Chapter 1

Introduction to statistics

Statistics ıs the art of learning from data. It is concerning with the collection of data, their subsequent descriptıon and their analysis, which often leads to the drawing of conclusions,

Exp: to check the effect of a new tablet/drug on a group of patients,

Data vs Information:

Data is the raw material that is to be processed for information or for collection of details. It is unorganized data or facts that are to be processed. Data is plain fact and it has to be processed for further information. Data is alone enough to get details and find the meaning of something. Data is the computers language. Data is useless unless it is processed or has been made into something. Data has no meaning when it has not been interpreted. Data comes in figures, dates and numbers and is not processed.

Examples of Data

- Student Data on Admission Forms: When students get admission in a college. They fill admission form. This form contains raw facts (data of student) like name, father's name, address of student etc.
- Data of Citizens: During census, data of all citizens is collected.
- Survey Data: Different companies collect data by survey to know the opinion of people about their product.
- Students Examination data: In examination data about obtained marks of different subjects for all students is collected.

Information is processed data. The data that can be made useful is known as information. Information is basically the data plus the meaning of what the data was collected for. Data does not depend upon information but information depends upon data. It cannot be generated without the help of data. Information is something that is being conveyed. Information is meaningful when data is gathered and meaning is generated. Information cannot be generated without the help of data. Information is the meaning that has been formed with the help of data and that meaning makes sense because of the data that has been collected against the word. Information is processed and comes in a meaningful form.

Descriptive statıstics:

The part of statistics concerned with the description and sumarization of data.

Exp: at the end of the experiment, the data should be described,

İnferential statistics:

The part of statistics concerned wıth the drawing of conclusions from data. To be able to draw a conclusion from the data, we must take into account the possibility of chance.

To be able to draw logical conclusions from data, it is usually necessary to make some assumptions about the chances or probabilities of obtainıng the different data values. The totality of these assumptions is reffered to as a probability model to describe the data, an understanding of statistical inference requires some knowledge of the theory of probability.

Population:

The total collection of all the elements that we are interested is called a population (a collection of elements of interest).

Sample:

A subgroup of the population that will be studied in detail is called a sample.

Probability model:

The mathematical assumptions relating to the lıkelihood of different data values.

Representative means sample:

The sample is chosen in such a way that all possible choices of the k members are equally likely (lottery)

Stratified random sampling:

In this type of sample, first the population is stratified into sub-populations and then the correct number of elements is randomly chosen from each of the subgroups.

Exp: 2000 students in one school, 300 first-year class, 500 in second year class, 600 in 3^{rd} class, and 600 in the 4^{th} year class. We want to select 40 students, so we select 10 students from every class,

Chapter 2

Describing data Sets

In this chapter we learn methods for presenting and describing sets of data. We introduce different types of tables and graphs, which enable us to easily use key features of a data set.

Frequency:

Is the number of times that a given value occurs in a data set

Frequency table:

Present each distinct value with its frequency of occurrence.

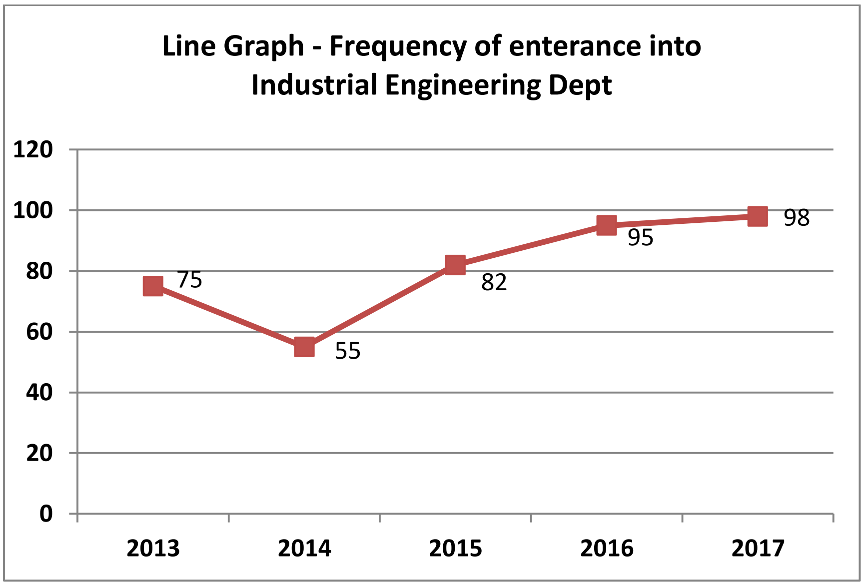
Frequency table of students that enter to our department in different years

year	2013	2014	2015	2016	2017
Frequency	75	55	82	95	98

Line Graph:

Data from a frequency table can be graphically pictured by a line graph, which plots the successive values on the horizontal axis and indicates the corresponding frequency by the height of a vertical line.

