
# Estimating a Mean Difference for Paired Data

Julie Deeke

Statistics Course Developer

# Twin Education Levels

Twin Days in Twinsburg, Ohio annually since 1976

**Variable:** Education Level of Twins

# Paired Values

- Want to treat the two sets of values simultaneously

# Paired Values

- Want to treat the two sets of values simultaneously
- Other ways paired data arise:
  - Measurements collected on the same individual

# Paired Values

- Want to treat the two sets of values simultaneously
- Other ways paired data arise:
  - Measurements collected on the same individual

# Paired Values

- Want to treat the two sets of values simultaneously
- Other ways paired data arise:
  - Measurements collected on the same individual

# Paired Values

- Want to treat the two sets of values simultaneously
- Other ways paired data arise:
  - Measurements collected on the same individual
  - Measurements collected on matched individuals

# Paired Values

- Want to treat the two sets of values simultaneously
- Other ways paired data arise:
  - Measurements collected on the same individual
  - Measurements collected on matched individuals


# Paired Values

- Want to treat the two sets of values simultaneously
- Other ways paired data arise:
  - Measurements collected on the same individual
  - Measurements collected on matched individuals

# Paired Values

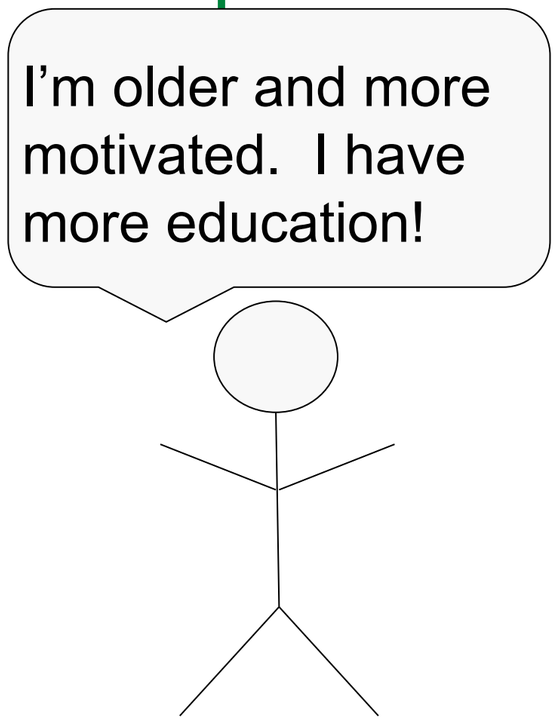
- Want to treat the two sets of values simultaneously
- Other ways paired data arise:
  - Measurements collected on the same individual
  - Measurements collected on matched individuals
- Variable: Difference of measurements within pairs

# Research Question

What is the *average* difference between the older twin's and younger twin's self-reported education?

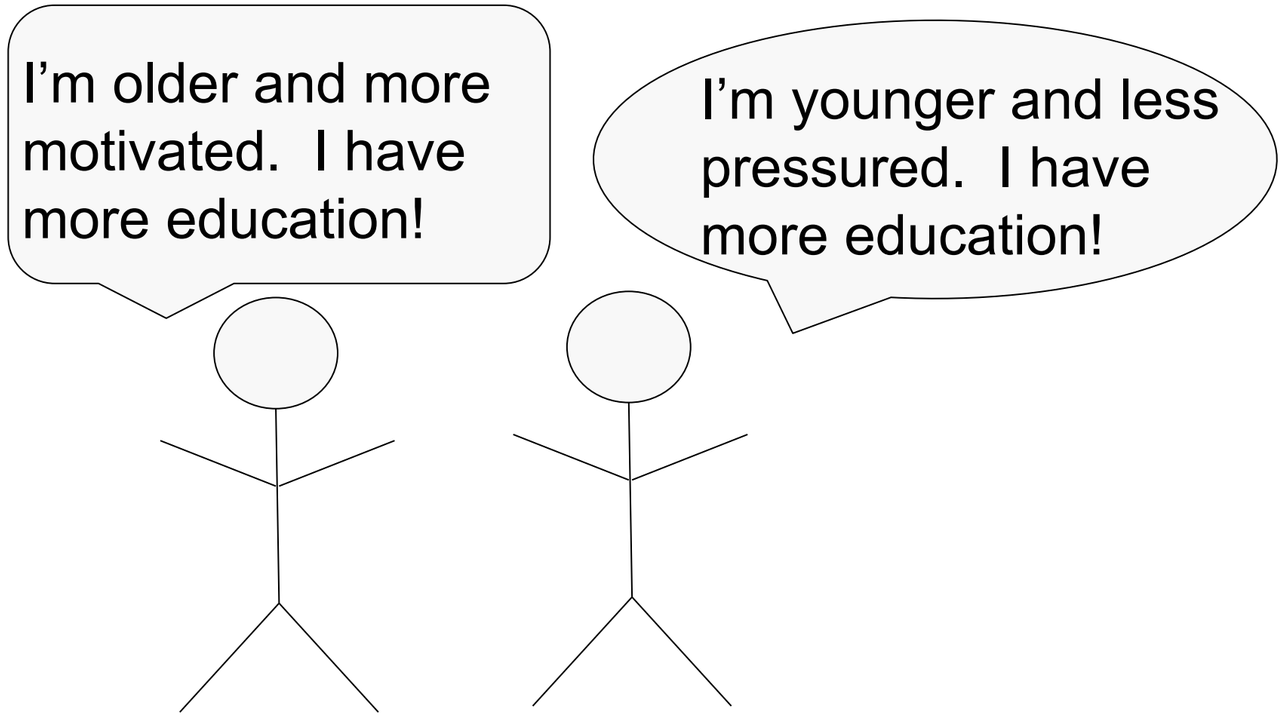
# Research Question

What is the *average* difference between the older twin's and younger twin's self-reported education?



