STATISTICS

Read OEEC&M 3rd Ed., Chapter 13, pages 307-413
2nd Ed., Chapter 42, pages 1071-1080

Use of Statistics

- Summarize data
- Hypothesis testing
- Measuring and evaluating risk
- Testing for compliance
- Identify sources of variation (error) and account for them

Analytical Limitations

- Specificity of the System
  - What is measured
- Positive Interference
  - Two materials behave similarly to yield high results
  - Measured value > true value
- Negative Interference
  - Two materials act in opposite ways to yield low results
  - Measured value < true value
- Sensitivity
  - Change in concentration needed to cause a change in response reading
  - Slope of the concentration – response curve
- Limits of Detection
  - Lowest value statistically different from blank
  - 3 times the standard deviation of the blank signal
  - Lowest standard actually run
    - Below it are less than values
    - Zero
    - Non-detection
- Limits of Quantitation
  - Level above which quantitative results can be obtained with confidence
  - Signal is 10 times the standard deviation of blank
- Precision
  - Closeness in agreement among results
  - Must have good precision
- Repeatability
  - Closeness in agreement among results in same lab, equipment, analyst
- Reproducibility
  - Closeness in agreement among labs
- Accuracy
  - Closeness in agreement between true value and measured value
  - Can be adjusted by calibration, use of standards and suitable blanks
- Bias
  - Difference between the true value and the measured value

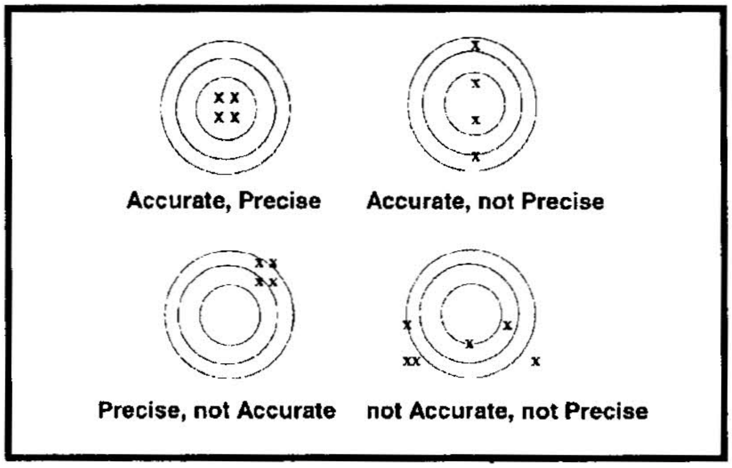


Figure 42.3 — Precision and accuracy.

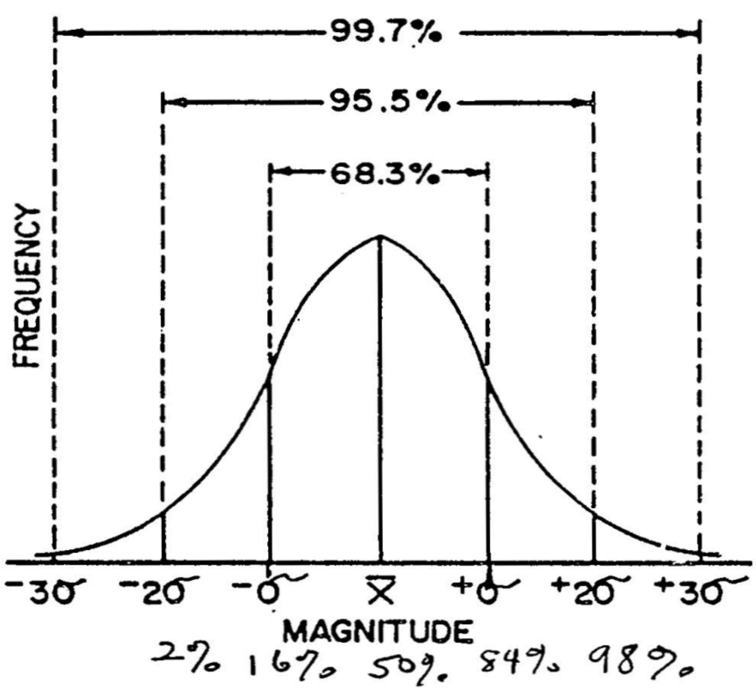
Describing the Sample Population

- If a determination is repeated a number of times under similar conditions the observed values (x_i) will be distributed randomly about an average as the result of errors. For large n, as $n \rightarrow \infty$, a plot of the relative frequency against the magnitude will produce a symmetrical bell-shaped (Gaussian or Normal) curve.
- The distribution is completely defined by the mean (x) and standard deviation (σ)

Mean $x = \Sigma x_i/n$

Standard Deviation $\sigma = \sqrt{}\Sigma(x_i - x)^2/n\text{-}1$

Gaussian or Normal Curve



Testing for Compliance

- Assumes a normal distribution

Confidence Interval (CI)

- Can determine a CI for the mean
- $CI_{95\%}$, 95% of all observations fall within $\pm 2\sigma$ of mean
- Used to test for differences among sites, populations, conditions
- Used to forecast the number of samples required to determine differences between means
- Used to test for compliance; does the $CI_{95\%}$ overlap the standard or does it lay outside?

