Probability Histograms
Normal Approximation

Gaston Sanchez

Probability Histogram

Chapter 18

Probability histogram represents chance by area (generally in %)

Empirical Histograms

-vs-

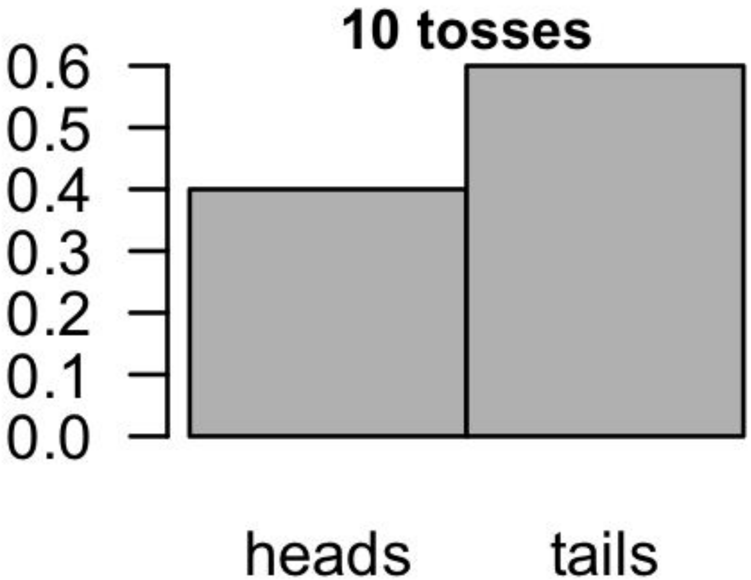
Probability Histograms

Kerrich's Coin Tosses

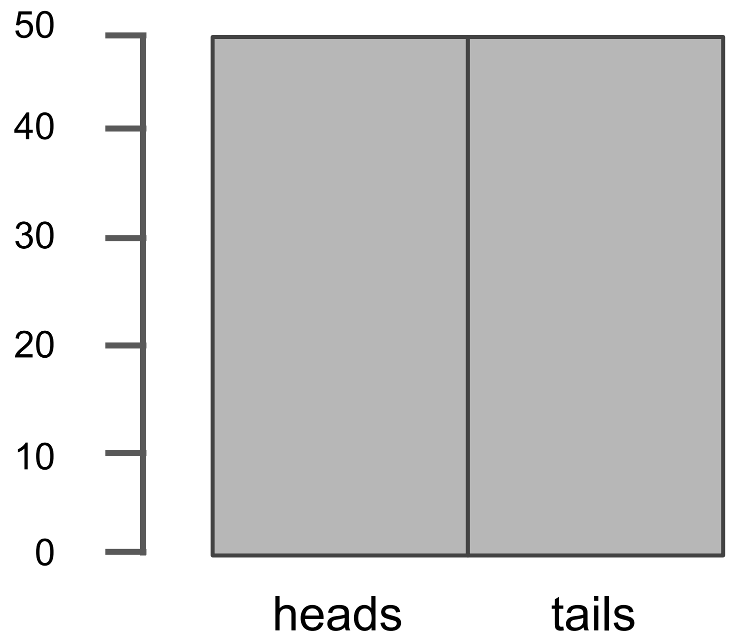
# tosses	# heads	Percent heads	Percent tails
10	4		
100	44		
1000	502		
10000	5067		

