# Day 02

# Variation, error, and uncertainty

- Variation, error
- Bootstrapping, confidence & prediction interval
- Uncertainty laundering: pretending we know more than we do
- Error propagation: when small errors become big problems

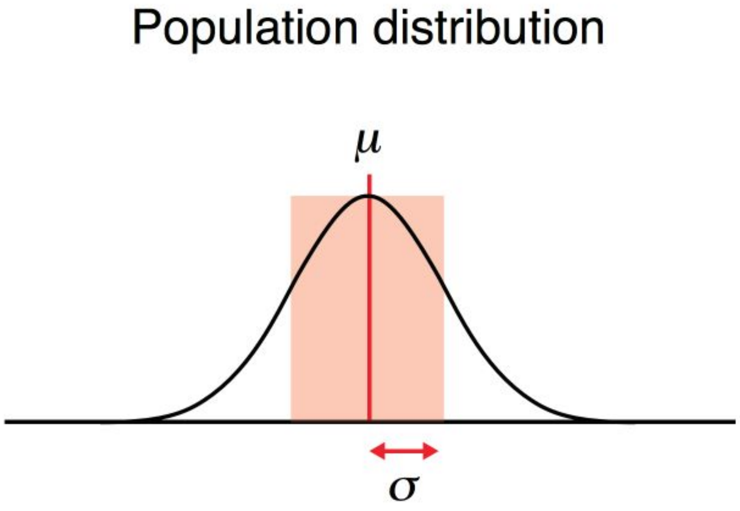
# Whether we are right vs. the chances of being wrong

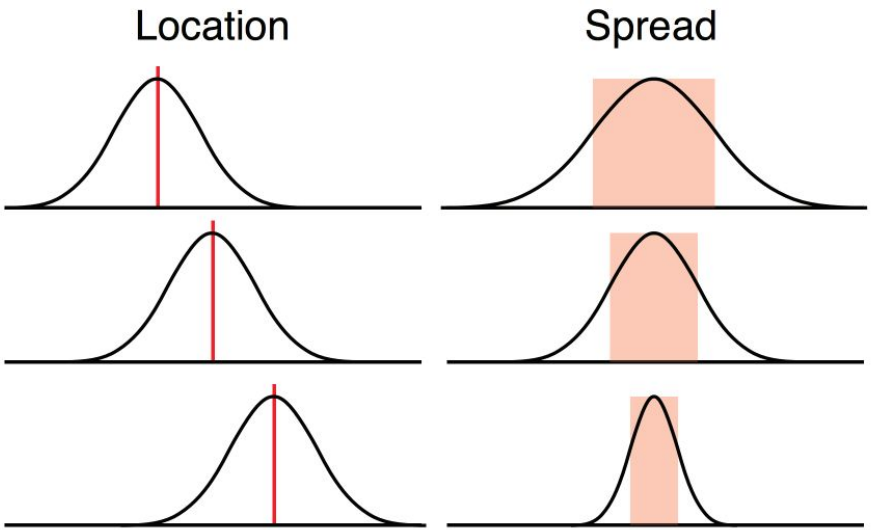
Repeated measurements → Range of values.

Statistics helps us by helping with:

- Modeling the role of chance
- Represent data as estimates with errors

# Population distribution



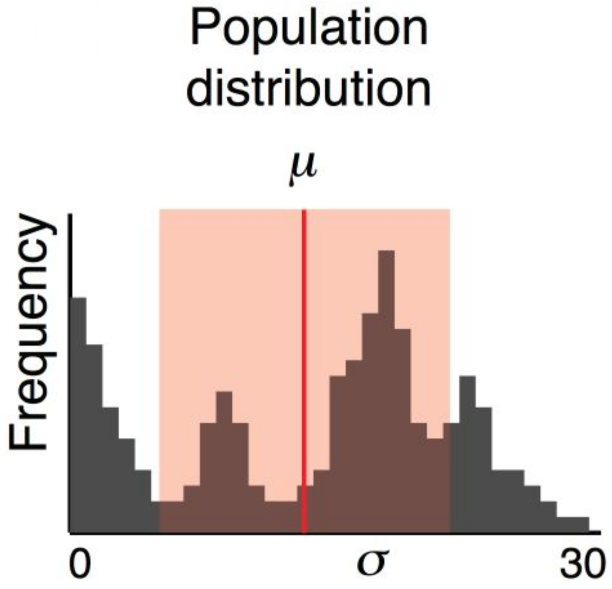




$\mu$: Population mean | $\sigma$: Population standard deviation

These are, of course, hard to calculate because it is hard to collect data about the entire population.

Krzywinski & Altman 2013 Nat. Methods 10:809

# Estimating population parameters by sampling




Samples

$X_1 = [1,9,17,20,26]$

$X_2 = [8,11,16,24,25]$

$X_3 = [16,17,18,20,24]$

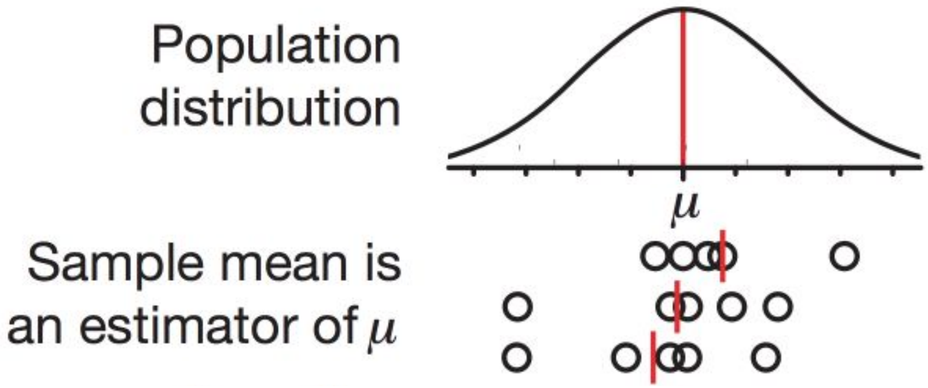
...