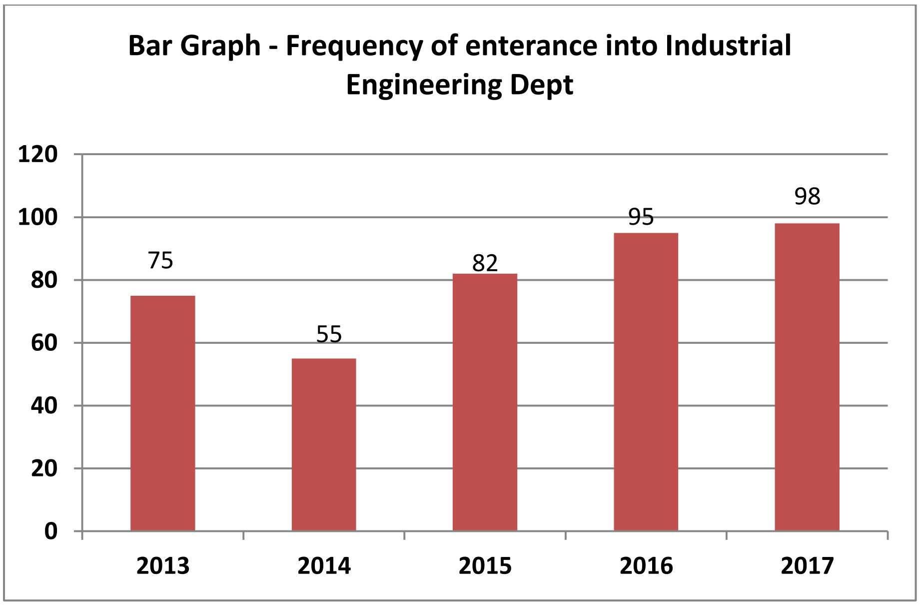
Example: draw line graph for frequency table



Bar Graph:

Sometimes the frequencies are represented not by lines but rather by bars having some thickness. These graphs called bar graphs.

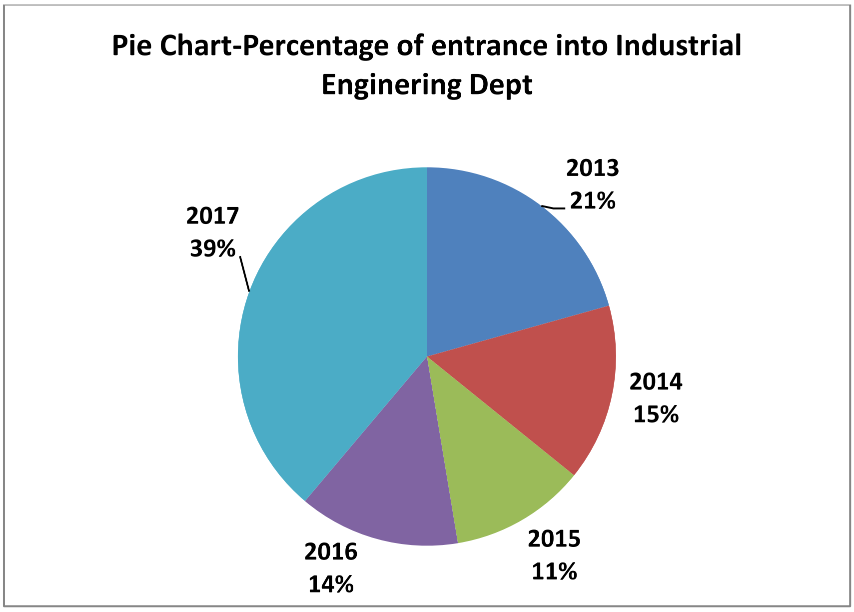
Example: draw Bar graph for frequency table



Pie chart:

This chat is often used to plot relative frequency when the data are non-numeric. A circle is constructed and then is sliced up into distinct sectors, one for each different data value,

Example: draw Pie chart for frequency table



Example: the following data represent the waiting time queue for 50 passengers.

