<u>Problem</u>:

Assume that among diabetics the fasting blood level of glucose is approximately normally distributed with a mean of 105mg per 100ml and an SD of 9 mg per 100 ml. What proportion of diabetics having fasting blood glucose levels between 90 and 125 mg per 100 ml ?

NORMAL DISTRIBUTION AND ITS APPLICATION

INTRODUCTION

- Statistically, a population is the set of all possible values of a variable.
- Random selection of objects of the population makes the variable a random variable (it involves chance mechanism)

Example: Let 'x' be the weight of a newly born baby.

'x' is a random variable representing the weight of the baby.
The weight of a particular baby is not known until he/she is born.

Discrete random variable:

If a random variable can only take values that are whole numbers, it is called a discrete random variable.

Example: No. of daily admissions
No. of boys in a family of 5
No. of smokers in a group of 100 persons.

Continuous random variable:

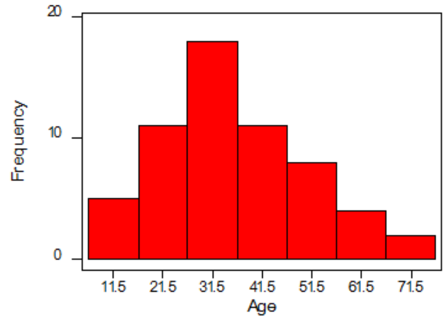
If a random variable can take any value, it is called a continuous random variable.

Example: Weight, Height, Age & BP.

The Normal or Gaussian distribution is the most important continuous probability distribution in statistics.

The word 'normal' here does not mean 'ordinary' or 'common' nor does it mean 'disease-free'.
It simply means that the distribution conforms to a certain formula and shape.

Normal Disribution and its application



- A Histogram
  - Values on the x-axis (horizontal)
  - Numbers on the y-axis (vertical)

- Normal distribution is defined by a particular shape
  - Symmetrical
  - Bell-shaped

<u>Gaussian Distribution:</u>

Many biologic variables follow this pattern

- Hemoglobin, Cholesterol, Serum Electrolytes, Blood pressures, age, weight, height

One can use this information to define what is normal and what is extreme

In clinical medicine 95% or 2 Standard deviations around the mean is normal

- Clinically, 5% of "normal" individuals are labeled as extreme/abnormal
  - We just accept this and move on.

Table 9.3 Example of a Normal Distribution—Distribution of 1000 Men in a Village According to Their Height

Height inches	No. of men of given height			
61–62	2			
62–63	5			
63–64	17			
64–65	43			
65–66	86			
66–67	152			
67–68	193	690	955	997
68–69	197			
69–70	148			
70–71	91			
71–72	45			
72–73	16			
73–74	4			
74–75	1			
Total	1000			

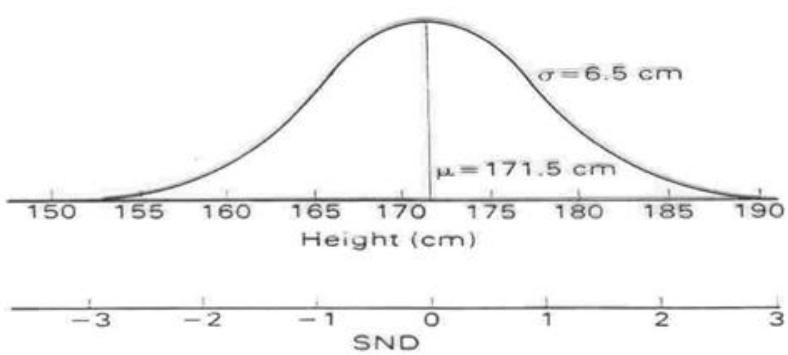
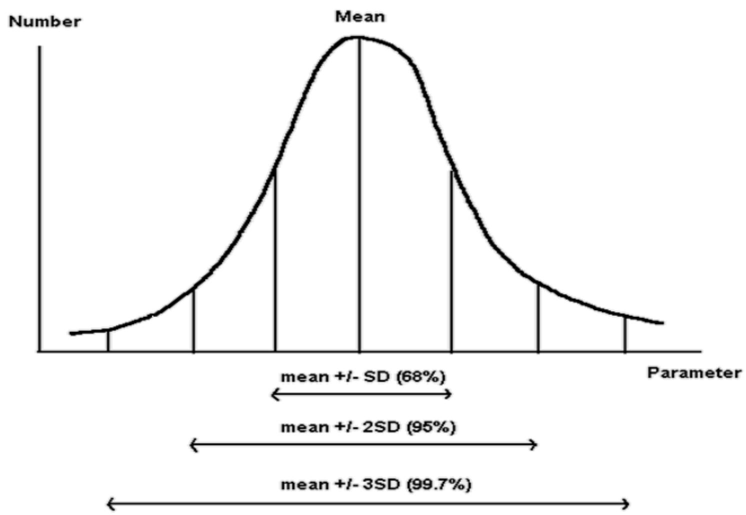


Fig. 5.2 Relationship between normal distribution in original units of measurement and in standard normal deviates. SND = (height − 171.5)/6.5. Height = 171.5 + (6.5 × SND).