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

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

Krzywinski & Altman 2013 Nat. Methods 10:809

# Estimating population parameters by sampling



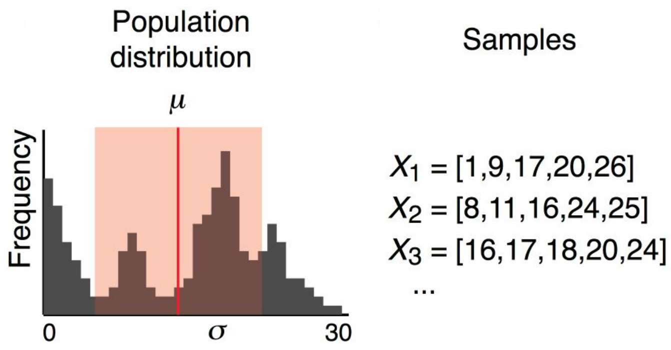

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

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

Krzywinski & Altman 2013 Nat. Methods 10:809

# Standard deviation

- Error bars based on **s.d.** → spread of your data.
- Useful as predictors of the range of new samples.
- Only indirectly supports visual assessment of differences in values:
  - **s.d.** bars reflect the variation of the data
  - They do not reflect the error in your measurement.

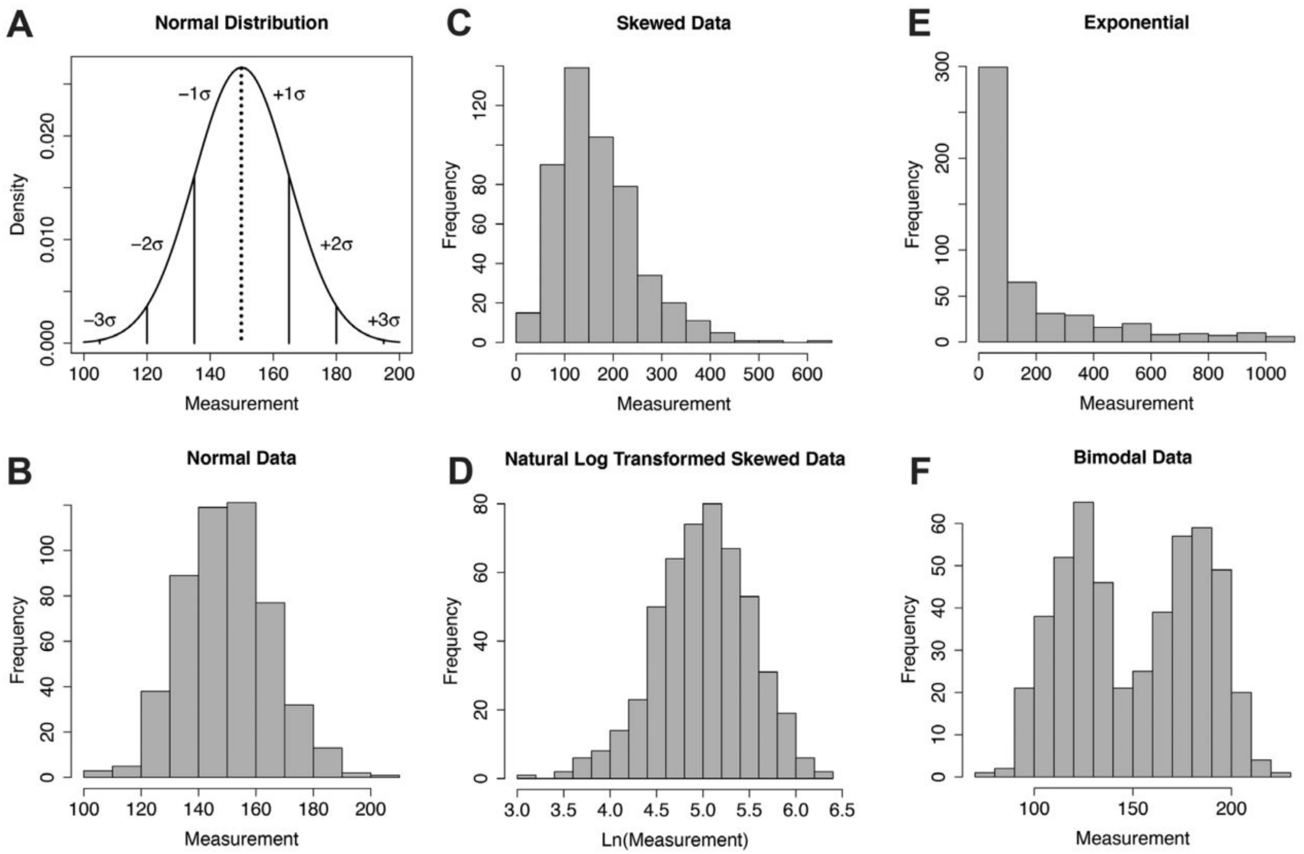



$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

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

Krzywinski & Altman 2013 Nat. Methods 10:921

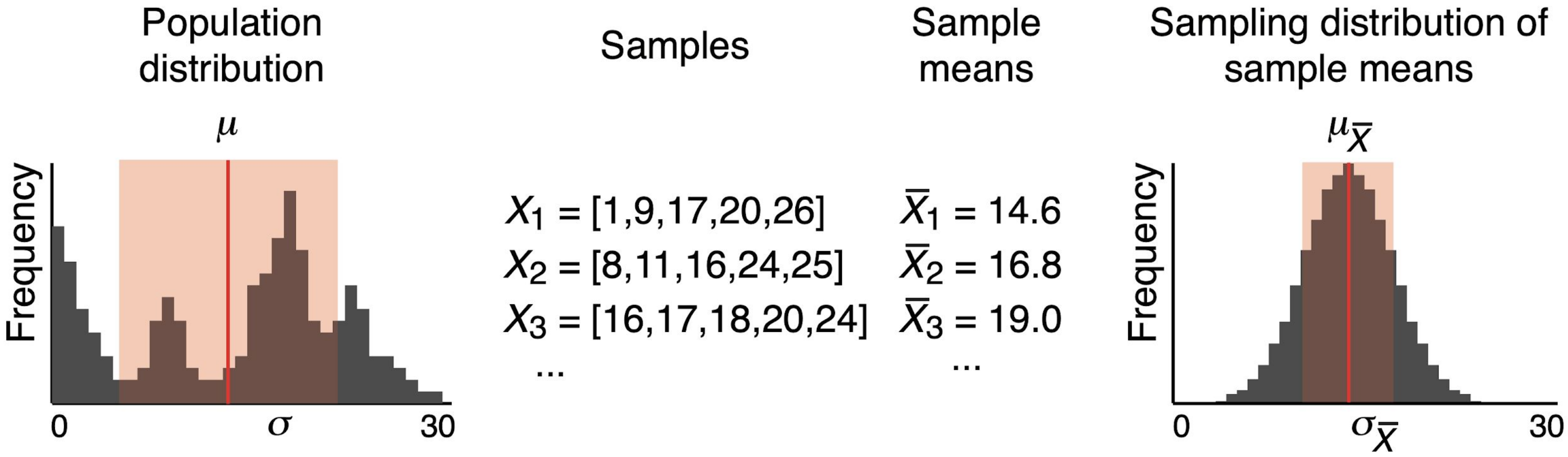
# Population distribution




https://www.molbiolcell.org/doi/10.1091/mbc.E15-02-0076

# Standard error of the mean

- Error bars based on **s.e.m.** → spread of the ***means*** of independent measurement samples, not the sample you collected (your data).
- s.e.m. = standard deviation of the means



