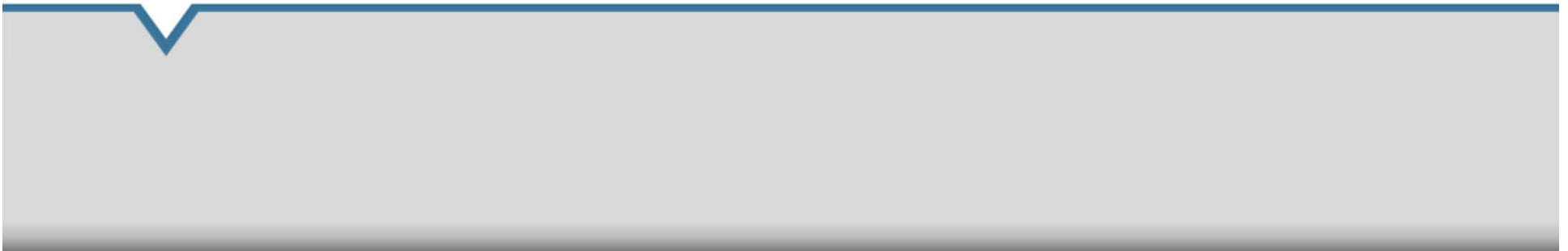
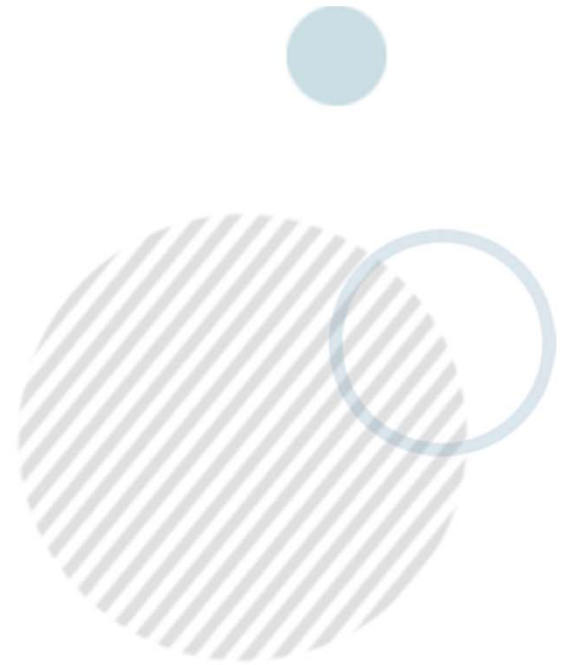


불확도 이해



통계 기초

데이터의 수량화 정리

중심경향을 나타내는 척도

평균값(mean: $\bar{x}$): 중심경향의 가장 보편적인 지수로서 집단내의 모집단을 더해서 데이터 개수로 나눈 값. $\bar{x} = \frac{\sum x}{n}$

중앙값(median: $\tilde{x}$): 데이터를 크기순으로 배열했을 때 그 중앙의 하나의 값(데이터의 수가 홀수일 경우) 또는 중앙의 둘의 값(짝수일 경우)의 산술 평균 값

최빈값(mode): 빈도수가 가장 많은 데이터의 값



통계 기초

분포의 산포를 나타내는 척도

산포는 품질수준을 유지할 수 있는지 또는 품질을 불안하게 하는 외적 원인이 존재하는지를 분석하는데 도움을 준다.

범위(Range): 한 조(set)의 여러 데이터 중 최대값과 최소값의 차이. $R = x_{max} - x_{min}$
보통 데이터가 10개 이하일 때 변동의 척도로서 관리도에 많이 사용되나 큰 단점은 두 데이터(최대값과 최소값)만 고려하고 조 내의 다른 데이터의 값의 영향은 고려되지 않는다는 점이다.