Characteristics of Normal Distribution

- Symmetrical about mean, μ
- Mean, median, and mode are equal
- Total area under the curve above the x-axis is one square unit
- 1 standard deviation on both sides of the mean includes approximately 68% of the total area
- 2 standard deviations includes approximately 95%
- 3 standard deviations includes approximately 99%

- Normal distribution is completely determined by the parameters μ and σ
  - Different values of μ shift the distribution along the x-axis
  - Different values of σ determine degree of flatness or peakedness of the graph

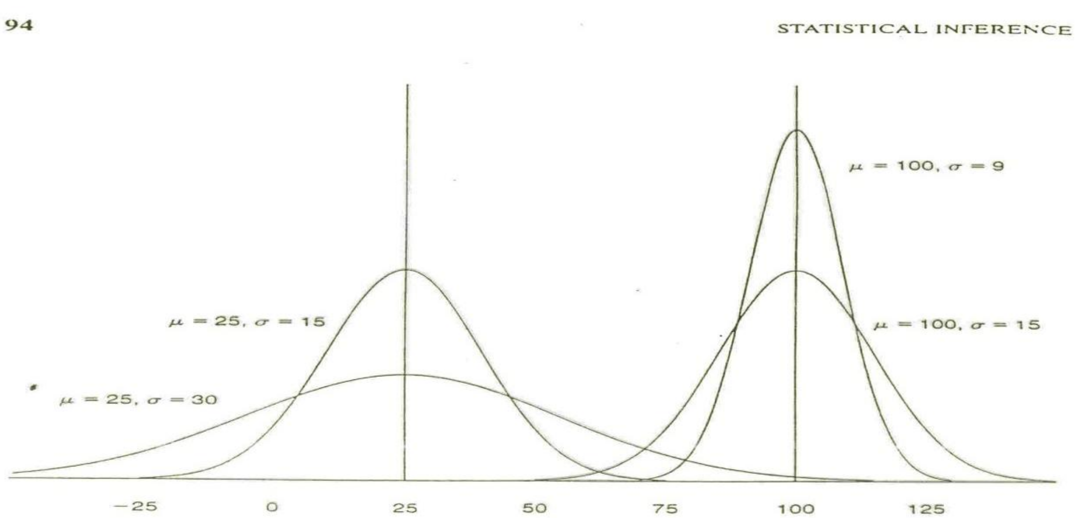


Figure 4.8. Examples of normal distributions.

Uses of Normal Distribution

- It's application goes beyond describing distributions
- It is used by researchers.
- The major use of normal distribution is the role it plays in statistical inference.
- The z score along with the t –score, chi-square and F-statistics is important in hypothesis testing.
- It helps managers to make decisions.

What's so Great about the Normal Distribution?

If you know two things, you know everything about the distribution
- Mean
- Standard deviation

You know the probability of any value arising

Standardised Scores

My diastolic blood pressure is 100
- So what ?

Normal is 90 (for my age and sex)
- Mine is high
  - But how much high?

Express it in standardised scores
- How many SDs above the mean is that?

Mean = 90, SD = 4 (my age and sex)

$$\frac{\text{My Score - Mean Score}}{\text{SD}} = \frac{100\text{-}90}{4} = 2.5$$

This is a *standardised score*, or *z-score*

Can consult tables (or computer)
- See how often this high (or higher) score occur

Measure Of position

❖ *z* Score (or standard score)

The number of standard deviations that a given value x is above or below the mean

z score

Sample

$$z = \frac{x - \bar{x}}{s}$$

Population

$$z = \frac{x - \mu}{\sigma}$$

Round to 2 decimal places

<u>Standard Scores:</u>

The Z score makes it possible, under some circumstances, to compare scores that originally had different units of measurement.

<u>Z Score:</u>

Suppose you scored a 60 on a numerical test and a 30 on a verbal test. On which test did you perform better?

