Description of Data II
(Normal distribution and its application)

Objectives:

- Able to understand the concept of Normal distribution.
- Able to calculate the z-score for quantitative variable.
- Able to apply the concept in the interpretation of a clinical data..

Team Members: Munirah Aldofyan – Njoud Alenezy

Team Leaders: Mohammed ALYousef & Rawan Alwadee

Revised By: Basel almeflh

Dr. Shaffi Ahmed

Resources:

- 436 Lecture Slides + Notes

Important – Notes

436researchteam@gmail.com

Editing file

Feedback link

- **Normal distribution and its application**
  - **Gaussian Distribution:** the most important continuous probability distribution in statistics.
  - **Characteristics of Normal Distribution**
    - Symmetrical about mean, μ
    - 3 standard deviations includes approximately 99%
    - 2 standard deviations includes approximately 95%
    - Total area under the curve above the x-axis is one square unit
    - Mean, median, and mode are equal
    - 1 standard deviation on both sides of the mean includes approximately 68% of the total area
  - **Uses of Normal Distribution:**
    - It's application goes beyond describing distributions
    - It is used by researchers.
    - It helps managers to make decisions.
    - The z score is important in hypothesis testing.
    - The major use of normal distribution is the role it plays in statistical inference.

If you know two things,

1-Mean

2-Standard deviation

•you know everything about the distribution

•You know the probability of any value arising

✓ **Standardized Scores:**

$$\frac{my\ score\ - mean\ score}{SD}$$

✓ **Measures of Position:**

- *z* Score (or standard score)

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

436 Research Team

- The Normal or Gaussian distribution is the most important continuous probability distribution in statistics.
- The term “Gaussian” refers to ‘Carl Freidrich Gauss’ who develop this distribution.
- The word ‘normal’ here does not mean ‘ordinary’ or ‘common’ nor does it mean ‘disease-free’.
- It simply means that the distribution confirms to a certain formula and shape.

Gaussian Distribution

- 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			

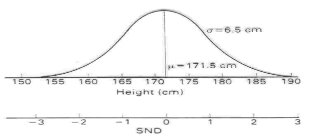
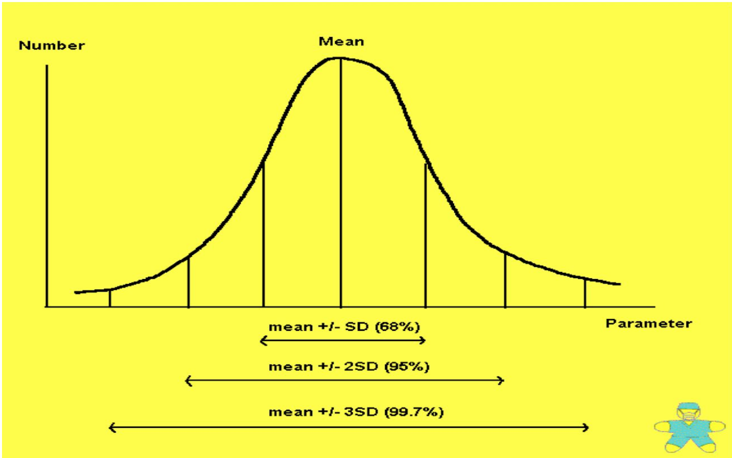


Fig. 5.2 Relationship between normal distribution in original units of measurement and in standard normal deviates. SND = (height − 171.5)/6.5. Height = 171.5 + (6.5 × SND).



Characteristics of Normal Distribution

- Symmetrical about mean, μ (mean will be in the center)
- 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%

For example, 68% of the distribution is within one standard deviation of the mean and approximately 95% of the distribution is within two standard deviations of the mean. Therefore, if you had a normal distribution with a mean of 50 and a standard deviation of 10, then 68% of the distribution would be between 50 - 10 = 40 and 50 +10 =60. Similarly, about 95% of the distribution would be between 50 - 2 x 10 = 30 and 50 + 2 x 10 = 70. The symbol for the population standard deviation is σ; the symbol for an estimate computed in a sample is s.
this example is taken from here

Uses of Normal Distribution

- It's application goes beyond describing distribution
- It is used by researchers.
- The major use of normal distribution is the role it plays in statistical inference.

Most of the statically theory based on this concept

- The z score is important in hypothesis testing.
- It helps managers to make decisions.

What's so Great about the Normal Distribution?

- If you know two things,
  - Mean
  - Standard deviation

you know everything about the distribution

You know the probability of any value arising

Standardized Scores: describe nature of data how it distributed

- 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 standardized scores
  - how much SDs above the mean is that?

