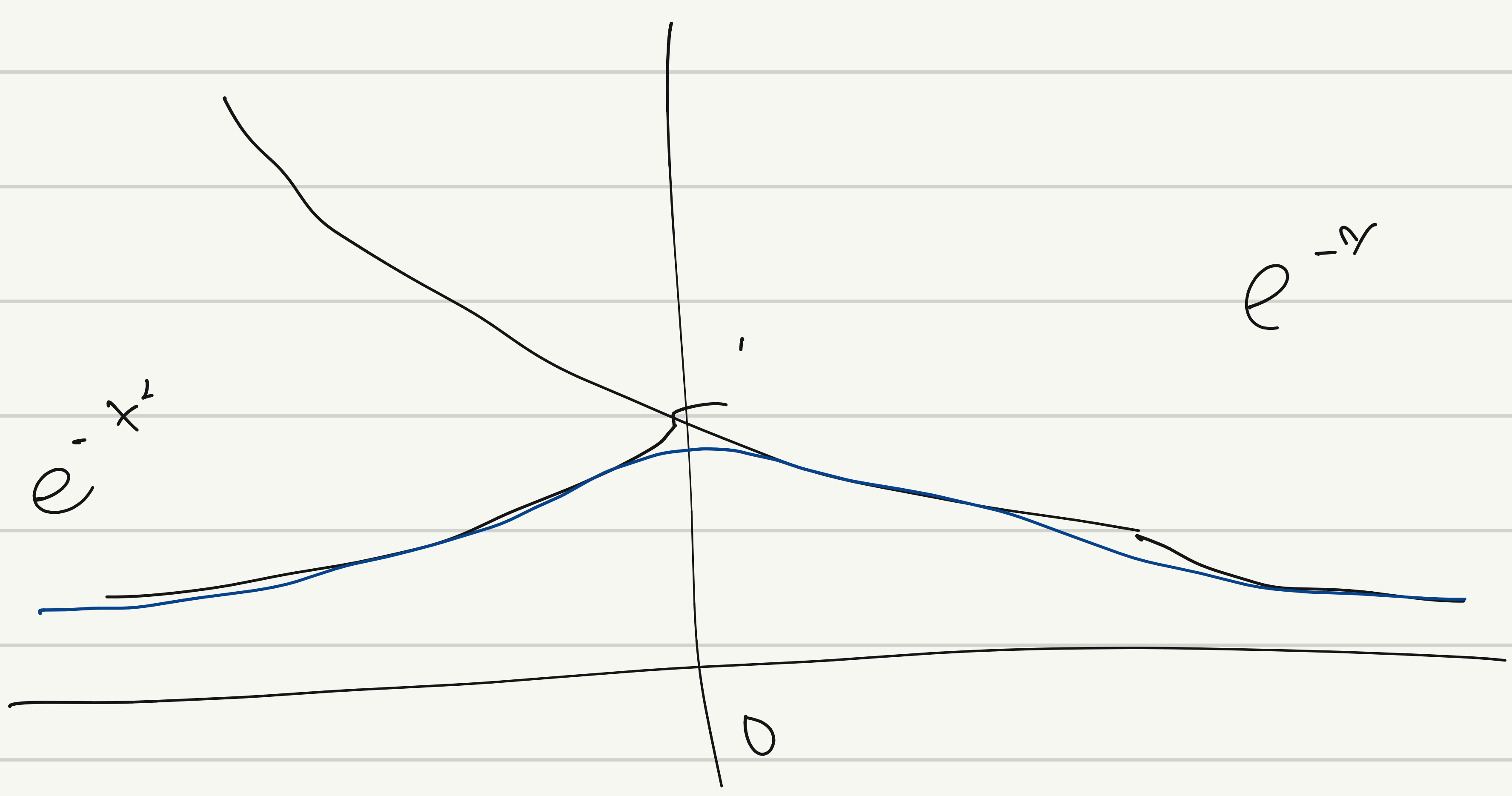
◎ p-value가 α.value 보다 낮으므로 기각.