- First, we need to know how other people did on the same tests.
  - Suppose that the mean score on the numerical test was 50 and the mean score on the verbal test was 20.
  - You scored 10 points above the mean on each test.
  - Can you conclude that you did equally well on both tests?
  - You do not know, because you do not know if 10 points on the numerical test is the same as 10 points on the verbal test.

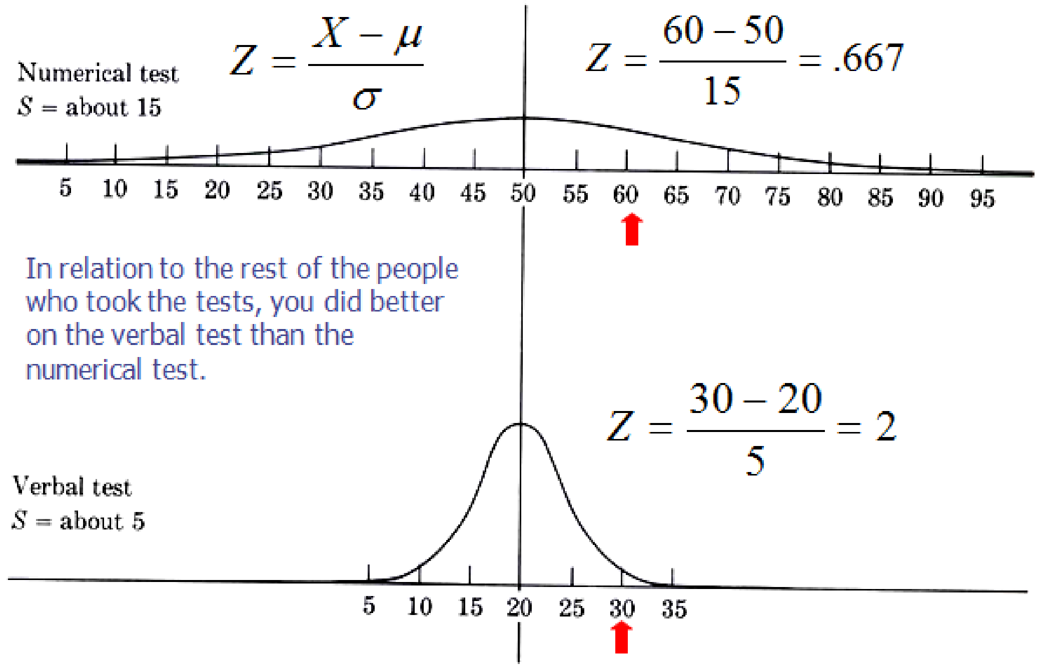
- Suppose also that the standard deviation on the numerical test was 15 and the standard deviation on the verbal test was 5.
  - Now can you determine on which test you did better?

To find out how many standard deviations away from the mean a particular score is, use the Z formula:

Population: Sample:

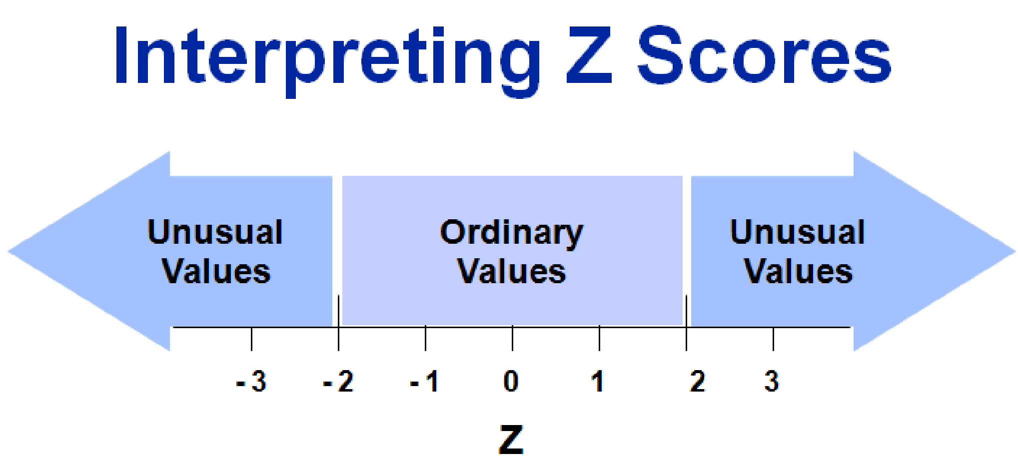
$$z = \frac{X - \mu}{\sigma} \qquad z = \frac{X - \bar{X}}{s}$$

Z Score



Z score

- Allows you to describe a particular score in terms of where it fits into the overall group of scores.
    - Whether it is above or below the average and how much it is above or below the average.
- A standard score that states the position of a score in relation to the mean of the distribution, using the standard deviation as the unit of measurement.
    - The number of standard deviations a score is above or below a mean.

