2

HYPERGEOMETRIC DISTRIBUTION

If a population of size N contains K items of "successes" and N−K "failures", then the probability of the hypergeometric random variable X, the number of successes in a random sample of ~~xx~~ size n is:

$$\cancel{h(x; N, n, K) =}$$

$$P(X=k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

Where,

K = number of successes in the population

k = number of observed successes

N = population size

n = number of draws

Example 1

A deck ~~card~~ of cards contains 20 cards: 6 red cards and 14 black cards. 5 cards are drawn random without replacement. What is the probability that exactly 4 red cards are drawn?

Example 1

The average number of homes sold by the ACME Realty company is 2 homes per day. What is the probability that exactly 3 homes will be sold tomorrow?

Solution

$\mu = 2$, since 2 homes are sold per day, on average.

$x = 3$

$e = 2.71828$

$$P(x;\mu) = \frac{(e^{-\mu})(\mu^{x})}{x!} = \frac{2.71828^{-2} \times 2^{3}}{3!}$$

$$P(3;2) = \frac{(0.13534)(8)}{6} = 0.180$$

Variance and standard deviation.

The variance of a set of data points is the mean squared deviation from the mean.

$$\text{Variance} = \frac{\sum (x-\bar{x})^2}{n}$$

$$S^2 = \frac{\sum (x-\bar{x})^2}{n}$$

Standard deviation is the positive square root of the variance

$$S \text{ or } \sigma = \sqrt{\text{variance}}$$

Example 1: Obtain the standard deviation of the following scores 5, 3, 7, 6, 11, 10

Solution

$$\bar{x} = \frac{3+5+6+7+10+11}{6} = \frac{42}{6} = 7$$

x	$x-\bar{x}$	$(x-\bar{x})^2$	x^2
3	-4	16	9
5	-2	4	25
6	-1	1	36
7	0	0	49
10	3	9	100
11	4	16	121
		46	340

$$\text{Standard deviation} = \sqrt{\frac{\sum (x-\bar{x})^2}{n}} = \sqrt{\frac{46}{6}}$$

$$S = \sqrt{7.667} = 2.7689$$

Coefficient of variation (CV)

$$CV = \left(\frac{S}{\bar{x}} \times 100\right)\%$$

2

Poisson Distribution

It is usually used in situations where the probability of a rare event occurring of a large samples.
Suppose we conduct a poisson experiments in which the average number of successes within a given region is M. Thus, the poisson distribution is

$$p(x;M) = \frac{(e^{-M})(M^{x})}{x!}$$

where,

x = actual number of success that result from the experiment

$e = 2.71828$

The poisson distribution has the following properties

(1) The mean of the distribution is equal to M

(2) The variance is also equal to M

Poisson ~~binomial~~ distribution can therefore be used to approximate for the binomial distribution when p is very small and n is very large

Solution

The probability of choosing exactly 4 red cards is

p(4 red cards) = # samples with 4 red cards and 1 black card / # of possible 4 card samples

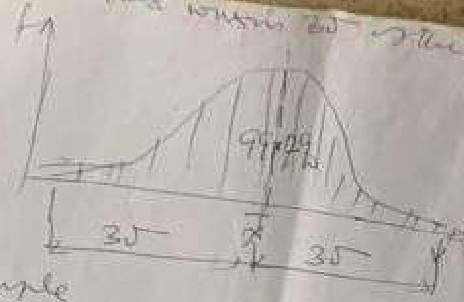
$$P(X=k)\ \frac{\binom{6}{4}\binom{14}{1}}{\binom{20}{5}}$$

$$\frac{{}^{6}C_{4}\times {}^{14}C_{1}}{{}^{20}C_{5}} = \frac{15\times 14}{15504} = \underline{\underline{0.0135}}$$

Bernoulli distribution

The Bernoulli distribution: ~~which~~ it takes ~~take~~ value "1" with ~~[illegible]~~ probability p and value "0" with probability $q = 1-p$

① Values within 3σ of the mean



Example

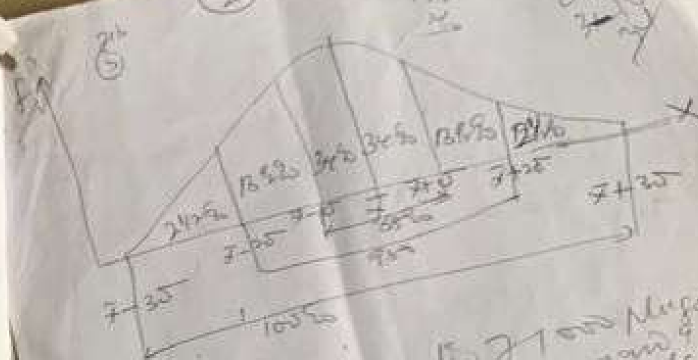
A manufacturing runs to produce ~~500~~ 1000 bolts of normal lengths 32.5mm ~~[illegible]~~. If means and standard deviations are 32.58mm and 0.06mm respectively. Calculate the range of variability of 68%, 95% and 99.7%.

