UCLA Statistics

STATISTICS 10
Introduction to Statistical Reasoning

Bingling Wang
Ph.D., Lecturer, Department of Statistics

PROBABILITY DISTRIBUTIONS

Random Variable

Map outcomes of random processes to numbers

When the value of a variable is determined by a chance event,
that variable is called a random variable (r.v.).

- **Discrete random variable**: numerical values that can be listed or counted
  - The number of phone calls you will receive during the day
  - The number of bus lines that go to UCLA

- **Continuous random variable**: take any values in an interval and are often measurements.
  - The exact weight of a newborn baby
  - The exact time it takes to commute to school by bus

Probability Distribution

Probability distribution describes a random variable (r.v.)

Depending on the variable type, we take different approaches to display or analyze the probability distribution of the variable.

Tables Graphs Equations

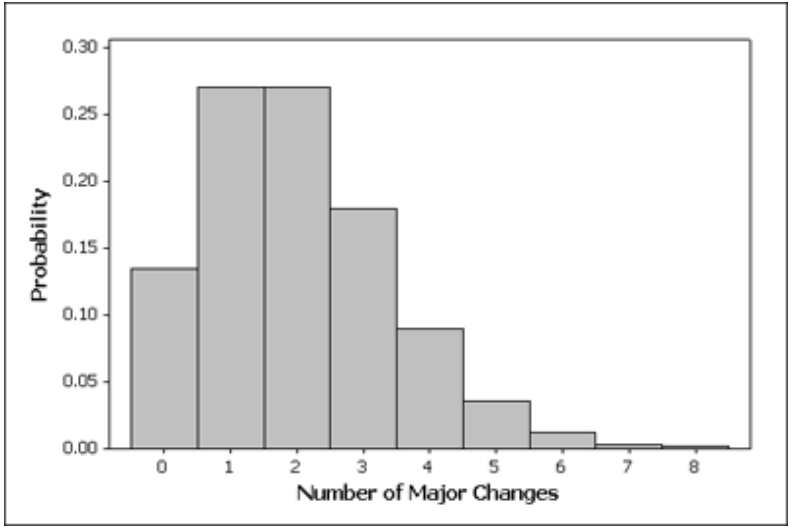
- The outcomes are random
- Probabilities are assigned to all possible outcome
- All probability values must be between 0 and 1
- The sum of the probabilities of all possible outcomes must be 1.

DISCRETE PROBABILITY DISTRIBUTION

Example

Consider the random variable “The number of times a student changes major at a college.”

X=# changes in major	0	1	2	3	4	5	6	7	8
Probability	0.135	0.271	0.271	0.180	0.090	0.036	0.012	0.003	0.002



What is the probability that a college student will change majors at most once?
What is the probability that a college student will change majors at least twice?

Exercise

Suppose you roll a fair 6-sided die:

- You win 5 points if you roll a 5 or a 6.
- You lose 1 point if you roll a 1.
- For any other outcome you win or lose nothing.

Create a table showing the probability distribution for the points you will win by playing this game.

Let X be the random variable that represents the number of points you win.

X	P(x)
-1	
0	
5	

Example

Probability Distributions Can Also Be Equations

Example: suppose we roll a fair die until we see a 6.

X = number of trials

X	1	2	3	.....
Probability	$\frac{1}{6}$	$\frac{5}{6}\times\frac{1}{6}$	$\frac{5}{6}\times\frac{5}{6}\times\frac{1}{6}$	

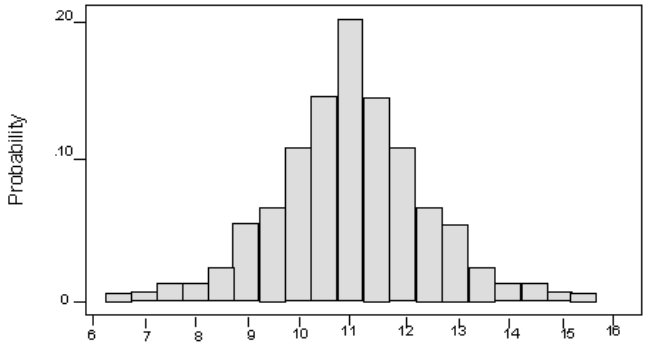
$$P(x) = \left(\frac{5}{6}\right)^{x-1}\left(\frac{1}{6}\right)$$

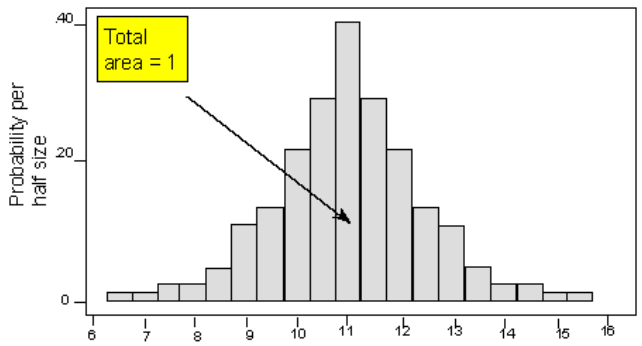
$$P(x = 10) = \left(\frac{5}{6}\right)^{10-1}\left(\frac{1}{6}\right) = 0.0323$$

