Inferential Statistics and Probability

Inferential Statistics

- *Population*: a set of examples
- *Sample*: a proper (not whole) subset of a population
- Goal: Estimate some statistic about the population based on statistics about the sample
- Key fact: If the sample is *random*, it tends to exhibit the same properties as the population from which it is drawn

An Example

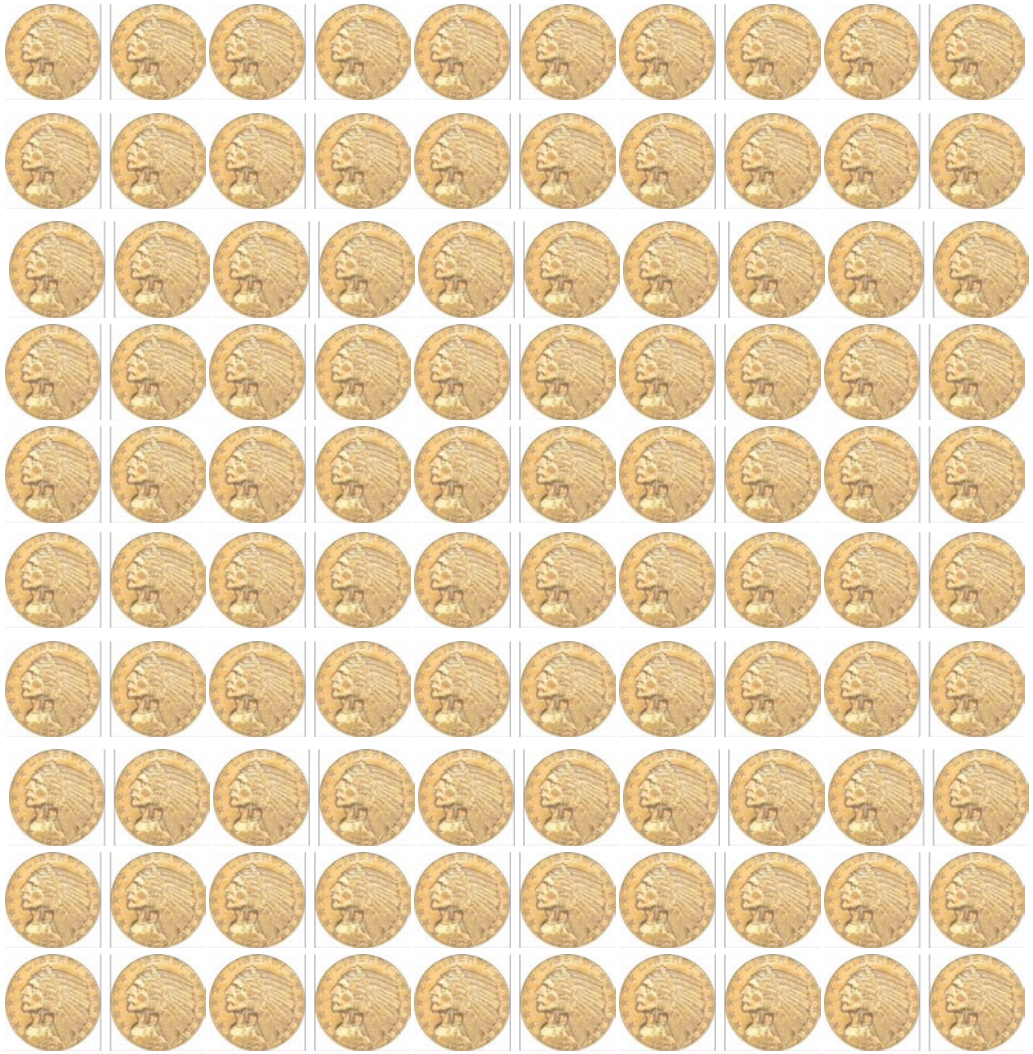
- Given a single coin, estimate fraction of heads you would get if you flipped the coin an infinite number of times

How confident would you be about answering 1.0?

Two Flips

Now confident would you be about answering 1.0?

100 Flips



How about now?

100 Flips

How do you feel about 52/100?

Why the Difference in Confidence?

- Confidence in our estimate depends upon two things
- Size of sample (100 versus 2)
- Variance of sample (all heads versus 52 heads)
- As the variance grows, we need larger samples to have the same degree of confidence

Roulette

CC-BY Antoinetav

10M Spins of the Wheel

```
10000000 spins of Fair Roulette
Expected return betting red = -0.03566%
Expected return betting black = 0.03566%
Expected return betting 2 = 0.11312

10000000 spins of Fair Roulette
Expected return betting red = -0.00342%
Expected return betting black = 0.00342%
Expected return betting 2 = 0.10628

10000000 spins of Fair Roulette
Expected return betting red = -0.07232%
Expected return betting black = 0.07232%
Expected return betting 2 = -0.28792%
```

Law of Large Numbers

- In repeated independent tests with the same actual probability *p* of a particular outcome in each test, the chance that the fraction of times that outcome occurs differs from *p* converges to zero as the number of trials goes to infinity

Bernoulli

the average of the results obtained from a large number of trials should be close to the expected value, and will tend to become closer as more trials are performed.

