

- 검정 (testing)

- 검정 (testing)은 데이터 뒤에 숨어있는 현상현황의 분포와 속성에 대한 가설의 진위를 검증적으로 증명하는 작업이다.

예) 어떤 트레이더의 일일 수익률은 다음과 같다.

-2.5%, -5%, 4.3%, -3.7%

이 트레이더는 계속해서 돈을 잃을 가능성이 아닌가!

- 가설 (Hypothesis)

- 현상현황에 대한 어떤 주장은 가설이라고 하며, 이 가설을 증명하는 작업은 검정 (testing) 이라고 한다. 특히 현상현황의 속성 값이 특정한 값을 가지는 가설을 검증하는 것은 속성 검정 (parameter testing) 이라고 한다.

- 속성 검정에서 가장 널리 사용되는 가설 중 하나는 속성 값이 특정한 변수 값(주로 0) 이라는 가설이다.

$$H: \theta = 0.$$

- 이 가설은 회귀 분석 (regression) 에러 항에 사용되며, 회귀계수 값이 0이면 독립변수 (independent variable) 가 종속변수 (dependent variable) 의 영향을 미치지 않는다는 의미가 된다.

- 대립가설과 귀무가설의 예)

- (1) 만약 θ 가 0이 아니라는 것을 증명하고 싶다면 귀무가설과 대립가설은 다음과 같다.

$$H_0: \theta = 0, H_a: \theta \neq 0$$

- (2) 만약 θ 가 0보다 크다는 것을 증명하고 싶다면 귀무가설과 대립가설은 다음과 같다.

$$H_0: \theta = 0, H_a: \theta > 0$$

- (3) 만약 θ 가 0보다 작다는 것을 증명하고 싶다면 귀무가설과 대립가설은 다음과 같다.

$$H_0: \theta = 0, H_a: \theta < 0$$

- 양방향으로 같은 형태의 대립가설을 가지는 경우는 양방향 검정 (two-tailed testing),
단방향으로 같은 형태의 대립가설을 가지는 경우는 단방향 검정 (one-tailed testing) 이라고 하며 이에 따라 취할 유의수준을 구하는 방법이 달라지게 된다.

- 검정통계량

- 검정 (testing)을 하려면 즉, 귀무가설이 맞거나 틀린지를 증명하려면 어떤 증거가 있어야 한다. 이 증거에 해당하는 숫자는 검정 통계량 (test statistics) 이라고 한다.

- 검정 통계량의 예는 들면 다음과 같다.

- "어떤 방식이 낫다." 라는 가설을 증명하려면 회귀 분석의 결과로 차이를 나타내는 계수의 유의성을 살펴야 한다고 가정하자. 이때 해당 수치가 검정 통계량이 된다.

- 데이터 분석에 검정 통계량 t 는 표준 데이터로 묶은 계수값을 말한다.

$$t = f(x_1, x_2, x_3)$$

- 검정 통계량은 특정 변수의 표준 데이터에 대한 계수 값이므로 정확하게 예측할 수 없다.
 즉, 불확실하다.

- 유의수준

- 검정 통계량이 때로는 검정 통계량 분포를 알고 있다면 실제 데이터에서 계산한 검정 통계량 값이 분포에서 어느 위치에 위치하고 있는지 알 수 있다. 이 위치를 나타내는 것이 유의수준 (p-value) 이다.

- 특히 우리는 "귀무가설이 맞음에도 불구하고 대립가설을 옳다고 판단하여 검정 통계량과 같은 (혹은 대립가설과 다른) 표준 데이터가 부적합하다"를 의미한다.

- 즉 유의수준의 값이 매우 작다는 것은 귀무가설이 맞다는 가정하에 현재의 검정 통계량 값이 나온 가능성이 매우 작다는 의미이다. 따라서 유의수준의 값이 매우 작으면 귀무가설이 옳다고 대립가설을 채택할 수 있다.

- 유의 수준과 기각역

- 제정한 유의수준에 대해 귀무가설을 기각할지 채택할지를 정할 수 있는 기준점으로 유의 수준 (level of significance) 이라고 한다. 일반적으로 사용되는 유의 수준은 1%, 5%, 10% 등이다.

- 유의수준에 대해 제정한 통계량은 기각역 (critical value) 이라고 한다.