8, 4, 6, 3, 7, 3, 7, 5, 4, 8, 3, 7, 15, 16, 15, 8, 4, 4, 3, 3, 9, 5, 12, 8, 7, 5, 9, 3, 8, 9, 22, 10, 3, 37, 7, 6, 8, 3, 5, 16, 4, 15, 3, 12, 6, 8, 12, 12, 3, 5

A) Construct a frequency table for these data
B) Using a line graph, plot the data
C) Using a bar graph, plot the data
D) Suppose we define the following categories for waiting time
   Waiting time ≤ 6, $6 < t \leq 12$, $12 < t \leq 20$, $t > 20$. Plot the relative frequencies using a pie chart.

Histogram:

A graph, which the data are divided into class intervals, whose frequencies are shown in a bar graph.

- The end points of a class interval are called the class boundaries
- the class interval $a - b$ contains, $a \leq x < b$
- Number of intervals, $k = \sqrt{n}$ (round it up), where is the number of observations
- Class intervals=Range/$\sqrt{n}$(use exact value of $\sqrt{n}$), where range=$x_{max} - x_{min}$

To construct a histogram from a data set:

1- Arrange the data in increasing order
2- Choose class intervals so that all data points are covered
3- Construct a frequency table/relative frequency table
4- Draw the bar graphs having heights determined by the frequencies in step 3

Relative frequency:

if f represents the frequency of Occurrence of some data value x then the relative frequency $\boldsymbol{f_n}$ where n is the total number of observations in data set.

Example: By information of waiting time (previous example),

1- Find appropriate number of classes and class interval
2- Construct frequency table and histogram
3- Is it symmetric or not. If not, what is its skewness?

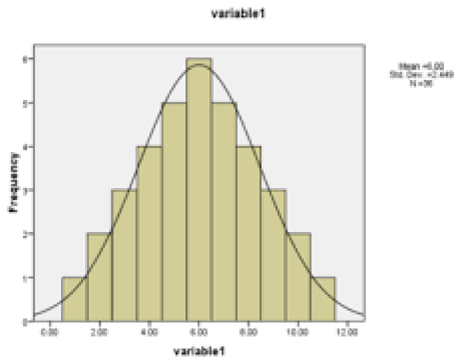
Example: consider the following data:

4.1, 3.3, 2.4, 3.7, 2.1, 0.8, 2.1, 10, 3.6, 2.1, 5.1, 3.1, 2.2, 2.4, 2.2, 1.9, 1.8, 5.8, 2.9, 4.5, 5, 7.1

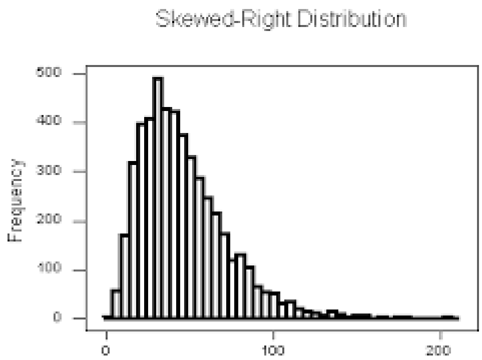
1- Find appropriate number of classes and class interval
2- Construct frequency table and draw histogram by using relative frequency
3- Is it symmetric or not. If not, what is its skewness?

Symmetry and Skewness:

- A histogram is perfectly symmetric if it's right half is a mirror image of left half

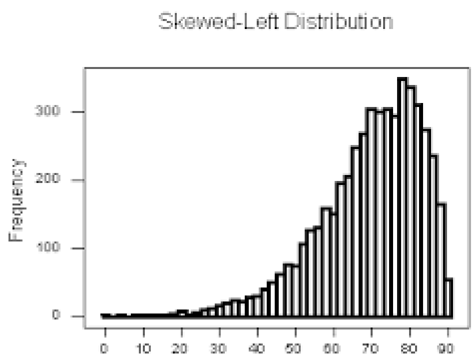


- Histograms that are not symmetric are referred to as skewed.
- A histogram with a long right-hand tail is said to be skewed to the right or positively skewed.



