MORU Tropical Health Network

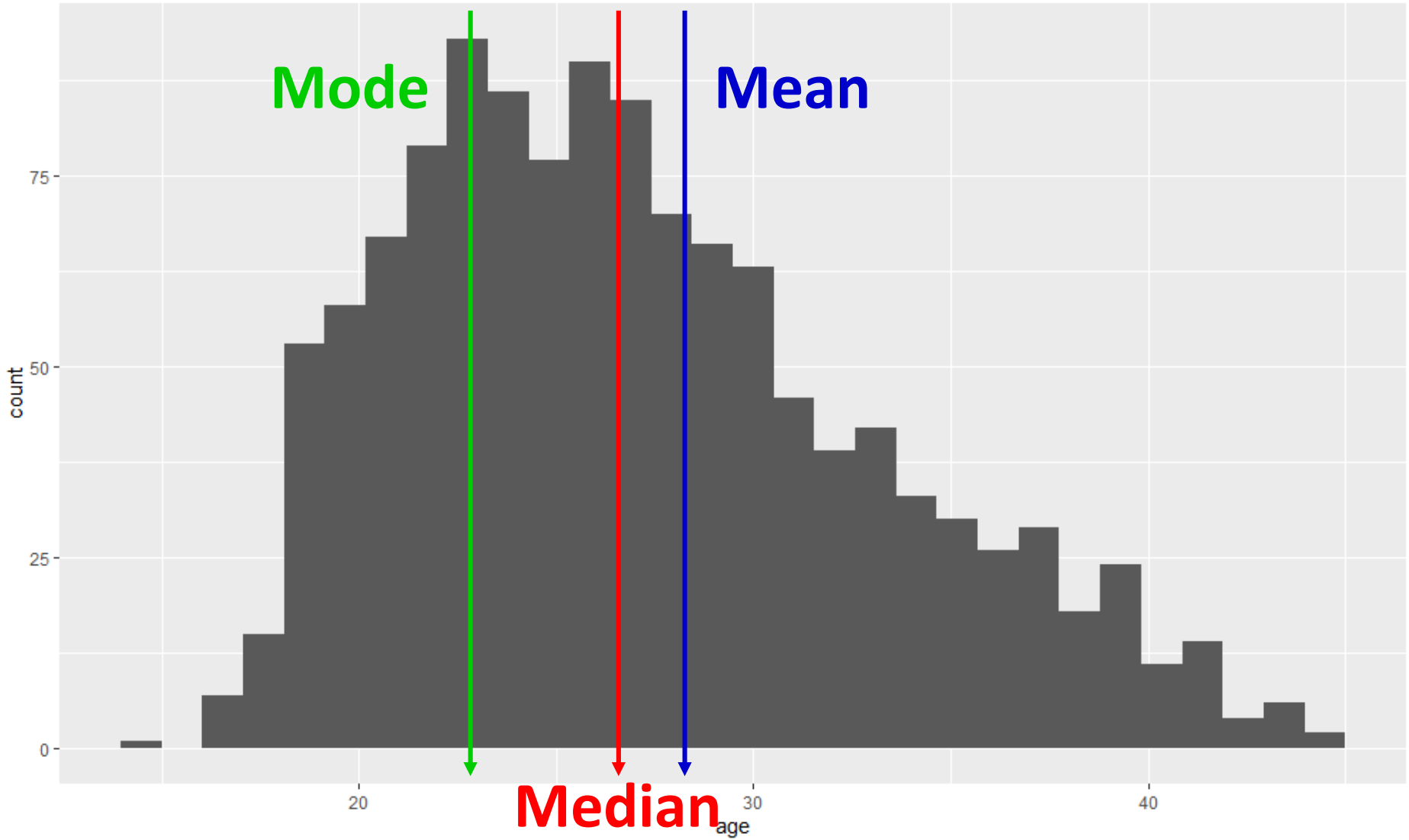
Basic R programing

11-12 Jan 2021

Lecture 2 (13:00-14:30): Descriptive data analysis

Dr. Palang Chotsiri
palang@tropmedres.ac

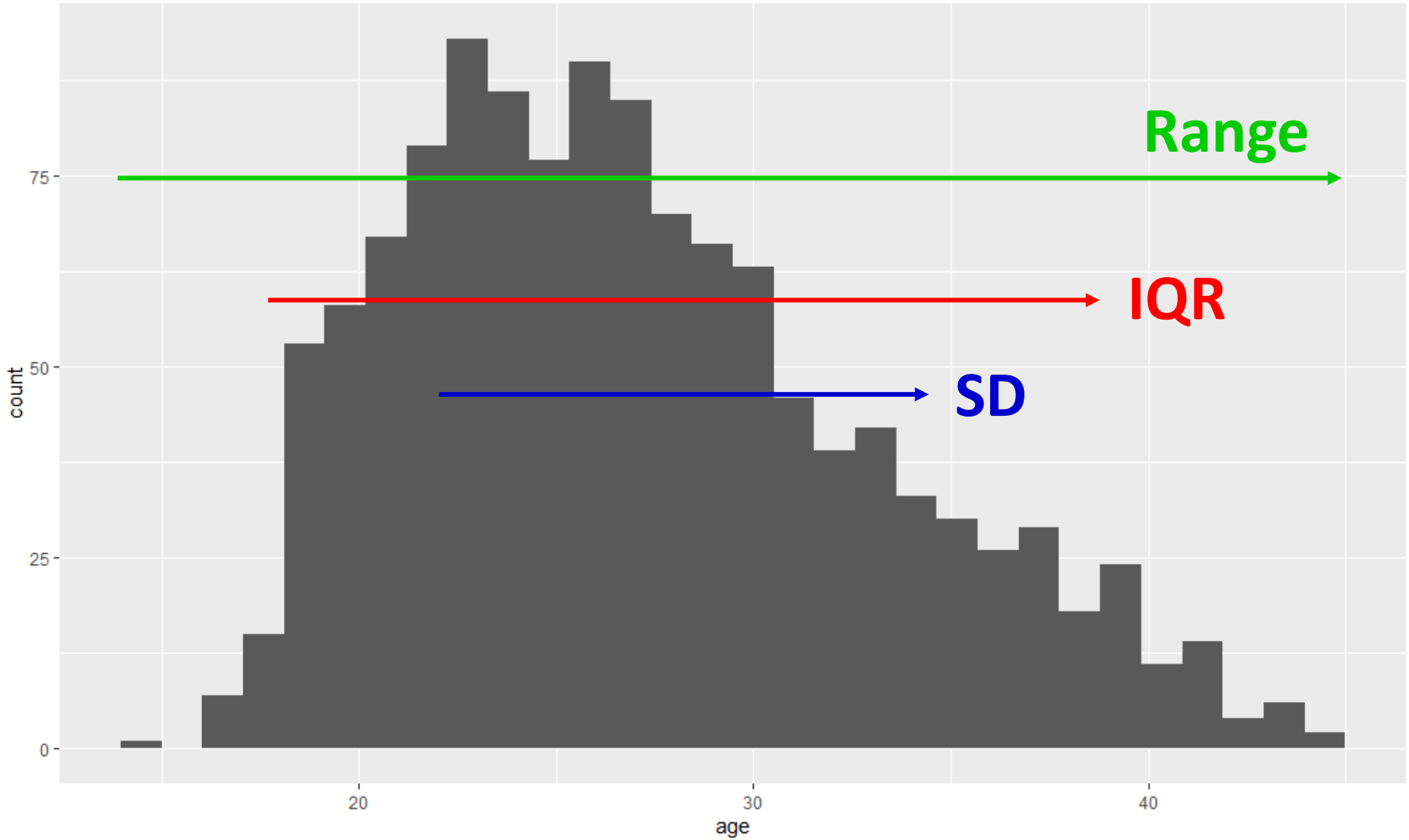
Central tendency



Function to calculate Modes

```
>Modes <- function(x) {
>  ux <- unique(x)
>  tab <- tabulate(match(x, ux))
>  ux[tab == max(tab)]
>}
```

Measure of spread



Measure of spread

- Range

```
> max(age, na.rm = TRUE) - min(age, na.rm = TRUE)
```

- Inter Quantile Range (IQR)

