우연오차와 그의 처리

목차

1. 정규분포 (Normal Distribution)

- 가우스 분포 (Gaussian Distribution).
- 평균을 중심으로 대칭을 이루는 종모양의 연속 확률 분포.
- 확률변수 X의 확률 밀도 함수.

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\pi-\mu)^2}{2\sigma^2}}, -\infty < x < \infty$$

y : 오차발생의 잦음도

x : 매번의 측정치 (확률변수)

- 정규분포 N (μ, σ^2)

표준정규분포 (Standard Normal Distribution)

: 정규분포에서 평균이 0이고 분산이 1인 경우.

N (0,1)

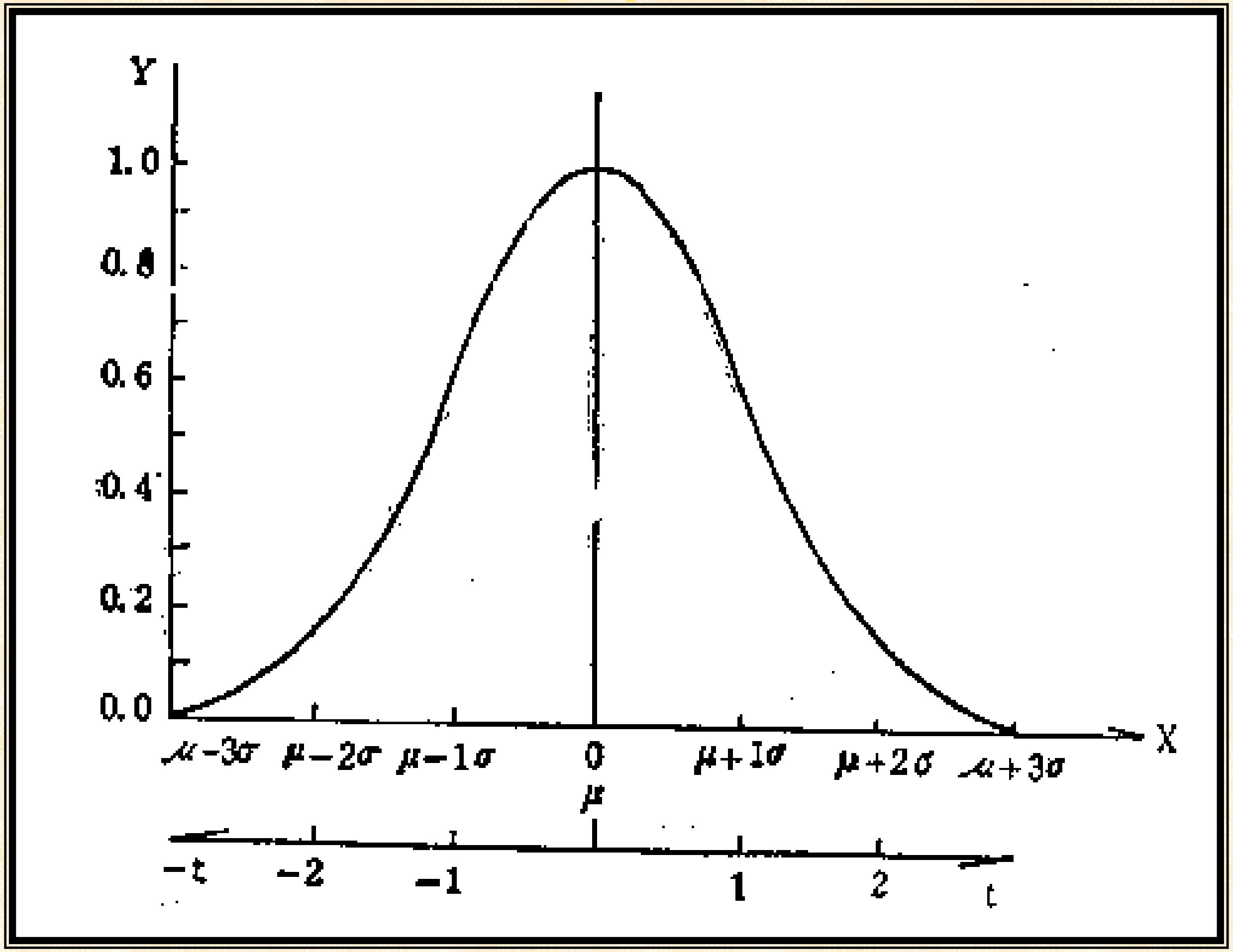
표준화 변수 T를 도입.

$$t = \frac{x - \mu}{\sigma} \qquad \varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

정규곡선면적표(正規曲線面積表)

