Expected Value, **Standard Error**, and a **Normal Approximation**

What does the average from the NTM look like?

Numbered-Ticket Model

… *is like* drawing ____ times from the box ___________ with replacement and averaging the draws.

Example #1

Rolling a die 40 times and averaging the numbers shown

…is like…

drawing ____ times from the box ______________ with replacement and averaging the draws.

Numbered-Ticket Model

… *is like* drawing ____ times from the box ____________ with replacement and averaging the draws.

Example #2

Rolling a die 40 times and computing the proportion of aces

…is like…

drawing ____ times from the box ______________ with replacement and averaging the draws.

Example #1

Rolling a die 40 times and averaging the numbers shown

...is like...

drawing 40 times from the box 1, 2, 3, 4, 5, 6 with replacement and averaging the draws.

If "averaging," use the same numbers

Example #2

Rolling a die 40 times and computing the proportion of aces

...is like...

drawing 40 times from the box 0, 0, 0, 0, 0, 1 with replacement and averaging the draws.

If "computing the proportion," use the INDICATOR TRICK!!

Fact

If we execute the box model, the result is an <u>average</u>.

Numbered-Ticket Model

… *is like* drawing ____ times from the box __________ with replacement and averaging the draws.

Question

What can we say about this (yet to be produced) average?

Question

What can we say about this (yet to be produced) average?

the expected value

the standard error

The average will be about _____ give or take _____ or so.

Suppose an actual list of numbers.

The entries in that list are about [the average] give or take [the SD] or so.

Suppose a hypothetical list of numbers that we generate by executing the NTM an infinite number of times.

The (hypothetical, long-run) average is the expected value.

The (hypothetical, long-run) SD is the standard error.

We can think of the expected value and standard error as a "long-run" average and SD of a chance process.

	actual list of numbers	**chance process**
typical value	average	expected value
give or take	SD	standard error

Equations

$$\textbf{expected value for average} = \textbf{average of tickets in the box}$$

$$\textbf{SE for average} = \frac{\textbf{SD of tickets in the box}}{\sqrt{\textbf{number of draws}}}$$

Helpful Hints

Suppose the tickets are "big-small" so that each ticket is either big B or small S (e.g., the tickets 2, 2, 2, 2, 14, 14), then

$$\textbf{SD of big-small tickets} = (\mathbf{B} - \mathbf{S}) \times \sqrt{(\textbf{fraction that are B}) \times (\textbf{fraction that are S})}$$

Suppose the are "0-1" so that each ticket is either 0 or 1 (e.g., the box 0, 0, 0, 1), then

$$\textbf{SD of 0-1 tickets} = \sqrt{(\textbf{fraction that are 0}) \times (\textbf{fraction that are 1})}$$

Example #1

Rolling a die 40 times and averaging the numbers shown

...is like...

drawing **40** times from the box **1, 2, 3, 4, 5, 6** with replacement and averaging the draws.

the expected value ↓   the standard error ↓

The avg will be about 3.5 give or take _____ or so.

expected value for average = average of tickets in the box

= 3.5

Example #1

Rolling a die 40 times and averaging the numbers shown

...is like...

drawing __40__ times from the box __1, 2, 3, 4, 5, 6__ with replacement and averaging the draws.

the expected value — the standard error

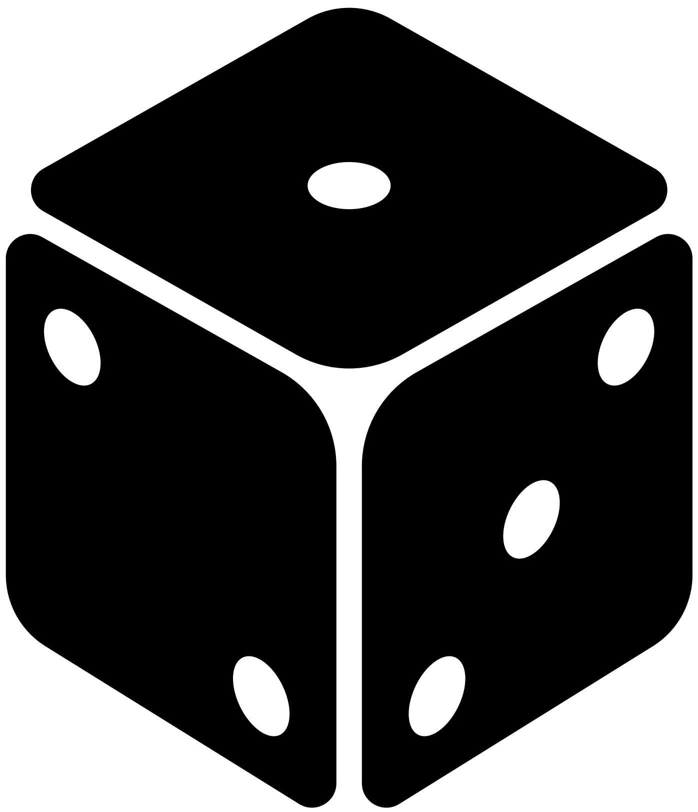
The sum will be about __3.5__ give or take __0.27__ or so.