Gambler's Fallacy

- If deviations from expected behavior occur, these deviations are likely to be evened out by opposite deviations in the future
- Probability of 15 consecutive reds
  - 1/32,378
- Probability of 25 consecutive reds
  - 1/33,554,432
- Probability of 26 consecutive reds
  - 1/67,108,865
- Probability of 26 consecutive reds when previous 25 rolls were red
  - 1/2

Regression to the Mean

- Following an extreme random event, the next random event is likely to be less extreme
- If you spin a fair roulette wheel 10 times and get 100% reds, that is an extreme event (probability = 1/1024)
- It is like that in the next 10 spins, you will get fewer than 10 reds
- So, if you look at the average of the 20 spins, it will be closer to the expected mean of 50% reds than to the 100% you saw in the first 10 spins

Francis Galton, 1885

TABLE I.

NUMBER OF ADULT CHILDREN OF VARIOUS STATURES BORN OF 205 MID-PARENTS OF VARIOUS STATURES.

(All Female heights have been multiplied by 1·08).

Heights of the Mid-parents in inches.	Heights of the Adult Children.													Total Number of		Medians.
	Below	62·2	63·2	64·2	65·2	66·2	67·2	68·2	69·2	70·2	71·2	72·2	73·2	Above	Adult Children.	Mid-parents.	
Above ..	..	..	..	..	..	..	..	..	..	..	..	1	3	..	4	5	..
72·5	..	..	..	..	..	..	..	1	2	1	2	7	2	4	19	6	72·2
71·5	..	..	..	..	1	3	4	3	5	10	4	9	2	2	43	11	69·9
70·5	1	..	1	..	1	1	3	12	18	14	7	4	3	3	68	22	69·5
69·5	..	..	1	16	4	17	27	20	33	25	20	11	4	5	183	41	68·9
68·5	1	..	7	11	16	25	31	34	48	21	18	4	3	..	219	49	68·2
67·5	..	3	5	14	15	36	38	28	38	19	11	4	..	..	211	33	67·6
66·5	..	3	3	5	2	17	17	14	13	4	..	..	..	..	78	20	67·2
65·5	1	..	9	5	7	11	11	7	7	5	2	1	..	..	66	12	66·7
64·5	1	1	4	4	1	5	5	..	2	..	..	..	..	..	23	5	65·8
Below ..	1	..	2	4	1	2	2	1	1	..	..	..	..	..	14	1	..
Totals ..	5	7	32	59	48	117	138	120	167	99	64	41	17	14	928	205	..
Medians ..	..	..	66·3	67·8	67·9	67·7	67·9	68·3	68·5	69·0	69·0	70·0	..	..	..	..	..

NOTE.—In calculating the Medians, the entries have been taken as referring to the middle of the squares in which they stand. The reason why the headings run 62·2, 63·2, &c., instead of 62·5, 63·5, &c., is that the observations are unequally distributed between 62 and 63, 63 and 64, &c., there being a strong bias in favour of integral inches. After careful consideration, I concluded that the headings, as adopted, best satisfied the conditions. This inequality was not apparent in the case of the Mid-parents.

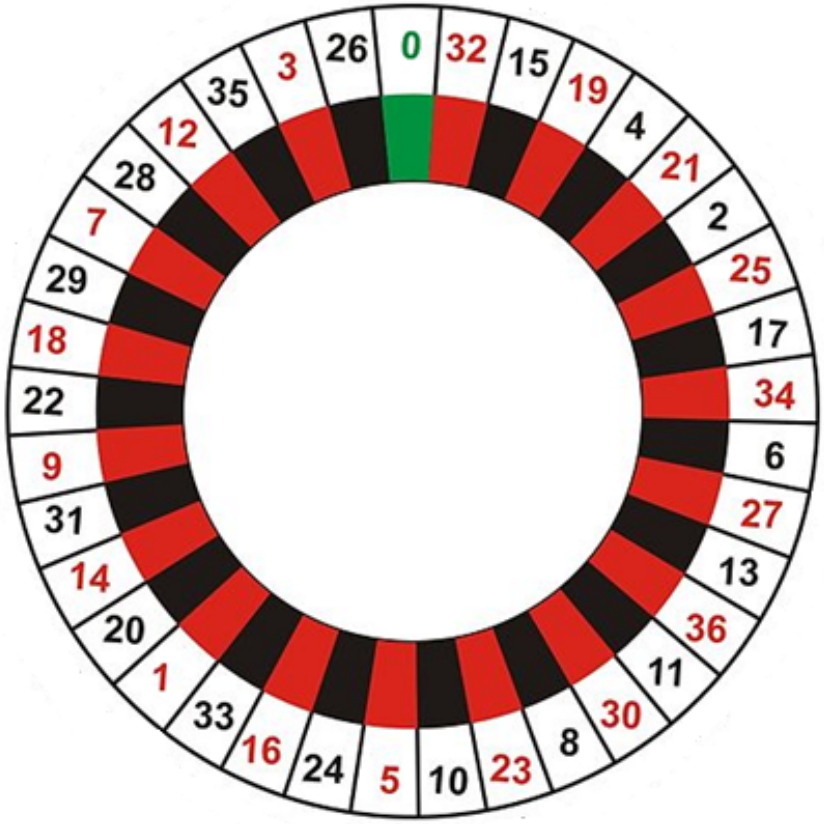
Back to Roulette

```
10000000 spins of Fair Roulette
Expected return betting red = -0.03566%
Expected return betting black = 0.03566%
Expected return betting 2 = 0.11312

10000000 spins of Fair Roulette
Expected return betting red = -0.00342%
Expected return betting black = 0.00342%
Expected return betting 2 = 0.10628

10000000 spins of Fair Roulette
Expected return betting red = -0.07232%
Expected return betting black = 0.07232%
Expected return betting 2 = -0.28792%
```

Casinos Not in the Business of Being Fare



CC-BY Darsie (modified)