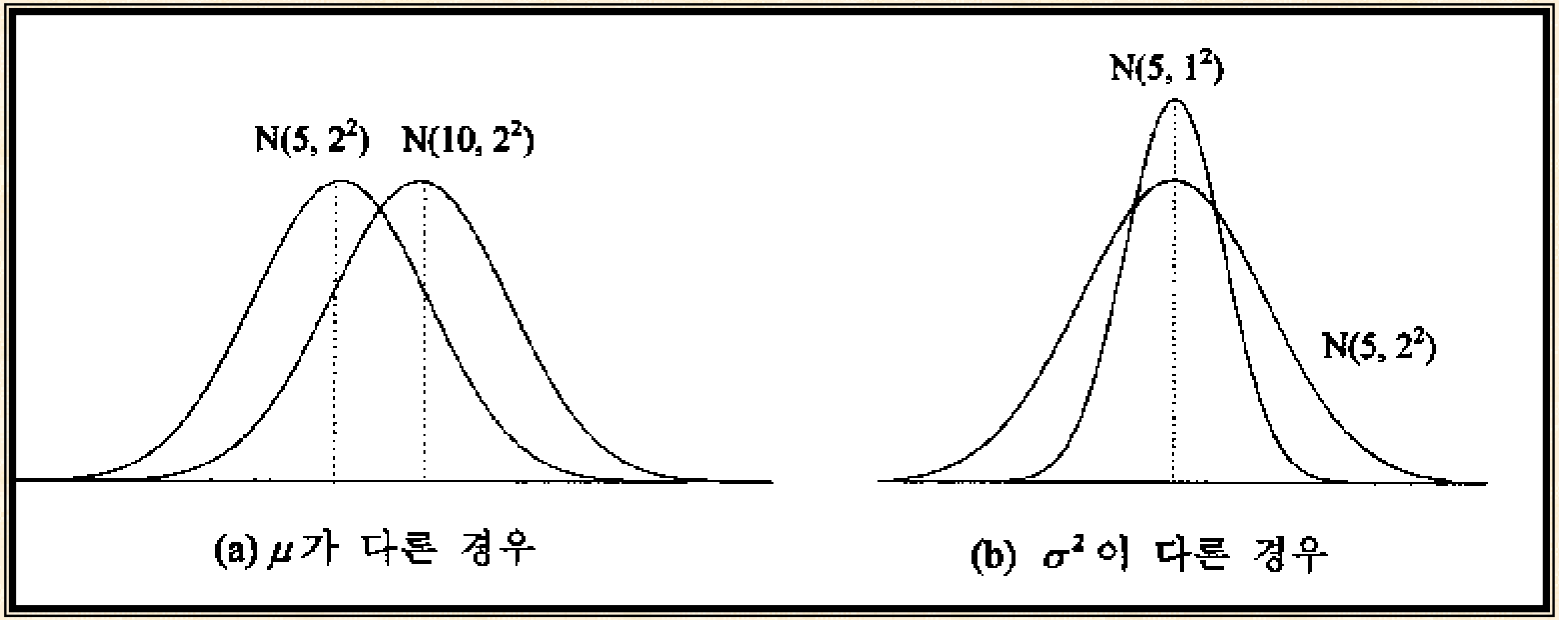
$$\Phi(t) = \int_0^t \phi(t)\,dt$$

t	.00	.01	.02	.03	.04
0.0	.00000	.00399	.00798	.01197	.01595
0.1	.03983	.04380	.04776	.05172	.05567
0.2	.07926	.08317	.08706	.09095	.09483
0.3	.11791	.12172	.12552	.12930	.13307
0.4	.15542	.15910	.16276	.16640	.17003
0.5	.19146	.19497	.19847	.20194	.20540
0.6	.22575	.22907	.23237	.23565	.23891
0.7	.25804	.26115	.26424	.26730	.27035
0.8	.28814	.29103	.29389	.29673	.29955
0.9	.31594	.31859	.32121	.32381	.32639
1.0	.34134	.34375	.34614	.34850	.35083



Y
1.0
0.8
0.6
0.4
0.2
0.0
X
μ-3σ
μ-2σ
μ-1σ
0
μ
μ+1σ
μ+2σ
μ+3σ
-t
-2
-1
1
2
t

μ와 σ^2의 변화에 따른 정규분포의 형태



x 가 $\mu \pm 1\sigma$ 사이의 값을 취할 확률

$$P_r(\mu - 1\sigma \le x < \mu + 1\sigma) = P_r(|t| \le 1)$$
$$= 2 \times 0.34134 = 0.68268$$

$$P_r(\mu - 2\sigma \le x < \mu + 2\sigma) = P_r(|t| \le 2)$$
$$= 2 \times 0.47725 = 0.95450$$

