우도비 검정

정의

- 개요

1. 네이만-피어슨 정리에 의거하여, 우도비 검정 $\frac{\mathrm{L}(w)}{\mathrm{L}(\Omega)} = \mathrm{g(y)} \leq \mathrm{k}$은

1) 최량검정이고
2) 모든 점에서 $P_{\theta=\mathrm{w}}[X \in C] = \alpha \leq P_{\theta=\Omega}[X \in C]$인 불편검정이다.

2. 즉, 이 성질을 이용하면 항상 효과적인 측정 기준을 구할 수 있다.

정의

- T-검정

1. 독립인 확률변수 X~N(θ_1,σ^2) , Y ~ N(θ_2,σ^2) 을 고려하자.
   즉 분산은 같고 평균은 다르다.

2. H_0 : θ_1 = θ_2 VS H_1 :$\theta_1 > \theta_2$ or $\theta_1 < \theta_2$ 를 검정한다.

1) 모수의 공간을 각각 정의하면

(1) Ω = {(θ_1, θ_2, σ) : -∞<θ_1<∞, $-\infty<\theta_2<\infty$, $0<\sigma<\infty$}

(2) w = {(θ_1, θ_2, σ) : $-\infty<\theta_1 = \theta_2<\infty$, $0<\sigma<\infty$}

2) 우도비 검정을 정의하면

(1) $$\frac{\mathrm{L}(w)}{\mathrm{L}(\Omega)} = \frac{\prod_{i=1}^{n} f(x_i;\theta_1,\sigma) \prod_{i=1}^{m} f(y_i;\theta_1,\sigma)}{\prod_{i=1}^{n} f(x_i;\theta_1,\sigma) \prod_{i=1}^{m} f(y_i;\theta_2,\sigma)} = \frac{\prod_{i=1}^{n} f(x_i;\theta_1,\sigma) \prod_{i=1}^{m} f(y_i;\theta_1,\sigma)}{\prod_{i=1}^{n} f(x_i;\theta_1,\sigma) \prod_{i=1}^{m} f(y_i;\theta_2,\sigma)}$$

$$= \frac{\left[\frac{1}{\sqrt{2\pi}\sigma}\right]^{\frac{n+m}{2}} \exp\left(-\frac{\sum(x_i-\theta_1)^2+\sum(y_i-\theta_1)^2}{2\sigma^2}\right)}{\left[\frac{1}{\sqrt{2\pi}\sigma}\right]^{\frac{n+m}{2}} \exp\left(-\frac{\sum(x_i-\theta_1)^2+\sum(y_i-\theta_2)^2}{2\sigma^2}\right)}$$

정의

- T-검정

3) 각각의 우도를 최대화하기 위해 MLE 추정량을 구입해서 대입하면

(1) $\frac{\partial log\mathrm{L}(w)}{\partial\vec{\theta}} = \left[\frac{\partial log\mathrm{L}(w)}{\partial\theta_1}, \frac{\partial log\mathrm{L}(w)}{\partial\sigma}\right] = [0,0]$

- $\frac{\partial log\mathrm{L}(w)}{\partial\theta_1} = \sum(x_i - \theta_1) + \sum(y_i - \theta_2) = 0$
- $\frac{\partial log\mathrm{L}(w)}{\partial\sigma^2} = \frac{1}{\sigma^2}\left[\sum(x_i - \theta_1)^2 + \sum(y_i - \theta_1)^2\right] - (n + m) = 0$
- 따라서 $\theta_1 = \frac{\sum x_i + \sum y_i}{n+m}$, $\sigma^2 = \frac{\left[\sum(x_i - \theta_1)^2 + \sum(y_i - \theta_2)^2\right]}{n+m}$
- 따라서 $\mathrm{L}(\widehat{w}) = \left[\frac{e^{-1}}{\sqrt{2\pi}\sigma}\right]^{\frac{n+m}{2}}$

정의

- T-검정

3) 각각의 우도를 최대화하기 위해 MLE 추정량을 구입해서 대입하면

(2) $\frac{\partial log\mathrm{L}(w)}{\partial \vec{\theta}} = \left[\frac{\partial log\mathrm{L}(w)}{\partial \theta_1}, \frac{\partial log\mathrm{L}(w)}{\partial \theta_2}, \frac{\partial log\mathrm{L}(w)}{\partial \sigma^2}\right] = [0,0,0]$

