Unit 3
Probability

Dr. Hamza Aduraidi

Probability

Probability theory developed from the study of games of chance like dice and cards. A process like flipping a coin, rolling a die or drawing a card from a deck is called a ***probability experiment***. An ***outcome*** is a specific result of a single trial of a probability experiment.

Probability distributions

- Probability theory is the foundation for ***statistical inference***. A ***probability distribution*** is a device for indicating the values that a random variable may have.
- There are two categories of random variables. These are:
  - ***discrete random variables***, and
  - ***continuous random variables***.

Discrete Probability Distributions

- Binomial distribution – the random variable can only assume 1 of 2 possible outcomes. There are a fixed number of trials and the results of the trials are independent.
  - i.e. flipping a coin and counting the number of heads in 10 trials.
- Poisson Distribution – random variable can assume a value between 0 and infinity.
  - Counts usually follow a Poisson distribution (i.e. number of ambulances needed in a city in a given night)

Discrete Random Variable

- A discrete random variable X has a finite number of possible values. The probability distribution of X lists the values and their probabilities.

Value of X	x_1	x_2	x_3	…	x_k
Probability	p_1	p_2	p_3	…	p_k

1. Every probability p_i is a number between 0 and 1.
2. The sum of the probabilities must be 1.

- Find the probabilities of any event by adding the probabilities of the particular values that make up the event.

Example

- The instructor in a large class gives 15% each of A's and D's, 30% each of B's and C's and 10% F's. The student's grade on a 4-point scale is a random variable X (A=4).

Grade	F=0	D=1	C=2	B=3	A=4
Probability	0.10	.15	.30	.30	.15

- What is the probability that a student selected at random will have a B or better?
- ANSWER: P (grade of 3 or 4)=P(X=3) + P(X=4)

  = 0.3 + 0.15 = 0.45

Continuous Variable

- A ***continuous probability distribution*** is a ***probability density function***.
- The area under the smooth curve is equal to 1 and the frequency of occurrence of values between any two points equals the total area under the curve between the two points and the x-axis.

Normal Distribution

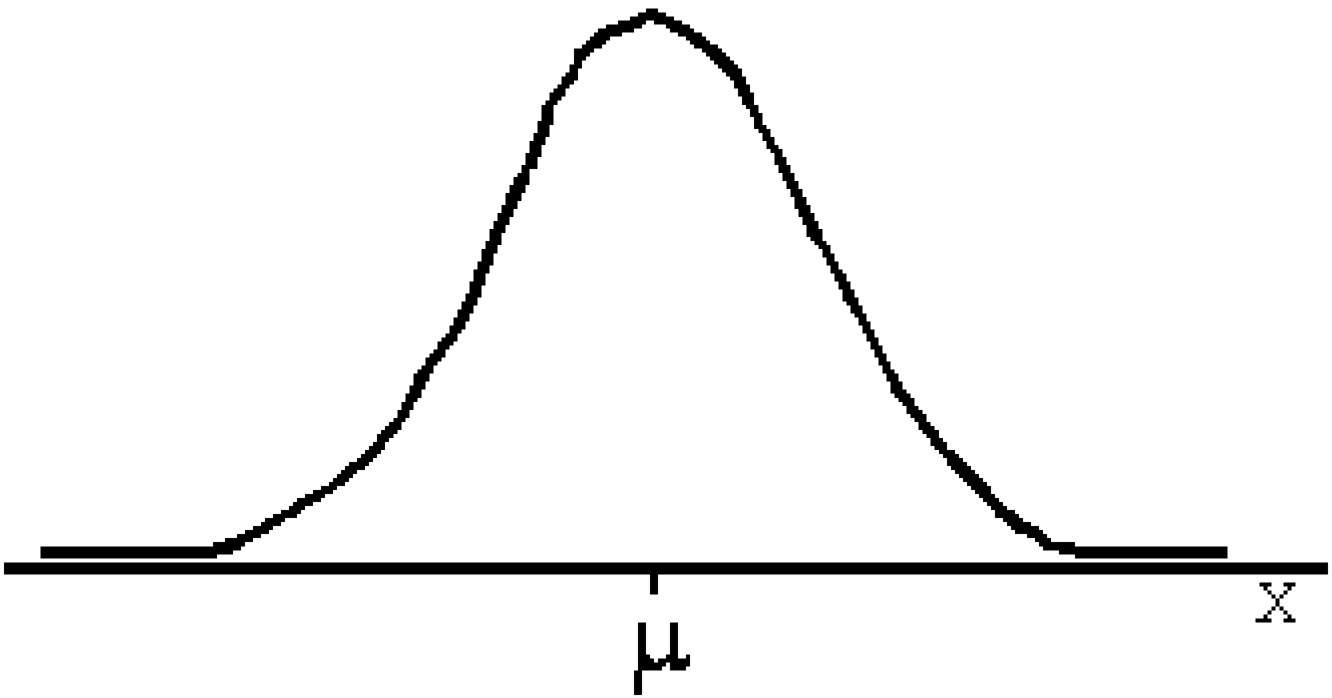
- Also called bell shaped curve, normal curve, or Gaussian distribution.
- A normal distribution is one that is unimodal, symmetric, and not too peaked or flat.
- Given its name by the French mathematician Quetelet who, in the early 19^{th} century noted that many human attributes, e.g. height, weight, intelligence appeared to be distributed normally.

