SAFOSO

Animal Health Matters.
For Safe Food Solutions.

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs,
Education and Research EAER
State Secretariat for Economic Affairs SECO

Measures of disease spread

Marco De Nardi

Objectives

1. Describe the following measures of spread: range, interquartile range, variance, and standard deviation

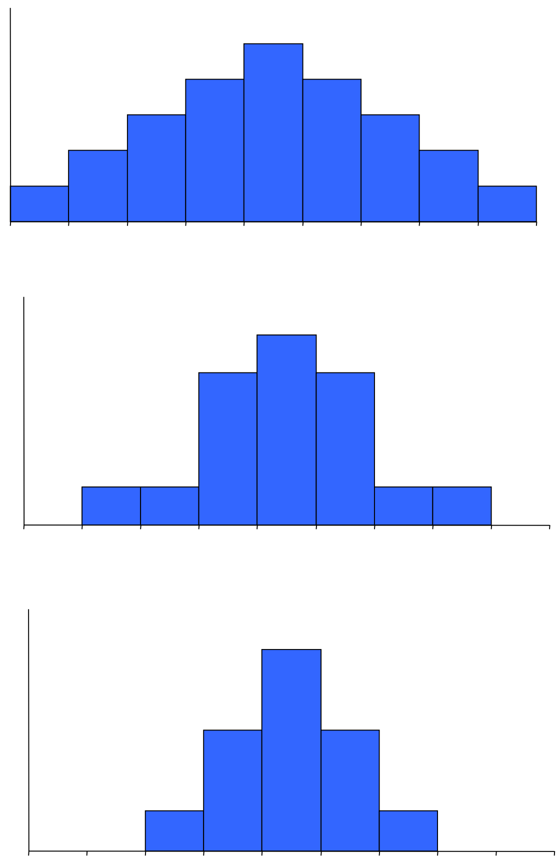
2. Discuss examples using measures of spread in epidemiology

Objective

Describe the following measures of spread: range, interquartile range, variance, and standard deviation

Understanding “Spread”

Age in 3 student classes (2nd primary)

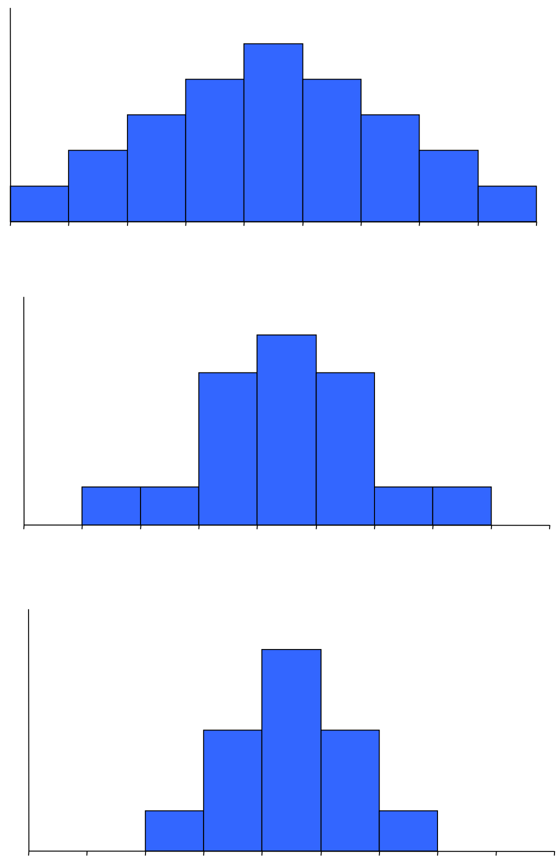


Same center

but …

different spread

Understanding "Spread"



Age in 3 student classes (2nd primary)

Frequency distribution is a representation, either in a graphical or tabular format, that indicate the way data are distributed over a range of values, by showing the number and percentage of individuals within wach group of values.

In this case they all have high frequencies in the centre of the distribution and low frequencies at the 2 extremes (**upper** and **lower tails** of the distribution)

Measures of Spread

Definition: **Measures that quantify the amount of variation, or dispersion, in a dataset**

Also known as:

- "Measure of dispersion"
- "Measure of variation"

Common measures

- **Range**
- **Interquartile range**
- **Variance**
- **Standard deviation**

Range

Definition: difference between largest and smallest values

- In **statistics**, the range is reported as a single number and is the result of subtracting the maximum from the minimum value

- In **epidemiology**, the range is often the largest and smallest observations in the sample, rather than the difference between the largest and smallest

