BU7142 Foundations of Business Analytics

Lecture 3

Random Variables and Normal Distribution

Dr Yufei Huang

Contents

- Discrete Random Variables
- Continuous Random Variables
- Normal Distribution

Random Variables

- Definition. A variable is called random if its value depends on chance.

- Notation. We denote random variables by capital letters, e.g. X,Y,…, and the lower cases, x, y,…, denote the outcomes (actual values).

- Examples

1. Assign to the result of a coin toss experiment the numbers:

   X=0 if 'tail' is obtained, X=1 if 'head' is obtained

   The result of a coin toss depends on chance and is a random variable.

2. Let Y be the random variable: "service time when calling customer service"

   The values Y can take are: from 0 to ∞

   The actual result of Y depends on chance.

Types of Random Variables

1. Discrete Random variables

Definition. A discrete random variable is a variable which can assume at most a countable number of values.

Examples: coin toss, the result on the face of a die.

2. Continuous random variables

Definition. A continuous random variable may take on any value in an interval of numbers (i.e., its possible values are uncountably infinite).

Example: service time.

Discrete Random Variable

Probability Distribution

- Definition. A probability distribution for a discrete random variable is a mutually exclusive listing of all possible numerical outcomes for that variable such that a particular probability of occurrence is associated with each outcome.

- If x is a possible outcome of a discrete random variable X, we can denote the probability of x, by P(x) or P(X=x).

- The probability distribution of a discrete random variable X must satisfy the following two conditions.

  1. $P(x) \geq 0$ for all possible values x of a random variable X
  2. $\sum_{\text{all } x} P(x) = 1$

Cumulative Distribution

- We use $F(x)=P(X \leq x)$ to denote the probability of all outcomes from X that are smaller or equal to x. F(x) is called the cumulative distribution function.

- Example. x is the outcome of a fair die

x=outcome of a fair die	P(x)	F(x)	Reward
1	1/6	1/6	£6
2	1/6	2/6	£12
3	1/6	3/6	£18
4	1/6	4/6	£24
5	1/6	5/6	£30
6	1/6	1	£36

- *Question: What is the reward value that you are expecting?*

Expected Value

- The expected value of X is the mean of its outcomes, weighed by the probability of each outcome.

- Let X be a random variable. Denote its possible outcomes by $x_1, \ldots, x_N$. The expected value of X, μ, is given by:

$$\mu = E(X) = \sum_{i=1}^{N} x_i P(x_i).$$

- *Question: what's the difference between expected value and average value?*

The expected value of a function of a random variable

Definition. Let h(X) be a function of a discrete random variable X. Denote the possible outcomes of X by $X_1,\ldots, X_N$. Then

$$E[h(X)] = \sum_{i=1}^{N} h(X_i)P(X_i).$$

Example. If h(X)=aX+b then

$$E[h(X)] = \sum_{i=1}^{N} h(X_i)P(X_i) = \sum_{i=1}^{N} (aX_i + b)P(X_i) = \sum_{i=1}^{N} aX_iP(X_i) + \sum_{i=1}^{N} bP(X_i)$$

$$\underset{\sum_i cg(i) = c\sum_i g(i)}{=} a\sum_{i=1}^{N} X_iP(X_i) + b\sum_{i=1}^{N} P(X_i) \underset{\sum_{i=1}^{N} P(X_i)=1}{=} aE(X) + b\cdot 1 = aE(X) + b.$$

Therefore, if h(X)=aX+b then E[h(X)]=aE(X)+b.

The expected value of a function of a random variable

$$E[h(X)] = \sum_{i=1}^{N} h(X_i)P(X_i).$$

Example. If $h(X)=X^2$ then

$$E[h(X)] = E[X^2] = \sum_{i=1}^{N} X_i^2 P(X_i).$$

Example. If $h(X)=(X-m)^2$ then

$$E[h(x)] = E[(X-m)^2] = \sum_{i=1}^{N} (X_i - m)^2 P(X_i).$$

Variance

- The variance of a random variables measures its spread.

- Definition. Let X be a discrete random variable. Denote its possible outcomes by $X_1, \ldots, X_N$. Let $\mu = E(X)$. Then the variance of X is given by

$$\sigma^2 = V(X) = E[(X-\mu)^2] = \sum_{i=1}^{N} (X_i - \mu)^2 P(X_i).$$

probar que estas dos son iguales

- It can also be written as:

$$\sigma^2 = V(X) = E[(X-\mu)^2] = E(X^2) - [E(X)]^2$$

Standard Deviation

- The standard deviation of X is given by

$$\sigma = SD(X) = \sqrt{V(X)} = \sqrt{E[(X-\mu)^2]} = \sqrt{\sum_{i=1}^{N}(X_i-\mu)^2 P(X_i)}$$

$$= \sqrt{E(X^2)-(E(X))^2}$$

- Remark. Both standard deviation and variance can be considered as risk measures.

Example

x=outcome of a fair die	P(x)	F(x)	Reward
1	1/6	1/6	£6
2	1/6	2/6	£12
3	1/6	3/6	£18
4	1/6	4/6	£24
5	1/6	5/6	£30
6	1/6	1	£36

- Expected reward:

$$\mu = E(X) = \sum_{i=1}^{N} x_i P(x_i)$$

$$= 6*(1/6) + 12*(1/6) + 18*(1/6) + 24*(1/6) + 30*(1/6) + 36*(1/6)$$

$$= 1 + 2 + 3 + 4 + 5 + 6 = 21$$

- Note. 21 is different from all possible values of reward.

Example

x=outcome of a fair die	P(x)	F(x)	Reward
1	1/6	1/6	£6
2	1/6	2/6	£12
3	1/6	3/6	£18
4	1/6	4/6	£24
5	1/6	5/6	£30
6	1/6	1	£36

- Variance of reward:

$$\sigma^2 = V(X) = E[(X-\mu)^2] = \sum_{i=1}^{N} (X_i - \mu)^2 P(X_i) = 105$$

$$\sigma^2 = V(X) = E[(X-\mu)^2] = E(X^2) - [E(X)]^2 = 105$$

Bernoulli Distribution

- Definition. A Bernoulli variable with parameter p is a random variable which has only two possible outcomes, usually denoted by 0 and 1, and satisfies:
  - $P(X = 0) = 1-p$, and
  - $P(X = 1) = p$.

- Notation. If X follows Bernoulli distribution with parameter p, we write: X~BER(p). p can be considered the probability of success.

- Example. Coin toss follows X~Ber(0.5).

- Remark. If X is a Bernoulli variable with parameter p then
  - $E(X)=p$ ⇨ $E(X)=1*p+0*(1-p)$
  - $E(X^2)=p$ ⇨ $E(X^2)=1^2*p+0^2*(1-p)=p$
  - $V(X)=p(1-p)$ ⇨ $V(X)=E(X^2)-E(X)^2=p-p^2=p(1-p)$.

pistaca para demostrar las dos de arriba

Binomial Distribution

- Definition. A binomial random variable is a random variable that counts the number of successes in many independent, identical Bernoulli trials, that is:

$$X=X_1+X_2+\ldots+X_n,$$

where X_i for i=1,…n is a Bernoulli trial with probability p.

- Independence. The results of each trial do not depend on each other.

- Notation: A binomial random variable is denoted by $X \sim B(n,p)$.

- Example. The number of defective computer chips in a sample of 100 chips.

- Remark. Binomial distribution measures the number of successes.

Binomial Distribution

- Theorem. Let X~B(n,p) be a binomial random variable. Then for x=0,1,2,…n:

$$1.\ P(X=x)=\binom{n}{x}p^x(1-p)^{n-x},$$

$$2.\ E(X)=np,$$

$$3.\ V(X)=np(1-p),$$

- Example. The probability that a chip is not defective is 0.6. What is the probability that when 5 chips are manufactured, there are 2 defective chips?

  Solution. P=0.6, n=5. 2 defects out of 5, so 3 good chips.

$$P(X=3)=\binom{5}{3}0.6^3(1\text{-}0.6)^2=\frac{5!}{3!2!}0.6^3(0.4)^2$$

$$=\frac{3!4\cdot 5}{3!2!}0.6^3(0.4)^2=10\cdot 0.6^3(0.4)^2=0.3456.$$

- Using Binomial Distribution Calculation Table

Binomial Distribution

- Exercise (Aczel, Ex. 3.39). A management graduate is applying for nine jobs, and believes that she has in each of the nine cases a constant and independent 0.48 probability of getting an offer. What is the probability that she will have at least three offers?

Solution. p=0.48, n=9. $P(X=x)=\binom{n}{x}p^x(1-p)^{n-x}$

$$P(X\geq 3)=1-P(X<3)=1-[P(X=0)+P(X=1)+P(X=2)]$$

$$=1-\binom{9}{0}0.48^0(1-0.48)^9-\binom{9}{1}0.48^1(1-0.48)^8-\binom{9}{2}0.48^2(1-0.48)^7$$

$$=1-1*1*0.52^9-\frac{9!}{8!1!}*0.48*0.52^8-\frac{9!}{7!2!}*0.48^2*0.52^7$$

$$=1-0.52^9-\frac{8!*9}{8!*1!}*0.48*0.52^8-\frac{7!*8*9}{7!*2!}*0.48^2*0.52^7$$

$$=1-0.52^9-9*0.48*0.52^8-4*9*0.48^2*0.52^7=0.8889$$

Mathematical Background: Combinatorics

Definition. Let n be a natural number. n!=1·2 ·… ·n.