Two-sided Tests

- Are usually used in statistics; z-value =1.96; standard error = $\sigma/\sqrt{n}$
  $CI_{(95\%)} = x \pm 1.96\ \sigma/\sqrt{n}$

One-sided Tests

- NIOSH & OSHA use a one-sided test
- Need to be 95% confident the observed value is above the PEL before they cite
- Use z=1.645
- $CI_{(95\%)} = x \pm 1.645\ \sigma/\sqrt{n}$

NIOSH & OSHA

- Simplify the calculations
- Coefficient of Variation (CV)
  - Variation relative to the mean
  - σ/x
  - Published for every method
- Standardization
  - Unitize
  - Divide equation by the standard
- $CI_{(95\%)} = x/STD \pm 1.645\ \sigma/STD\sqrt{n} = x/STD \pm 1.645CV$
- Sampling and Analytical Error (SAE) = 1.645CV
- $CI_{(95\%)} = x/STD \pm SAE$

Null Hypothesis for the Compliance Officer

- Is the exposure below the standard? If not cite
- Samples are collected to decide if it is possible to reject the null hypothesis that compliance exists
- Wants to minimize Type I error
  - The probability of declaring non-compliance when the true state is compliance

Null Hypothesis for the Employer

- State of non-compliance exists

- Employer's responsibility is to protect worker health
- Collect samples to see if it is possible to reject the null hypothesis that non-compliance exists
- Wants to minimize Type II error
    - The probability of declaring compliance when the true state is non-compliance

The compliance officer and the employer use one-sided statistical tests, but the compliance officer is interested in the lower confidence limit and the employer in the upper confidence limit

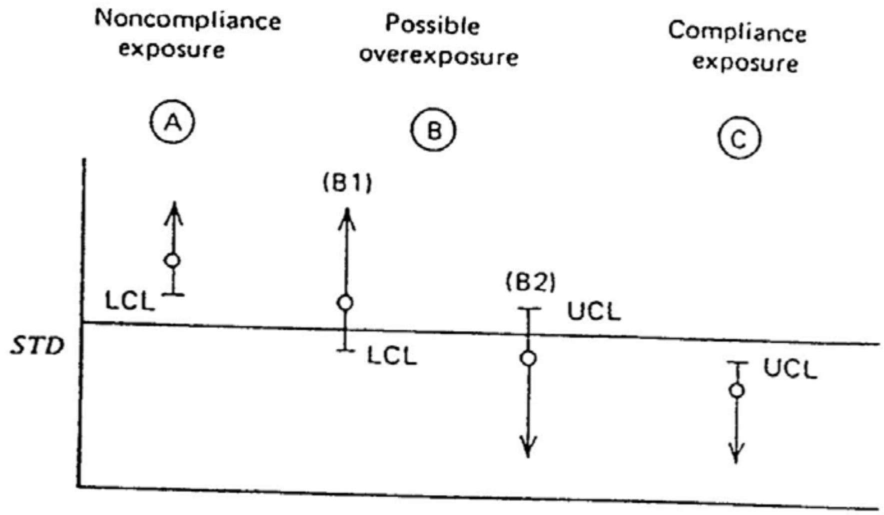
Cumulative Error

- Must consider all the errors involved in a measurement
- Cumulative error may have to be calculated
- Errors are not additive