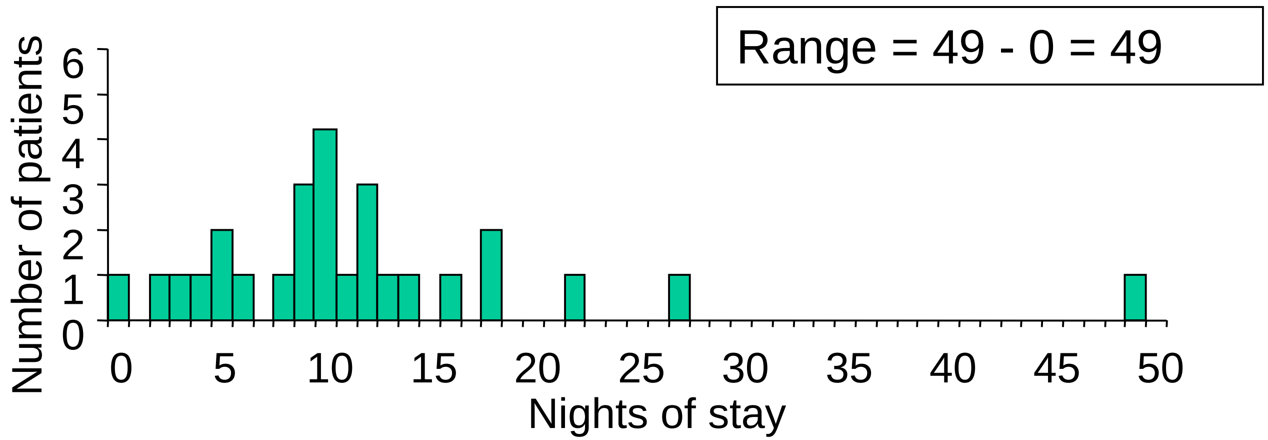
Finding the Range of Length of Stay Data

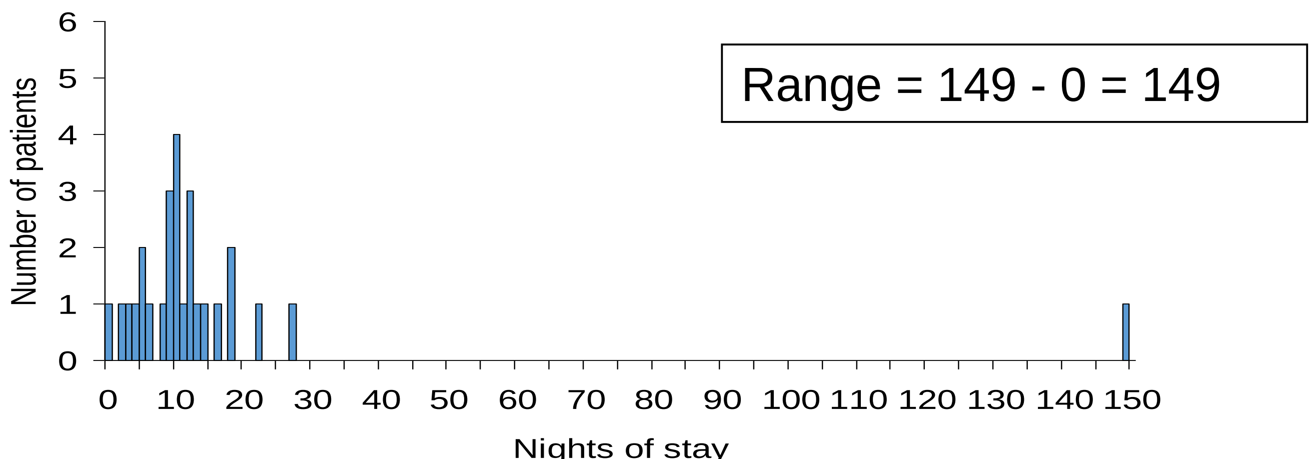
Number of nights spent in the hospital for 27 patients following infection from Staphilococcus spp…

0, 2, 3, 4, 5, 5, 6, 8, 9,
9, 9, 10, 10, 10, 10, 11, 12, 12,
12, 13, 14, 16, 18, 18, 22, 27, 49

What is the range of values?

Is the Range Sensitive to Outliers?





Interquartile Range (IQR)

Definition: **The difference between the first and third "quartiles"** *(defined on next slide)*

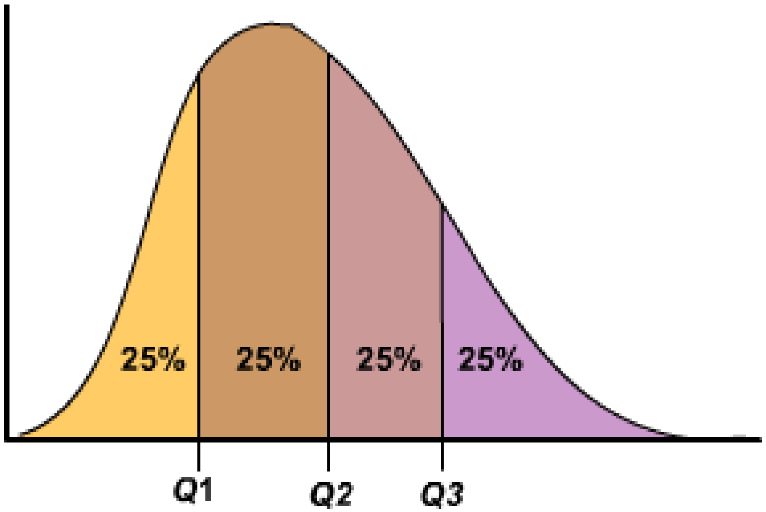
Properties / Uses:

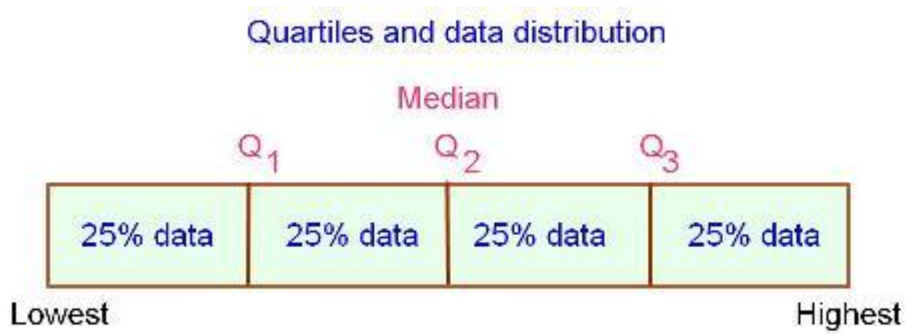
- A modification of the range that is less sensitive to outliers
- Calculated as the difference between two data values, but not the two most extreme

Quartiles

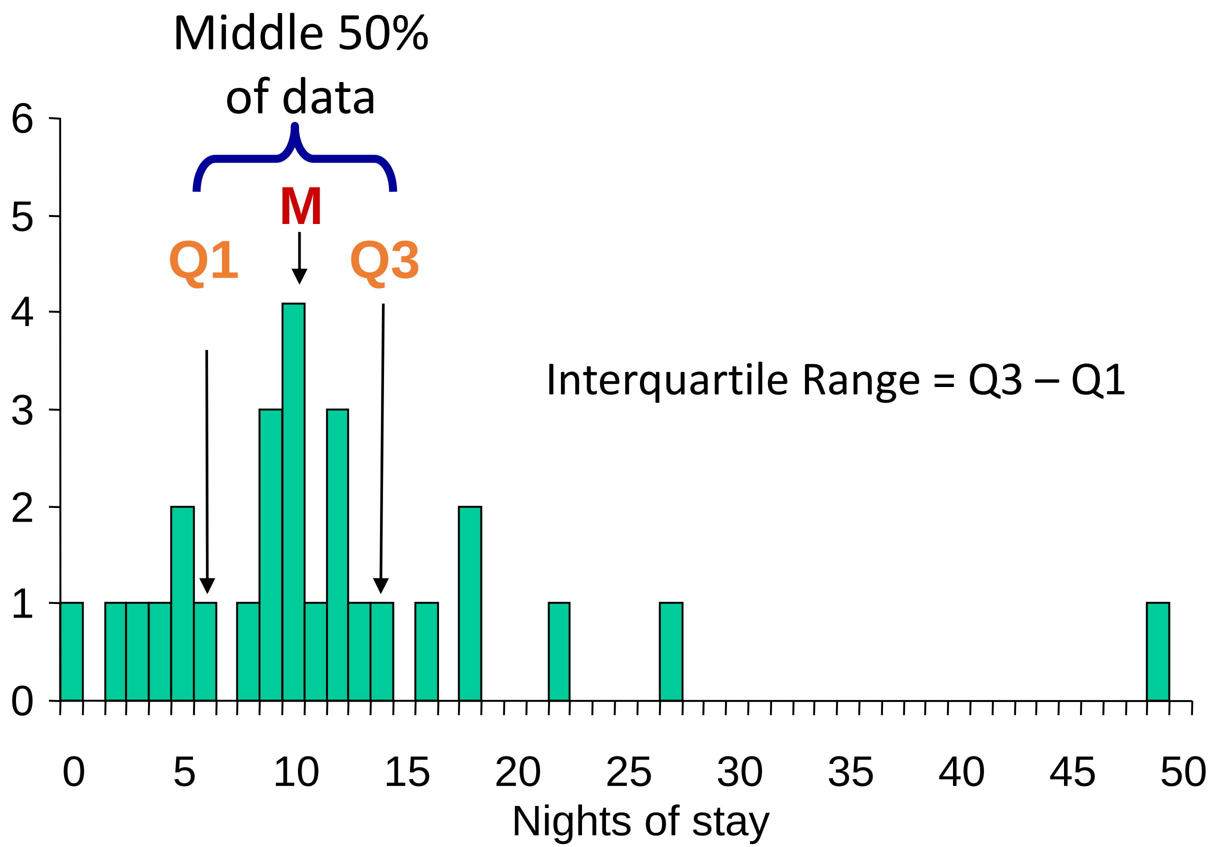
- Definition: **Quartiles are the values that split the data into four equal parts**

- 25% of observations are below the first quartile (Q1)
- 25% of observations are between Q1 and Q2 (median)
- 25% of observations are between Q2 (median) and Q3
- 25% of observations are above Q3


25%
25%
25%
25%
Q1
Q2
Q3


Quartiles and data distribution
Median
Q1
Q2
Q3
25% data
25% data
25% data
25% data
Lowest
Highest

Interquartile Range — Length of Stay Data



Interquartile Range = Q3 – Q1

Calculating the Quartile Positions in a Dataset with N Observations

Position of Q1: $1*(n+1) / 4$

Position of Q2 (Median) : $2*(n+1) / 4$

Position of Q3: $3*(n+1) / 4$

Example: In a dataset with 55 observations, the position of the third quartile (Q3) would be:

$3 * (55+1) / 4 = 168 / 4 = 42$

So the third quartile is the **value** in the **42nd position** in the ordered list

Determining IQR in Length of Stay Data

Number of nights in the hospital for 27 patients

Q1

0, 2, 3, 4, 5, 5, 6, 8, 9,

Q2

9, 9, 10, 10, 10, 10, 11, 12, 12,

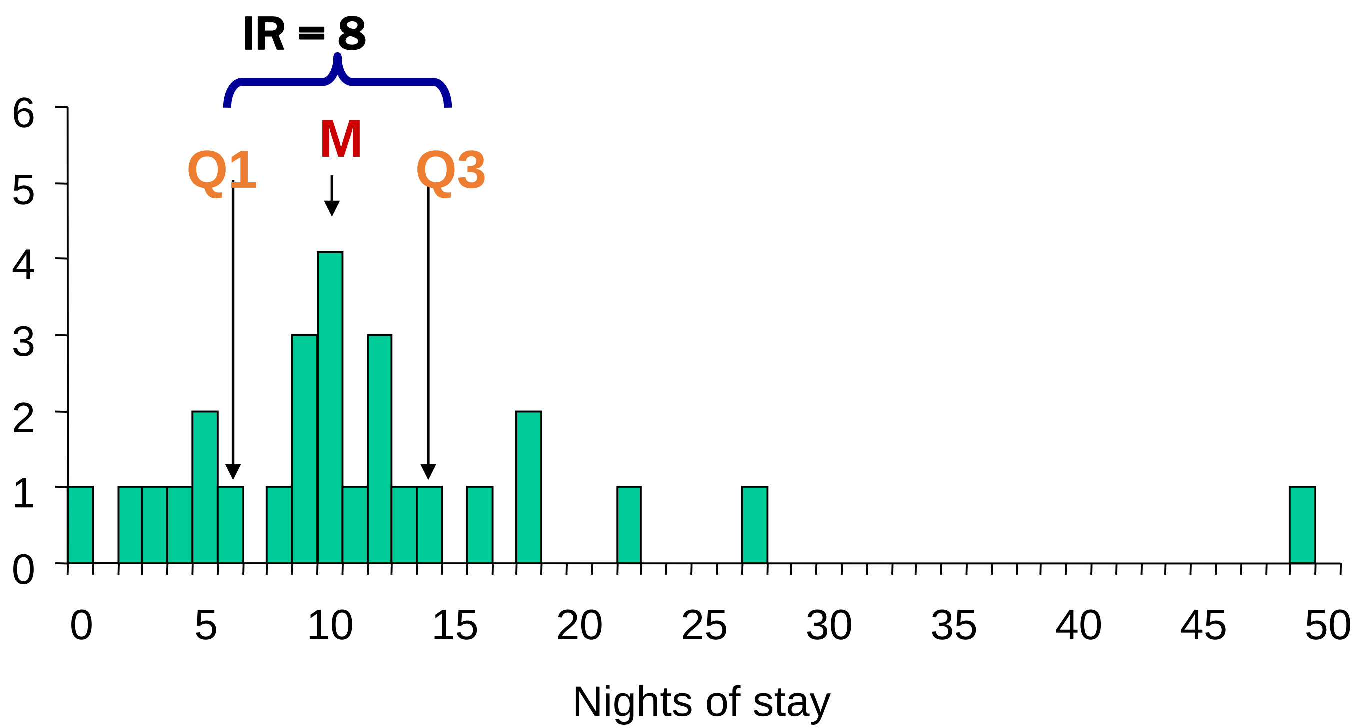
12, 13, 14, 16, 18, 18, 22, 27, 49

Q3

	Position		Value
Q1:	1*(27+1) / 4 = 7th	→	6
Q2 (Median):	2*(27+1) / 4 = 14th	→	10
Q3:	3*(27+1) / 4 = 21st	→	14