편차자승합(Sum of squares: S): 각 데이터의 값과 평균값과의 차를 편차라고 하는데 개개의 데이터가 집단의 중심으로부터 어느 정도 거리에 있는가를 표시한 값이다. 그런데 집단 전체를 알기 위해 이들 편차를 합치면 0이 되므로 이를 자승하여 합한 것을 편차자승합이라 한다. $S = \sum_i^n (x_i - \bar{x})^2$

분산(Variance: s^2 또는 σ^2): 편차자승합은 데이터 전체에 대한 분포를 표시하고 있어 데이터의 수가 많아지면 그 값도 커진다. 따라서 편차자승합을 데이터의 수로 나누어 데이터 단위당의 분포로 표시할 수 있는데 이를 분산이라 한다. $s^2 = \frac{\sum_i^n (x_i - \bar{x})^2}{n}$

통계 기초

불편분산(unbiased variance: V or s^2) : 분산과 다른 점은 분모가 n 이 아니고 $n-1$ 인데 통계적 계산에서 n 을 분모로 사용하면 모분산을 과소평가하게 되는 것으로 $n-1$ 을 사용함으로써 이 과소평가를 시정시켜 줌으로써 실제 모분산에 보다 접근하기 위해 사용.

$$V = \frac{\sum_i^n (x_i - \bar{x})^2}{n - 1}$$

표준편차(Standard Deviation: s or σ) : 분산의 제곱근으로 분산은 편차의 제곱으로 계산되기 때문에 원래 자료의 단위보다 큰 단위로 표시된다. 따라서 분산의 제곱근인 표준편차를 구하면 원래의 자료 단위로 환산되어 같은 단위로 측정된 평균이나 다른 통계량과 쉽게 비교할 수 있는 이점이 있다. $s = \sqrt{s^2}$

분산계수(Coefficient of Variance: C.V.) : 표준편차를 평균값으로 나눈 값으로 백분율로 나타낸다. 측정단위가 다른 두 종류의 데이터를 직접 비교하는 것은 곤란하다. 이 경우 분산계수를 사용하면 편리한데 측정단위가 다르더라도 동일한 데이터는 같은 산포의 측정값을 나타내기 때문이다. $C.V. = \frac{s}{\bar{x}} \times 100$

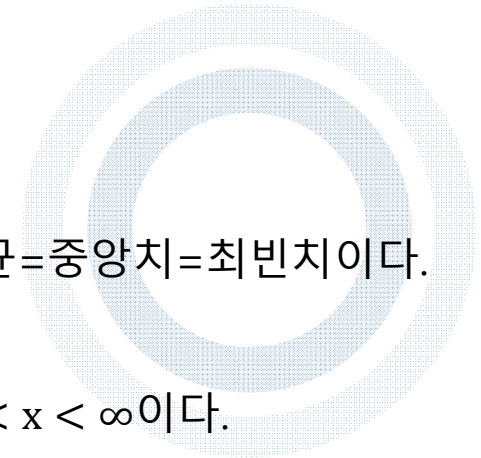
표준편차(s) vs 표준오차($s/\sqrt{n}$)?

