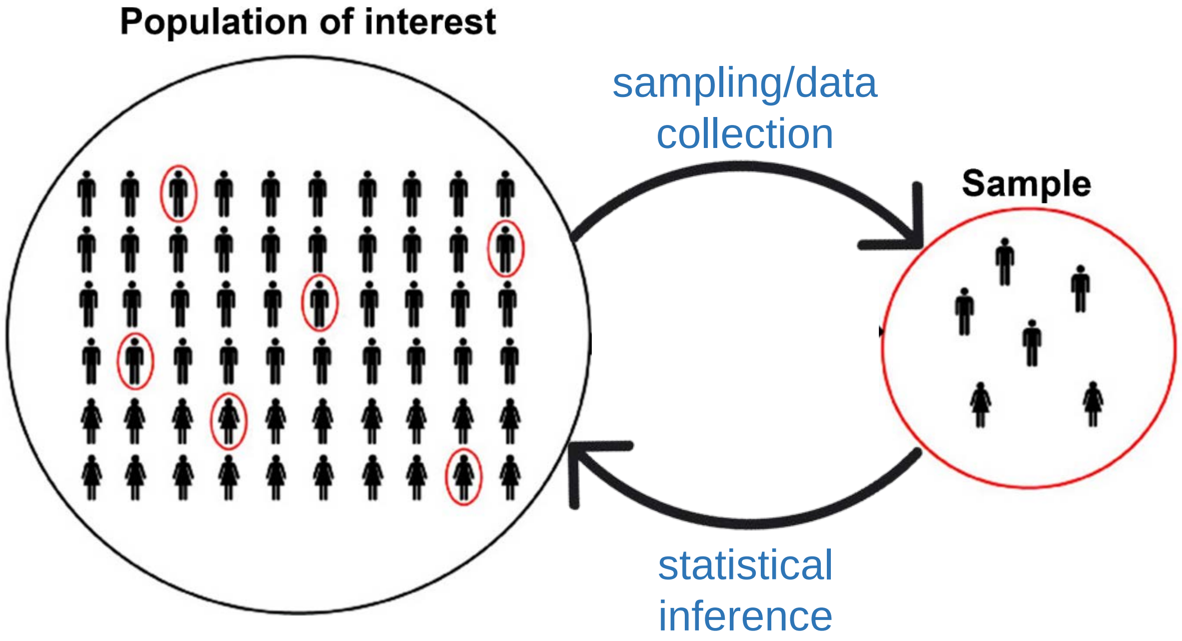
Continuous data – 1 group estimation

Notation

- μ = population mean
- σ = population standard deviation

- $\bar{x}$ = sample mean
- s = sample standard deviation

μ is fixed



$\bar{x}$ varies (would be different with a different sample of the population)

Goal: What is the population mean?

- Can't measure the entire population so take a sample instead
- What is our best guess of the population mean?
  - The sample mean, of course!
- So does $\bar{x} = \mu$? Not exactly. What would make us more sure that $\bar{x}$ is close to μ?
  - More subjects or fewer subjects in the sample?
  - Sample observations more spread out or closer together?

Low Birth Weight Data

- Information on 100 low birth weight infants born in two teaching hospitals in Boston, Massachusetts

Variable	Description
sex	Sex of the baby (Male, Female)
gestage	Gestational age at time of birth (weeks)
length	Length of the baby (cm)
birthwt	Birth weight of the baby (g)
headcirc	Baby's head circumference (cm)
apgar	Apgar score (integers, min=0, max=10). This is a scoring system used for assessing the clinical status of a newborn. 7 or higher is generally considered normal, 4-6 is low, and 3 or below is critically low.

Find the dataset (lowbwt.xlsx) and the full data dictionary (lowbwt Data Dictionary.pdf) in the Data Module on the Canvas site