Examples.

3!=1·2·3=6

4!=1·2·3·4=24.

100!=9.3326e+157

Theorem. The number of ways to choose k elements from a set of n elements is given by $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

By definition, $\binom{n}{0} = 1$

Example. Calculate $\binom{5}{3}$

Solution. $\binom{5}{3} = \frac{5!}{3!(5-3)!} = \frac{5!}{3!2!} = \frac{1\cdot 2\cdot 3\cdot 4\cdot 5}{1\cdot 2\cdot 3\cdot 1\cdot 2} = \frac{4\cdot 5}{1\cdot 2} = 10.$

Example

In how many ways can a manager choose a team of 3 people from a department with 7 people to perform a certain task?

Solution.

N=7, k=3.

$$\binom{7}{3} = \frac{7!}{3!(7-3)!} = \frac{7!}{3!4!} = \frac{1\cdot 2\cdot 3\cdot 4\cdot 5\cdot 6\cdot 7}{1\cdot 2\cdot 3\cdot 1\cdot 2\cdot 3\cdot 4} = \frac{5\cdot 6\cdot 7}{1\cdot 2\cdot 3} = 5\cdot 7 = 35$$

Therefore, there are 35 possible teams of 3 people the manager should choose from.

Denote the employees by 1,2,..7.
The possible teams are:

1 2 3	1 2 4	1 2 5	1 2 6	1 2 7
1 3 4	1 3 5	1 3 6	1 3 7	
1 4 5	1 4 6	1 4 7		
1 5 6	1 5 7			
1 6 7				
2 3 4	2 3 5	2 3 6	2 3 7	
2 4 5	2 4 6	2 4 7		
2 5 6	2 5 7			
2 6 7				
3 4 5	3 4 6	3 4 7		
3 5 6	3 5 7			
3 6 7				
4 5 6	4 5 7			
4 6 7				
5 6 7				

Continuous Random Variable

Continuous Random Variable

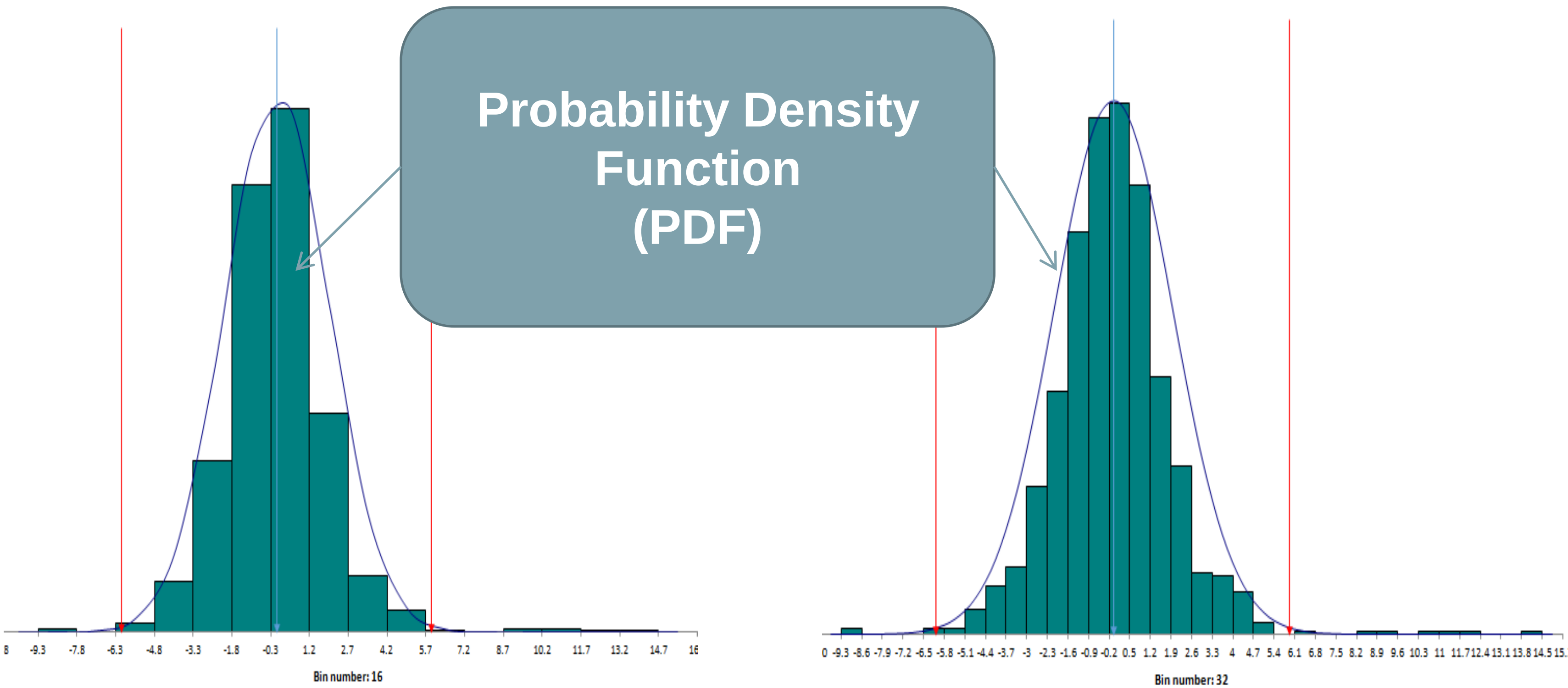
Definition. A continuous random variable is a random variable that can take on any value in an interval of numbers.

Examples:

- Time (e.g. time required to complete a stage in a project)
- Investment risk
- Profit (though one may argue that money has a finite resolution)

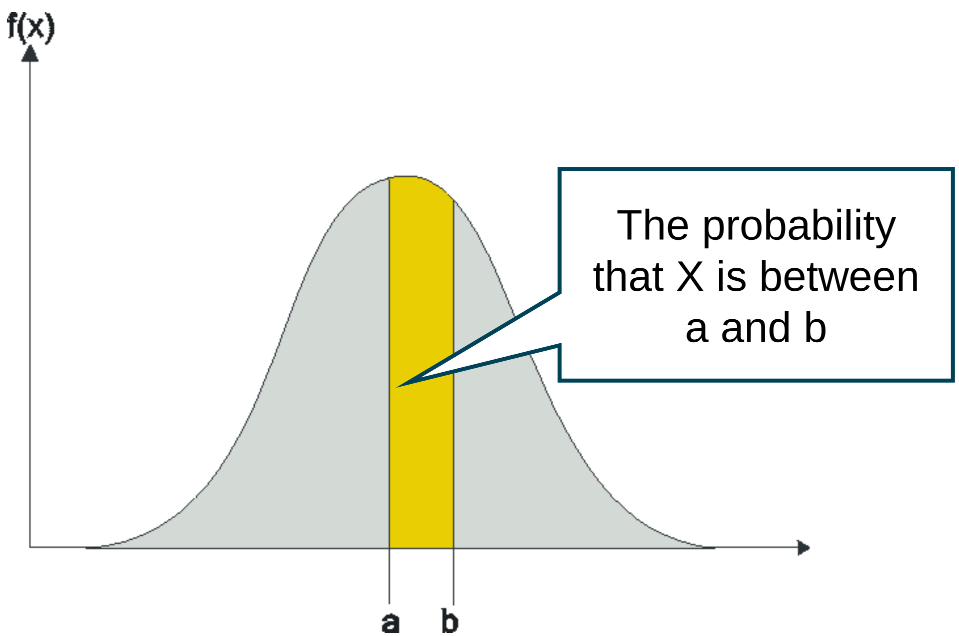
Probability Density Function

- As the number of outcomes increase, a probability distribution of the discrete random variable resembles the probability density function of a continuous variable



Probability Density Function

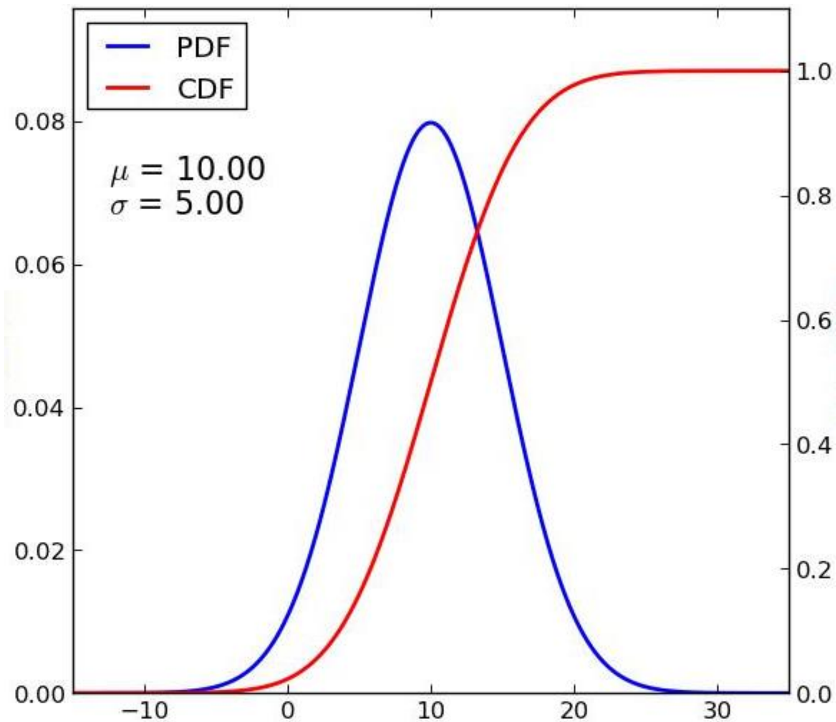
- Properties of the probability density function f(x):
  1. $f(x) \geq 0$ for all x
  2. The total area under f(x) is 1
  3. The probability that X is between a and b is the area under f(x) between a and b.



