신뢰구간

정의

- *정의*

1. $X_1, \ldots, X_n$을 pdf $f(x)$를 갖는 확률변수 x의 확률표본이라고 하자.
어떤 통계량 L = L($X_1, \ldots, X_n$)과 U = U($X_1, \ldots, X_n$)을 정의할 때

1) $(1-a) = p[\theta \in (L, U)]$ (단, 0<a<1)

2) 즉, 모수 θ가 L,U로 정의되는 구간 사이에 포함될 확률이 $(1-a)$%라고 할 때,
이 때 L,U 사이의 구간을 신뢰구간이라고 한다.

3) 이 때, $(1-a)$를 신뢰구간의 신뢰계수 라고 한다.

정의

- μ의 신뢰구간

1. 중심극한정리

1) $X_1, \dots, X_n$이 평균 μ, 분산 σ^2인 X의 확률표본이라고 할 때

$$\lim_{n \to \infty} \frac{\overline{x} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

2. 중심극한 정리를 이용하여

1) $(1 - a) = p[\theta \in (L, U)]$ 를 정의하면

(1) $(1 - a) = p[-t_{a/2,n-1} < T < t_{a/2,n-1}]$

(2) T = $\frac{\overline{x}-\mu}{\sigma/\sqrt{n}}$ 이므로,

$p[-t_{a/2,n-1} < \frac{\overline{x}-\mu}{\sigma/\sqrt{n}} < t_{a/2,n-1}]$

$= p[-t_{a/2,n-1} \cdot \sigma/\sqrt{n} < \overline{x} - \mu < t_{a/2,n-1} \cdot \sigma/\sqrt{n}]$

$= p[-t_{a/2,n-1} \cdot {}^{\sigma}/_{\sqrt{n}} - \overline{x} < -\mu < t_{a/2,n-1} \cdot \sigma/\sqrt{n} - \overline{x}]$

$= p[\overline{x} - t_{a/2,n-1} \cdot {}^{\sigma}/_{\sqrt{n}} < \mu < \overline{x} + t_{a/2,n-1} \cdot \sigma/\sqrt{n}]$

정의

- 평균차에 대한 신뢰구간

1. μ_x와 μ_y를 갖는 확률분포 X와 Y가 존재할 때, X와 Y에서 각각 선출한 확률표본을 $[X_1, \dots, X_n], [Y_1, \dots, Y_n]$ 이라고 하자.

1) 이 때, μ_x- μ_y = a라고 할 때, $\bar{x}$ - $\bar{y}$ = $\hat{a}$ 는 a에 대한 불편추정량이다.

2) 두 분포의 표본평균의 차 $\bar{x}$ - $\bar{y}$ = $\hat{a}$ 의 분산을 구하면

(1) var($\bar{x}$ - $\bar{y}$) $= \mathrm{Var}(\bar{x})\ (-1)^2\ \mathrm{Var}(\bar{y})$ + 0 = $\mathrm{Var}(\bar{x})$ + $\mathrm{Var}(\bar{y})$

(2) 이 때, $\bar{x}$ ~ N(μ_x, $\frac{\sigma_x^2}{n_x}$) 이고 $\bar{y}$ ~ N(μ_y, $\frac{\sigma_y^2}{n_y}$) 임이 알려져 있으므로,

- $\mathrm{Var}(\bar{x})$ + $\mathrm{Var}(\bar{y})$ = $\frac{\sigma_x^2}{n_x}$ + $\frac{\sigma_y^2}{n_y}$

정의

- 평균차에 대한 신뢰구간(z-zcore)

(3) 정규분포의 가법성에 따라, $\hat{a}$ 와 a로 구성되어 있는 새로운 통계량을 구성하면

- W = $\dfrac{(\overline{x} - \overline{y}) - (\mu_x - \mu_y)}{\sqrt{\dfrac{\sigma_x^2}{n_x} + \dfrac{\sigma_y^2}{n_y}}}$ ~ N(0,1) 이다.

