PROBABILITIES, RANDOM VARIABLES & THEIR DISTRIBUTIONS

3.1 Probabilities

- In any random experiment (e.g. fertilizer trials on yield), there is always uncertainty on whether a particular event (like desired yield) will or will not occur. We measure this using ***probability***. This is the likelihood that an event will occur.
- Probabilities range from 0 – 1 or 0 – 100% (in percentage)
- If we are certain an event will occur, we say the probability is 1 or 100%. If we are sure the event will not occur, then probability is zero. If our probability is ¼ or 0.25, then there is a 25% chance of the event occurring.
- If we toss a coin, the probability of getting "heads" i.e. P(H) is ½ or 50% and probability of getting "tails" i.e. P(T) is also ½ or 50%. This is because there are only 2 possible outcomes.
- P(Event) = # of possibilities meeting your condition / total number of outcomes

3.2 Random Variables

- A random variable is an alphabetical character (a capital letter) which associates each outcome of an event with a real number.
- It can take on different numbers because it is a function.
- **X** = {1 if Heads, 0 if tails}. In this case, **X** is the random variable.
- Random variables can be associated with discrete (countable) outcomes or continuous (measurable) outcomes.
- Simple example:
- If **X** = # of heads after 2 flips of a coin, what is the probability distribution of the different possible values?

Possible outcomes:

- [H, H], [H, T], [T, H] and [T, T]. Therefore:

a) P(X = 0) = ¼ or 25%
b) P(X=1) = 2/4 or 50%
c) P(X =2) = ¼ = 25%

This is a distribution of probabilities for the random variable **X**.

- Ideally in Statistics, we try to find the probability that a random variable (X) will equal a particular value.
- For example, what is the probability you will get 5 heads or 0 heads if you Toss a coin 7 times? Or what is the probability that you will have 2 successes in terms of yield if you Use 4 fertilizer trials?
- There are two important additional measures of a random variable:

a) The expected value of a random variable represented by E(X) indicates the average value or mean. $E(X) = \mu$

b) The variance of a random variable represented by Var(X) is a measure of the spread of the distribution around its mean.

3.3 Distributions under Random Variables

There are different distributions under different types of random variables. The most common are:

a) **Discrete** random variable distributions

- Binomial Distribution
- Poisson Distribution

b) **Continuous** random variable distributions

- Normal Distribution
- Standard Normal Distribution

Binomial Distribution

- The binomial distribution or binomial random variable (X) represents the possible number of successes in a given number of trials. E.g. if we toss a coin 20 times, how often will we observe “H” which is our success in this case.
- The random variable X is represented as:
- X ~ Binomial (n, p)
- This entells that X is distributed binomially, with “n” number of trials and “p” probability of successes
- To find the probability that our random variable X is equal to a particular number, we use:
- P(X = x) = $\binom{n}{x} p^x q^{n-x}$

Where: x = number of successes, n = number of trials, p = probability of success, and q = (1-p) = probability of failure

- p =denotes the probability of success in *one* of the n trials.

- q =denotes the probability of failure in one of the n trials.

- $P(x)$ =denotes the probability of getting exactly x successes among the n trials.

- x = denotes a specific number of successes in n trials, so x can be any whole number between 0 and n, inclusive.

This starts the count of number of ways event can occur.

This is the probability of success for x trials.

$$P(x) = \frac{n!}{(n-x)!\,x!} p^{x} q^{n-x}$$

This ends the count of number of ways event can occur.

This deletes duplications.

This is the probability of failure for the x trials.

Characteristics of a binomial distribution

1)It has fixed (finite) number of trials n.

2)Only 2 possible outcomes, success (p) and failure (1-p)

3)Probability of success remains constant through each trial

4)Trials are independent of each other

•There are two important additional measures of a random variable:

a)The expected value of a random variable represented by E(X) indicates the average value or mean. $E(X) = \mu$

b)The variance of a random variable represented by Var(X) is a measure of the spread of the distribution around its mean.

