Probability Distribution

Pro Dr Najlaa Fawzi

What Is a Probability Distribution?

A probability distribution is a statistical function that describes all the possible values and likelihoods that a random variable can take within a given range.

This range will be bounded between the minimum and maximum possible values, but precisely where the possible value is likely to be plotted on the probability distribution depends on several factors. These factors include the distribution's mean (average), standard deviation, skewness, and kurtosis.

PROBABILITY DISTRIBUTION

There are two types:

Probability distribution of continuous variables.

Probability distribution of discrete variables.

Probability Distribution

There are different theoretical types of probability distribution :

1- Binomial prob distribution

2- Poisson prob distribution

3- Normal prob distribution

4- Skewed prob distribution

Binomial prob distribution

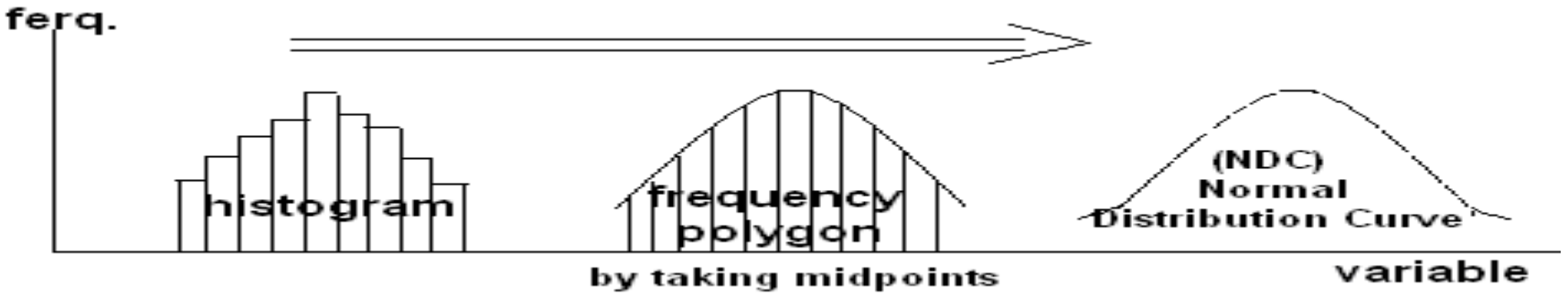
The binomial distribution, for example, evaluates the probability of an event occurring several times over a given number of trials and given the event's probability in each trial.

Coin and figuring the probability of that coin coming up heads in 10 straight flips. A binomial distribution is *discrete*, as opposed to continuous, since only 1 or 0 is a valid response.

The Poisson distribution **is the discrete probability distribution of the number of events occurring in a given time period, given the average number of times the event occurs over that time period.**

A Poisson Process is a model for a series of discrete event where the average time between events is known, but the exact timing of events is random.

A Poisson distribution can be used to analyze the probability of various events regarding how many customers go through the drive-through. It can allow one to calculate the probability of a pause in activity (when there are 0 customers coming to the drive-through) as well as the probability of a spell of activity (when there are 5 or more customers coming to the drive-through).
This information can, in turn, help a manager plan for these events with staffing and scheduling.



NORMAL DISTRIBUTION (*of continuous variables*):

it is the most important distribution in statistics and mostly used. Parameters include population mean (μ) which is the measure of central tendency, and the standard deviation (δ) which is the measure of dispersion.

If we have a group of continuous variables with certain class interval,

we can represent them by histogram and frequency polygon.

we have a group of variables which is huge and the class interval is very small so the frequency polygon will take a shape of very smooth curve and that curve is called "normal distribution curve" or Gaussian curve .

Characteristics of normal distribution curve include:

1. It is used for continuous variables only.
2. Symmetrical around its mean i.e. the right side is equal to the left.
3. The mean, mode and median are equal.
4. Total probability under the curve (area under the curve - AUC -) equals to one.
5. 50% of AUC lies to the right of the median & 50% to the left.