$$\text{SE for average} = \frac{\text{SD of tickets in the box}}{\sqrt{\text{number of draws}}} = \frac{???}{\sqrt{40}} = \frac{1.71}{6.32} = 0.27$$

```
> x <- c(1, 2, 3, 4, 5, 6)
> sqrt(mean((x - mean(x))^2))
[1] 1.707825
```

I rolled a die 40 times and got an average of 3.53. That's right in line with our claim that the average will be about 3.5 give or take 0.27 or so.

I did it nine more times and got 2.90, 3.43, 3.58, 146, 3.65, 3.15, 3.40, 3.20, and 3.45. Again, that's right in line with our claim that the average will be about 3.5 give or take 0.27 or so.



Example #3

Tossing a coin 100 times and computing the proportion of heads

...is like...

drawing 100 times from the box 0, 1 with replacement and averaging the draws.

the expected value / the standard error

The average will be about 0.5 give or take 0.05 or so.

$$\text{SE for average} = \frac{\text{SD of tickets in the box}}{\sqrt{\text{number of draws}}} = \frac{\sqrt{0.5 \times 0.5}}{\sqrt{100}} = \frac{0.5}{10} = 0.05$$

trick: SD of a 0-1 box = sqrt(frac 1s × frac 0s)

Two things to notice:

1. The SD of a 0-1 box is EASY!
2. The SD of the box {0, 1} (i.e., a single 0 and a single 1) equals 0.5.

the expected value

the standard error

The average will be about ___0.5___ give or take ___0.05___ or so.