Kerrich's Coin Tosses

# tosses	# heads	Percent heads	Percent tails
10	4	40%	60%
100	44	44%	56%
1000	502	50.2%	49.8%
10000	5067	50.67%	49.33%


10 tosses
0.6
0.5
0.4
0.3
0.2
0.1
0.0
heads
tails

Probability Histogram: one coin toss



Demo 1:
roll dice and adding points

Example: pair of dice

Roll a pair of dice, and add up the points

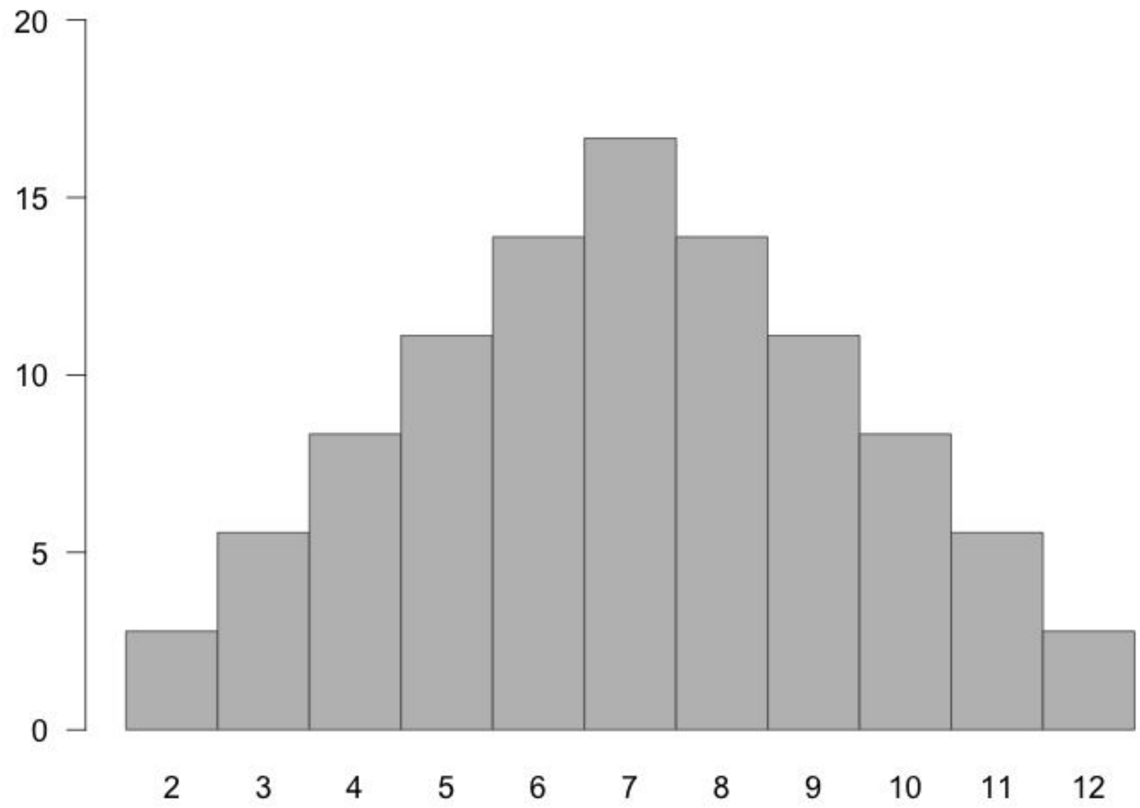
1	2	3	4	5	6

\# of draws = 2