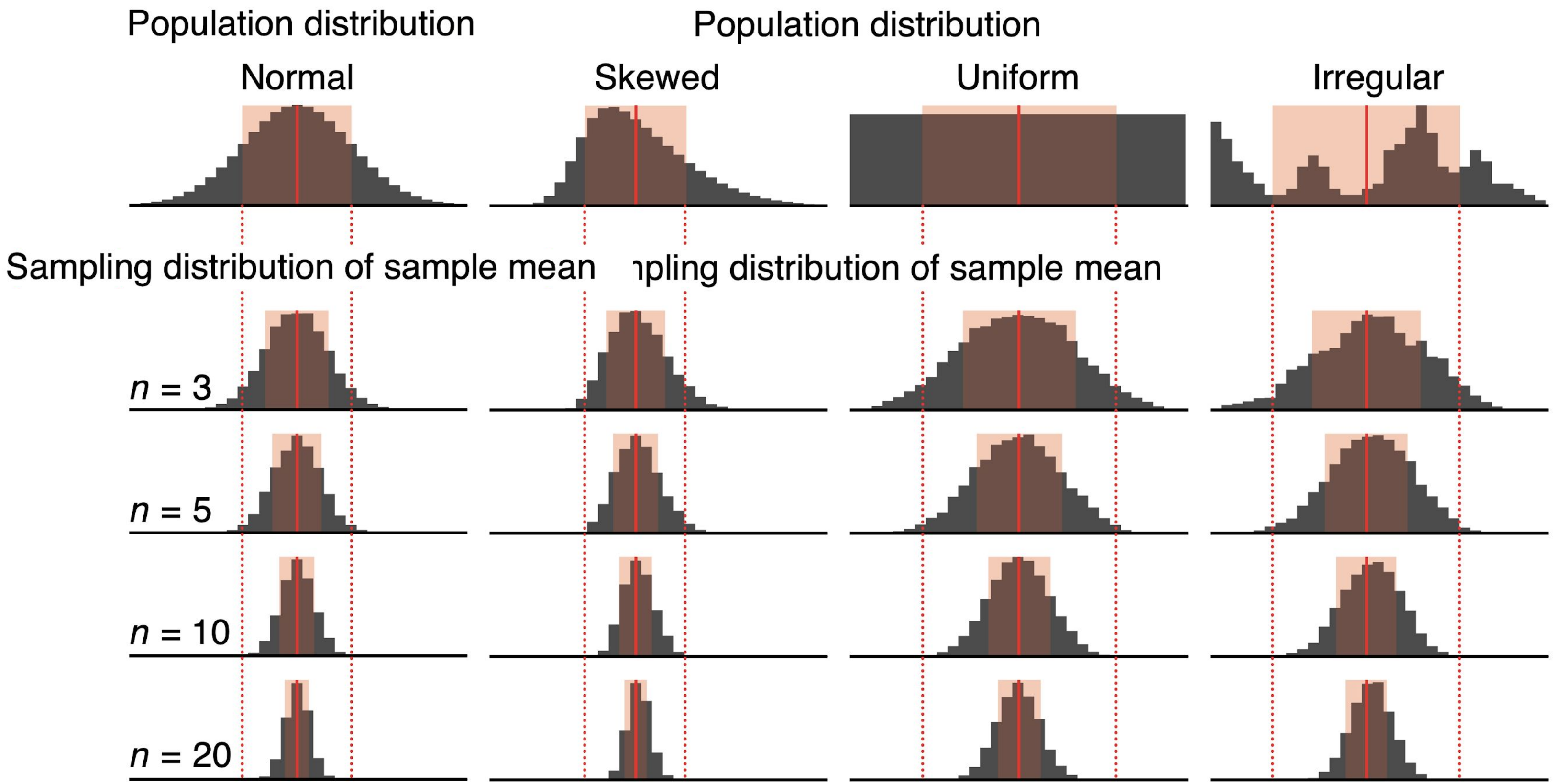

Krzywinski & Altman 2013 Nat. Methods 10:921

# Standard error of the mean

- Error bars based on **s.e.m.** → spread of the ***means*** of independent measurement samples, not the sample you collected (your data).
- s.e.m. = standard deviation of the means
- s.e.m. << s.d. of individual samples
- In rare cases, can be estimated using a formula: s.e.m. = s.d. / $\sqrt{n}$
  - Rest of the times, use bootstrapping.
- Dependent on sample size:
  - Shrinks as we perform more measurements.

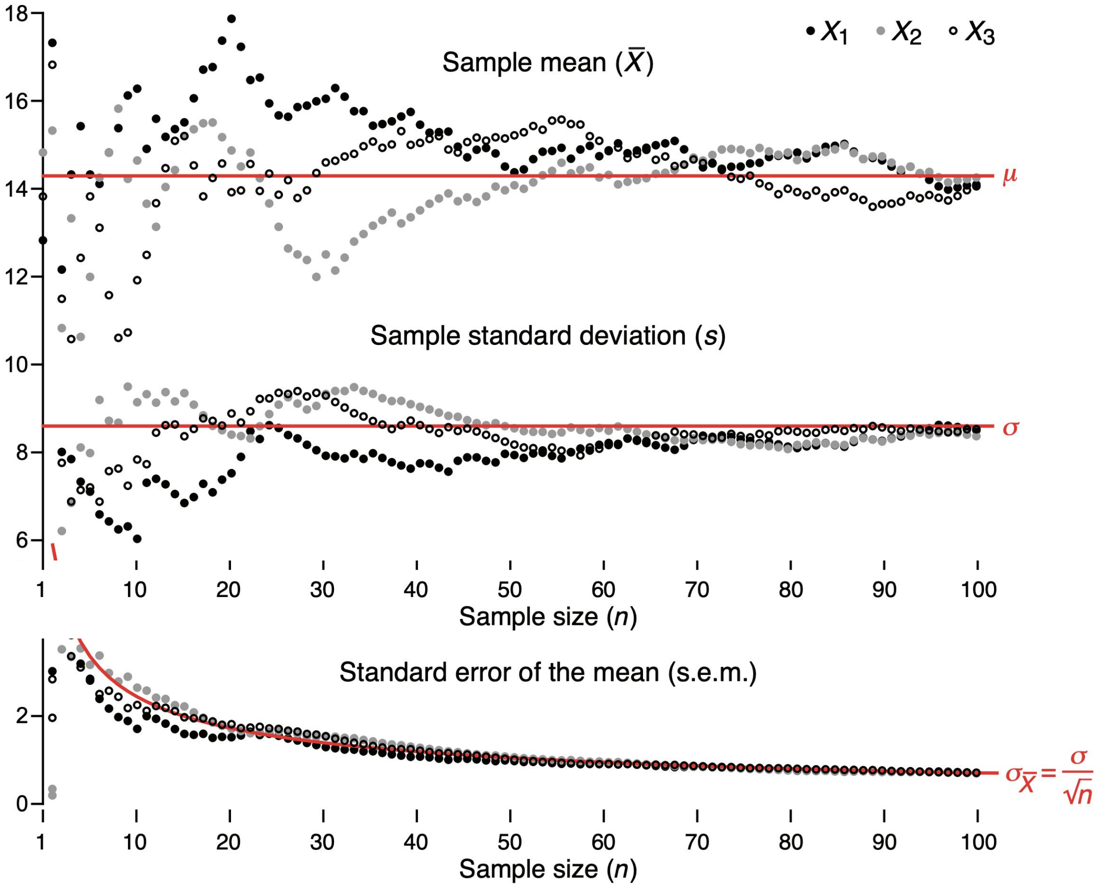
Krzywinski & Altman 2013 Nat. Methods 10:921

# Standard error of the mean




Krzywinski & Altman 2013 Nat. Methods 10:921

# Standard error of the mean




Krzywinski & Altman 2013 Nat. Methods 10:921

# Let’s write code to calculate mean, s.d., and s.e.m. of a sample

## Instructions

1. Generate 1000 random numbers from a normal distribution with mean = 0 & s.d. = 1. Let these 1000 numbers represent the population.
2. Randomly choose 10 numbers from these 1000.
   These 10 numbers represent a sample from the population.
3. Calculate the sample **mean**.
4. Calculate the sample **s.d.**
5. Calculate **s.e.m.** using the formula (s.d. / $\sqrt{n}$).