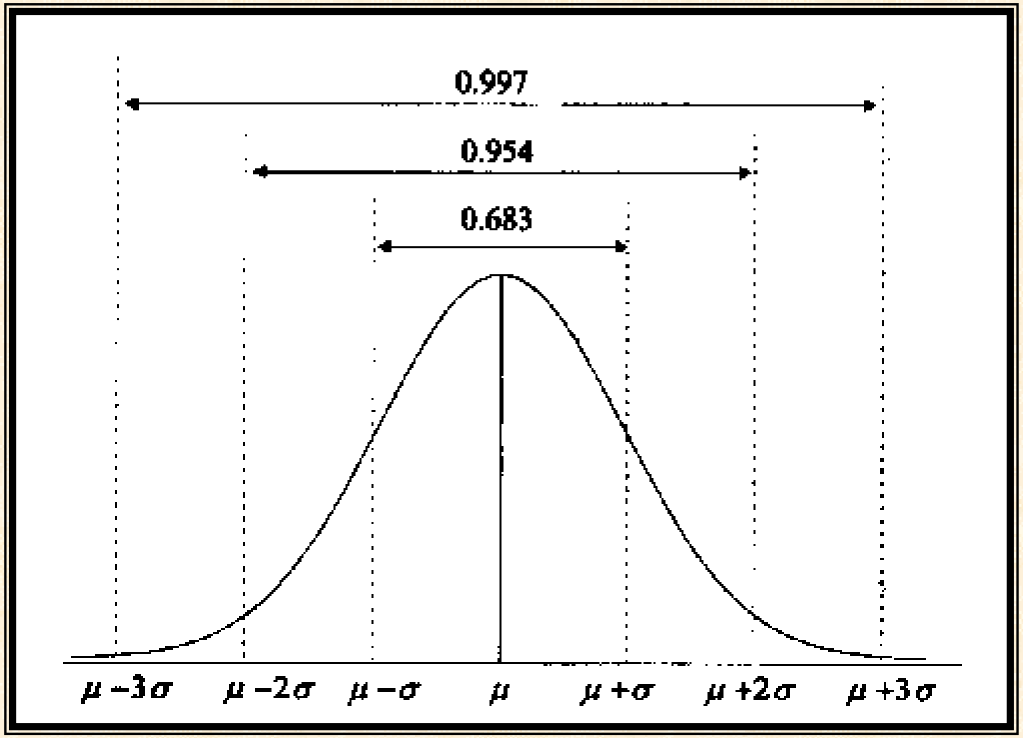
$$P_r(\mu - 3\sigma \le x < \mu + 3\sigma) = P_r(|t| \le 3)$$
$$= 2 \times 0.49865 = 0.99730$$