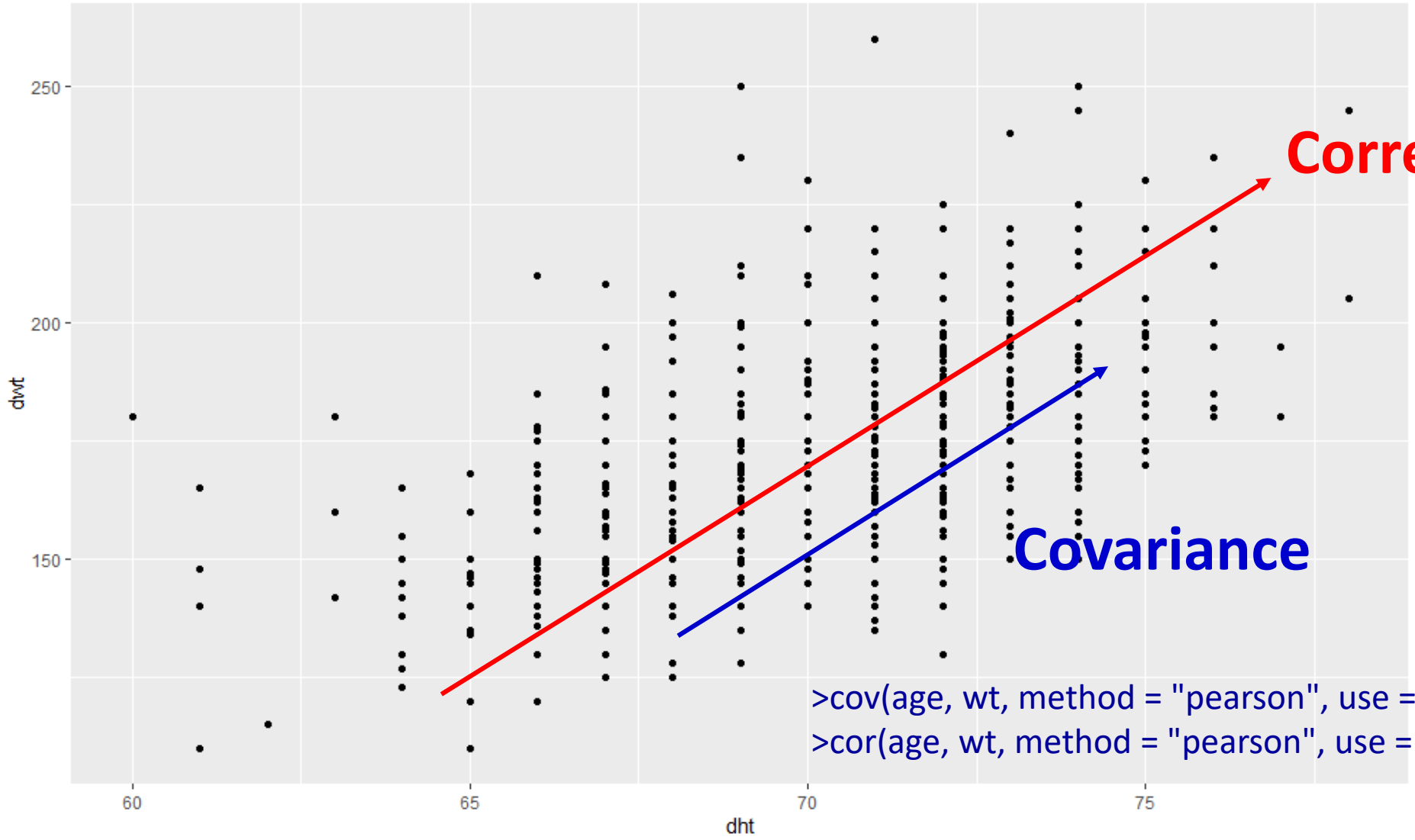
```
> Q3 <- quantile(age, 0.75, na.rm = TRUE)
> Q1 <- quantile(age, 0.25, na.rm = TRUE)
> Q3-Q1

> IQR(age, na.rm = TRUE)
```