① At 68% of the bolts are likely to have lengths between 32.58 − 0.06 and 32.58 + 0.06, i.e between 32.52 and 32.64mm

② At 95% of the bolts are likely to have lengths between 32.58 − 0.12 and 32.58 + 0.12 i.e between 32.38 and 32.62mm.

③ At 99.7% of the bolts are likely to have lengths between 32.58 − 0.18 and 32.58 + 0.18 i.e between 32.32 and 32.76mm.

②



Example: ~~[illegible]~~ If the diameters of 1000 plugs gave a mean of 74.82mm and standard deviation of 0.14mm. Calculate

ⓐ the number of plugs likely to have diameters less than 74.54mm.

ⓑ the number of plugs likely to have diameters between 74.68mm and 75.10mm.

Solution:

Given: $\bar{x} = 74.82$mm, $\sigma = 0.14$mm, $N = 1000$ plugs

ⓐ $74.54 = 74.82 - x(0.14)$ $\quad 74.82 - 74.54 = 0.28$

$x(0.14) = 0.28$

$x = \frac{0.28}{0.14} = 2$

~~General~~ General steps

Step 1: Identify the parts of the problem

1. the mean (average or μ)
2. the standard deviation (σ)
3. population size
4. sample size (n)
5. a number associated with greater than ($\bar{x}$) **NOTE**: this is the sample mean

Step 2: Draw a graph. Label the center with the mean

Step 3: Use the following formula to find the z-score. Plug in the numbers from step 1

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

$\bar{x}$ = mean
μ = mean of population
n = sample size
σ = standard deviation

standard error $= \frac{s}{\sqrt{n}}$

$$z = \frac{\bar{x} - \mu}{\text{standard error}}$$

$74.54 = \bar{x} - 2\sigma$

∴ number of rings with diameters less than 74.54 mm = 2½ percent × 1600 = 40

(b) $74.68 = 74.82 - 0.14 = \bar{x} - \sigma$

$75.10 = 74.82 + 2(0.14) = \bar{x} + 2\sigma$

∴ Numbers of rings between these values = (34 + 34 + 13½) percent = 81½ percent of 1600 = 1304

The Central Limit Theorem and Mean

An essential component of the central limit theorem is that the average of your sample means will be the population mean. In other words, add up the means from all of your samples, find the average and that average will be your actual population mean. Similarly, if you find the average of all of the standard deviations in your sample you will find the actual standard deviation of a population.

Three central limit theorem cases

1. We want to find the probability that the mean is greater than a certain number
2. We want to find the probability that the mean is less than a certain number
3. We want to find the probability that the mean is between a certain set of numbers either side of the mean

A

(3)

Probability

Definition

Probability is a measure of the likelihood that a particular event will occur in any one trial or experiment, carried out in prescribed conditions.

Notations

The probability that a certain event A occurs is denoted by $P(A)$ and also ~~equal~~ ~~success~~ = (P success).

Success or failure

When an event occurs in any one trial, is called success, but when it fails to occur is called failure.

In N trials, there are X successes, there will also be $(N-X)$ failures

~~Thus~~ $\frac{X}{N} = P(\text{success})$ and $\frac{N-X}{N} = P(\text{failure})$

$P(\text{failure}) = P(\text{not } A)$

$\therefore \quad P(A) + P(\text{not } A) = 1$ —— (7)

The event not A is called the compliment of event A and is often written $\bar{A}$

i.e. $P(A) + P(\bar{A}) = 1$

TYPES OF PROBABILITY

The determination of probability may be undertaken from two approaches:

(a) empirical (or experimental) probability

(b) classical (or theoretical) probability

Empirical probability. It is based on previous (known) results & examples: number of ~~cheques paid~~ tons of steel production per day. number of tons of maize harvested at experimental farm at Ikole. etc.