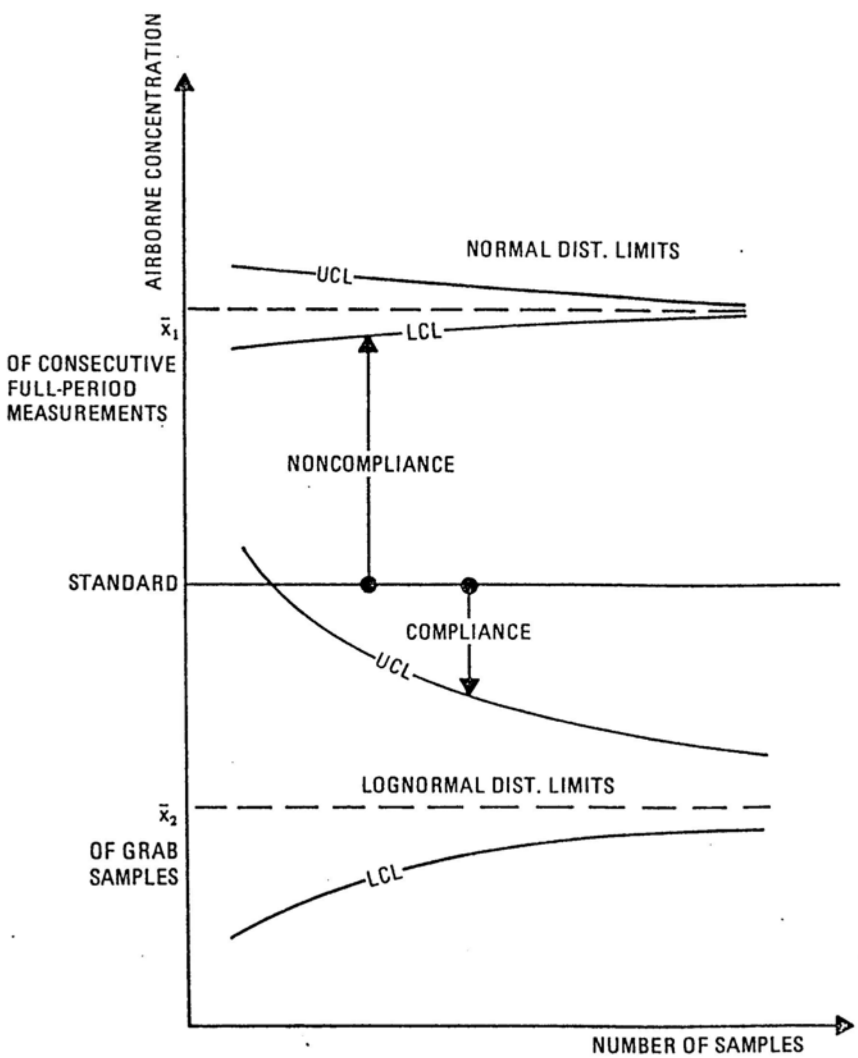
$$E_c = \sqrt{E_{flow}^2 + E_{time}^2 + E_{meas}^2 + E_{other}^2}$$

$$CV_T = \sqrt{CV_a^2 + CV_s^2}$$

ACTION \ TRUE STATE	COMPLIANCE WITH STANDARD	NONCOMPLIANCE WITH STANDARD
DECLARE COMPLIANCE	No Error	Type II Error
DECLARE NONCOMPLIANCE	Type I Error	No Error



Noncompliance exposure
Possible overexposure
Compliance exposure
A
B
C
(B1)
(B2)
LCL
LCL
UCL
UCL
STD



AIRBORNE CONCENTRATION
NORMAL DIST. LIMITS
UCL
LCL
$\bar{x}_1$
OF CONSECUTIVE FULL-PERIOD MEASUREMENTS
NONCOMPLIANCE
STANDARD
COMPLIANCE
UCL
LOGNORMAL DIST. LIMITS
$\bar{x}_2$
OF GRAB SAMPLES
LCL
NUMBER OF SAMPLES

EXPOSURE CLASSIFICATION

Full Period Single Sample Measurement

- n=1
- $CI_{(95\%)}$= (TWA/PEL) ± 1.645 CV

Full Period Consecutive Samples, Uniform Exposure

- All sample intervals equal
  - $CI_{(95\%)}$= (TWA/PEL) ± 1.645 CV/$\sqrt{}$n
- Unequal sample intervals
  - $CI_{(95\%)}$= (TWA/PEL) ± 1.645 CV$\sqrt{}(T_1)^2 + (T_2)^2 + \ldots (T_n)^2/\Sigma T_i$
- Non-uniform exposure

$CI_{(95\%)}$= (TWA/PEL) ± 1.645 CV$\sqrt{}(T_1X_1)^2 + (T_2X_2)^2 + \ldots (T_nX_n)^2$/STD (ΣT_i) ($\sqrt{}1+CV^2$)

Partial Period Consecutive Samples

- Compare the unitized $CI_{(95\%)}$ with the Partial Period Limit(PPL) instead of unity
- PPL = period of STD/total time of sample collection

Many Samples, Normal Distribution

- Compute the mean, σ, $CI_{(95\%)}$
- NIOSH/OSHA,
  - Standardize or unitize (divide each observation by the standard)
  - Compute the mean, σ, $CI_{(95\%)}$ of the unitized data set

Many Samples, Log-Normal Distribution

- Take the $\log_{10}$ of each observation
- Compute the mean, σ, $CI_{(95\%)}$ of the log transformed data
  - Antilog of the mean is the geometric mean
  - Antilog of σ is the geometric standard deviation
- NIOSH/OSHA
  - Standardize or unitize (divide each observation by the standard)
  - Take the $\log_{10}$ of each unitized observation
  - Compute the mean, σ, $CI_{(95\%)}$ of the log transformed unitized data set
  - Antilog of the mean is the geometric mean; antilog of σ is the geometric standard deviation