통계 기초

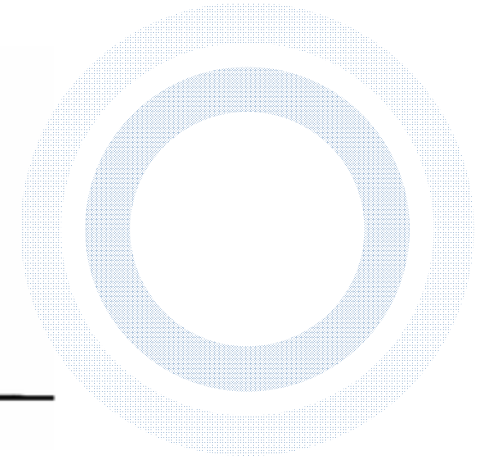
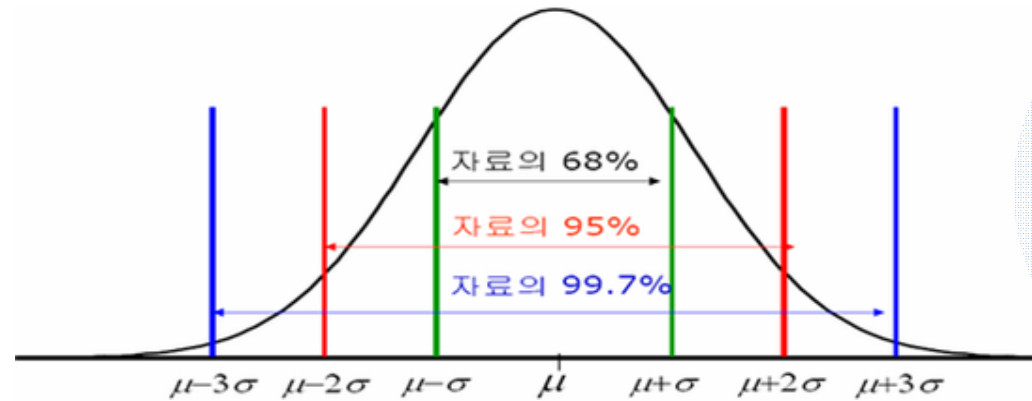
정규분포의 특징

1. 정규곡선은 종모양을 나타낸다.
2. 정규분포는 평균을 중심으로 좌우대칭을 이룬다. 따라서 평균=중앙치=최빈치이다.
3. 정규분포의 형태와 위치는 평균과 표준편차가 결정한다.
4. 정규곡선은 x축에 닿지 않으므로 확률변수 X 의 범위는 $-\infty < x < \infty$ 이다.
5. 정규곡선 밑의 면적은 1이다. 따라서 평균을 중심으로 오른쪽 또는 왼쪽으로 곡선 밑의 면적은 0.5이다.
6. 정규곡선 밑의 두 점 사이의 면적은 정규확률변수가 이들 두 점 사이를 취할 확률이다.
7. 평균과 다른 어떤 점 사이 정규곡선 밑의 면적은 그 점이 평균으로부터 떨어져 있는 표준편차의 수의 함수이다.

확률변수 X 가 평균 μ 와 분산 σ^2 으로 정규분포를 따른다면 $X \sim N(\mu, \sigma^2)$ 으로 표기할 수 있다.



통계 기초

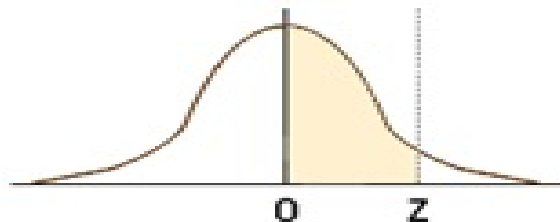


표준정규분포란 확률변수 Z 가 평균 $\mu=0$, 분산 $\sigma^2=1$ 인 정규분포를 따르는 것을 말하는데, Z 분포라고도 부르며 $Z \sim N(0,1)$ 로 표현한다.

$Z = \frac{x - \mu}{\sigma}$ 를 통해 정규확률변수 X 를 표준정규확률변수 Z 값으로 변형한다.

