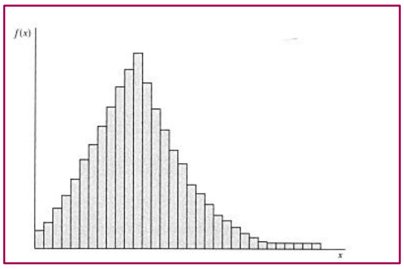
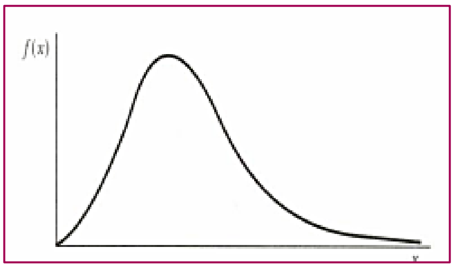
Probability Distributions

- ***Continues distribution***
  - A continuous variable is one that can assume any value within a specified interval of values.
  - Consequently, between any two values assumed by a continuous variable, there exist an infinite number of values.
  - In Normal distributions, as the number of possible outcomes (observations, n) approaches infinity, and the width of the class intervals approaches zero, the frequency polygon approaches a smooth curve.





  - The most important of continuous probability distributions is the normal distribution (sometimes referred to as ***Gaussian distribution***
  - $f(x) = \frac{1}{\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$, $(-\infty < x < \infty)$
  - The parameters of distribution are: μ: mean and σ: the standard deviation. π=3.14159, e=2.71828

- ***Characteristics of the normal distribution***

1. It is considered the most prominent probability distribution in statistics
2. Arises as the outcome of the ***central limit theorem***, which states that under mild conditions (no extremities) the sum of a large number of random variables is distributed approximately normally (concentration, weight, temperature).
3. It is symmetrical about its mean, μ (number of values > μ equals the number of values < μ).
4. The curve on either side of μ is a *mirror image* of the other side.
5. The *mean*, the *median*, and the *mode* are all equal.
6. The total *area under the curve* above the x-axis *is one square unit*. This characteristic follows from the fact that the normal distribution is a probability distribution (AUC = 1= Probability).
7. The relative frequency *(probability)* of occurrence of values between any two points (a, b) on the x-axis is equal to the *total area bounded by the curve* (gray area).

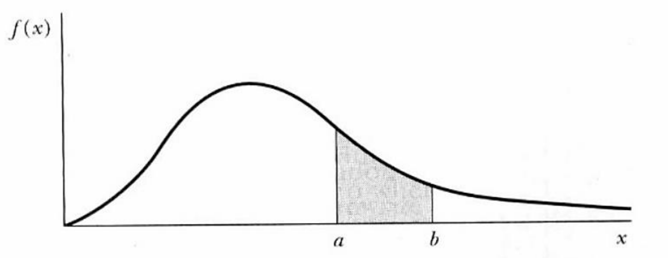
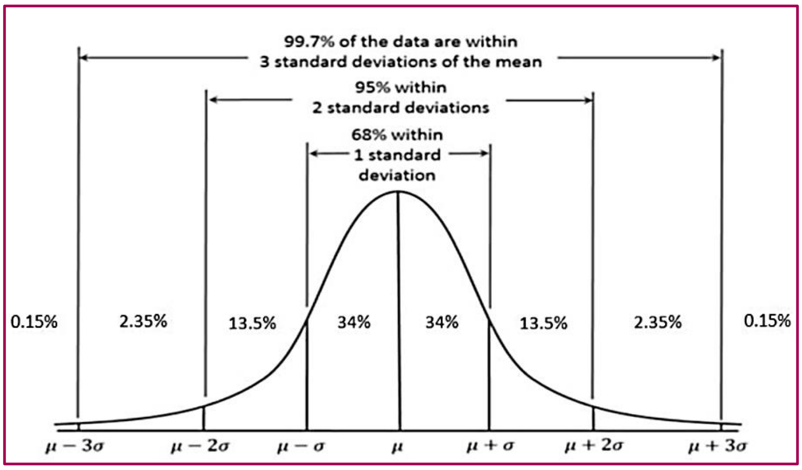


FIGURE 4.5.3 Graph of a continuous distribution showing area between *a* and *b*.

8. If we erect perpendiculars a distance of 1 *standard deviation from the mean in both directions*, the area enclosed by these perpendiculars, the x-axis, and the curve will be approximately ***68 percent*** of the total area.

9. If we extend these lateral boundaries a distance of 2 *standard deviations* on either side of the mean, approximately ***95 percent*** of the area will be enclosed.

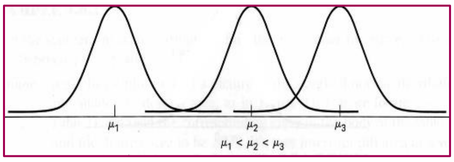
10. If we extend these lateral boundaries a distance of 3 *standard deviations* will cause approximately ***99.7 percent*** of the total area to be enclosed.