- **if the mean is bigger than the median that is the histogram is skewed right.**

- A histogram with a long left-hand tail is said to be skewed to the left or negatively skewed.



- **if the mean is smaller than the median that is the histogram is skewed left.**

Chapter 3

Probability

Probability: The word probability is commonly used term that relates to the chances that a particular event will occur when some experiments are performed.

An experiment is any process that produces an observation or outcome,

$$Probability = \frac{Number\ of\ Favorite\ Outcomes}{Number\ of\ Possible\ Outcomes}$$

An experiment is any process that produces an observation or outcome,

Sample space: The set of all possible outcomes of an experiment is called the sample space.

Example:

- In one gender survey: $sample\ space: s = [m, f]$
- Selecting one person: $sample\ space: s = [m, f]$
- Flipping a coin: $sample\ space: s = [h, t]$
- Rolling a dice: $sample\ space: s = [1, 2, 3, 4, 5, 6]$
- Rolling two dice:
  $sample\ space: s = [(1,1), (1,2), \ldots (1,6), (2,1), (2,2), \ldots (2,6), \ldots, (6,1), (6,2), \ldots (6,6)]$

Event: any set of outcomes of a sample space is called an event

Example: After rolling two dice, we are ınterested to reach sum of the two dice is equal to 5, find the members of this event?

$\boldsymbol{A} \cup \boldsymbol{B}$**:** is called the **union** of events A and B to consist of all outcomes that are in A or in B or in both A and B.

$\boldsymbol{A} \cap \boldsymbol{B}$: is called **intersection** of events A and B to consist of all outcomes that are both in A and B.

Null event: the event without any outcomes. (∅)

Exclusive events: If the intersection of A and B is null event then we say that A and B are disjoint or mutually exclusive.

Complement of an event like A will occur when A does not occur or the set of outcomes that do not belong to A.

Complement of A = A^c

Example: in an experiment we roll a dice. We are ınterested to know about:

S= Sample space- A=even numbers-B=odd numbers- C=bigger than 3- D=smaller than or equal to 5

Find members of every event and also $A \cap B$, $B \cup C$, $A \cup (B \cap C)$, $(A \cup B)^C$, $(B \cap C)^C$

Example: A fair dice is rolled once and a fair coin is flipped once.

- Find members of event that either the dice will land on 3 or that the coin will land on heads?
- Find members of event that either the dice will land on 3 or that the coin will land on heads?

Properties of probability

1- $0 \leq P(A) \leq 1 \quad \forall A$
2- P(S)=1, where s is sample space
3- If intersection of A and B is empty ($A \cap B = \emptyset$)
$P(A \cup B) = P(A) + P(B)$
$A \cup A^C = S \quad \rightarrow P(A \cup A^C) = 1$
$P(A^C) = 1 - P(A)$
$P(\emptyset) = 0$
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Example: If A and B are disjoint events and $P(A) = 0.2$ & $P(B) = 0.5$

Find: $P(A^c) \quad P(A \cup B) \quad P(A \cap B) \quad P(A^c \cap B)$

Example: Ali has 40% chance of receiving A grade in statistics, 50% chance of receiving A in physics and 86% chance of receiving A in eıther statistics or physics.

Find the probability that he

- Does not receive A in either statistics or physics
- Receive A in both courses

Example: suppose two dice are rolled. Find the probability that the sum of dice is 6

Example: A man has 10 keys to open 10 doors. Find the probability that

- The first key opens a chosen door
- All of 10 keys are tried

Conditional probability:

We are often interested in determining probabilities when some partial information concerning the outcomes of the experiment is available. In such situations the probabilities are called conditional probabilities.

Assume that A and B are two events, $P\left({}^{A}/_{B}\right) = Probability\ of\ A\ given\ B = \frac{P(A \cap B)}{P(B)}$

Example: consider rolling two dice

- What is the probability that there will be at least one 6 in outcomes?
- We know that the summation is 8, now what is the probability that we have at least one 6?
- We know that the difference is 2, now what is the probability that we have exact one 2?
-

Multiplication rule:

$$P(A \cap B) = P(A).P\left({}^{B}/_{A}\right) = P(B).P({}^{A}/_{B})$$

Example: suppose that 2 people are randomly chosen from a group of 4 women and 6 men.