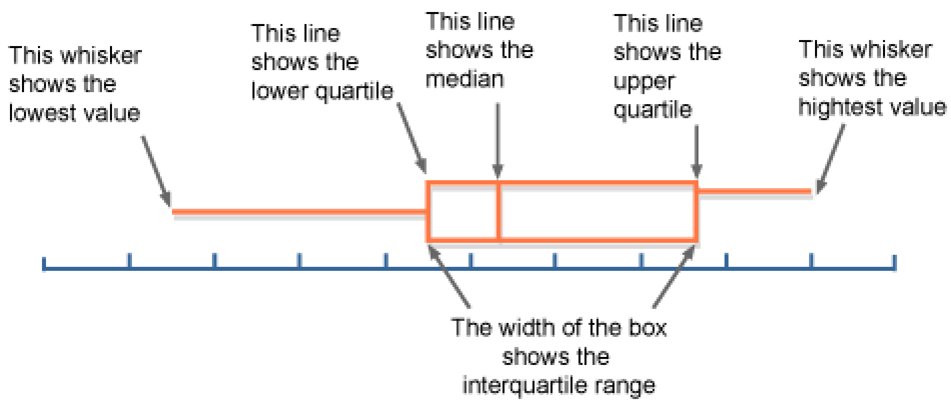
Interquartile range: 14 - 6 = 8

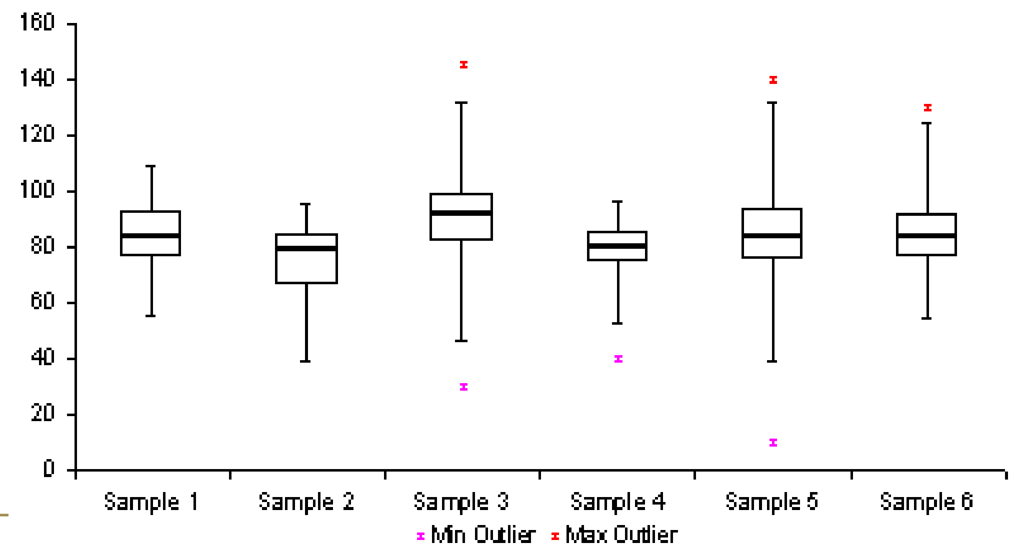
Interquartile Range — Length of Stay Data



How to represent quartiles and interquartiles?

Box and whiskers plot





Variance and Standard Deviation

General Description: deviations of each sample observation from the mean.

Variance

- Average of squared deviations from mean
- Sum $(x_i - \text{mean})^2 / (n-1)$

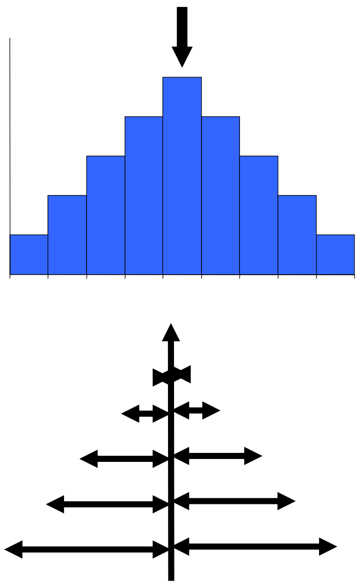
Standard deviation

- Square root of variance

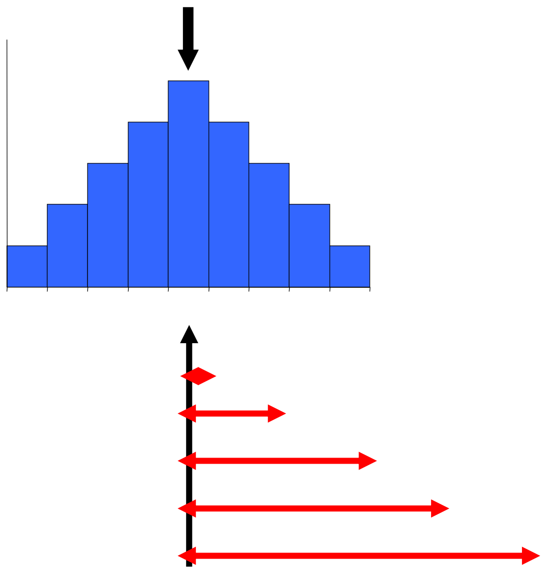
The variance and standard deviation are related measures that quantify how closely clustered the observed values are to the mean.

Variance and Standard Deviation

Mean



Mean



When describing how far each observation is from the mean, some values will be positive and some will be negative

<u>Variance and Standard Deviation</u> describe this dispersion of data by squaring the values to eliminate negative numbers

Equations for Variance and Standard Deviation

- x : mean
- x_i : value
- n : number of data points
- s^2: variance
- s : standard deviation

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

$$s = \sqrt{\frac{\sum (x_i \quad x)^2}{(n-1)}}$$

Standard Deviation – Properties / Uses

When data are approximately normally distributed (bell-shaped curve), then:

- 68.3% of the data fall within 1 SD of the mean
- 95.5% of the data fall within 2 SD of the mean
- 95.0% of the data fall within 1.96 SD of the mean
- 99.7% of the data fall within 3 SD of the mean

