Faculty of Management
Faculté de gestion

MBA - Base Camp
Management Statistics

Seminar Leader: Derek Hart

Base Camp
Session 3

Session Material

August 7/2015

Binomial Probability Distribution

Poisson Probability Distribution

Continuous Probability Distributions

Normality in Regression Modeling (Assumptions)

Read: **Statistics for Management and Economics**

Gerald Keller, 10th edition
Chapter 7 (7.4, and 7.5), Chapter 8 (8.1, 8.2)

DESAUTELS Faculty of Management
Faculté de gestion

Binomial Probability Distribution

McGill

Situation...

The Children's Hospital in Montreal, Quebec recently reported that only 40% of the children who suffer extreme head trauma in an accident survive. In March of 1995, the Children's had nine children who were hospitalized by this type of injury. As part of the effort to coordinate medical facilities, the hospital staff had to arrive at several difficult decisions.

What is the probability that all nine will survive?

The Bernoulli Process

1. Repetitive and statistically independent trials

Notation: n = the predetermined number of trials.

2. Each trial results in one of two mutually exclusive outcomes.

Notation: Traditionally called SUCCESS and FAILURE

3. The probability of success in any one trial is the same from trial to trial.

Notation: $P(\text{success}) = p$ and $P(\text{failure}) = q$

also $p + q = 1$

Binomial Random Variable

Define X = The number of successes in n trials.

then,

the probability distribution of X is called the *Binomial Probability Distribution*.

Notation: X~ B(n, p) [X has a Binomial Distribution with n trials and a probability of success equal to p].

and P(X=x/ p) means the probability that the RV X takes on a value x(successes).

Situation…

The Children's Hospital in Montreal, Quebec recently reported that only 40% of the children who suffer extreme head trauma in an accident survive. In March of 1995, the Children's had nine children who were hospitalized by this type of injury. As part of the effort to coordinate medical facilities, the hospital staff had to arrive at several difficult decisions.

What is the probability that all nine will survive?

Probability Calculation

All nine survive means 9 successes
that is,

$$\boxed{P(s \cap s \cap s \ldots \cap s)}$$

The intersection of nine successes,

but each trial is independent.

So we obtain, $\boxed{P(S) * P(S) * P(S) * \ldots. * P(S)}$

Or 0.4 * 0.4 * 0.4…… *0.4

$(0.4)^9 = 0.000262$

DESAUTELS Faculty of Management
Faculté de gestion

McGill
Base Camp

Probability Calculation

Calculate the probability that 1 survives.

That is, $P(X=1/p=.4)$

$$P(S \cap F \cap F \cap .... \cap F) = P(S) * P(F) * P(F) * ... * P(F)$$

Or $= 0.4 * 0.6 * 0.6 * ... * 0.6$

$= (0.4)^1 * (0.6)^8$

But there are nine ways of finding one success and each has a probability equal to $(0.4)*(0.6)^8$

Therefore

$P(X=1/p=.4) = 9*(0.4)^1*(0.6)^8 = 0.0605$

Binomial Probability Function

In general, if X~ B(n,p)
then

$$P(X = x) = C_x^n * p^x * q^{n-x}$$

Where

$$C_x^n = \frac{n!}{(n-x)! * x!}$$

$$\therefore \quad P(X = 5 / p = 0.4) = C_5^9 \times (0.4)^5 \times (0.6)^4$$

$$= \frac{9!}{(9-5)! \times 5!} \times (0.4)^5 \times (0.6)^4 = 0.1672$$

Probability Distribution

If the probability for each number of successes is calculated, they can be listed in a tabular format.

X	P(X=x)
0	0.0101
1	0.0605
2	0.1612
3	0.2508
4	0.2508
5	0.1672
6	0.0743
7	0.0212
8	0.0035
9	0.0003
Total	1.000

Binomial Probability Calculation

Given X~ B(0.4,9)

Calculate P(X>3/p=0.4)

If we refer to the table on the previous slide

$$
\begin{aligned}
P(X>3) &= P(X=4)+P(X=5)+..+P(X=8)+P(X=9) \\
&= 0.2508 + 0.1672+\ldots+ 0.0035 + 0.0003 \\
&= 0.5173
\end{aligned}
$$

that is, there is a 51.73% chance that more than 3 children will survive.

Alternative Approach

$$P(X>3/p=0.4) = 1 - P(X \leq 3)$$
$$= 1 - [\, P(X=3)+P(X=2)+\ldots+P(X=0)]$$
$$= 1 - [\, 0.2508+0.1612+\ldots+0.0101]$$
$$= 1 - 0.4827$$
$$= 0.5173$$

Expected Value and Standard Deviation

The expected number that will survive can be calculated by using the E(X) formula.

X	P(X=x)	X*P(X)
0	0.0101	0
1	0.0605	0.0605
2	0.1612	0.3224
3	0.2508	0.7524
4	0.2508	1.0032
5	0.1672	0.836
6	0.0743	0.4458
7	0.0212	0.1484
8	0.0035	0.028
9	0.0003	0.0027
	1.000	3.5994

The expected value is 3.5994 or 3.6

Note: E(X) = n p

= 9*0.4

= 3.6

E(X) and SD(X)

If X is a Binomial random variable with a probability of success equal to p and the number of trials n, then

$$\boxed{\begin{array}{l} E(X) = np \\ and \\ SD(X) = \sqrt{npq} \end{array}}$$

Therefore, in the hospital example

$$E(X) = np = 0.4 \times 9 = 3.6$$

$$SD(X) = \sqrt{9 \times 0.4 \times 0.6} = \sqrt{2.16}$$
$$= 1.47$$

Faculty of Management
Faculté de gestion

Poisson Probability Distribution

Poisson Distribution

Let X = the number of occurrences of an event in a
given interval t

Examples:

number of accidents in Montreal per year

number of car arrivals at a tollbooth in one minute

number of defects in 100m of textiles

Characterized by a number of discrete, independent successes occurring on a continuous scale.