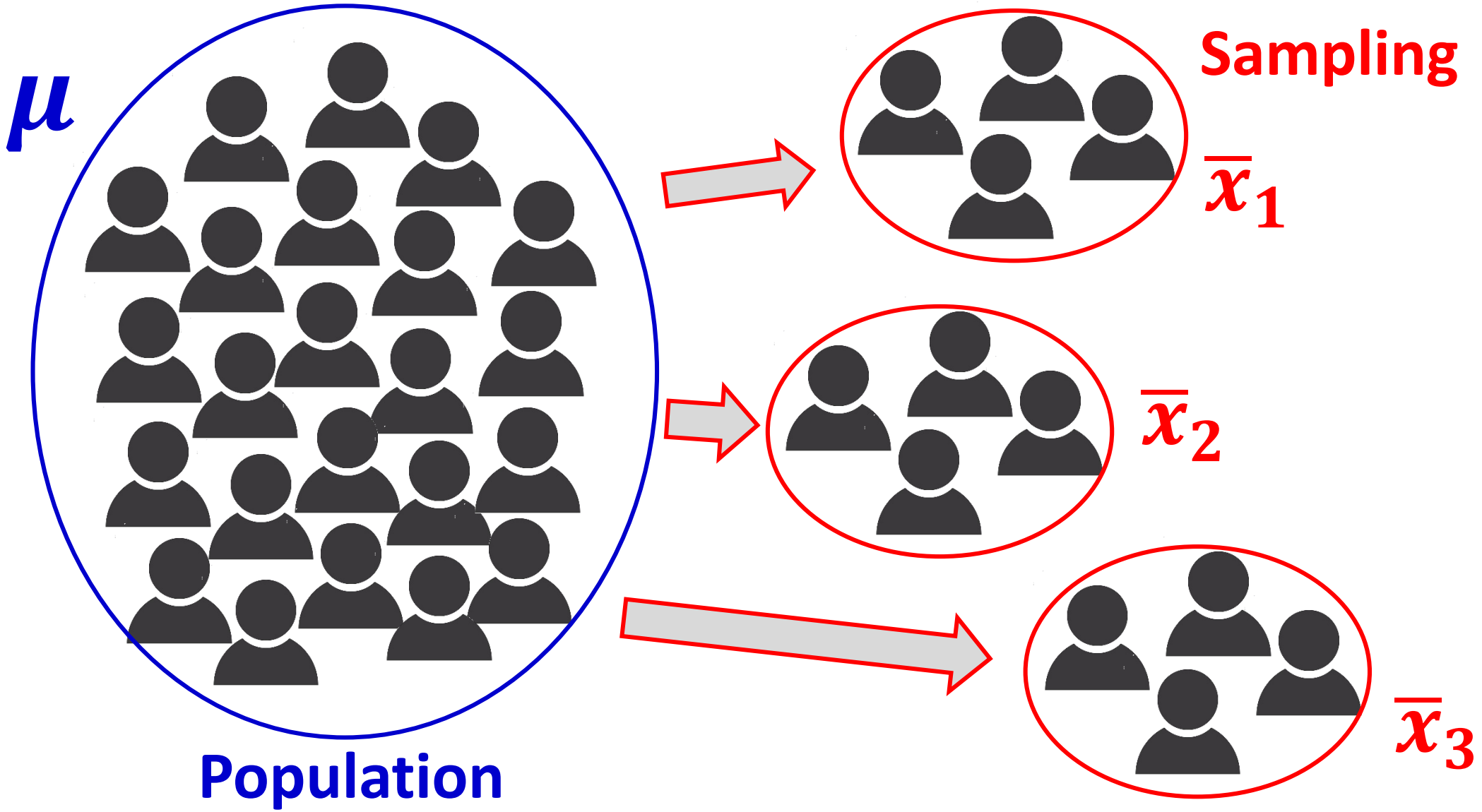
- Standard Deviation (SD)

```
> sd(age, na.rm = TRUE)
> var(age, na.rm = TRUE)
```

Measure of correlation




```
>cov(age, wt, method = "pearson", use = "complete.obs")
>cor(age, wt, method = "pearson", use = "complete.obs")
```

