

BINOMIAL DISTRIBUTION

Assume that we have two flavours of lays in the market

Red Lays (Tomato,chilli flavour) Green Lays (Cream and onion flavour)

soo which flavour is more loved by the people? or are they both eually loved?

To know this we can conduct a survey by asking some random people which flavour they prefer

CASE 1 :

Let say everybody prefered RED Lays and one person prefered GREEN Lays, then its obvious that RED Lays is much prefered flavour

CASE 2:

But what if 4 people prefered RED Lays and 3 prefered GREEN Lays then we dont have enough confidence to say that RED Lays is the most prefered Lays.

same way what if 3 people prefered RED Lays and 4 prefered GREEN Lays then we dont have enough confidence to say that Green Lays is the most prefered Lays either.

or it could be like this few people may not have preference and the result we got are due to some small sample size

- To understand this we must know what exatly we need to expect if there is no preference
- Then we determine if our survey results fits into that expectations if not we can clearly reject the idea of both lays are equally loved

BINOMIAL DISTRIBUTION tell us what to expect if no preference

- We will be using the binomial distribution to model what to expect

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

- Then we check how well model fits the data, will reject if model is poor fit

Let take a simple example where we did survey on 3 people if they like RED Lays or the GREEN Lays

Assume that

- 1st person prefered RED Lays
- 2nd person prefered RED Lays
- 3rd person prefered GREEN Lays

if people didnt prefer one flavour over other then we will assume that 50% chance of people preferring RED Lays and 50% of chance of people preferring GREEN Lays also we can calculate probability of 2 people randomly picking RED Lays and 1 people randomly picking GREEN Lays

x : No. of people who prefered RED Lays

n : Total number of samples

p : Probability of people picking RED Lays

1-p : Probability of people picking GREEN Lays

Gathering the values for each variable and fitting them in the formulae we get

for the above example x=2; n=3; p=0.5

$$pr(x = 2|n = 3, p = 0.5) = \frac{3!}{2!(3 - 2)!}0.5^2(1 - 0.5)^{3-2}$$
$$pr(x = 2|n = 3, p = 0.5) = 0.375$$

The probability that 2 out of 3 people randomly preferring RED Lays is **0.375**