↑ 유의확률 ↑ 기각역

◎ Z 검정

- 양측검정 : "×2" 확률적인 의미에서 2배 ∴ $p = 0.5 \rightarrow 0.025$

◎ 정규분포의 특징



- 그 자료의 μ, σ^2 을 알면 그 사건 전체를 설명 가능.
- independently, identically, distributed

- μ를 중심으로 σ만큼 떨어짐.
- AUC가 확률 → 적분해서 구함.
- 연속형이니까

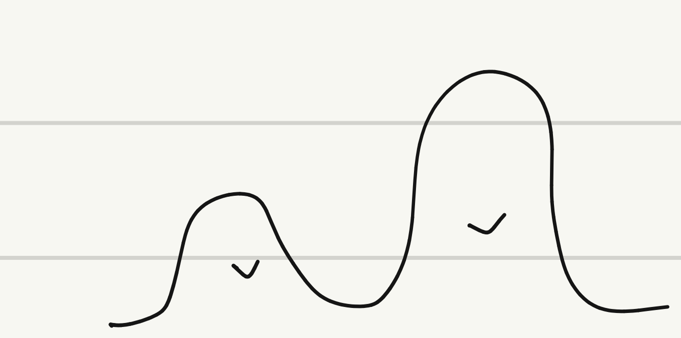
  $x \sim N(0,1)$

  $P(x=2) = 0$ ← 항상 구간으로 나타내 확률 구해.

- parametric distribution

$x \sim \underbrace{N(\mu, \sigma^2)}$ ← 모수 분포 → $x \sim Bin\underbrace{(n, p)}$

※ 비모수 분포 ← Gaussian Mixture Model



$\rightarrow \lambda_1 f_1(x) + \lambda_2 f_2(x)$

③ 정규분포의 표준화.

$$P(X > x_0) = \int_{x_0}^{\infty} \cdots$$

$$X \sim N(\mu, \sigma^2), \quad Z = \frac{X-\mu}{\sigma}, \quad Z \sim N(0, 1)$$

↓

$$P(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

⑥

$$\underbrace{P\left(\left|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right| < z_{\alpha/2}\right)} = 0.95, \quad \alpha = 0.05$$

우리가 궁금한 것. 모집단 μ의 신뢰구간

$$-z_{\alpha/2} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}$$

모집단의 표준편차

$$\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

실제로 모집단의 σ를 알 방법이 거의 없음

그래서 t-test를 진행 ($\sigma \to S$)

⑦ Student's t-test

개발자가 가명을 사용

t 분포를 기반으로 t 검정을 하겠다.

- 정규분포는 하나의 분포를 사용. ← ~~일~~(1) 변량 (one-sample) — 단일 표본

↓

~~이~~(2) 변량 (two-sample) — 두 표본

} t-test

① 독립 일표본 T 검정 (one-sample t-test)

- 정규분포와 student's t-분포와의 관계
  - 모집단의 분산을 모르는 경우.
  - 자료수가 많지 않은 경우.