Cumulative Distribution Function

- Definition. The Cumulative Distribution Function (CDF) of a continuous random variable is $F(x)=P(X \leq x)$.

- The meaning of the cumulative distribution function:
  - The area under f (x) between the smallest possible value of X and x
  - The probability that $X \leq x$.



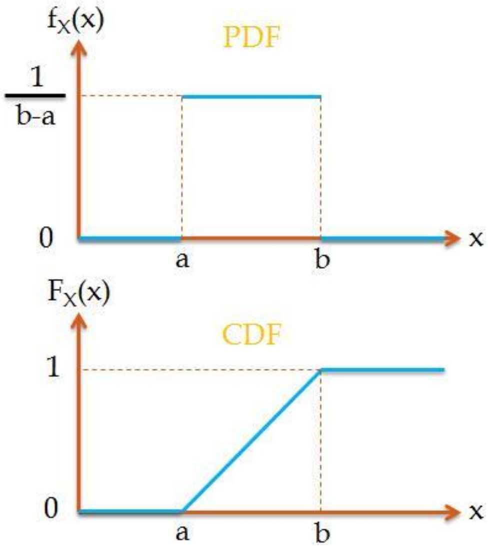
Uniform Distribution

- Definition. If X follows a uniform distribution between a and b, then we write: X~U(a,b).

- The probability density function of the uniform distribution:

$$f(x)=\begin{cases}\dfrac{1}{b-a} & a\le x\le b\\ 0 & \text{Otherwise}\end{cases}$$

$$F(x)=\begin{cases}0 & x<a\\ \dfrac{x-a}{b-a} & a\le x\le b\\ 1 & x>b\end{cases}$$

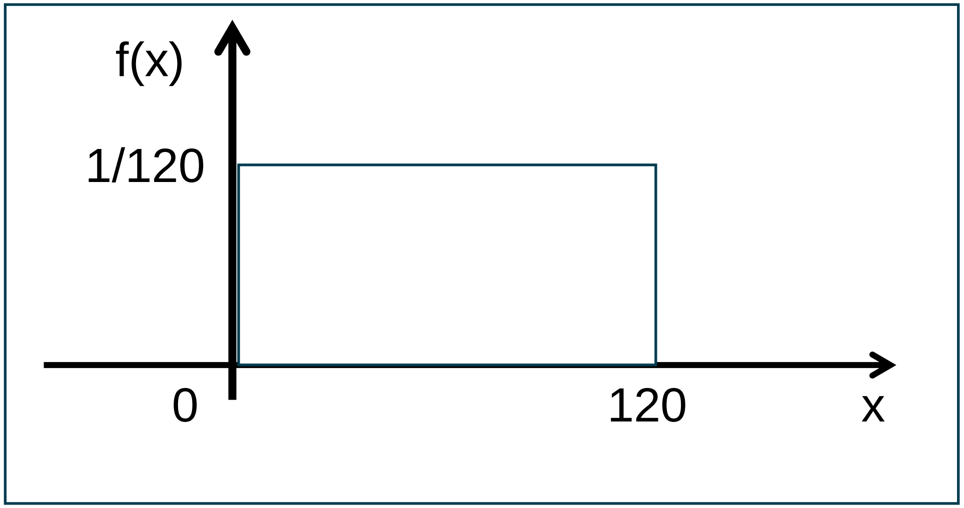


- Remark. If X~U(a,b) and a ≤ c ≤ d ≤ b then:
  1. P(c≤x ≤d)=(d-c)/(b-a)
  2. E(X)=(a+b)/2
  3. $V(X)=(b-a)^2/12$.

Uniform Distribution

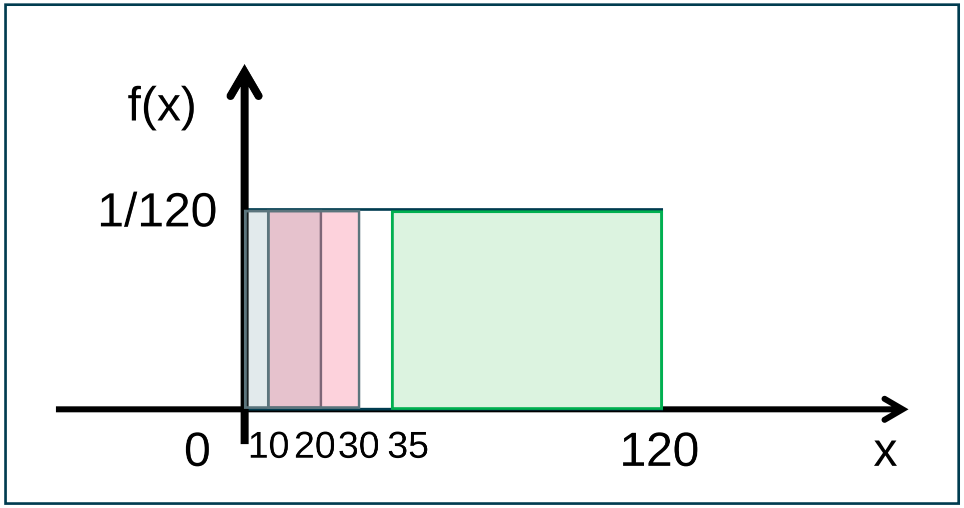
- Example. Assume that the time between two online orders of customers has a uniform distribution between 0 to 120 seconds. What is the probability that the time between two online orders is:

1. less than 20 seconds?
2. between 10 and 30 seconds?
3. more than 35 seconds?
4. What are the mean and standard deviation of the time between online orders?



Uniform Distribution

- Solution.
  1. $P(X \leq 20) = P(0 \leq X \leq 20) = ((20-0) \cdot 1/120 = 1/6 = 0.1667.$
  2. $P(10 \leq X \leq 30) = (30-10) \cdot 1/120 = 1/6 = 0.1667.$
  3. $P(X \geq 35) = (120-35) \cdot 1/120 = 0.7083.$
  4. $E(X) = (a+b)/2 = (0+120)/2 = 60.$ $V(X) = (b-a)^2 /12 = (120-0)^2 /12 = 1200.$
     $SD = 1200^{0.5} = 34.641.$



Normal Distribution

Normal Distribution: PDF

- The probability density function of a normal distribution with mean μ and standard deviation σ is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

- Example: A normal distribution with mean 3 and standard deviation 5:

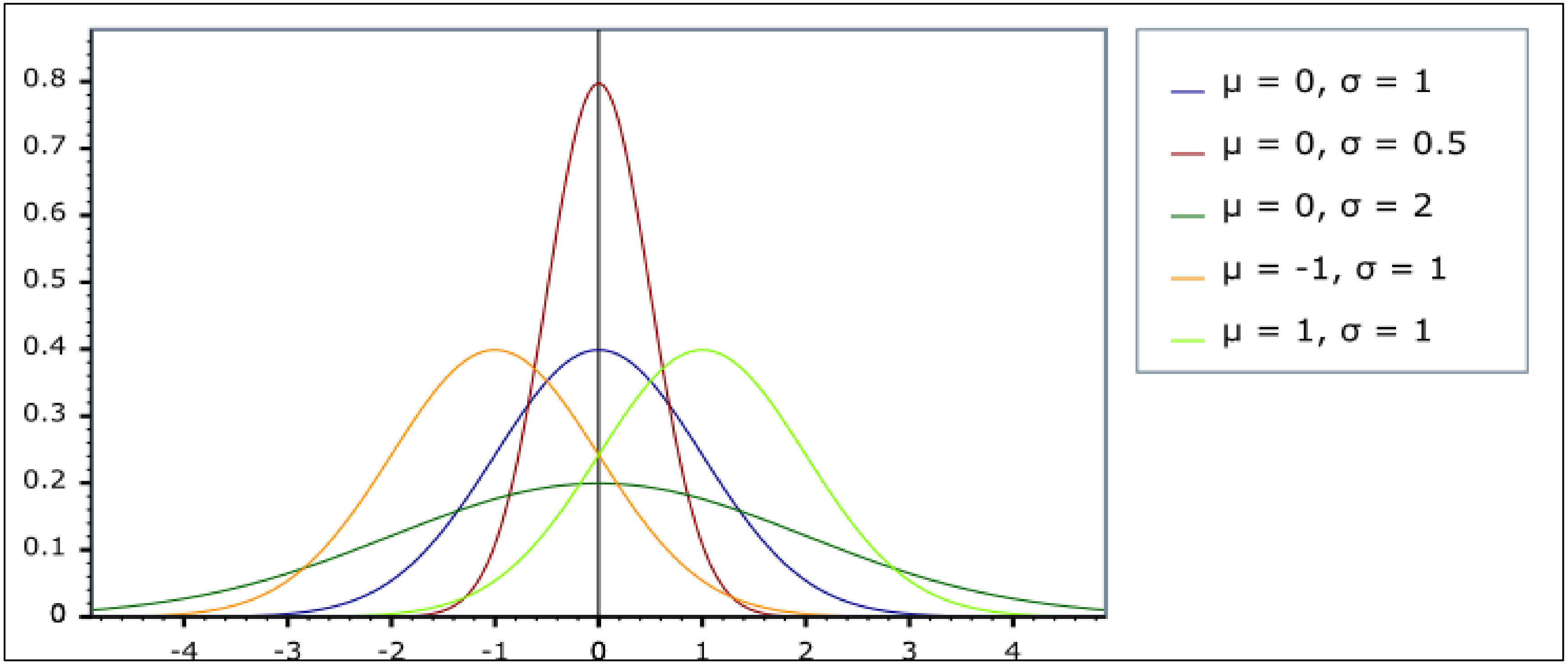
$$f(x) = \frac{1}{\sqrt{2\pi}5} e^{-\frac{1}{2}\left(\frac{x-3}{5}\right)^2}$$

- Example: A normal distribution with mean 4 and standard deviation 6

$$f(x) = \frac{1}{\sqrt{2\pi}6} e^{-\frac{1}{2}\left(\frac{x-4}{6}\right)^2}$$

Examples of Normal Distribution

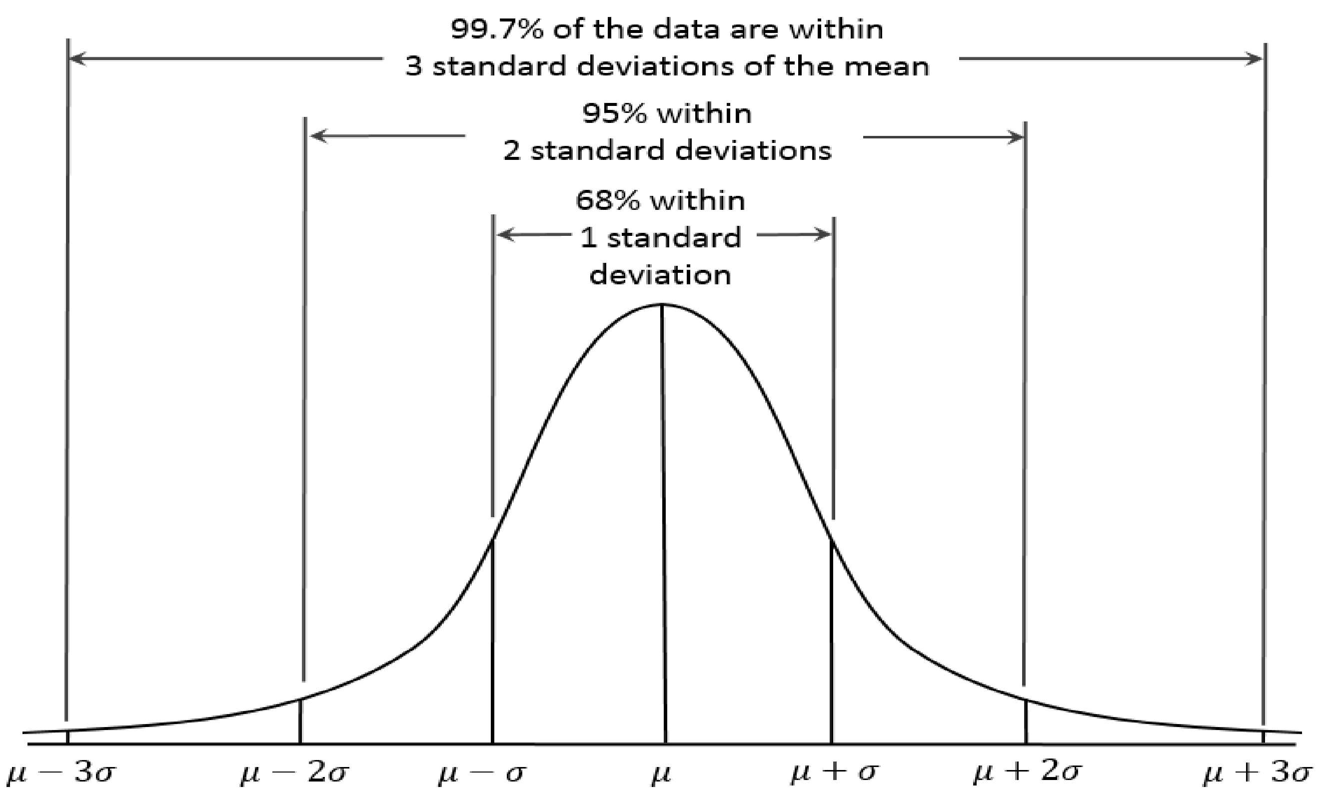
$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$



- μ determines the location of the maximum of the function
- σ determines how wide it is.

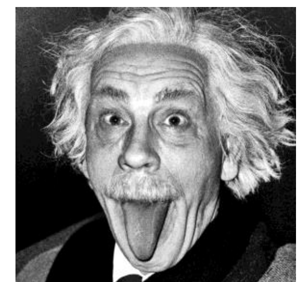
Properties of Normal Distribution

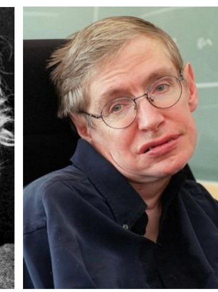
- The total area under the density curve = 1
- 68.26% of the area under the curve is between μ-σ, μ+σ,
- 95.44% of the area under the curve is between μ-2σ, μ+2σ,
- 99.72% of the area under the curve is between μ-3σ, μ+3σ.

