CORPORATE

INSTITUTE OF SCIENCE AND TECHNOLOGY , BHOPAL

MATHEMATICS FORMULA GUIDE BY : Prof. Akhilesh jain

STATISTICS FORMULAS

Descriptive Statistics:

Term	Meaning	Population Formula	Sample Formula
Mean	Average	$\mu = \frac{\sum_{i=1}^{N} X_i}{N}$	$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$
Median	The middle value – half are below and half are above		
Mode	The value with the most appearances		
VARIANCE	The average of the squared deviations between the values and the mean	$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu)^2$	$s^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$
Standard Devi ation	The square root of Variance, thought of as the "average" deviation from the mean.	$\sigma = \sqrt{\sigma^2}$	$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}}$
Coefficient of Varia tion	The variation relative to the value of the mean		$CV = \frac{s}{\overline{X}}$

Probability Terms:

Term	Meaning	Notation	Example* (see footnote)
Probability	For any event A, probability is represented within $0 \le P \le 1$. P= P(A)= Number of favorable event / Total Event	P()	
Random Experiment	A process leading to at least 2 possible outcomes with uncertainty as to which will occur.		Rolling a dice
Event	A subset of all possible outcomes of an experiment.		Events A and B
Intersection of Events	Let A and B be two events. Then the intersection of the two events is the event that both A and B occur (logical AND).	$A \cap B$	The event that a 2 appears
Union of Events	The union of the two events is the event that A or B (or both) occurs (logical OR).	$A \cup B$	The event that a 1, 2, 4, 5 or 6 appears
Complement	Let A be an event. The complement of A is the event that A does not occur (logical NOT).	$\overline{A}$	The event that an odd number appears
Mutually Exclusive Events	A and B are said to be mutually exclusive if at most one of the events A and B can occur.		A and B are not mutually exclusive because if a 2 appears, both A and B occur
Collectively Exhaustive Events	A and B are said to be collectively exhaustive if at least one of the events A or B must occur.		A and B are not collectively exhaustive because if a 3 appears, neither A nor B occur
Basic Outcomes	The simple indecomposable possible results of an experiment. One and exactly one of these outcomes must occur. The set of basic outcomes is mutually exclusive and collectively exhaustive.		Basic outcomes 1, 2, 3, 4, 5, and 6

AKHILESH JAIN

CORPORATE

INSTITUTE OF SCIENCE AND TECHNOLOGY , BHOPAL

MATHEMATICS FORMULA GUIDE BY : Prof. Akhilesh jain

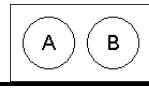
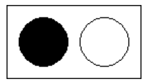
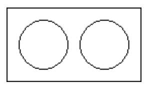
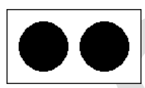
STATISTICS FORMULAS

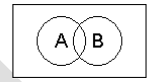
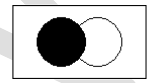
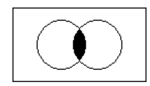
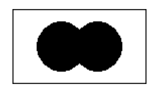
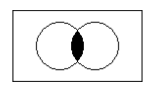
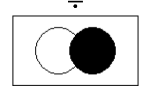
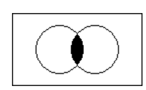
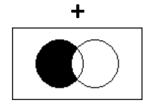
Sample Space	The totality of basic outcomes of an experiment.		{1,2,3,4,5,6}

* Roll a fair die once. Let A be the event an even number appears, let B be the event a 1, 2 or 5 appears

Probability Rules:

If events A and B are mutually exclusive

Term	Equals	Area: 
P(A)=	P(A)	
P($\overline{A}$)=	1 - P(A)	
P(A∩B)=	0	
P(A∪B)=	P(A) + P(B)	

Term	Equals	Venn: 
P(A)=	P(A)	
P($\overline{A}$)=	1 - P(A)	
P(A∩B)=	P(A) * P(B) *only if A and B are independen*	
P(A∪B)=	P(A) + P(B) – P(A∩B)	
P(A|B)= ***[Bayes' Law***: *P(A holds given that B holds)]*	$\frac{P(A\cap B)}{P(B)}$	 ÷ 
P(A∩B) = P(A|B) * P(B)		
P(A∩B) = P(B|A) * P(A)		
P(A)=	P(A∩B) + P(A∩ $\overline{B}$) = **P(A|B)P(B) + P(A|** $\overline{B}$ **)P(** $\overline{B}$ **)**	 + 

General probability rules:

1) If **P(A|B) = P(A)**, then A and B are **independent events**! (for example, rolling dice one after the other).