Length of Stay Data

Number of nights spent in the hospital for 27 patients

$(0 - 12)^2 = 144$	$(9 - 12)^2 = 9$	$(12 - 12)^2 = 0$
$(2 - 12)^2 = 100$	$(9 - 12)^2 = 9$	$(13 - 12)^2 = 1$
$(3 - 12)^2 = 81$	$(10 - 12)^2 = 4$	$(14 - 12)^2 = 4$
$(4 - 12)^2 = 64$	$(10 - 12)^2 = 4$	$(16 - 12)^2 = 16$
$(5 - 12)^2 = 49$	$(10 - 12)^2 = 4$	$(18 - 12)^2 = 36$
$(5 - 12)^2 = 49$	$(10 - 12)^2 = 4$	$(18 - 12)^2 = 36$
$(6 - 12)^2 = 36$	$(11 - 12)^2 = 1$	$(22 - 12)^2 = 100$
$(8 - 12)^2 = 16$	$(12 - 12)^2 = 0$	$(27 - 12)^2 = 225$
$(9 - 12)^2 = 9$	$(12 - 12)^2 = 0$	$(49 - 12)^2 = 1369$

Sum = 2370; **Var** = 2370/26 = 91.15; **SD** = $\sqrt{91.15}$ = **9.55**

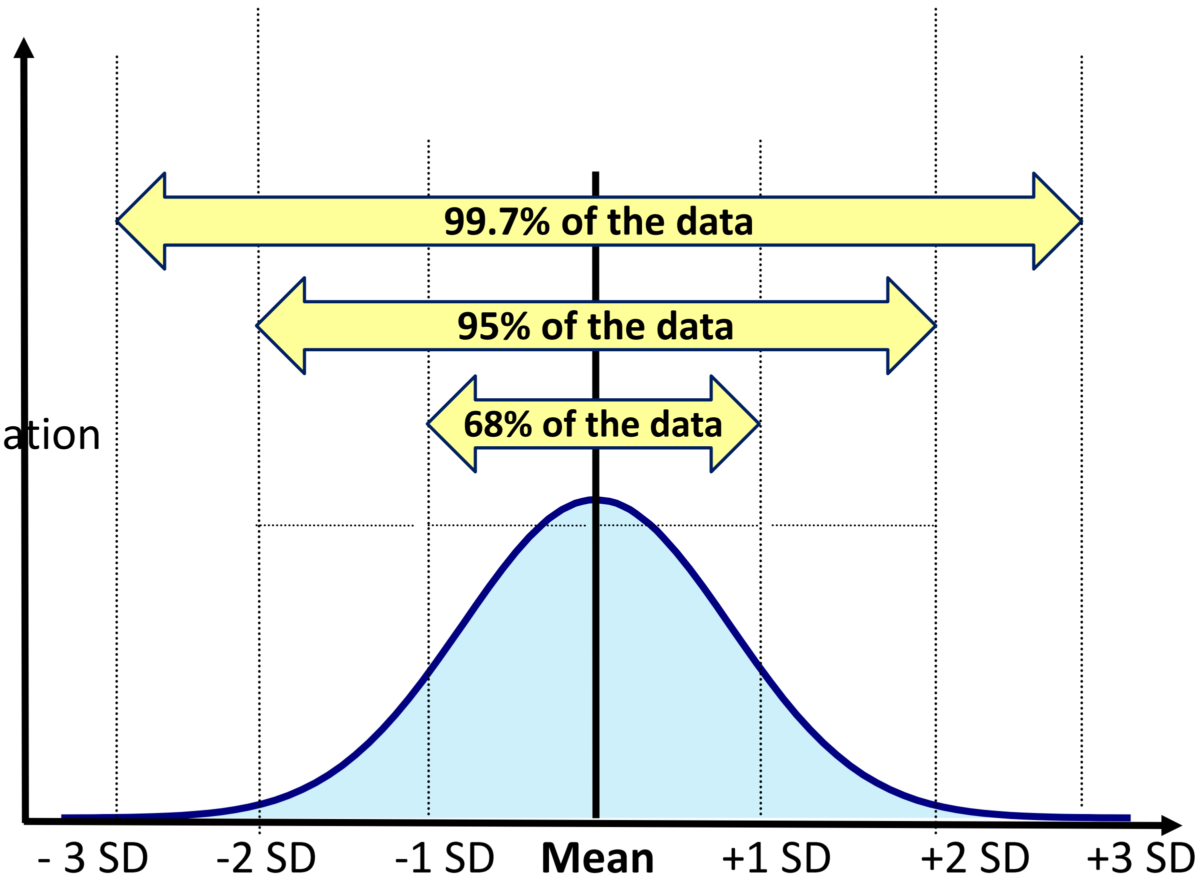
Standard Deviation – Properties / Uses

When data are approximately normally distributed (bell-shaped curve), then:

- 68.3% of the data fall within 1 SD of the mean
- 95.5% of the data fall within 2 SD of the mean
- 95.0% of the data fall within 1.96 SD of the mean
- 99.7% of the data fall within 3 SD of the mean

Normal Distribution

SD = Standard deviation



Objective 2

Discuss examples using measures of spread in epidemiology

Example: Using the Range in Epidemiology

The range is commonly used when reporting:

- Incubation period
- Duration of illness
- Age of cases

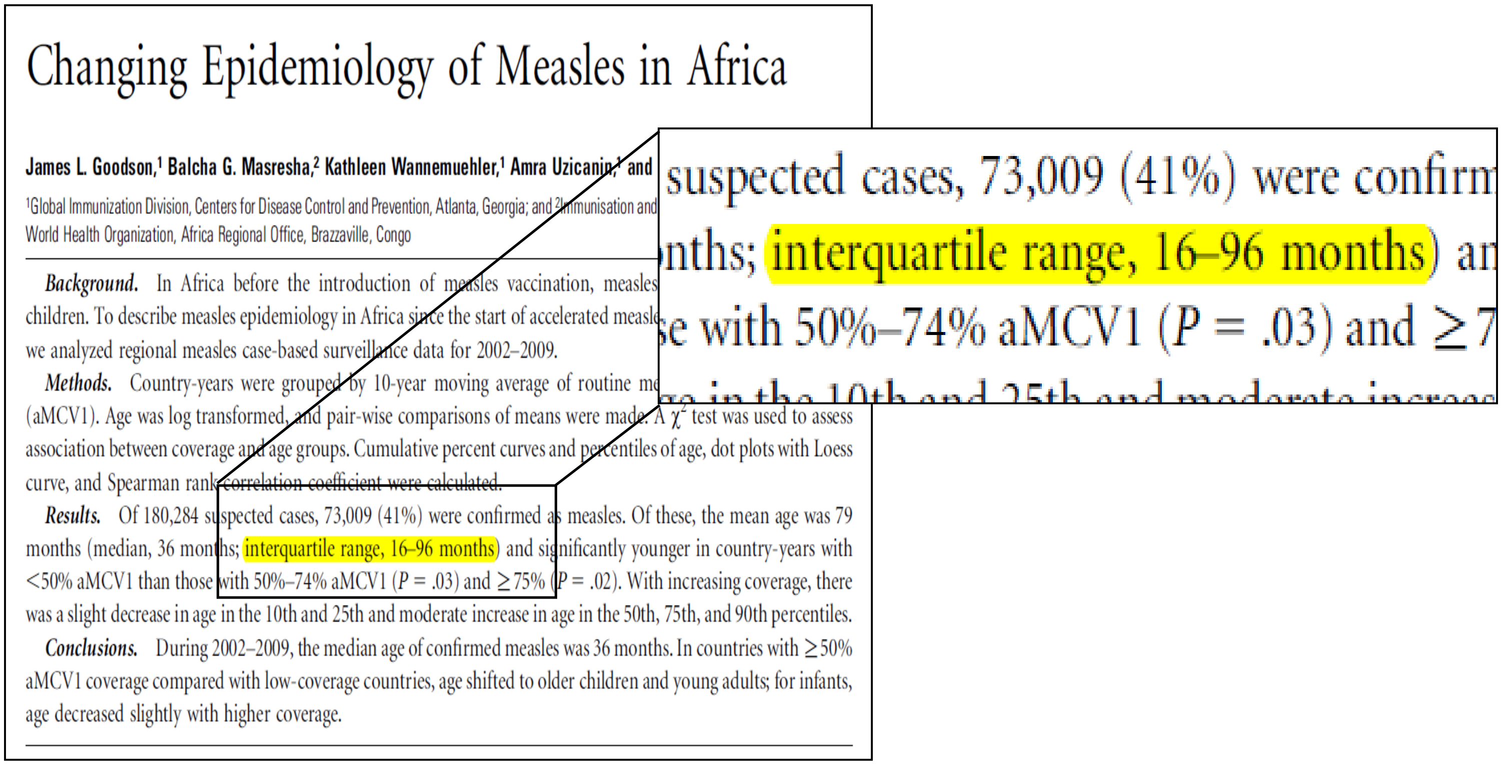
Summary of characteristics of selected biological warfare agents

Agent	Incubation period (days)	Person to person transmission	Morbidity/mortality if untreated	Diagnosis
Ba. anthracis	1–5	No	High/high	Culture, serology, ELISA, PCR
Y. pestis	2–3	Yes	High/high	Culture, serology, ELISA, PCR
Brucella sp.	5–60	No	High/low	Culture, serology, ELISA, PCR
Bu. mallei	3–7	No	High/low	Culture, serology, ELISA, PCR
Bu. pseudomallei	3–7	No	High/low	Culture, serology, ELISA, PCR
Botulinum toxin	1–5	No	High/high	ELISA, mouse inoculation for toxin detection
Variola virus	7–17	Yes	High/high	Serology, virus isolation, ELISA, PCR
Viral hemorrhagic fever	4–21	Yes	High/high	Serology, virus isolation, ELISA, PCR

Source:Thavaselvam and Vijayaraghavan. J Pharm Bioallied Sci. 2010 Jul-Sep; 2(3): 179–188

Example: Using the IQR in Epidemiology

The interquartile range (IQR) is reported when the presence of outliers can skew other measures



Changing Epidemiology of Measles in Africa

James L. Goodson,[1] Balcha G. Masresha,[2] Kathleen Wannemuehler,[1] Amra Uzicanin,[1] and

[1]Global Immunization Division, Centers for Disease Control and Prevention, Atlanta, Georgia; and [2]Immunisation and
World Health Organization, Africa Regional Office, Brazzaville, Congo

Background. In Africa before the introduction of measles vaccination, measles
children. To describe measles epidemiology in Africa since the start of accelerated measle
we analyzed regional measles case-based surveillance data for 2002–2009.