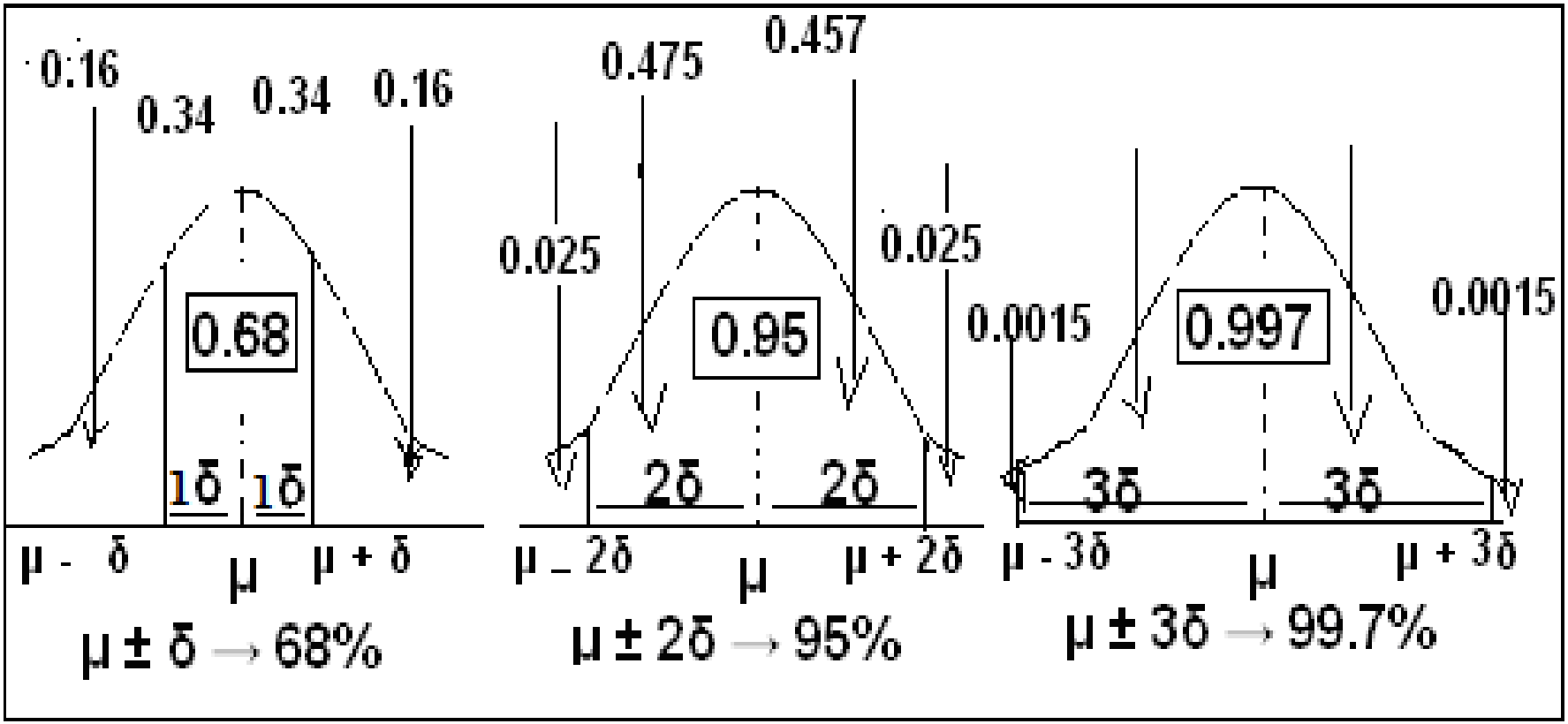
6.Probability Limits around the mean: If you move by one standard deviation (δ) away from the mean on each side; the AUC limited by ±

1δ equals to 68% of total AUC, & so: μ ± 1δ → 68%,

μ ± 2δ ⟶ 95%,

μ ± 3δ ⟶ 99 .7%, and as 99.7% ≈ 1(or100%)

so AUC ≈ 6δ (3 on each side of μ).