Ceiling Values

- $CI_{(95\%)}$= (Highest Value/Ceiling) ± 1.645CV

Time-Weighted Average Exposure Calculation

- Uniform exposure
  - TWA = $[C_1T_1 + C_2T_2 + \ldots C_nT_n]/\Sigma T$
- Non-uniform exposure
  - Calculate the mean for each non-uniform exposure group
  - Calculate TWA

Determining Sample Size

TABLE A-1. SAMPLE SIZE FOR TOP 10% ($\tau = 0.1$) AND CONFIDENCE 0.90 ($\alpha = 0.1$) (USE $n = N$ if $N \le 7$)

Size of group (N)	8	9	10	11-12	13-14	15-17	18-20	21-24	25-29	30-37	38-49	50	∞
Required No. of measured employees (n)	7	8	9	10	11	12	13	14	15	16	17	18	22

TABLE A-2. SAMPLE SIZE FOR TOP 10% ($\tau = 0.1$) AND CONFIDENCE 0.95 ($\alpha = 0.05$) (USE $n = N$ if $N \le 11$)

Size of group (N)	12	13-14	15-16	17-18	19-21	22-24	25-27	28-31	32-35	36-41	42-50	∞
Required No. of measured employees (n)	11	12	13	14	15	16	17	18	19	20	21	29

TABLE A-3. SAMPLE SIZE FOR TOP 20% ($\tau = 0.2$) AND CONFIDENCE 0.90 ($\alpha = 0.1$) (USE $n = N$ if $N \le 5$)

Size of group (N)	6	7-9	10-14	15-26	27-50	51-∞
Required No. of measured employees (n)	5	6	7	8	9	11

TABLE A-4. SAMPLE SIZE FOR TOP 20% ($\tau = 0.2$) AND CONFIDENCE 0.95 ($\alpha = 0.05$) (USE $n = N$ if $N \le 6$)

Size of group (N)	7-8	9-11	12-14	15-18	19-26	27-43	44-50	51-∞
Required No. of measured employees (n)	6	7	8	9	10	11	12	14

For instance, with a ceiling standard defined for a 15-minute period, there are 32 discrete nonoverlapping periods in an 8-hour work shift. Thus, with $N = 32$ and with the use of Technical Appendix A, the following appropriate sample sizes are determined:

15-Minute period

At least one period from:	Confidence level	Sample at least:
Top 20%	0.90	9 periods
Top 20%	0.95	11 periods
Top 10%	0.90	16 periods
Top 10%	0.95	19 periods

Where the ceiling standard is defined for a 10-minute period, there would be 48 periods and the following sample sizes are appropriate:

10-Minute period

At least one period from:	Confidence level	Sample at least:
Top 20%	0.90	9 periods
Top 20%	0.95	12 periods
Top 10%	0.90	17 periods
Top 10%	0.95	21 periods

Very short time samples may sometimes be taken, as with a 3-minute detector tube or spot readings with a direct-reading meter. Then the appropriate number of samples to take is given by equation 5 of Technical Appendix A, and the results are:

Less than a 5-minute period

At least one period from:	Confidence level	Sample at least:
Top 20%	0.90	10 periods
Top 20%	0.95	13 periods
Top 10%	0.90	22 periods
Top 10%	0.95	28 periods

Random Sampling

- Statistics assume random sampling
- Eliminates bias
- Assign each worker, shift, day, time interval, etc. a number
- Use a random number generator or chart to select which employee, shift, day to monitor