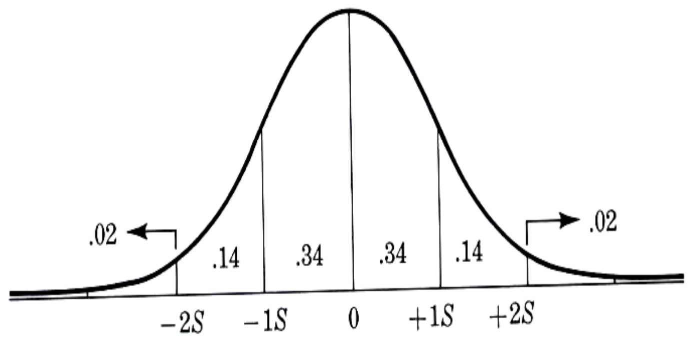


Properties of Z scores

The standard deviation of any distribution expressed in Z scores is always one.

- In calculating Z scores, the standard deviation of the raw scores is the unit of measurement.

The Standard Normal Curve



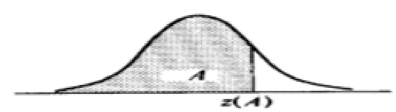
The Standard Normal Table

Using the standard normal table, you can find the area under the curve that corresponds with certain scores.

The area under the curve is proportional to the frequency of scores.

The area under the curve gives the probability of that score occurring.

Entry is area A under the standard normal curve from $-\infty$ to $z(A)$

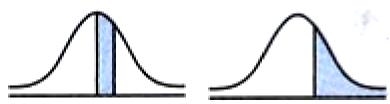


z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Reading the Z Table

- Finding the proportion of observations <u>between</u> the mean and a score when
    - Z = 1.80

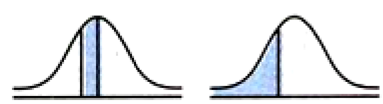
A	B	C
z		



1.76	.4608	.0392
1.77	.4616	.0384
1.78	.4625	.0375
1.79	.4633	.0367
$\boxed{1.80}$	$\boxed{.4641}$	.0359
1.81	.4649	.0351
1.82	.4656	.0344
1.83	.4664	.0336
1.84	.4671	.0329
1.85	.4678	.0322

- Finding the proportion of observations <u>between</u> a score and the mean when
    - Z = -2.10

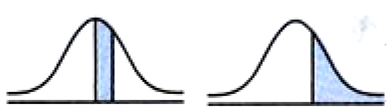
2.07	.4808	.0192
2.08	.4812	.0188
2.09	.4817	.0183
$\boxed{2.10}$	$\boxed{.4821}$	.0179
2.11	.4826	.0174
2.12	.4830	.0170
2.13	.4834	.0166



−z		
A′	B′	C′

- Finding the proportion of observations <u>above</u> a score when
    - Z = 1.80

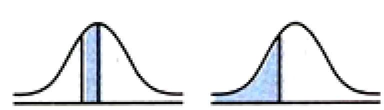
A	B	C
z		



1.77	.4616	.0384
1.78	.4625	.0375
1.79	.4633	.0367
$\boxed{1.80}$	.4641	$\boxed{.0359}$
1.81	.4649	.0351
1.82	.4656	.0344
1.83	.4664	.0336
1.84	.4671	.0329

- Finding the proportion of observations <u>below</u> a score when
    - Z = -2.10

2.07	.4808	.0192
2.08	.4812	.0188
2.09	.4817	.0183
$\boxed{2.10}$	.4821	$\boxed{.0179}$
2.11	.4826	.0174
2.12	.4830	.0170
2.13	.4834	.0166



−z		
A′	B′	C′

Z scores and the Normal Distribution

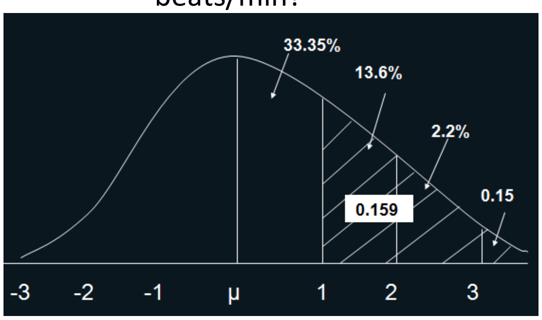
- Can answer a wide variety of questions about any normal distribution with a known mean and standard deviation.
- Will address how to solve two main types of normal curve problems:
    - Finding a proportion given a score.
    - Finding a score given a proportion.