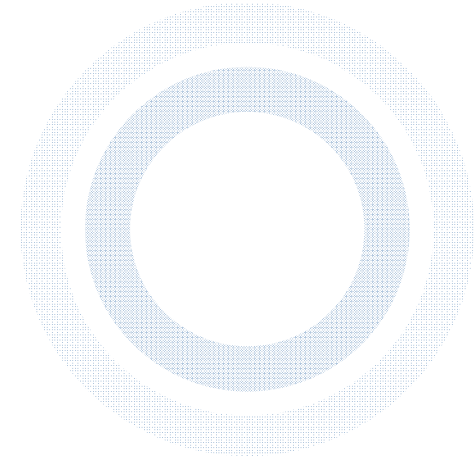
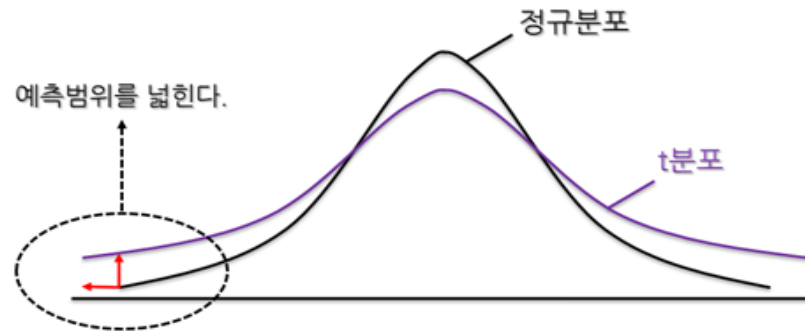
예를 들어 평균 50, 표준편차 5인 정규분포를 표준정규분포로 변환하면

$x=50$ 일 때 $Z=(50-50)/5=0$, $x=60$ 일 때 $Z=(60-50)/5=2$ 이다.



통계 기초

t분포



모집단의 평균과 표준편차를 모르고 표본크기가 작은 경우에($n < 30$) 모평균 μ 에 대한 신뢰구간의 설정은 t통계량을 이용해야 한다.

t분포는 표준정규분포(Z)와 유사한데, 즉 두 분포는 평균 0을 중심으로 좌우대칭이며 종모양을 나타낸다. 그러나 t분포는 표준정규분포보다 큰 분산을 갖는다. 그 이유는 모표준편차 σ 대신에 표본표준편차 s 를 사용하기 때문이다.

t통계량을 사용하여 신뢰구간을 설정하게 되면 Z통계량을 사용할 때보다 동일한 신뢰수준에서 그의 추정구간이 넓어지는데 이것은 표본표준편차 s 를 사용하는 데서 오는 추정상의 오류를 보상해 주기 위함이다.

불확도 이해

불확도의 개념

계측에 의한 값이나 계산에 의한 값 등 어떠한 자료를 이용해 도출된 추정치는 계측기에 의한 불확실성, 계측 당시 환경 조건이 표준 조건과 차이에 의한 불확실성, 산정식에 의한 불확실성 등 다양한 불확실성 요인에 의해 영향을 받게 된다.

이에 따라 추정치는 미지의 참값과의 편차(bias)를 보이게 되며, 추정치가 반복 측정값인 경우는 평균값을 중심으로 무작위(random)로 분산되는 양상을 보인다.

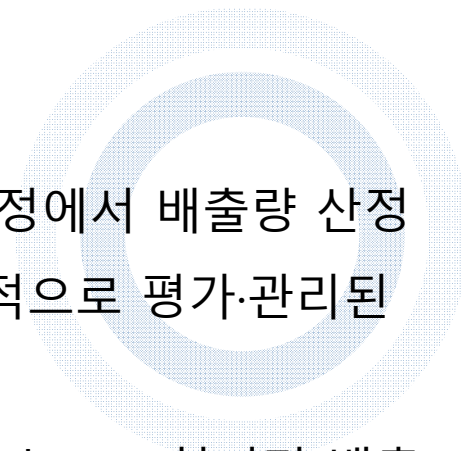
이러한 편차와 분산을 유발하는 불확실성 요인을 정량화하여 불확도(Uncertainty)로 표현하고 있다.

불확도 이해

불확도 관리 목적 및 범위

불확도는 온실가스 배출량의 신뢰도 관리와 제도 운영과정에서 배출량 산정과 관련된 방법론 및 방법 변경의 타당성을 입증하는 목적으로 평가·관리된다.