TABLE 3.2. TABLE OF RANDOM NUMBERS FOR PARTIAL SAMPLING*

ROW / COLUMN	1				5					10					15					20					25
1	05	57	23	06	26	23	08	66	16	11	75	28	81	56	14	62	82	45	65	80	36	02	76	55	63
	37	78	16	06	57	12	46	22	90	97	78	67	39	06	63	60	51	02	07	16	75	12	90	41	16
	23	71	15	08	82	64	87	29	01	20	46	72	05	80	19	27	47	15	76	51	58	67	06	80	54
	42	67	98	41	67	44	28	71	43	08	13	47	76	30	26	72	33	69	92	51	95	23	26	85	70
	05	83	03	84	32	62	83	27	48	83	03	13	84	90	20	20	50	87	74	83	51	62	10	23	30
6	60	46	18	41	25	74	73	51	72	90	40	52	85	41	20	89	48	98	27	38	81	33	83	82	94
	32	80	64	75	81	98	09	40	64	89	29	99	46	35	69	91	50	73	75	92	90	56	82	93	24
	79	86	53	77	78	06	62	37	48	82	71	00	78	21	65	65	88	45	82	44	78	83	22	78	09
	45	13	23	32	01	09	46	36	43	66	37	15	35	04	88	79	83	53	19	13	91	59	81	81	87
	20	60	97	48	21	41	84	22	72	77	99	81	83	30	46	15	90	26	51	73	66	34	99	40	60
11	67	91	44	83	43	25	56	33	28	80	99	53	27	56	19	80	76	32	53	95	97	53	09	61	98
	86	50	76	93	86	35	68	45	37	83	47	44	52	57	66	59	64	16	48	39	26	94	54	66	40
	56	73	58	38	23	36	10	95	16	01	10	01	59	71	55	99	24	88	31	41	00	73	13	80	62
	55	11	50	29	17	73	97	04	20	39	20	22	[illegible]	[illegible]	43	00	15	10	12	35	09	11	00	89	05
	25	54	35	87	92	52	04	49	73	96	57	53	57	08	93	09	69	87	83	07	46	39	50	37	85
16	41	48	67	79	44	57	40	29	10	34	58	63	51	18	07	41	02	39	79	14	40	68	10	01	61
	03	97	71	72	43	27	36	24	59	88	82	87	26	31	11	44	28	58	99	47	83	21	35	22	88
	90	24	83	48	07	41	56	68	11	14	77	75	48	68	08	90	89	63	87	00	06	18	63	21	91
	98	98	97	42	27	11	80	51	13	13	03	42	91	14	51	22	15	48	67	52	09	40	34	60	85
	74	20	94	21	49	96	51	69	99	85	43	76	55	81	36	11	88	68	32	43	08	14	78	05	34
21	94	67	48	87	11	84	00	65	93	56	43	99	21	74	84	13	56	41	90	96	30	04	19	68	73
	58	18	84	82	71	23	66	33	19	29	65	17	90	84	24	31	75	36	14	83	86	22	70	86	89
	31	47	28	24	88	49	28	69	78	62	23	45	53	38	78	65	87	44	91	93	91	62	76	09	20
	45	62	31	06	70	92	73	27	83	57	15	64	40	57	56	54	42	35	40	93	55	82	08	78	87
	[illegible]	49	87	12	27	41	07	91	72	64	63	42	06	66	82	[illegible]	28	36	45	31	99	01	03	35	76
26	69	37	22	23	46	10	75	83	62	94	44	65	46	23	65	71	69	20	89	12	16	56	61	70	41
	93	67	21	56	98	42	52	53	14	86	24	70	25	18	23	23	56	24	03	86	11	06	46	10	23
	77	56	18	37	01	32	20	18	70	79	26	85	77	89	28	17	77	15	52	47	15	33	35	12	75
	37	07	47	79	60	75	24	15	31	63	25	93	27	66	19	53	52	49	98	45	12	12	06	09	32
	72	08	71	91	73	46	39	60	37	58	22	25	20	84	30	02	03	62	68	58	38	04	06	69	94
31	55	22	48	46	72	50	14	24	47	67	84	37	32	84	82	64	97	13	69	85	20	09	80	46	75
	69	24	99	90	70	29	34	25	33	29	12	69	90	50	38	93	84	32	28	95	03	65	73	90	12
	91	86	77	18	21	91	66	11	84	65	48	75	26	94	51	40	[illegible]	53	36	39	77	69	05	25	07
	[illegible]	43	94	96	80	61	34	28	46	28	11	48	48	94	50	65	06	63	71	06	19	35	05	32	56
	58	78	02	85	80	29	67	27	44	07	57	23	20	28	22	62	97	59	62	[illegible]	41	72	70	71	07
36	33	75	88	51	00	33	56	15	84	34	28	50	16	65	12	81	56	43	54	14	63	37	74	97	59
	58	60	37	45	62	09	95	93	16	59	35	22	91	78	04	97	98	80	20	04	38	97	13	92	30
	72	13	12	95	32	87	99	32	83	65	40	[illegible]	92	57	22	68	98	79	16	23	53	56	56	07	47
	22	21	[illegible]	16	10	52	57	71	40	49	95	25	55	36	95	57	25	25	77	05	38	05	62	57	77
	97	94	83	67	90	64	74	88	17	22	38	[illegible]	04	33	49	38	47	57	[illegible]	87	15	39	43	87	03
41	09	03	68	53	63	29	27	31	66	53	39	34	88	87	04	35	83	69	52	74	93	16	52	01	65
	29	95	61	42	65	05	72	27	28	18	09	85	24	59	46	03	91	55	58	62	51	71	47	37	38
	[illegible]	95	78	93	47	[illegible]	38	36	33	95	05	90	25	72	85	23	23	30	70	[illegible]	55	93	25	84	83
	44	52	20	81	21	57	57	85	00	47	26	13	87	22	45	72	03	[illegible]	75	23	38	38	56	77	97
	68	[illegible]	12	15	88	02	18	74	56	79	[illegible]	53	63	41	77	15	07	39	87	11	19	25	62	19	30
46	29	35	77	60	29	09	25	09	42	28	07	15	43	67	55	29	58	75	84	06	19	54	31	16	53
	54	13	39	19	29	64	97	73	71	[illegible]	78	03	24	92	93	86	69	76	74	28	08	98	84	38	23
	75	16	85	64	64	93	85	68	08	84	15	41	57	84	45	[illegible]	70	13	17	60	47	80	10	13	00
	36	47	17	08	78	03	92	85	18	42	95	48	27	37	99	98	81	94	44	72	05	95	42	31	17
	29	[illegible]	08	21	91	23	76	72	84	98	26	23	66	54	86	88	95	14	82	57	17	99	16	28	99

