Hanyang University, Ansan, 425-791, Korea
calltimy@hanyang.ac.kr

Probability and Statistics / 확률과 통계

강의노트 17

Estimation 추정

121. 추정의 필요 - 실생활에서 부딪히는 문제는 데이터의 양이 너무 많아서 원하는 정보를 얻을 수 없다는 것

122. 표본크기와 표준오차(표본오차)
압정공장에서 대량생산되는 압정의 경우 : 성공확률 p, 실패확률 1-p 인 베르누이 시행 -> 이항분포

-1. 무작위로 n 개의 압정을 표본으로,그 중 무결함 압정 x 개
-2. 표본에서 성공확률을 $\hat{p}$ 으로 한다. 전체 공장 생산품을 대상으로 한 성공확률은 p 이다. ($\hat{p}=\frac{x}{n}$, 1000개 압정중 우량품이 832개라면 $\hat{p}$ = .832)
-3. [Q] $\hat{p}$ 는 좋은 근사치인가?
-4. [reQ] 1000개의 압정으로 된 표본을 많이 취해서 각 표본마다 $\hat{p}$ 의 값을 측정하면 $\hat{p}$ 의 값이 p를 중심으로 어떤식으로 분포할까?
-5. $E[\hat{p}]=p$, $\sigma[\hat{p}]=\frac{\sqrt{p(1-p)}}{\sqrt{n}}$
-6. n이 크면 $\hat{p}$ 는 정규분포를 따른다.
-7. $\hat{p}$ 는 정규분포에 가까우므로 측정값의 약 68%가 정확한 값 p 로부터 표준편차의 범위내에 위치한다는 결론을 내릴 수 있다.

123. n=1000, p=0.85 인 압정에서의 $\hat{p}$ 의 표준편차는...

$$\sigma(\hat{p})=\sqrt{\frac{(.85)(.15)}{1000}}=.0113$$

그래서 측정값은 약 68%가 아래 구간내에 있다고 예상할 수 있다.

$.8387 \leq \hat{p} \leq .8613$

X~B(n,p)
E[X] = np
V[X] = np(1-p)
$\hat{p}=\frac{X}{n}$

$$V[X]=V[np]=n^2V[p]$$

$$V[\hat{p}]=\frac{p(1-p)}{n} \quad \text{... (아래)}$$

$$\sigma[\hat{p}]=\frac{\sqrt{p(1-p)}}{\sqrt{n}}$$

$$V[\hat{p}]=V\left[\frac{X}{n}\right]=\frac{1}{n^2}V[X]=\frac{1}{n^2}np(1-p)=\frac{p(1-p)}{n}$$

124. $\hat{p}$ 의 표준편차는 표본오차의 척도이고, 이항분포 $\hat{p}$ 에서 표본오차는 $\sqrt{n}$ 에 반비례한다. (p=0.85) 표본크기를 4배 증가하시키면 분산은 두배($\sqrt{4}$) 감소한다.

n	1	4	16	25	100	10000
$\sqrt{n}$	1	2	4	5	10	100
$\sigma(\hat{p})$	0.3571	0.1785	0.0893	0.0714	0.0357	0.0036

125. 일반적인 통계 진행의 4단계

1) 미지의 매개변수로 모집단을 정의한다.
2) 이론적인 표본분포와 표준편차에 대한 추정치를 구한다.
3) 무작위로 표본을 추출, 추정치를 구한다.
4) 결과와 표본오차를 함께 보고한다.

126. 평균의 표본분포

금남 오이의 평균 길이를 알고 싶다. n개의 오이를 무작위로 뽑아서 길이를 잰다. -> x1, x2, x3, ... xn

$$E[X_i] = \mu$$
$$\sigma(X_i) = \sigma$$

표본평균 $\overline{X} = \dfrac{X_1 + X_2 + X_3 + ... + X_n}{n}$

$$E[\overline{X}] = \mu$$
$$\sigma(\overline{X}) = \frac{\sigma}{\sqrt{n}}$$

∴ 측정한 표본평균의 산포도는 $1/\sqrt{n}$ 에 비례한다.

127. 중심극한정리를 이용한 경우 두가지 문제

1) 표본의 크기가 커야 한다.
2) 표준편차 σ 를 알아야 한다.

128. 표본이 작고 표준편차가 알려져 있지 않을 때 : t-분포

표본 표준편차를 통해 σ 추정

$$s = \frac{1}{n-1}\sum_{i-1}^{n}(x_i - \overline{x})^2$$

s 는 표본의 표준편차이고,

$z = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}$ 에서

σ 대신 s를 넣어, 새로운 t를 정의한다.

$$t = \frac{\overline{X} - \mu}{s/\sqrt{m}}$$