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

$$\frac{my\ score\ - mean\ score}{SD} = \frac{100 - 90}{4} = 2.5$$

- This is a standardised score, or z-score
- Look z tables (or computer)
  - See how often this high (or higher) score occur

Measures of Position

- *z* Score (or standard score)

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

meaning that my BP is 100 and the mean is 90, the SD is 4, how many SDs do i need to reach my BP from the mean? 10 is the difference, 10 divided by 4 = 2.5 , so i need 2.5 times

//(90+4)=1 SD,, (90+8)= 2 SD,, (90+10)=2.5 SD

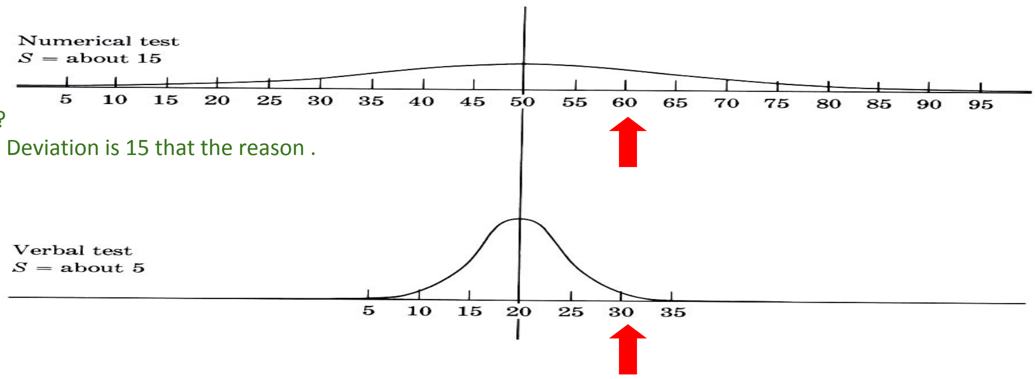
Standard Scores

The Z score makes it possible, under some circumstances, to compare scores that originally had different units of measurement. Means weight will be measure in kilo height in cm will be another units but having different variable in data but using the standard score I mean we can convert into z score... this is the advantage

Z Score

- 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?

Verbal is better because you almost 2 standard deviations.



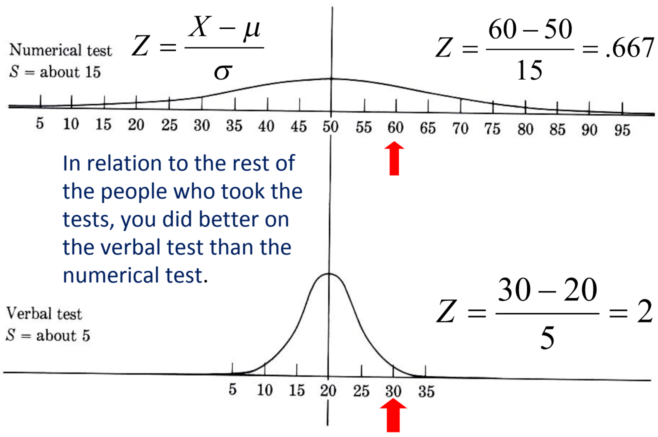
Why flat?
Standard Deviation is 15 that the reason .

Z score

To find out how many standard deviations away from the mean a particular score is, use the Z formula:
important to remember that the standard deviation is the unit of measurement .

Population: $$Z = \frac{X - \mu}{\sigma}$$

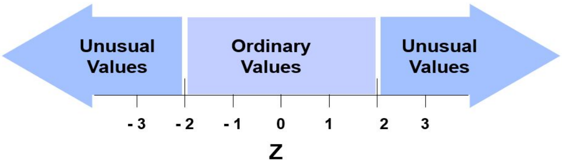
Sample: $$Z = \frac{X - \overline{X}}{S}$$



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.

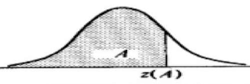
Interpreting Z Scores



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

Standard Normal Table

Table A[a] (Continued)
PROPORTIONS OF AREA UNDER STANDARD NORMAL CURVE FOR VALUES OF z