# Let's write code to <u>empirically</u> calculate s.e.m

**Instructions**

1. Generate 1000 random numbers from a normal distribution with mean = 0 & s.d. = 1
2. Repeat the following a 100 times:
    a. Randomly sample 10 numbers from the population of 1000 numbers
    b. Record their means
3. Calculate the s.d. of these 100 means.

What does this give you?

Recall: **s.e.m.** → spread of the ***means*** of independent measurement samples.

# Let's write code to <u>empirically</u> calculate s.e.m. of a sample

Given 10 numbers that represent a sample.
We have no access to the entire population.

```
30, 37, 36, 43, 42, 43, 43, 46, 41, 42
```

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 43 | 36 | 46 | 30 | 43 | 43 | 43 | 37 | 42 | 42 | 43 | 37 | 36 | 42 | 43 | 43 | 42 | 43 | 42 | 43 |
| 43 | 41 | 37 | 37 | 43 | 43 | 46 | 36 | 41 | 43 | 43 | 42 | 41 | 43 | 46 | 36 | 43 | 43 | 43 | 42 |
| 42 | 43 | 37 | 43 | 46 | 37 | 36 | 41 | 36 | 43 | 41 | 36 | 37 | 30 | 46 | 46 | 42 | 36 | 36 | 43 |
| 37 | 42 | 43 | 41 | 41 | 42 | 36 | 42 | 42 | 43 | 42 | 43 | 41 | 43 | 36 | 43 | 43 | 41 | 42 | 46 |
| 42 | 36 | 43 | 43 | 42 | 37 | 42 | 42 | 42 | 46 | 30 | 43 | 36 | 43 | 43 | 42 | 37 | 36 | 42 | 30 |
| 36 | 36 | 42 | 42 | 36 | 36 | 43 | 41 | 30 | 42 | 37 | 43 | 41 | 41 | 43 | 43 | 42 | 46 | 43 | 37 |
| 43 | 37 | 41 | 43 | 41 | 42 | 43 | 46 | 46 | 36 | 43 | 42 | 43 | 30 | 41 | 46 | 43 | 46 | 30 | 43 |
| 41 | 42 | 30 | 42 | 37 | 43 | 43 | 42 | 43 | 43 | 46 | 43 | 30 | 42 | 30 | 42 | 30 | 43 | 43 | 42 |
| 46 | 42 | 42 | 43 | 41 | 42 | 30 | 37 | 30 | 42 | 43 | 42 | 43 | 37 | 37 | 37 | 42 | 43 | 43 | 46 |
| 42 | 43 | 43 | 41 | 42 | 36 | 43 | 30 | 37 | 43 | 42 | 43 | 41 | 36 | 37 | 41 | 43 | 42 | 43 | 43 |

## Instructions

1. Create 1000 *bootstrap* samples:
   a. Each time, sample 10 numbers *with replacement*
   b. Calculate the mean of each bootstrap sample
2. Calculate the **s.d.** of these means.

This is the **s.e.m.** of your sample estimated using bootstrapping!

# Bootstrap methods: A practical tool for robust uncertainty

**Bootstrapping**

- Uses the original sample as a surrogate for the population, drawing repeated random samples with replacement to build an empirical distribution.
- Works when parametric assumptions (like normality) fail or when no textbook formulae exist for a complex statistic.
- Particularly valuable for estimating ratios, medians, or the 95% data range of non-symmetric distributions.

Fails if the original sample is small (<100), biased (not selected at random), or not independent.

Plot bootstrap distributions (e.g., rug plots):

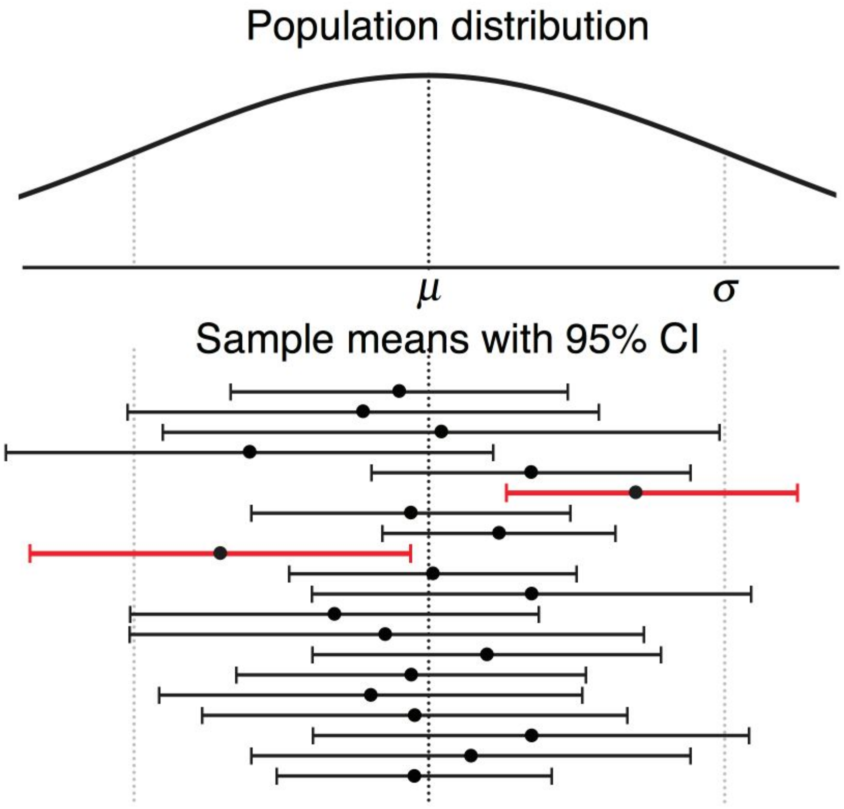
- Helps identify issues like discreteness, where a few unique values dominate the results.