Expectation (E)

Expectation (E) is defined as the product of the number of trials N and the probability p(A) that the event A will occur in any one trial, i.e

$E = N \times p(A)$ ——— (2)

Addition Law of probability

For a given ~~probabilities~~ events A and B, if A and B are mutually exclusive the ~~probability~~
~~$P(A \text{ or } B) = P(A) + P(B)$~~
of either event A or event B occurring but not both. i.e.

$P(A \text{ or } B) = P(A) + P(B)$

If the events A and B are non ~~mutually~~ exclusive so that events A and B can occur together

$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

For example

The probability of scoring a multiple of 3 (i.e. 3 or 6) $= \frac{2}{6}$ P(A)

the probability of scoring a multiple of 2 (i.e. 2, 4, 6) $= \frac{3}{6} = P(B)$

probability of scoring a multiple of 3 or a multiple 2 ~~=~~ $P(A) + P(B) = \frac{2}{6} + \frac{3}{6} = \frac{5}{6}$

If event is non-exclusive.

$P(A) + P(B) - P(A \text{ and } B)$

$\frac{5}{6} - \frac{1}{6} = 2/3$

Independent events and dependent events

(a) events are independent when the occurrence of one event does not affect the occurrence of the occurrence of the second event the probability

(b) events are dependent when one event does affect the probability of the occurrence of the second

Measures of Central tendency/Location

(2)

An average is a value which is representative of a set of data. Since such values tend to lie within a set of data arranged according to magnitude, averages are also called measures of central tendency or measure of location. Several averages can be defined, the most common being the arithmetic mean, the median and the mode. Others are the geometric and harmonic means.

The Arithmetic Mean (A.M)

The arithmetic mean of a set of n observations $x_1, x_2 --- x_n$ denoted by $\bar{x}$ (and read x bar) is defined by

$$\bar{x} = \frac{x_1 + x_2 + --- + x_n}{n} = \frac{\sum x_i}{n} \quad \text{—(i)}$$

If the observations $x_1, x_2 --- x_k$ occur $f_1, f_2 --- f_k$ times respectively (where $f_1, f_2 --- f_k$ are corresponding), the A.M. is given by

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + --- + f_kx_k}{f_1 + f_2 + --- + f_k} = \frac{\sum fx}{\sum f} \quad \text{—(ii)}$$

Example I: The A.M. of the following numbers 10, 14, 8, 12 and 6 is given by

$$\bar{x} = \frac{10+14+8+12+6}{5} = 50/5 = 10$$

Example II: If 5, 8, 6 and 2 occurs with frequencies 3, 2, 4, 1 respectively, the A.M. is given

$$\bar{x} = \frac{(5\times3)+(8\times2)+(6\times4)+(2\times1)}{3+2+4+1} = 57/10 = 5.7$$

(4)

Dispersion

Dispersion or variability is a way of describing how spread out a set of data is, when a data set has a large value.

Measures of Dispersion

Measure of dispersion are Range, standard deviation, variance, mean absolute deviation, coefficient of variation.

The Range. The simplest measure of dispersion, is the difference between the highest value and the lowest value.

$$\text{The range} = V_{max} - V_{min} \qquad (1)$$

The mean Absolute Deviation

$$MAD = \frac{\Sigma|x-\bar{x}|}{n} \qquad (ii)$$

If however, we have a frequency distribution then

$$MAD = \frac{\Sigma f|x-\bar{x}|}{\Sigma f} \qquad (iii)$$

where $\Sigma f = n$ = number of observations and $|x-\bar{x}|$ means modulus of deviation $|x-\bar{x}|$ indicating that we ignore the signs of the deviations and take them all to be positive

(5)

Conditional probability

Conditional probability of one event B occurring when given an event A has already taken place vice versa. In case of an event B stated above. It is denoted by the symbol $P(B/A)$. If A and B are independent events, the fact that event A has already occurred will not affect the probability of event B.

$$\therefore P(B/A) = P(B)$$

If A and B are dependent events, then event A having occurred will affect the probability of the occurrence of event B.

For example

A box contains six 15 ohms resistors and twenty 40 ohms resistors. The resistors are all unmarked and of the same physical size.

