

Generell statistikk - intro

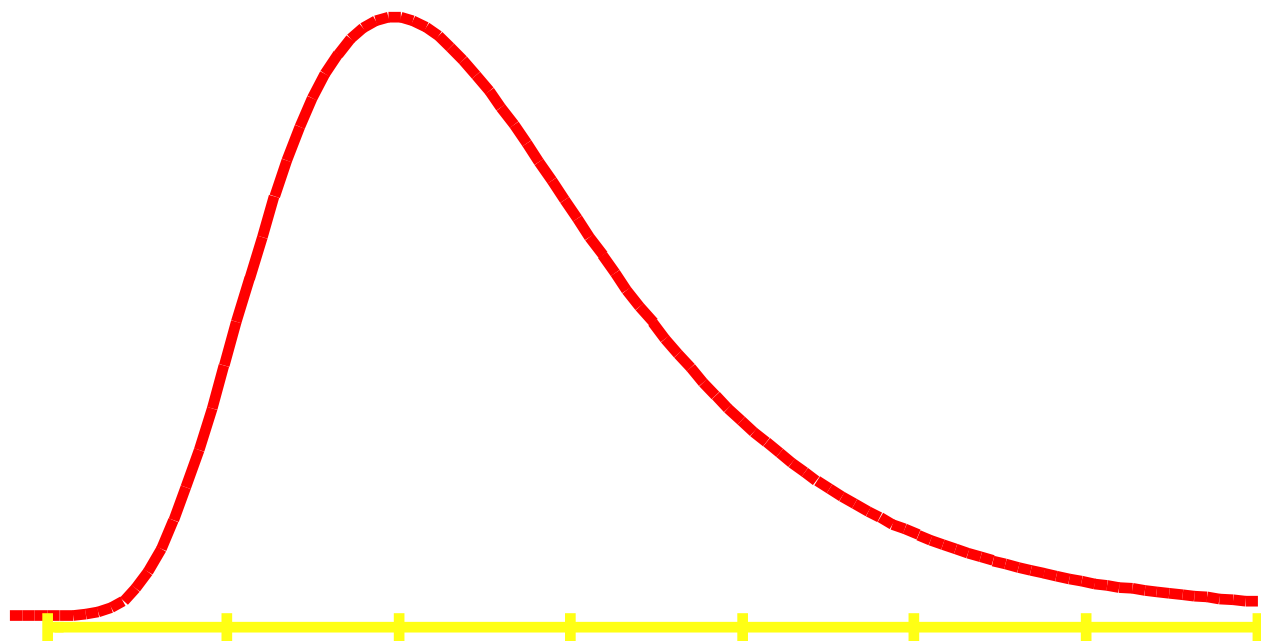
Hans Thore Smedbold

Workshop NS-EN 689,

Workshop hos Trondheim Kommune 20.06.2019

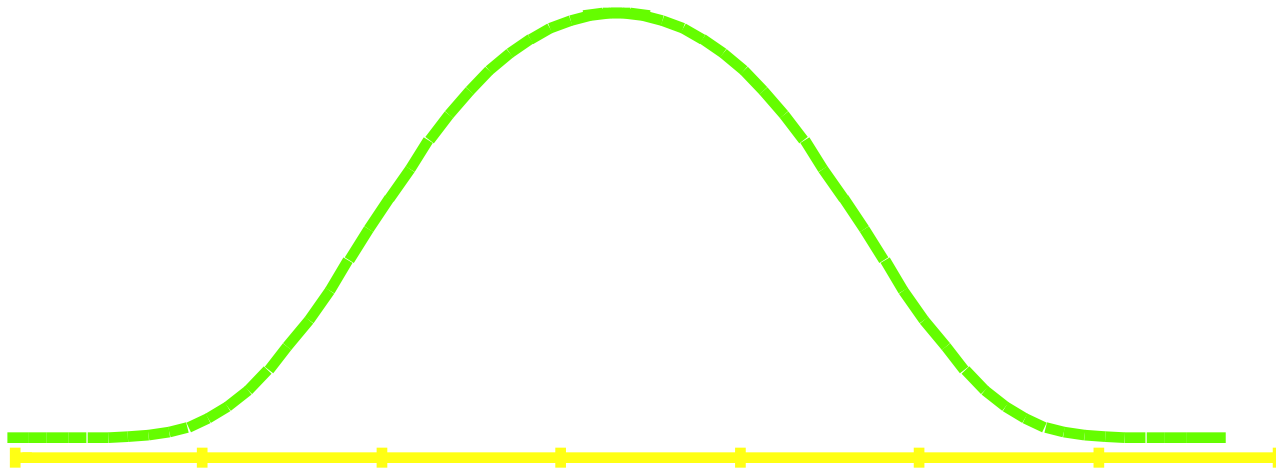
Fordeling av yrkeshygieniske data

- ▶ Erfaringsmessig viser det seg at miljødata har en tendens til å være log-normalfordelte