- What is the probability that both are women?
- What is the probability that one is woman and one is man?

Independency:

- Events A and B are independent, $P(A \cap B) = P(A).P(B)$,
- Then $P\left({}^{A}/_{B}\right) = P(A)$ & $P\left({}^{B}/_{A}\right) = P(B)$
- If $A_1, A_2, \dots, A_n$ are independent, then $P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1).P(A_2), \dots (P(A_n)$

Example: A couple is planning to have children, assume that each child is equally likely to be of either gender and the gender of the children is independent. Find the probability that:

- All three children will be girl?
- At least one child will be girl?

Example: A game club has 120 members, 40 member play chess, 56 members Play Bridge, 26 members play both chess & bridge.
If a member of the club is randomly chosen, find the conditional probability that he or she

- Plays chess given that he or she plays bridge?
- Plays bridge given that he or she plays chess?

Remark

Contribution of n things taken r

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

It represents the number of different groups of size r that can be selected from a set of size n when the order of selection is not important.

Note:

- $\binom{n}{r} = \binom{n}{n-r}$
- $0! = 1$
- $\binom{n}{1} = n$

Example: How many different groups of size 3 can be chosen from a set of 6 people?

Example: A committee of 4 people is selected from a group of 5 men and 7 women. If the selection is made randomly.

- What is the probability that the committee will consist of 2 men and 2 women?

Example: if 4 workers are assigned to 4 jobs,

- How many different assignments are possible?
- How many assignments are possible if workers 1 & 2 are both able to do jobs 1 & 2 and workers 3 & 4 are able to do jobs 3 & 4?

Example: A delivery company has 12 trucks which 4 of them have faulty brakes, if an inspector randomly chosen 2 trucks for brake check,

- what is the probability that none of them has faulty brake?

Chapter 5

Discrete random variable

Random variable: is a function that associates a real number with each element in the sample space.

Discrete Random Variable: A random variable whose possible values constitute a sequence of disjoint on the number line.

$$\sum_{i=1}^{n} P(X = x_i) = 1$$

Example: suppose that X is a random variable that takes one of the values 1, 2, or 3. If $P(X = 1) = 0.4$ & $P(X = 2) = 0.25$, what is the $P(X = 3)$?

Example: A shipment of parts contains 10 items of which 2 are defective. two of these items are randomly chosen and inspected, let X denote the number of defectives. Find the probability mass function of X.

Example: A supervisor in a manufacturing plant has three men and three women working for him. He wants to choose two workers for a special job. Not wishing to show any biases in his selection, he decides to select the two workers at random. Let Y denote the number of women in his selection. Find the probability mass function of Y.

Expectation value of a discrete random variable

If X is discrete random variable having a probability mass function $P(X = x_\iota)$, then $E(X) = \sum_{i=1}^{n} x_i p(X = x_i)$

Example: in rolling a fair dice, where X is the side facing up, find $E(X)$?

Properties of expectation value:

- $E(cX) = c.E(X)$ x is random variable and c is constant value
- $E(X + c) = c + E(X)$
- $E(X \pm Y) = E(X) \pm E(Y)$ X & Y are random variables
- $E[\sum_{i=1}^{n} x_i] = \sum_{i=1}^{n} E(x_i)$

Example: Suppose that a radio contains six transistors, two of which are defective. Three transistors
are selected at random, removed from the radio, and inspected. Let Y equal the number of defectives observed, where Y = 0, 1, or 2. Find the expected value for Y.

Example: Roll a dice. If the side that comes up is odd, you win the $ equivalent of that side. If it is even, you lose $4. Find expected value of earning?

Example: three people are randomly chosen from a group of 10 men and 20 women. Let X denotes the number of men chosen and let Y denotes the number of women chosen.