$$X \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

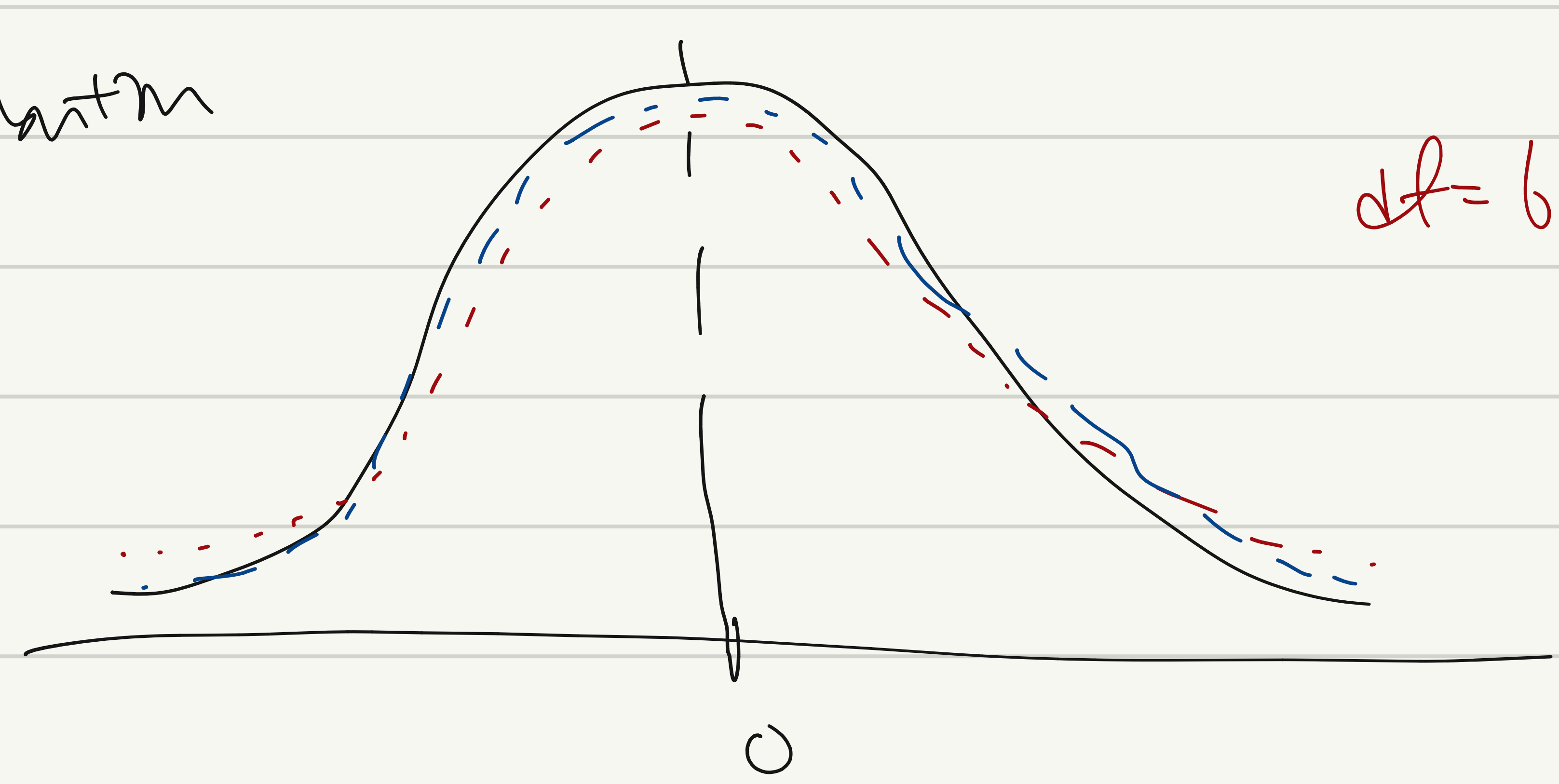
↓

자유도 df (degree of freedom)

$$\frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t(n-1)$$

← 데이터의 개수 (n)에 따라 t 분포가 변한다.

T-distribution



df=6 ← 7개로부터 s^2을 계산

⇒ 자료가 많아질수록 $N(0,1)$와 유사해 진다.
(자유도의 개수가 많아지면)

◎ T 분포에서 95%의 신뢰구간

$$P\left(\left|\frac{}{s/\sqrt{n}}\right| < t_{(n-1,\,\alpha/2)}\right) = 0.95$$

↓

$$\bar{X} - t_{(n-1,\,\alpha/2)}\frac{s}{\sqrt{n}} < \mu < \bar{X} + t_{(n-1,\,\alpha/2)}\frac{s}{\sqrt{n}}$$

※ 자료 $n \to \infty$ (30개 이상)
⇒ 모수 분포가 정규가 아니더라도 t-검정을 사용한다.
⇓
자료가 많으면 $s \to \sigma$로 수렴

CLT
중심극한정리

⑨ t 분포

t 분포 조건 ⟶ $n < 30$, $X \overset{iid}{\sim} N(\mu, \sigma^2)$

⑩ Shapiro-Wilk test

$SS = \Sigma (x_i - \bar{x})^2$, $s^2 = \dfrac{\Sigma (x-\bar{x})^2}{n-1}$

$a_1 (x_{10} - x_1)$
$+ a_2 (x_9 - x_2)$
$+$
$\vdots$

$b = \Sigma a_n (x_n - x)$

W (Shapiro-Wilk) $= \dfrac{b^2}{SS} = 1$

※ t-test 하기 전에 반드시 검정해야 함.

⑪ 독립 이표본 T 검정 (Two-Sample test)

· 얻어진 두 랜덤 표본으로부터 두 정규모집단이 같은 평균(μ)을 가지는지를 검증

※ n은 달라도 됨

· 대립 가설 종류
연구자 가설
다르다.