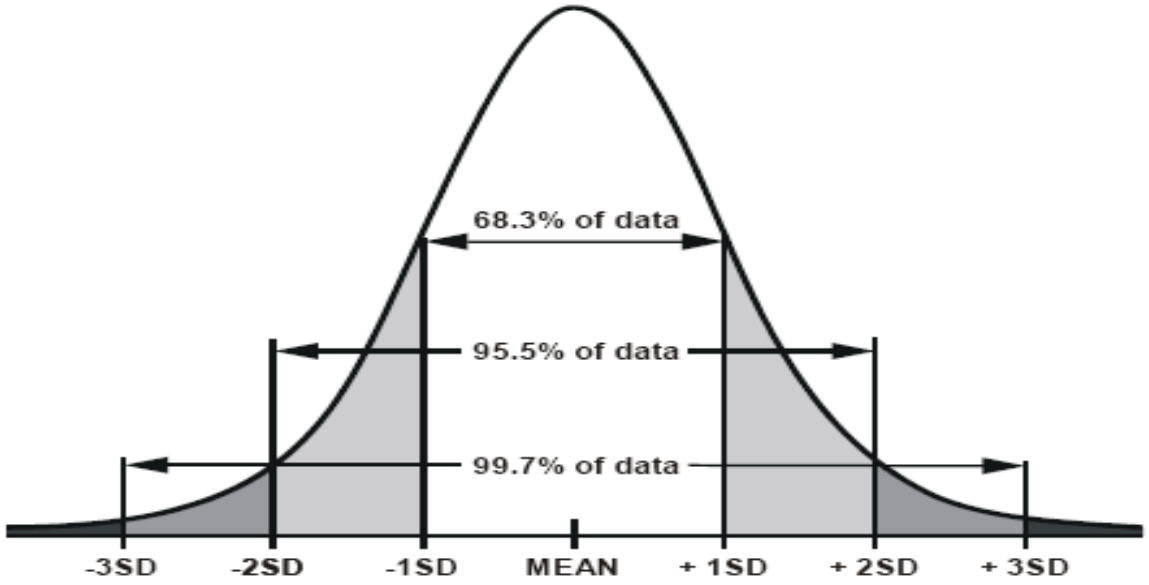

0.16
0.34
0.34
0.16
0.68
1δ
1δ
μ - δ
μ
μ + δ
μ ± δ → 68%
0.025
0.475
0.457
0.025
0.95
2δ
2δ
μ – 2δ
μ
μ + 2δ
μ ± 2δ → 95%
0.0015
0.997
0.0015
3δ
3δ
μ - 3δ
μ
μ + 3δ
μ ± 3δ → 99.7%

7.Different values of μ and δ shift the graph of distribution along X & Y axes.

If we change μ while keeping δ constant, the curve will shift to the right on increasing μ & to the left on decreasing μ.

On changing δ and keeping μ constant; the curve will become more flat on increasing δ and narrower on decreasing δ without any shifting the curve to any side. μ1 < μ2 < μ3

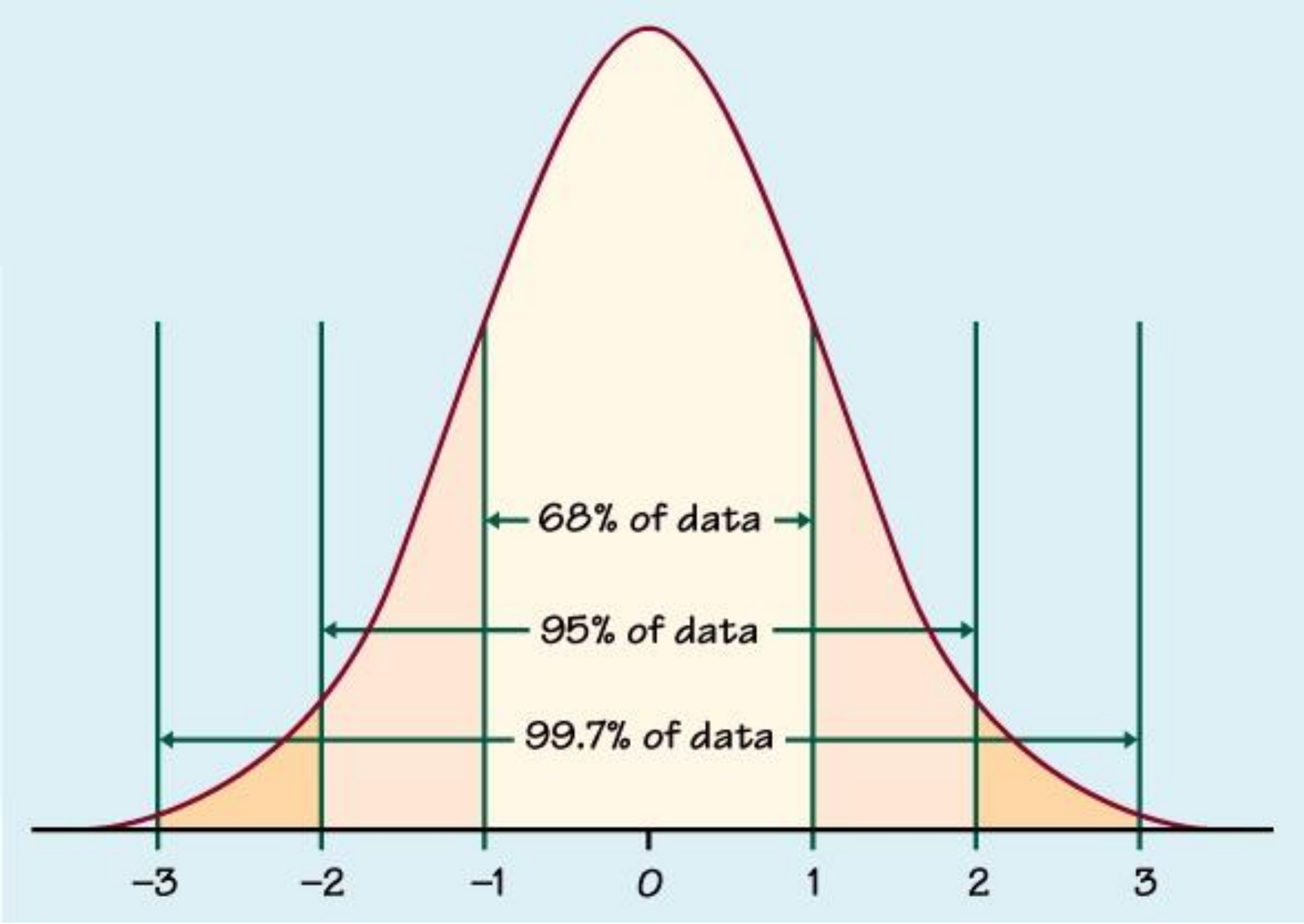
Figure 3.9
Areas under the normal curve that lie between 1, 2, and 3 standard deviations on each side of the mean



