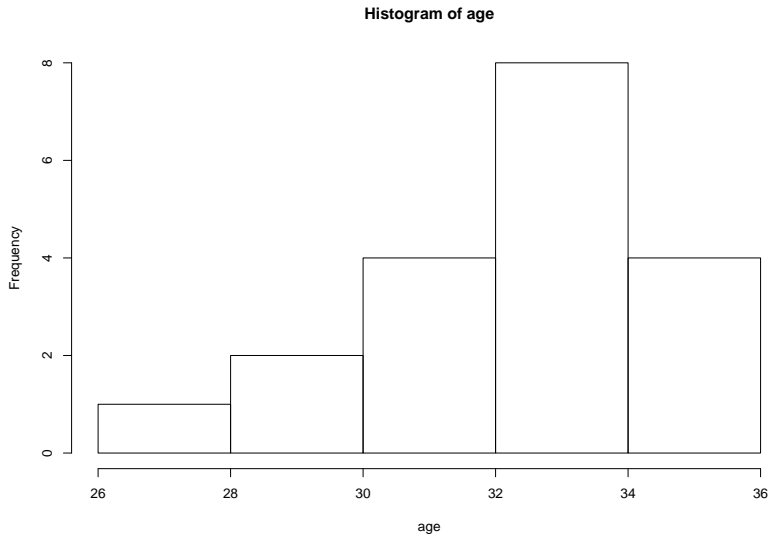


Descriptive statistics

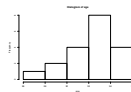
Guess my age

Graph your estimates

```
hist(age)
```

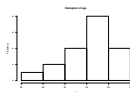


Summarise that distribution



- **Central tendency / location**

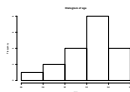
Summarise that distribution



► Central tendency / location

► mean: $\frac{a_1 + a_2 + a_3}{n}$

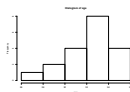
Summarise that distribution



► Central tendency / location

- mean: $\frac{a_1 + a_2 + a_3}{n}$
- median (50% percentile)

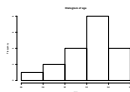
Summarise that distribution



► Central tendency / location

- mean: $\frac{a_1 + a_2 + a_3}{n}$
- median (50% percentile)
- mode (most frequent value)

Summarise that distribution

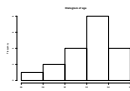


► Central tendency / location

- mean: $\frac{a_1 + a_2 + a_3}{n}$
- median (50% percentile)
- mode (most frequent value)

► Variation / Spread

Summarise that distribution



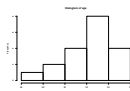
► Central tendency / location

- mean: $\frac{a_1 + a_2 + a_3}{n}$
- median (50% percentile)
- mode (most frequent value)

► Variation / Spread

- min, max, range

Summarise that distribution



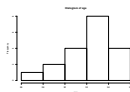
► Central tendency / location

- mean: $\frac{a_1 + a_2 + a_3}{n}$
- median (50% percentile)
- mode (most frequent value)

► Variation / Spread

- min, max, range
- quantiles (quartiles, percentiles...)

Summarise that distribution



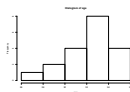
► Central tendency / location

- mean: $\frac{a_1 + a_2 + a_3}{n}$
- median (50% percentile)
- mode (most frequent value)

► Variation / Spread

- min, max, range
- quantiles (quartiles, percentiles...)
- standard deviation: $SD = \sqrt{\frac{\sum (x - \mu)^2}{n - 1}}$

Summarise that distribution



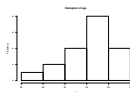
► Central tendency / location

- mean: $\frac{a_1 + a_2 + a_3}{n}$
- median (50% percentile)
- mode (most frequent value)

► Variation / Spread

- min, max, range
- quantiles (quartiles, percentiles...)
- standard deviation: $SD = \sqrt{\frac{\sum (x - \mu)^2}{n - 1}}$
- standard error: $SEM = \frac{SD}{\sqrt{n}}$

Summarise that distribution



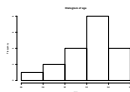
► Central tendency / location

- mean: $\frac{a_1 + a_2 + a_3}{n}$
- median (50% percentile)
- mode (most frequent value)