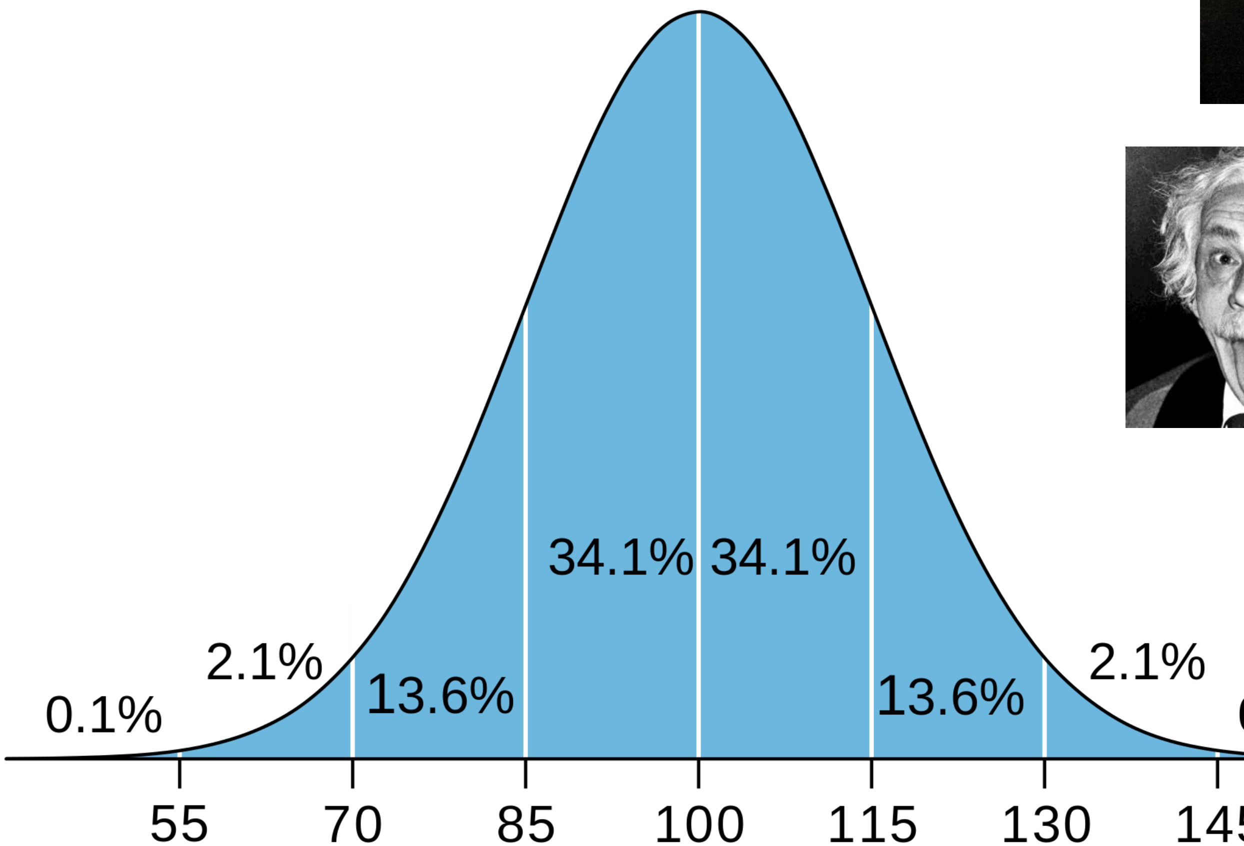


Example: IQ Distribution

- It is a Normal distribution: $f(x) = \frac{1}{\sqrt{2\pi 15}} e^{-\frac{1}{2}\left(\frac{x-100}{15}\right)^2}$







Example

- Watch

https://www.youtube.com/watch?v=dr1DynUzjq0

The Sum of Normal Distributions

- Theorem. Let $X_1, X_2, \ldots, X_n$ be normally distributed independent random variables. Then the sum $S=X_1+X_2+\ldots+X_n$ is also normally distributed with
  $E(S)=E(X_1)+E(X_2)+\ldots+E(X_n)$
  $V(S)= V(X_1)+V(X_2)+\ldots+V(X_n)$.

- Example. Let X_1, X_2, X_3, X_4 be normally distributed independent random variables, suppose
  $E(X_1)=7$, $E(X_2)=5$, $E(X_3)=3$, $E(X_4)=1$, and $V(X_1)=2$, $V(X_2)=4$, $V(X_3)=6$, $V(X_4)=8$.
  Find the distribution of the sum $S=X_1+X_2+X_3+X_4$. Report the mean and variance of S.

- Solution. As the variables are independent and normally distributed, S is also normally distributed with
  $E(S)=7+5+3+1=16$ and
  $V(S)=2+4+6+8=20$.

Linear Combination of Normal Distributions

- Theorem. Let X_1, X_2,…, X_n be normally distributed random variables, which are independent. Then their linear combination $L=a_1X_1+a_2X_2+\ldots+a_nX_n$ is also normally distributed with

  $E(L)=a_1E(X_1)+a_2E(X_2)+\ldots+a_nE(X_n)$ and

  $V(L)= a_1^2V(X_1)+a_2^2V(X_2)+\ldots+a_n^2V(X_n)$.

- Example. Let X_1, X_2, X_3 , X_4 be normally distributed random variables, which are independent. Assume that

  $E(X_1)=7$, $E(X_2)=5$, $E(X_3)=3$, $E(X_4)=1$, and $V(X_1)=2$, $V(X_2)=4$, $V(X_3)=6$, $V(X_4)=8$.

  Find the distribution of the linear combination $L=2X_1+X_2+3X_3+X_4$. Report the mean, variance, and standard deviation of L.

- Solution. As the variables are independent and normally distributed, L is also normally distributed with

  $E(L)=2\cdot7+5+3\cdot3+1=25$ and

  $V(L)=2^2\cdot2+4+3^2\cdot6+8=74$. $SD(L)=74^{0.5}=8.6023$.

Generalisation

Theorem. Let X_1, X_2,…, X_n be normally distributed random variables, which are independent. Then $Q=a_1X_1+a_2X_2+\ldots+ a_nX_n+b$ is also normally distributed with:

$$E(Q)=a_1E(X_1)+a_2E(X_2)+\ldots+a_nE(X_n)+b$$

$$V(Q)= a_1^2V(X_1)+a_2^2V(X_2)+\ldots+a_n^2V(X_n).$$

Example

A cost accountant needs to forecast the unit cost of a product for next year.

- Each unit of the product requires 12 hours of labour and 5.8 pounds of raw material.
- Each unit of the product is assigned an overhead cost of $184.50.
- The cost of an hour of labour next year will be normally distributed with an expected value of $45.75 and a standard deviation of $1.80.
- The cost of the raw material per pound will be normally distributed with an expected value of $62.35 and a standard deviation of $2.52.

Find the distribution of the unit cost of the product. Report its expected value, variance, and standard deviation.

Example

Solution. Denote:
L: the unit cost of labour
M: the unit cost of the raw material
Q: unit cost of the product

Then: $Q = 12L + 5.8M + 184.50$.

If L does not influence the M, we can assume that the two are independent. Therefore Q is a normal random variable and:

$E(Q) = 12 * 45.75 + 5.8 * 62.35 + 184.5 = \$1095.13$
$V(Q) = 12^2 * 1.80^2 + 5.8^2 * 2.52^2 = 680.19$
$SD(Q) = 680.19^{0.5} = \$26.08$.

Standard Normal Distribution

- Definition. The standard normal random variable Z is the normal random variable with mean 0 and standard deviation1. That is:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

- Notation. Z~N(0,1).

Calculation of Probabilities

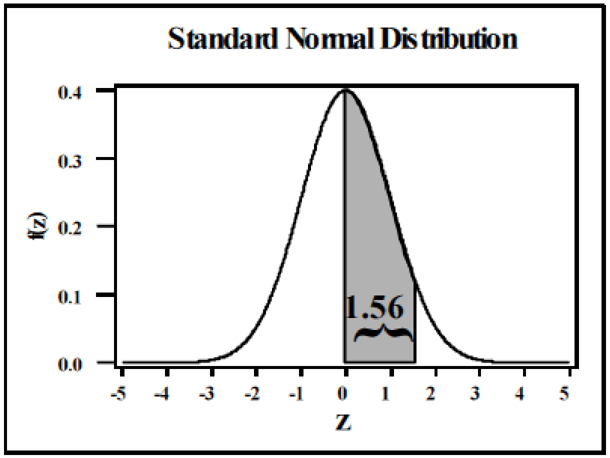
- Example. Assume that the standardized daily price change of an asset is normally distributed with mean 0 and sd 1. What is the probability that the daily price change is between 0 and 1.56?

- Solution. we are interested in $P(0<Z<1.56)$. That is, the area marked in grey.

- In Aczel, page 152 there is a table which gives the area between 0 and each value of z

  (also available in Moodle)

Using Standard Normal Distribution Table



Standard Normal Probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

Look in row labeled 1.5 and column labeled .06 to find $P(0 \leq Z \leq 1.56) =$ 0.4406

Exercises: using the table

- Exercise 1. What is P(0<Z<0.67)?

  P(0<Z<0.67)=0.2486.

- Exercise 2. What is P(0.56<Z<0.93)?

  P(0.56<Z<0.93)

  =P(0<Z<0.93)-P(0<Z<0.56)

  =0.3238-0.2123=0.1115.

- Exercise 3. Calculate P(0.24<Z<0.33).

  P(0.24<Z<0.33)

  =0.1293-0.0948=0.0345.

FIGURE 4–4 The Table Area TA for a Point *z* of the Standard Normal Distribution

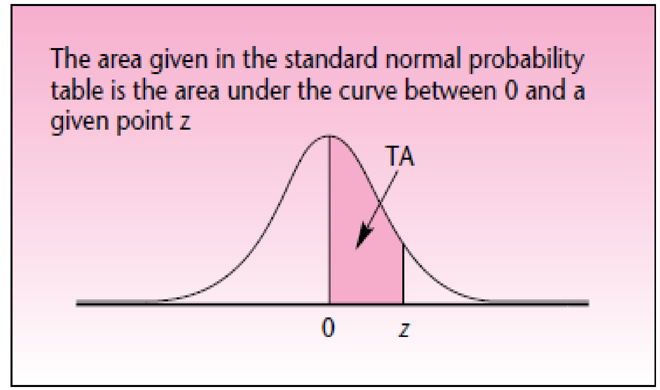


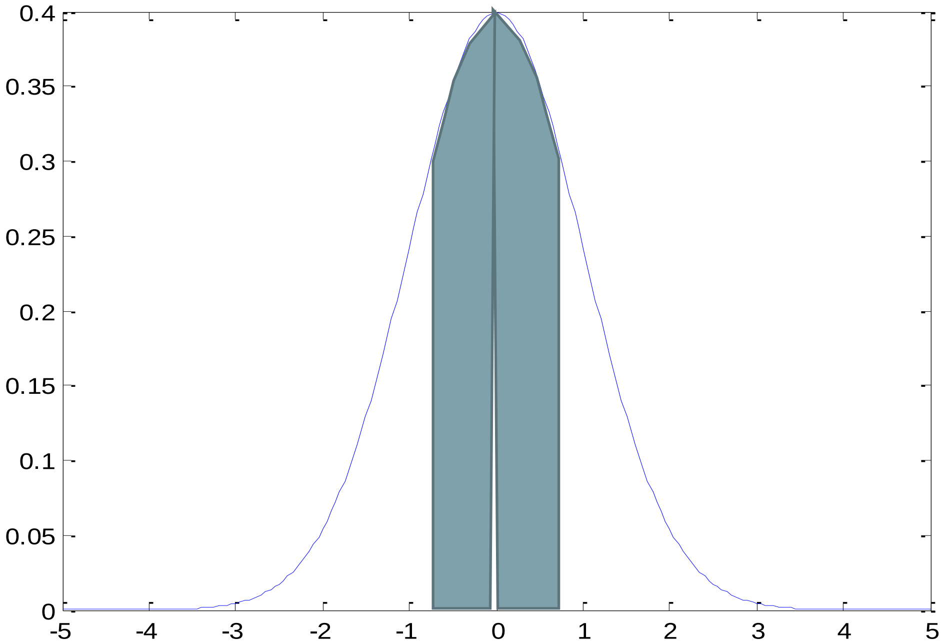
TABLE 4–1 Standard Normal Probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389