The mean and standard deviation can be presented as a sort of shorthand to describe normally distributed data. Consider, for example, serum cholesterol levels of a representative sample of several thousand men in their mid-30's. We could list the serum cholesterol level for each man, or show a frequency distribution, or simply report the mean value and standard deviation. The frequency distribution is shown in Table 3.4. We can further summarize these data by reporting a mean of 213 and a standard deviation of 42.

Σ = (Greek letter sigma) = sum of
n or N = the number of observations
f_i = frequency of x_i

x_i = i-th observation (x_1=1st observation, x_4=4th observation)


68% of data
95% of data
99.7% of data
−3
−2
−1
0
1
2
3

Example: If population mean of systolic blood pressure is 120 mmHg with population standard deviation of 10 mmHg. What is the probability of getting a patient with systolic BP

a) Between 120 and 130 mmHg

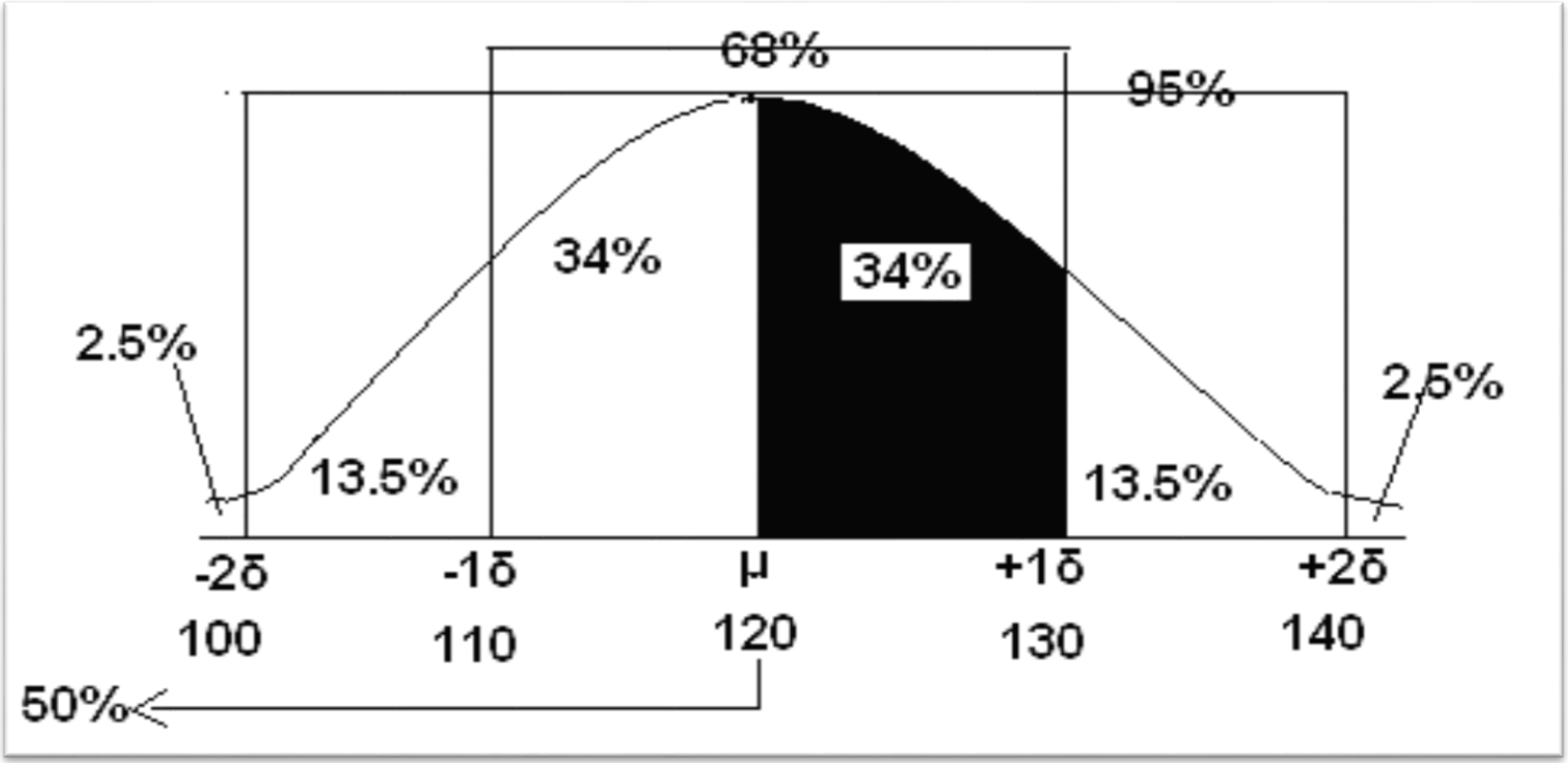
b) < 120mmHg

c) < 100 mmHg

d) > 140mmHg

e) Between 120 and 125 mmHg

f) > 135mmHg


68%
95%
34%
34%
2.5%
2.5%
13.5%
13.5%
-2δ
-1δ
μ
+1δ
+2δ
100
110
120
130
140
50%

a.From 120 to 130 we move one standard deviation, so the probability is 34% (0.34) (i.e. half the 68%).

b. Probability of less than 120 mmHg is 50%.

c. Probability of less than 100 mmHg is 2.5%.

d. Probability of more than 140 mmHg is 2.5%.

In an ideal distribution of values , the mean , median ,and mode are equal within the population under study .

If a set of values includes disproportionately larger number of high values than low or vice-versa, the most common value {mode} may differ from the average value .

Asymmetry curves are generated and these , are said to be skewed .

Under these circumstances , the median may be more representative .

If the difference [mean – median] is positive , or if the mean is greater than the mode , the curve is positively skewed .

If the [mean– median] is negative , or its mean is less than its mode, the curve is negatively skewed.

Coefficient of skew ness in symmetrical distribution {NDC} is = zero

The distribution is said to be **positively skewed** when it has large number of low scores and small number of very high scores whereas **negatively skewed** distribution have a relatively large number of high scores and small number of low scores .

In a positive (right) skewed curve **Mean > Median > Mode**

In a negative (left) skewed curve **Mean < Median < Mode**