- Find E(X)
- Find E(Y)
- Find E(X+Y)

Variance of a discrete random variable

If X is discrete random variable having a probability mass function $P(X = x_i)$, then $var(X) = E[(X-\mu)^2] = E(X^2) - \mu^2 = \sum_{i=1}^{n}(x_i - \mu)^2 P(X = x_i)$

Example: Find $var(X)$, when the random variable X is such that

$$X = \begin{cases} 1 & with\ probability\ p \\ 0 & with\ probability\ 1-p \end{cases}$$

Properties of variance:

- $var(cX) = c^2 . var(X)$ x is random variable and c is constant value
- $var(X + c) = var(X)$
- $var(X \pm Y) = var(X) + var(Y)$ X & Y are independent random variables
- $var[\sum_{i=1}^{n} x_i] = \sum_{i=1}^{n} var(x_i)$
- $Standard\ Deviation\ (SD) of\ X = \sqrt{var(X)}$

Example: The annual gross earnings of a soccer player are a random variable with expected value of 400,000 $ and SD of 80,000 $. The manager of soccer player receives 15% of this amount. Determine the E(X) & SD(X) of the amount received by the manager?

Example: in rolling a fair dice, where X is the side facing up, find $var(X)$?

Example: Y=0.25X (X & Y are independent random variables). We know that $\mu_X = 18$ & $\sigma_X^2 = 0{,}01$. Find mean and variance of Y ?

Example: for a constant C, P(X=C)=1. Find var(X)

Continuous Random Variable:

A random variable is continuous if its probability is given by area under a curve. The curve is called a Probability Density Function (PDF) for the random variable.

Let X be a Continuous Random Variable with Probability Density Function f(X). let a & b be any two numbers when a<b. then

- $p(a \leq x \leq b) = p(a < x < b) = p(a \leq x < b) = p(a < x \leq b) = \int_a^b f(x)dx$
- $p(a \leq x) = p(a < x) = \int_a^{+\infty} f(x)dx$
- $p(x \leq b) = p(x < b) = \int_{-\infty}^{b} f(x)dx$
- $\int_{-\infty}^{+\infty} f(X) = 1$

Expectation value and variance of a discrete random variable

$$E(X) = \int_{-\infty}^{+\infty} Xf(X)dx$$

$$var(X) = \int_{-\infty}^{+\infty} X^2 f(X)dx - [E(X)]^2$$

Example: A hole is drilled in a sheet metal and then a shaft is inserted through the hole. The shaft clearance is equal to difference between the radius of the hole and radius of the shaft. Let the random variable X denotes the clearance in millimeters. The probability density function of X is as below.

$$f(X) = \begin{cases} 1.25(1 - x^4) & 0 < x < 1 \\ 0 & otherwise \end{cases}$$

What is the probability that the shaft clearance is larger than 0.8mm?

Example: The total number of hours, measured in units of 100 hours, that a family runs a vacuum cleaner over a period of one year is a continuous random variable X that has the density function

$$f(x) = \begin{cases} x & 0 < x < 1 \\ 2 - x & 1 < x < 2 \\ 0 & elsewhere \end{cases}$$

Find the probability that over a period of one year, a family runs their vacuum cleaner

- Less than 120 hours
- Between 50 and 100 hours
- Expectation value of X
- Standard deviation of X

Example: consider the density function

$$f(x) = \begin{cases} k\sqrt{x} & 0 < x < 1 \\ 0 & elsewhere \end{cases}$$

- Evaluate k
- Find the probability that X is between 0.3 and 0.6

Chapter 6

Discrete and continuous distributions

Binomial distribution

a binomial random variable with parameters n and p represents the number of success with probability p independently. If X is such a random variable, then

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0,1,\dots,n$$

Expectation value and variance of binomial distribution:

If X is a discrete binomial random variable with parameters n and p, then

$$E(X) = np \qquad var(X) = np(1-p) = npq$$

Example: suppose you will be attending 6 hockey games. If each game independently will go to overtime with probability 0.1, find the probability that:

- At least one of the games will go to the overtime
- At most one of the games will go into overtime

Example: A fair dice is to be rolled 20 times, find the expected value and variance of the number of times:

- 6 appears
- 5 or 6 appears
- An odd number appears

Example: At a certain airport, 70% of the flights arrive on time. A sample of 10 flights is studied. Find

- Probability that exactly 8 flights were on time
- Probability that less than or equal to 8 flights were on time

Hypergeometric distribution

(Binomial distribution in which the trials are not ındependent)

The probability distribution of the hyper geometric random variable is the number of success in a random sample of size n selected from N items of which k are labeled success and n-k labeled failure.

