Probability

Probability

A number that represents the likelihood an event will occur.

Probabiltity vs Proportion

Probability

- Represents the chance of some event occurring
- Theoretical
- Event has not occurred

Proportion

- Summarizes how frequently some event actually occurred
- Empirical
- Event has occurred

Probabiltity vs Proportion Coin Example

If we flip a fair coin, the probability that it lands on heads is 1/2 or 50%.

But if we flip a coin 20 times and count the number of times it lands on heads, lets say 12 times, then our proportion is 12/20 or 60%.

Coin Example

Let's flip some coins!

```

coin_flip <- function(i = NULL) {
      flip <- rbinom(1, 1, 0.5) # (# of random numbers, # of trials, probab
      flip <- ifelse(flip == 1, "Tails", "Heads")
      return(flip)
  }

coin_flip()
```

Card Example

	A	2	3	4	5	6	7	8	9	10	J	Q	K
♢													
♡													
♤													
♧													

Discrete Probability

Discrete Probability

A type of probability that deals with the likelihood an event will occur within a finite (limited) possible outcomes.

Binomials

Trials (i.e., an act with an different outcomes) that have exactly two possible outcomes.

- Coin flips
- Chances of success for a free-throw shooter in basketball, where 1 = a basket made and 0 = a miss.

Binomial probability distribution contains all the possible results over a set of trials and lists the probability of each result.

The Binomial Coefficient and the Probability Distribution

Criteria

1.) Fixed Trials

2.) Independent Trials

3.) Fixed Probability of Success

4.) Two Mutuallly Exclusive Outcomes

The Binomial Coefficient and the Probability Distribution

p(x) = probability of x occurring $p(x) = \binom{n}{x} p^x q^{n-x}$

x = number of successes

n = sample size

p = probability event will occur

q = probability event will not occur

Breakdown of "n choose x"

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

n = sample size
x = number of successes
! = factorial

What is Q?

$$q = (p - 1)$$

p = probability event will occur
q = probability event will **not** occur

Example

Do people prefer Qdoba or Chipotle? Let's say asked three people what they prefer.

Example

But we are conditioning on order here.

$0.5 * 0.5 * 0.5 = 0.125$

$0.5 * 0.5 * 0.5 = 0.125$

$0.5 * 0.5 * 0.5 = 0.125$

$0.125 + 0.125 + 0.125 = ?$

Let's work the formula

Do people choose Qdoba or Chipotle? Let's say asked 3 people what they prefer (with a probability (p) of 0.50 for choosing Qdoba) - 2 people chose Qdoba and 1 chose Chipotle.

What is the probability that people preferred Qdoba?

$$p(x) = (\frac{n!}{x!(n-x)!})p^x(1-p)^{n-x}$$

x = ?

n = ?

p = ?

Let's attack our 'n choose x' first

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

$$\binom{3}{2} = \frac{3!}{2!(3-2)!}$$

$$\frac{6}{2} = 3$$

Let's continue

$$p(x) = (3)p^x q^{n-x}$$

is the probability that Qdoba was chosen 2 out of 3 times.

p^x

Let's continue

$$p(x) = (3)0.5^2(1-p)^{n-x}$$

1.) This is the probability that someone will prefer Chipolte

$$(1 - 0.50) = 0.50$$

2.) This corresponds to the one person that preferred Chipotle

$$(3 - 2) = 1$$

Final Version

Before simplified:

$$p(x) = (\frac{3!}{2!(3-2)!})0.5^2(1-0.5)^{3-2}$$

Simplified:

$$p(x) = (3)(0.25)(0.5)$$

$$p(x) = 0.375$$

Recap

The Binomial Coefficient or 'n choose k'

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

The Binomial Probability Distribution

$$p(x) = \left(\frac{n!}{x!(n-x)!}\right)p^x(1-p)^{n-x}$$

Now that we covered the surface of binomial distributions, let's dig a little deeper, with multiple binomial coefficient calculations.