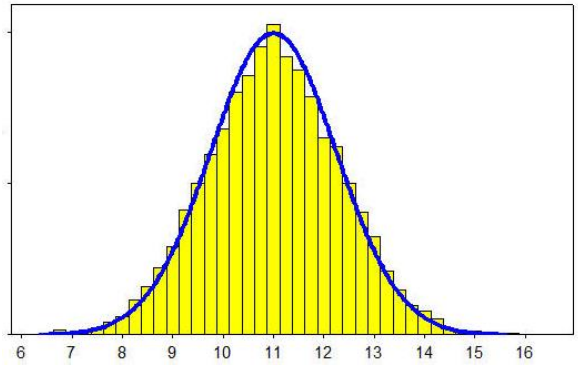
$$\mathrm{P}(\mathrm{x} \leq 3) = \frac{1}{6} + \left(\frac{5}{6}\right)\left(\frac{1}{6}\right) + \left(\frac{5}{6}\right)^{2}\left(\frac{1}{6}\right) = 0.4213$$

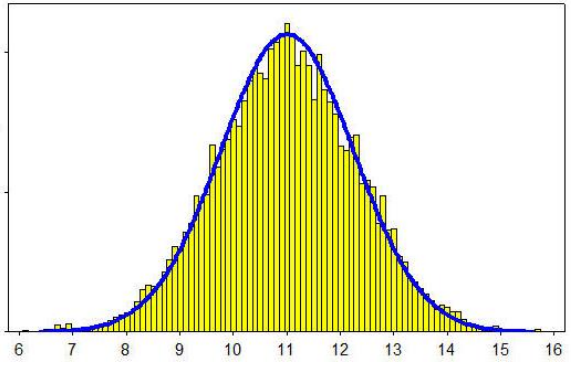
CONTINUOUS PROBABILITY DISTRIBUTION

Transition to Continuous Random Variables



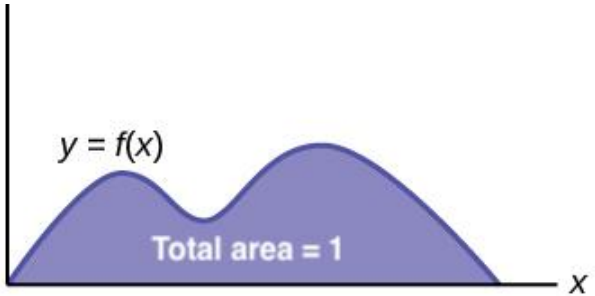


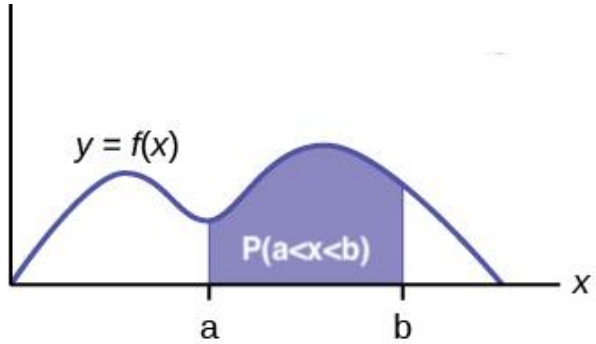




Continuous Probability Distribution

- The continuous probability distribution is usually visualized by a curve
  -- **Probability density curve**
  - The density curve cannot lie below the x-axis
- The probability for a continuous random variable is calculated as
  -- **Area under the curve**
  - The total area under the curve must equal 1
  - The probability of any single value occurring is zero



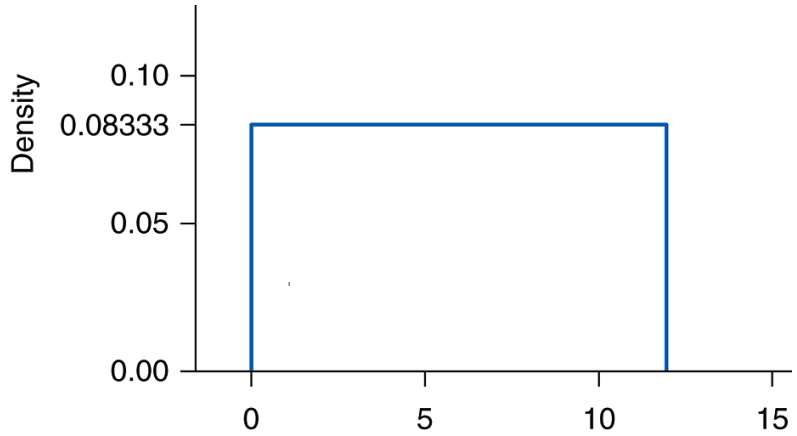


Finding Probabilities

Example:

A bus arrives at a certain stop every 12 minutes. The graph shows the probability distribution for wait times before the bus arrives.