온실가스 배출량은 활동자료, 배출계수 등 매개변수의 함수로 표현되며 배출량 불확도는 활동자료와 배출계수 불확도를 합성하여 결정한다.



불확도 이해

불확도의 종류

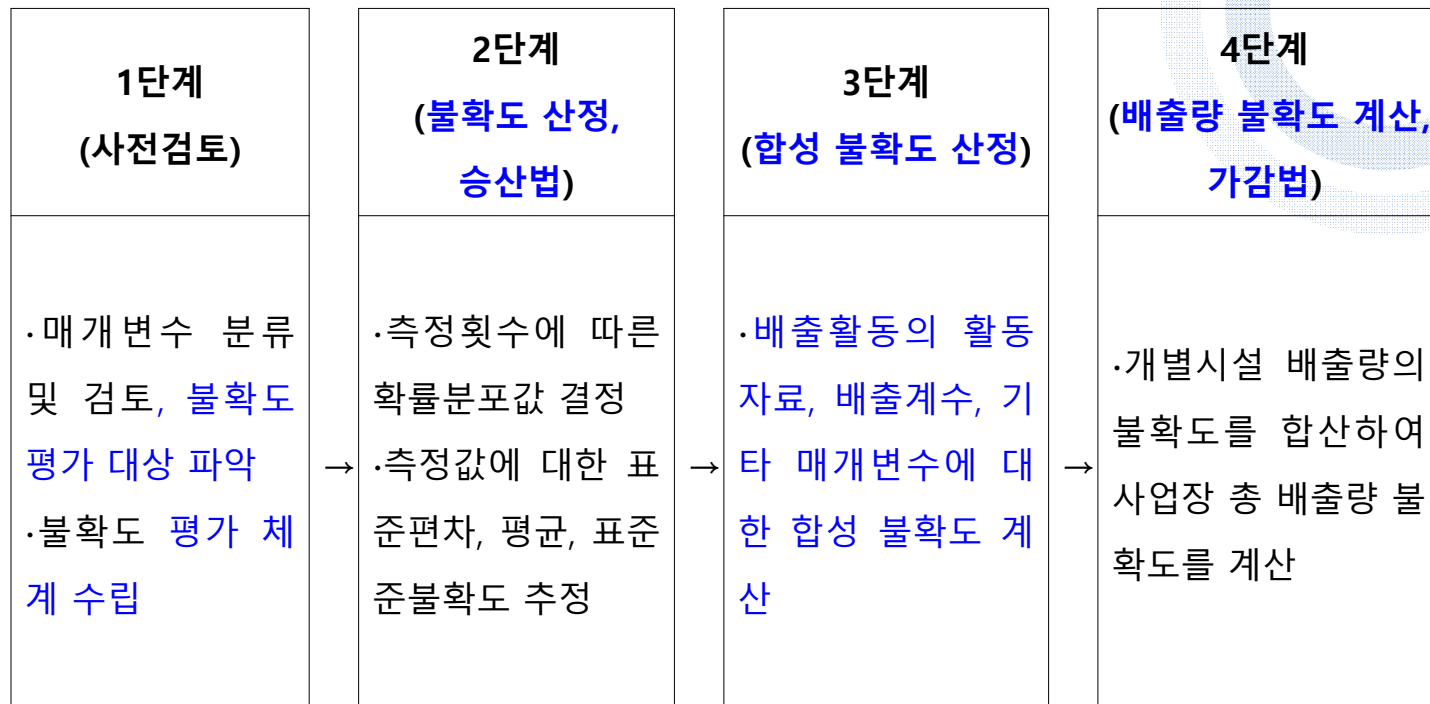
불확도는 표준불확도, 확장불확도, 상대불확도, 합성불확도 등으로 구분

1. 표준불확도는 반복 측정값의 표준오차로서 표현
2. 확장불확도는 표준불확도에 신뢰구간을 특정 짓는 포함인자를 곱하여 결정
3. 상대불확도는 불확도를 비교 가능한 값으로 환산하기 위해 불확도를 최적 추정값(평균)으로 나누고 100을 곱하여 백분율로 표현
4. 합성불확도는 여러 불확도 요인이 존재하는 경우 각 인자에 대한 불확도를 합성