	0.0 · · · · ·
·	
·	
·	
·	
1.0	.3414

평균 μ , 표준편차 σ 인 참값으로부터 측정치 x 가

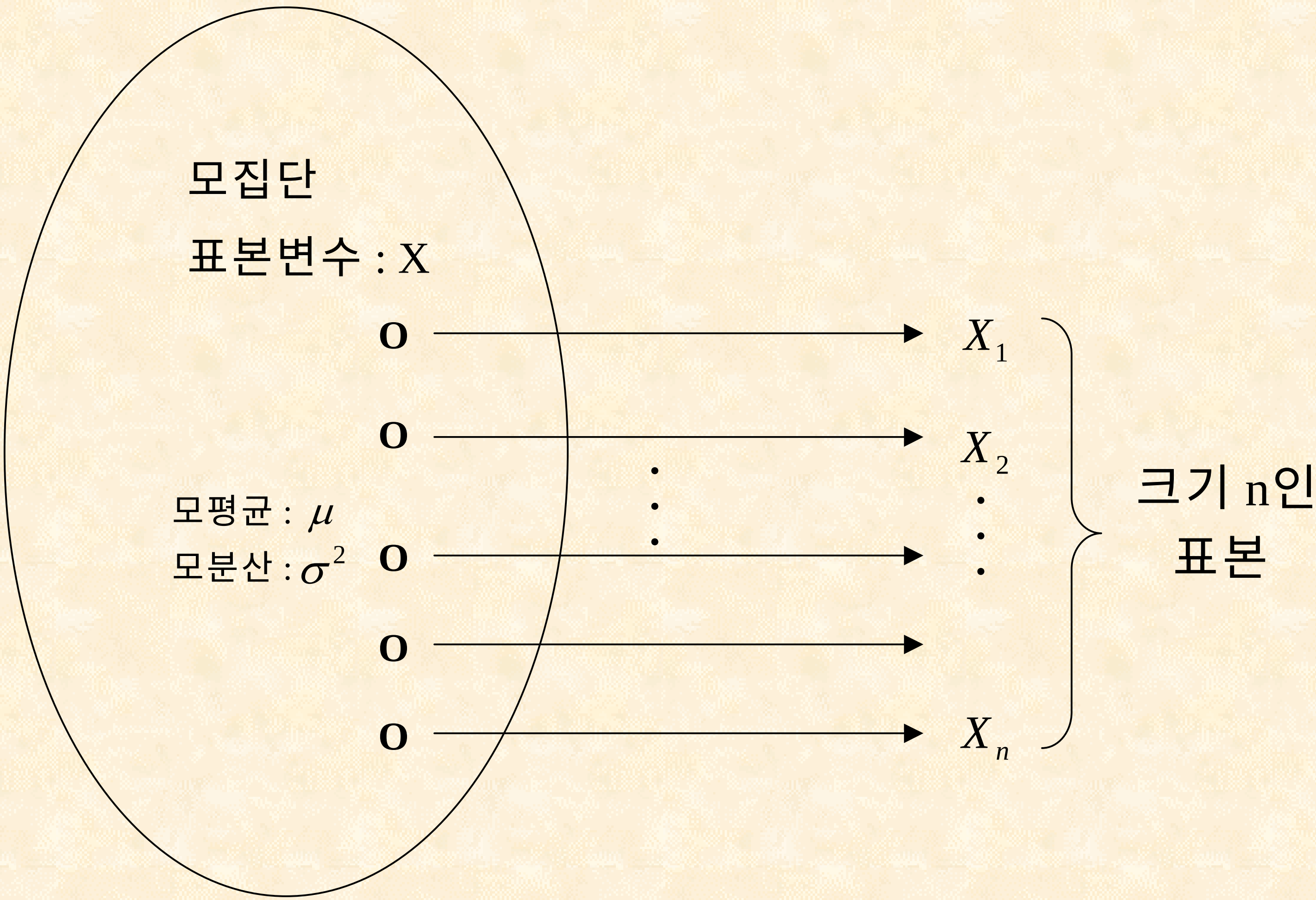
1. $\mu \pm 1\sigma$ 사이의 값을 취할 확률은 68.27%
2. $\mu \pm 2\sigma$ 사이의 값을 취할 확률은 95.45%
3. $\mu \pm 3\sigma$ 사이의 값을 취할 확률은 99.73%