2) If there are *n* possible outcomes which are equally likely to occur:

P(outcome i occurs) = $\frac{1}{n}$ for each $i \in$ [1, 2, ..., *n*]

**Example: Shuffle a deck of cards, and pick one at random. P(chosen card is a 10 ♦) = 1/52.*

3) If event A is composed of ***n* equally likely basic outcomes**:P(A) =

$$\frac{Number\ \ of\ \ Basic\ \ Outcomes\ \ in\ \ A}{n}$$

Example: Suppose we toss two dice. Let A denote the event that the sum of the two dice is 9. P(A) = 4/36 = 1/9, because there are 4 out of 36 basic outcomes that will sum 9.

Probability

1. Additive Rule: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
2. Multiplicative Rule: $P(A \cap B) = P(A)P(B)$, *if A and B are independent*

AKHILESH JAIN

3. Complement Rule: $P(\overline{A}) = 1 - P(A)$

Conditional Probability

1. Definition: $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$
2. Multiplicative Rule: $P(A \cap B) = P(A)P(B \mid A) = P(B)P(A \mid B)$

Binomial Distribution

$X \sim B(n, p)$

1. $P(X = r) = \binom{n}{r} p^r (1-p)^{n-r} = \dfrac{n!}{r!(n-r)!} p^r (1-p)^{n-r}$
2. **Mean** = μ = np; **Variance** = $\sigma^2 = np(1-p) = npq$, **Standard Deviation** = $\sqrt{npq}$

Normal Distribution

X~ N(μ, σ)

1. Standard normal: $Z = \dfrac{X - \mu}{\sigma}$

Sample size : sample size to estimate the parameter μ to within B units with (1-α)100% confidence:

$$n = \left[\frac{z_{\alpha/2}\sigma}{B}\right]^2$$

Test statistics for μ and p

1. z-test for μ : $z = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}$
2. t-test for $\mu : t = \dfrac{\overline{X} - \mu}{s/\sqrt{n}}$ (d.f. = n–1)
3. z-test for p : $z = \dfrac{\hat{p} - p}{\sqrt{qp/n}}$, $np \geq 5$ and $nq \geq 5$(where q = 1 – p)

Test statistics for $\mu_1 - \mu_2$ and $p_1 - p_2$

1. z-test for $\mu_1 - \mu_2$: $z = \dfrac{\overline{X}_1 - \overline{X}_2 - \mu_1 - \mu_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$
2. t-test for $\mu_1 - \mu_2$ when σ_1, σ_2 unknown and $\sigma_1 = \sigma_2$: $t = \dfrac{\overline{X}_1 - \overline{X}_2 - \mu_1 - \mu_2}{\sqrt{s_p^2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$,

   where d.f. = $n_1 + n_2 - 2$ and $S_p^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$
3. t-test for μ_D (for matched pairs): $t = \dfrac{\overline{X}_D - \mu_D}{S_D/\sqrt{n_D}}$, where d.f. = $n_D - 1$

AKHILESH JAIN

4. z-test for $p_1 - p_2$:
$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

(where H_0: $p_1 - p_2 = 0$ and $\hat{p}_1 = \frac{X_1}{n_1}; \hat{p}_2 = \frac{X_2}{n_2}; \hat{p} = \frac{X_1 + X_2}{n_1 + n_2}; \hat{q}_1 = 1 - \hat{p}_1; \hat{q}_2 = 1 - \hat{p}_2$, and

all of $n_1\hat{p}_1, n_1\hat{q}_1, n_2\hat{p}_2, n_2\hat{q}_2 \geq 5$)

Central limit theorem $z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}, \quad n \geq 30$

CONFIDENCE INTERVALS

Confidence interval for a mean (large samples) $\bar{x} - z_c \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_c \frac{\sigma}{\sqrt{n}}$

Confidence interval for a mean (Small samples) $\bar{x} - t_c \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_c \frac{s}{\sqrt{n}}$

Confidence interval for a proportion (where *np* > 5 and *nq* > 5)

$$\frac{r}{n} - z_c \sqrt{\frac{\frac{r}{n}\left(1 - \frac{r}{n}\right)}{n}} < p < \frac{r}{n} + z_c \sqrt{\frac{\frac{r}{n}\left(1 - \frac{r}{n}\right)}{n}}$$