불확도 이해

온실가스 측정 불확도 산정절차



불확도 이해

1. 측정값에 대한 통계량(표본 평균과 표본 표준편차), 표준불확도, 확장불확도 계산

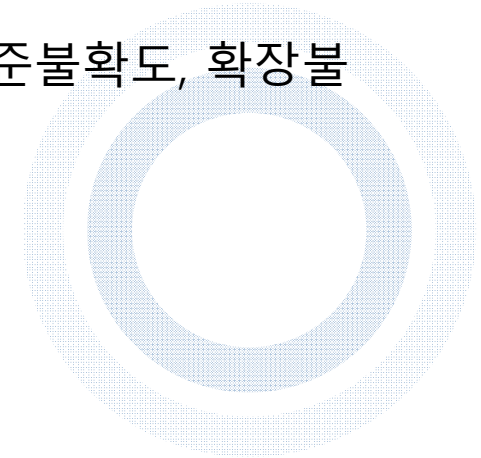
$$\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k$$
$$s = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2}$$

표준불확도(표준오차)

$$U_s = \frac{s}{\sqrt{n}}$$

확장불확도는 95% 신뢰수준

$$U_p = t \times \frac{s}{\sqrt{n}}$$



불확도 이해

2. 각 매개변수에 대한 상대확장불확도($U_{r,p}$)계산

$$U_{r,p} = \frac{U_p}{x} \times 100$$

관리업체가 보고해야 할 불확도는 상대확장불확도

3. 합성 불확도 산정(승산법)

배출량이 여러 매개변수의 곱으로 표현되는 경우 합성방법 중의 하나인 승산법에 따라 각 매개변수의 상대불확도를 합성

$$U_{r,E} = \sqrt{U_{r,A}^2 + U_{r,B}^2 + U_{r,C}^2 + U_{r,D}^2 + \dots}$$

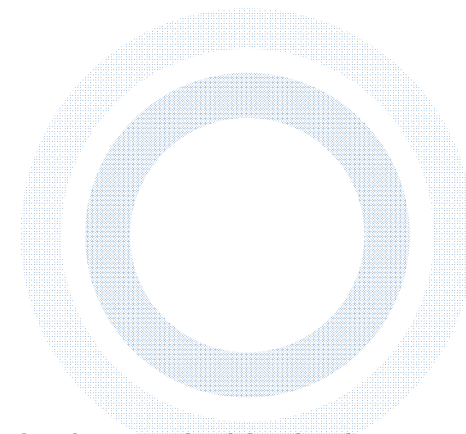
$U_{r,E}$: 배출량(E)의 상대확장불확도(%)

$U_{r,A}$: 활동자료(A)의 상대확장불확도(%)

$U_{r,B}$: 배출계수(B)의 상대확장불확도(%)

$U_{r,C}$: 매개변수 C 의 상대확장불확도(%)

$U_{r,D}$: 매개변수 D 의 상대확장불확도(%)



불확도 이해

4. 불확도의 조합(가감법)

사업장 혹은 관리업체의 온실가스 배출량은 개별 배출원 혹은 배출시설의 합으로 표현되며, 합으로 표현되는 값에 대한 불확도는 가감법에 따라 개별 불확도를 합성하여 산정한다.

$$U_{r,E_T} = \frac{\sqrt{\sum (E_i \times U_{r,E_i} / 100)^2}}{E_T} \times 100$$

U_{r,E_T} : 사업장/배출시설 총 배출량(E_T)의 상대확장불확도(%)