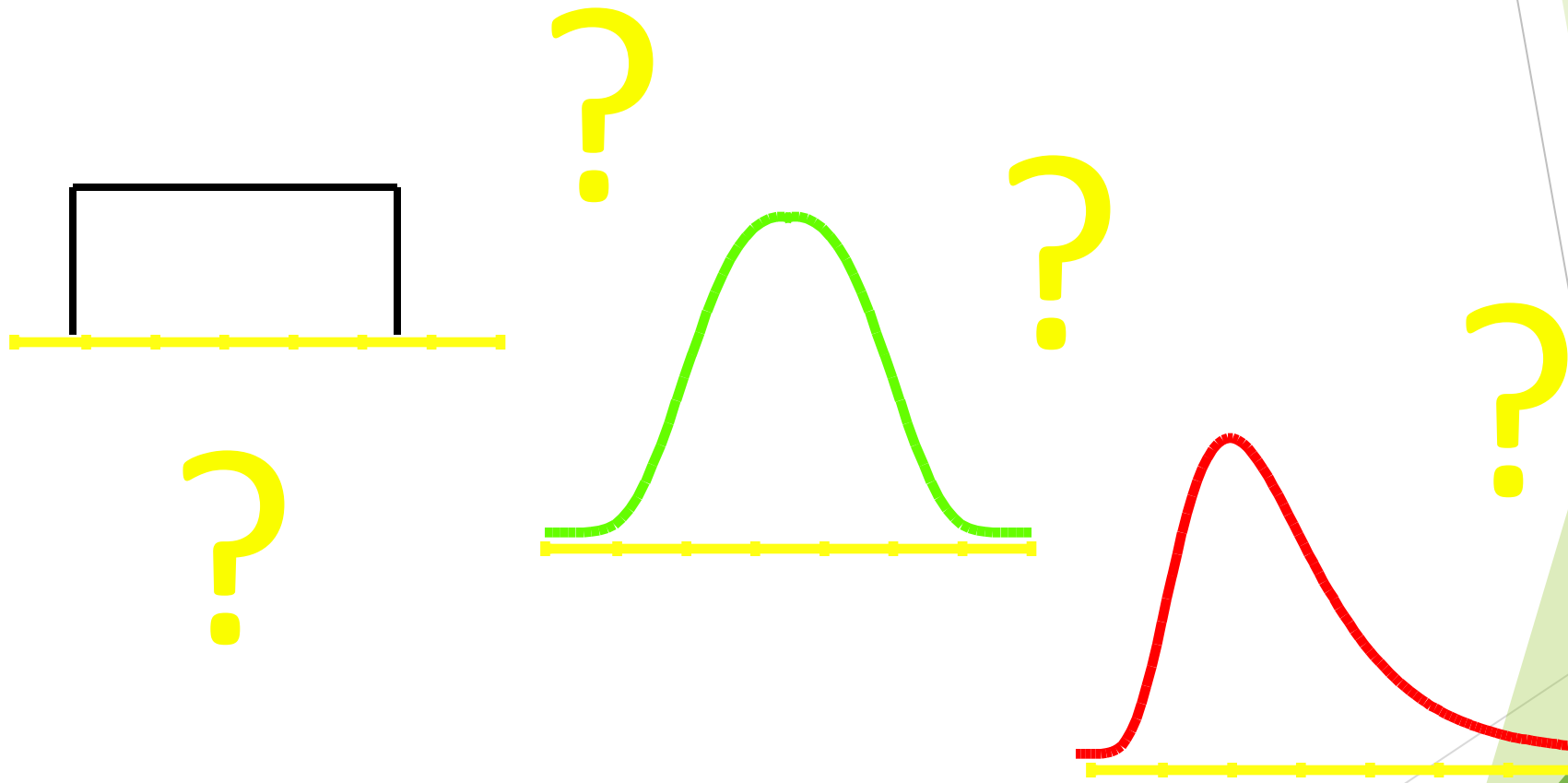


Fordeling av prøvetakings / analyse feil

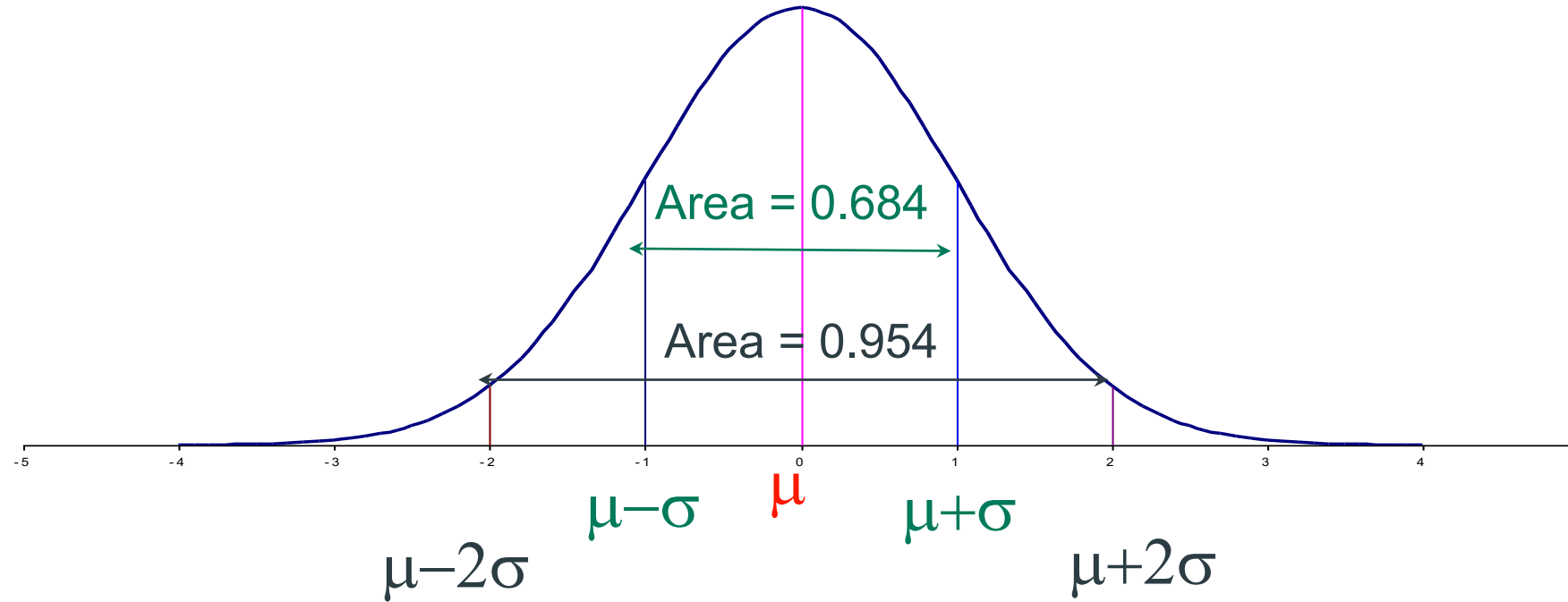
- ▶ Industrial hygiene sampling and analytical error tends to be normally distributed



Sjekk fordelingen



Normalfordeling



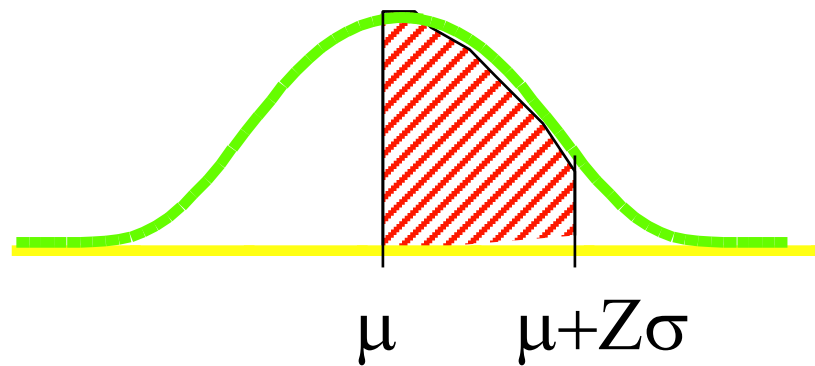
Hvis en kan anta normal fordeling, har den noen kjente egenskaper.

