CASE STUDIES IN STATISTICAL THINKING

Introduction to swimming data

Justin Bois
Lecturer, Caltech

The 2015 FINA World Championships

Photo by Chan-Fan, CC-BY-SA-4.0

Strokes at the World Championships

- Freestyle
- Breaststroke
- Butterfly
- Backstroke

Events at the World Championships

- Defined by gender, distance, stroke
- Example: men's 200 m freestyle

Rounds of events

- **Heats**: First round
- **Semifinals**: Penultimate round in some events
- **Finals**: The final round; the winner is champion

Data source

Data are freely available from OMEGA at omegatiming.com

Domain-specific knowledge is

- Imperative
- An absolute pleasure

CASE STUDIES IN STATISTICAL THINKING

Let's practice!

CASE STUDIES IN STATISTICAL THINKING

Do swimmers go faster in the finals?

Justin Bois
Lecturer, Caltech

Michael Phelps's personal bests

Event	Time	Venue	Date	Round
100 m free	47.51	Beijing	2008-08-11	Final
200 m free	1:42.96	Beijing	2008-08-12	Final
400 m free	3:47.79	Indianapolis	2005-04-01	Final
100 m back	53.01	Indianapolis	2007-08-03	Final
200 m back	1:54.65	Indianapolis	2007-08-01	Final
100 m breast	1:02.57	Columbia	2008-02-17	Final
200 m breast	2:11.30	San Antonio	2015-08-10	Final
100 m fly	49.82	Rome	2009-08-01	Final
200 m fly	1:51.51	Rome	2009-29-07	Final
200 m IM	1:54.16	Shanghai	2011-07-28	Final
400 m IM	4:03.84	Beijing	2008-08-10	Final

Sarah Sjöström's personal bests

Event	Time	Venue	Date	Round
50 m free	23.67	Budapest	2017-07-29	Semifinal
100 m free	51.71	Budapest	2017-07-23	Final
200 m free	1.54.08	Rio de Janeiro	2016-08-09	Final
400 m free	4.06.04	Amiens	2014-03-16	Final
50 m back	27.80	Borås	2017-06-30	Final
100 m back	59.98	Eindhoven	2015-04-05	Final

Your question

Do swimmers swim faster in the finals than in other rounds?

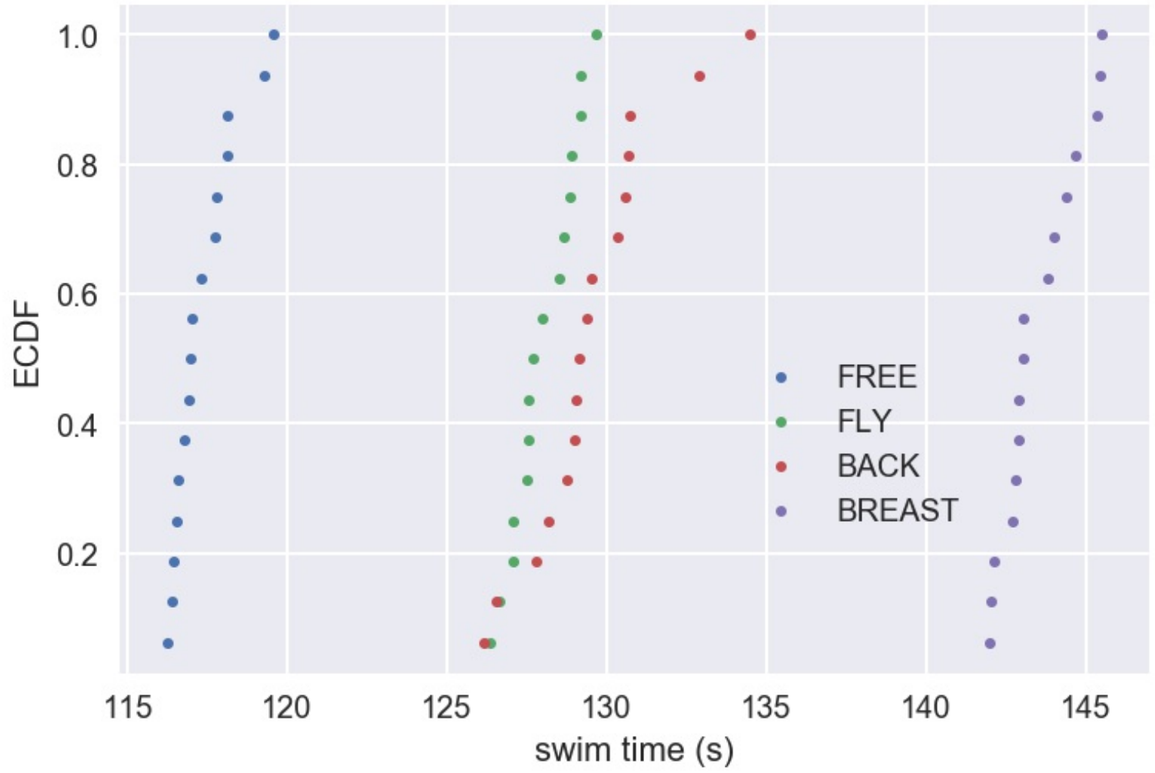
- Individual swimmers, or the whole field?
- Faster than heats? Faster than semifinals?
- For what strokes? For what distances?

