finding things out

problems:

① seeing through randomness

② knowing when you have

solutions

① seeing through randomness:

- ✔ large numbers
- ✔ averaging

② knowing when you have

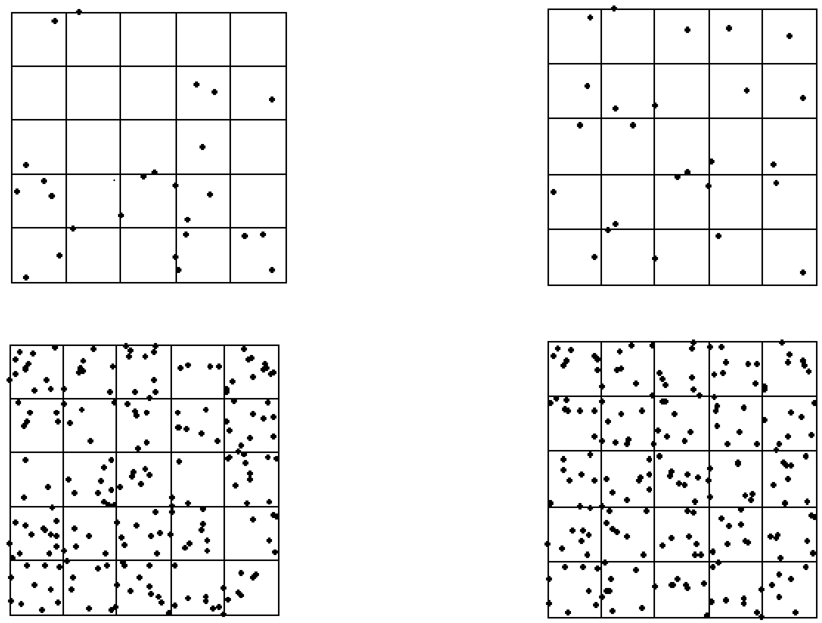
- ✔ measuring variability
- ✔ statistical significance

large numbers

really random | too uniform

25 drops

200 drops



- more drops ⇒ less difference

summing

- add up several races
- some random effects cancel
- overall difference:
  large: 83-76 = 7
- proportionate difference:
  small: 7/83 < 10%
- average scores close:
  8.3–7.6 = 0.7

heads	tails
5	10
10	7
8	10
10	4
10	7
9	10
10	3
10	5
6	10
5	10
83	76

fairness and independence

- fairness:
  each outcome has same probability
  i.e. probability of head = $^1/_2$
- independence:
  each toss has same probability

fairness affects average

- number of heads $\approx$ n $\times$ prob(head)

- but <u>not</u> exact . . .
  . . . the world <u>is</u> very random

unbiased coin

10 series of 20 tosses, prob(head)=0.5:

1:	TTTHHHTHHHTHHHTTHTHH	12
2:	HTTTTHTHHHTHHTTHHTHH	11
3:	HHTTTTTHHTTHHTHHTHTT	9
4:	THTTTHHHHHTHTHHTHHHH	13
5:	HTTTTHTTTHHTHHHHTTTT	8
6:	TTTTHTTTHHHHTTHTTTHH	8
7:	TTTTHHHTTHTHTHTHHTTT	8
8:	THHHTHTTHHTHTHHTHTTH	11
9:	HTTTTTTTTHHTTTHTHHHH	8
10:	HHHTTHHHHTTTHHTHTTTH	11
	average	9.9 $\sigma = 1.9$

biased coin

prob(head) = 0.8:

1:	HHHHHHHHHTHHHHHHHHTH	18
2:	THTHHTHHHHHHHHHHHHHH	17
3:	THHHTHTHHHHHHHHHHHHT	16
4:	HTHHHHHHHHHTHHHHTHHH	17
5:	HHHTHHHHHHHHHHHHTHHH	18
6:	HHHTHHHHTHHTHTHHHHHT	15
7:	HTTTHTHTHHHHHTHTHHHH	13
8:	HHTTHHTHHHHTHHTTTHTH	12
9:	HHHHHHHHHHHHHHHHTHHH	19
10:	HTHHHHHHHHHHHHHTHHHH	18
	average	16.3 $\sigma = 2.3$

independence affects variability

- independence: context doesn't matter

  e.g. prob(head after head) = prob(head after tail)

- positive correlation:

  things vary together

  e.g. prob(head after head) > prob(head after tail)

- negative correlation:

  things vary in opposite way

  e.g. prob(head after head) < prob(head after tail)

positive correlation

p(H) = p(T),   but p(H after H) > p(H after T)

1:	HTTTTTTTTTTTTTTTTTTT	1
2:	TTHHHHHTHHHHHHTHHTTT	13
3:	HHHHHTTHHHHHHTTHTTHH	14
4:	HTHHTTHHHHHHHHHHHHTT	15
5:	HHHHHHTHHHTTTTTTTTTT	9
6:	TTTHTTTTTHTHHHTTHHHH	9
7:	TTTTTTTTTHTTHTTTTTTT	2
8:	HHHHHTTTTTTTTTTTTTHH	7
9:	TTTTTHTHHHHHHHHTHTHH	12
10:	HHHHHHHTTTHHHHHTTHTH	14
	average	9.6 σ = 5.0

- long runs of heads and tails
- high variability of head count

negative correlation

p(H) = p(T), but p(H after H) < p(H after T)

1:	TTHTHHTHHTHTHTHTTHTH	10
2:	TTHHTHTHTHTTHTHHTHHT	10
3:	HTHTTHHTTHTHTHTHTHTH	10
4:	THTHHTTHTHTTHTHTHTTH	9
5:	HTHTHTHTHTHHTTHHTTHT	10
6:	TTHTHHTTHHTHTHTHTHTH	10
7:	TTHHTTHTHTHTHTHTHTHH	10
8:	HTHTHTTHTHHTTHTTHHHT	10
9:	HHHTHHHTHTHTTHHHTHHT	13
10:	THHHHTHTTHTTHHTTHTHT	10
	average	10.2

$\sigma = 1.0$

- alternating heads and tails
- low variability of head count