If X is such a random variable, then

$$H(X,N,n,k) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} \qquad x = 0,1,\ldots,k$$

Expectation value and variance of Hypergeometric distribution:

If X is a discrete hyper geometric random variable with parameters n and p, then

$$E(X) = \frac{nk}{N} \qquad var(X) = \left(\frac{N-n}{N-1}\right)\frac{nk}{N}\left(1-\frac{k}{N}\right) = (\frac{N-n}{N-1})npq$$

Example: Draw 6 cards from a deck without replacement. What is the probability of getting two hearts?

Example: A crate contains 50 light bulbs of which 5 are defective and 45 are not. A Quality Control Inspector randomly samples 4 bulbs without replacement. Let *X* = the number of defective bulbs selected. Find the probability mass function, *f*(*x*), of the discrete random variable *X*.

Find mean and variance X.

Poisson distribution

A random variable X has Poisson distribution with parameterλ, if probability mass function of that random variable is:

$$p(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!} \qquad x = 0,1,2,\ldots$$

Expectation value and variance of Poisson distribution:

If X is a discrete Poisson random variable with parameter λ, then

$$E(X) = var(X) = \lambda$$

Example: If electricity power failures occur according to a Poisson distribution with an average of 3 failures every twenty weeks.

- calculate the probability that there will not be more than one failure during a particular week.
- Find expectation and variance of electricity power failures during a particular week

Note: Poisson random variable arise approximations to Binomial random variable, if the number of trials is large and probability of success is small ($\lambda = np$)

Example: the probability of producing a defective item is 0.1.

- What is the probability that a sample of 10 items will contain at most 1 defective item?
- What is the Poisson approximation for this probability?

Geometric distribution

Suppose that repeated independent Bernoulli trials each one having probability of success P are to be performed. Let X be the number of trials needed until the first success occurs. We say that X follows the geometric probability distribution with parameter p .

Probability mass function of X

$p(X = x) = (1 - p)^{x-1}p \qquad x = 1,2,\ldots$

What are the Differences between the Geometric and the Binomial Distributions?

- The most obvious difference is that the Geometric Distribution does not have a set number of observations, n.
- The second most obvious is the question being asked:
  Binomial: asks for the probability of a certain number of success
  Geometric: asks for the probability of the first success

Expectation value and variance of Gmetric distribution:

If X is a discrete geometric random variable with parameters x and p, then

$$E(X) = \frac{1}{p} \qquad var(X) = \frac{1-p}{p^2} = \frac{q}{p^2}$$

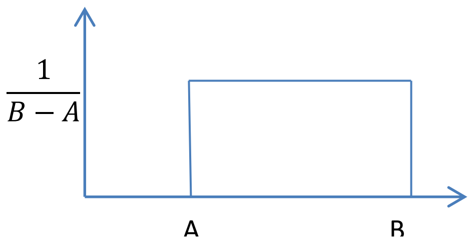
Example: If a production line has a 20% defective rate.

- What is the probability that first defective comes in third selection?
- What is the average number of inspections to obtain the first defective?

Uniform distribution

The density function of continuous uniform random variable X on the interval $[A, B]$ is

$$f(X, A, B) = \begin{cases} \frac{1}{B-A} & A \leq X \leq B \\ 0 & otherwise \end{cases}$$



Expectation value and variance of Uniform distribution:

If X is a continuous Uniform random variable with parameters x and p, then

$$E(X) = \frac{A+B}{2} \qquad var(X) = \frac{(B-A)^2}{12}$$

Example: Suppose that X is a Uniform R.V over the interval (0,2). find

- $p(X > {}^{1}/_{3}) =$
- $p(0.3 \leq X > 0.9) =$
- E(X) & Var(X)

Exponential distribution

The Random Variable X has an exponential distribution with parameter λ, if it's density function is given by

$$f(X) = \begin{cases} \lambda e^{-\lambda x} & x > 0, \lambda > 0 \\ 0 & otherwise \end{cases}$$

Expectation value and variance of Uniform distribution:

If X is a continuous Exponential random variable with parameters x and λ, then

$$E(X) = \frac{1}{\lambda} \qquad var(X) = \frac{1}{\lambda^2}$$

Example: If jobs arrive every 15 seconds on average, $\lambda = 4$ per minute,

- what is the probability of waiting less than or equal to 30 seconds

Example: Accidents occur with a Poisson distribution at an average of 4 per week. i.e.λ= 4

- Calculate the probability of more than 5 accidents in any one week
- What is the probability that at least two weeks will pass between accident?