Properties of the Distribution

- The number of occurrences in two non-overlapping intervals are independent.
- The probability that a success will occur in an interval is the same for all intervals of equal size and proportional to the size of the interval.
- The probability that two or more successes will occur in an interval approaches zero as the interval becomes smaller.

Poisson Probability Function

λ = the average rate of occurrences in the given process per unit length(per second, per meter, etc.)

X = the number of successes within a defined length t (i.e. every 10 seconds, every 20 meter, etc.)

therefore the average (expected) number of occurrences within the length t is given by $\mu = \lambda t$

$$P(X = x) = \frac{\mu^{x} e^{-\mu}}{x!}, x = 0,1,2,3,\ldots.$$

Example:

Suppose in a carpet manufacturing process, an average of 2 flaws occur per 10 running meters of material. What is the probability that a given 10-m segment will have one or fewer flaws?

Note: *Since we have 2 flaws per 10-m therefore*

λ = 2/10 = 0.2 flaws per meter

since t = 10 μ = 10 * 0.2 = 2

$P(X \leq 1) = P(X=0) + P(X=1) =$

$= 0.1353 + 0.2707$

$= 0.4060$

Poisson Probability Tables

These tables are found at the back of the Textbook

x\mean	mean=2
0	0.1353
1	0.2707
2	0.2710
etc	

Note: Many tables are given in the cumulative form therefore $P(X=x) = P(X \leq x) - P(X \leq x-1)$

Question: P(more than one flaw in 20 - m of material)
$\mu = \lambda t = 0.2 \times 20 = 4$ flaws per 20 meters

$P(X > 1) = 1 - P(X \leq 1) = 1 - 0.09158$
$= 0.9084$

Note: The value of λ must be adjusted to correspond to the interval under consideration

P(2 flaws in 20 meters) $\neq$ P(1 flaw in 10 meters)

$\mu = 4$ 0.1465 $\neq$ 0.2707 $\mu = 2$

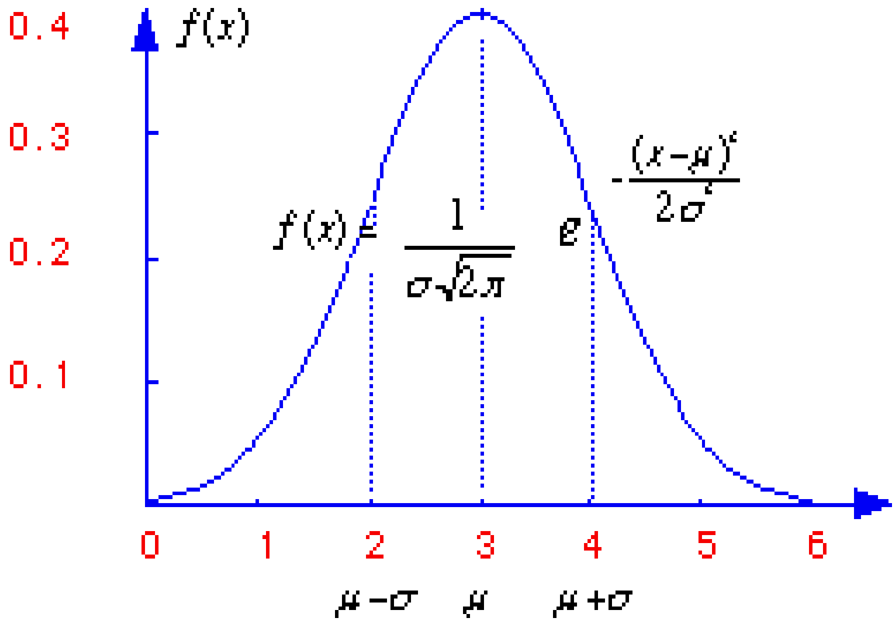
Normal Probability Distribution

Probability Distribution of a Continuous Random Variable

Examples of Normally Distributed RV's

Human weight, height, IQ
Wingspan of Dragon-flies
Cost of Textbooks
Monthly returns on equity
Etc….

The Gaussian Equation



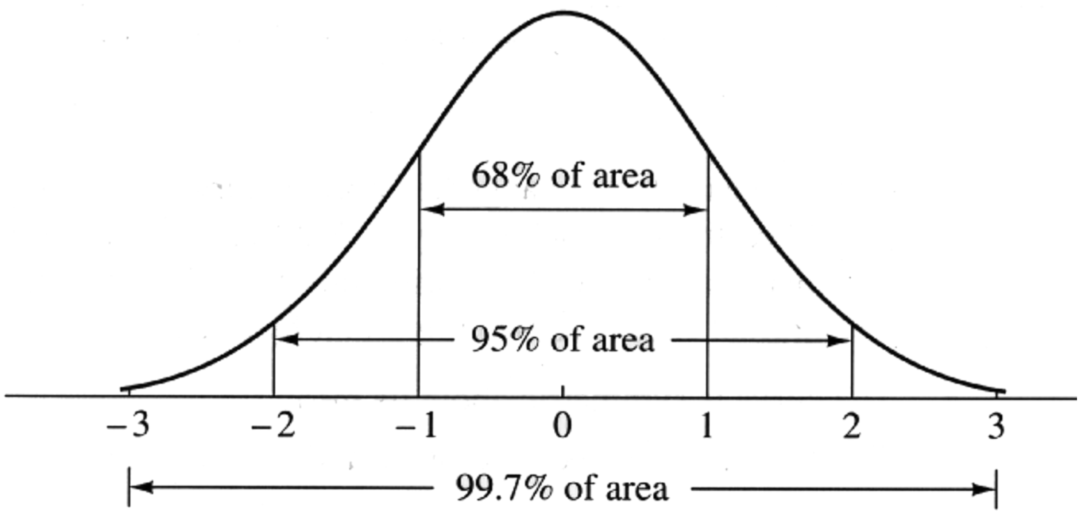
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

f(x) has two main properties:

- **f(x) is nonnegative.**
- **The total area under the curve is equal to 1.**

The Empirical Rule

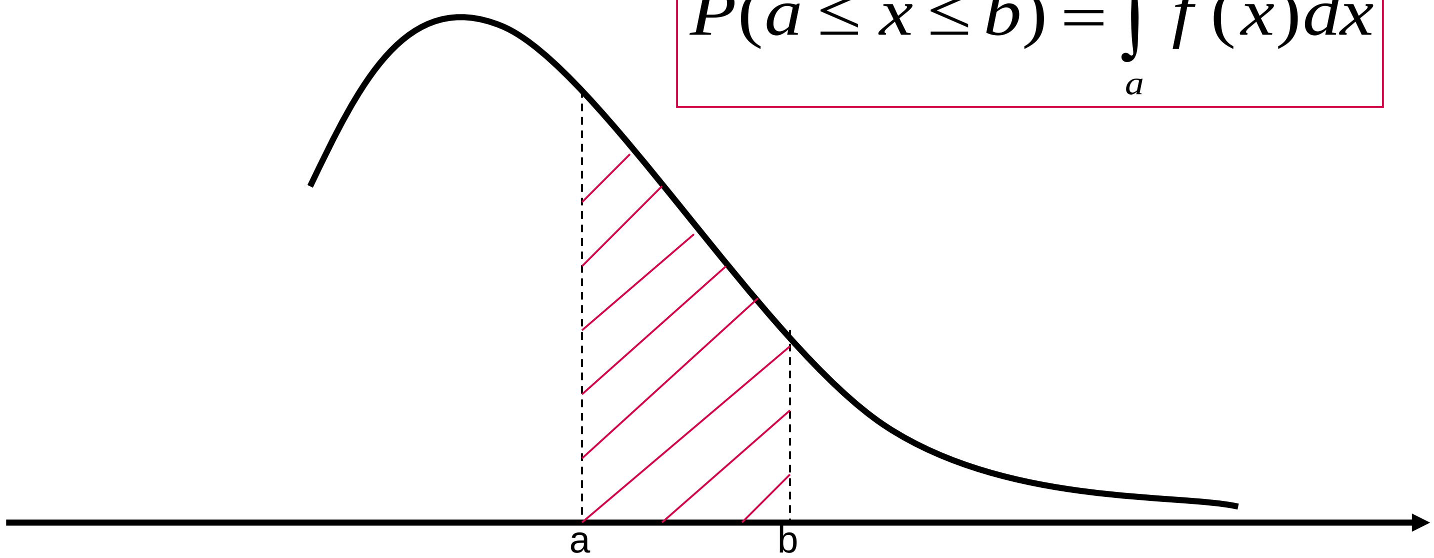


- if 68 % of the data is in the interval (x ± s)
- if 95 % of the data is in the interval (x ± 2s), and
- if 99.7 % (i.e. virtually all) of the data is in the interval (x ± 3s)

Calculating Probability

Normal curve f(x)

$$P(a \le x \le b) = \int_a^b f(x)dx$$



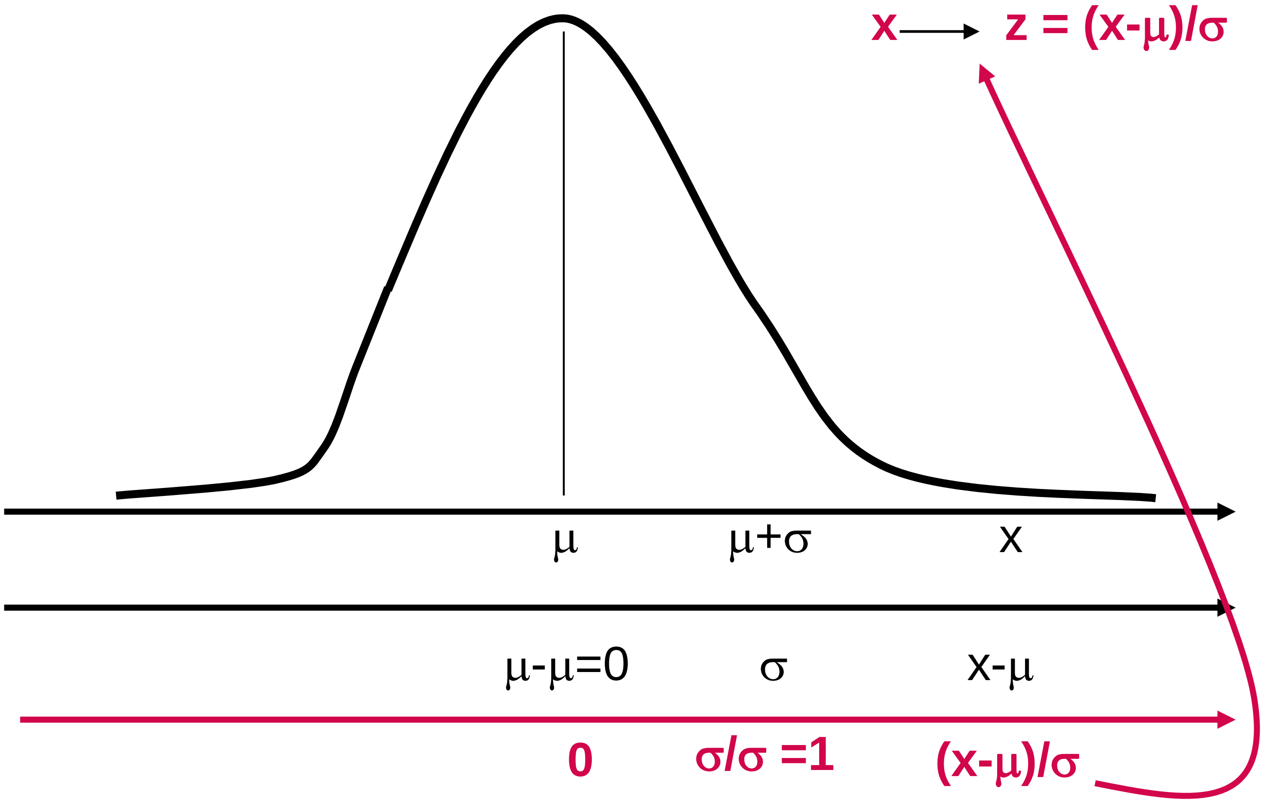
The probability that a randomly selected x lies between a and b is equal to the **area under the curve between a and b**.