- $\frac{\partial log\mathrm{L}(w)}{\partial \theta_1} = \sum(x_i - \theta_1) = 0$
- $\frac{\partial log\mathrm{L}(w)}{\partial \theta_2} = \sum(y_i - \theta_2) = 0$
- $\frac{\partial log\mathrm{L}(w)}{\partial \sigma^2} = \frac{1}{\sigma^2}\left[\sum(x_i - \theta_1)^2 + \sum(y_i - \theta_2)^2\right] - (n+m) = 0$
- 따라서 $\theta_1 = \frac{\sum x_i}{n}$, $\theta_2 = \frac{\sum y_i}{m}$,$\sigma'^2 = \frac{\left[\sum(x_i - \theta_1)^2 + \sum(y_i - \theta_2)^2\right]}{n+m}$
- 따라서 $\mathrm{L}(\hat{\Omega}) = \left[\frac{e^{-1}}{\sqrt{2\pi}\sigma'^2}\right]^{\frac{n+m}{2}}$

정의

- T-검정

3) MLE 추정량을 투입한 최대우도함수의 우도비를 정의하면

(1) $\frac{\mathrm{L}(\widehat{w})}{\mathrm{L}(\widehat{\Omega})} = \frac{\left[\frac{e^{-1}}{\sqrt{2\pi}\sigma^2}\right]^{\frac{n+m}{2}}}{\left[\frac{e^{-1}}{\sqrt{2\pi}\sigma\prime^2}\right]^{\frac{n+m}{2}}} = \left[\frac{\sigma^2}{\sigma\prime^2}\right]^{\frac{n+m}{2}}$

4) $\frac{\mathrm{L}(\widehat{w})}{\mathrm{L}(\widehat{\Omega})} = \mathrm{g(y)} = \Lambda$ 에서 y는 결합 정규분포에 대한 충분통계량이고,
이 함수를 이용한 최량기각역을 정의하면

(1) $\mathrm{p[g(y)} = \Lambda \leq$ k] = α 에서 $\mathrm{p[y} = [\Lambda]^{\frac{2}{n+m}} \leq g^{-1}$(k)] = α

정의

- T-검정

5) 이 때, $[\Lambda]^{\frac{2}{n+m}} = \dfrac{\left[\sum(x_i-\theta_1)^2+\sum(y_i-\theta_2)^2\right]}{\sum\{x_i-[\frac{n\bar{x}+m\bar{y}}{(n+m)}]\}^2+\sum\{y_i-[\frac{n\bar{x}+m\bar{y}}{(n+m)}]\}^2}$ 에서

(1) $\sum\{x_i-[\frac{n\bar{x}+m\bar{y}}{(n+m)}]\}^2 = \sum\{(x_i-\bar{x}) - [\bar{x} - \frac{n\bar{x}+m\bar{y}}{(n+m)}]\}^2 = \sum(x_i-\bar{x})^2 - n(\bar{x} - \frac{n\bar{x}+m\bar{y}}{(n+m)})^2$

(2) $\sum\{y_i-[\frac{n\bar{x}+m\bar{y}}{(n+m)}]\}^2 = \sum\{(y_i-\bar{y}) - [\bar{y} - \frac{n\bar{x}+m\bar{y}}{(n+m)}]\}^2 = \sum(y_i-\bar{y})^2 - n(\bar{y} - \frac{n\bar{x}+m\bar{y}}{(n+m)})^2$

6) 또한

(1) $\sum(x_i-\bar{x})^2 - n(\bar{x} - \frac{n\bar{x}+m\bar{y}}{(n+m)})^2 = \frac{m^2n}{(n+m)^2}(\bar{x}-\bar{y})^2 + \sum(x_i-\bar{x})^2$

(2) $\sum(y_i-\bar{y})^2 - n(\bar{y} - \frac{n\bar{x}+m\bar{y}}{(n+m)})^2 = \frac{m^2n}{(n+m)^2}(\bar{x}-\bar{y})^2 + \sum(y_i-\bar{y})^2$

(3) 이를 $[\Lambda]^{\frac{2}{n+m}}$에 대입하면

- $\dfrac{\left[\sum(x_i-\theta_1)^2+\sum(y_i-\theta_2)^2\right]}{\frac{m^2n}{(n+m)^2}(\bar{x}-\bar{y})^2+\sum(y_i-\bar{y})^2+\sum(x_i-\bar{x})^2}$

정의

- T-검정

7) 정리하면

(1) $\dfrac{\frac{m^2 n}{(n+m)^2}(\overline{x}-\overline{y})^2}{\frac{\left[\sum(x_i-\overline{x})^2+\sum(y_i-\overline{y})^2\right]}{n+m-2}}$ 는 $\dfrac{\sqrt{n}\bar{x}}{\sqrt{\sigma^2}}$ 꼴을 따르고, 이는 T(n+m-2)인 T 분포이다.