0.44 0.46

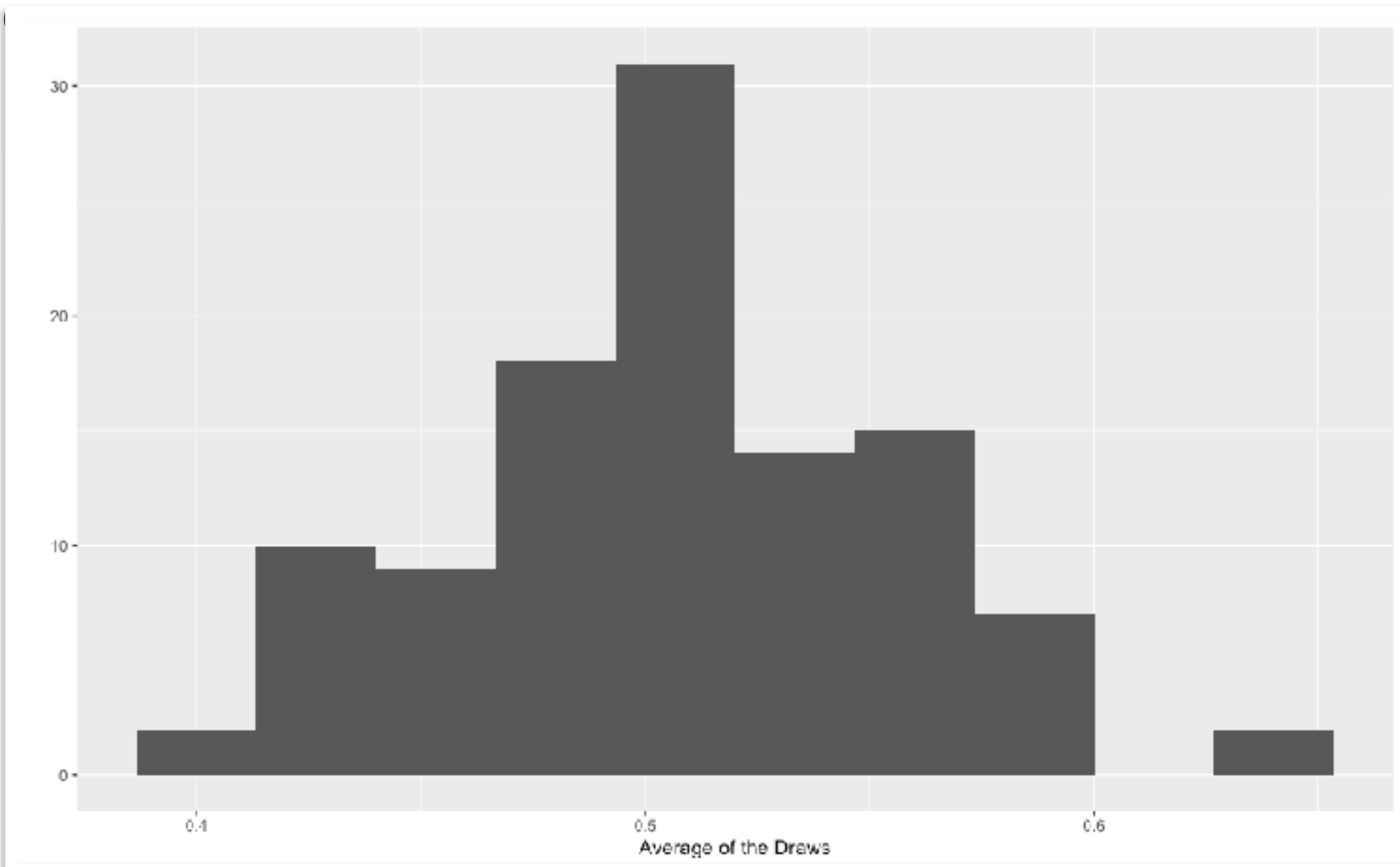
0.53 0.55

0.51 0.49

0.55 0.52

0.50 0.49

0.44 0.48



0.56 0.53

0.48 0.48

0.44 0.45

0.37 0.48

0.42 0.59

0.55 0.48

Highlights, Again

- If we have a chance process, we can sometimes describe it with a **numbered-ticket model**.
- If we have a box model, then we compute the **expected value** and the **standard error** for the sum:

$$\textbf{expected value for average} = \textbf{average of box}$$

and

$$\textbf{SE for average} = \frac{\textbf{SD of box}}{\sqrt{\textbf{number of draws}}}.$$

- With the expected value and standard error, we can fill in the following: **The average will be about ___ give or take ___ or so.**
- The SD of 0-1 box (i.e., a box with only 0s and 1s) has an EASY formula:

$$\textbf{SD of 0-1 box} = \sqrt{(\textbf{fraction that are 0}) \times (\textbf{fraction that are 1})}.$$

- The SD of the box {0, 1} (i.e., a single 0 and a single 1) is $\sqrt{(0.5) \times (0.5)} = 0.5$

Rolling a die 40 times and averaging the numbers shown

...is like...

drawing <u>40</u> times from the box <u>1, 2, 3, 4, 5, 6</u> with replacement and averaging the draws.

Question

What's the chance that the average is more than 4?

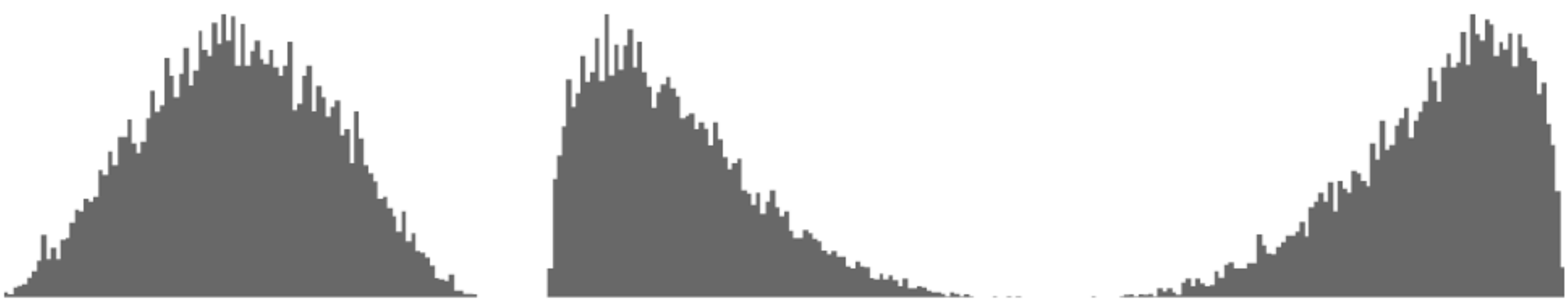
To make things concrete,

let's excuse this box model <u>once</u>.