E_T : 사업장/배출시설의 총 배출량(이산화탄소 환산 톤)

E_i : E_T 에 영향을 미치는 배출시설/배출활동(i)의 배출량(이산화탄소 환산 톤)

U_{r,E_i} : E_T 에 영향을 미치는 배출시설/배출활동(i)의 상대확장불확도(%)

산정 tip

[예시 4] 동일한 측정기기를 이용하여 반복측정에 의한 불확도 산정예시

- 연간 코크스 사용량(i)을 측정기기를 활용하여 동일 무게를 12회 측정

측정횟수(n)	1	2	3	4	5	6	7	8	9	10	11	12
표준무게기준	54	52	55	53	50	57	42	45	56	8	68	52

① 평균($\bar{x}$) : $\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k = 53.5 \text{ ton}$

② 표준편차(s) : $s = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2} = 6.56 \text{ ton}$

- ③ 95% 신뢰구간에서 측정횟수($n=12$)에 따른 확률분포값(t_i) 도출
- 확률분포(t_i) = 2.20

표본횟수(n)에 따른 확률분포(t)를 구하기 위한 표준정규분포표

<표-51> 신뢰구간 및 측정횟수에 표준정규분포표

측정 횟수(n)	신뢰구간					
	68.27%	90%	95%	95.45%	99%	99.73%
2	1.84	6.31	12.71	13.97	63.66	235.8
3	1.32	2.92	4.30	4.53	9.92	19.21
4	1.20	2.35	3.18	3.31	5.84	9.22
5	1.14	2.13	2.78	2.87	4.60	6.62
6	1.11	2.02	2.57	2.65	4.03	5.51
7	1.09	1.94	2.45	2.52	3.71	4.90
8	1.08	1.89	2.36	2.43	3.50	4.53
9	1.07	1.86	2.31	2.37	3.36	4.28
10	1.06	1.83	2.26	2.32	3.25	4.09
11	1.05	1.81	2.23	2.28	3.17	3.96
12	1.05	1.80	2.20	2.25	3.11	3.85
13	1.04	1.78	2.18	2.23	3.05	3.76

- ④ 코크스 사용량(i)에 대한 표준불확도(U_i) 계산

$$U_i = \frac{s_i \times t_i}{\sqrt{n}} \times \frac{1}{x} \times 100(\%) = \frac{6.56 \times 2.20}{\sqrt{12}} \times \frac{1}{53.5} \times 100(\%) = 7.79\%$$

- ⑤ 연간 코크스 사용량의 표준불확도는 7.79% 이다.

- ⑥ 불확도 표현은 53.5 ton $\pm$ 7.79 % (신뢰수준 95 %, $k=2.20$)

측정횟수	측정치(톤)	비고
1	5.1	신뢰수준 95% K=2.78
2	5.1	
3	5.2	
4	4.9	
5	5.0	

활동자료에 대한 불확도(%)는 얼마인가? (단, 비중은 1이다.)

[정답]

1) 평균 = $5.1 + 5.1 + 5.2 + 4.9 + 5.0 = 25.3 / 5 = 5.06$

2) 표본측정값의 표준편차

$$= \sqrt{[(1 \div 4) \times \{(5.1 - 5.06)^2 + (5.1 - 5.06)^2 + (5.2 - 5.06)^2 + (4.9 - 5.06)^2 + (5.0 - 5.06)^2\}]} \\ = 0.114$$

3) 불확도(%)

$$= (0.114 \times 2.78 \div \sqrt{5}) \times (1 \div 5.06) \times 100 \\ = 2.8\%$$

어떤 용액의 각 요인에 관한 상대불확도는 다음과 같다고 함.

U1 : 0.17%

U2 : 0.084%

U3 : 0.13%

그러면 이 용액의 불확도는 얼마인가?

답: 0.23%

$$U = \sqrt{(0.17)^2 + (0.084)^2 + (0.13)^2}$$