DESAUTELS Faculty of Management Faculté de gestion

McGill Base Camp

The Normal Z-value

We start with a normal RV $X \sim N(\mu, \sigma)$
which is standardized into $Z \sim N(0, 1)$



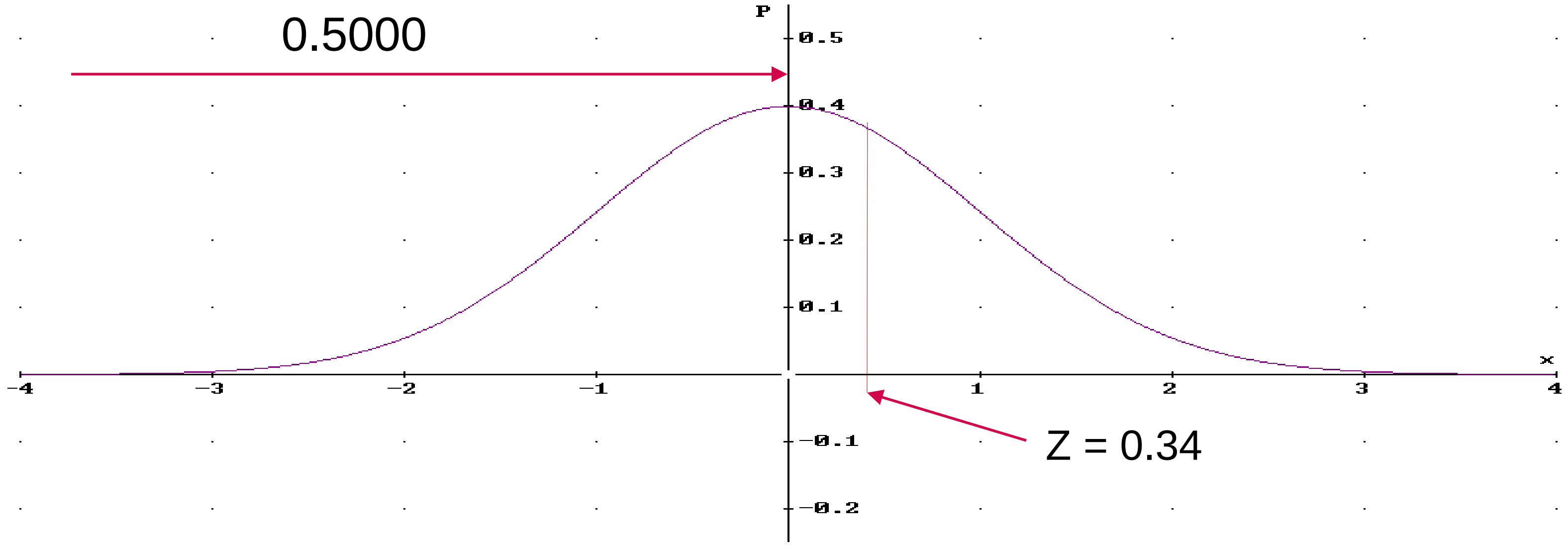
DESAUTELS Faculty of Management Faculté de gestion

McGill Base Camp

TABLE A Standard normal probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681