Sum: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12

Probability Histogram

sum	P(sum)
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36



Demo 2:
roll dice and multiplying points

Example

100 coin tosses, find chance of exactly 50 heads

Binomial: $n = 100$, $p = 0.5$, $k = 50$

P(50 heads in 100 tosses) =

$$\binom{100}{50} (\tfrac{1}{2})^{50} (1-\tfrac{1}{2})^{50} = 7.96\%$$

Demo 3:
Tossing coins

Normal Approximation

Example

100 coin tosses, find chance of exactly 50 heads

Binomial: $n = 100$, $p = 0.5$, $k = 50$

EV = 50

SE = 5

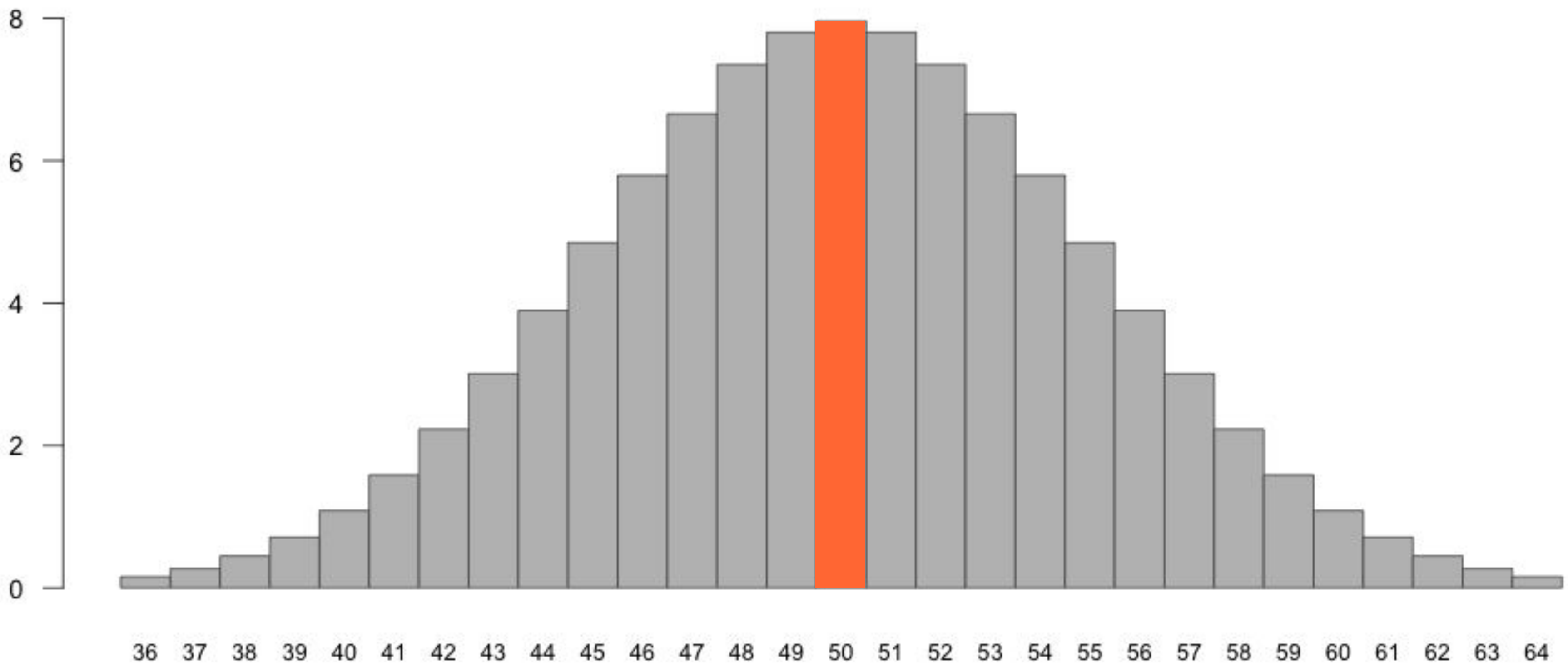
Example

100 coin tosses, find chance of exactly 50 heads

Binomial: $n = 100$, $p = 0.5$, $k = 50$

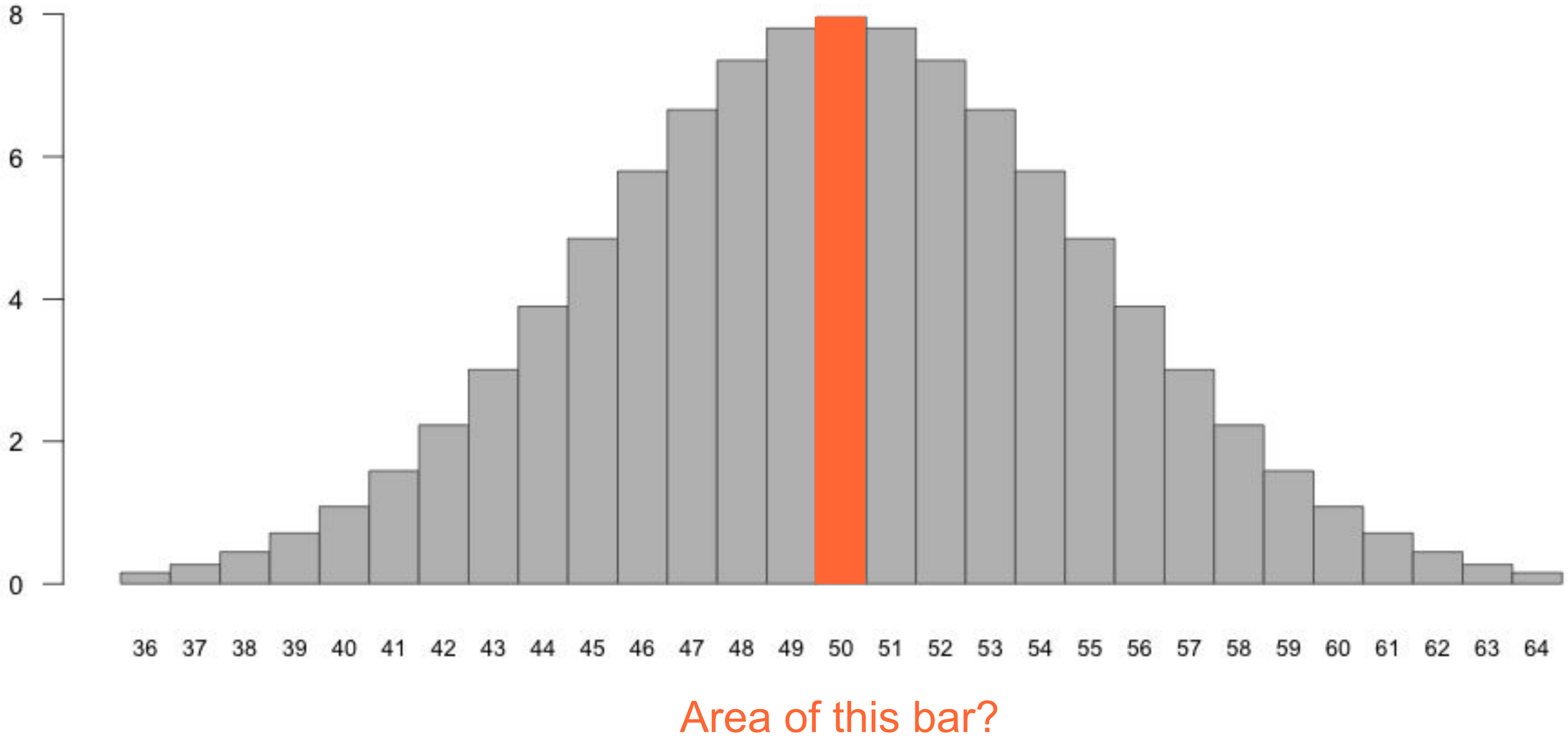
P(50 heads in 100 tosses) =

$$\binom{100}{50} (½)^{50} (1-½)^{50} = 7.96\%$$



Area of this bar?