Example: Length of LBW Infants

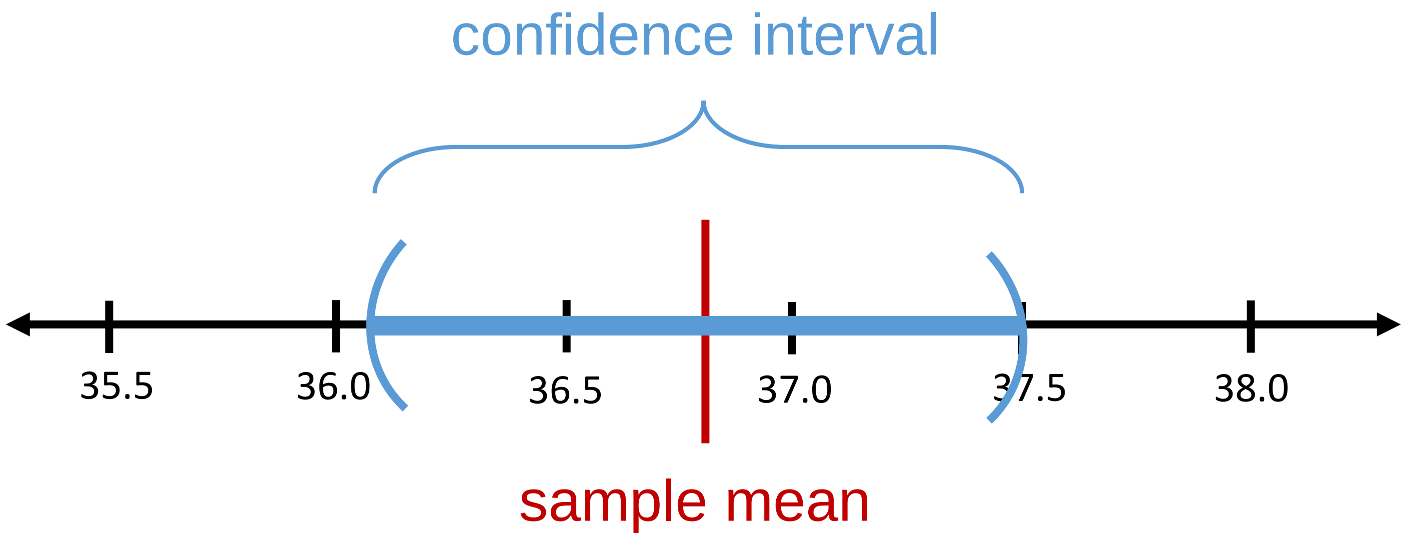
- Low birthweight is a term used to describe babies who are born weighing less than 2,500 grams (5 pounds, 8 ounces).
- What is the average length of all low birth weight infants in the United States?

Sample summary statistics for length variable:

mean	sd	0%	25%	50%	75%	100%	n
36.82	3.571435	20	35	38	39	43	100

Confidence Interval

- Instead of just giving an estimate of the population mean, we can give an interval that we're pretty sure includes the population mean



Example: Length of LBW Infants

Sample summary statistics for length variable:

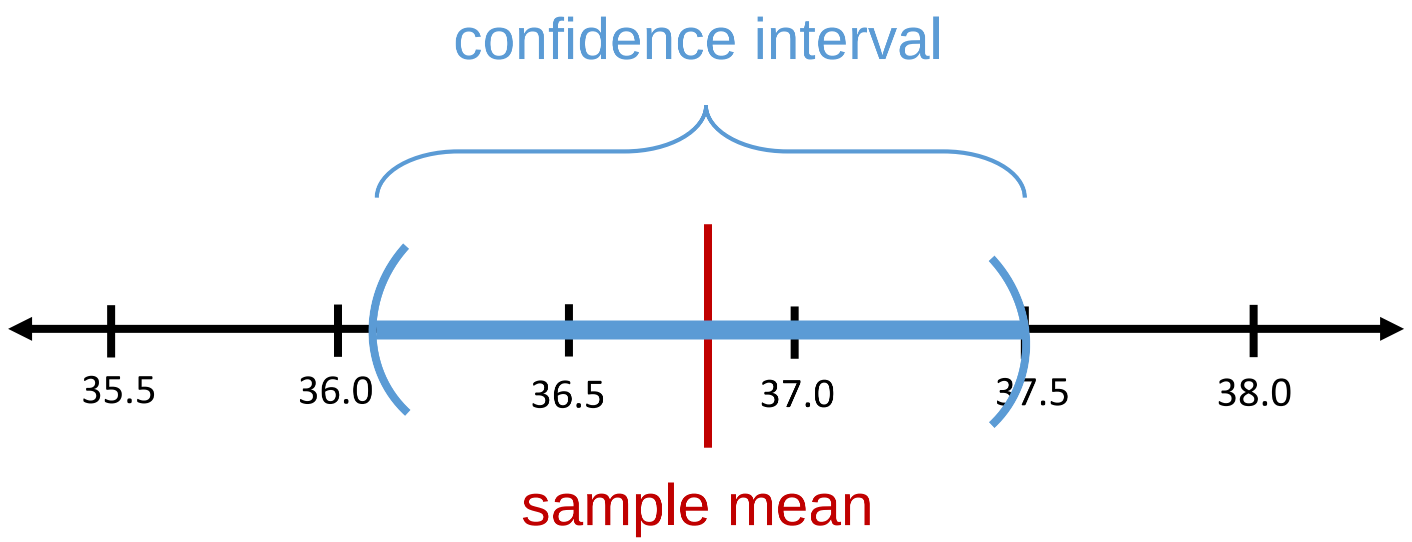
```
 mean       sd 0% 25% 50% 75% 100%   n
36.82 3.571435 20  35  38  39   43 100
```

95% confidence interval for mean length:

```
95 percent confidence interval:
 36.11135 37.52865
sample estimates:
mean of x
    36.82
```

- Interpretation: “We are 95% confident that the average length of low birth weight infants in the United States is between 36.1 cm and 37.5 cm.”