Lets review exponent rules

Exponent Rules For $a \neq 0, b \neq 0$	
Product Rule	$a^x \times a^y = a^{x+y}$
Quotient Rule	$a^x \div a^y = a^{x-y}$
Power Rule	$\left(a^x\right)^y = a^{xy}$
Power of a Product Rule	$(ab)^x = a^x b^x$
Power of a Fraction Rule	$\left(\frac{a}{b}\right)^x = \frac{a^x}{b^x}$
Zero Exponent	$a^0 = 1$
Negative Exponent	$a^{-x} = \frac{1}{a^x}$
Fractional Exponent	$a^{\frac{x}{y}} = \sqrt[y]{a^x}$

Example 2

Let's imagine there are 4 Formula-1 races that will occur in June and July, Max Verstappen of Red Bull Racing has a 60% chance of winning.

Assuming that the races are independent of each other, what is the probability that he will win 0 races, 1 race, 2 races, 3 races, or all 4 races?

$$p(x) = (\frac{n!}{x!(n-x)!})p^x(1-p)^{n-x}$$

Let's break this down with 4 steps

Step 1

What do we know?

n = 4 races

x = He will win: 0, 1, 2, 3, 4 races

p = 0.60

- Since x has multiple outcomes, we must calculate the binomial coefficient for each.

STEP 2

Calculate 'n choose x' for each x

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

- 0
- 1 $\binom{4}{0} = \frac{4!}{0!(4-0)!} = 1$
- 2 $\binom{4}{1} = \frac{4!}{1!(4-1)!} = 4$
- 3 $\binom{4}{2} = \frac{4!}{2!(4-2)!} = 6$
- 4 $\binom{4}{3} = \frac{4!}{3!(4-3)!} = 4$

$\binom{4}{4} = \frac{4!}{4!(4-4)!} = 1$

Step 3

Plug in and solve the rest

$(Step\ 2)\ p^x q^{n-x}$

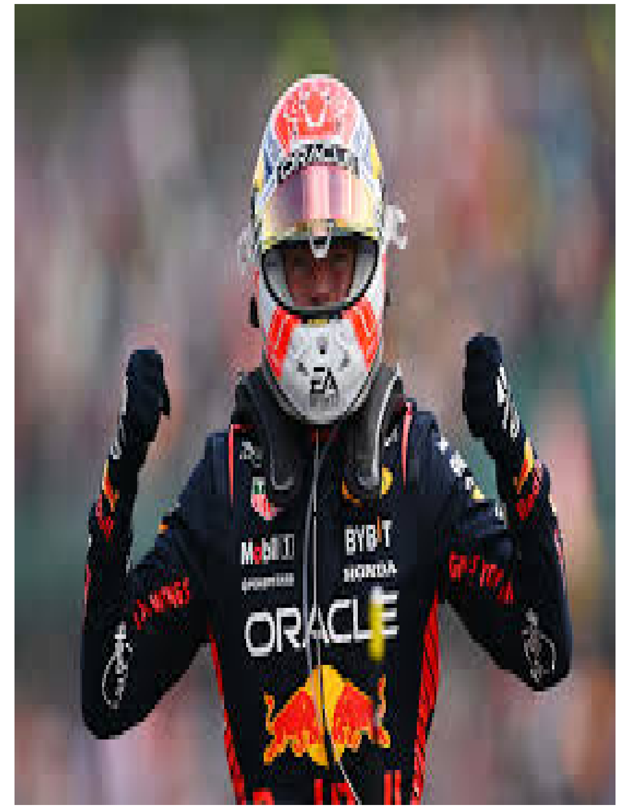
- $(1)0.60^0(1-0.60)^{4-0} = 0.0256$
- $(4)0.60^1(1-0.60)^{4-1} = 0.1536$
- $(6)0.60^2(1-0.60)^{4-2} = 0.3456$
- $(4)0.60^3(1-0.60)^{4-3} = 0.3456$
- $(1)0.60^4(1-0.60)^{4-4} = 0.1296$

STEP 4

Summarize the result

Max Verstappen has a ...

2.56% chance to win 0 races.
15.36% chance to win 1 races.
34.56% chance to win 2 races.
34.56% chance to win 3 races.
12.96% chance to win 4 races.