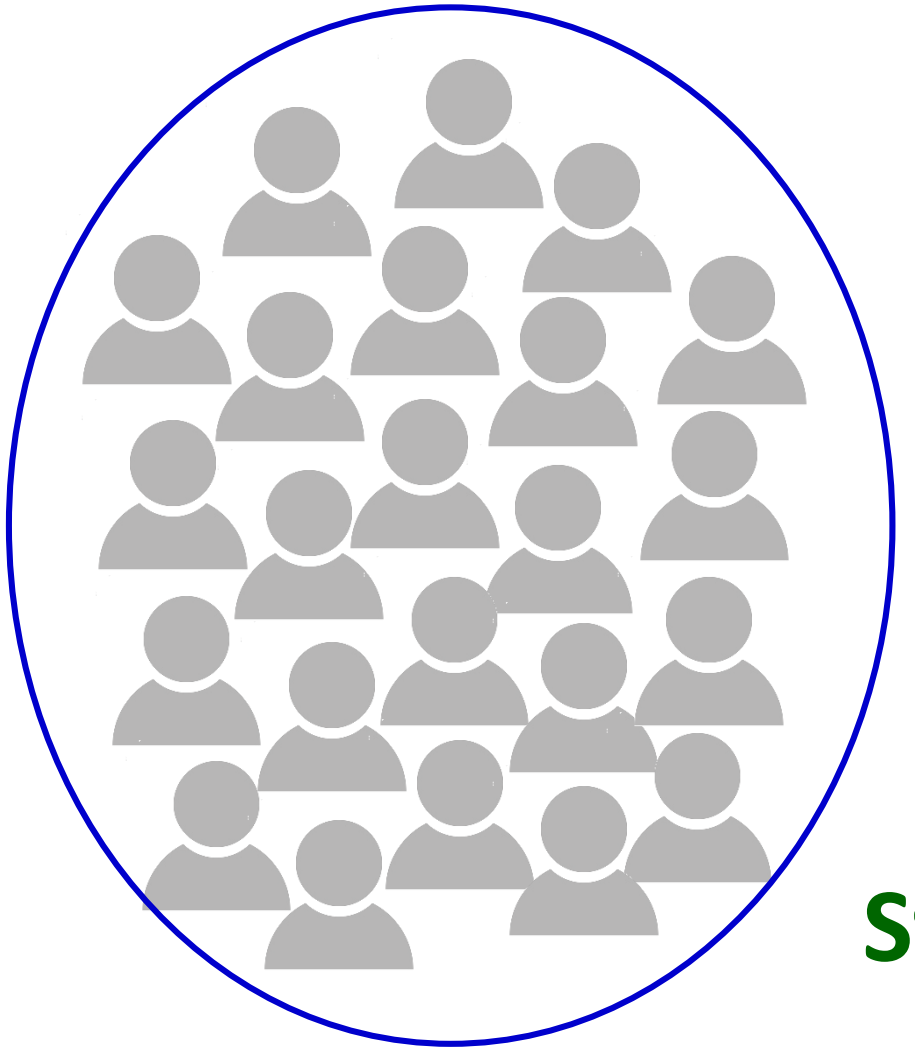

μ
Sampling
$\overline{x}_1$
$\overline{x}_2$
$\overline{x}_3$
Population

How do you know that the $\bar{x}_1 = \mu$?

95% Confidence Interval

Guess $\pm$ 2 $\times$ SEs

Mean $\pm$ 2 $\times$ SD



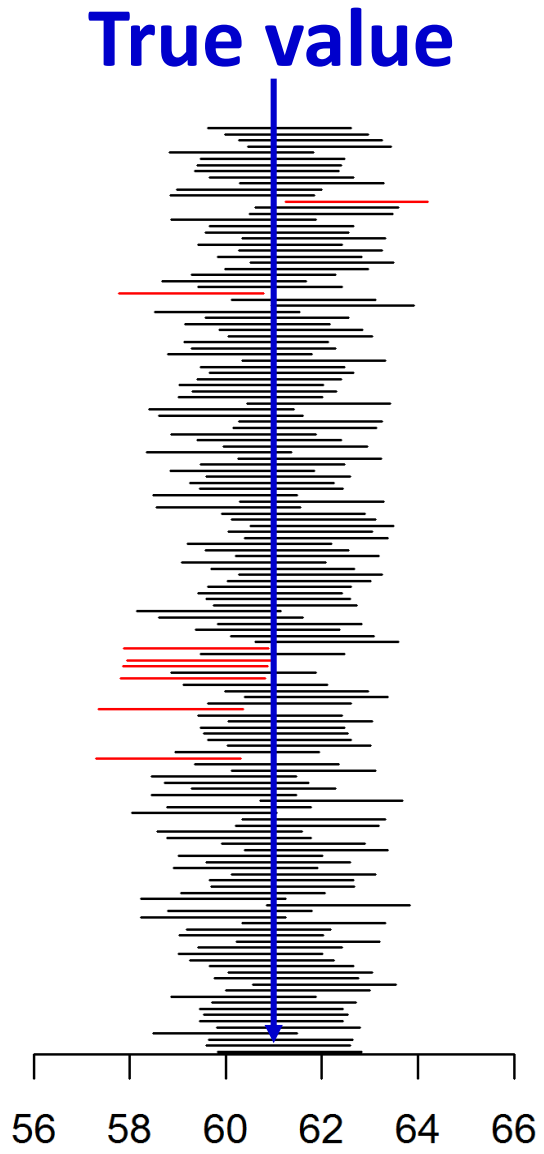
Population

Sampling

$\overline{x} \pm 2 \cdot SD$

μ

Statistical Inference



95% of researcher constructing 95% confidence intervals will have the true value inside their interval

5% will get wrong.

95% of the confidence intervals YOU construct in your lifetime will be “right”

Summary

- We **do not know** a population parameter.
- We pick some individual from the population and calculate the statistical parameter of the sample.
- We infer that the population parameter = sample parameter.
- We allow ourselves for the wrong inference by using the 95% confidence interval.