Populasjon vs. Utvalg standardavvik

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}}$$

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

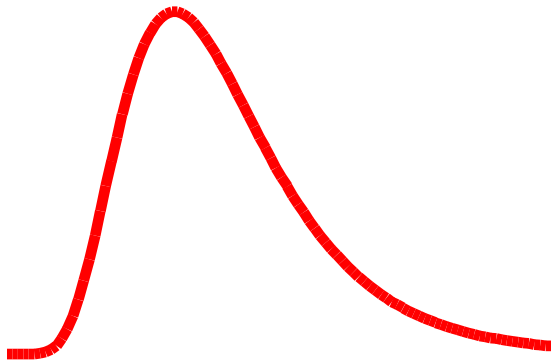
Normal fordeling



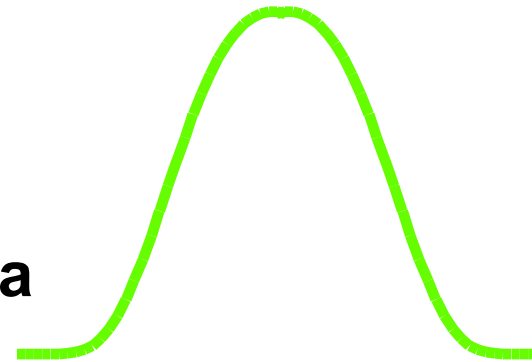
Z	Område under normal kurven fra μ til $\mu + Z\sigma$
1.0	0.34
1.645	0.45
1.96	0.475
2.0	0.477
∞	0.5