Normal Distribution

- The normal curve is unimodal and symmetric about its mean (μ).
- In this distribution the mean, median and mode are all identical.
- The standard deviation (σ) specifies the amount of dispersion around the mean.
- The two parameters μ and σ completely define a normal curve.

Normal Distribution

- Also called a Probability density function. The probability is interpreted as "area under the curve."
- The random variable takes on an infinite # of values within a given interval
- The probability that X = any particular value is 0. Consequently, we talk about intervals. The probability is = to the area under the curve.
- The area under the whole curve = 1.


μ
x

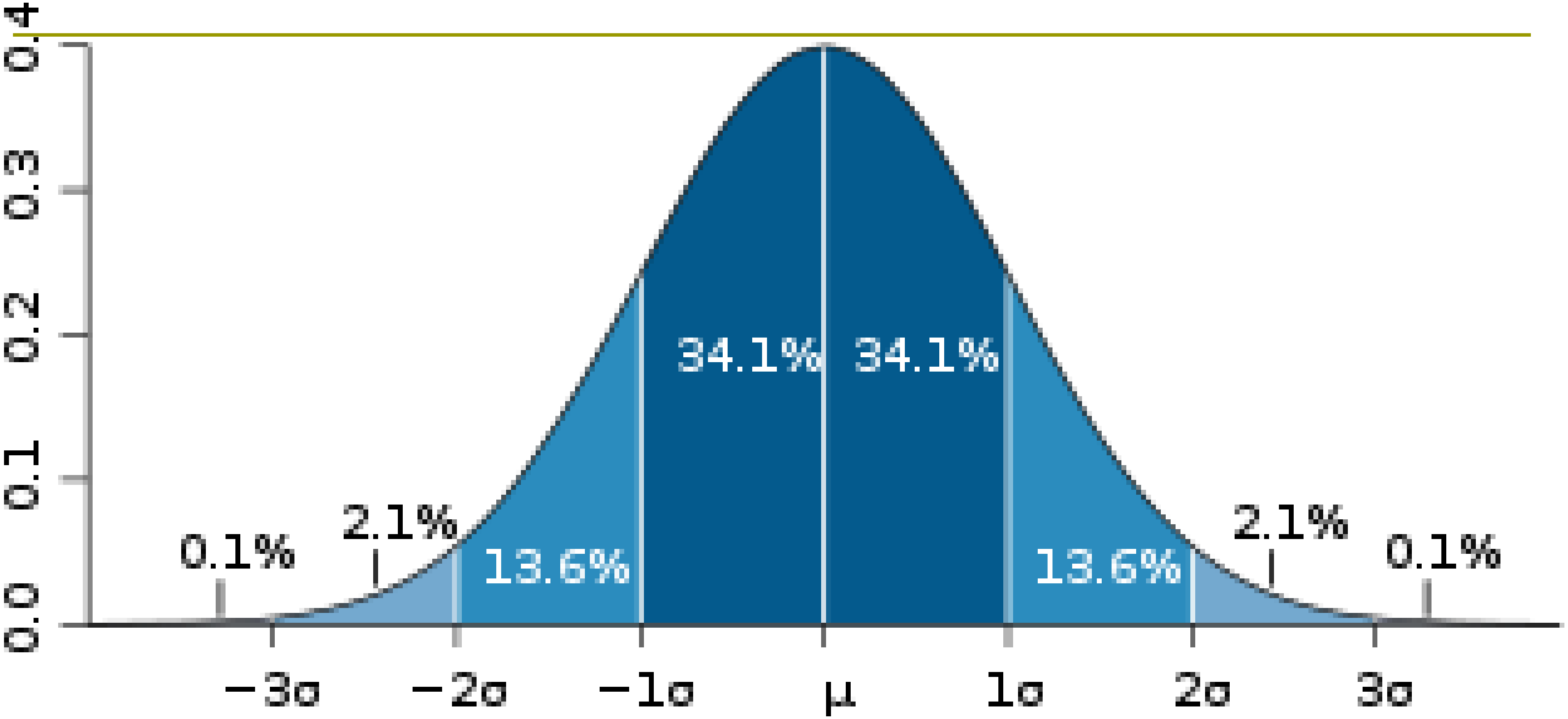
Properties of a Normal Distribution

1. It is symmetrical about μ .
2. The mean, median and mode are all equal.
3. The total area under the curve above the x-axis is 1 square unit. Therefore 50% is to the right of μ and 50% is to the left of μ.
4. Perpendiculars of:

   $\pm 1\ \sigma$ contain about 68%;
   $\pm 2\ \sigma$ contain about 95%;
   $\pm 3\ \sigma$ contain about 99.7%

   of the area under the curve.

The normal distribution



The Standard Normal Distribution

- A ***normal distribution*** is determined by μ and σ. This creates a family of distributions depending on whatever the values of μ and σ are.
- The ***standard normal distribution*** has

  $\mu=0$ and $\sigma=1$.

Standard Z Score