It’s not clear how to proceed.

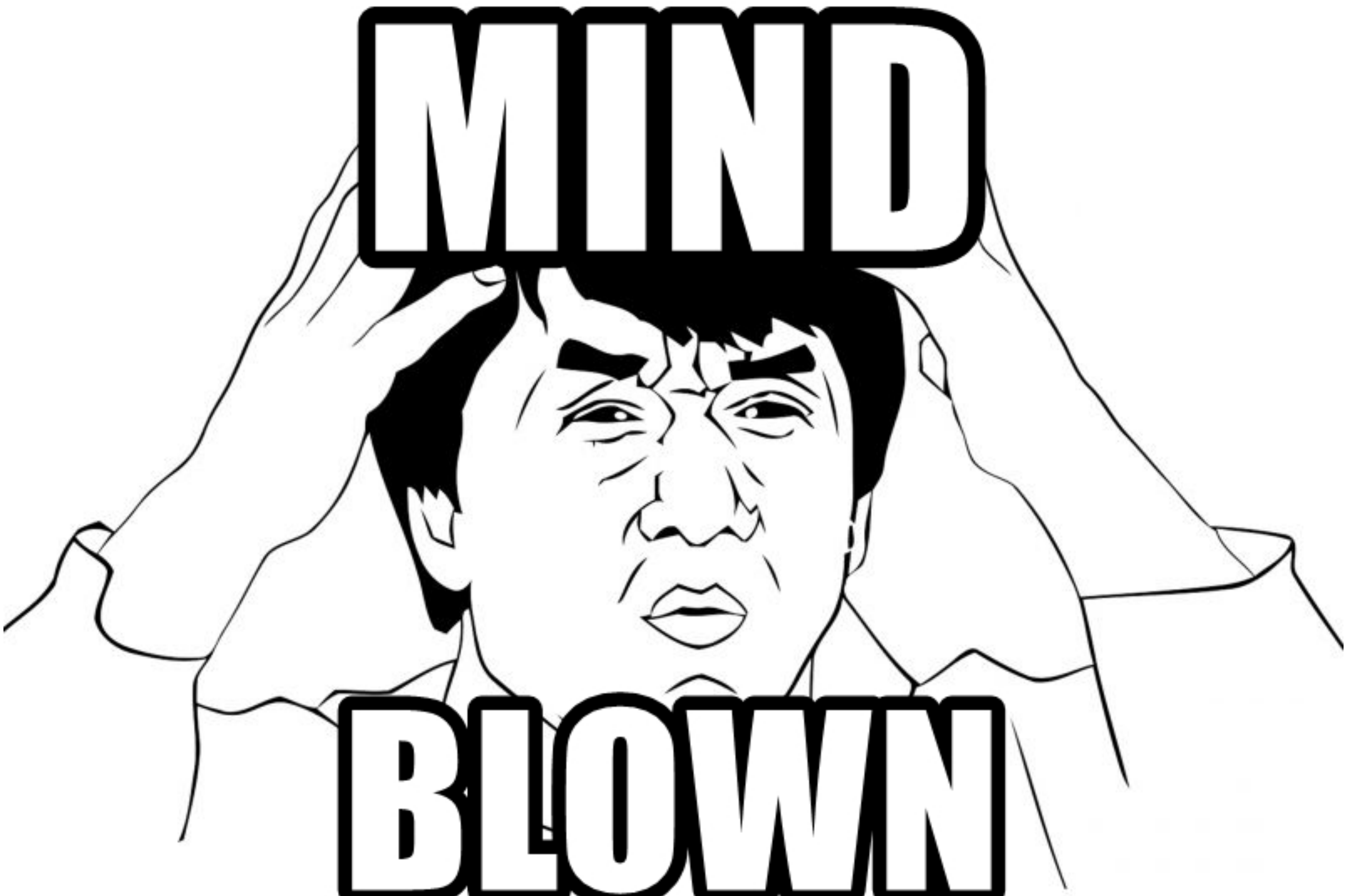
We know the expected value and SE.

But we don’t know the shape of the long-run histogram.