Exercises

Assuming the normal heart rate (H.R) in normal healthy individuals is normally distributed with Mean = 70 and Standard Deviation =10 beats/min

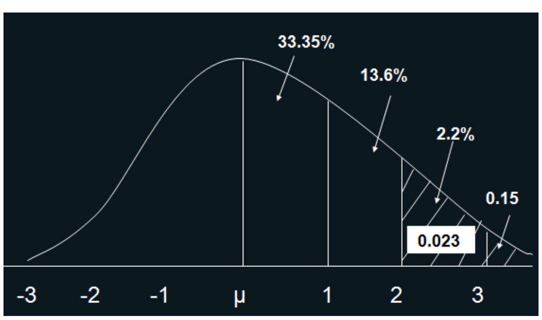
Exercise # 1

1) What area under the curve is above 80 beats/min?



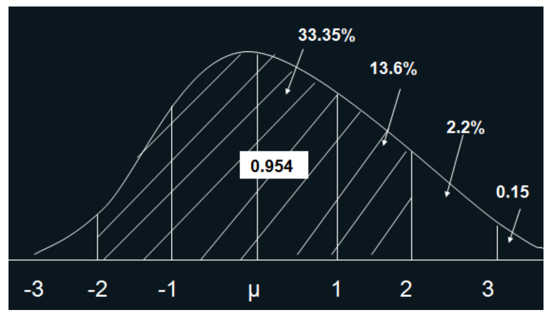
Exercise # 2

2) What area of the curve is above 90 beats/min?



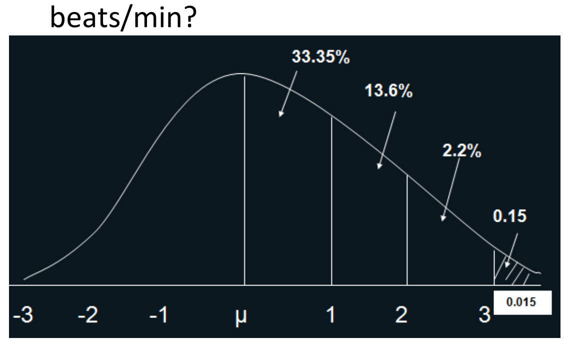
Exercise # 3

3) What area of the curve is between 50-90 beats/min?



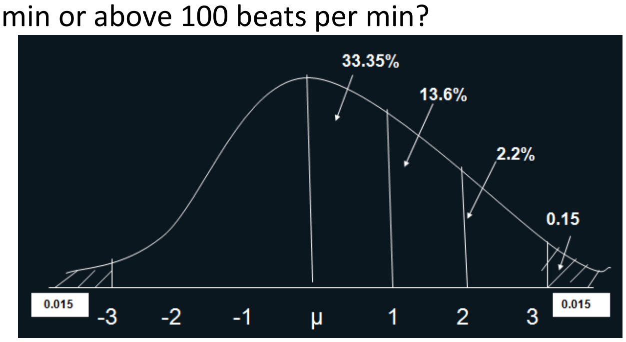
Exercise # 4

4) What area of the curve is above 100 beats/min?



Exercise # 5

5) What area of the curve is below 40 beats per min or above 100 beats per min?



Exercise:

Assuming the normal heart rate (H.R) in normal healthy individuals is normally distributed with Mean = 70 and Standard Deviation =10 beats/min

Then:

1) What area under the curve is above 80 beats/min?
   Ans: 0.159 (15.9%)

2) What area of the curve is above 90 beats/min?
   Ans: 0.023 (2.3%)

3) What area of the curve is between 50-90 beats/min?
   Ans: 0.954 (95.4%)

4) What area of the curve is above 100 beats/min?
   Ans: 0.0015 (0.15%)

5) What area of the curve is below 40 beats per min or above 100 beats per min?
   Ans: 0.0015 for each tail or 0.3%

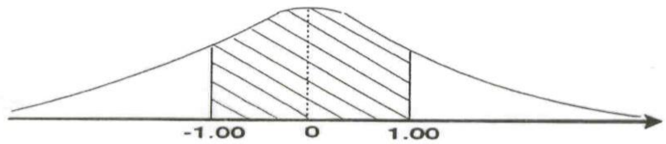
Problem:

Assume that among diabetics the fasting blood level of glucose is approximately normally distributed with a mean of 105mg per 100ml and an SD of 9 mg per 100 ml. What proportion of diabetics having fasting blood glucose levels between 90 and 125 mg per 100 ml ?