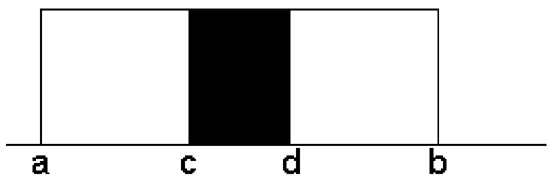
Counting Rules:

Term	MEANING	FORMULA	Example
Basic Counting Rule	The *number* of ways to pick *x* things out of a set of *n* (with no regard to order). The *probability* is calculated as 1/*x* of the result.	$\binom{n}{x} = \frac{n!}{x!(n-x)!}$	The *number* of ways to pick 4 specific cards out of a deck of 52 is: 52!/((4!)(48!)) = 270,725, and the *probability* is 1/270,725 = 0.000003694
Bernoulli Process	For a sequence of *n* trials, each with an outcome of either success or failure, each with a probability of *p* to succeed – the probability to get *x* successes is equal to the Basic Counting Rule formula (above) times $p^x(1-p)^{n-x}$.	$P(X = x \mid n, p) = \left(\frac{n!}{x!(n-x)!}\right) p^x (1-p)^{n-x}$	If an airline takes 20 reservations, and there is a 0.9 probability that each passenger will show up, then the probability that **exactly** 16 passengers will show is: $\frac{20!}{16!\,4!}(0.9)^{16}(0.1)^4$ = 0.08978
Bernoulli Expected Value	The expected value of a Bernoulli Process, given *n* trials and *p* probability.	E(X) = np	In the example above, the number of people expected to show is: (20)(0.9) = 18

AKHILESH JAIN

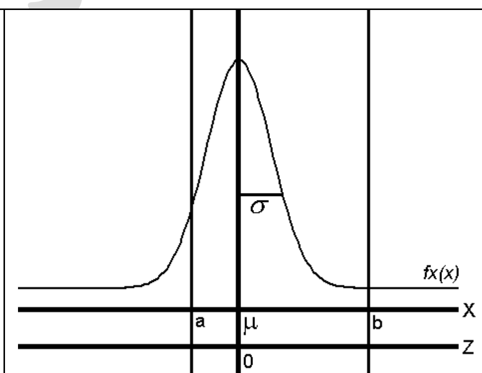
Bernoulli Variance	The variance of a Bernoulli Process, given *n* trials and *p* probability.	Var(X) = np(1 - p)	In the example above, the Bernoulli Variance is (20)(0.9)(0.1) = 1.8
Bernoulli Standard Deviation	The standard deviation of a Bernoulli Process:	$\sigma(X) = \sqrt{np(1-p)}$	In the example above, the Bernoulli Standard Deviation is $\sqrt{1.8}$ = 1.34

Uniform Distribution:



Term/Meaning	Formula
Expected Value	$\overline{X} = \frac{(a+b)}{2}$
Variance	$\sigma_X^2 = \frac{(b-a)^2}{12}$
Standard Deviation	$\sigma_X = \frac{(b-a)}{\sqrt{12}}$
Probability that X *falls* between *c* and *d*	$P(c \le X \le d) = \frac{d-c}{b-a}$

Normal Distribution:

Probability Density Function: $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ *where* $\pi \cong 3.1416$ *and* $e \cong 2.7183$ *P(a ≤ X ≤ b) = area under* $f_X(x)$ *between a and b:* $P(a \le X \le b)$ $= P\left[\left(\frac{a-\mu}{\sigma}\right) \le Z \le \left(\frac{b-\mu}{\sigma}\right)\right]$	Standard Deviations away from the mean: $Z = \frac{X-\mu}{\sigma}$ *(Z and σ are swappable!)*	

Correlation:

- If X and Y are two different sets of data, their correlation is represented by $Corr_{(XY)}$, r_{XY}, or ρ_{XY} (rho).
- If Y increases as X increases, $0 < \rho_{XY} < 1$. If Y decreases as X increases, $-1 < \rho_{XY} < 0$.
- The extremes $\rho_{XY} = 1$ and $\rho_{XY} = -1$ indicated perfect correlation – info about one results in an exact prediction about the other.
- If X and Y are completely uncorrelated, $\rho_{XY} = 0$.
- The **Covariance** of X and Y, $Cov_{(XY)}$, has the same sign as ρ_{XY}, has unusual units and is usually a means to find ρ_{XY}.