# Confidence interval

- **CI** is an interval estimate that indicates the reliability of a measurement.
    - The 95% CI bar captures the population mean 95% of the time.
- Just like s.e.m., CI can be calculated using a bootstrapping technique.

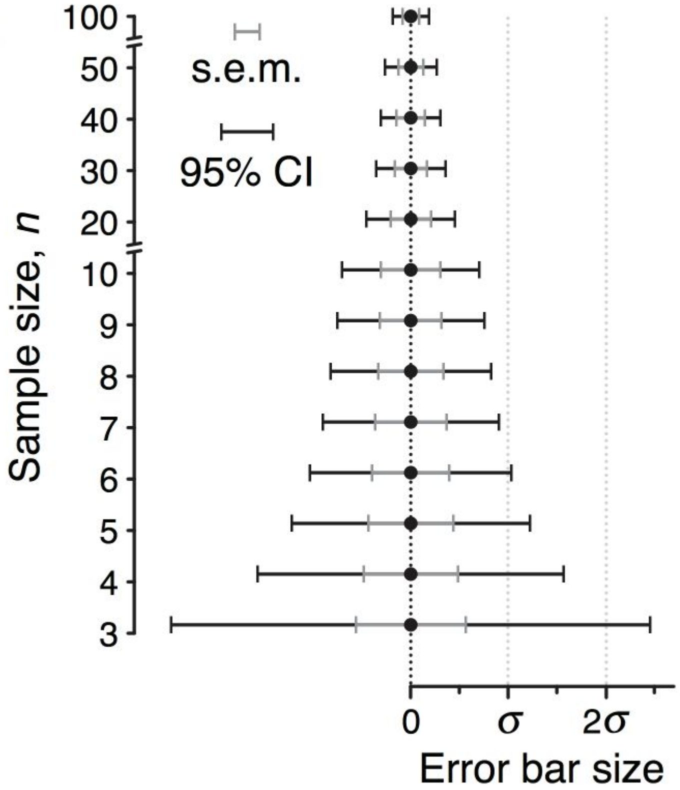
30, 37, 36, 43, 42, 43, 43, 46, 41, 42

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 43 | 36 | 46 | 30 | 43 | 43 | 43 | 37 | 42 | 42 | 43 | 37 | 36 | 42 | 43 | 43 | 42 | 43 | 42 | 43 |
| 43 | 41 | 37 | 37 | 43 | 43 | 46 | 36 | 41 | 43 | 43 | 42 | 41 | 43 | 46 | 36 | 43 | 43 | 43 | 42 |
| 42 | 43 | 37 | 43 | 46 | 37 | 36 | 41 | 36 | 43 | 41 | 36 | 37 | 30 | 46 | 46 | 42 | 36 | 36 | 43 |
| 37 | 42 | 43 | 41 | 41 | 42 | 36 | 42 | 42 | 43 | 42 | 43 | 41 | 43 | 36 | 43 | 43 | 41 | 42 | 46 |
| 42 | 36 | 43 | 43 | 42 | 37 | 42 | 42 | 42 | 46 | 30 | 43 | 36 | 43 | 43 | 42 | 37 | 36 | 42 | 30 |
| 36 | 36 | 42 | 42 | 36 | 36 | 43 | 41 | 30 | 42 | 37 | 43 | 41 | 41 | 43 | 43 | 42 | 46 | 43 | 37 |
| 43 | 37 | 41 | 43 | 41 | 42 | 43 | 46 | 46 | 36 | 43 | 42 | 43 | 30 | 41 | 46 | 43 | 46 | 30 | 43 |
| 41 | 42 | 30 | 42 | 37 | 43 | 43 | 42 | 43 | 43 | 46 | 43 | 30 | 42 | 30 | 42 | 30 | 43 | 43 | 42 |
| 46 | 42 | 42 | 43 | 41 | 42 | 30 | 37 | 30 | 42 | 43 | 42 | 43 | 37 | 37 | 37 | 42 | 43 | 43 | 46 |
| 42 | 43 | 43 | 41 | 42 | 36 | 43 | 30 | 37 | 43 | 42 | 43 | 41 | 36 | 37 | 41 | 43 | 42 | 43 | 43 |