Your question

Do individual female swimmers swim faster in the finals compared to the semifinals?

Events: 50, 100, 200 meter freestyle, breststroke, butterfly, backstroke

Diff'rent strokes



Fractional improvement

$$f = \frac{\text{semifinals time} - \text{finals time}}{\text{semifinals time}}$$

Your question(s)

Original question:

- Do swimmers swim faster in the finals than in other rounds?

Sharpened questions:

- What is the fractional improvement of individual female swimmers from the semifinals to the finals?

- Is the observed fractional improvement commensurate with there being no difference in performance in the semifinals and finals?

CASE STUDIES IN STATISTICAL THINKING

Let's practice!

CASE STUDIES IN STATISTICAL THINKING

How does the performance of swimmers decline over long events?

Justin Bois
Lecturer, Caltech

More swimming background

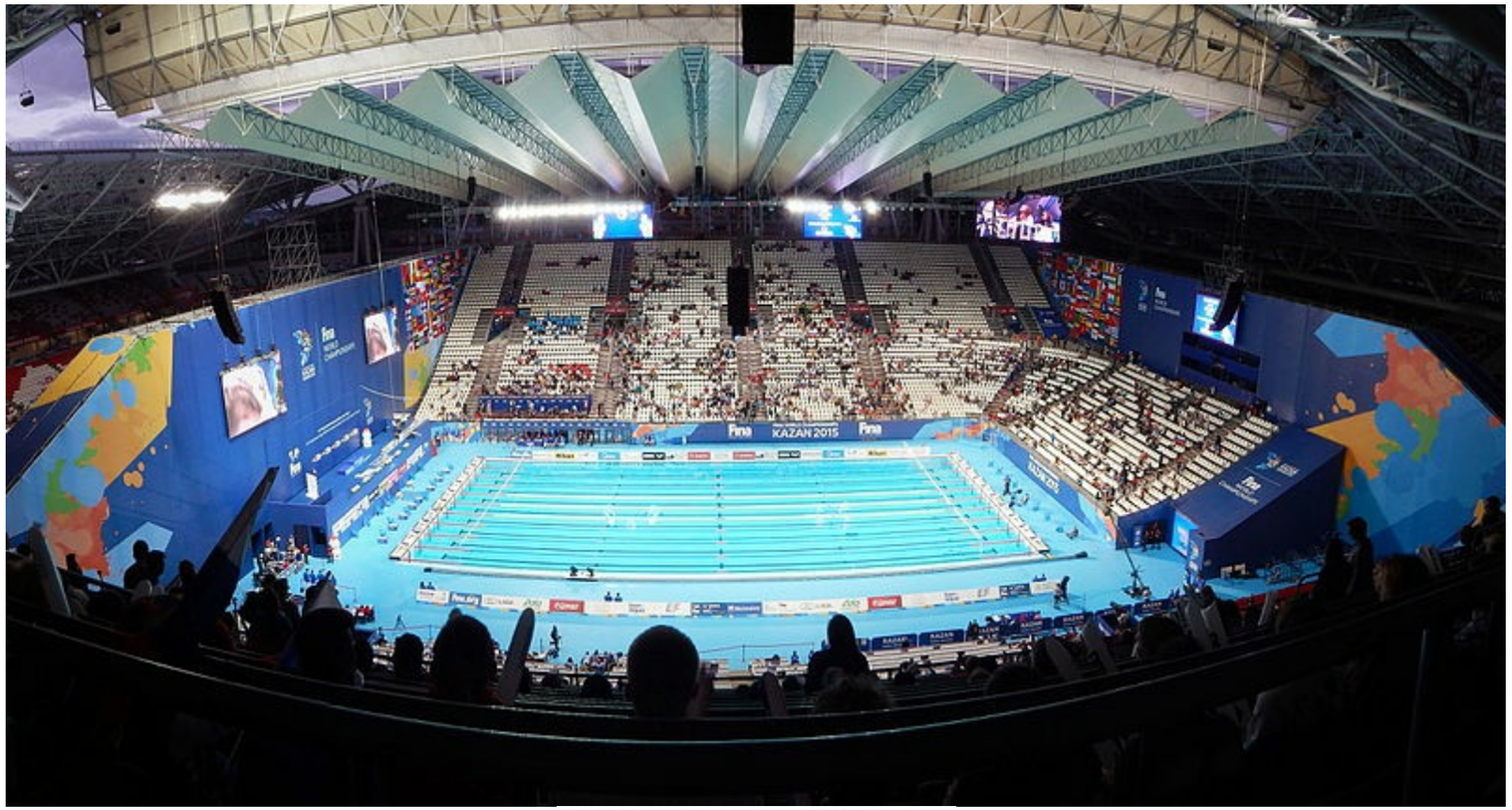
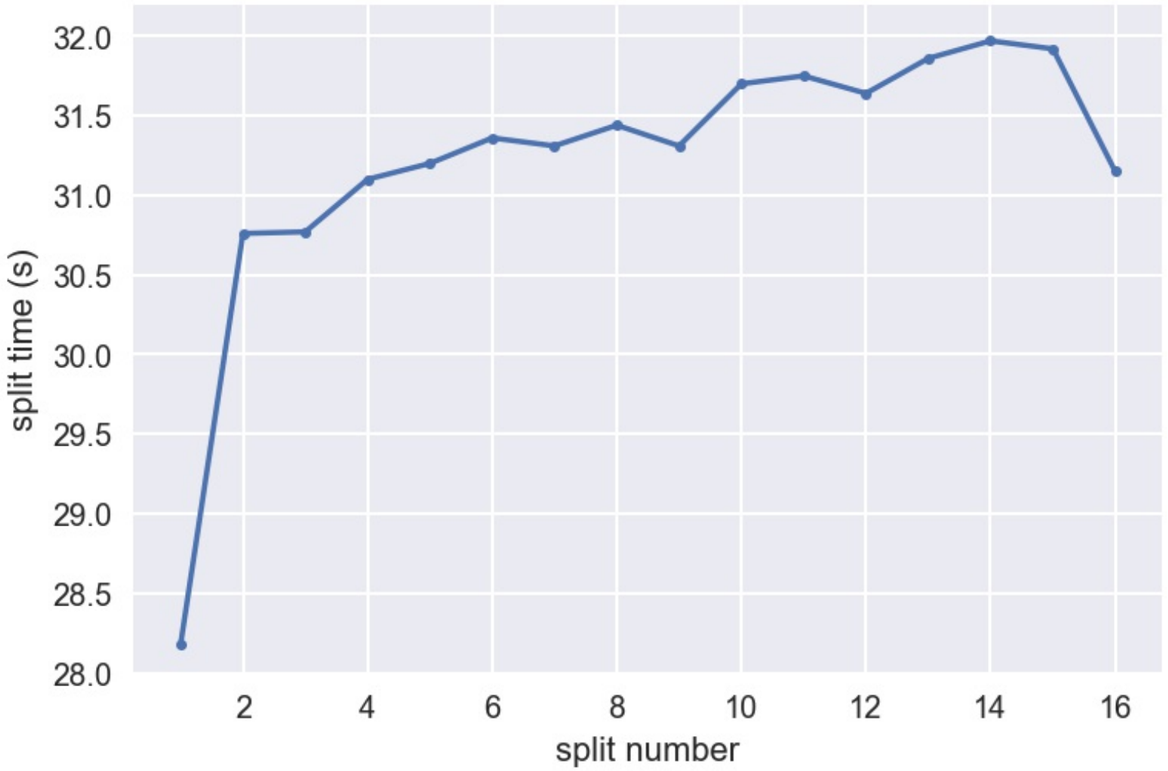