정규분포의 확률밀도함수와 구간 확률



원시료(모집단) : 시료전체

공시료(표본) : 실험에 제공된 시료



표본 평균

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

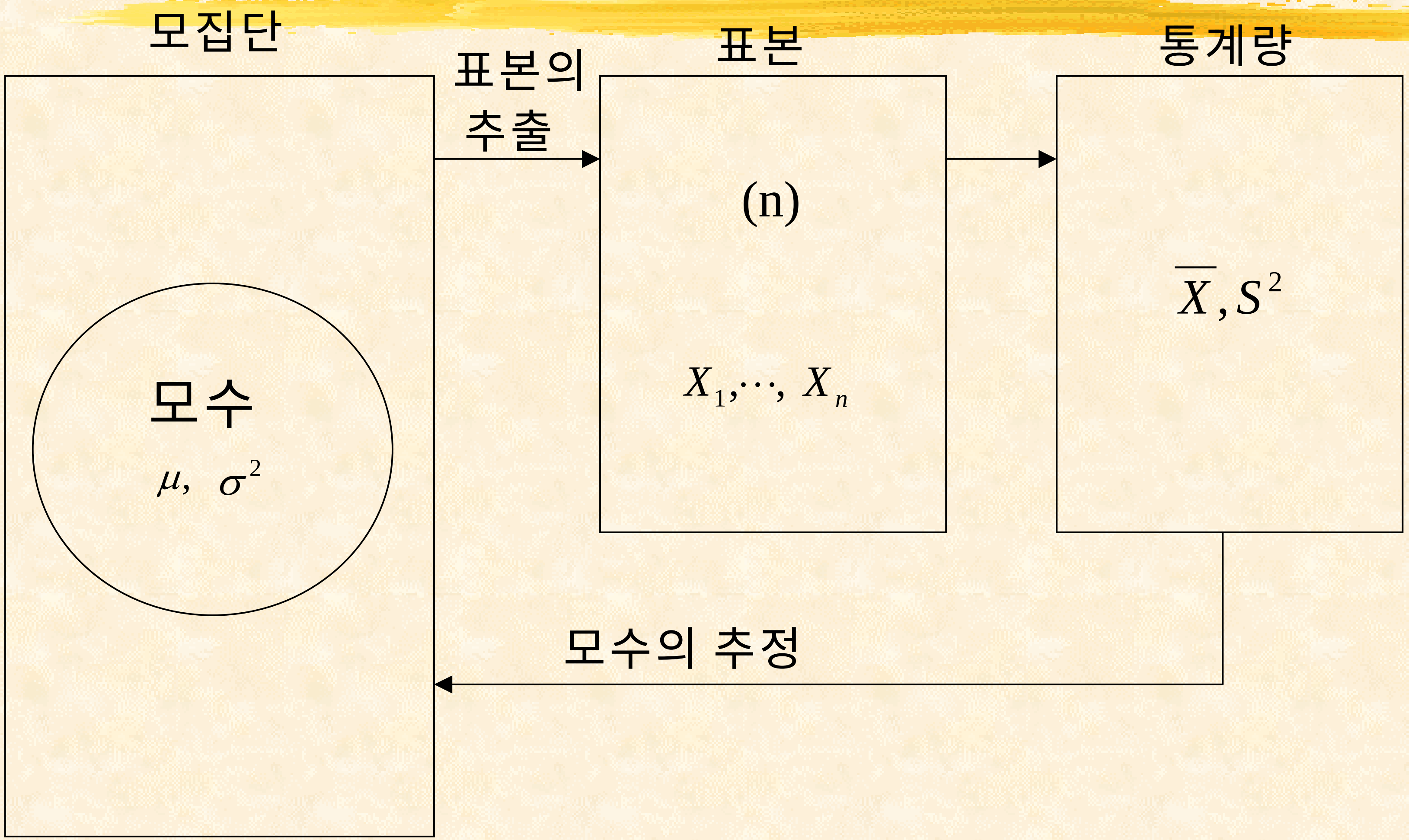
표본 분산

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

표본 표준 편차

$$S_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

<모수에 대한 추정 과정>



1) 원시료의 평균 추정

95%의 확률 : $\mu \pm 1.96\sigma$

99&의 확률 : $\mu \pm 2.58\sigma$

신뢰계수 95%인 신뢰한계

분석치의 평균이 $\bar{x}$일 때 원시료의 평균

$$\bar{x} - 1.96\sigma_{\bar{x}} < \mu < \bar{x} + 1.96\sigma_{\bar{x}} \quad (\text{표준오차} : \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}})$$

분석치의 개수 n이 충분히 클 때 σ_x 대신 s_x 사용.

$$\bar{x} - 1.96\frac{s_x}{\sqrt{n}} < \mu < \bar{x} + 1.96\frac{s_x}{\sqrt{n}}$$ (95% 신뢰구간)

$$\bar{x} - 2.58\frac{s_x}{\sqrt{n}} < \mu < \bar{x} + 2.58\frac{s_x}{\sqrt{n}}$$ (99% 신뢰구간)

$$\mu = \bar{x} \pm 1.96\frac{s_x}{\sqrt{n}}$$ (95% 신뢰구간)

$$\mu = \bar{x} \pm 2.58\frac{s_x}{\sqrt{n}}$$ (99% 신뢰구간)

측정치의 수가 적을 경우 ⟶ t분포에 의존

예제 1) 아연의 원자량 65.37

질량 분광기로 조사 질량수 64, 66, 67, 68, 70의 동위원소 검출. 질량수 70의 것도 아연 원자량으로 보아도 괜찮은가?

$$\bar{x} = \frac{64+66+67+68+70}{5} = 67$$

(단, 표준 오차 2.23)

$$\sigma_{\bar{x}} 2.23$$

$$\mu = 65.37$$

$$\left|\frac{\bar{x}-\mu}{\sigma_x}\right| = \left|\frac{67-65.37}{2.23}\right| = 0.73 < t_{0.05} = 1.96$$

원자량 70도 아연원자량에 포함시켜야 한다.

예제 2) 어떤 공업 원료 속에 들어 있는 철의 함량 측정 위해 25개 분석 평균치 32.40mg, 표준 편차가 0.20.

