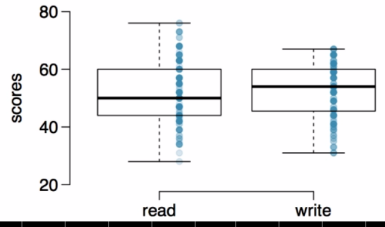
Paired observations

200 observations were randomly sampled from the High School and Beyond survey. Each student took a reading and writing test. Do the average reading and writing test scores differ?



Analyzing paired data

Two sets of observations that have this special correspondence are called paired.

To analyze paired data, analyze the difference.

	id	read	write	diff
1	70	57	52	5
2	86	44	33	11
3	141	63	44	19
4	172	47	52	-5
...	...	...	...	
200	137	63	65	-2

Hypothesis test for paired data

Null hypothesis

Average reading and writing scores are equal

H_0: $\mu_{diff} = 0$

Alternative hypothesis

Average reading and writing scores are different

The t distribution test statistic

The test statistic for inference on a small sample (n < 50) mean is the T statistic with df = n - 1:

$$T_{df} = \frac{\text{point estimate} - \text{null value}}{SE}$$

$$SE = \frac{\text{sample standard deviation}}{\sqrt{\text{sample size}}}$$

Interpretation

p = 0.387 means:

There's a 38.7% chance of observing a difference of 0.545 or more in a sample of 200 students if the true average difference between reading and writing scores is 0

Based on this, we do not reject H_0: the difference of 0.545 is reasonably explained by random chance.

Question

Each student took a reading and writing test, with scores as shown on the right.

- Are the reading and writing scores likely to be independent?
- If not, what type of association would we expect to see?

	id	read	write
1	70	57	52
2	86	44	33
3	141	63	44
4	172	47	52
...	...	...	...
200	137	63	65

Hypothesis test for paired data

Parameter of interest:

Average difference between reading and writing scores of **all** high school students

μ_{diff}

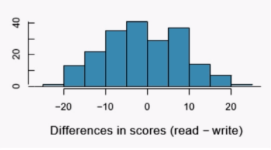
Point estimate:

Average difference between reading and writing scores of sampled high school students

$\bar{x}_{diff}$

Conditions

Less than 10% of all high school students were sampled, and the distribution of differences is:

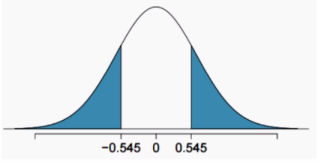


Are the conditions for a hypothesis test satisfied?

- Independence: sample size sufficiently small
- Normality: little apparent skew

Hypothesis test for paired data

200 students were sampled. The observed average difference, $\bar{x}_{diff}$, is -0.545. The standard deviation of the difference is 8.887. Reject H_0?




```
> se <- 8.887 / sqrt(200)
> t <- (-.545 - 0) / se
> df <- 200 - 1
> pt(t, df = df) * 2
[1] 0.3868365
```

Confidence interval paired data

Conf. interval = pt. estimate ± t* × standard error

200 students were sampled. The observed average difference, $\bar{x}_{diff}$, is -0.545. The standard deviation of the difference is 8.887. What is the 95% confidence interval?

```
> se <- 8.887 / sqrt(200)
> df <- 200 - 1
> margin <- qt(.975, df) * se
> -0.545 + c(-margin, +margin)
[1] -1.7841889  0.6941889
```

Confidence interval contains 0
← cannot reject H_0

Comparison with two-sample test

- A paired test is more similar, mathematically, to a one-sample test
- When analyzing paired data, some variation between individuals is cancelled out when the samples are subtracted, reducing SE
- When analyzing independent samples, variation between individuals adds together
- Result: paired studies have greater power