$H_1 : \mu_1 \neq \mu_2$, $H_0 : \mu_1 = \mu_2$ 양측검정, Two-tailed
$H_1 : \mu_1 > \mu_2$, $H_0 : \mu_1 \leq \mu_2$ 단측검정, Upper-tailed
$H_1 : \mu_1 < \mu_2$, $H_0 : \mu_1 \geq \mu_2$ Lower-tailed

· 정규 분포를 따르는 두 확률 변수의 차이 분포

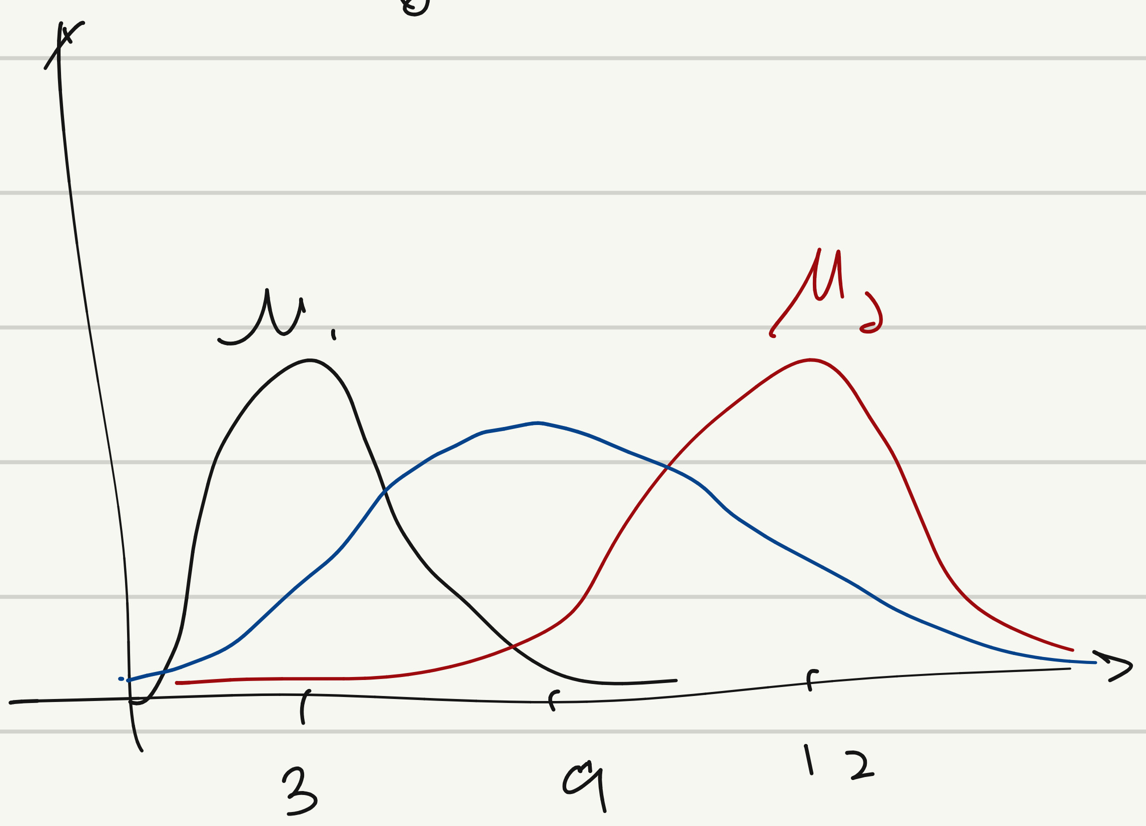
$X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$

$X_1 - X_2 \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$, X_1 X_2

· 정규 분포를 따르는 두 평균의 차이 분포

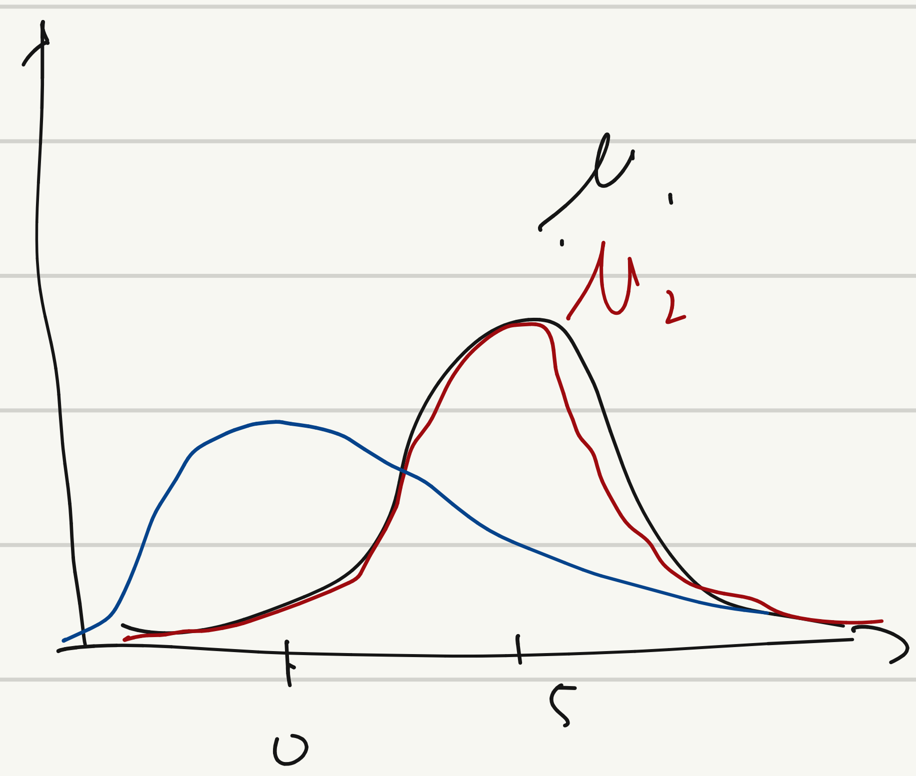
$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}\right)$, Y_1 Y_2