A z	B	C	A z	B	C	A z	B	C
1.68	.4535	.0465	2.24	.4875	.0125	2.80	.4974	.0026
1.69	.4545	.0455	2.25	.4878	.0122	2.81	.4975	.0025
1.70	.4554	.0446	2.26	.4881	.0119	2.82	.4976	.0024
1.71	.4564	.0436	2.27	.4884	.0116	2.83	.4977	.0023
1.72	.4573	.0427	2.28	.4887	.0113	2.84	.4977	.0023
1.73	.4582	.0418	2.29	.4890	.0110	2.85	.4978	.0022
1.74	.4591	.0409	2.30	.4893	.0107	2.86	.4979	.0021
1.75	.4599	.0401	2.31	.4896	.0104	2.87	.4979	.0021
1.76	.4608	.0392	2.32	.4898	.0102	2.88	.4980	.0020
1.77	.4616	.0384	2.33	.4901	.0099	2.89	.4981	.0019
1.78	.4625	.0375	2.34	.4904	.0096	2.90	.4981	.0019
1.79	.4633	.0367	2.35	.4906	.0094	2.91	.4982	.0018
1.80	.4641	.0359	2.36	.4909	.0091	2.92	.4982	.0018
1.81	.4649	.0351	2.37	.4911	.0089	2.93	.4983	.0017
1.82	.4656	.0344	2.38	.4913	.0087	2.94	.4984	.0016
1.83	.4664	.0336	2.39	.4916	.0084	2.95	.4984	.0016
1.84	.4671	.0329	2.40	.4918	.0082	2.96	.4985	.0015
1.85	.4678	.0322	2.41	.4920	.0080	2.97	.4985	.0015
1.86	.4686	.0314	2.42	.4922	.0078	2.98	.4986	.0014
1.87	.4693	.0307	2.43	.4925	.0075	2.99	.4986	.0014
1.88	.4699	.0301	2.44	.4727	.0073	3.00	.4987	.0013
1.89	.4706	.0294	2.45	.4929	.0071	3.01	.4987	.0013
1.90	.4713	.0287	2.46	.4931	.0069	3.02	.4987	.0013
1.91	.4719	.0281	2.47	.4932	.0068	3.03	.4988	.0012
1.92	.4726	.0274	2.48	.4934	.0066	3.04	.4988	.0012
1.93	.4732	.0268	2.49	.4936	.0064	3.05	.4989	.0011
1.94	.4738	.0262	2.50	.4938	.0062	3.06	.4989	.0011
1.95	.4744	.0256	2.51	.4940	.0060	3.07	.4989	.0011
1.96	.4750	.0250	2.52	.4941	.0059	3.08	.4990	.0010
1.97	.4756	.0244	2.53	.4943	.0057	3.09	.4990	.0010
1.98	.4761	.0239	2.54	.4945	.0055	3.10	.4990	.0010
1.99	.4767	.0233	2.55	.4946	.0054	3.11	.4991	.0009
2.00	.4772	.0228	2.56	.4948	.0052	3.12	.4991	.0009
2.01	.4778	.0222	2.57	.4949	.0051	3.13	.4991	.0009
2.02	.4783	.0217	2.58	.4951	.0049	3.14	.4992	.0008
2.03	.4788	.0212	2.59	.4952	.0048	3.15	.4992	.0008
2.04	.4793	.0207	2.60	.4953	.0047	3.16	.4992	.0008
2.05	4798	.0202	2.61	.4955	.0045	3.17	.4992	.0008
2.06	.4803	.0197	2.62	.4956	.0044	3.18	.4993	.0007
2.07	.4808	.0192	2.63	.4957	.0043	3.19	.4993	.0007
2.08	.4812	.0188	2.64	.4959	.0041	3.20	.4993	.0007
2.09	.4817	.0183	2.65	.4960	.0040	3.21	.4993	.0007
2.10	.4821	.0179	2.66	.4961	.0039	3.22	.4994	.0006
2.11	.4826	.0174	2.67	.4962	.0038	3.23	.4994	.0006
2.12	.4830	.0170	2.68	.4963	.0037	3.24	.4994	.0006
2.13	.4834	.0166	2.69	.4964	.0036	3.25	.4994	.0006
2.14	.4838	.0162	2.70	.4965	.0035	3.30	.4995	.0005
2.15	.4842	.0158	2.71	.4966	.0034	3.35	.4996	.0004
2.16	.4846	.0154	2.72	.4967	.0033	3.40	.4997	.0003
2.17	.4850	.0150	2.73	.4968	.0032	3.45	.4997	.0003
2.18	.4854	.0146	2.74	.4969	.0031	3.50	.4998	.0002
2.19	.4857	.0143	2.75	.4970	.0030	3.60	.4998	.0002
2.20	.4861	.0139	2.76	.4971	.0029	3.70	.4999	.0001
2.21	.4864	.0136	2.77	.4972	.0028	3.80	.4999	.0001
2.22	.4868	.0132	2.78	.4973	.0027	3.90	.49995	.00005
2.23	.4871	.0129	2.79	.4974	.0026	4.00	.49997	.00003
−z A′	B′	C′	−z A′	B′	C′	−z A′	B′	C′