Term	Formula	Notes
Correlation	$Corr(XY) = \frac{Cov(XY)}{\sigma_X \sigma_Y}$	Used with Covariance formulas below
Covariance	$Cov(XY) = E[(X-\overline{X})(Y-\overline{Y})]$ *(difficult to calculate)*	Sum of the products of all sample pairs' distance from their respective means multiplied by their respective probabilities
	$Cov(XY) = E(XY) - E(X)E(Y)$	Sum of the products of all sample pairs multiplied by their respective probabilities, minus the product of both means

AKHILESH JAIN

Hypothesis Testing:

Test Type	Test Statistic	Two-tailed H_a	Two-tailed Critical Value	Lower-tail H_a	Lower-tail Critical Value	Upper-tail H_a	Upper-tail Critical Value
Single μ ($n \geq 30$)	$z_0 = \dfrac{\overline{X}-\mu_0}{s/\sqrt{n}}$	$\mu \neq \mu_0$	$\pm z_{\alpha/2}$	$\mu < \mu_0$	$-z_\alpha$	$\mu > \mu_0$	$+z_\alpha$
Single μ ($n < 30$)	$t_0 = \dfrac{\overline{X}-\mu_0}{s/\sqrt{n}}$	$\mu \neq \mu_0$	$\pm t_{(n-1,\alpha/2)}$	$\mu < \mu_0$	$-t_{(n-1,\alpha)}$	$\mu > \mu_0$	$+t_{(n-1,\alpha)}$
Single p ($n \geq 30$)	$z_0 = \dfrac{\hat{p}-p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$	$p \neq p_0$	$\pm z_{\alpha/2}$	$p < p_0$	$-z_\alpha$	$p > p_0$	$+z_\alpha$
Diff. between two μs	$z_0 = \dfrac{(\overline{X}-\overline{Y})-0}{\sqrt{\dfrac{s_x^2}{n_x}+\dfrac{s_Y^2}{n_Y}}}$	$\mu_X - \mu_Y \neq 0$	$\pm z_{\alpha/2}$	$\mu_X - \mu_Y < 0$	$-z_\alpha$	$\mu_X - \mu_Y > 0$	$+z_\alpha$
Diff. between two ps	$z_0 = \dfrac{(\hat{p}_X-\hat{p}_Y)-0}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{n_X+n_Y}{n_X n_Y}\right)}}$	$p_X - p_Y \neq 0$	$\pm z_{\alpha/2}$	$p_X - p_Y < 0$	$-z_\alpha$	$p_X - p_Y > 0$	$+z_\alpha$

Classic Hypothesis Testing Procedure			
Step		Description	Example
1	**Formulate Two Hypotheses**	The hypotheses ought to be mutually exclusive and collectively exhaustive. The hypothesis to be tested (the null hypothesis) always contains an equals sign, referring to some proposed value of a population parameter. The alternative hypothesis never contains an equals sign, but can be either a one-sided or two-sided inequality.	H_0: $\mu = 0$ H_A: $\mu < 0$
2	**Select a Test Statistic**	The test statistic is a standardized estimate of the difference between our sample and some hypothesized population parameter. It answers the question: "*If the null hypothesis were true, how many standard deviations is our sample away from where we expected it to be*?"	$\dfrac{\overline{X}-\mu_0}{s/\sqrt{n}}$
3	**Derive a Decision Rule**	The decision rule consists of regions of rejection and non-rejection, defined by critical values of the test statistic. It is used to establish the probable truth or falsity of the null hypothesis.	We reject H_0 if $\overline{X} < \mu_0 - z_\alpha \dfrac{\sigma}{\sqrt{n}}$.
4	**Calculate the Value of the Test Statistic; Invoke the Decision Rule in light of the Test Statistic**	Either reject the null hypothesis (if the test statistic falls into the rejection region) or do not reject the null hypothesis (if the test statistic does not fall into the rejection region.	$\dfrac{\overline{X}-\mu_0}{s/\sqrt{n}}$ $= \dfrac{-0.21-0}{0.80/\sqrt{50}}$

AKHILESH JAIN

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

t-table

d. f.	0.100	0.050	0.025	0.010	0.005
1	3.078	6.314	12.706	31.821	63.656
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763

AKHILESH