► Variation / Spread

- min, max, range
- quantiles (quartiles, percentiles...)
- standard deviation: $SD = \sqrt{\frac{\sum (x - \mu)^2}{n - 1}}$
- standard error: $SEM = \frac{SD}{\sqrt{n}}$
- coefficient of variation ($CV = SD / \text{mean}$)

Summarise that distribution



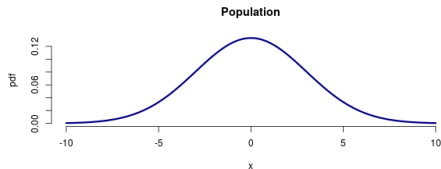
► Central tendency / location

- mean: $\frac{a_1 + a_2 + a_3}{n}$
- median (50% percentile)
- mode (most frequent value)

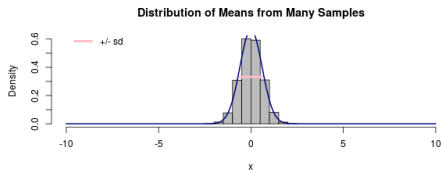
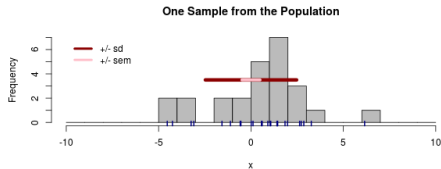
► Variation / Spread

- min, max, range
- quantiles (quartiles, percentiles...)
- standard deviation: $SD = \sqrt{\frac{\sum (x - \mu)^2}{n - 1}}$
- standard error: $SEM = \frac{SD}{\sqrt{n}}$
- coefficient of variation ($CV = SD / \text{mean}$)
- confidence intervals

Relationship between SD and SEM

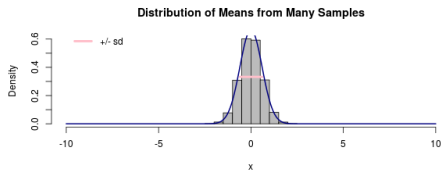
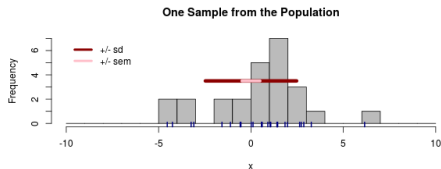
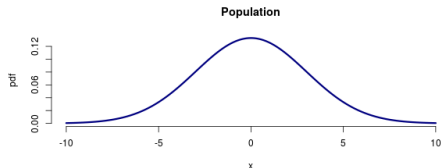


► SD quantifies scatter in population

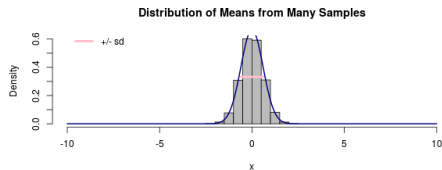
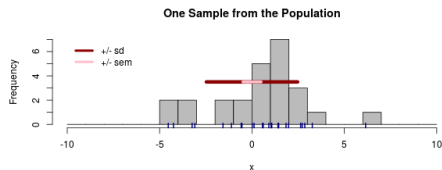
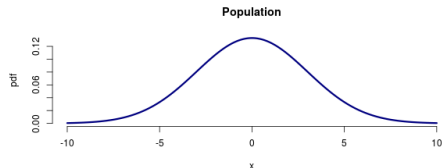


Relationship between SD and SEM

- ▶ SD quantifies scatter in population
- ▶ SEM quantifies uncertainty in parameter estimate (population mean)

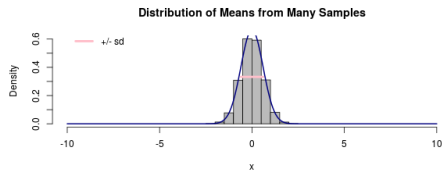
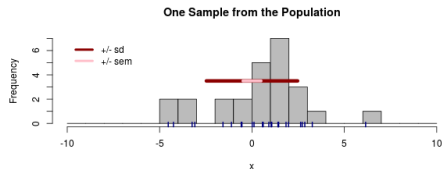
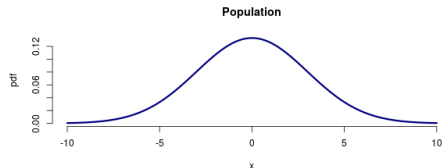