Example: Length of LBW Infants



- Interpretation: “We are 95% confident that the average length of low birth weight infants in the United States is between 36.1 cm and 37.5 cm.”

Confidence Interval Formula

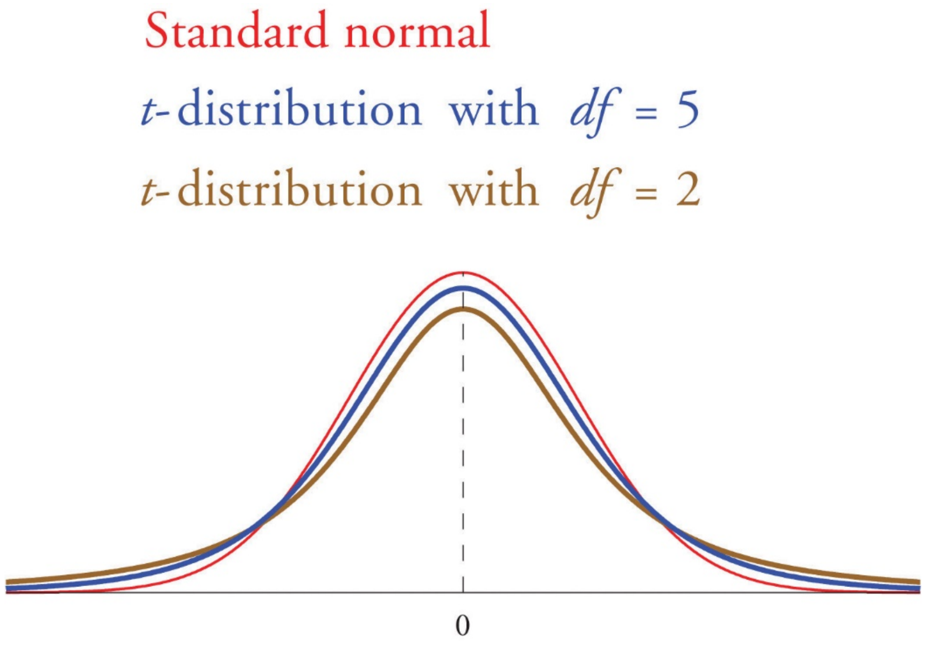
Confidence interval for μ:

$$\bar{x} \pm t_{df}^{*} \times \frac{s}{\sqrt{n}}$$

- $\bar{x}$ = sample mean
- t_{df}^{*} = critical value (represents our level of confidence)
- s = sample standard deviation
- n = sample size

t Distribution

- Like the normal distribution but a little shorter and wider
- **Degrees of freedom** (df) is a number that controls how wide it is (smaller df = wider)
  - Sample size determines df
- The extra width basically adds a little fudge factor to the normal distribution to account for the fact that you don't know the population standard deviation



Features of CI for Population Mean

- Centered at sample mean
- Increase sample size ➔ CI narrower
- Increase confidence level ➔ CI wider
- Increase variability in the sample ➔ CI wider

CI for μ: $\bar{x} \pm t^*_{df} \times \frac{s}{\sqrt{n}}$

- $\bar{x}$ = sample mean
- t^*_{df} = critical value (represents our level of confidence)
- s = sample standard deviation
- n = sample size

Example: Length of LBW Infants

Sample summary statistics for length variable:

```
 mean        sd 0% 25% 50% 75% 100%   n
36.82 3.571435 20  35  38  39   43 100
```

95% confidence interval for mean length:

```
95 percent confidence interval:
 36.11135 37.52865
sample estimates:
mean of x
    36.82
```

- What could we do to make the confidence interval narrower?

"We are 95% confident that the average length of low birth weight infants in the United States is between 36.1 cm and 37.5 cm."

Interpretation

- Remember: The CI is for the population mean, not the sample mean. We already know the sample mean!

WRONG interpretation:

- "We are 95% confident that the sample mean is between ____ and ____."