Z-Table



P(0 < Z < 0.34) = 0.6331 – 0.5000

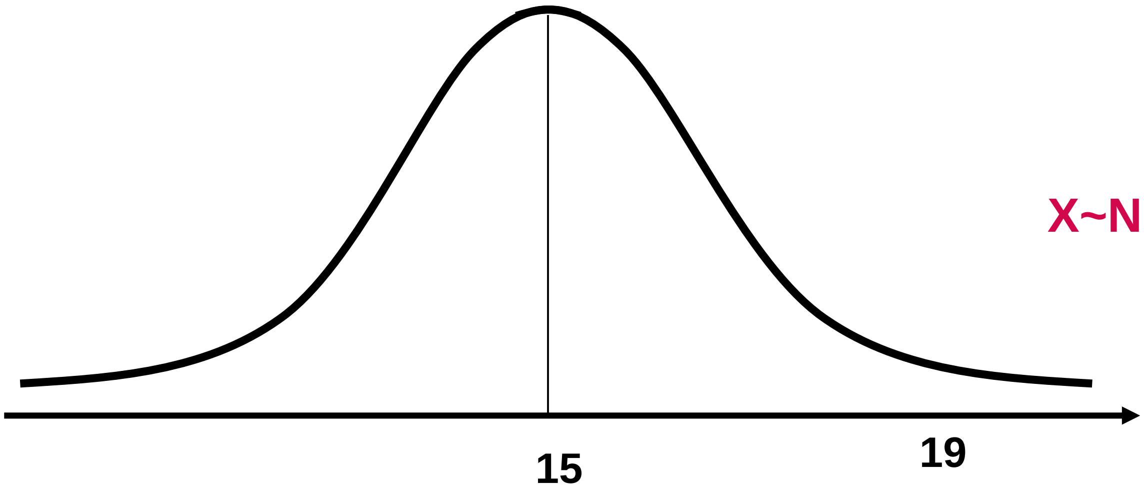
= 0.1331

TABLE A Standard normal probabilities (*continued*)

z	.00	.01	.02	.03	.04
0.0	.5000	.5040	.5080	.5120	.5160
0.1	.5398	.5438	.5478	.5517	.5557
0.2	.5793	.5832	.5871	.5910	.5948
0.3	.6179	.6217	.6255	.6293	$\boxed{.6331}$
0.4	.6554	.6591	.6628	.6664	.6700
0.5	.6915	.6950	.6985	.7019	.7054
0.6	.7257	.7291	.7324	.7357	.7389
0.7	.7580	.7611	.7642	.7673	.7704
0.8	.7881	.7910	.7939	.7967	.7995
0.9	.8159	.8186	.8212	.8238	.8264
1.0	.8413	.8438	.8461	.8485	.8508
1.1	.8643	.8665	.8686	.8708	.8729
1.2	.8849	.8869	.8888	.8907	.8925
1.3	.9032	.9049	.9066	.9082	.9099

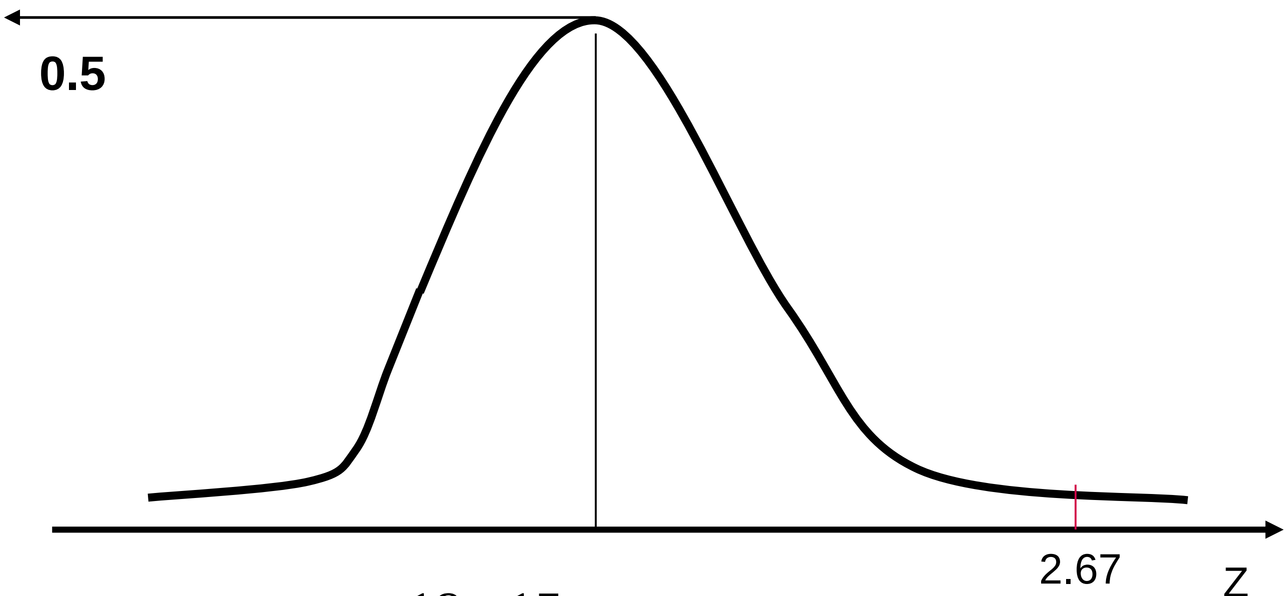
Example

The time required to complete a task on a manufacturing line is normally distributed with a mean time of 15 minutes and a standard deviation of 1.5 minutes. If the next stage of the process must begin at most 19 minutes from the beginning of this task, what is the probability that the system stays within the time frame?



X~N(15,1.5)