Methods. Country-years were grouped by 10-year moving average of routine me
(aMCV1). Age was log transformed, and pair-wise comparisons of means were made. A χ^2 test was used to assess association between coverage and age groups. Cumulative percent curves and percentiles of age, dot plots with Loess curve, and Spearman rank correlation coefficient were calculated.

Results. Of 180,284 suspected cases, 73,009 (41%) were confirmed as measles. Of these, the mean age was 79 months (median, 36 months; interquartile range, 16–96 months) and significantly younger in country-years with <50% aMCV1 than those with 50%–74% aMCV1 ($P = .03$) and ≥75% ($P = .02$). With increasing coverage, there was a slight decrease in age in the 10th and 25th and moderate increase in age in the 50th, 75th, and 90th percentiles.

Conclusions. During 2002–2009, the median age of confirmed measles was 36 months. In countries with ≥50% aMCV1 coverage compared with low-coverage countries, age shifted to older children and young adults; for infants, age decreased slightly with higher coverage.

suspected cases, 73,009 (41%) were confirm
nths; interquartile range, 16–96 months) ar
se with 50%–74% aMCV1 (P = .03) and ≥ 7

Example: Using the Standard Deviation in Epidemiology

The standard deviation is often reported along with the mean to show variability in the data

Table 1. Body composition and severity of disease among HIV positive and HIV negative individuals with/without tuberculosis

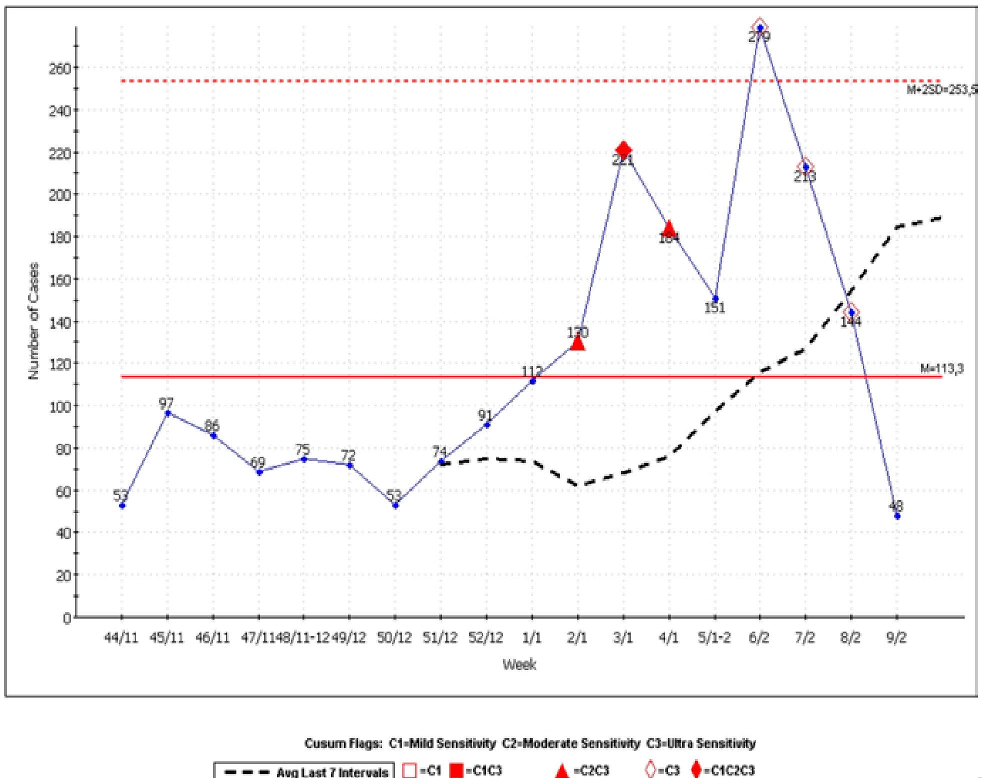
Characteristic [mean, (SD)][c]	HIV positive with TB (n=31)		HIV negative with TB (n=32)	
	Men (n=10)	Women (n=21)	Men (n=18)	Women (n=14)
Age in yrs	30.9 (4.6)	29.2 (5.9)	26.0 (7.3)	26.3 (4.6)
BMI kg/m^2	18.4 (1.7)	18.6 (3.0)	18.2 (2.0)	20.3 (4.3)
LMI in kg/m^2	16.6 (1.3)	15.4 (0.9)[b]	16.6 (1.5)	16.1 (0.9)
FMI in kg/m2	1.8 (0.6)	3.3 (2.1)	1.6 (0.8)	4.6 (3.6)[a]
Severity TBscore	6.5 (1.5)	5.6 (2.6)	6.8 (2.3)	5.9 (2.4)

Source: Mupere et al. BMC Public Health 2012, 12:1050

Example: Using the Standard Deviation in Epidemiology

The standard deviation helps detect unusual increases in disease and can be used to establish alert thresholds

Post-Flooding Diarrhea Outbreak in Jakarta, Indonesia, Jan–Feb 2002



← 2 SD

← Mean

Source: Chretien et al, PLoSsmedicine, Mar 2008 Vol 5(3).

SAFOSO

Animal Health Matters.
For Safe Food Solutions.

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs,
Education and Research EAER
State Secretariat for Economic Affairs SECO

Thanks