Photo by Chan-Fan, CC-BY-SA-4.0

More swimming background

- **Split**: The time is takes to swim one length of the pool

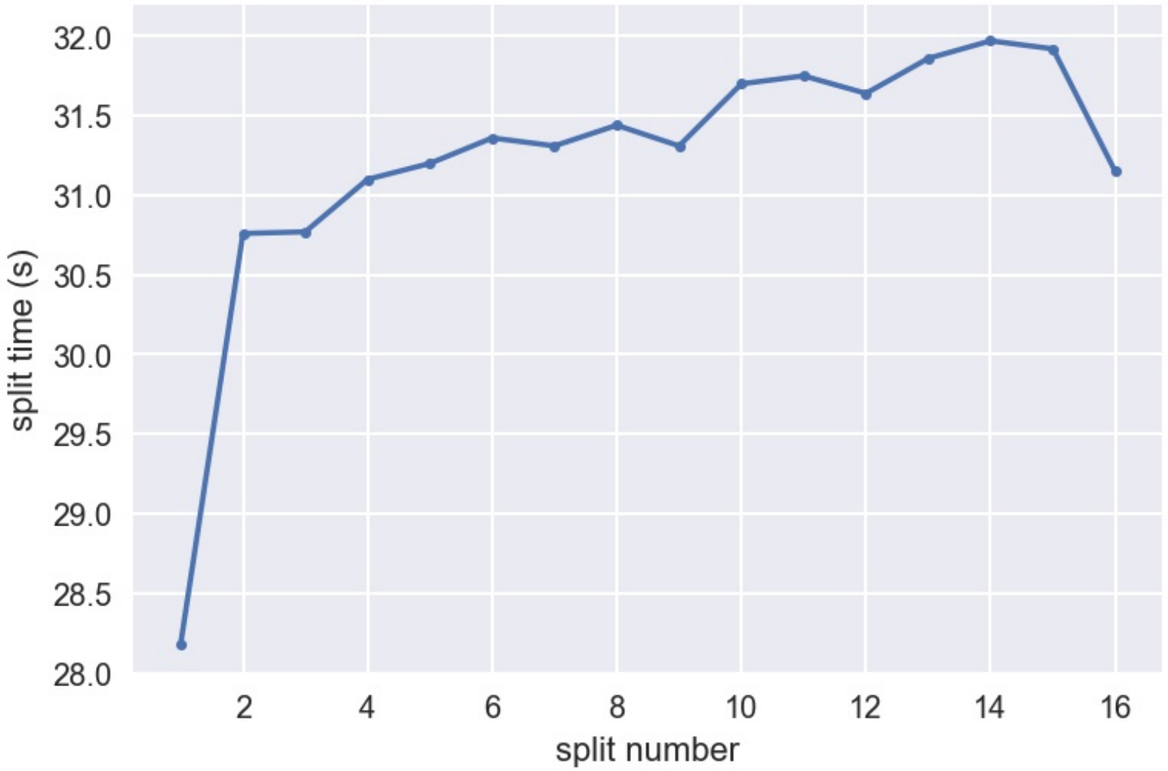
More swimming background



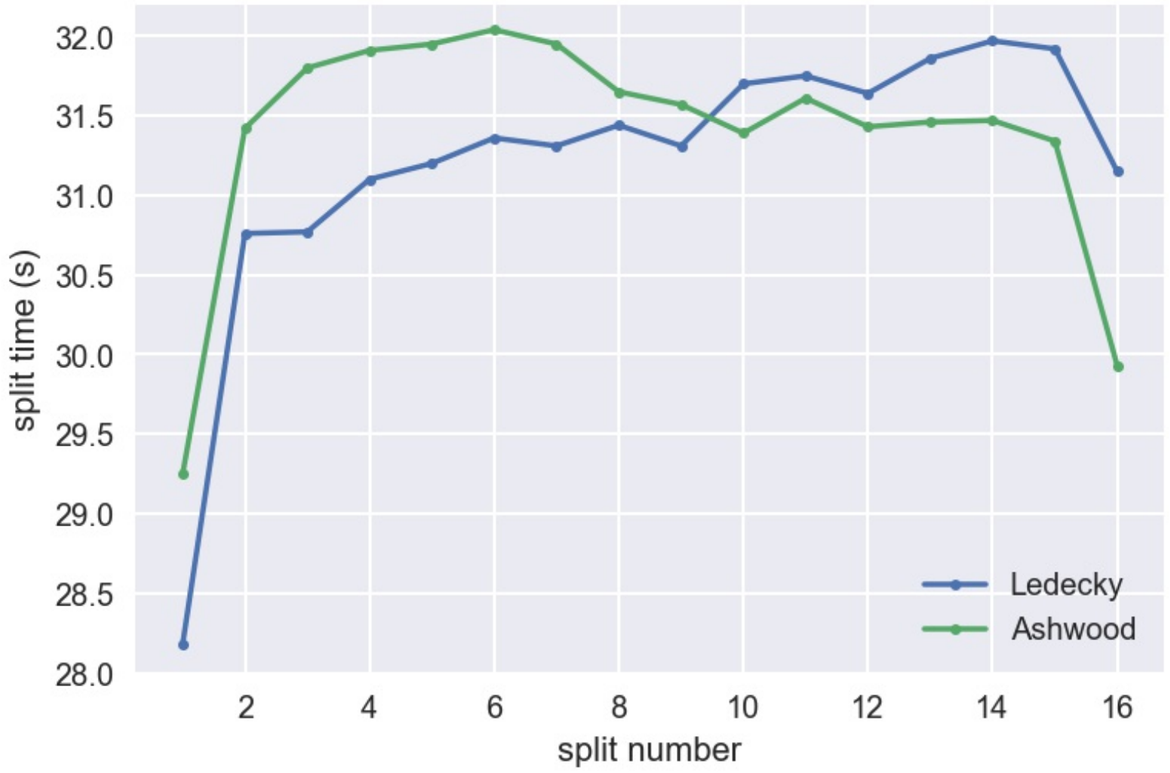
More swimming background

Image: Miho NL, CC-BY-3.0

More swimming background



Slowing down



Quantifying slowdown

- Use women's 800 m freestyle heats
- Omit first and last 100 meters
- Compute mean split time for each split number
- Perform linear regression to get slowdown per split
- Perform hypothesis test: can the slowdown be explained by random variation?

Hypothesis tests for correlation

- Posit null hypothesis: split time and split number are completely uncorrelated
- Simulate data assuming null hypothesis is true

```
scrambled_split_number = np.random.permutation(split_number)
```

- Use Pearson correlation, ρ, as test statistic

```
rho = dcst.pearson_r(scrambled_split_number, splits)
```

- Compute p-value as the fraction of replicates that have ρ at least as large as observed

CASE STUDIES IN STATISTICAL THINKING

Let's practice!