For a binomial distribution,

a) $E(X) = np$

b) $Var(X) = npq$ or $np(1-p)$

Examples

1)If we flip a coin 10 times, what is the probability that we observe 7 Heads? Assume the coin is fair. Calculate expected value and variance.

2)In a given maize variety, it is known that germination ratio is 80%. What is the probability of 6 seeds germinating out of 10 seeds planted? Calculate expected value and variance.

3)At UTH, it was found that generally, out of 100 births, 25 babies turn out to be boys. What is the probability of getting 10 boys in 30 births?

4)Observations over a long period of time have shown that on average, 1 out of 10 items produced by a given process is defective. If we select 5 items,

a)What would be the expected number of defective items out of the 5?

b)What would be the variance?

c)What is the probability of observing at most one defective item?

d)What is the probability of observing 3 or more defective items?

Poisson Distribution

- The Poisson distribution or Poisson random variable X expresses the probability of a given number of events occurring in a fixed interval of time.
- This is represented as: X ~ Poisson (λ)
- λ is called the mean or constant rate/ mean rate
- The probability that over a specific time interval, you observe x many occurrences is given by:
- P(X = x) = $(e^{-\lambda})$ (λ^{x}) / x!
- Where:
- e: A constant equal to approximately 2.71828. (Actually, e is the base of the natural logarithm system.)
- λ : The mean number of successes that occur in a specified region.
- x: The actual number of successes that occur in a specified region
- In literature, λ is used interchangeably with μ

The Poisson Distribution

$$P(X) = \frac{e^{-m} . m^{x}}{X!}$$

Where,
X = 1,2,3,4.....
e = 2.7183 (the base of natural logarithms)
m = the mean of Poisson distribution i.e. the average number of occurrence of an event.

Characteristics of the Poisson Distribution

1) The average number of successes (μ) that occurs in a specified region is known.
2) Used for very rare occurences
3) The number of successes in various intervals are independent

3) The probability that a success will occur is proportional to the size of the interval meaning that it is the same for all intervals of equal size.

4) The expected value/ mean and the variance are both equal to λ

Q- The average number of accidents at a particular intersection every year is 18.
(a) Calculate the probability that there are exactly 2 accidents there this *month*.

There are 12 months in a year, so $\lambda = \frac{18}{12}$

$= 1.5$ accidents per **month**

$$P(X = 2) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

$$= \frac{e^{-1.5}1.5^{2}}{2!}$$

$$= \mathbf{0.2510}$$

(b) Calculate the probability that there is at least one accident this month.

$P(X \geq 1) = P(X=1) + P(X=2) + P(X=3) + \ldots$ Infinite.

So… Take the complement: **P(X=0)** $= \dfrac{e^{-\lambda}\lambda^{x}}{x!}$

$$= \frac{e^{-1.5}1.5^{0}}{0!}$$

$$= \frac{e^{-1.5}\times 1}{1}$$

$$= \mathbf{0.22313}$$

So $P(X \geq 1) = 1 - P(X=0)$

$= 1 - 0.22313$

$= 0.7769$

c What is the probability that there are *more than 2* accidents <u>in a particular month</u>

b $P(X > 2) = 1 - P(X \le 2)$

There are an infinite number of cases so instead consider $X \le 2$

$= 1 - [\, P(X = 0) + P(X = 1) + P(X = 2)\,]$

$= 1 - [\, 0.2231 + 0.3347 + 0.2510\,]$

$= 1 - 0.8088$

$= 0.1912$

Examples

1) Suppose that over 10 minutes on average, we observe 6 black cars. What is the probability that over a time interval from 0 to 10 minutes, the number of black cars observed is 8?

2) If a typist makes errors at random at a mean rate of 2 per page. Find the probability of making:

   a) no errors

   b) more than 3 errors

3) If the number of animals arriving at a watering point to drink water is on average 2 animals every 3 minutes.

   a) What is the probability that 5 animals will arrive in 9 minutes?

   b) What is the probability that 5 or more animals will arrive in 9 minutes?