Example cont'd



$$P(x \le 19) = P(z \le \frac{19-15}{1.5})$$

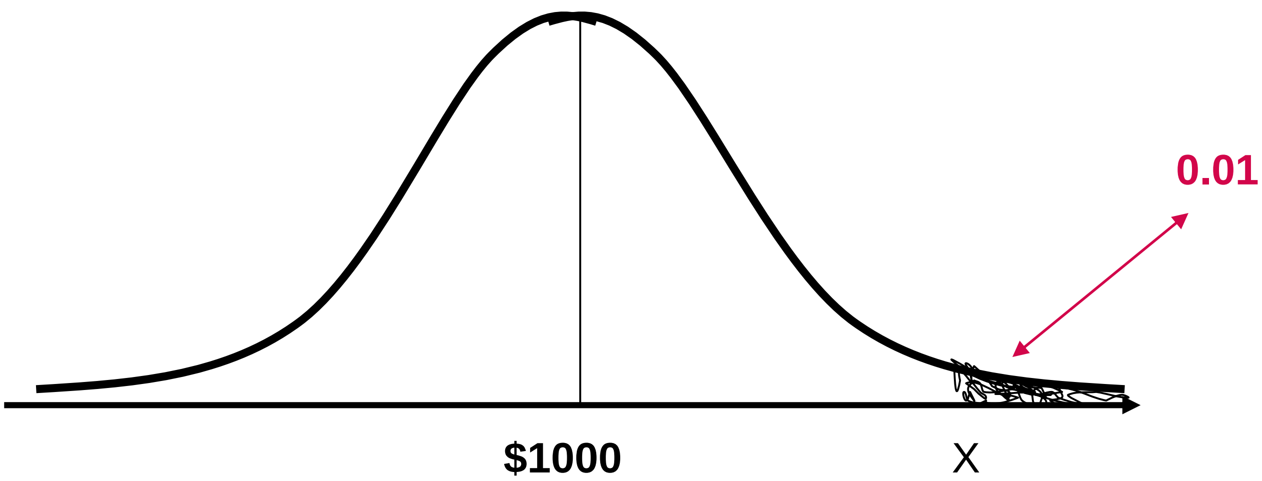
$$P(z \le 2.67) = 0.9962$$

That is, there is a 99.62% probability that the process will stay within the time frame

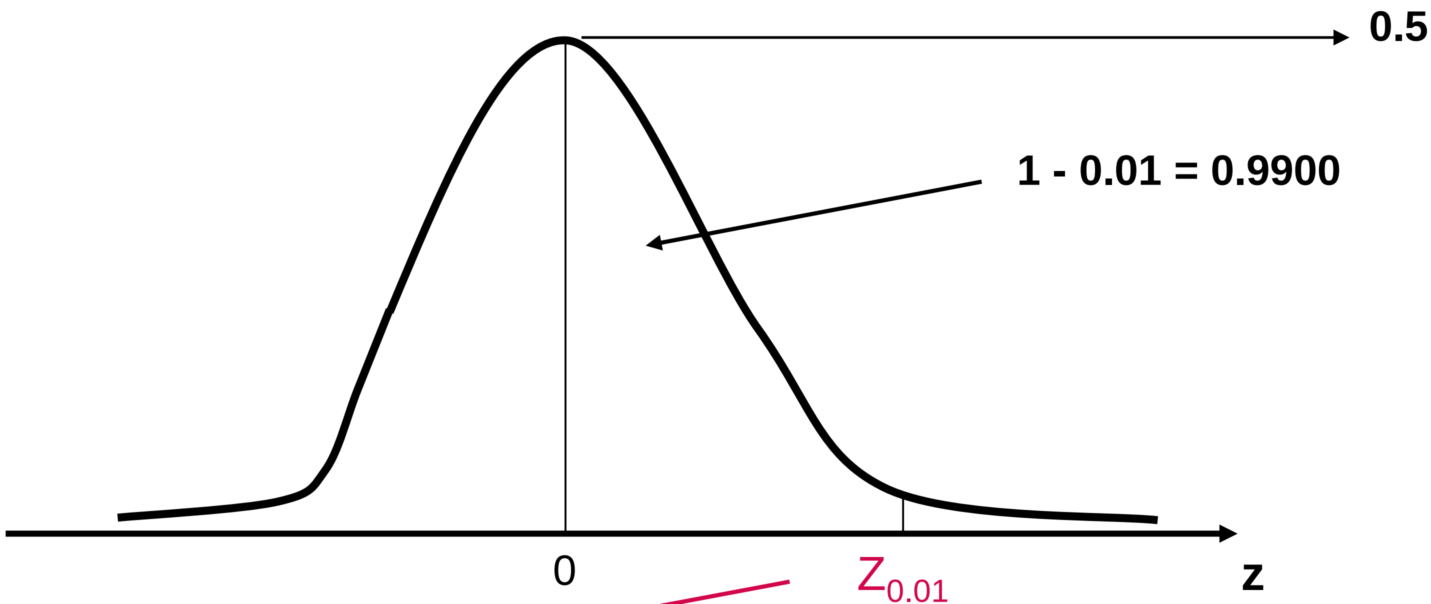
Example

The top 1% of bonuses paid to a particular employee group fall into a category called, "unusually high". If a company's bonuses are normally distributed with a mean of $1000 and standard deviation of $250, find the cutoff beyond which the bonus is unusually high.

X~N($1000,$250)



Example cont'd



$Z_{0.01} = 2.33$

But $Z = (X - \mu)/\sigma$

therefore $2.33 = (X - 1000) / 250$

or $X = 1000 + (2.33)(250)$

$X = \$ 1582.50$

Normal Probability Distribution

Example of Normally Distributed RV's

The time required to load a moving van is normally distributed with a mean of 1 hour and 45 minutes and a standard deviation of 18 minutes. If the time required to drive a moving van from the warehouse to the airport is also normally distributed with a mean of half an hour and a standard deviation of 12 minutes, determine the probability than a van can be loaded and driven to the airport in less than 2 hours. Assume that the sum of normal random variables is also normally distributed. What other assumption is necessary to obtain your answer?