그룹간 평균이 서로 다를때



$\mu_1 \neq \mu_2$

그룹간 평균이 서로 같을 때



$\mu_1 = \mu_2$

⊙ 두 평균 차이 분포에 대한 정규화.

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

$$* \quad \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)}} = Z \sim N(0, 1)$$

if, $\sigma_1 = \sigma_2$, then $\dfrac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

⊙ 두 모평균의 차이 신뢰구간.

$$(\bar{X}_1 - \bar{X}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + z_{\alpha/2}\sqrt{\cdots}$$

⊙ T 분포에서 두 모평균의 차이 신뢰구간. $(n = n_1 + n_2)$

$$(\bar{X}_1 - \bar{X}_2) - t_{(n-2, \alpha/2)}\sqrt{\frac{s_1^2 + s_2^2}{n}} < \mu_1 - \mu_2 < \bar{X}_1 + \bar{X}_2 + (\ldots)$$

if $n_1 = n_2$,

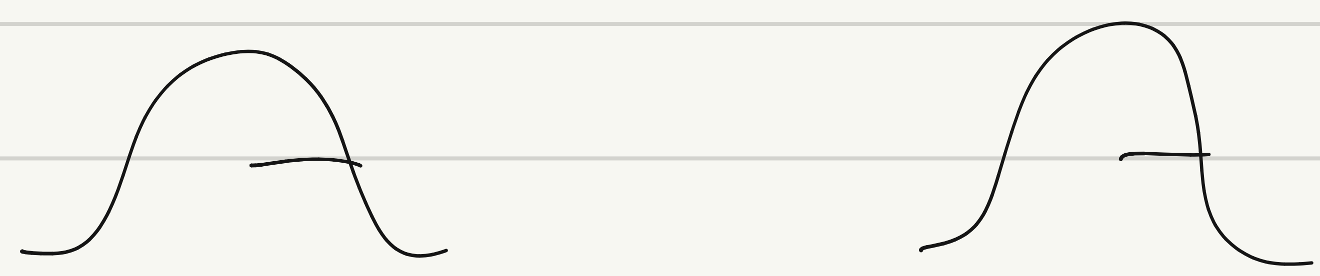
$$-t_{(n-2, \alpha/2)}\, s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

합동 표본 분산 $= S_p$

⑩ 합동 표본분산 (Pooled variance) ※ (그룹 1, 2가 같은 분산을 가지고 있을 때에만)

$$S_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

→ 기각역이 좁아지고 분석이 용이해짐.



μ_1 μ_2

$$s_1^2 = \frac{\Sigma(x_i-\bar{x})^2}{n_1-1} \qquad s_2^2 = \frac{\Sigma(y_i-\bar{y})^2}{n_2-1}$$

$$S_p^2 = \frac{\Sigma(x_i-\bar{x})^2 + \Sigma(y_i-\bar{y})^2}{n_1+n_2-2}$$

자유도를 2개를 뺌

⑪ T분포, 카이^2 분포, F분포

⑫ 대응표본 T검정 (Paired T-test)

- 서로 대응되는 표본 측정치 간에 평균에 차이가 있는지를 검정.
- 표본 간에는 독립성 유지

ex) n_b n_f

160 - 116 → 44

152 - 136 → 16

⋮ ⋮ -16

40

$\bar{D}$

$$t = \frac{\bar{D}}{S_D}$$