- The ***standard z score*** is obtained by creating a variable z whose value is

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

- Given the values of μ and σ we can convert a value of x to a value of z and find its probability using the table of normal curve areas.

Z Score

- Is a standard score that indicates how many SDs from the mean a particular values lies.
- Z = Score of value – mean of scores divided by standard deviation.

Standard Normal Scores

- How many standard deviations away from the mean are you?

Standard Score (Z) = $\dfrac{\textit{Observation} - \textit{mean}}{\textit{Standard deviation}}$

"Z" is normal with mean 0 and standard deviation of 1.

Standard Normal Scores

A standard score of:

- **Z = 1:** The observation lies one SD above the mean
- **Z = 2:** The observation is two SD above the mean
- **Z = -1:** The observation lies 1 SD below the mean
- **Z = -2:** The observation lies 2 SD below the mean

Standard Normal Scores

- Example: Male Blood Pressure, mean = 125, s = 14 mmHg
  - BP = 167 mmHg

$$Z = \frac{167 - 125}{14} = 3.0$$

  - BP = 97 mmHg

$$Z = \frac{97 - 125}{14} = -2.0$$

What is the Usefulness of a Standard Normal Score?

- It tells you how many SDs (s) an observation is from the mean
- Thus, it is a way of quickly assessing how "unusual" an observation is
- *Example:* Suppose the mean *BP* is 125 mmHg, and standard deviation = 14 mmHg
  - Is 167 mmHg an unusually high measure?
  - If we know Z = 3.0, does that help us?