Normal Probability Distribution

Let x = load time   x ~ N(105, 18)
Let y = drive time y ~ N(30, 12)

Since x and y are normally distributed, therefore (x + y) is normal

E(x + y) = 105 + 30 = 135

x and y must be assumed independent

V(x + y) = $18^2 + 12^2$ = 468 therefore SD(x + y) = 21.633

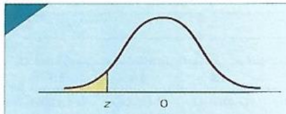
(x + y) ~ N(135, 21.633)

so P(x+y < 120) = P(z< (120 – 135)/ 21.633) = P(z< -0.69) = 0.2451

i.e. **the probability that the total time is less than 2 hours is 0.2451**

Normal Probabilities

TABLE 3 **Cumulative Standardized Normal Probabilities**



$P(-\infty < Z < z)$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
−0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

The Nscore Plot
or
The Normal Probability Plot

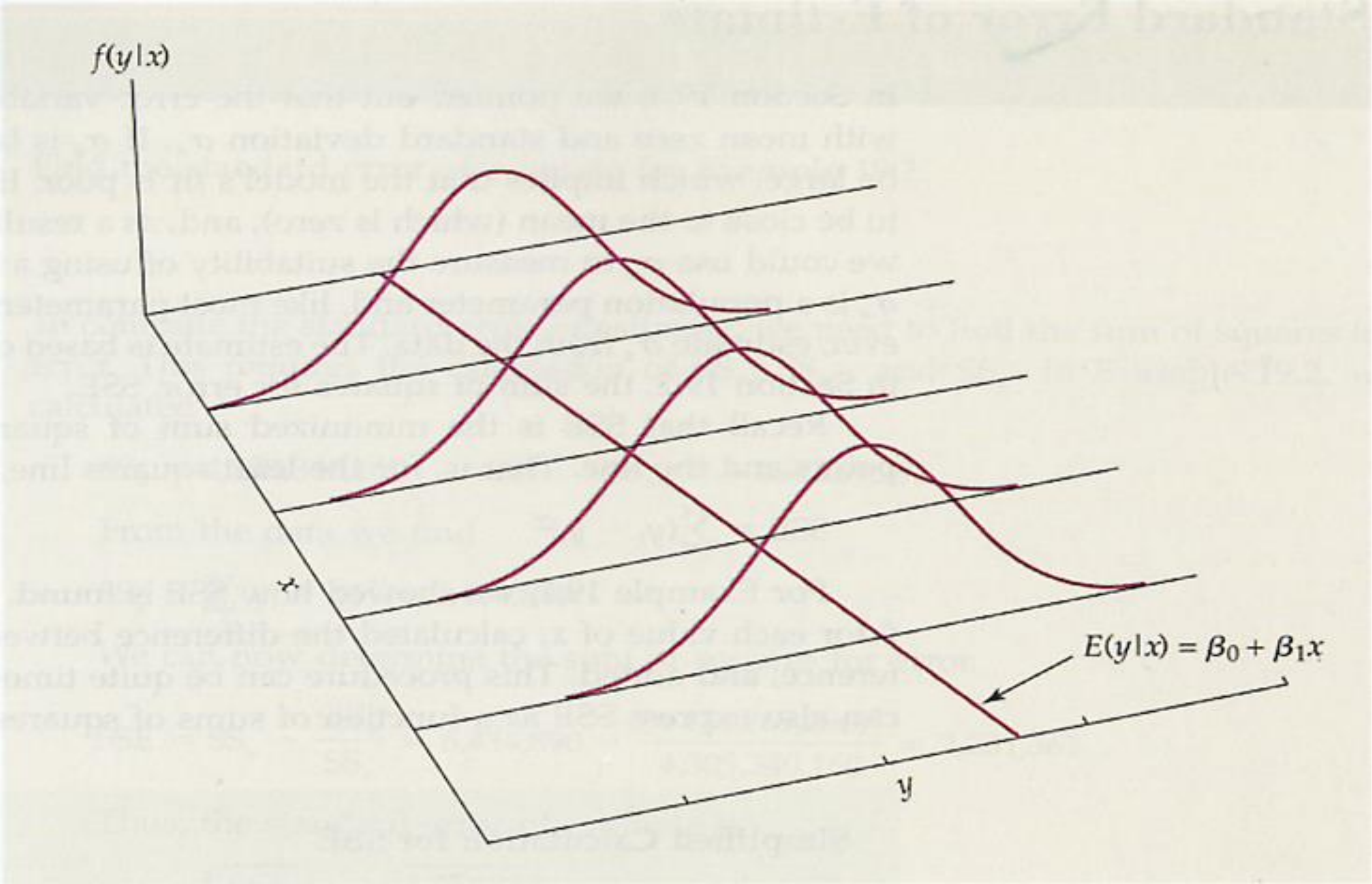
Recall: Assumptions of the Model

$$\boxed{y_i = b_0 + b_1 x_i + e_i}$$

Random Variable

1. **The e_i's are bell-shaped (normal).**
2. The mean of the e_i's is equal to zero.
3. The standard deviation of the e_i's is the same for all x_i.
4. The e_i's are statistically independent.

Assumptions



Normal Scores (Nscores) or Normal Probability Plot

Recall if X is Normally Distributed, then $X = \mu + Z * \sigma$

that is, X and Z are related in a linear equation

If they are plotted we would get a straight line.