Kurtosis is a measure of the degree to which a distribution is "**peaked" or flat** in comparison to a normal distribution whose graph is characterized by a bell- shaped appearance.

Symmetric curves may show kurtosis in accordance with whether they are unusually **peaked [leptokurtic],** or **flattened [platykurtic]**

The ideal curve is **mesokurtic.**

B is called the **coefficient of kurtosis.**

Its value for normal curve is three ; if the B is higher than three , the curve is more flattened . If B is lower than 3, degree of flatness is lower.

Preferences for the use of normal distribution :

1- The ND has been extensively and accurately tabulated consequently ,

if it seems to apply fairly well to a problem , the investigator has many time – saving tables ready at hand.

2- The distribution of many variables are approximately normal ,

such as height, body mass index, weight.

3- With measurements whose distributions are not normal , a simple transformation of the scale of measurements may induce approximate normality , square root , and the log X are often used as transformations.

Those transformations are found useful for flexible use of some tests of significance like student's ' t – test ' .

4- Even if the distribution of sample averages tends to become normal , under a wide variety of conditions , as the size of the sample increases . This is the single most important reason for the use of the normal distribution.

A Z-score

is a numerical measurement used in statistics of a value's relationship to the mean (average) of a group of values, measured in terms of standard deviations from the mean.

If a Z-score is 0, it indicates that the data point's score is identical to the mean score.

A Z-score of 1.0 would indicate a value that is one standard deviation from the mean.

Z-scores may be positive or negative, with a positive value indicating the score is above the mean and a negative score indicating it is below the mean.

Z- score : The location of any element in a normal distribution can be expressed in terms of how many SD it lies above or below the mean of the distribution .

Is the relative deviate or standard normal deviate .

Z score

It represents the distance of a value (x) from the (mean) in units of SD

It is a standard score, scores converted into standard units.

This is for presentation of a measurement (X) from the reference mean or median in terms of SD.

It is as measurement Z= (X- mean)/ SD

Therefore , also known as standardized scores or SD scores.

The Z-score, is the number of standard deviations a given data point lies from the mean.

Here is how to interpret z-scores.