이 원료가 함유하는 철분의 평균 조성을 99% 및 95%의 신뢰한계로써 구하라.

$$\bar{x} = 32.40, s = 0.20, n = 25$$

(95% 신뢰구간) $$32.40 \pm 1.96\frac{0.20}{\sqrt{25}} = 32.40 \pm 0.08$$

32.48mg에서 32.32mg 사이

(99% 신뢰구간) $$32.40 \pm 1.96\frac{0.20}{\sqrt{25}} = 32.40 \pm 0.08$$

32.48mg에서 32.32mg 사이

가장 정확한 분석방법판단

측정치의 평균 $\bar{x}$ 을 모수의 평균 μ 와 비교(평균의 검정

차 $\left|\bar{x}-\mu\right|$ 크면 유의적

$$\frac{\left|\bar{x}-\mu\right|}{\sigma_{\bar{x}}} \geq 1.96 \qquad \text{유의수준 } 5\%$$

$$\frac{\left|\bar{x}-\mu\right|}{\sigma_{\bar{x}}} \geq 2.58 \qquad \text{유의수준 } 1\% \quad \left(\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\right)$$

예제 3) 어떤 균일한 합금 분석에서 Mn함량 20회 분석 평균 =32.37% 진짜 Mn함량=32.16% 표준편차=0.15%, 5%의 유의수준으로써, 이 분석 방법이 정확하였는가를 가려내어라.

$$\frac{\left|\bar{x}-\mu\right|}{S_x/\sqrt{n}} \geq 1.96$$

$$\bar{x} = 32.37$$

$$S_x = 0.15$$

$$\frac{\left|\bar{x}-\mu\right|}{S_x/\sqrt{n}} = \frac{32.37-32.16}{0.15/\sqrt{20}} = 6.27 \rangle \ t_{0.05} = 1.96$$

유의적, 분석방법이 부정확

2) 측정치의 평균 $\overline{x}$와 원시료의 평균 μ 와의 유의성 검정

*분석 평균치의 차의 유의성 검정

$$\frac{\left|\overline{x_1} - \overline{x_2}\right|}{\sigma_{\overline{x_1} - \overline{x_2}}} \geq 1.96 = t_{0.05}$$ (유의수준 5%)

$$\frac{\left|\overline{x_1} - \overline{x_2}\right|}{\sigma_{\overline{x_1} - \overline{x_2}}} \geq 2.58 = t_{0.01}$$ (유의수준 1%)

$$\sigma_{\overline{x_1} - \overline{x_2}} = \sqrt{\sigma_{\overline{x_1}}^{\;2} + \sigma_{\overline{x_2}}^{\;2}} = \sqrt{\frac{\sigma_{\overline{x_1}}^{\;2}}{n_1} + \frac{\sigma_{\overline{x_2}}^{\;2}}{n_2}}$$

예제 4) 어떤 경합금 분석 Al함량

무게분석 25회 평균 10.02%, σ = 0.5%

분광분석 15회 평균 10.12%, σ = 0.4%

95% 유의성으로써 이 두 방법에 차이에 있다고 보는가?

$$\overline{x_1} = 10.12$$

$$\overline{x_2} = 10.02$$

$$\sigma_{\overline{x_1}-\overline{x_2}} = \sqrt{\sigma_{\overline{x_1}}{}^2 + \sigma_{\overline{x_2}}{}^2} = \sqrt{\frac{\sigma_{\overline{x_1}}{}^2}{n_1} + \frac{\sigma_{\overline{x_2}}{}^2}{n_2}} = \sqrt{\frac{(0.40)^2}{15} + \frac{(0.50)^2}{25}} = 0.144$$

$$\frac{\left|\bar{x}_1 - \bar{x}_2\right|}{\sigma_{\bar{x}_1 - \bar{x}_2}} = \left|\frac{0.10}{0.144}\right| = 0.69 \langle \ 1.96$$

유의적이 아니다.
분석방법이 다른데서 오는 결과의 차이가 있다고 할 수 없다.

실제로는 $\sigma_{\bar{x}_1}$ 이나 $\sigma_{\bar{x}_2}$ 가 알려져 있는 예가 비교적 드물다.
따라서 이 경우 σ_x 대신 S_x에 의해 계산(F분포)

2.T-분포

1) 원시료의 평균 추정

$$t = \frac{\bar{x} - \mu}{S_x / \sqrt{n-1}}$$

자유도 $\nu = n-1$ 인 t-분포를 한다.

μ 를 추정함에 있어서 그의 신뢰구간 및 신뢰한계

$$\bar{x} - t_{\nu,p}\frac{s_x}{\sqrt{n-1}} \langle \mu \langle \bar{x} + t_{\nu,p}\frac{s_x}{\sqrt{n-1}}$$

$t_{\nu,p}$: ν는 자유도 p : 확률

cf.) $n_1=n_2$의 경우

t 분포표에서 $t_{\nu,p}$ 의 값을 자유도 $\nu = 2(n-1)$대신 $\nu = n-1$
에 대응하는 값을 사용

t—분 포 표

자유도 ν \ P	0.1	0.05	0.02	0.01	0.001
1	6.314	12.706	31.821	63.657	636.619
2	2.920	4.303	6.965	9.925	31.598
3	2.353	3.182	4.541	5.841	12.941
4	2.132	2.776	3.747	4.604	8.610
5	2.015	2.571	3.365	4.032	6.859