You can use Binomial Probability or you can use Normal Approximation

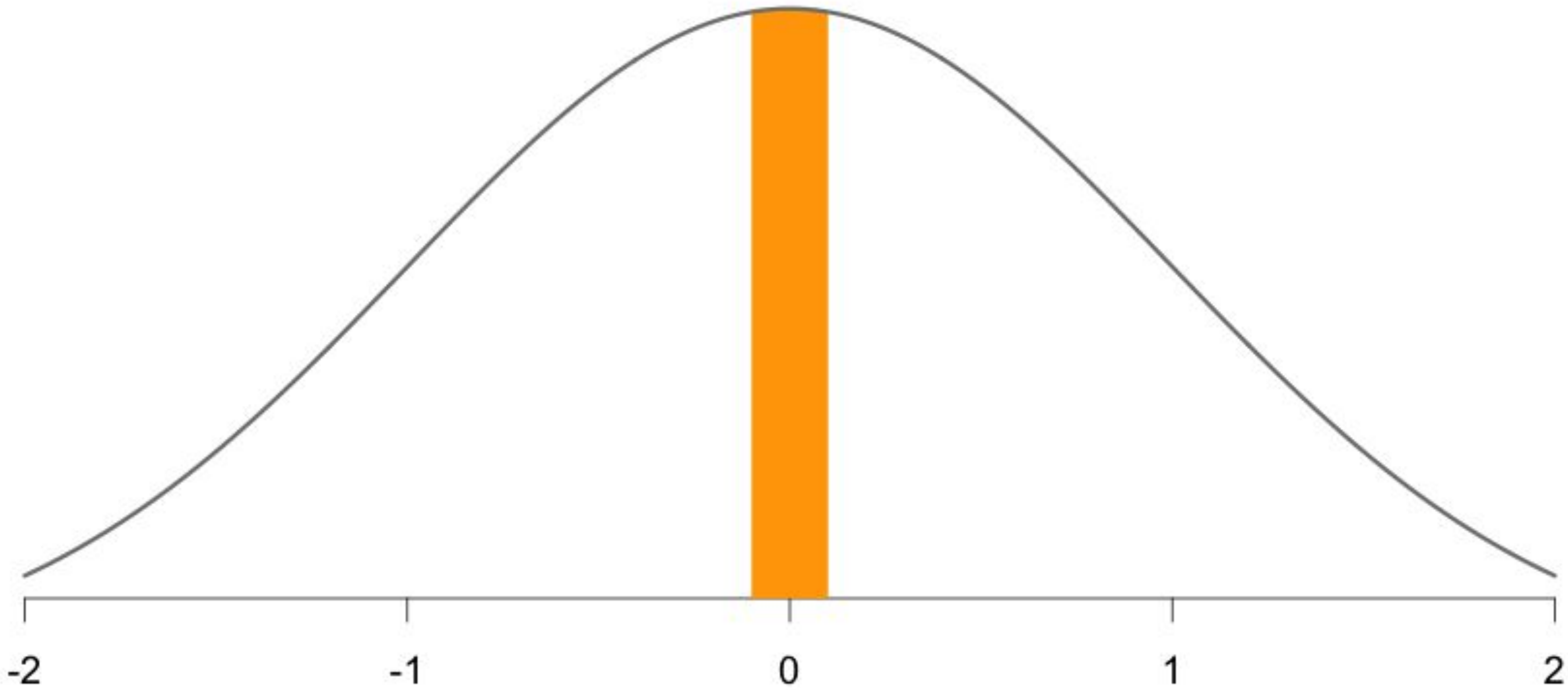


$$SU_1 = \frac{49.5 - 50}{5} = -0.1 \qquad SU_2 = \frac{50.5 - 50}{5} = 0.1$$

$$SU_2 = \frac{50.5 - 50}{5} = 0.1$$

$$SU_1 = \frac{49.5 - 50}{5} = -0.1$$