Transformering av data

Log-normal



Normal



log av data

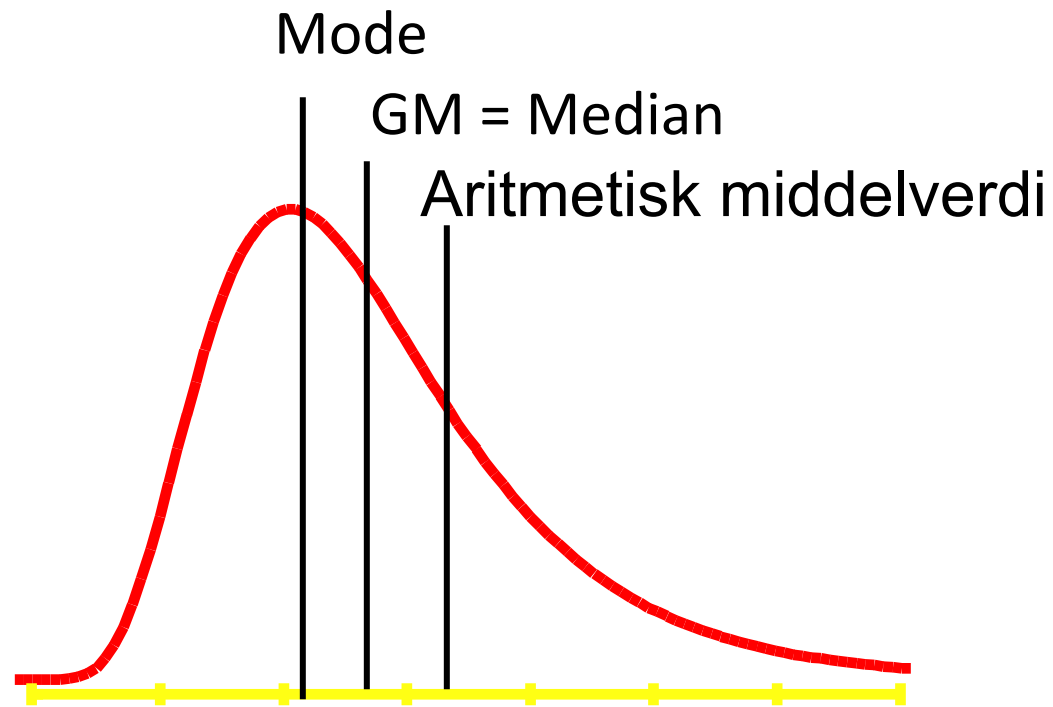


anti-log av data

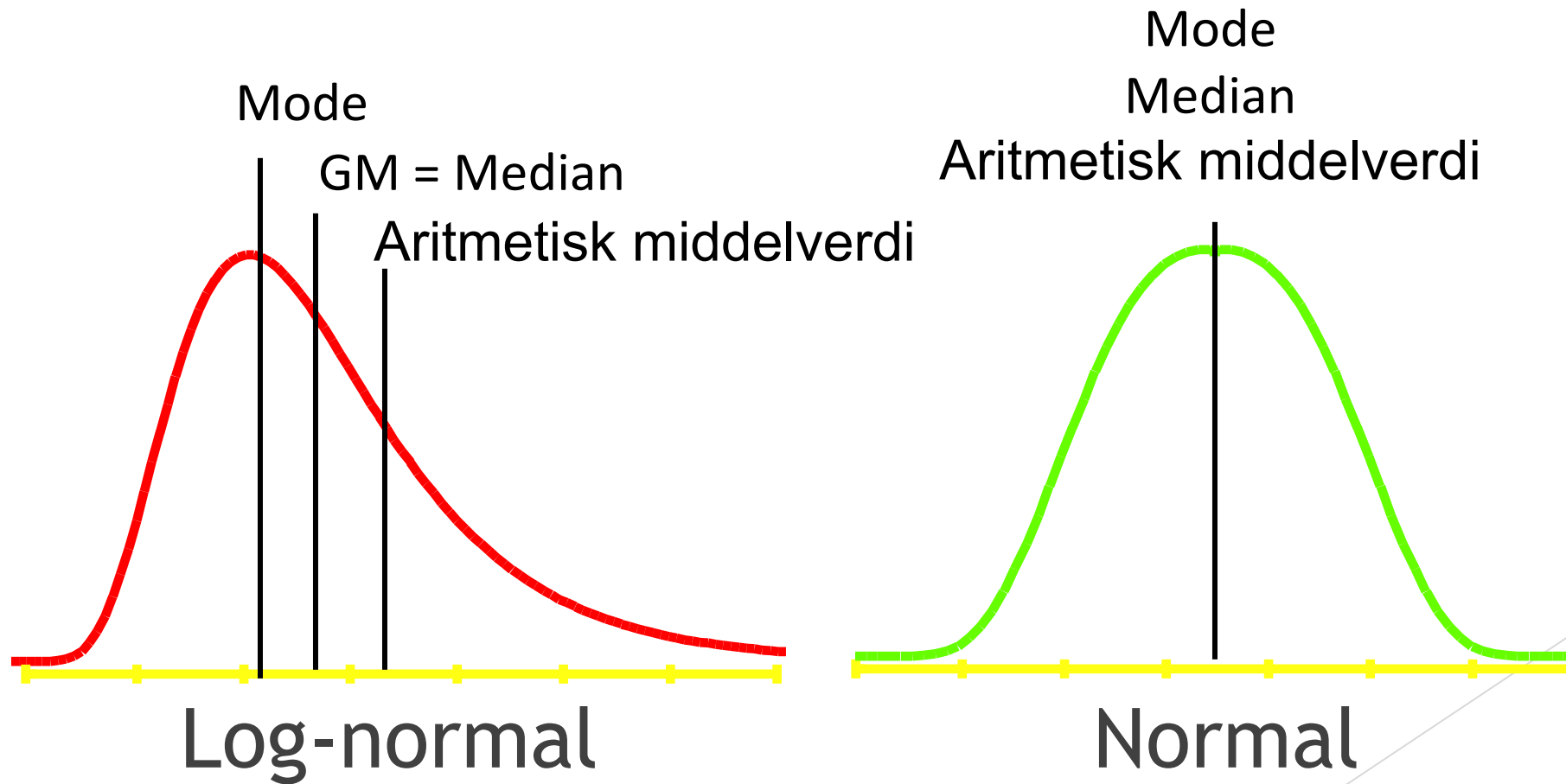


Egenskaper ved en log-normalfordeling

$$\text{Mode} < \text{Median} = \text{GM} < \text{AM}$$



Forskjeller i egenskaper



Eksempel: Utregning av GM and GSD

	X_i	$\ln X_i$	$(\ln X_i - \sum \ln X_i / n)^2$
1	52	3.95	1.85
2	114	4.74	0.32
3	205	5.32	0.0001
4	365	5.90	0.35
5	780	6.66	1.82
sum=			4.34
mean=	303	5.31	