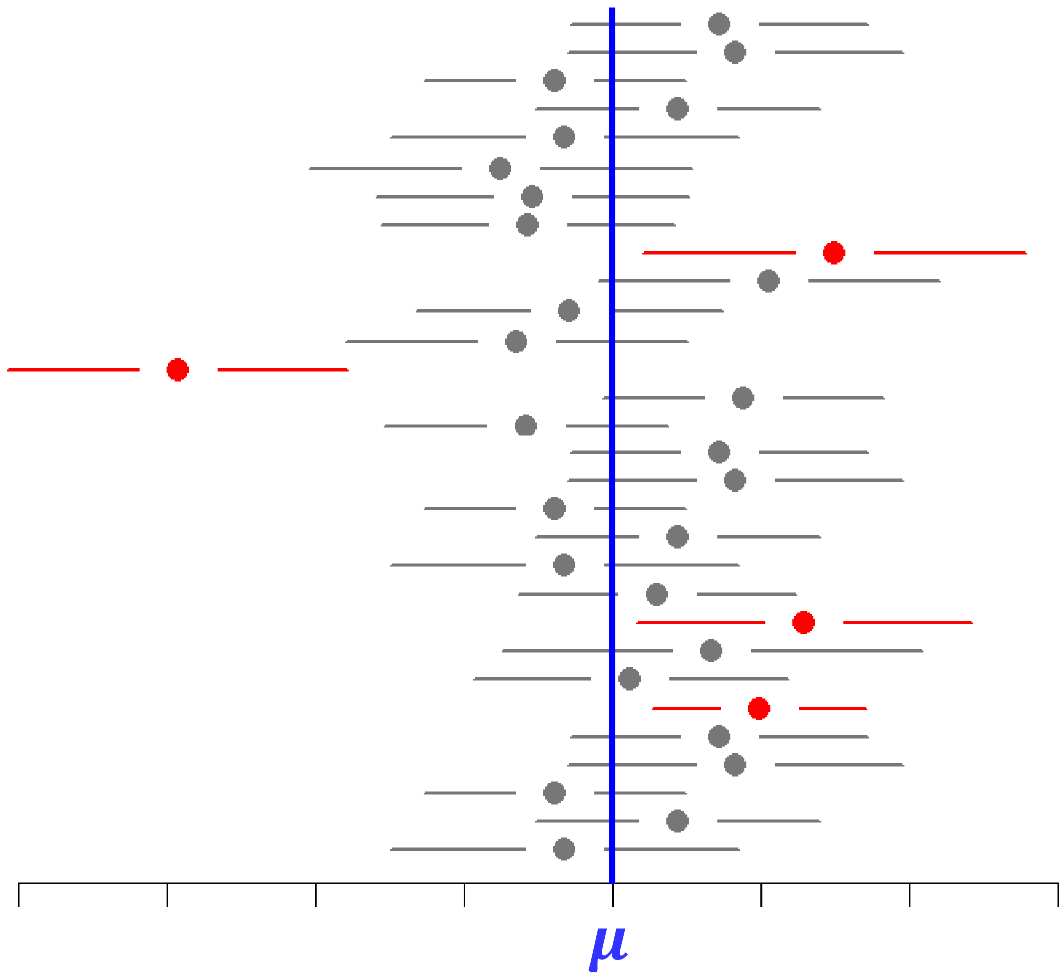
Interpretation

- Remember: μ is fixed! $\bar{x}$ and s (and therefore our confidence interval) change with new samples

WRONG interpretations:

- "There is a 95% chance that the population mean is between ____ and _____."
- "The probability that the population is between ____ and ____ is 95%."

Interpretation



- Suppose we repeated the process of taking a sample, calculating the sample mean, and calculating a 95% confidence interval 100 times. We would expect 95 of the 100 intervals to contain the true population mean.

Interpretation

- "We are 95% confident that the mean variable in the population is between ____ and ____."
  - "95%": change to appropriate level if not 95%
  - "variable": write out what the variable is
  - "in the population": write out what the population is
  - "___ and ___": fill in lower bound and upper bound – don't forget units!

Low birth weight infant example: "We are 95% confident that the average length of low birth weight infants in the United States is between 36.1 cm and 37.5 cm."

Plausible Values

- The CI contains what we consider to be plausible values for the population mean
- Gives us a standardized way to answer the question, “Is the population mean equal to ___?”
- If the value of interest is not in the CI, we’re confident that the answer is “No”.

Example: Apgar Score in LBW Infants

- Suppose that the average Apgar score of normal birth weight infants is 7 points. Is the average Apgar score for low birth weight infants different than that of normal birth weight infants?

95% confidence interval for Apgar score in low birth weight data:

```
95 percent confidence interval:
 5.767767 6.732233
sample estimates:
mean of x
     6.25
```

We are 95% confident that the average Apgar score for low birth weight infants is between 5.8 and 6.7 points. Thus, we can say with 95% confidence that it appears as if the average Apgar score of low birth weight infants is different than that of normal birth weight infants.

Important Points

- Sample mean is used to estimate population mean
- Interpretation of confidence interval for the mean
- How the variability of the data, the sample size, and your confidence level affect the width of the confidence interval