Use the distribution to find the probability of a wait time between 0 and 10 minutes.



NORMAL PROBABILITY DISTRIBUTION

The Normal Distribution

The most widely used probability model for continuous numerical variables.

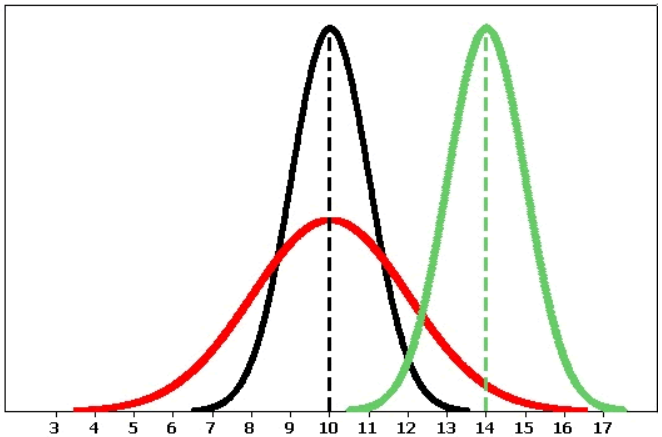
Also called Gaussian distribution / Laplace-Gauss distribution

Defined by the **mean** and the **standard deviation**

Notation: $N(\mu, \sigma)$

-- a Normal distribution with mean μ, and standard deviation σ

- **Characteristics of the distribution**:
  - Unimodal
  - Symmetric
  - Bell shaped



Finding Normal Probabilities

The first and helpful step is usually to:
sketch the curve, label it appropriately, and shade the area of interest.

1. Using the exact probability density function (pdf)

pdf for normal distribution:

$$P(a < x < b) = \int_a^b f(x)\,dx \qquad f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

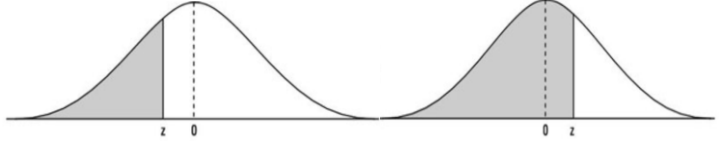
2. With computer technology

In R: function **pnorm(q, mean, sd)** computes probabilities $P(x \leq q)$ from a normal distribution with specified mean and sd.

3. With standard normal table -- Z-table

Standard Normal: the normal model with mean 0 and standard deviation 1, $N(0, 1)$.

Z Table



$Z \sim N(0,1)$ Number in the table represents the left-tail probability $P(Z \leq z^*)$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.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

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
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

z	.00
−5.0	.000000287
−4.5	.00000340
−4.0	.0000317
−3.5	.000233

z	.00
3.5	.999767
4.0	.9999683
4.5	.9999966
5.0	.999999713

Finding Normal Probabilities with Z-table

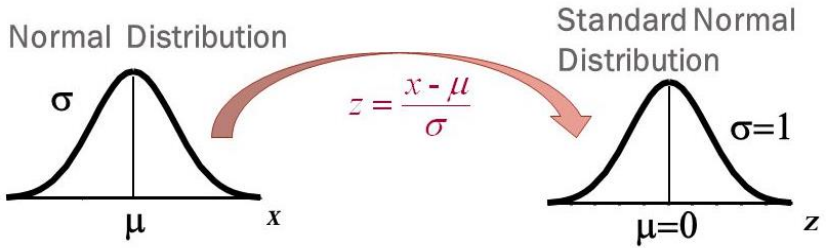
This requires us to convert the normal distribution $N(\mu, \sigma)$ to a standard normal distribution $N(0,1)$ by converting the normal values to standard units — **standardization / z-transformation.**

Steps:

$$X \sim N(\mu, \sigma)$$

1. Convert the values to standard units, $z = \frac{x-\mu}{\sigma}$.
2. Find the probability associated with the z-score using the Z-table.

Standardizing data into z-scores does not change the shape of the distribution.



Z Table Exercises

Suppose X is a normal random variable, $X \sim N(\mu, \sigma)$.

1. What is the probability of X taking a value less than 2.8 standard deviations below its mean?

2. What is the probability of X taking a value more than 0.75 standard deviations above its mean?

3. What is the probability of X taking a value between 1 standard deviation below and 2 standard deviation above its mean?

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
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

Normal Probability Exercise

Suppose the distribution of teenagers' weights follows a normal distribution $N(125, 10)$.

What is the probability that a randomly selected teenager weighs between 135.5 and 140.6 pounds?

1. Convert the values to standard units;

2. Use the Z-table to find the area under the curve between the two values.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.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