(4) 이를 이용하면

- $(1-a) = p[(\overline{x} - \overline{y}) - z_{a/2,n-1}\sqrt{\dfrac{\sigma_x^2}{n_x} + \dfrac{\sigma_y^2}{n_y}} < (\mu_x - \mu_y) < (\overline{x} - \overline{y}) - z_{a/2,n-1}\sqrt{\dfrac{\sigma_x^2}{n_x} + \dfrac{\sigma_y^2}{n_y}}]$

정의

- 평균차에 대한 신뢰구간(t-score)

1. W = $\frac{(\overline{x} - \overline{y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}$ ~ N(0,1) 에서

1) X의 표준편차 S_x 와 Y의 표준편차 S_Y 는

(1) $\frac{(n-1)S_x^2}{6^2}$ ~ $x^2(n_x$-1) 이고 $\frac{(n-1)S_y^2}{6^2}$ ~ $x^2(n_y$-1) 이므로

(2) S_x 와 S_Y 의 가중평균 S_p 를 정의하면

- $S_p = \frac{(n_x\text{-}1)S_x^2+(n_y\text{-}1)S_y^2}{n_x\text{-}n_y-2}$,

(3) 이 때, V = $\frac{(n-2)S_p^2}{6^2}$ ~ x^2(n-2) 이다.

정의

- 평균차에 대한 신뢰구간(t-score)

2. 이 때, T분포의 변환함수를 정의하면

1) T = $\frac{W}{\sqrt{\frac{V}{(n-2)}}}$ 에서, W ~ N(0,1) 이고, V ~ χ^2(n-2) 이므로

2) T = $\frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y) \Big/ \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}{\sqrt{\frac{(n-2)S_p^2}{6^2} \Big/ (n-2)}}$ ~ t(n-2)를 따른다.

(1) 위 식을 정리하면

$$= \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

정의

- 평균차에 대한 신뢰구간(t-score)

2. T 스코어를 이용한 신뢰구간은

1) $(1-a)$

$$= p[(\overline{x} - \overline{y}) - t_{a/2,n-1}S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} < (\mu_x - \mu_y) < (\overline{x} - \overline{y}) - t_{a/2,n-1}S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}]$$

2) 이 처럼, $\frac{\overline{x}-\mu}{\sigma/\sqrt{n}}$ ~ t(n) 이라는 점과, $\lim_{n\to\infty} t(n)$ ~N(0,1) 을 이용하여
<u>모수의 신뢰구간을 측정 가능하게 하는 통계량들</u>을 피벗 확률변수라고 한다.

예제

- $X_1, \ldots, X_{10}$이 N(μ_x, σ^2)을 따르고, $Y_1, \ldots, Y_7$이 N(μ_Y, σ^2)를 따르는 확률표본이라고 할 때 $\overline{x}$ = 4.2, $\overline{y}$ = 3.4 이고, $\frac{S_x^2}{n_x-1}$ = 49, $\frac{S_y^2}{n_y-1}$ =32 라고 한다면, 이 두 확률변수의 평균차이의 90% 신뢰구간을 구하라

1. $(1-0.1) = p[(4.2 - 3.4) - t_{0.1/2,15} S_p \sqrt{\frac{1}{10} + \frac{1}{7}} < (\mu_x - \mu_y) < (4.2 - 3.4) + t_{0.1/2,15} S_p \sqrt{\frac{1}{10} + \frac{1}{7}}]$

1) 이 때 가중분산 S_p 는

2) $S_p = \sqrt{\frac{(10-1)s_x^2+(7-1)s_y^2}{10+7-2}} = \sqrt{\frac{9\cdot 46+6}{10+7-2}}$=0.33

2. $p[0.8 - t_{0.1/2,15} \cdot 0.33 \cdot \sqrt{\frac{1}{10} + \frac{1}{7}} < (\mu_x - \mu_y) < 0.8 + t_{0.1/2,15} \cdot 0.33 \cdot \sqrt{\frac{1}{10} + \frac{1}{7}}]$

1) 이 때, $t_{0.1/2,15}$=2.015 이므로,

(1) $p[0.8 - 2.015 \cdot 0.33 \cdot \sqrt{\frac{1}{10} + \frac{1}{7}} < (\mu_x - \mu_y) < 0.8 + 2.015 \cdot 0.33 \cdot \sqrt{\frac{1}{10} + \frac{1}{7}}]$

$= p[3.51 < (\mu_x - \mu_y) < 5.11]$