*Reproduced from Table A-36 of Natrella (3.1), with permission of the Rand Corporation, "A Million Random Digits," The Free Press, 1955.

Plot of Percent Frequency versus the Concentration

- Graphic technique that tests for normalcy and pictorially illustrates the sample data and distribution
- Bell-shaped curve, normal distribution
- Skewed curve, log-normal distribution

Plot of Cumulative Frequency versus Concentration

- When plotted using probability paper
    - Straight line, normal distribution
    - 50 percentile is the mean
    - 84%/50% = slope = σ

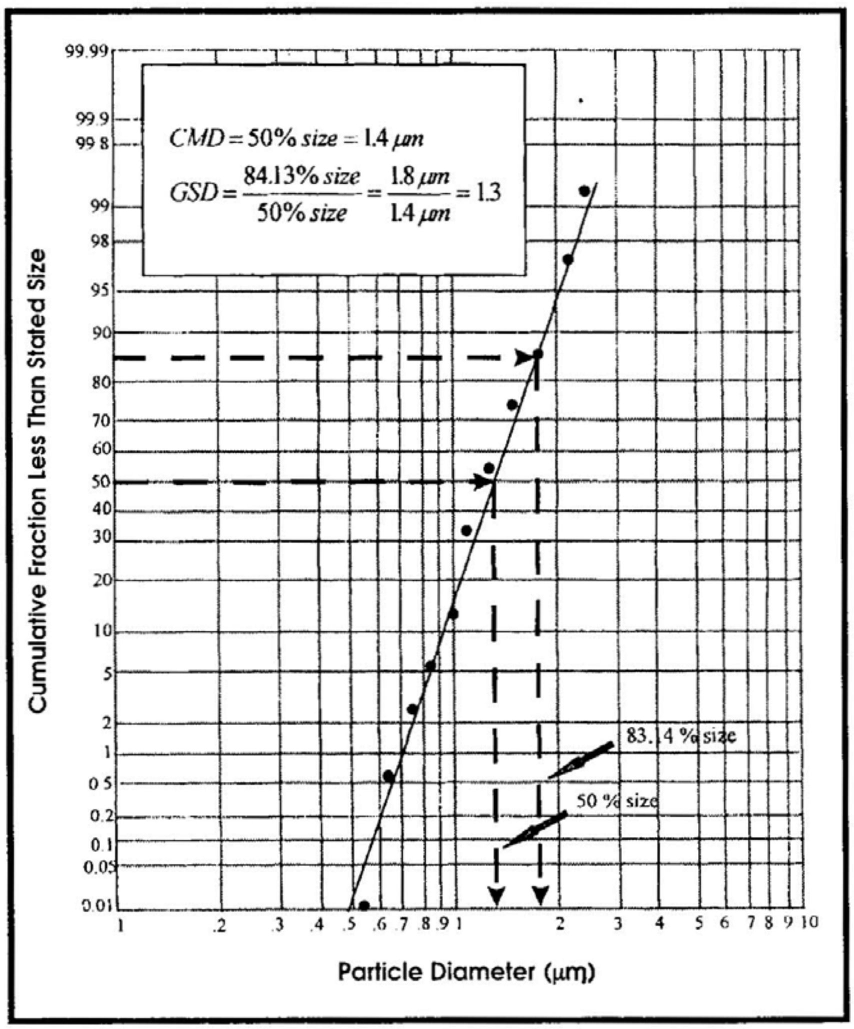


Figure 12.14 — Aerosol size distribution cumulative probability plot. The aerosol particle sizes are approximately lognormally distributed if a straight line provides a good fit to the measurement data. CMD and GSD are determined from the plotted line. Count-based distributions result from optical measurement instruments that estimate particle size by examining particle light scattering behavior.