# Research Question

What is the *average* difference between the older twin's and younger twin's self-reported education?



# Research Question

What is the *average* difference between the older twin's and younger twin's self-reported education?

**Population** - All identical twins

# Research Question

What is the *average* difference between the older twin's and younger twin's self-reported education?

**Population** - All identical twins
**Parameter of Interest** - Population mean difference of self-reported education level $\mu_d$

# Research Question

What is the *average* difference between the older twin's and younger twin's self-reported education?

**Population** - All identical twins
**Parameter of Interest** - Population mean difference of self-reported education level $\mu_d$

difference = older - younger

# Research Question

What is the *average* difference between the older twin's and younger twin's self-reported education?

**Population** - All identical twins
**Parameter of Interest** - Population mean difference of self-reported education level $\mu_d$

difference = older - younger

# Research Question

What is the *average* difference between the older twin's and younger twin's self-reported education?

**Population** - All identical twins
**Parameter of Interest** - Population mean difference of self-reported education level $\mu_d$

**Construct a 95% confidence interval for the mean difference of self-reported education for a set of identical twins.**

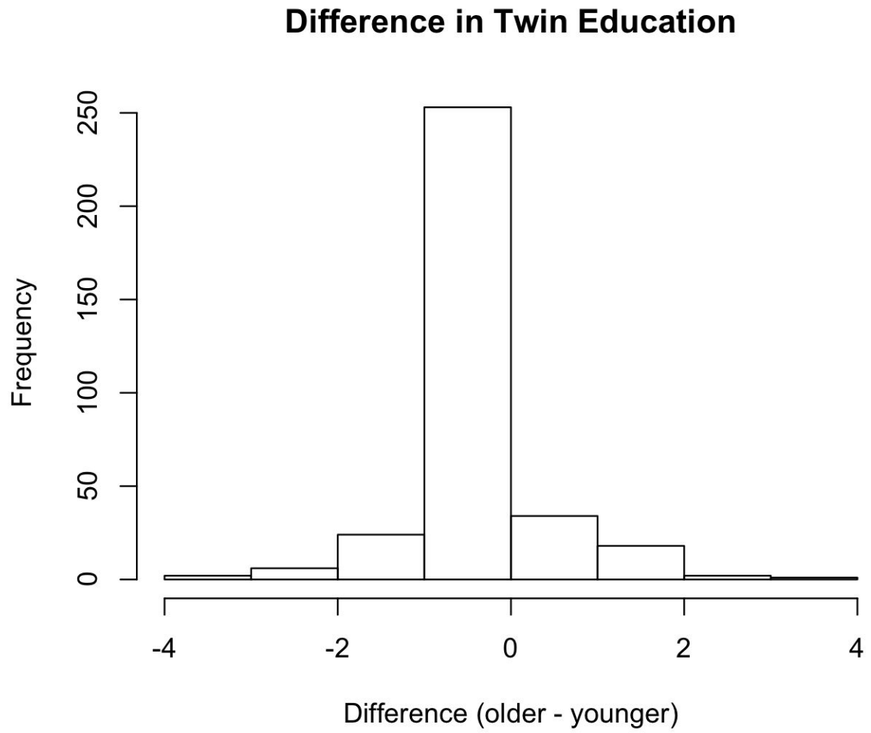
# Difference Calculation

## Difference = older twin - younger twin

| Older twin education | Younger twin education | Difference (older - younger) |
|---|---|---|
| 16 | 16 | 0 |
| 18 | 16 | 2 |
| 12 | 12 | 0 |
| 14 | 14 | 0 |
| 13 | 15 | -2 |