Symmetry of Normal Distribution:

- Example:

  P(-0.82<Z<0.82)=2*P(0<Z<0.82)= 2*0.2939=0.5878



Tricks

- If $a>0$, $P(0<Z<a)$ is the value written in the table.
- If $a>0$, $P(-a<Z<0)=P(0<Z<a)$ (symmetry)
- $P(Z<0)=P(-\infty<Z<0)= P(0<Z< \infty)=P(Z>0)=0.5$
- If $a>0$, $P(Z<a)=P(-\infty<Z<0)+P(0<Z<a)=0.5+P(0<Z<a)$.
- $P(Z>a)=1-P(Z<a)$

Exercises

1. P(Z<0.66)

=0.2454+0.5=0.7454

2. P(Z<-0.66)

=P(Z>0.66) (symmetry)

=1-P(Z<0.66)

=1-0.7454=0.2546

3. P(-0.23<Z<0.34)

=P(-0.23<Z<0)+P(0<Z<0.34)

=P(0<Z<0.23)+ P(0<Z<0.34)

=0.0910+0.1331=0.2241.

4. Find z such that P(0<Z<z) is 0.27.

The nearest value is z=0.74.

FIGURE 4–4 The Table Area TA for a Point *z* of the Standard Normal Distribution

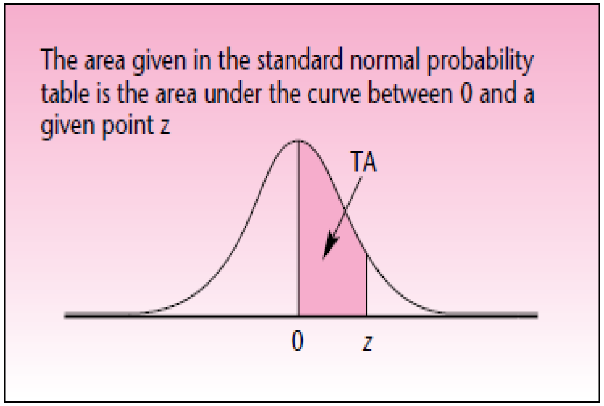


TABLE 4–1 Standard Normal Probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389

Caution: Different Tables

Normal Deviate z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-4.0	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
-3.9	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
-3.8	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
-3.7	.0001	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
-3.6	.0002	.0002	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
-3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
-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
-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
-2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
-2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
-2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
-2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
-2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
-2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
-2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
-1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
-1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
-1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
-1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
-1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
-1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
-1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
-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
-.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
-.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
-.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
-.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
-.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
-.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
-.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483

FIGURE 4–4 The Table Area TA for a Point *z* of the Standard Normal Distribution

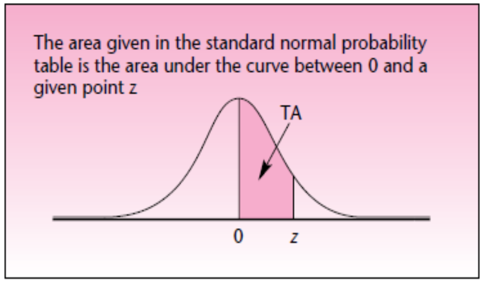


TABLE 4–1 Standard Normal Probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389

Why Standard Normal Distribution?

- In order to calculate probabilities of the type $P(X<x)$ or $P(a \leq X \leq b)$, we need to calculate the area under part of a normal distribution of X

- In general, it is difficult for a "normal" Normal distribution

- However, for the standard normal distribution Z, there are tables that help us calculate the probabilities.

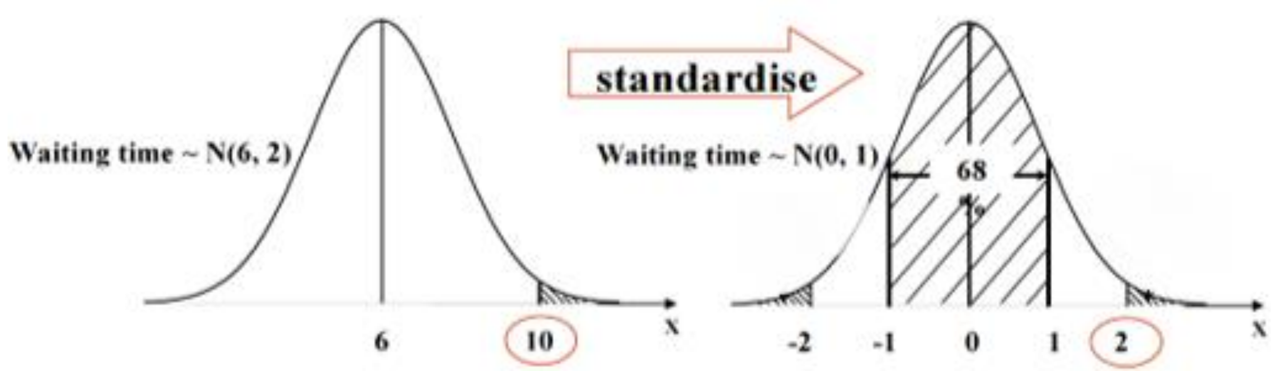
- To calculate the area under the relevant normal distribution X, we will transform the question from X to Z and use the result of the table for Z.

How to transform the question from X to Z? Namely, how to standardize normal distributions?

Standardization of Normal Distribution

$$X = \mu + \sigma Z, \quad Z = (X - \mu)/\sigma$$

Example. Suppose the waiting time for customer service is normally distributed with a mean of 6 min. and standard deviation of 2 min. What is the probability that a customer will wait more than 10 minutes?



$$P(Time>10)=P(Z>(10-6/2)=P(Z>2)=2.3\%$$

Example

- Let X be the amount of money spent by customers in a shop. Assume that $X \sim N(50, 25^2)$. What is the probability that a customer spends more than £100 in the shop?

- Solution.

$$P(X > 100) = P\left(Z > \frac{100 - \mu}{\sigma}\right)$$

$$= P\left(Z > \frac{100 - 50}{25}\right) = P(Z > 2)$$

$$= 1 - P(Z < 2)$$

$$= 1 - [P(-\infty < Z < 0) + P(0 < Z < 2)] = [1 - (0.5 + 0.4772)] = 0.0228.$$

Exercise

The number of shares traded daily on the New York Stock Exchange (NYSE) is referred to as the *volume of* trading. Assume that the number of shares traded on the NYSE is a normally distributed random variable, with a mean of 1.8 billion and a standard deviation of 0.15 billion.

For a randomly selected day, what is the probability that the volume is below 1.35 billion?

Solution.

X~N(1.8,0.15²).

$$P(X < 1.35) = P\left(Z < \frac{1.35 - \mu}{\sigma}\right) = P\left(Z < \frac{1.35 - 1.8}{0.15}\right)$$

$$P(Z < -3) = 0.5 - P(0 < Z < 3) = 0.5 - 0.4987 = 0.0013.$$

The Inverse Approach: example

- The length of phone calls in a company is normally distributed, with the mean 50 seconds and the standard deviation 12 seconds. What is the duration of a call which 80% of all calls are longer?

