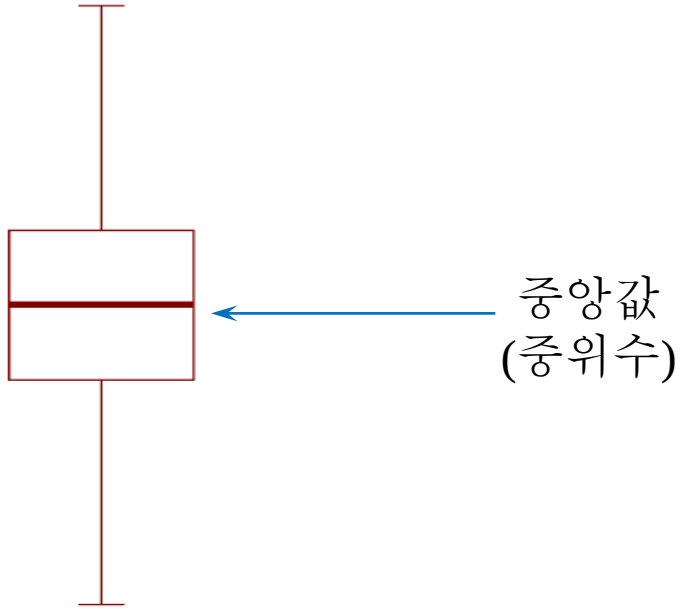
  → 제2사분위수 (Q2) : α = 50%에 해당하는 분위수.

  → 제3사분위수 (Q3) : α = 75%에 해당하는 분위수.

I 중위수, 최저값, 최고값

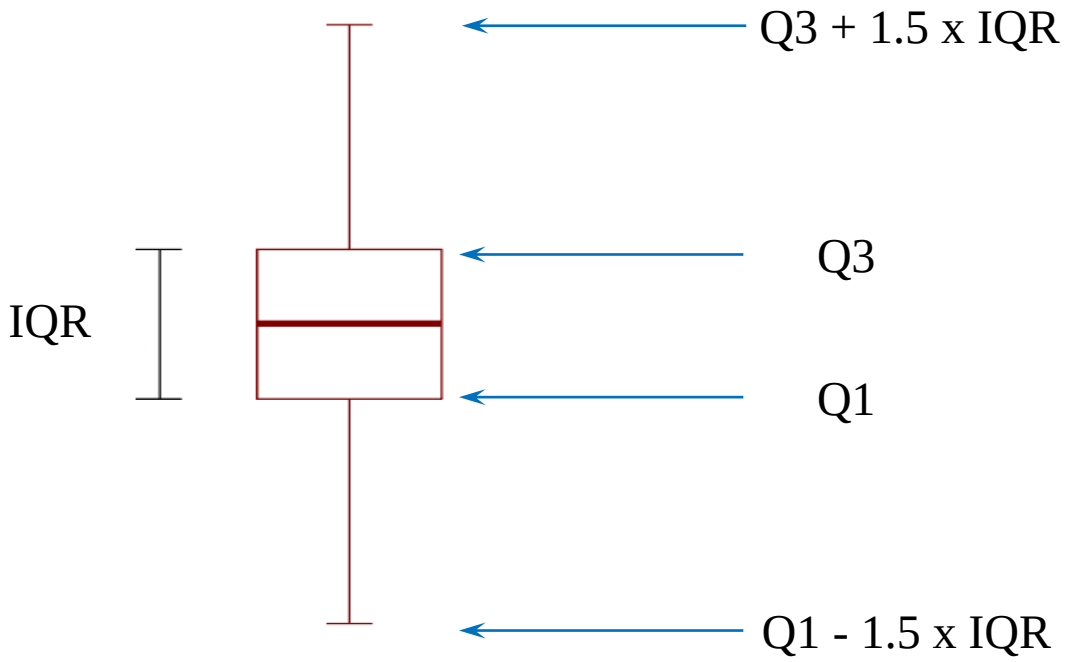
- 중위수(median) = 50% 백분위수.
- 최고값(maximum) = 100% 백분위수.
- 최저값(minimum) = 0% 백분위수.

I 상자그림 (Boxplot)



FAST CAMPUS
ONLINE
장순용 강사.

I 상자그림 (Boxplot)



FAST CAMPUS
ONLINE
장순용 강사.

I 끝.

감사합니다.

FAST CAMPUS
ONLINE
장순용 강사.