정의

- Z검정과 T검정

1. T검정은 Z검정으로 분포수렴한다.

1) H_0 :θ=θ_1 VS H_1 : $\theta \neq \theta_1$ 을 검증할 때

2) $Tn = \frac{\sqrt{n}(\bar{x}-\theta_1)}{s_n}$ 에서 기각역 c = ($Tn \geq$ t$_{\frac{\alpha}{2},n-1}$) 을 검증한다.

(1) 이 때, 표준편차 S는 모편차 σ에 확률 수렴하는 성질을 이용하면

(2) $\frac{\sigma}{S_n}\frac{\sqrt{n}(\bar{x}-\theta_1)}{s_n} \xrightarrow{D} \frac{\sigma}{\sigma}\frac{\sqrt{n}(\bar{x}-\theta_1)}{s_n}$ = $\frac{\sqrt{n}(\bar{x}-\theta_1)}{s_n}$

3) 일반적으로, t검정은 z검정에 비해 더 보수적이므로, 정규성을 가정하기 어려운 경우 t검정을 쓰면 더 강건한 검정을 할 수 있다.

정의

- F검정

1. 어떤 확률변수 X,Y가 각각 N(θ_1, θ_3) 와 N(θ_2, θ_4)인 분포에서 각각 $X_1, \dots, X_n$과 $Y_1, \dots, Y_n$을 추출했다고 하자.
2. 이 때, H_0 : θ_3 = θ_4 를 검정하면

1) 이 때, 실험공간과 모수공간은 각각

(1) Ω = {($\theta_1, \theta_2, \theta_3, \theta_4$) : $-\infty < \theta_1$, $\theta_2 < \infty$, $0 < \theta_3, \theta_4 < \infty$}

(2) w = {($\theta_1, \theta_2, \theta_3, \theta_4$) : $-\infty < \theta_1 = \theta_2 < \infty$, $0 < \theta_3 = \theta_4 < \infty$}

2) $\frac{\mathrm{L}(\hat{w})}{\mathrm{L}(\hat{\Omega})} = \frac{\sum(x_i - \overline{x})^2/(n-1)}{\sum(y_i - \overline{y})^2/(m-1)}$ = F 이다.

(1) 이 때, 관심모수인 $\theta_3 = \theta_4$, 즉 H_0라면

(2) $P_{H_0}\left(|\mathrm{F}| \geq \mathrm{C}_{n-1,m-1}\right) = \alpha$ 에서 기준 α와 자유도 n-1, m-1의 F분포로 가설 채택-기각을 결정한다.

정의

- F검정과 T검정의 혼합

1. 어떤 확률변수 T,F가 각각 N(θ_1, θ_3) 와 N(θ_2, θ_4)를 따른다고 하자.
2. 이 때, 각각을 따르는 우도비 검정 통계량을 정의한다.

1) $P_{\theta \in w}\left(|T| \geq t_{n-1,m-1}\right) = \alpha_1$

2) $P_{\theta \in w}\left(|F| \geq c_{\frac{a}{2},n+1,m+1}\right) = \alpha_2$ 를 정의할 때

3. 이 때, 다음이 성립한다.

1) $\overline{x}, \overline{y}, \sum(x_i - \overline{x})^2, \sum(y_i - \overline{y})^2$ 는 각각 $\theta_1, \theta_2, \theta_3, \theta_4$에 대한 결합완비충분통계량이다.

2) F = $\dfrac{\sum(x_i-\overline{x})^2/(n-1)}{\sum(y_i-\overline{y})^2/(m-1)}$ 은 위치규모불변통계량으로,

결합완비충분통계량 $\overline{x}, \overline{y}, \sum(x_i - \overline{x})^2, \sum(y_i - \overline{y})^2$와는 독립이다.

3) 한편, $\dfrac{\frac{m^2 n}{(n+m)^2}(\overline{x}-\overline{y})^2}{\frac{\left[\sum(x_i-\overline{x})^2 + \sum(y_i-\overline{y})^2\right]}{n+m-2}}$ 는 완비충분통계량으로 이루어진 함수로, 보조통계량인 F와는 독립이다.

따라서 이 둘을 결합하여 가설검정을 수행할 수 있다.