$$P(-0.1 < Z < 0.1) = ?$$



Area under normal curve between -0.1 and 0.1

= 7.97%

One more example

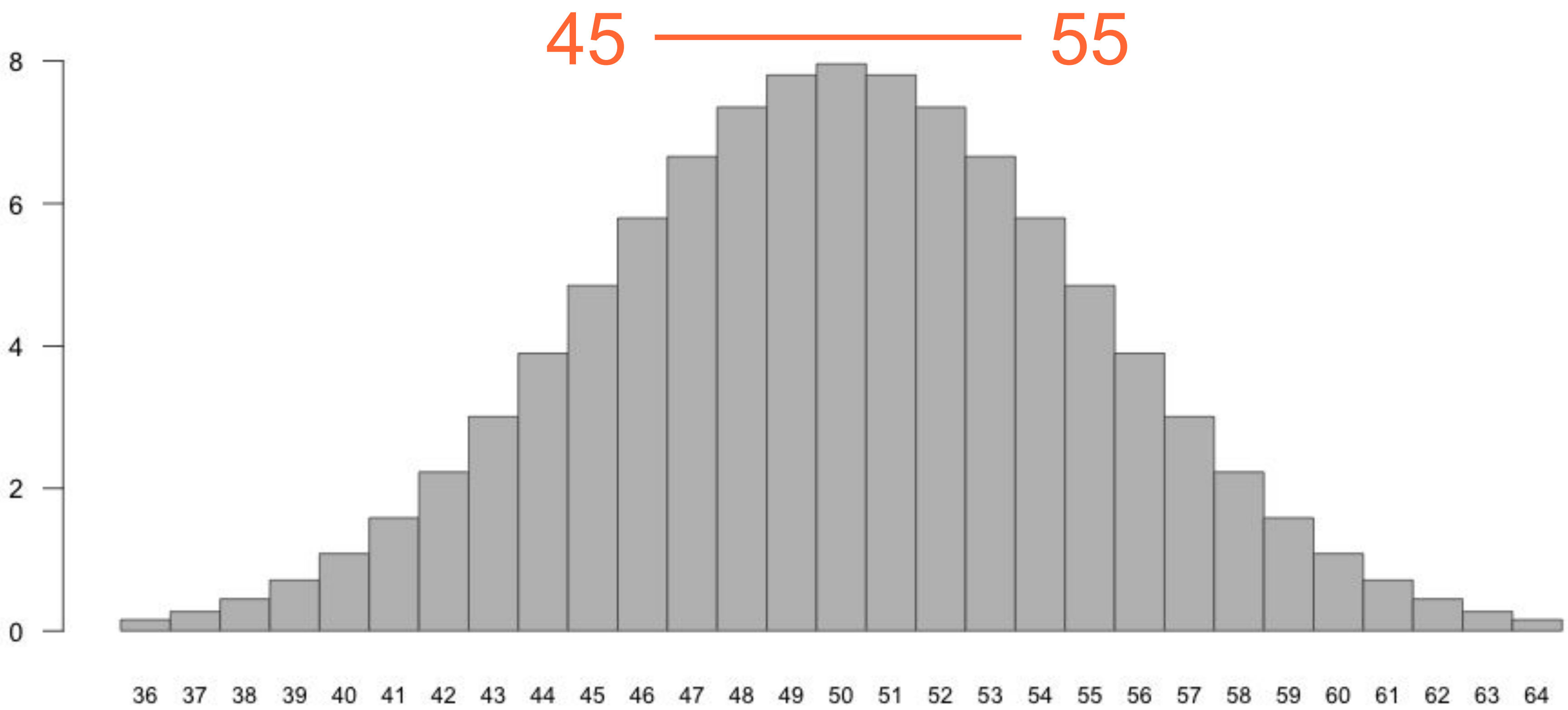
Example

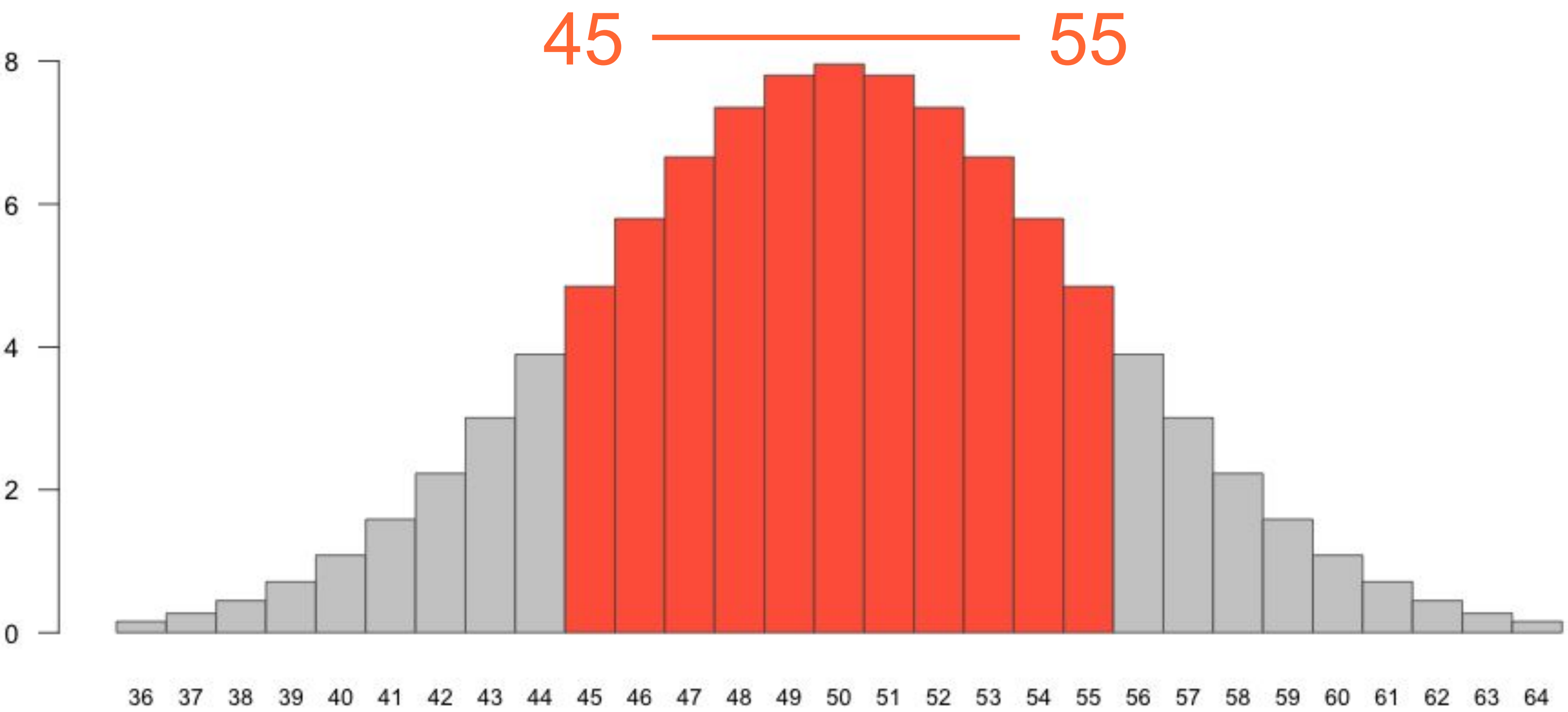
100 coin tosses, find chance of getting 45 to 55 heads inclusive