Example Problems

1. A charcoal tube and personal pump were used to sample for 2-chloroacetophone (PEL=0.05 ppm). The lab reported 0.04 ppm with a coefficient of variation (CV) of 0.09. What is the compliance status?
   $CI_{(95\%)} = x/STD \pm 1.645CV = 0.04/0.05 \pm 1.645(0.09) = 0.8 \pm 0.148$
   $0.65 \leq CI_{(95\%)} \leq 0.95$
   UCL< 1, so compliance

2. If in the above example the sample time was 6 hours, what is the compliance status?
   PPL = 8/6 = 1.33
   UCL< 1.33, so compliance

3. A direct reading meter was used to monitor for ozone (PEL=0.1 ppm). Thirty five randomly selected readings were selected. What is the compliance status?

0.084	0.062	0.127	0.057	0.101	0.072	0.077
0.145	0.084	0.101	0.105	0.125	0.076	0.043
0.079	0.078	0.067	0.073	0.069	0.084	0.061
0.066	0.085	0.080	0.071	0.103	0.075	0.070
0.048	0.092	0.066	0.109	0.110	0.057	0.107

   Standardize

0.84	0.62	1.27	0.57	1.01	0.72	0.77
1.45	0.84	1.01	1.05	1.25	0.76	0.43
0.79	0.78	0.67	0.73	0.69	0.84	0.61
0.66	0.85	0.80	0.71	1.03	0.75	0.70
0.48	0.92	0.66	1.09	1.10	0.57	1.07

   Calculate:
   Mean= 0.831
   $\sigma = 0.230$
   $CI_{(95\%)} = x \pm 1.645\ \sigma/\sqrt{n} = 0.831 \pm 1.645(0.230)/\sqrt{35} = 0.831 \pm 0.0639$
   $0.767 \leq CI_{(95\%)} \leq 0.895$
   UCL< 1, so compliance

4. An employee is exposed to H_2S for about 16 short periods each work shift. The ceiling value is 20 ppm. Five randomly selected measurements were taken, and the results were 12, 14, 13, 16 and 15 ppm with a CV of 0.12. What is compliance status with the ceiling standard?
   $CI_{(95\%)} = (\text{Highest Value/Ceiling}) \pm 1.645CV = 16/20 \pm 1.645(0.12) = 0.8 \pm 0.197$
   $0.603 \leq CI_{(95\%)} \leq 0.997$
   UCL< 1, so compliance

5. Calculate the TWA.

Time of exposure (T_t)	*Average exposure concentration (ppm)*
1 hour	250
3 hours	100
4 hours	50
Total T = 8 hours	

Then the TWA for the 8-hour workday will be

$$TWA = \frac{(1)\ (250) + (3)\ (100) + (4)\ (50)}{8} = 94 \text{ ppm}$$

6. Calculate the TWA (non-uniform exposure situation).

Operation	*Duration*	*Sample*	*Results (of each 5-min sample)*
Solvent room	0800-1030	A	110 ppm
		B	180
		C	90
		D	120
		E	150
Printer feed	1030-1630	F	50
		G	35
		H	60
		I	40

The solvent room average exposure is

$$\overline{x}_1 = \frac{(110 + 180 + 90 + 120 + 150)}{5} = 130 \text{ ppm}$$

The printer feed average exposure is

$$\overline{x}_2 = \frac{(50 + 35 + 60 + 40)}{4} = 46 \text{ ppm}$$

Then the TWA exposure for the 8-hour shift (excluding 30 minutes for lunch) is

$$TWA = \frac{(2.5 \text{ hr})\ \ (130 \text{ ppm}) + (5.5 \text{ hr})\ \ (46 \text{ ppm})}{8 \text{ hr}}$$

$$= 72 \text{ ppm}$$

7. There are 20 assembly line workers in an electronics manufacturing facility. How many workers need to be monitored?
   - All workers exposed at the action level or higher.
   - 15 to have 95% confidence that at least one of the top 10% exposures were monitored.
   - 13 to have 90% confidence that at least one of the top 10% exposures were monitored.
   - 10 to have 95% confidence that at least one of the top 20% exposures were monitored.
   - 8 to have 90% confidence that at least one of the top 20% exposures were monitored.

8. A worker is exposed to ethylene oxide (STEL=5ppm). How many samples assure compliance with the short-term exposure limit?
   There are 32 discrete 15-minute intervals each work shift.
   - 19 to have 95% confidence that at least one of the top 10% exposures were monitored.
   - 16 to have 90% confidence that at least one of the top 10% exposures were monitored.
   - 11 to have 95% confidence that at least one of the top 20% exposures were monitored.
   - 9 to have 90% confidence that at least one of the top 20% exposures were monitored.