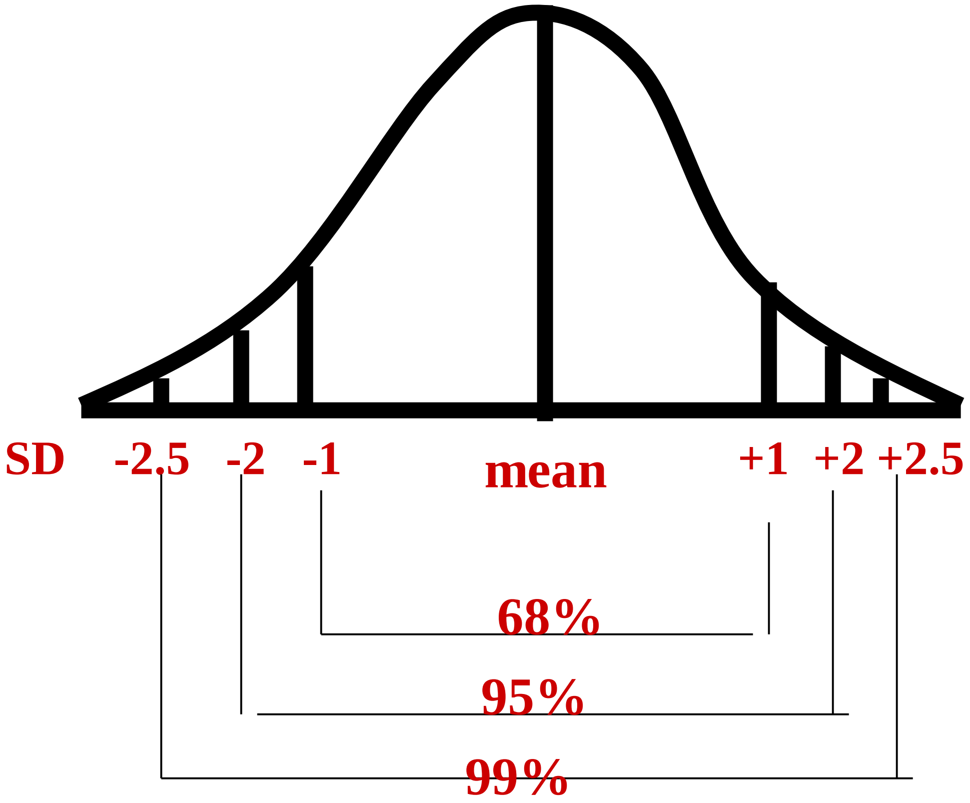
Standardizing Data: Z-Scores

- We can convert the original scores to new scores with $\overline{X} = 0$ and $s = 1$.
- This will give us a pure number with no units of measurement.
- Any score below the mean will now be negative.
- Any score at the mean will be 0.
- Any score above the mean will be positive.

Standardizing Data: Z-Scores

- No matter what you are measuring, a Z-score of more than +5 or less than – 5 would indicate a very, very unusual score.
- For standardized data, if it is normally distributed, 95% of the data will be between ±2 standard deviations about the mean.
- If the data follows a normal distribution,
  - **95% of the data will be between -1.96 and +1.96.**
  - **99.7% of the data will fall between -3 and +3.**
  - **99.99% of the data will fall between -4 and +4.**

Relationship between SD and frequency distribution



Importance of Normal Distribution to Statistics

- Although most distributions are not exactly normal, most variables tend to have approximately normal distribution.
- Many inferential statistics assume that the populations are distributed normally.
- The normal curve is a probability distribution and is used to answer questions about the likelihood of getting various particular outcomes when sampling from a population.

Why Do We Like The Normal Distribution So Much?

- There is nothing "special" about standard normal scores
  - These can be computed for observations from any sample/population of continuous data values
  - The score measures how far an observation is from its mean in standard units of statistical distance
- But, if distribution is not normal, we may not be able to use Z-score approach.

Normal Distribution

- Q Is every variable normally distributed?
- A Absolutely not
- Q Then why do we spend so much time studying the normal distribution?
- A Some variables are normally distributed; a bigger reason is the “Central Limit Theorem”!!!!!!!!!!!!!!!!!!!!!!!!!!!???????????
?

Central Limit Theorem

- describes the characteristics of the "**population of the means"** which has been created from the means of an infinite number of random population samples of size (N), all of them drawn from a given "**parent population**".
- It predicts that regardless of the distribution of the parent population:
  - The **mean** of the population of means is always equal to the mean of the parent population from which the population samples were drawn.
  - The **standard deviation** of the population of means is always equal to the standard deviation of the parent population divided by the square root of the sample size (N).
  - The distribution of means will increasingly approximate a **normal distribution** as the size N of samples increases.

Central Limit Theorem

- A consequence of Central Limit Theorem is that if we average measurements of a particular quantity, the distribution of our average tends toward a normal one.
- In addition, if a measured variable is actually a combination of several other uncorrelated variables, all of them "contaminated" with a random error of any distribution, our measurements tend to be contaminated with a random error that is normally distributed as the number of these variables increases.
- Thus, the Central Limit Theorem explains the ubiquity of the famous bell-shaped "Normal distribution" (or "Gaussian distribution") in the measurements domain.

Normal Distribution

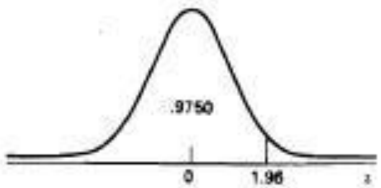
Note that the normal distribution is defined by two parameters, μ and σ . You can draw a normal distribution for any μ and σ combination. There is one normal distribution, Z, that is special. It has a μ = 0 and a σ = 1. This is the Z distribution, also called the *standard normal* distribution. It is one of trillions of normal distributions we could have selected.