- **A z-score less than 0 represents an element less than the mean.**

- **A z-score greater than 0 represents an element greater than the mean.**

- **A z-score equal to 0 represents an element equal to the mean.**

- **A z-score equal to 1 represents an element that is 1 standard deviation greater than the mean; a z-score equal to 2, 2 standard deviations greater than the mean; etc.**

A z-score equal to -1 represents an element that is 1 standard deviation less than the mean; a z-score equal to -2, 2 standard deviations less than the mean; etc.

Z - Scale: for each value of z there is a specified probability illustrated in z table, this is done by using a table to obtain a probability associated with any normal distribution provided that the mean and the standard deviation are known.

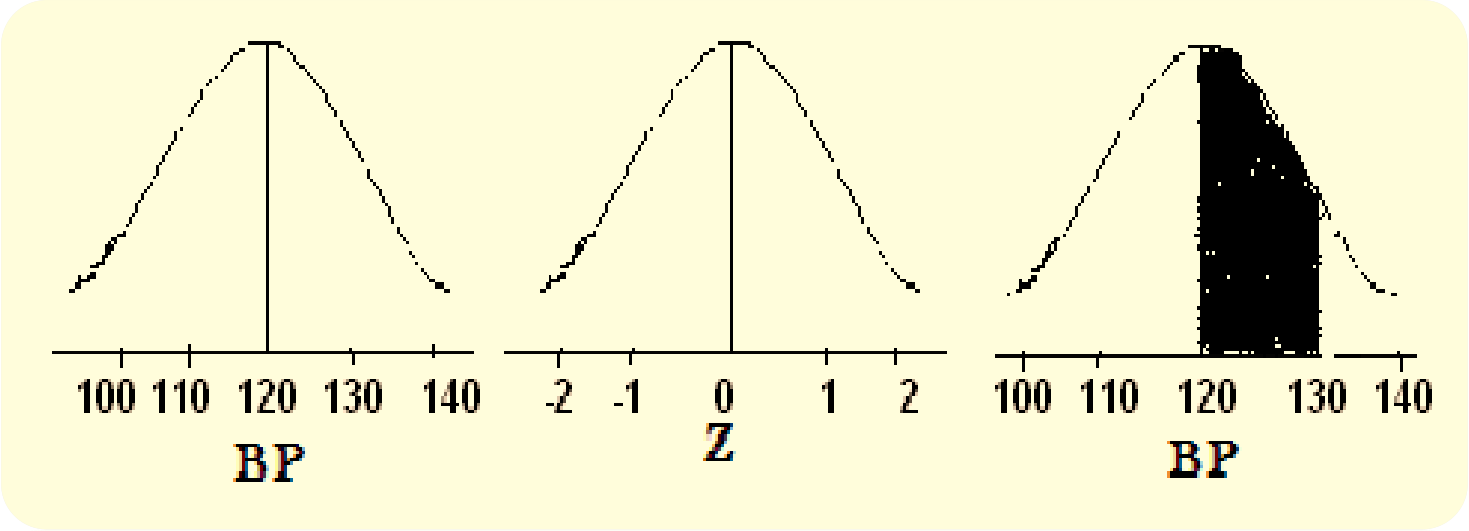
So any continuous variable follows the normal distribution follows Z – scale, e.g. weight, height, s. cholesterol, etc.

Where: x is the value that you want to know its proportion.

$$Z = \text{no. of standard deviations} = (x - \mu) / \delta$$

If we go back to the same previous example of SBP:

Z = no. of δ from μ to x = 120-120/10 = 0 for 120
= 130-120/10 = 1 for 130
= 140-120/10 = 2 for 140
= 110-120/10 = -1 for 110
= 100-120/10 = -2 for 100
= 125-120/10 = .5 for 125
= 140-120/10 = 2 for 140
= 135-120/10 = 1.5 for 135



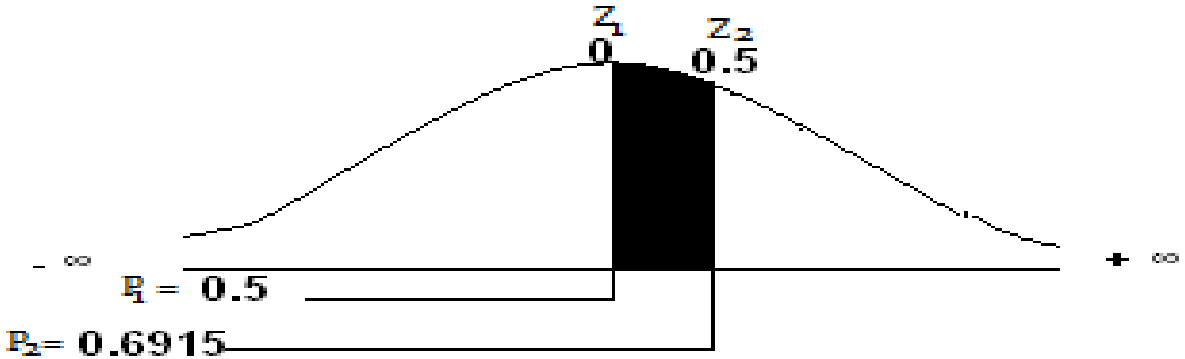
For probability of population proportion of SBP lying between 120 and 125 mmHg;

z = 120-125/10 = 0.5

then from Z- table; we take the probability positioned for z value of 0.5 which is 0.6915.