# Difference Summary

## Difference = older twin - younger twin

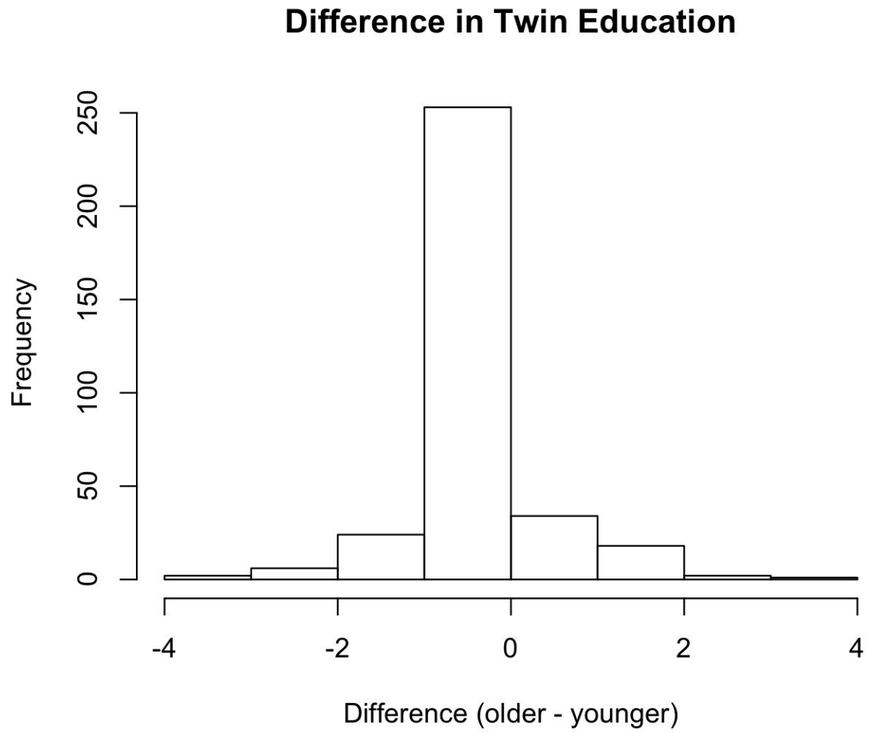



*n* = 340 observations

Minimum = -3.5 years
Maximum = 4 years
72.1% had a difference of 0 years

# Difference Summary

## Difference = older twin - younger twin




*n* = 340 observations

Minimum = -3.5 years
Maximum = 4 years
72.1% had a difference of 0 years

Mean = 0.0838 years
Standard Deviation = 0.7627 years

# Confidence Interval Basics

**Best Estimate ± Margin of Error**

# 95% Confidence Interval Calculations

**Best Estimate ± Margin of Error**

**Sample mean difference ± "a few" · estimated standard error**

$$\bar{x}_d \pm t^* \left(\frac{s_d}{\sqrt{n}}\right)$$

**$t^*$ multiplier comes from a t-distribution with *n – 1* degrees of freedom**

**95% confidence**

**$n = 25 \rightarrow t^* = 2.064$**

**$n = 1000 \rightarrow t^* = 1.962$**

# Mean Difference Confidence Interval

Mean = 0.084 years
Standard Deviation = 0.76 years

*n* = 340 observations ⟶ t* = 1.967

$$\bar{x}_d \pm t^* \left(\frac{s_d}{\sqrt{n}}\right)$$

# Mean Difference Confidence Interval

Mean = 0.084 years
Standard Deviation = 0.76 years

*n* = 340 observations ⟶ t* = 1.967

$$\bar{x}_d \pm t^* \left(\frac{s_d}{\sqrt{n}}\right)$$

$$0.084 \pm 1.967\ (0.76/\sqrt{340})$$

# Mean Difference Confidence Interval

Mean = 0.084 years
Standard Deviation = 0.76 years

*n* = 340 observations ⟶ t* = 1.967

$$\bar{x}_d \pm t^* \left(\frac{s_d}{\sqrt{n}}\right)$$

$0.084 \pm 1.967\ (0.76/\sqrt{340})$

$0.084 \pm 1.967\ (0.04)$

$0.084 \pm 0.0814$

# Mean Difference Confidence Interval

Mean = 0.084 years
Standard Deviation = 0.76 years

*n* = 340 observations ⟶ t* = 1.967

$$\bar{x}_d \pm t^* \left(\frac{s_d}{\sqrt{n}}\right)$$

$0.084 \pm 1.967\ (0.76/\sqrt{340})$

$0.084 \pm 1.967\ (0.04)$

$0.084 \pm 0.0814$

(0.0025, 0.1652) years

# Interpreting the Confidence Interval

**“range of reasonable values for our parameter”**

# Interpreting the Confidence Interval

**“range of reasonable values for our parameter”**

With 95% confidence, the population mean difference of the older twin’s and younger twin’s self-reported education is estimated to be between 0.0025 years and 0.1652 years.

# IVQ

Is there a difference between education levels of the older and younger twin?

# Intervals for Differences

Is there a mean difference between the education level of twins?

If education levels are generally equal → mean difference is **0**

If education levels are unequal → mean difference is not **0**

Look for **0** in the range of reasonable values

# Assumptions

We need to assume that we have a **random sample of identical twin sets**.

**Population of differences is normal** (or a **large enough sample size** can help to bypass this assumption).

# Summary

Extension of the one mean confidence interval

~use difference variable now

# Summary

Extension of the one mean confidence interval

~use difference variable now

Data need to be paired to calculate a difference variable

~two measurements on same individual

~two measurements on similar, matched individuals

# Summary

Extension of the one mean confidence interval

~use difference variable now

Data need to be paired to calculate a difference variable

~two measurements on same individual

~two measurements on similar, matched individuals

0 in the confidence interval

~implies the mean difference is 0 → no true difference