Binomial: $n = 100$, $p = 0.5$, $k = 50$

$EV = 50$

$SE = 5$

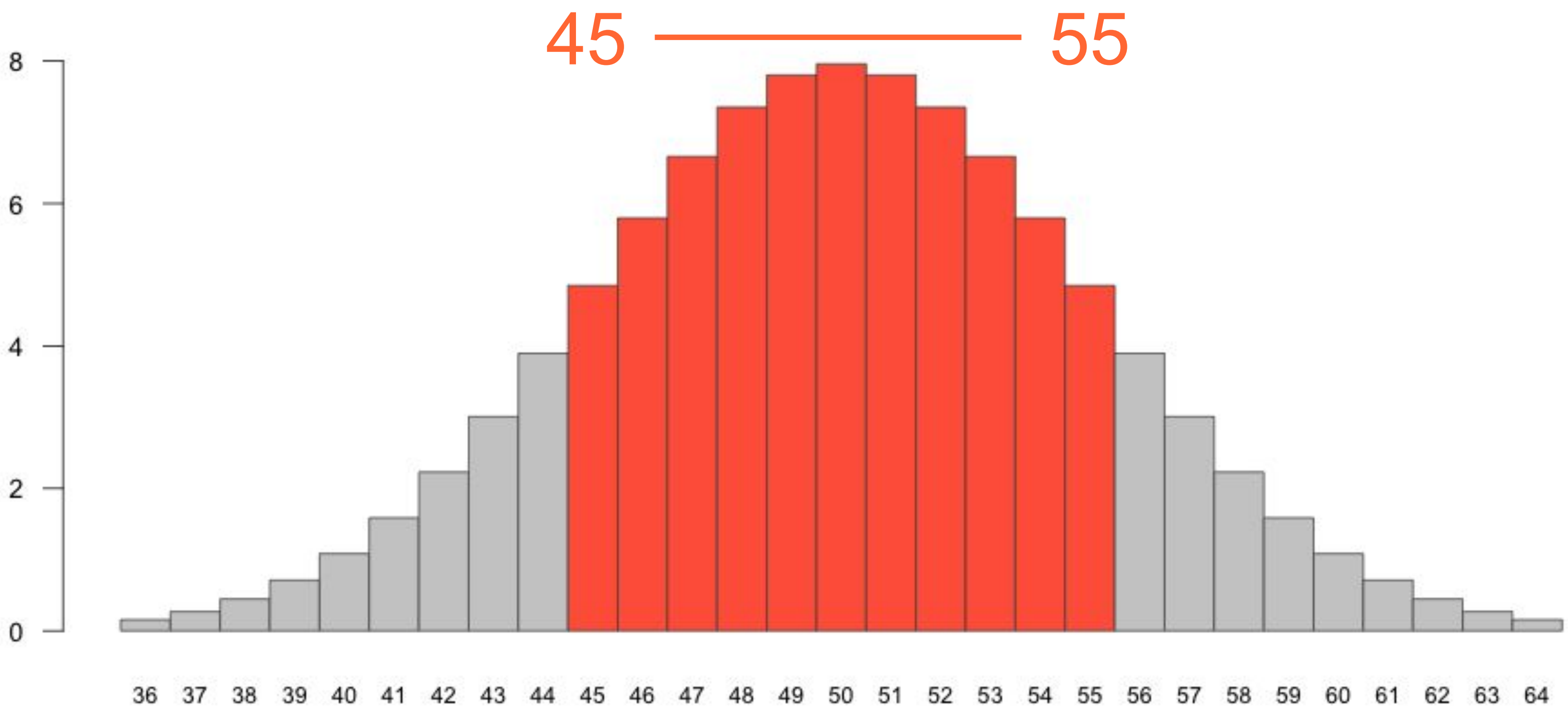

45
55
8
6
4
2
0
36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64



Area of these bars?