Z-table: contains numbers (probabilities) that represent AUC between specific intervals keeping in mind that the whole AUC= 1.

Z score is also called Critical Ratio.

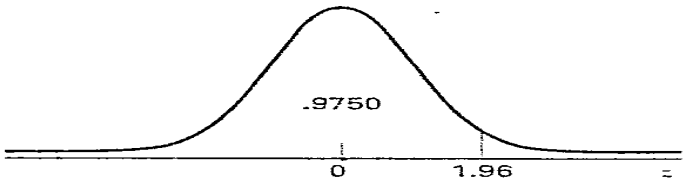


The probability for the area limited between 0 (for 120, where z table opposes z value of 0 with the probability of 0.5) and 0.5 (for 125, where z table opposes z value of 0.5 with the probability of 0.6915)

= 0.6915 – 0.5 = 0.1915

19.15% = proportion of people having SBP between 120 &125.

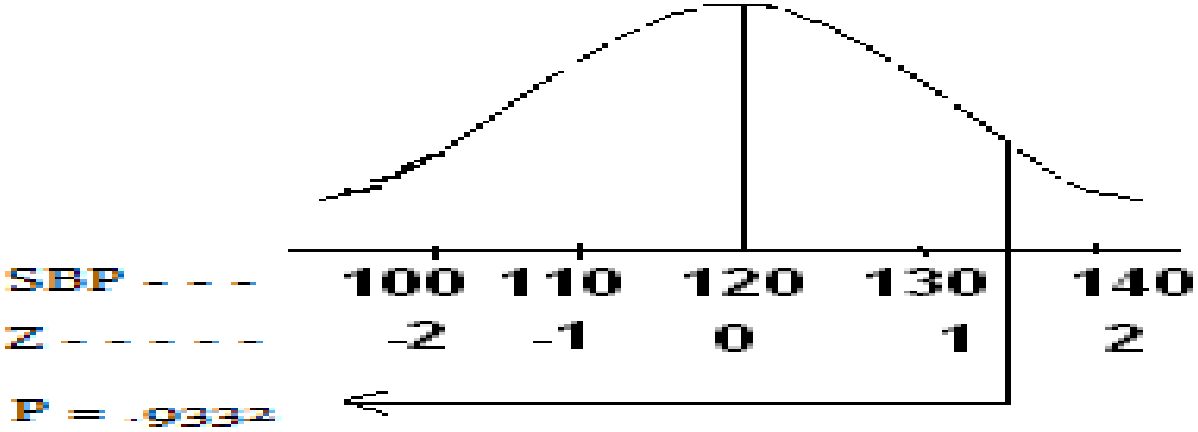
TABLE : Normal Curve Areas $P(z \leq z_0)$. Entries in the Body of the Table Are Areas Between $-\infty$ and z