Relationship between SD and SEM



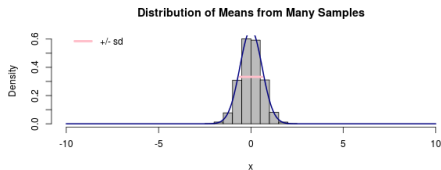
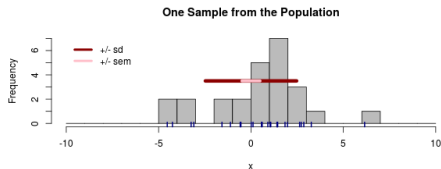
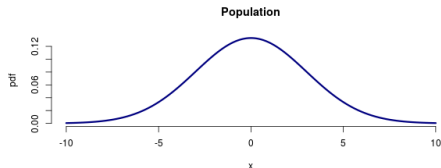
- ▶ SD quantifies scatter in population
- ▶ SEM quantifies uncertainty in parameter estimate (population mean)
- ▶ $SEM = SD/\sqrt{n}$

Relationship between SD and SEM



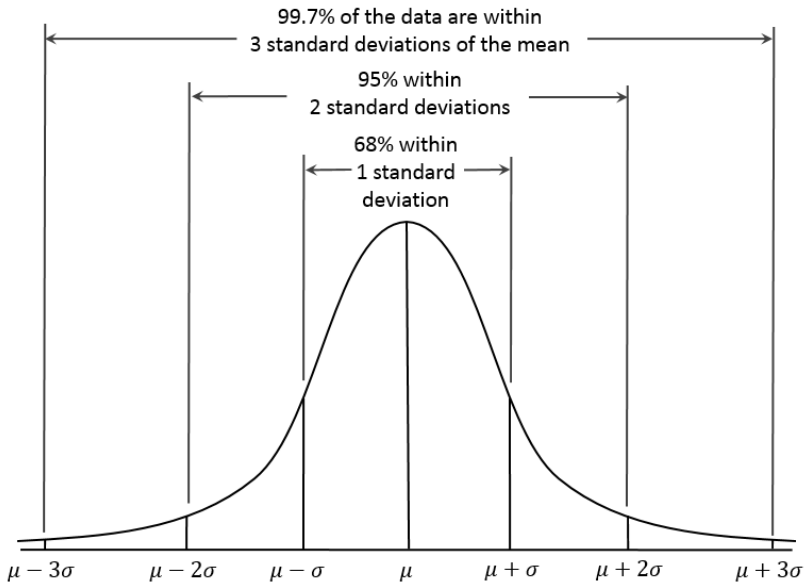
- ▶ SD quantifies scatter in population
- ▶ SEM quantifies uncertainty in parameter estimate (population mean)
- ▶ $SEM = SD/\sqrt{n}$
- ▶ SEM decreases with sample size (mean better known), SD does not.

Relationship between SD and SEM



- ▶ SD quantifies scatter in population
- ▶ SEM quantifies uncertainty in parameter estimate (population mean)
- ▶ $SEM = SD/\sqrt{n}$
- ▶ SEM decreases with sample size (mean better known), SD does not.
- ▶ https://gallery.shinyapps.io/sampling_and_stderr/

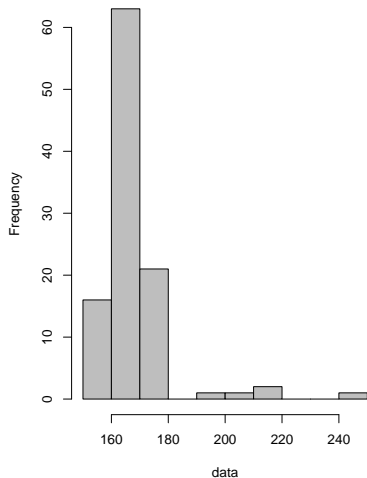
In a Normal distribution



What statistical descriptors are best? (and why)

<https://pollev.com/franciscorod726>

Histogram of data



Histogram of data