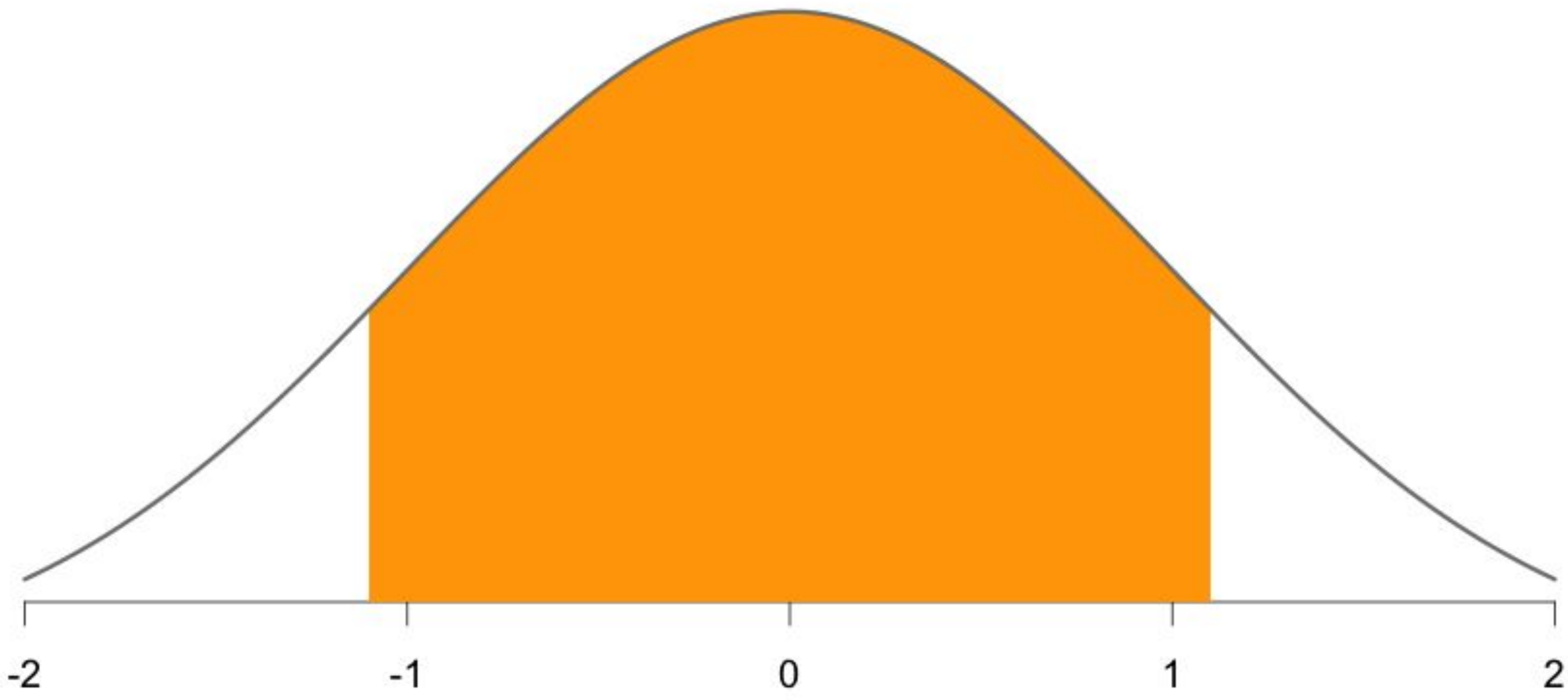
From binomial: 72.8747%

P(45H) + P(46H) + … + P(54H) + P(55H) = 72.8747%



Area of these bars?

$$SU_1 = \frac{44.5 - 50}{5} = -1.1 \qquad SU_2 = \frac{55.5 - 50}{5} = 1.1$$



Area under normal curve between -1.1 and 1.1
= 72.87%

When does Normal Approximation work?

As the number of draws from a box increases, the probability histogram for the sum of draws looks more like the normal curve.

The box does NOT matter ....

What's important is the number of draws

CLT (Central Limit Theorem)

Rule of thumb

When does Normal Approximation work?

1) At least 25 draws

2) EV ± 2SEs is in range of possible values