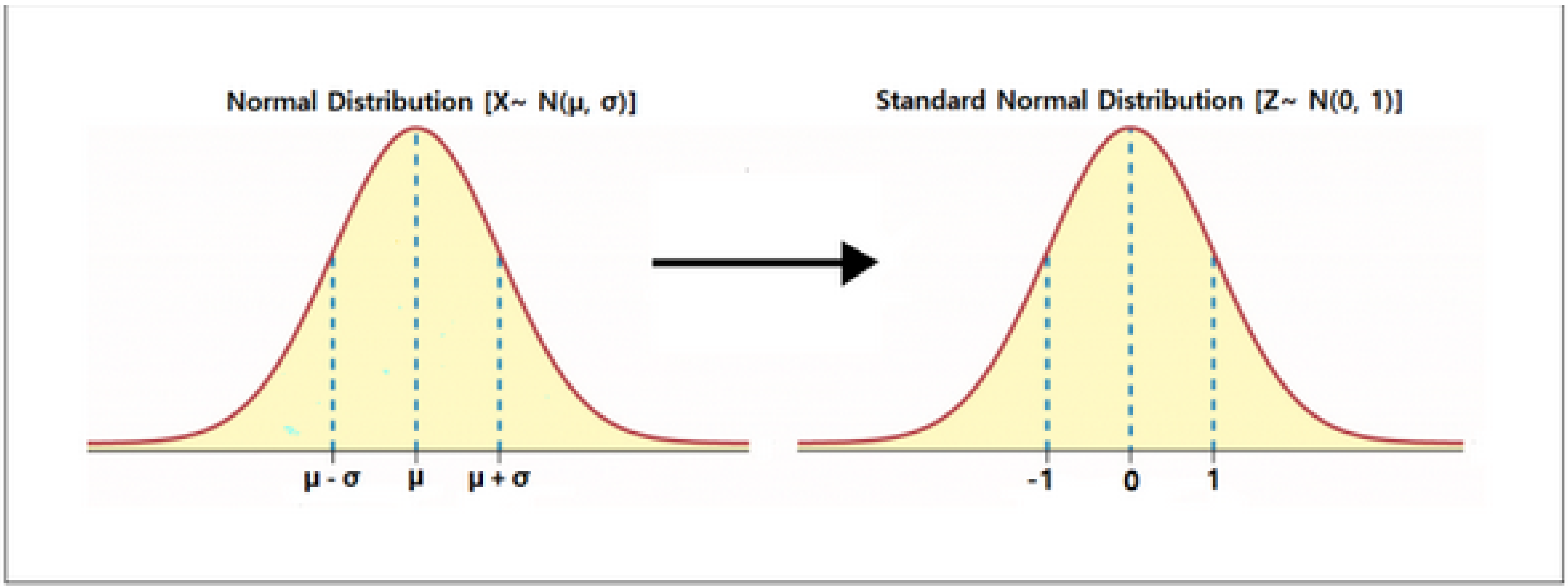


Continuous Probability: The Standard Normal Curve

Standard Normal Curve

Sometimes our distribution can have a lot of variance and thus can result in varying levels of kurtosis.

To compensate for this, we can standardize our scores by apply a z-score.



Z-SCORE

Tells you how many standard deviations a specific point is away from the mean of a distribution.

Positive z-score: A positive z-score indicates that the value (x) lies above the mean by a certain number of standard deviations.

Negative z-score: A negative z-score indicates that the value (x) lies below the mean by a certain number of standard deviations.

Zero z-score: A z-score of 0 means the value (x) is exactly equal to the mean of the data set.

Z-Score

$$z = \frac{x - \bar{x}}{\sigma}$$

= Raw score

x= Sample mean

$\bar{x}$= Standard deviation

σ

Z-Score Example

Imagine you have a class of 20 students and you give them a math test. The mean was 75 points, and the standard deviation (SD) is (+/-) 10 points.

One student scored 88 points on the test. Calculate the z-score for the student.

$$z = \frac{x - \bar{x}}{\sigma}$$

Z-Score Example

$$z = \frac{88 - 75}{10}$$

$$z = 1.30$$

The score is 1.30 standard deviations above the mean.

In other words, the student preformed better than 1.30 standard deviations compared to the average score in the class.

Have a Great Weekend!