예제5) A,B두 실험실에서 석회 질소의 같은 시료를 각각 분석

A실험실: 20.20%   B실험실: 20.00% 의 질소 함량

신뢰계수 95%로써 석회질소의 신뢰한계를 구하여라.

$$\bar{x} = \frac{20.20 + 20.00}{2} = 20.10$$

$$s_x = \sqrt{\frac{1}{2}\left\{(20.20 - 20.10)^2 + (20.00 - 20.10)^2\right\}} = 0.10$$

$\nu = 2 - 1 = 1$ 0.05에해당한값 $t = 12.71$

$$\frac{t \cdot s_x}{\sqrt{n-1}} = \frac{(12.71)(0.10)}{\sqrt{2-1}} = 1.27$$

$$20.10 - 1.27 \langle \ \mu \ \langle \ 20.10 + 1.27$$

분석결과는 $20.10 \pm 1.27\%$

2) 두 종류의 측정치의 평균 $\bar{x}$ 의 비교

a) 측정치의 모분산이 동일하다고 볼 수 있는 경우

만약 같은 시료를 가지고 각각 다른 방법으로 분석한 값이

$(x_1, x_2, \cdots, x_{n1})$ 및 $(y_1, y_2, \cdots, y_{n1})$면 $\sigma_1 = \sigma_2$ 또는 $\mu_1 = \mu_2$ 이므로

$$t = \frac{\bar{x} - \bar{y}}{\hat{\sigma}_{\bar{x}-\bar{y}}}$$

$\hat{\sigma}$ 는 불편추정치(unbiased estimate)

측정치의 평균이 원시료의 평균과 일치하면 측정치의 평균은 원시료 평균의 불편추정치라고 한다.

$$\hat{\sigma}^2 = \frac{n_1 s_1{}^2 + n_2 s_2{}^2}{n_1 + n_2 - 2} = \frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 + \sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2}{n_1 + n_2 - 2}$$

$$\hat{\sigma}_{\overline{x}-\overline{y}} = \hat{\sigma}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = \hat{\sigma}\sqrt{\frac{n_2 + n_1}{n_1 n_2}}$$

예제6)어느 시료중에 들어있는 Ti-함량을 정량

무게 분석법 7.97, 7.92, 7.93, 7.94%

비색법 7.96,7.94,7.98,7.92,7.95%

1%의 위험율로써 이들 분석방법의 차이를 판단

비색법 $\bar{x} = 7.95$

무게분석법 $\bar{y} = 7.94$

$$\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 = 0.0020 \qquad \sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 = 0.0014$$

$n_1 = 5,\ n_2 = 4$ 이므로

$$\hat{\sigma} = \sqrt{\frac{0.0020 + 0.0014}{5 + 4 - 2}} \cong 2.2 \times 10^{-2}$$

$$\hat{\sigma}_{\bar{x}-\bar{y}} = 2.2 \times 10^{-2} \sqrt{\frac{1}{5} + \frac{1}{4}} = 1.48 \times 10^{-2}$$

$$\left| \frac{\bar{x} - \bar{y}}{\hat{\sigma}_{\bar{x}-\bar{y}}} \right| = \left| \frac{7.95 - 7.94}{1.48 \times 10^{-2}} \right| \cong 0.68 \langle\ t_{7,0.01} = 3.499$$

두 분석 방법의 차를 인정할 수 없다.

b) 측정치의 모분산이 동일하다고 볼 수 없는 경우

Cochran 및 Cox에 의한 근사법을 사용

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\hat{\sigma}_{\bar{x}}^{\,2} + \hat{\sigma}_{\bar{y}}^{\,2}}}$$

$$\hat{\sigma}_{\bar{x}}^{\,2} = \frac{\hat{\sigma}_1^{\,2}}{n_1} \quad , \quad \hat{\sigma}_{\bar{y}}^{\,2} = \frac{\hat{\sigma}_2^{\,2}}{n_2}$$

$$\hat{\sigma}_1^{\ 2} = \frac{n_1 s_1^{\ 2}}{n_1 - 1} = \frac{\sum\left(x - \bar{x}\right)^2}{n_1 - 1} = \frac{x\text{의 평방합}}{\text{자유도}}$$

$$\hat{\sigma}_2^{\ 2} = \frac{n_2 s_2^{\ 2}}{n_2 - 1} = \frac{\sum\left(y - \bar{y}\right)^2}{n_2 - 1} = \frac{y\text{의 평방합}}{\text{자유도}}$$

$$t_p = \frac{\hat{\sigma}_{\bar{x}}^{\ 2} \cdot t_{\nu_1, p} + \hat{\sigma}_{\bar{y}}^{\ 2} \cdot t_{\nu_2, p}}{\hat{\sigma}_{\bar{x}}^{\ 2} + \hat{\sigma}_{\bar{y}}^{\ 2}}$$