Reading the Z Table

Finding the proportion of observations between the mean and a score when

Z = 1.80

Reading the Z Table

Finding the proportion of observations above a score when

Z = 1.80

A z	B	C
1.68	.4535	.0465
1.69	.4545	.0455
1.70	.4554	.0446
1.71	.4564	.0436
1.72	.4573	.0427
1.73	.4582	.0418
1.74	.4591	.0409
1.75	.4599	.0401
1.76	.4608	.0392
1.77	.4616	.0384
1.78	.4625	.0375
1.79	.4633	.0367
1.80	.4641	.0359
1.81	.4649	.0351
1.82	.4656	.0344
1.83	.4664	.0336
1.84	.4671	.0329
1.85	.4678	.0322
1.86	.4686	.0314
1.87	.4693	.0307
1.88	.4699	.0301
1.89	.4706	.0294
1.90	.4713	.0287
1.91	.4719	.0281
1.92	.4726	.0274

Reading the Z Table

Finding the proportion of observations between a score and the mean when

Z = -2.10

Reading the Z Table

Finding the proportion of observations below a score when

Z = -2.10

A′ −z	B′	C′
1.98	.4761	.0239
1.99	.4767	.0233
2.00	.4772	.0228
2.01	.4778	.0222
2.02	.4783	.0217
2.03	.4788	.0212
2.04	.4793	.0207
2.05	4798	.0202
2.06	.4803	.0197
2.07	.4808	.0192
2.08	.4812	.0188
2.09	.4817	.0183
2.10	.4821	.0179
2.11	.4826	.0174
2.12	.4830	.0170
2.13	.4834	.0166
2.14	.4838	.0162
2.15	.4842	.0158
2.16	.4846	.0154
2.17	.4850	.0150
2.18	.4854	.0146
2.19	.4857	.0143
2.20	.4861	.0139
2.21	.4864	.0136
2.22	.4868	.0132
2.23	.4871	.0129

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

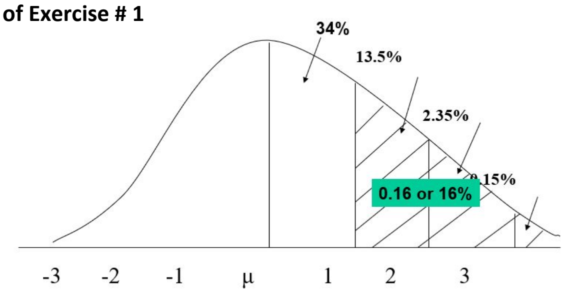
Then:

1) What area under the curve is above 80 beats/min? how much percentage of patients are their heart rate is 80 b\m and above?

Total are 100%.//1 standard deviation cover 68%, half of 68%(we did this step because the SD covers 10 above and below 70 and in this case we only want above so we take the half)

,34% Subtract from 50 = 16%(we did this step because 34% are above the mean (70) with a range of 10(70+10=80) but what about those above it?(and our case we only want above 80) we subtract from 50 because this 50% represents those above the mean and 34% are within (70--80) and we want above 80 which leaves us with the other part of the 50% (50-34=16%))..who have a beat above 80 beats/min.

Diagram of Exercise # 1



Exercise # 2

Then:

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

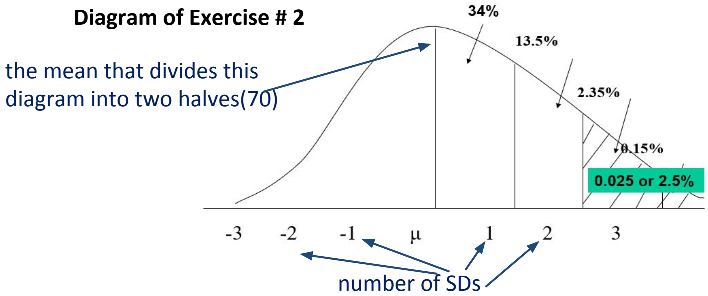
=2.5% Because we are asking 2 standard deviations.

2 standard deviations (90)

mean is 70

Probability .025 or 2.5%

Diagram of Exercise # 2



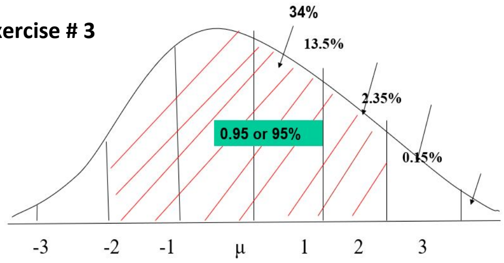
Exercise # 3

Then:

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

0.95 or 95% because of 2SD.