Solution. We have $X \sim N(50,12^2)$. First, we find z such that $P(Z>z)=0.8$, or, $P(Z<z)=0.2$

z is negative

$\Rightarrow P(Z<z)=0.5-P(z<Z<0)$

$=0.5-P(0<Z<-z)=0.2$

$\Rightarrow P(0<Z<-z)=0.5-0.2=0.3$

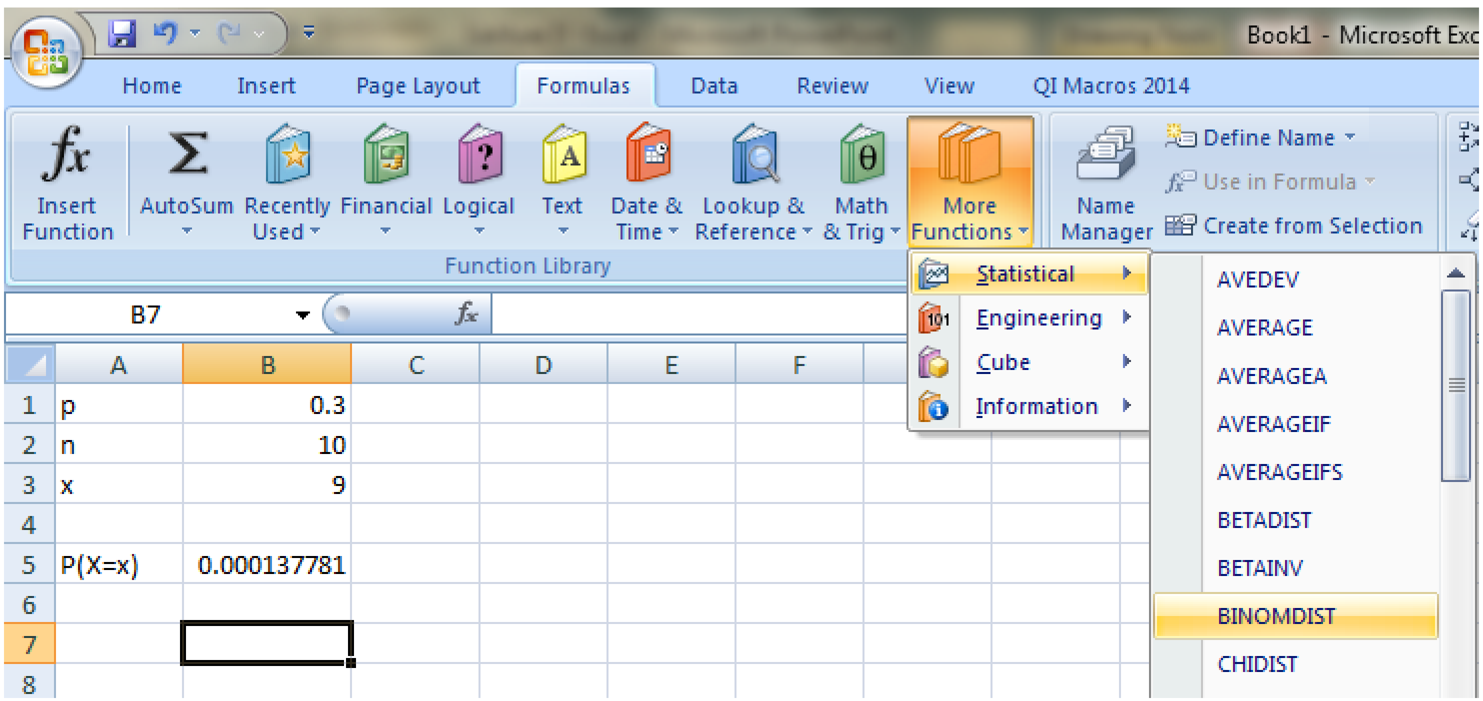
$\Rightarrow -z=0.84,\ z=-0.84$

$\Rightarrow X=\mu+\sigma z=50-12 \cdot 0.84=39.92$[sec]

Excel: Calculation of Binomial Distribution

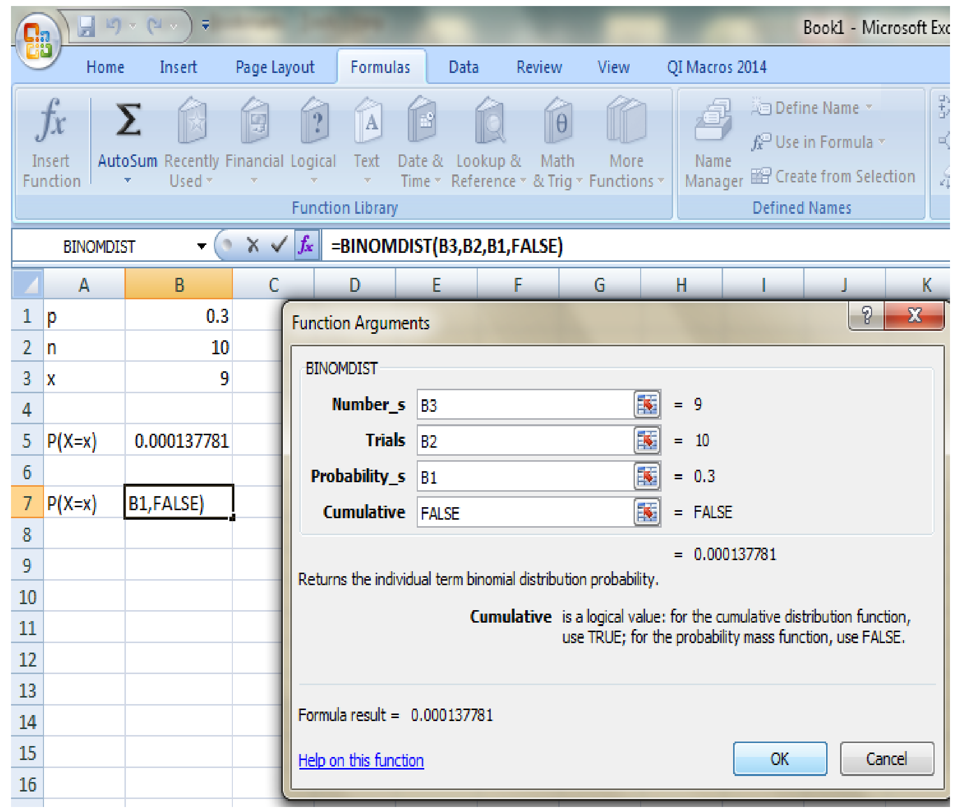
Excel Function: BinomDist

- To get to this function go to Formulas->More functions->Statistical->BinomDist
- We use X~B(10, 0.3) as an example.



Excel Function: BinomDist

- Number_s=number of successes (x)
- Trials=n
- Probability_s=probability of success
- Cumulative
  - FALSE for probability
  - TRUE for the cumulative probability

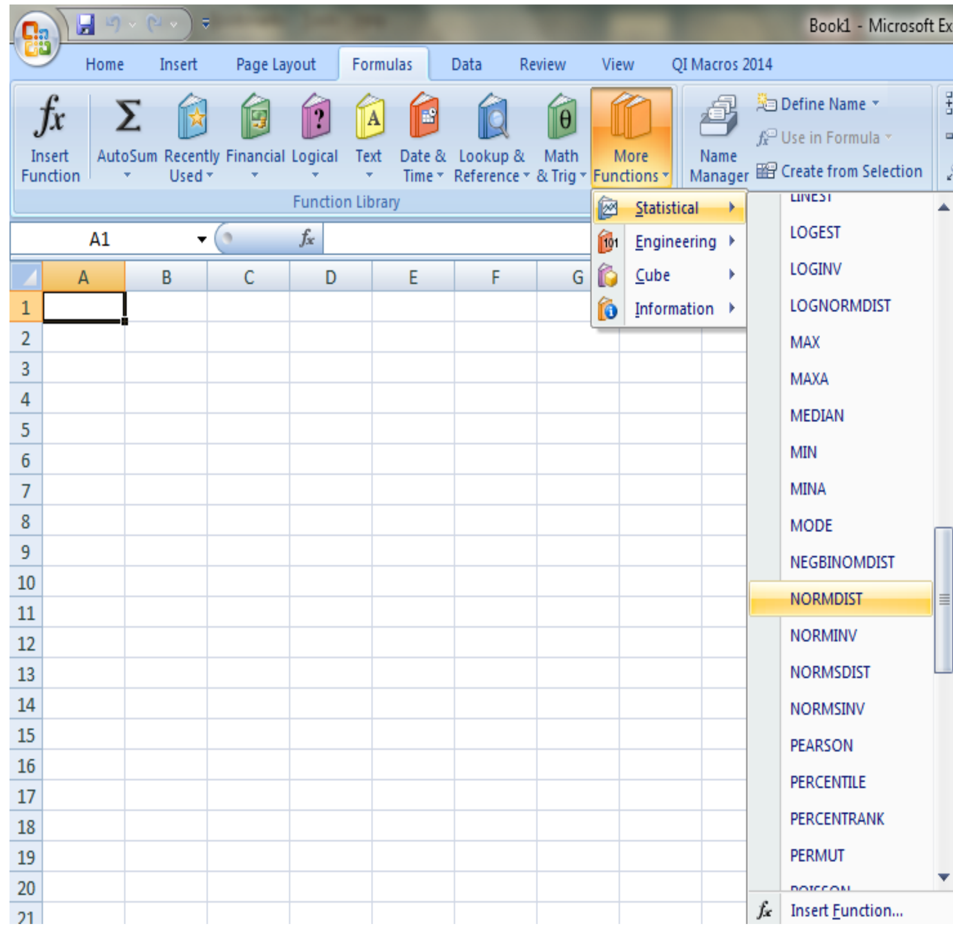


Excel: Calculation of Normal Distribution

Excel: Normal Distribution

- Choose: Formulas
  ->More functions
  ->Statistical

- Then choose NORM.DIST
  NORM.DIST(x,mean,std,true) calculates
  P(X<x)=P(-∞<X<x)
  for X~N(mean,std).