Standard Normal Variable

- It is customary to call a standard normal random variable Z.
- The outcomes of the random variable Z are denoted by *z.*
- The table in the coming slide give the area under the curve (probabilities) between the mean and *z.*
- The probabilities in the table refer to the likelihood that a randomly selected value Z is equal to or less than a given value of *z* and greater than 0 (the mean of the standard normal).

Table of Normal Curve Areas

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

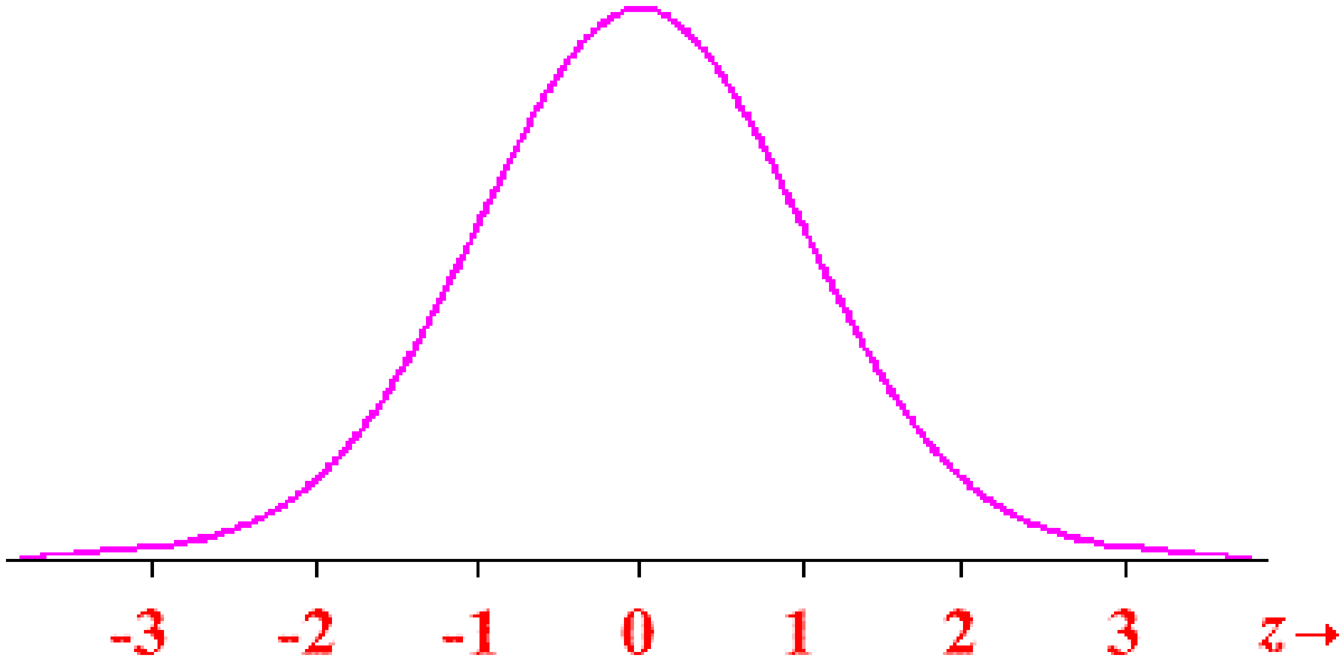


z	−0.09	−0.08	−0.07	−0.06	−0.05	−0.04	−0.03	−0.02	−0.01	0.00	z
−3.80	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	−3.80
−3.70	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	−3.70
−3.60	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0002	.0002	−3.60
−3.50	.0002	.0002	.0002	.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 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

