

통계학 개론

01. 서론

통계학 (statistics) | 통계적(과학적)인 방법을 가지고 사실에 의한 추론으로

의사결정을 하기 위한 일련의 방법

이론 통계학

수리 통계학
(응용)

기술 통계학 (descriptive) | 주어진 자료에서 μ, σ 같은 요약통계량

추론 통계학 (inferential) | 모론에서 표본을 가지고 (귀납적)

가설검정, 회귀 - 예측, 상관 분석

inductive 추론

표본 (3% - 5%) 추출 방법

변수 [양적 (숫자) | 질적 (범주)]

이산 연속

횡단 자료 (cross-section) | 같은 시간 & 같은 위치에 따른 원인에 대하여 유원 자료

시계열 자료 (time-series) | 다른 " & 다른 " " "

→ PPDAC cycle

Problem-plan-data-analysis-conclusion

02.

도수분포표, 막대·원

03.

그룹화 자료 (3% 범주), 비그룹화 (보통은 3% 범주)

제곱평균 제곱차 | 이분변에 이분변이 있을 경우 일률비율제거 후 제곱

제곱평균 제곱차 | $\frac{n_1 \cdot x_1 + n_2 \cdot x_2}{n_1 + n_2}$

기하평균 | $(x_1 \cdot x_2 \cdot \dots)^{1/n}$, 귀납적 공구나 값을 표본으로 같은 행상 연산

가중평균 | $\frac{\sum w_i x_i}{\sum w_i}$

산술평균

$$s^2 = \frac{1}{n} (x_1^2 + \dots + x_n^2) - \left(\frac{x_1 + \dots + x_n}{n} \right)^2$$

$$s^2 = \frac{1}{n-1} \quad \quad \quad = \frac{1}{n-1} \quad \quad \quad$$

CV (coefficient of variation) | 변동계수, 산술평균에 대한 변이계수
 $\frac{s}{\bar{x}}$ CV는 표본평균에 대한 두가지의 변이

레비슨트 정리 | μ 에 대해 $1 - \frac{1}{n}$ 은 제1 근사 $\mu \pm 1.5$ 에 놓인다

→ 이런 법칙도 적용가능. but. 경험법칙 only 중요함

04. 확률

대수 확률 | 어느 실험을 iter 하면 생길지가 확률에 관한

조건부 확률 | 조건 안에서 사건에 대한 확률

조건부 확률 | $P(A|B) = \frac{P(A \cap B)}{P(B)}$

베이지 정리 | $P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$

05.

이항분포 \rightarrow 베이지

nC_x (조합, combination) $\frac{n!}{x!(n-x)!}$

조합 \rightarrow 3가지 방법 다함 $\frac{n!}{x!(n-x)!}$

nPr (순열, permutation) $\frac{n!}{(n-x)!}$

H3 분포, r 분포
n 분포, x 분포

표준 H1에 의거하고 확률적 접근. $\frac{1}{2!}$ (2는 H1에 대한 전 (interval)에 대한 영향)

06. 정규분포 · 연속

정규분포 standardizing $z = \frac{x - \mu}{\sigma}$

07. 표본추론

$$\frac{n}{N} \leq 0.05 \Rightarrow \sigma_x = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

$$\hat{p} = p, \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

중심극한정리 = 표본 크기가 충분히 커면 정규분포 (Z)에 수렴

08. 추정 방법

표본 통계량의 함수로 모수를 추정함 \rightarrow 추정 (estimator)

여기에 사용된 표본 통계량 | 추정값

1) 점추정

추정값

점추정 | 추정값을 점추정 (point estimate)이라 함

구간추정 | 추정값을 구간으로 이것에 모수가 포함될 것이라고 함

σ를 모를 때 $\begin{cases} \text{모집단 분포, } n < 30 \\ n \geq 30 \end{cases} \Rightarrow \mu \text{ 추정시 } z\text{-분포}$
 $\begin{cases} \text{모집단 분포, } n < 30 \\ n \geq 30 \end{cases} \Rightarrow \mu \text{ 추정시 } t\text{-분포}$

2개의 한계 $z = z_{\alpha/2} = z_{\frac{\sigma}{\sqrt{n}}}$ / 신뢰도 $\mu \pm z$, 신뢰도 95% = 1 - α
 $\Rightarrow n = \frac{z^2 \sigma^2}{e^2}$

σ를 모를 때 $\begin{cases} \text{모집단 분포, } n < 30 \\ n \geq 30 \end{cases} \Rightarrow \mu \text{ 추정시 } t\text{-분포}$
 $\begin{cases} \text{모집단 분포, } n < 30 \\ n \geq 30 \end{cases} \Rightarrow \mu \text{ 추정시 } z\text{-분포}$

09. 가설검정

H_0 : 귀무가설, 모수에 대한 가설-가설이 전까지 관성

H_1 : 대립가설, " , 보통 주장하는 바

실제 (H1)
True False
예) True 0 1 제 1종 오류, H_0 를 영제로 H_0 가 맞을 때 (p-value)
False 1 0 제 2종 오류, H_1 를 영제로 H_1 가 맞을 때 (p-value)
 $1-p$ (검정력), p를 작게 만들수록 0에 가까워짐

i) 가설 ii) 분포 iii) p-value iv) 결론
ii) 분포통계량