Diagram of Exercise # 3



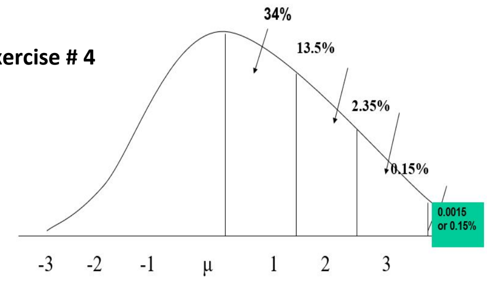
Exercise # 4

Then:

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

0.0015.(3 standard deviations). so small area in the extreme right side.

Diagram of Exercise # 4

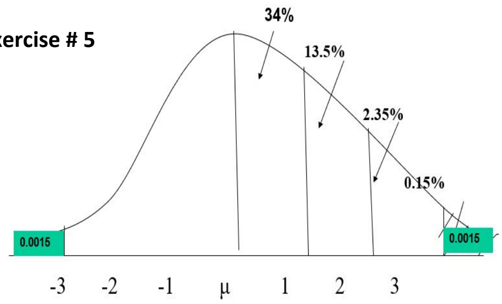


Exercise # 5

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

Extreme 3 standard deviations - extreme +3 standard deviations on upper side =because it crosses 3 standard Deviations

Diagram of Exercise # 5



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: 0.16 (16%)
2) What area of the curve is above 90 beats/min?
   Ans: 0.025 (2.5%)
3) What area of the curve is between 50-90 beats/min?
   Ans: 0.95 (95%)
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%

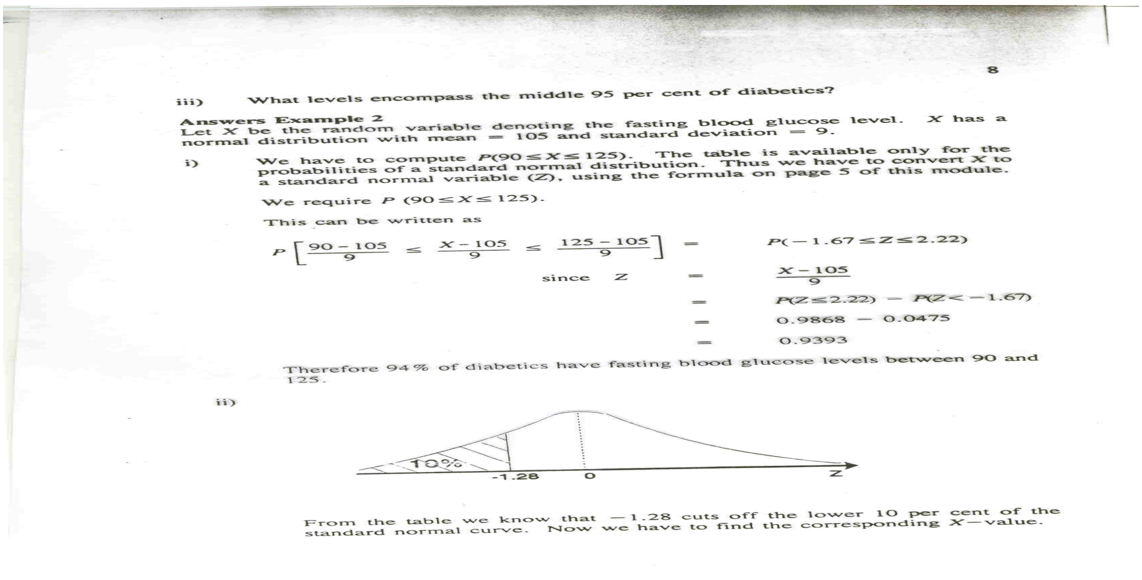
Problem:

Assume that among diabetics the fasting blood level of glucose is approximately normally distributed with a mean of 105 mg 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 ? we have to get 2 z score, it is not a whole number to apply the rule it is in fraction.

90 - 105 \ 9 = -1.6 this is one z score

125 - 105 \ 9 = 2.2 this is the second z score.

then you continue solving as in the picture below.



8

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 $P(90 \leq X \leq 125)$. 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 $P\ (90 \leq X \leq 125)$.

This can be written as

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

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

$$= P(Z \leq 2.22) - P(Z < -1.67)$$

$$= 0.9868 - 0.0475$$

$$= 0.9393$$

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.