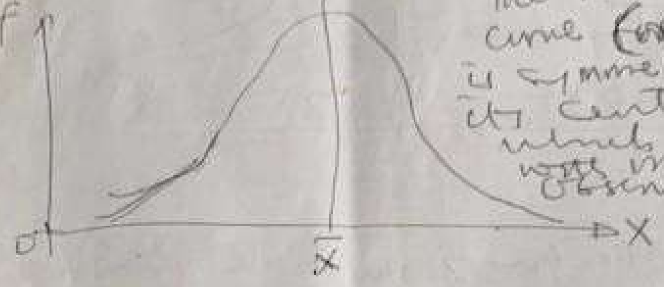
(a) If one resistor is picked out at random, determine the probability of its resistance being 15 ohms

(b) If the first resistor is found to be 15 ohms and it is returned on one side, find the probability that the second selected resistor will be of

Dr. O.E. Omofunmi

Statistics (1)

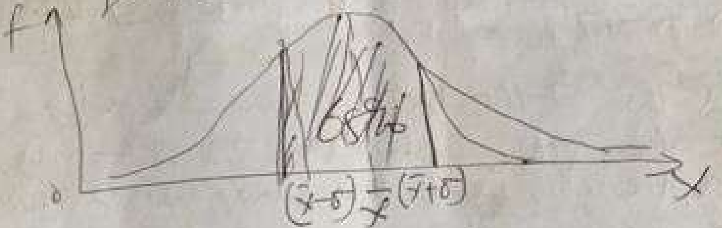
Normal distribution Curve



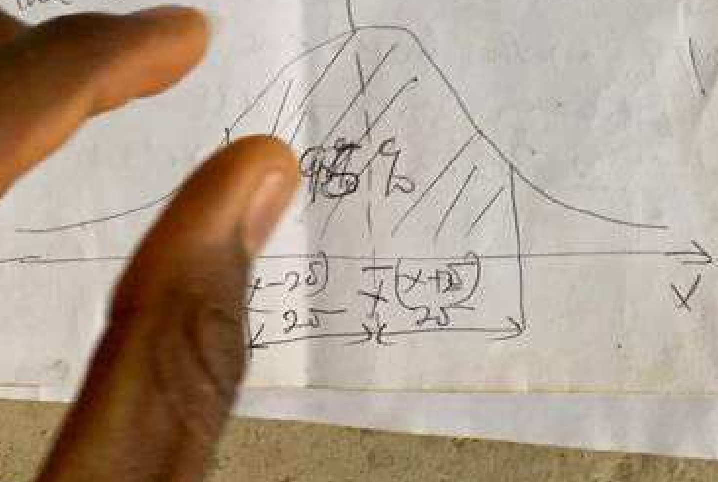
The normal distribution curve (or normal curve) is symmetrical about its centre line which coincides with mean $\bar{x}$ of the observations.

Values within standard deviation (σ) of the mean

① values within one (1σ) standard deviation of the mean.



② values within two (2σ) standard deviations of the mean



resistance 400 ohms.

solution

Let A = (15 ohm resistor) and B = (400 ohms resistor)

(a) N = 6 + 13 = 19.

Thus $P(A) = \frac{6}{19}$

(b) The box now contains five 15 ohm resistors and thirteen 400 ohm resistors

$p(B/A) = \frac{13}{18}$

The probability of getting a 15 ohm resistor at the first selection, retaining it, and getting a 400 ohm resistor at the second selection is

$p(A \text{ and } B/A) = p(A) + p(B/A)$

$= \frac{6}{19} + \frac{13}{18} = \frac{78}{342}$

If A and B are independent events

$p(A \text{ and } B) = p(A) \times p(B)$

and if A and B are dependent events

$p(A \text{ and } B) = p(A) \times p(B/A)$

(3)

Masses (kg/m²) x	no of lots f	Cum. no of lots (cf) cf
70 – 74	1	1
75 – 79	6	7
80 – 84	17	24
85 – 89	29	53
90 – 94	20	73
95 – 99	17	90
100 – 104	13	103
105 – 109	10	113
110 – 114	6	119
115 – 119	3	122
120 – 124	2	124
125 – 129	1	125
Total	125	

~~The above table shows that there are 53 lots out of 125 that requires less than 90kg/m²~~

The above table shows the following:

(i) the last column shows that there are 53 lots out of 125 that requires less than 90kg/m².

(ii) that 20 lots require between 89.5 and 94.5 kg/m²

(iii) that ~~[illegible]~~ fewer than half of the lots are less than 89.5 but more than half are less than 94.5 kg/m² and hence the median class is 90 – 94

The median is then, located between 90 and 94 by the following expression

$$\bar{X} = L_{1} + \left[\frac{\frac{n}{2} - F}{f_{1}}\right] C \quad \text{——— (ii)}$$

(4)

Classical probability

The classical approach to probability is based on a consideration of the theoretical number of ways in which it is possible for an event A to occur.