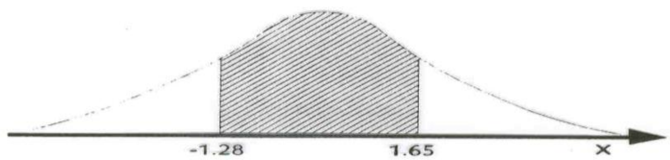
Answer \ next page

iii) Since the total area is unity,

$$
\begin{aligned}
P(-1.00 \le Z \le 1.00) &= 1 - \{P(Z < -1.00) + P(Z > 1.00)\} \\
&= 1 - (2 \times 0.1587) \\
&= 0.6826
\end{aligned}
$$



iv)



$$
\begin{aligned}
P(-1.28 \le Z \le 1.65) &= P(Z \le 1.65) - P(Z < -1.28) \\
&= 0.9505 - 0.1003 \\
&= 0.8502
\end{aligned}
$$

Note:
Since $P(Z=z) = 0$ for a particular value z, we can use $P(Z<z)$ or $P(Z \le z)$ as they are equal.

Example 2
Assume that among diabetics the fasting blood level of glucose is approximately normally distributed with a mean of 105 mg per 100 ml and an *SD* of 9 mg per 100 ml.

i) What proportion of diabetics have fasting blood glucose levels between 90 and 125 mg per 100 ml?

ii) What level cuts off the lower 10 per cent of diabetics?

iii) What levels encompass the middle 95 per cent of diabetics?

Answers Example 2
Let X be the random variable denoting the fasting blood glucose level. X has a normal distribution with mean = 105 and standard deviation = 9.

i) We have to compute $P(90 \le X \le 125)$. The table is available only for the probabilities of a standard normal distribution. Thus we have to convert X to a standard normal variable (Z), using the formula on page 5 of this module.

We require $P\ (90 \le X \le 125)$.

This can be written as

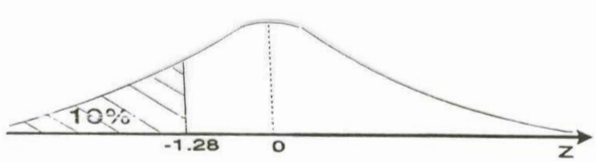
$$
P\left[\frac{90-105}{9} \le \frac{X-105}{9} \le \frac{125-105}{9}\right] = P(-1.67 \le Z \le 2.22)
$$

$$
\text{since } Z = \frac{X-105}{9}
$$

$$
\begin{aligned}
&= P(Z \le 2.22) - P(Z < -1.67) \\
&= 0.9868 - 0.0475 \\
&= 0.9393
\end{aligned}
$$

Therefore 94% of diabetics have fasting blood glucose levels between 90 and 125.

ii)



From the table we know that -1.28 cuts off the lower 10 per cent of the standard normal curve. Now we have to find the corresponding X-value.

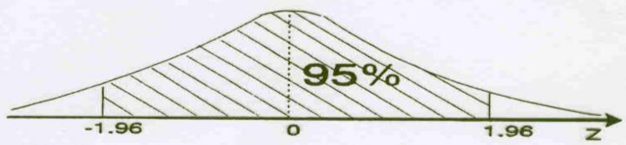
$$-1.28 = \frac{X - 105}{9}$$

$$X = -1.28 \times 9 + 105$$

$$= 93.5 \text{ mg/100 ml}$$

93.5 mg/100 ml cuts off the lower 10 per cent of diabetics.

iii)



We know from the tables that $P(-1.96 \leq Z \leq 1.96) = 0.95$.

Corresponding $X-$values are

$$-1.96 \times 9 + 105 = 87.4$$

and $\quad 1.96 \times 9 + 105 = 122.6$ mg per 100ml

Note:

A one-sided p percentage point of the standard normal distribution is the value z such that

$$P(Z \geq z) = p/100$$

and a two-sided p percentage point z is such that

$$P(Z \geq z) = P(Z \leq -z)$$

$$= (^{p}/_{2})/100$$

or equivalently

$$P(Z \geq z) + P(Z \leq -z) = p/100$$