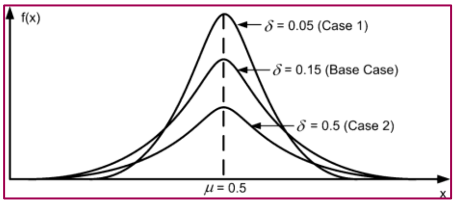
11. The normal distribution is completely determined by the parameters **μ** and **σ**.
   - ✓ Different values of μ, shift the graph of the distribution along the x-axis.
   - ✓ Different values of σ determine the degree of flatness or peakedness of the graph of the distribution (μ: location σ: shape).

★ Example:

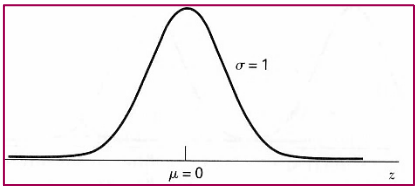
✓ 3 Normal distributions with different μ values and equal σ values.



✓ 3 Normal distributions with equal μ values and different σ values.

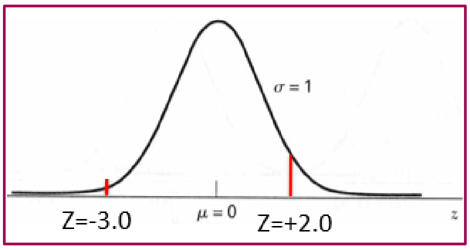


- ***The Standard Normal Distribution***
   - ➢ A standard normal distribution is a normal distribution which has a ***mean (μ)=0*** and a ***standard deviation (σ)=1***.

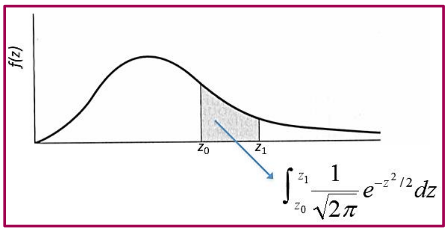


   - ➢ ***Z (standard score, z-score) = (x - μ)/σ***
   - ➢ Normal distribution $f(x) = \frac{1}{\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$
   - ➢ Is standardized to Standard Normal Distribution $f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$
   - ➢ ***What is z-score??***
     Z-scores are scores converted into the number of standard deviations that the values (x) are from the mean (μ): $z = \frac{x-\mu}{\sigma}$
   - ➢ ***Z-scoring is used to standardize scores*** in order to provide a way of comparing them (compare score of 3.0 in IELTS with a 580 in the old TOEFL exam??)

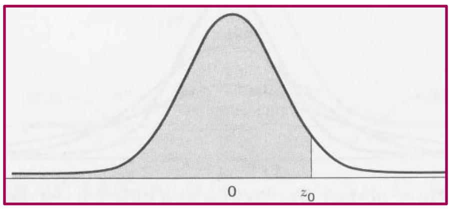
- ***Important note:*** *for z-scores: mean (μ)=0 and a standard deviation (σ)=1.*
- for any normal distribution we can get the standard score → so can use tables.
- Probability in normal distribution for any exact value = zero. P(X=1) = integration from X to X So it equals zero.
- We convert any x-score into z-score, because <u>we have only one table for standard normal distribution.</u>
  - ✓ z = 0: The value equals the mean.
  - ✓ z > 0: The value is above the mean.
  - ✓ z < 0: The value is below the mean.



- A z-score = ***+2.0*** means that the original score (*x*) was 2 standard deviations ***above*** the mean.
- A z-score = ***–3.0*** means that the original score was three standard deviations ***below*** the mean
- To find the ***probability*** that z takes on a value between any two points on the z-axis (z_0 and z_1), we must find the ***area*** bounded by the perpendiculars erected at these points, the curve and the horizontal z-axis.
- The area can be found by ***integrating*** the f(z) function between two values of the variable.
- In the standard normal distribution, the integral (area) is given by the equation



- There are *cumulative tables* that provide the results of all such integrations.
- These tables provide the *areas under the curve between* ***-∞ and z*** (cumulative relative frequency again!!).
- ***These tables are used to find the probability that a statistic is observed below, above, or between values on the standard normal distribution.***



Appendix II • Statistical Tables 689

TABLE D *(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

P(Z ≤ 1.0) = 0.8413

P(Z ≤ 1.57) = 0.9418

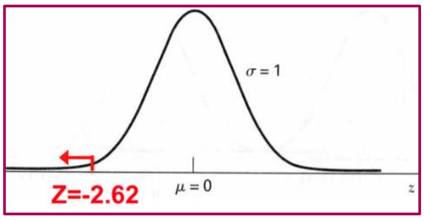
P(Z ≤ 1.96) = 0.9750

P(2.4 ≤ Z ≤ 2.5) = 0.9938-0.9918=0.002

P(Z > 3.0) =1-P(Z≤3)= 1-0.9987=0.0013

★ Example 1:

- As a part of a study of Alzheimer's disease, data reported were compatible with the hypothesis that brain *weights* of victims of the disease are normally distributed. From the reported data we may compute a *mean of* 1076.80g and a *standard deviation of* 105.76 g. Find the probability that a randomly selected patient will have a brain that weighs less than or equal to 800 grams P(x≤800).
- First, we need to calculate the z-score for x=800 g.
- $$z = \frac{800 - 1076.80}{105.76} = -2.62$$
- Now we can use the standard normal distribution table: What is the probability that randomly selected patients to have a brain weight less than or equal to z = -2.62
- P (x ≤800) = P (z ≤ -2.62) = 0.0044



z	−0.09	−0.08	−0.07	−0.06	−0.05	−0.04	−0.03	−0.02	−0.01	0.00
− 3.80	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
−3.70	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
−3.60	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0002	.0002
−3.50	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
−3.40	.0002	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003
−3.30	.0003	.0004	.0004	.0004	.0004	.0004	.0004	.0005	.0005	.0005
−3.20	.0005	.0005	.0005	.0006	.0006	.0006	.0006	.0006	.0007	.0007
−3.10	.0007	.0007	.0008	.0008	.0008	.0008	.0009	.0009	.0009	.0010
−3.00	.0010	.0010	.0011	.0011	.0011	.0012	.0012	.0013	.0013	.0013
−2.90	.0014	.0014	.0015	.0015	.0016	.0016	.0017	.0018	.0018	.0019
−2.80	.0019	.0020	.0021	.0021	.0022	.0023	.0023	.0024	.0025	.0026
−2.70	.0026	.0027	.0028	.0029	.0030	.0031	.0032	.0033	.0034	.0035
−2.60	.0036	.0037	.0038	.0039	.0040	.0041	.0043	.0044	.0045	.0047

★ Example 2:

- Suppose it is known that the **Dissolution time (D_t)** of certain pharmaceutical tablets is approximately normally distributed with a *mean of* 70 min and a *standard deviation of* 3 min.
- What is the probability that a tablet picked at random from this batch will have D_t between 65 and 74 min?

- $$z_{65} = \frac{(x-\mu)}{\sigma} = \frac{65-70}{3} = -1.67$$
- $$z_{74} = \frac{(x-\mu)}{\sigma} = \frac{74-70}{3} = +1.33$$
- $$P(65 \le x \le 74) = P(-1.67 \le z \le 1.33) = P(-\infty \le z \le 1.33) - P(-\infty \le z \le 1.67) = 0.9082 - 0.0475 = 0.8607$$

★ Example 3:

- In a batch of 10,000 tablets described in the previous example, ***how many tablets*** would you expect to have Dt ***at least 77min***?
- $$z_{77} = \frac{77-70}{3} = 2.33$$

  $$P(x \ge 77) = P(z \ge 2.33) = 1 - 0.9901 = 0.0099$$

Questions

- ***Q1. In a million marriage cases, what is the expected number of married couples with the same birthday?***

  There are 365 possible birthdays for each person (ignoring leap years). The probability that the second person in a couple has the same birthday as the first is 1/365.

  - Expected number of couples with the same birthday = (Total couples) / Probability of same birthday = 1,000,000 *(1/365) = 2739.73

  Also, what is the expected number of these partners that celebrate the 10th of October as their birthday? (Probability) = 1/365 *1/365

  Expected number of such couples: 1,000,000 *1/365 *1/365 = 7.51

- ***Q2. A ball is drawn from a bag containing 3 white and 3 black balls. After the ball is drawn, it is then <u>replaced,</u> and another ball is drawn. This goes on repeatedly.***

  What is the probability that of the first 4 balls drawn, exactly 2 are white?

  Each draw is independent, and the probability of drawing a white ball is 3/6 =1/2 = 0.5

  $P\,(2) = \binom{4}{2}\, 0.5^2\, (1-0.5)^{4-2} = 0.375$

- ***Q3. Determine the (binomial) probability (p) of getting 5 heads in 10 flips of a 2-sided coin using both formula and tables?***

  n = 10, p = 0.5, x = 5

  $P\,(5) = \binom{10}{5}\, 0.5^5\, (1-0.5)^{10-5} = 0.246$

- ***Q4. Quality control analyst at HIKMA Pharmaceutical concluded that the Omeprazole contents in OMERAL® tablets are normally distributed variable with mean μ = 30mg and standard deviation σ = 4mg.***

  a) Find Probability of selecting random tablets contain less than 40mg.

  $z = \frac{40-30}{4} = 2.5$, Using z-table, $P(Z < 2.5) = 0.9938$.

  b) Probability of selecting randomly tablets containing more than 21mg.

  $z = \frac{21-30}{4} = -2.25$, $P(z > -2.25) = 1 - p\,(z \le -2.25) = 1 - 0.0122 = 0.9878$

  c) Probability of selecting randomly tablets containing more than 30mg and less than 35 mg.

  $z_1 = \frac{30-30}{4} = 0$, $z = \frac{35-30}{4} = 1.25$, $p\,(30 < x < 35) = p\,(0 < z < 1.25) = 0.8944 - 0.5 = 0.3944$