9. For the following toluene (PEL = 100 ppm) exposure data in ppm, compute the mean, standard deviation and 95% confidence interval for the mean. Plot the cumulative percent frequency on probability paper. What percent of the time would you expect the exposure to exceed the standard? (10 points)

18	73	41	46	61	52	48	30	51	54
34	36	47	63	35	34	39	104	95	34
60	62	129	42	95	98	69	50	90	42

$x = 57.7$

$\sigma = 26$

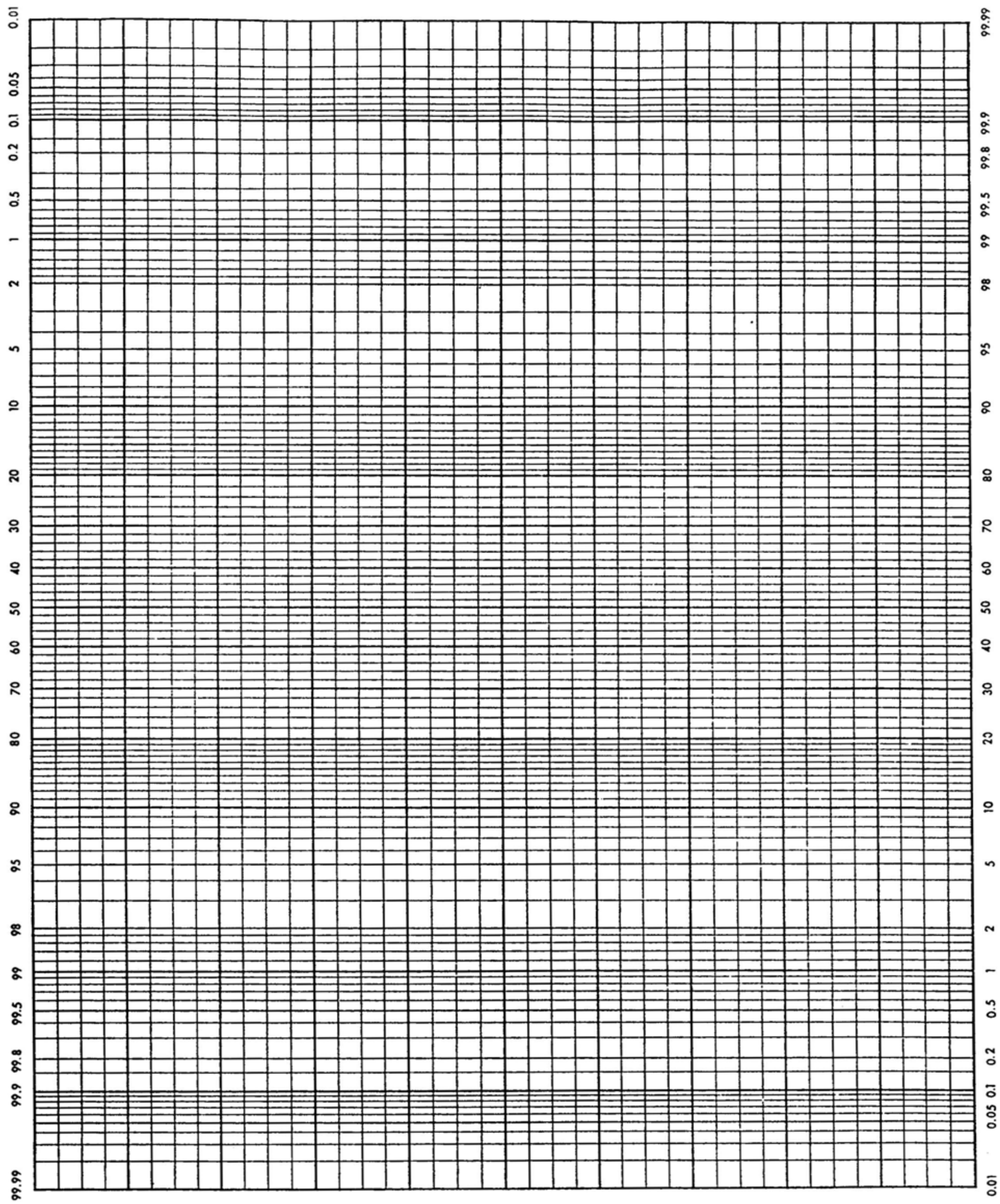
$CI_{(95\%)} = x \pm 1.645\ \sigma/\sqrt{n}$
$= 57.7 \pm 1.645\ (26)/\sqrt{30}$
$= 57.7 \pm 7.8$

$49.9 \leq CI_{(95\%)} \leq 65.5$

Compliance

Standard is exceeded @ 8% of the time

Interval	n	%	Σ%
0-19	1	3	3
20-39	7	23	26
40-59	10	33	59
60-79	6	20	79
80-99	4	13	92
100-119	1	3	95
≥120	1	3	98


0.01
0.05
0.1
0.2
0.5
1
2
5
10
20
30
40
50
60
70
80
90
95
98
99
99.5
99.8
99.9
99.99
99.99
99.9
99.8
99.5
99
98
95
90
80
70
60
50
40
30
20
10
5
2
1
0.5
0.2
0.1
0.05
0.01