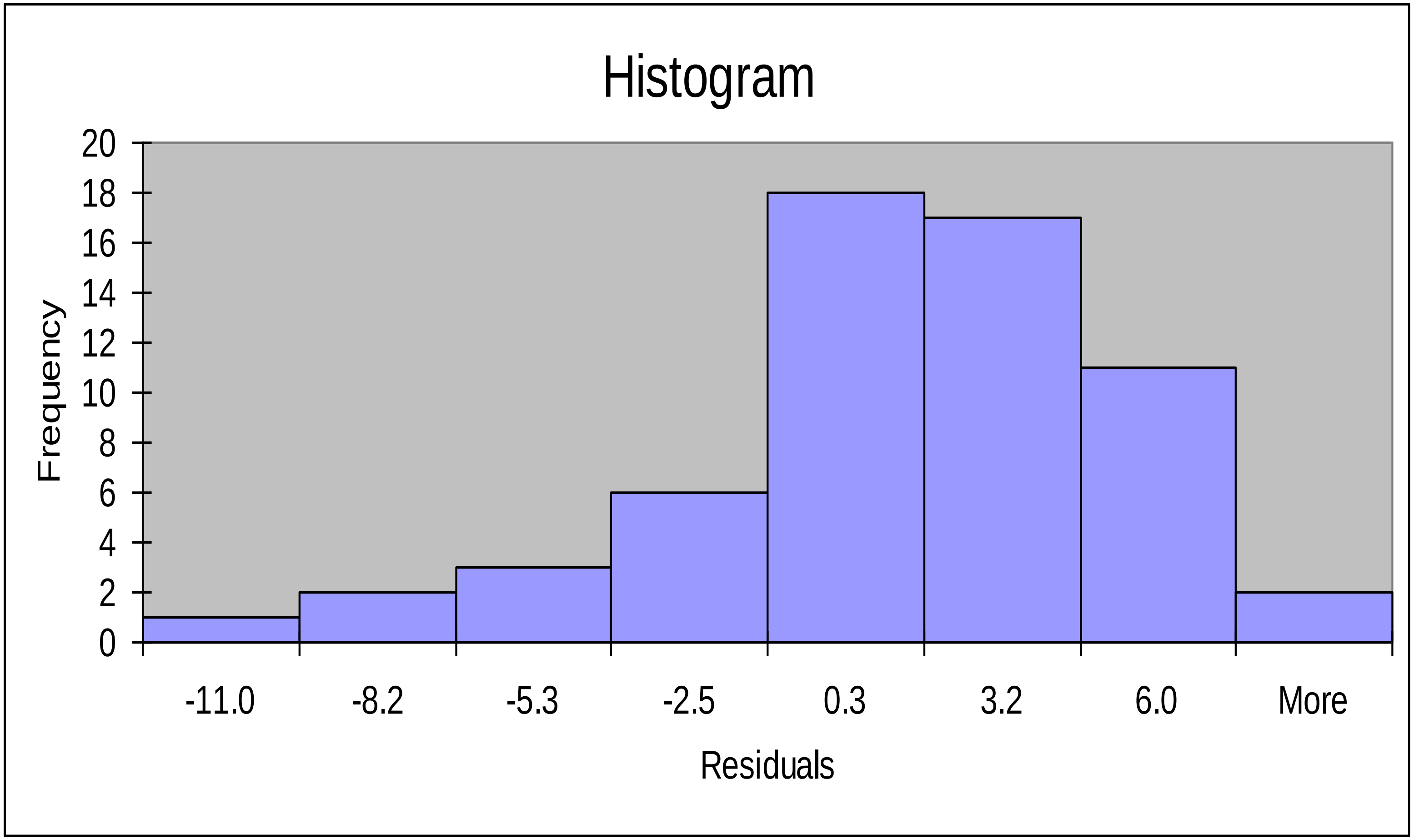
Given this fact, we can test how close a sample data set fits a normal distribution by plotting the ranked data against the ideal standardized scores. These scores are called the Nscores. The graph is called the Nscores plot or Normal Probability Plot

Example : Residuals for Market Price versus Area

RESIDUAL OUTPUT

Observation	Predicted	Residuals	Std. Residuals
1	30.075	-4.075	-1.032
2	30.387	-10.987	-2.783
3	30.497	-5.297	-1.342
4	31.103	-4.903	-1.242
5	32.645	-1.645	-0.417
6	32.664	1.936	0.491
7	32.939	3.461	0.877
8	33.196	-0.196	-0.050
9	33.251	4.149	1.051
10	33.582	8.818	2.234
11	33.747	-0.947	-0.240
12	33.765	-8.165	-2.069
13	34.151	0.649	0.164
14	34.463	1.337	0.339
15	34.592	-0.992	-0.251
16	34.830	-3.830	-0.970
17	34.959	4.241	1.074
18	35.234	0.766	0.194
19	35.455	-0.655	-0.166
20	35.473	-1.073	-0.272
21	35.657	2.343	0.594
22	35.822	-1.222	-0.310
23	35.895	-0.295	-0.075
24	36.262	-0.462	-0.117
25	36.722	2.878	0.729
26	36.850	-1.850	-0.469
27	37.015	0.585	0.148
28	37.364	3.836	0.972
29	37.401	-6.201	-1.571
30	37.456	-7.456	-1.889
31	37.511	-0.111	-0.028
32	37.603	0.397	0.101
33	38.227	-1.027	-0.260
34	38.246	5.754	1.458
35	38.264	5.936	1.504
36	39.072	4.528	1.147
37	39.072	-0.672	-0.170
38	39.311	2.889	0.732
39	39.476	0.924	0.234
40	39.604	0.796	0.202
41	39.733	3.867	0.980
42	39.806	1.594	0.404
43	39.825	-0.225	-0.057
44	39.898	1.902	0.482
45	39.972	4.828	1.223
46	39.972	-1.572	-0.398
47	40.155	3.445	0.873
48	40.247	2.553	0.647
49	40.320	0.280	0.071
50	40.706	0.894	0.226
51	40.816	1.984	0.503
52	41.404	-2.404	-0.609
53	41.881	-0.081	-0.021
54	42.010	6.390	1.619
55	43.222	-3.422	-0.867
56	43.442	3.758	0.952
57	44.342	0.858	0.217
58	46.857	-8.057	-2.041
59	49.924	-2.524	-0.639
60	53.633	-8.233	-2.086

Histogram of Residuals



Check for Normality Assumption