Krzywinski & Altman 2013 Nat. Methods 10:921

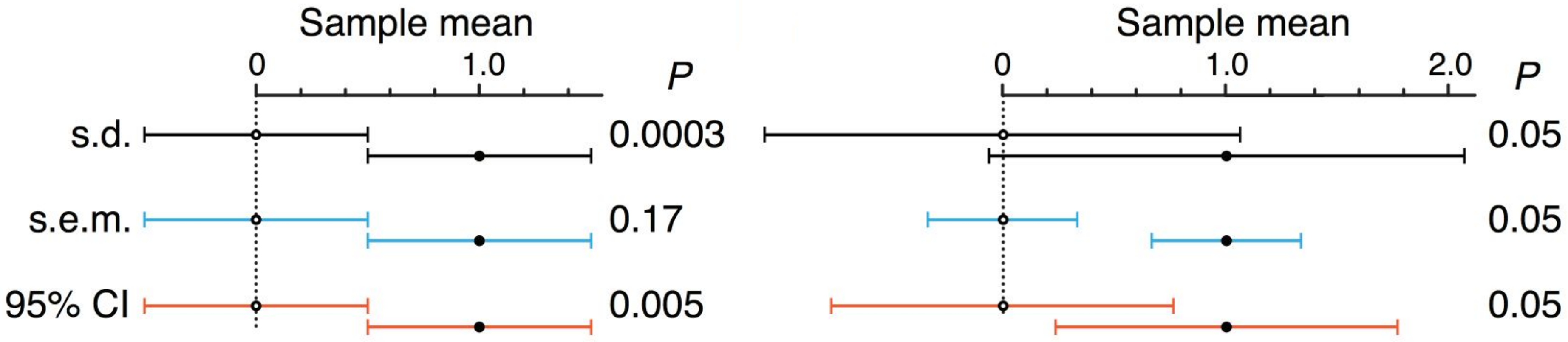
# Confidence interval




Krzywinski & Altman 2013 Nat. Methods 10:921

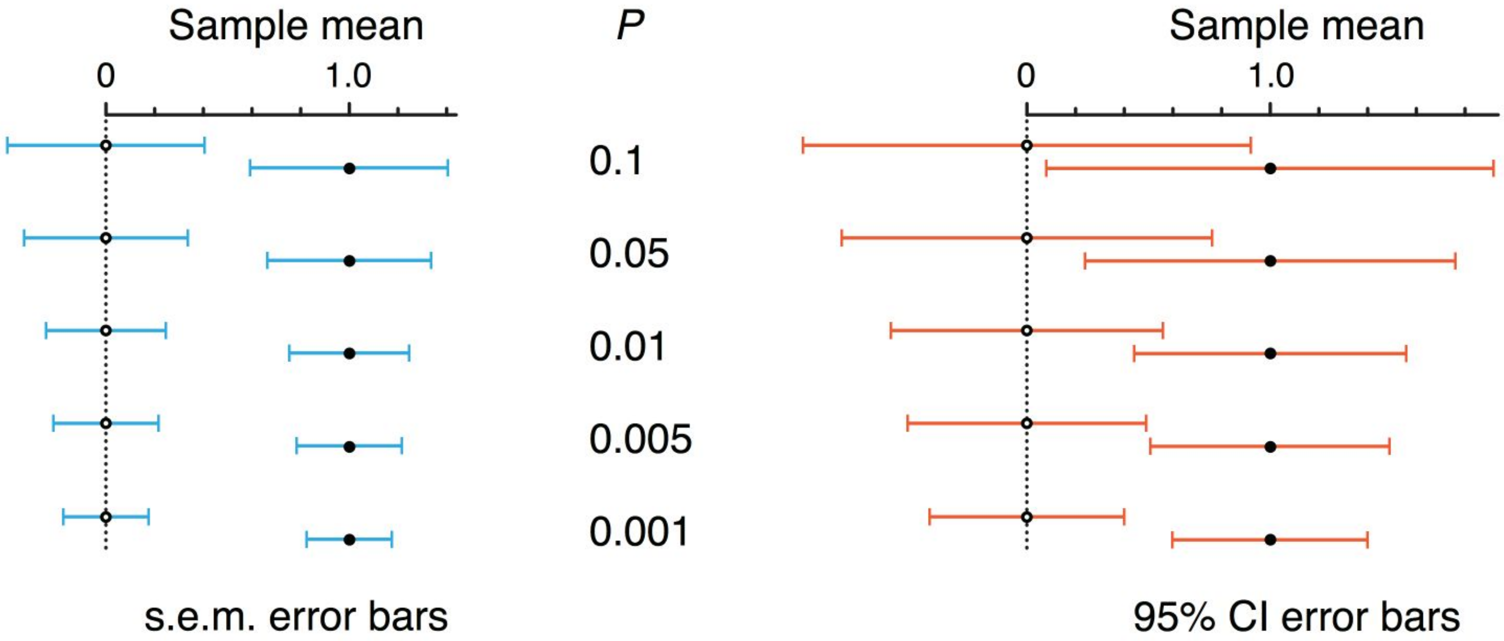
# The type of “spread” matters

- Non-overlapping ≠ “significant” difference
- Overlapping ≠ not “significant” difference
- It depends on the type of the error bar.




Krzywinski & Altman 2013 Nat. Methods 10:921

# Standard error of the mean, Confidence interval




Krzywinski & Altman 2013 Nat. Methods 10:921

# The confidence interval vs. prediction interval confusion

**The critical distinction**

Confidence Interval (CI): quantifies uncertainty about a population parameter (e.g., the mean)

Prediction Interval (PI) quantifies uncertainty about future individual observations (yet to be measured)

PIs are always wider than CIs. [Question: How do we use bootstrapping to calculate the PI?]