$$GM = e^{(\sum \ln X_i / n)}$$
$$= e^{5.31} = 203.1$$

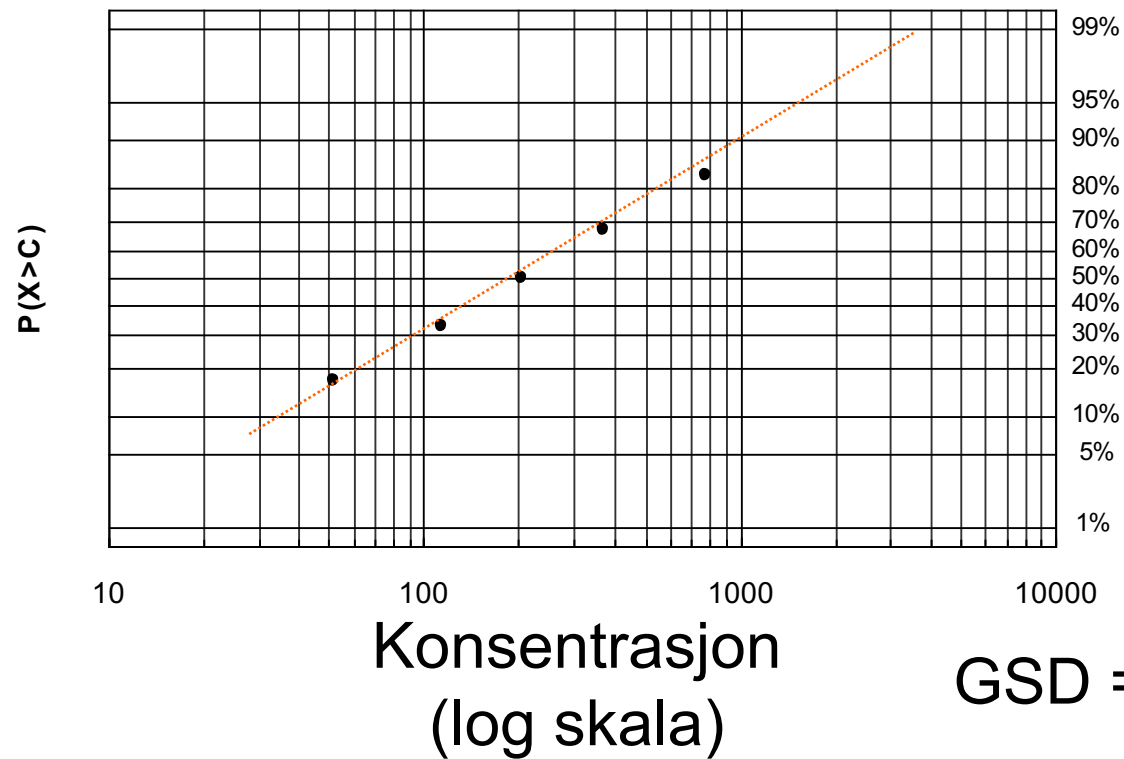
$$S_{\ln x} = \sqrt{\frac{\sum (\ln x_i - \bar{x}_{\ln x})^2}{n-1}}$$
$$= \sqrt{\frac{4.34}{5-1}} = 1.042$$

$$GSD = e^{S_{\ln x}} = e^{1.042} = 2.83$$

Ulike tester for normal/log-normalfordeling

- ▶ Statistisk beregning
 - ▶ W-test (Shapiro Wilks)
 - ▶ Filliben's metode
- ▶ Grafisk metode
 - ▶ Sannsynlighetsplott
 - ▶ Log-sannsynlighetsplott

Bruk av fordelingsplott til å estimere GM og GSD



← 84% punkt

← GM = 50% punkt

← 16% punkt

$GSD = 84\% \text{ punktet} / 50\% \text{ punktet}$
(eller $50\% \text{ punktet} / 16\% \text{ punktet}$)

Log-sannsynlighetsplott

- $GM = 50\%$ punkt = 200
- $GSD = 84\%$ punkt / 50% punkt
= $550 / 200 = 2,75$
- Estimer percentilene fra en linje gjennom punktene gir best tilpasning
- $Mode < GM = Median < Aritmetisk\ middelv\ddot{e}r\ddot{d}i$

Prøvetakingstid vs Eksponeringstid

- ▶ Bare data med samme prøvetakingstid bør analyseres sammen
 - ▶ Beregne tidsveid gjennomsnitt
 - ▶ Hva med rest tid - eksponert / ueksponert
 - ▶ Konservativ vurdering