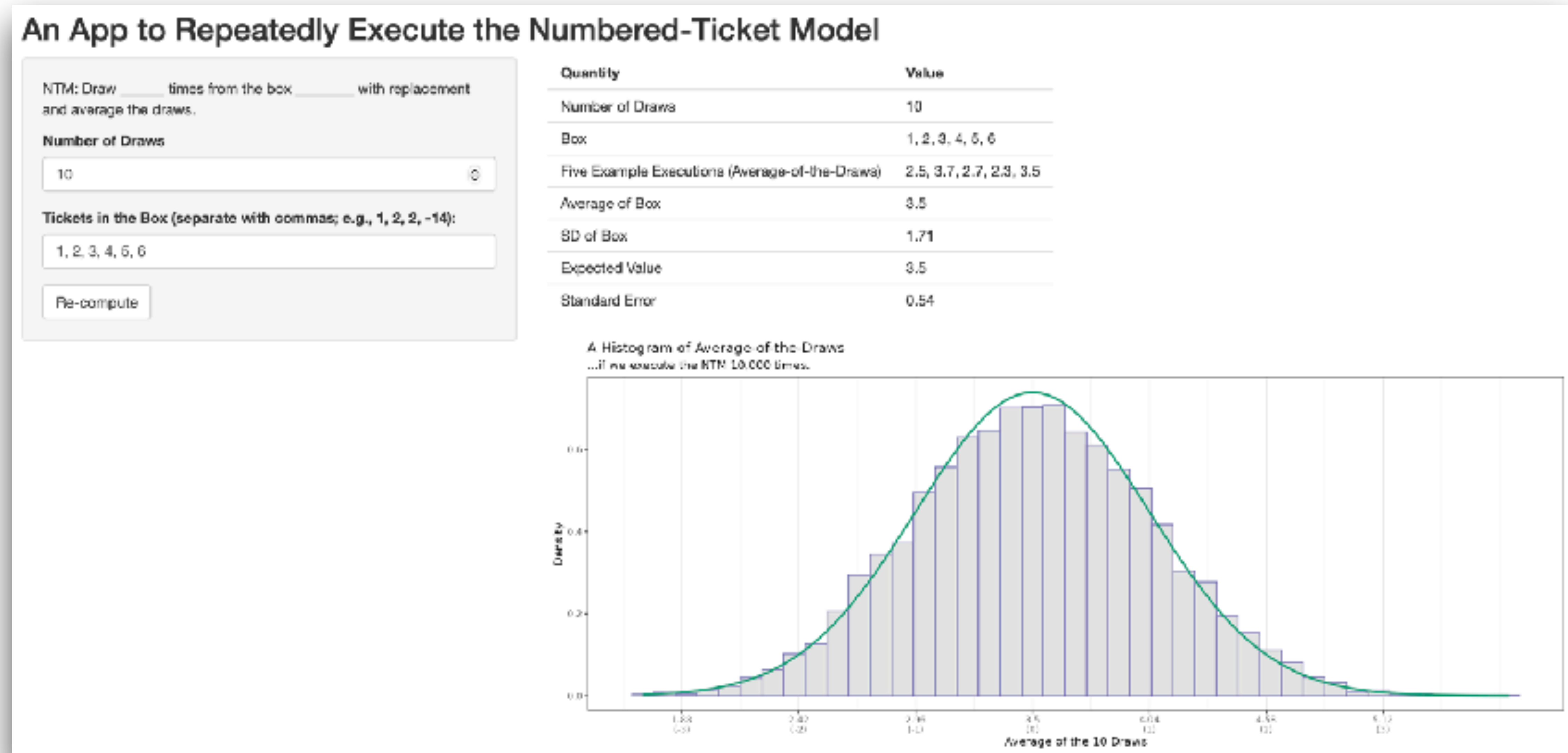
Surely the shape is an “it depends” kinda thing.

As long as the number of draws is sufficiently large,
the average follows the normal curve.