z	−0.09	−0.08	−0.07	−0.06	−0.05	−0.04	−0.03	−0.02	−0.01	0.00	z
−3.80	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	−3.80
−3.70	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	−3.70
−3.60	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0002	.0002	−3.60
−3.50	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	−3.50
−3.40	.0002	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	−3.40
−3.30	.0003	.0004	.0004	.0004	.0004	.0004	.0004	.0005	.0005	.0005	−3.30
−3.20	.0005	.0005	.0005	.0006	.0006	.0006	.0006	.0006	.0007	.0007	−3.20
−3.10	.0007	.0007	.0008	.0008	.0008	.0008	.0009	.0009	.0009	.0010	−3.10
−3.00	.0010	.0010	.0011	.0011	.0011	.0012	.0012	.0013	.0013	.0013	−3.00
−2.90	.0014	.0014	.0015	.0015	.0016	.0016	.0017	.0018	.0018	.0019	−2.90
−2.80	.0019	.0020	.0021	.0021	.0022	.0023	.0023	.0024	.0025	.0026	−2.80
−2.70	.0026	.0027	.0028	.0029	.0030	.0031	.0032	.0033	.0034	.0035	−2.70
−2.60	.0036	.0037	.0038	.0039	.0040	.0041	.0043	.0044	.0045	.0047	−2.60
−2.50	.0048	.0049	.0051	.0052	.0054	.0055	.0057	.0059	.0060	.0062	−2.50
−2.40	.0064	.0066	.0068	.0069	.0071	.0073	.0075	.0078	.0080	.0082	−2.40
−2.30	.0084	.0087	.0089	.0091	.0094	.0096	.0099	.0102	.0104	.0107	−2.30
−2.20	.0110	.0113	.0116	.0119	.0122	.0125	.0129	.0132	.0136	.0139	−2.20
−2.10	.0143	.0146	.0150	.0154	.0158	.0162	.0166	.0170	.0174	.0179	−2.10
−2.00	.0183	.0188	.0192	.0197	.0202	.0207	.0212	.0217	.0222	.0228	−2.00
−1.90	.0233	.0239	.0244	.0250	.0256	.0262	.0268	.0274	.0281	.0287	−1.90
−1.80	.0294	.0301	.0307	.0314	.0322	.0329	.0336	.0344	.0351	.0359	−1.80
−1.70	.0367	.0375	.0384	.0392	.0401	.0409	.0418	.0427	.0436	.0446	−1.70
−1.60	.0455	.0465	.0475	.0485	.0495	.0505	.0516	.0526	.0537	.0548	−1.60
−1.50	.0559	.0571	.0582	.0594	.0606	.0618	.0630	.0643	.0655	.0668	−1.50
−1.40	.0681	.0694	.0708	.0721	.0735	.0749	.0764	.0778	.0793	.0808	−1.40
−1.30	.0823	.0838	.0853	.0869	.0885	.0901	.0918	.0934	.0951	.0968	−1.30
−1.20	.0985	.1003	.1020	.1038	.1056	.1075	.1093	.1112	.1131	.1151	−1.20
−1.10	.1170	.1190	.1210	.1230	.1251	.1271	.1292	.1314	.1335	.1357	−1.10
−1.00	.1379	.1401	.1423	.1446	.1469	.1492	.1515	.1539	.1562	.1587	−1.00
−0.90	.1611	.1635	.1660	.1685	.1711	.1736	.1762	.1788	.1814	.1841	−0.90
−0.80	.1867	.1894	.1922	.1949	.1977	.2005	.2033	.2061	.2090	.2119	−0.80
−0.70	.2148	.2177	.2206	.2236	.2266	.2296	.2327	.2358	.2389	.2420	−0.70
−0.60	.2451	.2483	.2514	.2546	.2578	.2611	.2643	.2676	.2709	.2743	−0.60
−0.50	.2776	.2810	.2843	.2877	.2912	.2946	.2981	.3015	.3050	.3085	−0.50
−0.40	.3121	.3156	.3192	.3228	.3264	.3300	.3336	.3372	.3409	.3446	−0.40
−0.30	.3483	.3520	.3557	.3594	.3632	.3669	.3707	.3745	.3783	.3821	−0.30
−0.20	.3859	.3897	.3936	.3974	.4013	.4052	.4090	.4129	.4168	.4207	−0.20
−0.10	.4247	.4286	.4325	.4364	.4404	.4443	.4483	.4522	.4562	.4602	−0.10
0.00	.4641	.4681	.4721	.4761	.4801	.4840	.4880	.4920	.4960	.5000	0.00

TABLE *(continued)*

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	z
0.00	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359	0.00
0.10	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753	0.10
0.20	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141	0.20
0.30	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517	0.30
0.40	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879	0.40
0.50	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224	0.50
0.60	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549	0.60
0.70	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852	0.70
0.80	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133	0.80
0.90	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389	0.90
1.00	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621	1.00
1.10	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830	1.10
1.20	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015	1.20
1.30	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177	1.30
1.40	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319	1.40
1.50	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441	1.50
1.60	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545	1.60
1.70	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633	1.70
1.80	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706	1.80
1.90	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767	1.90
2.00	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817	2.00
2.10	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857	2.10
2.20	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890	2.20
2.30	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916	2.30
2.40	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936	2.40
2.50	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952	2.50
2.60	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964	2.60
2.70	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974	2.70
2.80	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981	2.80
2.90	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986	2.90
3.00	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990	3.00
3.10	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993	3.10
3.20	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995	3.20
3.30	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997	3.30
3.40	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998	3.40
3.50	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	3.50
3.60	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	3.60
3.70	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	3.70
3.80	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	3.80



The probability of finding people (or proportion of people) with SBP

≥ 135 mmHg is .9332, so probability of finding a person

with SBP > 135 mmHg is 1 - 0.9332 = 0.0668 = 6.68%.