Beskrive dag-til-dag eksponering og variasjon
i denne

- ▶ Fokus på høyeksponering
- ▶ Fokus på gjennomsnitt

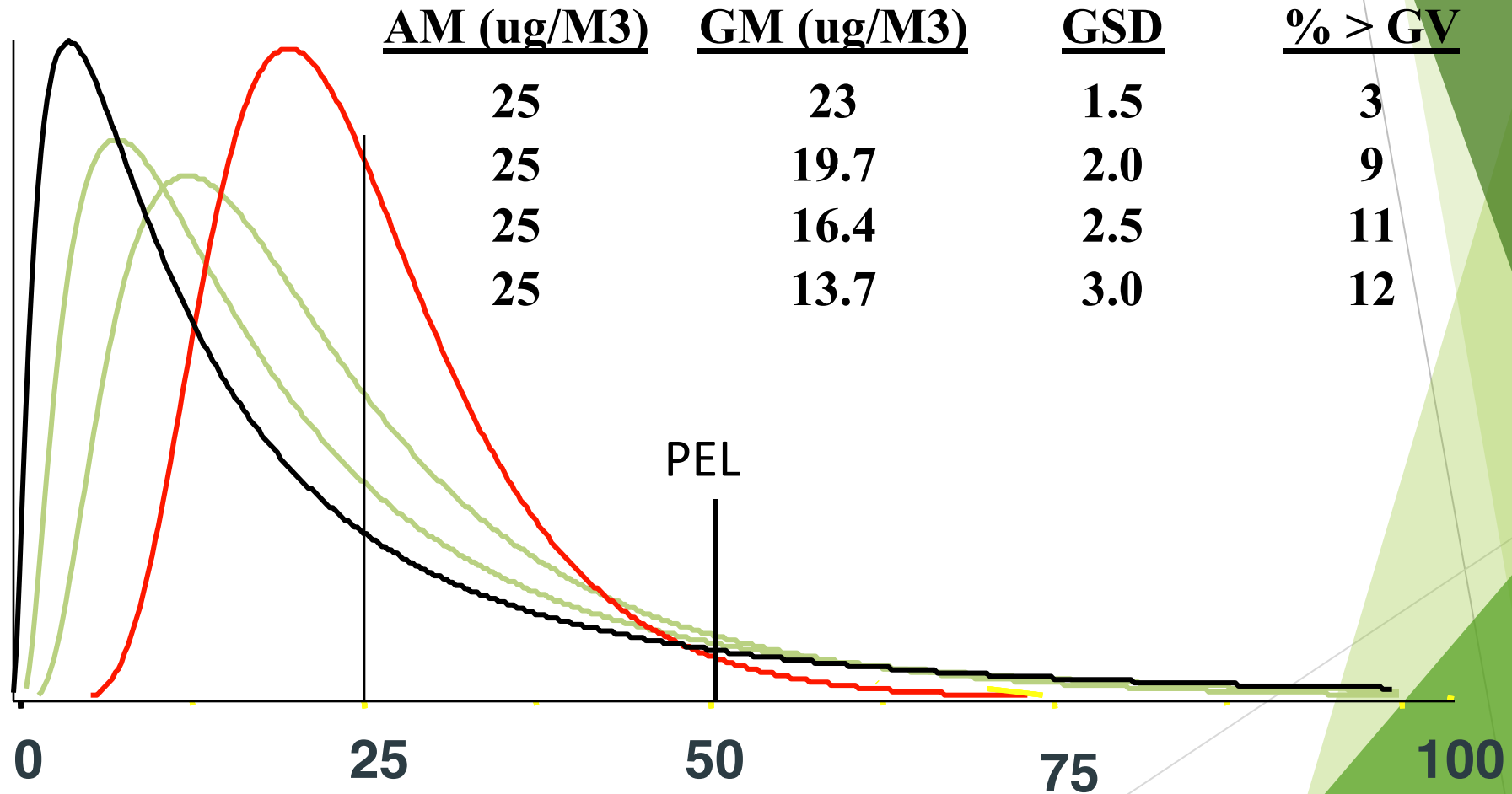
Fire eksponeringsfordelinger

Alle med AM på 25 ug/m³

<u>GM (ug/m³)</u>	<u>GSD</u>	<u>% > GV</u>
23	1.5	3
19.7	2.0	9
16.4	2.5	11
13.7	3.0	12

GV= 50 ug/m³

GM's og GSD's for ulike fordelinger



Percentiler av en log-normalfordeling

$$\text{Percentil} = e^{\left[\bar{x}_{\ln x} + Z s_{\ln x} \right]}$$

eller

$$\text{Percentil} = \text{GM} \cdot \text{GSD}^Z$$

Eksempel - øvre 95%il

$$\text{Percentil} = e^{\left[\bar{x}_{\ln x} + Z s_{\ln x} \right]}$$

$$Z_{0.95} = 1.645$$

$$95\%il = e^{\left[5.31 + 1.645 (1.04) \right]}$$

$$= 1120$$

EMK = Toleransegrense

Tolerance Limits are statistics below which we can have a stated confidence that a stated percentile of the entire population lies.

For a log normal distribution:

$$TL = e^{[\bar{x}_{\ln x} + K \cdot s_{\ln x}]}$$

(K from tolerance limit tables)

Another way of describing the tolerance limits -
the upper X% confidence limit on the Y%ile.