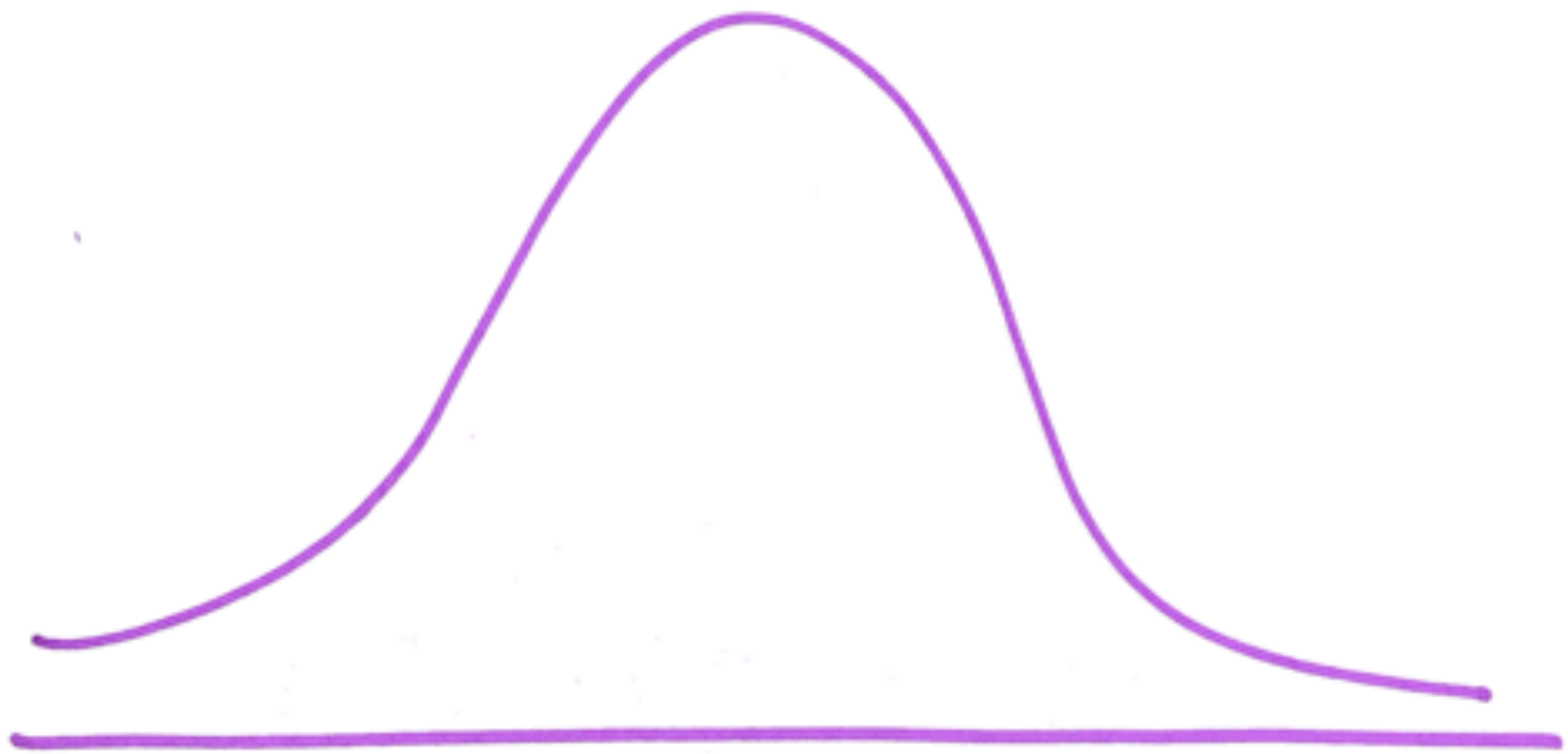
proof by example

Link



Normal Approximation

1. **Draw a picture!**
    i. **bell-shaped curve**
    ii. **label values** (expected value and quantity of interest, usually)
    iii. **shade area of interest**
2. **Convert to standard units (use expected value instead of average and standard error instead of SD).**
3. **Use rules.**
    i. **normal table (p. A-104)**
    ii. **100% rule**
    iii. **symmetry**



Practice Problem I

If I roll a die 50 times, what's the chance I get more than 25% sixes?

Initial guess?

What's the box model?

What's the expected value and standard error?

What's the normal approximation?

draw picture

convert to standard units

use rules

Practice Problem II

Suppose I give an exam with 9 true-false questions. A student isn't well-prepared, so they decide to guess on each.

What's the chance they pass (70% or more)?

Initial guess?

What's the box model?

What's the expected value and standard error?

What's the normal approximation?

draw picture

convert to standard units

use rules

Practice Problem III

What if the exam has 100 true-false questions?

Now what's the chance they pass (70% or more)?

Initial guess?

What's the box model?

What's the expected value and standard error?

What's the normal approximation?

draw picture

convert to standard units

use rules