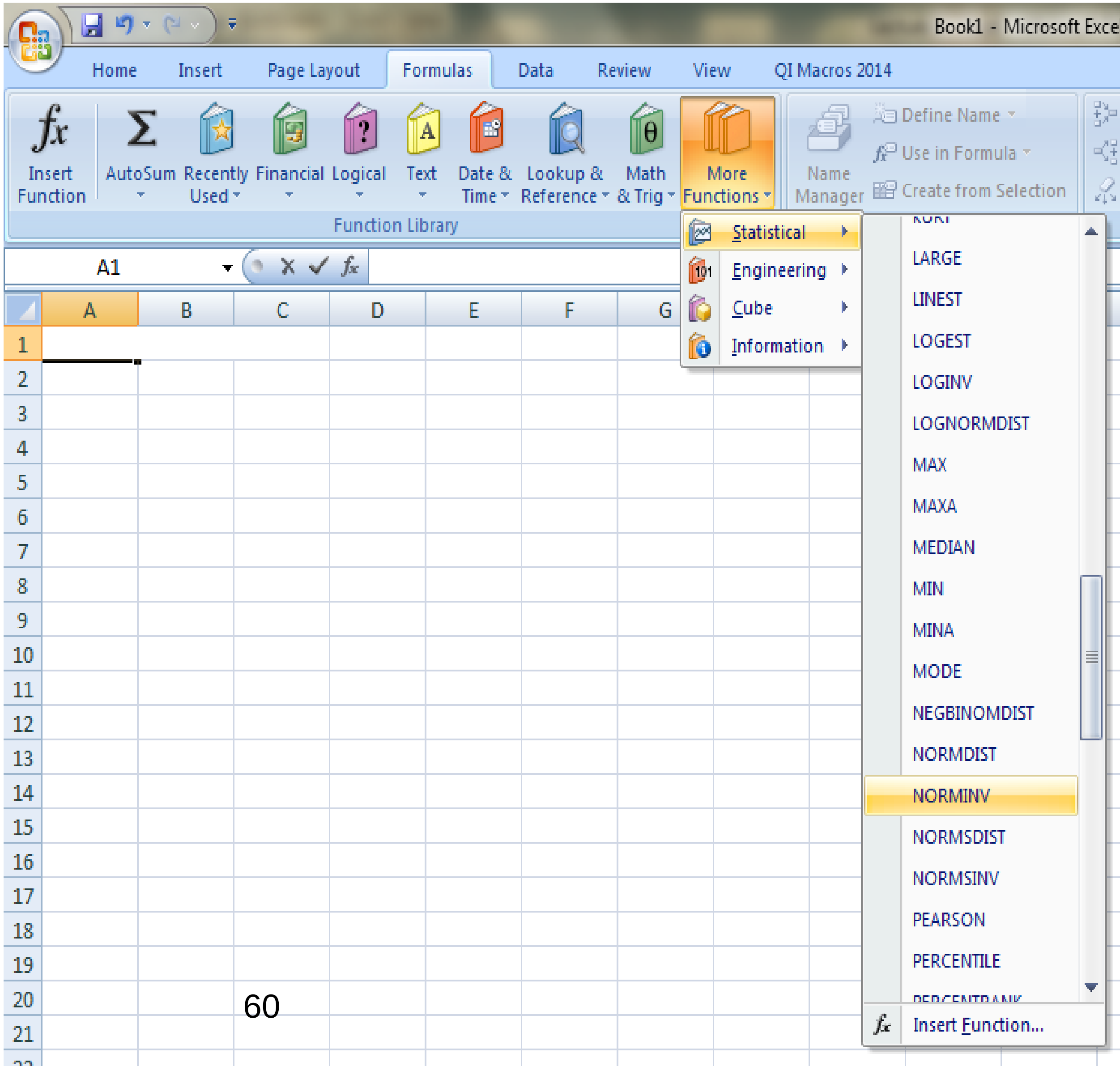


Inverse Approach

- Find the z value such that P(Z<z)=0.7454.
- Solution. Choose: Formulas

  ->More functions

  ->Statistical
- Then choose NORM.INV



Excel Exercise

- Exercise. A management graduate is applying for nine jobs, and believes that she has in each of the nine cases a constant and independent 0.48 probability of getting an offer. What is the probability that she will have at least three offers?

- Solution in Excel.

Excel Exercise

In the final exam of a course, suppose for each student, the probability of passing the final exam is 0.9, and we have 100 students in the class, uses Excel to calculate the following:

- the probability of exactly 80 students pass the final exam
- the probability of all students pass the final exam
- the probability of at most 20 students fail the exam
- generate the probability distribution function and the cumulative distribution function, and draw a picture containing the two functions

Seminar 3

Random Variable and Normal Distribution

Exercise 1

- If you flip a coin 3 times, what is the probability that you will get Heads on at least one of the 3 flips?

Exercise: Random Variables

- The US Census Bureau (http://www.census.gov/econ/susb/) published data about the number of firms and their employees in the US.

Mining, quarrying, and oil and gas extraction		Finance and insurance	
Employment size	Number of firms	Employment size	Number of firms
01: Total	21,408	01: Total	233,563
02: 0-4	12,411	02: 0-4	168,579
03: 5-9	3,206	03: 5-9	32,386
04: 10-14	1,443	04: 10-14	9,240
05: 15-19	897	05: 15-19	4,506
06: <20	17,957	06: <20	214,711
07: 20-24	548	07: 20-24	2,913
08: 25-29	404	08: 25-29	1,938
09: 30-34	293	09: 30-34	1,641
10: 35-39	218	10: 35-39	1,226
11: 40-44	189	11: 40-44	955
12: 45-49	164	12: 45-49	814
13: 50-74	453	13: 50-74	2,546
14: 75-99	203	14: 75-99	1,443
15: 100-149	244	15: 100-149	1,462
16: 150-199	127	16: 150-199	832
17: 200-299	150	17: 200-299	857
18: 300-399	75	18: 300-399	417
19: 400-499	35	19: 400-499	256
20: <500	21,060	20: <500	232,011
21: 500 +	348	21: 500 +	1,552

Exercise 2: For firms of less than 20 workers, what is the probability distribution of firm size in the mining, quarrying and oil and gas (MQOG) extraction is the US?

4	02: 0-4	12,411
5	03: 5-9	3,206
6	04: 10-14	1,443
7	05: 15-19	897

Exercise 3: We refer now only to MQOG firms with less than 20 workers.

A	B	C
Mining, quarrying, and oil and gas extraction		
Employment size	Number of firms	Probability distribution of firm size, for firms with <20 workers
02: 0-4	12,411	0.691151083
03: 5-9	3,206	0.178537618
04: 10-14	1,443	0.080358635
05: 15-19	897	0.049952665

1. What is the probability that the number of employees in an MQOG firm is between 10 and 20?

2. Calculate the cumulative distribution function.

Exercise 4

- Assume that in a survey about the number of cell phone each participant had, the following results were obtained

Number of cell phone	Percentage of participants
1	0.6
2	0.2
3	0.1
4	0.07
5	0.03

1. What is the expected number of cell phones per person?
2. What is the variance and standard deviation of the number of cell phones per person?

Binomial Distribution: Exercise 5

A commercial jet aircraft has five engines. For an aircraft in flight to land safely, at least two engines should be in working condition. Each engine has an independent reliability of p=90%.

a. What is the probability that 2 of the engines are in working condition?

b. What is the probability that an aircraft in flight can land safely?

Exercises 6: Binomial Distribution Table

Here are some table lookup exercises — find the following Binomial probabilities:

a) For Bin~(n = 8, p = 0.2), calculate $P(X = 4)$
b) For Bin~(n = 15, p = 0.6), calculate $P(X \geq 6)$
c) For Bin~(n = 10, p = 0.3), calculate calculate $P(X < 5)$
d) For Bin~(n = 10, p = 0.3), calculate $P(3 < X \leq 7)$

Exercise 7: Uniform Distribution

Exercise. The time a customer spends on the web site of a certain online shop is uniformly distribution between 20 minutes and 40minutes.

a. What is the probability that the spent time will be between 30 and 45 minutes?

b. What is the probability that the spent time will be less than 25 minutes?

c. What are the mean and standard deviation of the spent time?

Exercise 8: Normal Distribution

- The price of one of the parts of the main component of SeaStar250 is normally distributed with mean £20 and variance 4.

  The price of the second part of the same component is normally distributed with mean £30 and variance 5.

  Assume that the prices of the two parts are independent. What is the probability that the price of this component is between £52 and £55?

Thank You!

Any Questions?