Spesielle temaer

Spesielle temaer


- ▶ Analyse av eksponeringsdata ved bruk av Bayesiansk statistikk
- ▶ Optimalisering av prøvetakingsstrategier (PBEA)
- ▶ Håndtering av målinger under deteksjonsgrensen
- ▶ Estimering av eksponering ved hjelp av matematisk modellering
- ▶ Vurdering av direktevisende målinger
- ▶ Vurdering av langtidseksponering inkl. håndtering av dager med reel null eksponering
- ▶ Vurdering av tiltak

INVITASJON TIL KURS 4.-5. februar + 4.-5. mars 2009
Sted: Erik Blodøks, Tau

Videregående kurs for yrkeshygienikere
Statistikk og prøvetakingsstrategi

Kurset undervises i 2 + 2 dager + hjemmeoppgave.
Bestått eksamen gir 7,5 studiepoeng.

VELKOMMEN TIL KURS
Kurset bygger på boken "A Strategy for Assessing and Managing Occupational Exposures" (AIHA), i tillegg til litteratur knyttet til anvendelse av Bayesiansk statistikk innen yrkeshygiene. Kurset vil bli forelest av Paul Hewett (www.oesh.com).




Kurset er et tilbud til erfarne yrkeshygienikere som ønsker spesialisering for bedre å løse bedriftens utfordringer, kunne delta i utvelgning av regelverk og veiledning for industrien og for å kunne veilede andre yrkeshygienikere, begrenset antall plasser.
Eksamen 6. mars.
Mer informasjon om kurset kan fås ved henvendelse til Kjetil Jacobsen (kj@oehs.no), eller Hans Thore Smedbøld (hts@oehs.no).




KURSPROGRAM
Foreleser: Paul Hewett
Arrangør for kurset: Hans Thore Smedbøld (Univ. lektor II NTNU)

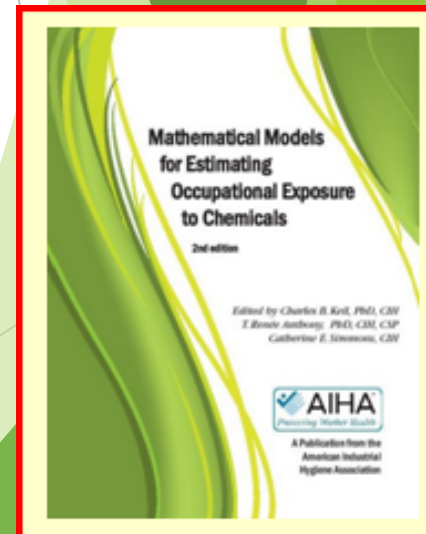
FAGLIG INNHOLD

1. Statistikk for yrkeshygienikere – del I: Grunnleggende
2. Statistikk for yrkeshygienikere – del II: Videregående
3. Vurderingsstrategier for eksponering
4. Utfallsbasert strategi for eksponeringsvurdering
5. Bayesiansk statistikk – del I: Grunnleggende
6. Bayesiansk statistikk – del II: Anvendelse
7. Rapportering av yrkeshygiene data

PRAKTISKE OPPLYSNINGER
Tid: 4.-5.2 + 4.-5.3-09 – start kl. 09.00
Sted: Erik Blodøks, Tau
Internett: <http://www.lilland.no/indexenik.htm>
Påmelding: til K. Jacobsen (kj@oehs.no), 402 49 6311 innen 1.01.09
Foreleser: Paul Hewett
Språk: engelsk med eksamen på norsk
Pris: kursavgift kr 6.000 – inklusive materiell, semesteravgift og kost og losj (fullpensjon)
Arrangør: Pol, Of og NTNU
Teknisk arrangør: OHS

ARRANGØRER:  NTNU
Det nasjonale universitet

VELKOMMEN!   



Målefeil vs variasjon

- ▶ Analytical

- ▶ Random errors in sampling and analysis

- ▶ Environmental

- ▶ Variability in exposures . . .

- ▶ From sampling period to sampling period

- ▶ From worker to worker within one period

- ▶ Random variation in contaminant generation, mixing, removal, worker mobility

Sampling and Analytical Variability: CV or Relative Standard Deviation

$$CV = S_r = \frac{\sigma}{\mu}$$

Estimate of the relative standard deviation, equal to S divided by the mean of a series of measurements. A measure of precision. Previously called CV (coefficient of variation).

Some CV's from NIOSH Methods

<u>Substance</u>	<u>CV</u>
Asbestos (PCM)	0.115 to 0.13
Asbestos (TCM)	0.11
Hydrocarbons (bp 36-126)	0.05 to 0.07
Hydrocarbons (aromatic)	0.05 to 0.09
Elements (ICP)	0.01 to 0.09
Naphthas	0.05
PNOC's, total	0.056
Silica, respirable (XRD)	0.09

NIOSH Manual of Analytical Methods, 4th ed.

Percent Contribution of Analytical Variability to Total Variability