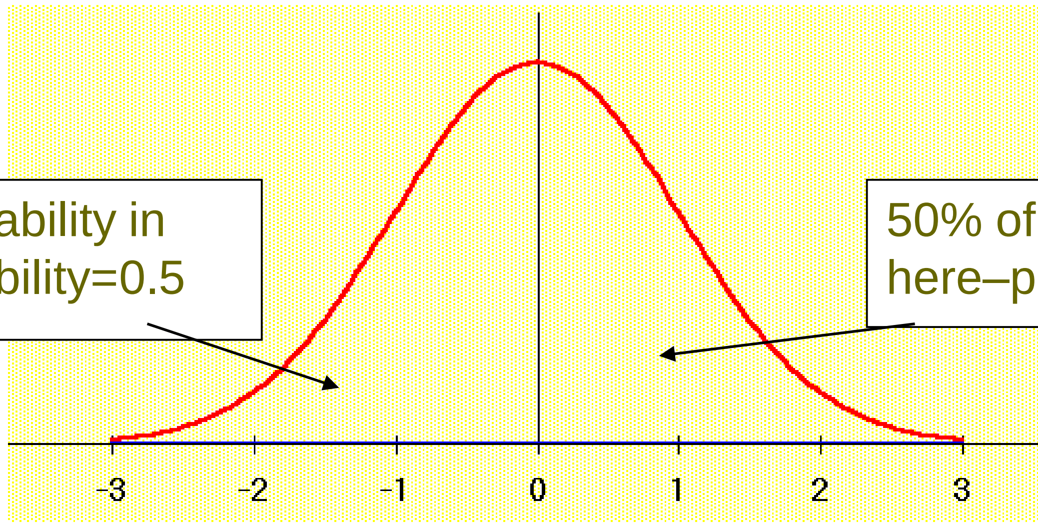
Standard Normal Curve



Standard Normal Distribution

50% of probability in here –probability=0.5

50% of probability in here–probability=0.5

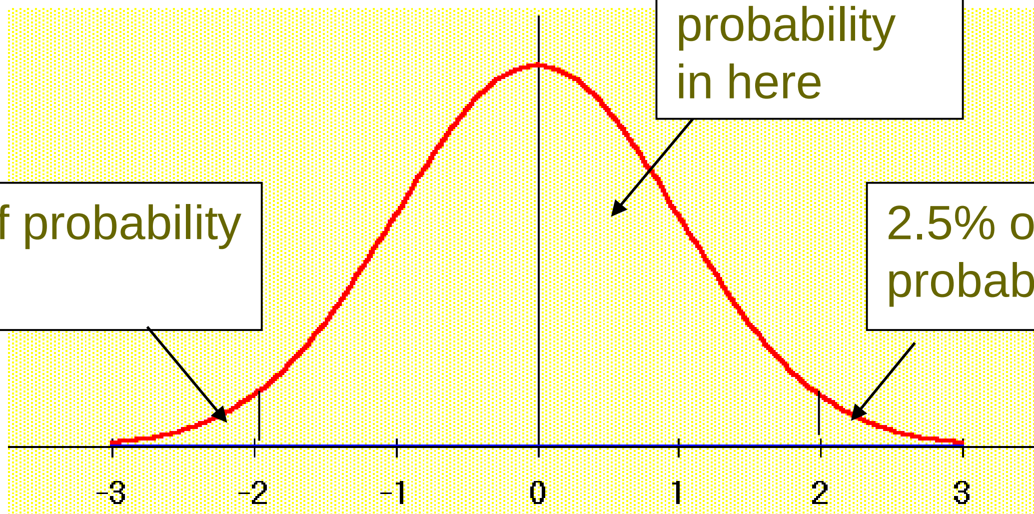


Standard Normal Distribution

95% of probability in here

2.5% of probability in here

2.5% of probability in here



Standard Normal Distribution with 95% area marked

Calculating Probabilities

- Probability calculations are always concerned with finding the probability that the variable assumes any value in an interval between two specific points *a* and *b.*
- The probability that a continuous variable assumes the a value between *a* and *b* is the area under the graph of the density between *a* and *b.*

Finding Probabilities

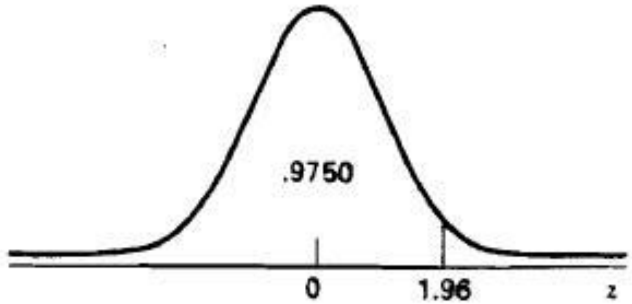
(a) What is the probability that $z < -1.96$?

(1) Sketch a normal curve

(2) Draw a line for $z = -1.96$

(3) Find the area in the table

(4) The answer is the area to the left of the line $P(z < -1.96) = .0250$



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

Finding Probabilities