Researchers often use a CI when a PI is required, which dangerously understates uncertainty, and leads to "dangerously narrow" expectations.

- E.g., in clinical settings, claiming "this patient will respond" based on a population mean CI ignores the wider variability captured by a PI.
- If a drug's effect has high variance, the mean might look favorable, but the treatment may be inappropriate for many individual patients.

CIs should be used to report the precision of estimates, but PIs must be used when the goal is to forecast individual outcomes. Check papers when you read them.

https://robjhyndman.com/hyndsight/intervals/

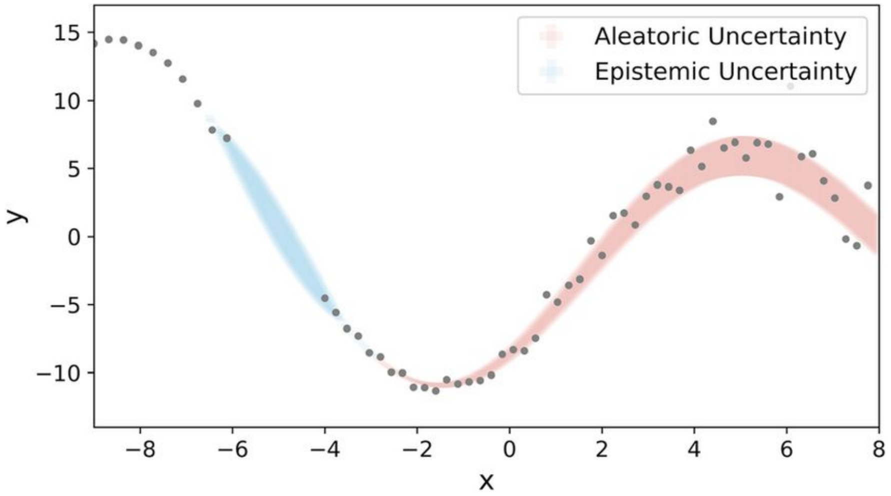
# Uncertainty laundering: Pretending we know more than we do

**Aleatoric vs. epistemic uncertainty**

List all possible sources of variation. Control them or measure them to avoid their being confounded with relationships among those items that are of primary interest.

It's important to distinguish between:

- **Aleatoric** uncertainty: irreducible randomness from
  measurement noise or biological variability
- **Epistemic** uncertainty: knowledge gaps regarding the correct model or parameter values (thus neglecting certain effects) or because particular data have been deliberately hidden




van der Bles et al., 2019 R Soc Open Sci.
https://en.wikipedia.org/wiki/Uncertainty_quantification

# Uncertainty laundering: Pretending we know more than we do

**Uncertainty laundering (h iding epistemic uncertainty)**

Researchers reporting only sampling uncertainty (standard errors or confidence intervals) while completely ignoring model uncertainty.

- Presenting a single model as "the truth", which ignores the fact that different, equally plausible models could lead to different results.
- Results quoted with an "aura of conviction" or excessive precision (e.g., reporting a p-value to five decimal places) when the underlying assumptions, such as normality, are known to be violated.

van der Bles et al., 2019 R Soc Open Sci.

# Uncertainty laundering: Pretending we know more than we do

## Communicating uncertainty

### What do you have uncertainty about (**object**)?

Underlying hypothesis behind your evidence, and/or about the specific numbers involved, and/or about categorical facts that you want to claim?

- Different levels of uncertainty about the fact that the mean global temperature has risen since 1850; that the temperature rise has been approximately 1.8°, or that the temperature rise has been the result of the 'greenhouse effect'.

### Why is there uncertainty (**source**)?

Unavoidable natural variation, because of the difficulties of measuring, because of limited knowledge about the underlying processes or because there is disagreement between experts?

- Thinking about this may help you identify more objects you are uncertain about, and about how your audience might perceive the uncertainty.

van der Bles et al., 2019 R Soc Open Sci.

# Uncertainty laundering: Pretending we know more than we do

**Communicating uncertainty**

For each object & source:

- Do you have both direct (specifically about that aspect) and indirect (quality of evidence) levels of uncertainty? You will probably want to, or have to, communicate them separately.

Key choices & considerations:

- Choose an expression of your uncertainties that suits the degree of precision you have.
- Consider your audience, their relationships to you and to the subject, and the effects you want to have on them.

Stating uncertainty does not necessarily undermine trust.

Important: keep your expressions of the magnitude of uncertainty clearly separate from the magnitude of any evidence you are trying to communicate.

- E.g. not confusing the effect of processed meat as a carcinogen, low; with the certainty that it is one, high.

van der Bles et al., 2019 R Soc Open Sci.
Communicating risks and benefits: an evidence-based user's guide. 2011 FDA.

# Error propagation: When small errors become big problems

## Compounding uncertainty

In multi-step analyses, small initial errors such as transcription mistakes or imprecise calibration can compound, leading to inconclusive or wildly varying results.

- E.g., Small errors in measurement can result in substantial bias in the regression coefficients of complex models.

Uncertainty propagates rapidly when using a model to predict values far from the calibration mean.

- E.g., assuming a linear trend in Olympic sprinting will continue indefinitely eventually leads to physically impossible "negative" times.