$\left| t \right| \rangle\ t_p$ $\bar{x}$ 와 $\bar{y}$ 의 차는 유의적

3. F-분포

표본 분산 s_1^2 및 s_2^2 를 가지고 모분산 σ_1^2 과 σ_2^2 가 같은지의 여부를 검정

$$F = \frac{V_1}{V_2}$$

V_1, V_2 :모분산의 불편추정치

$$V_1 = \frac{n_1 s_1^2}{n_1 - 1} \qquad V_2 = \frac{n_2 s_2^2}{n_2 - 1}$$

$$F > F_{\nu_1}^{\nu_2} \qquad \sigma_1^2 \neq \sigma_2^2$$

$$F < F_{\nu_1}^{\nu_2} \qquad \sigma_1^2 = \sigma_2^2$$

분석치의 횟수가 작을 경우

$$s_x = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}{n-1}}$$

$$\hat{\sigma}^2 = \frac{(n_1 - 1)s_x{}^2 + (n_2 - 1)s_y{}^2}{n_1 + n_2 - 2}$$

$$F = \frac{s_1{}^2}{s_2{}^2}$$

예제7)미지시료인 합금 분석알미늄침전 분석 결과

$Al(OH)_3$로 침전 Al-함량이 3.56, 3.58, 3.51, 3.55% 시료를

8-hydroxyquinoline으로 침전 후 브롬법으로 적정한 결과

3.50, 3.53, 3.56, 3.52, 3.54%

유의 수준 5%로써 두 방법의 차이를 인정할 수 있는지

(1) $$\bar{x} = 7.95$$

$$\bar{y} = 7.94$$

$$\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 = 0.0026 \qquad \sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 = 0.0020$$

$$V_x = \frac{0.0026}{3} = 0.00087$$

$$V_y = \frac{0.0020}{4} = 0.00050$$

$$F = \frac{0.00087}{0.00050} = 1.73 \langle \ F_3^4(0.05) = 6.59$$

분산의 차는 유의적이 아님을 알 수 있다.

따라서 지금 평균의 차의 유의성을 검정하면

$$\hat{\sigma}^2 = \frac{0.0026 + 0.0020}{3+4} = (2.56\times10^{-2})^2$$

$$\hat{\sigma}_{\bar{x}-\bar{y}} = \hat{\sigma}\sqrt{\frac{1}{n_1}+\frac{1`}{n_2}} = 2.56\times10^{-2}\sqrt{\frac{1}{4}+\frac{1}{5}} = 1.72\times10^{-2}$$

$$|t| = \frac{|\bar{x}-\bar{y}|}{\hat{\sigma}_{\bar{x}-\bar{y}}} = \frac{3.55-3.53}{1.72\times10^{-2}} = 1.16 \langle\ t_{7,0.01} = 3.499$$

역시 유의적이 아니므로 이들 두 분석 방법의 차는 인정할 수 없다.

(2)분석회수가 5회 이하이므로

$$s_x = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1}}$$

$$s_x{}^2 = 8.67\times10^{-4}$$

$$s_y{}^2 = 5.0\times10^{-4}$$

$$\hat{\sigma}^2 = \frac{(n_1 - 1)s_x{}^2 + (n_2 - 1)s_y{}^2}{n_1 + n_2 - 2}$$

$$= \frac{(4 - 1)(8.67 \times 10^{-4}) + (5 - 1)(5.0 \times 10^{-4})}{4 + 5 - 2}$$

$$= 6.57 \times 10^{-4} = (2.56 \times 10^{-2})^2$$

$$\hat{\sigma}_{\bar{x} - \bar{y}} = \hat{\sigma}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 2.56 \times 10^{-2}\sqrt{\frac{1}{4} + \frac{1}{5}} = 1.72 \times 10^{-2}$$

분산비: $F = \dfrac{8.67\times10^{-4}}{5.0\times10^{-4}} = 1.73 \langle\ F_4^3(0.05) = 6.59$

분산의 차는 유의적이 아니다.
따라서 평균의 차의 검정에 들어가면

$$|t| = \left|\frac{\bar{x}-\bar{y}}{\hat{\sigma}_{\bar{x}-\bar{y}}}\right| = \left|\frac{3.55-3.53}{1.72\times10^{-2}}\right| = 1.16\ \langle\ t_{7,0.01} = 3.499$$

역시 유의적이 아니어서 이 두 분석방법의 차를 인정할 수 없다.