The classical probability p of an event A occurring is defined by

$$P(A) = \frac{\text{number of ways in which event A occur}}{\text{total number of all possible outcomes}}$$

Note: In any trial, each separate possible result is called an outcome.

For instance: If, out of n possible outcomes of a trial, it is possible for an event A to occur in x ways and for the event A not to occur in y ways, then $n = x+y$ and $P(A) = \frac{x}{n} = \frac{x}{x+y}$

Mutually exclusive and mutually non-exclusive events:

(a) Mutually exclusive events: These events cannot occur together. For example, in rolling a die in the same single trials, event of throwing a six and that of throwing five cannot occur at the same time.

(b) Mutually non-exclusive events: ~~This~~ Events that occur simultaneously. In this case, there are pairs of events.

The Median

The median of a set of numbers arranged in order of magnitude is

(a) The middle value of the numbers if number of observations, n is odd

eg the median $\tilde{x}$ of 3, 9, 6, 1, 4 is obtained by first ordering as

1, 3, 4, 6, 9

$\Rightarrow \tilde{x} = 4$

(b) Taken to be the A.M of the two middle values if n is even. In such a situation we say that the median is ~~strictly~~ indeterminate eg The median of 8, 14, 3, 12, 1, 6 is

1, 3, 6, 8, 12, 14

$\tilde{x} = \frac{1}{2}(6+8) = 7$ which, unfortunately is none of the observations.

Approximating the median from a grouped frequency distribution.

Solution

$\frac{3q}{210}$

$p = 0.5$

$q = 1 - p = 1 - 0.5 = 0.5$

$n = 10$

$x = 6$

$P(x=6) = {}^{10}C_6 \times (0.5)^6 \times (0.5)^{10-6}$

$nC_x = 10C_6 = \frac{10!}{(10-6)! \times 6!} = \frac{10!}{4! \times 6!}$

$$\frac{10 \times \overset{3}{\cancel{9}} \times 8 \times 7 \times \cancel{6 \times 5 \times 4 \times 3 \times 2 \times 1}}{\cancel{4 \times 3 \times 2 \times 1} \times \cancel{6 \times 5 \times 4 \times 3 \times 2 \times 1}} = 210$$

$P(x=6) = 210 \times 0.5^6 \times 0.5^4$

$= 210 \times 0.015625 \times 0.0625$

$= 0.205078$

Tip: You can use the Combination calculator.

Binomial Distribution

~~If n~~

$mean = np$

$variance = npq$

$standard\ deviation = \sqrt{npq}$

where,

L_x = ~~lower~~ Lower class boundary of the median class

F = cumulative frequency less than the lower class limit of the median class

f_x = frequency of the median class

C = size or width of the median class

n = cf = number of observations.

$$\tilde{x} = 89.5 + \frac{\left[\frac{125}{2} - 53\right]5}{20} = 91.87 = 92 \text{ kg}$$

~~The~~ The Mode

The mode is defined as the observation with the highest frequency. At times we may have a bimodal (two modes) or a multimodal (a distribution ~~with~~ with many modes) distributions.

BINOMIAL DISTRIBUTION

Binomial formula can be used when the probability of success is very high.

Binomial Formula

$$b(x; n, P) = {}_nC_x * p^x * (1-p)^{n-x}$$

Where,

b = binomial probability

x = total number of "successes" (pass or fail, heads or tails etc)

p = probability of a success on an individual trial

n = number of trials

Note: The binomial distribution formula can also be written in a slightly different way,

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

Example 1

A coin is tossed 10 times. What is the probability of getting ~~exactly~~ exactly 6 heads?

Basic Statistics (1)

Frequency Distributions

Consider the following data on house occupations:

4, 6, 2 3, 6, 2 3, 2, 4, 3, 6, 4, 3, 5, 9, 3, 2, 2, 8
2, 9, 5, 4, 2, 8, 5, 4, 7, 5, 14, 3, 13, 5, 14, 4, 2, 1
7, 6, 15, 1, 3, 2, 3, 8, 3, 9, 4, 4

The above data is called raw data, which can be represented by the following distribution.

House-occupation distribution

Number of occupants (x)	Tally	No of Houses (f)
1	||	2
2	𝍸 ||||	9
3	𝍸 𝍸	10
4	𝍸 |||	8
5	𝍸	5
6	||||	4
7	||	2
8	|||	3
9	|||	3
13	|	1
14	||	2
15	|	1
Total		50