

TRAINING COURSES ON APPLIED ECONOMETRIC ANALYSIS
(SUMMER SCHOOL) FOR YOUNG ECONOMISTS /
RESEARCHERS ORGANIZED BY WIUT AND IFPRI
JUNE 4-15, 2018

Review of Probability

Farrukh Ataev
Lecturer, WIUT
6 June 2018

Outline

- **Session 1 Review of probability**
 - (Tue 5 June at 13:30-15:00)
- **Session 2 Review of probability (continued)**
 - (Tue 5 June at 15:30-17:00)
- **Session 3 Probability distributions**
 - (Wed 6 June at 9:00-10:30)
- **Session 4 Probability distributions (continued)**
 - (Wed 6 June at 11:00-12:30)

Session 3 outline

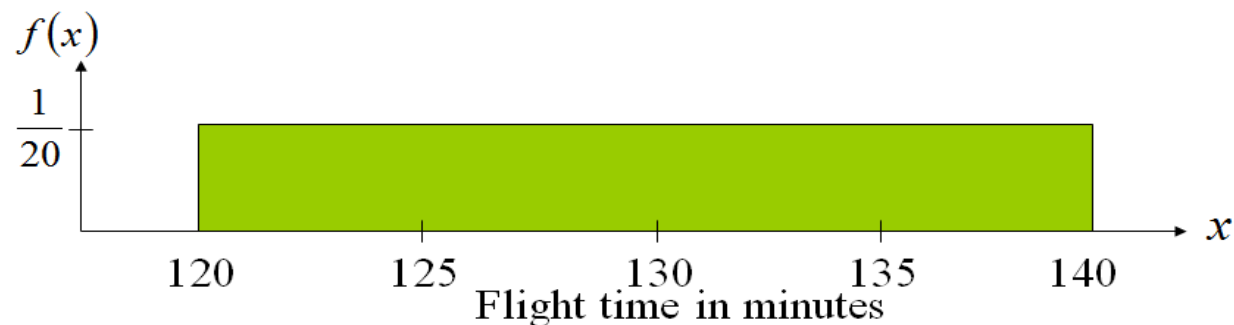
- ❖ Probability distribution (discrete and continuous)
- ❖ Discrete probability distribution
 - ❖ Binomial, Poisson
- ❖ Continuous probability distribution
 - ❖ Uniform, Normal (Standard normal), F, Chi-squared

Continuous: Uniform probability distribution

➤ It is a continuous probability distribution associated with situations in which all values of the random variable are equally likely.

➤ Continuous uniform probability function:
$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{for } a \leq x \leq b \\ 0, & \text{elsewhere} \end{cases}$$

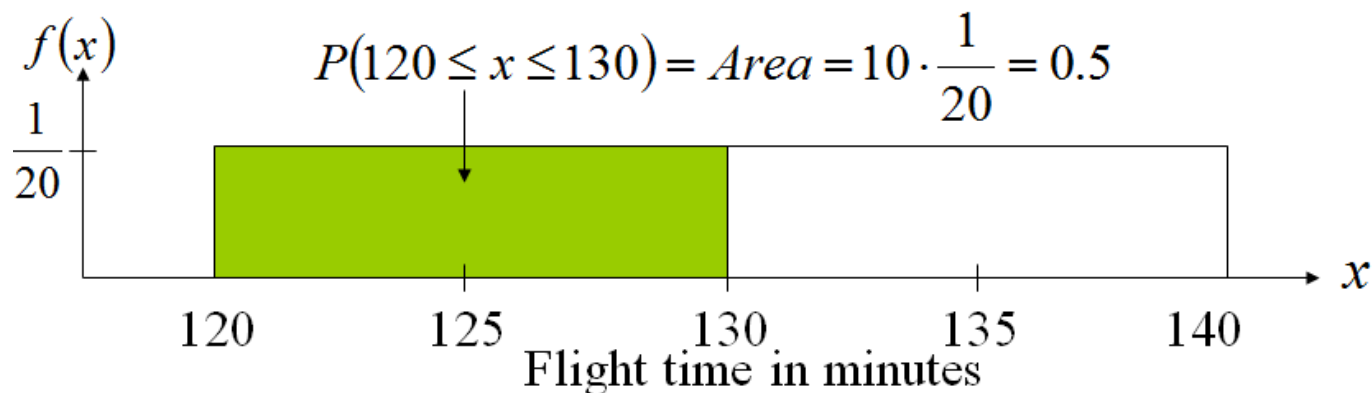
Example: Consider the random variable x that represents the total flight time of an airplane traveling from Chicago to New York. Suppose that the flight time can be any value in the interval from 120 minutes to 140 minutes.



Probability versus probability function

The probability ($P(c \leq x \leq d)$) of the random variable taking on a value within a given interval from c to d is defined to be the area under the graph of the probability function between c and d . It means the probability of the random variable taking on a particular value is zero, since the area under the graph $f(x)$ at a single point is zero.

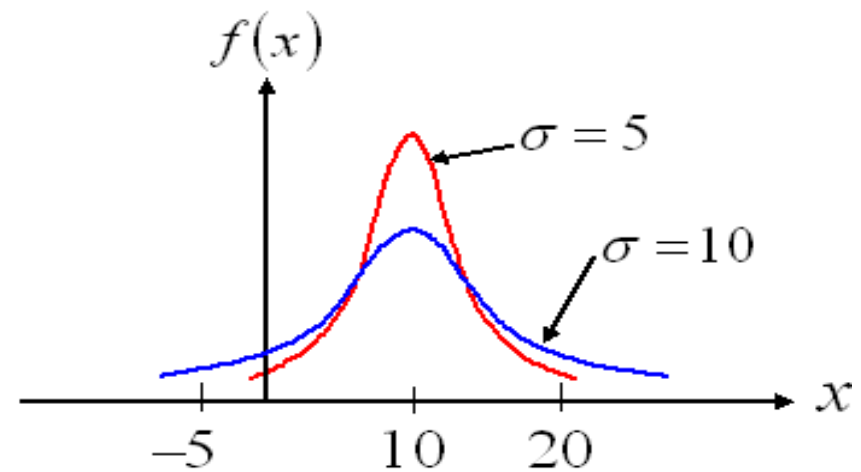
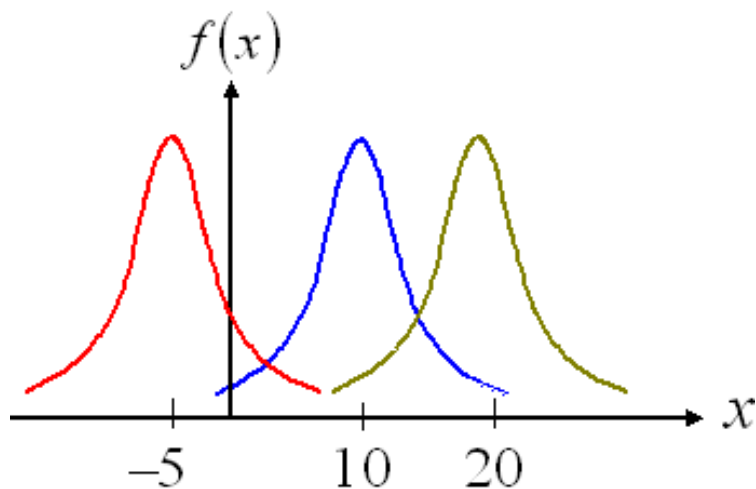
Example (cont.): The probability of a flight time between 120 and 130 is:



Exercise: Find the probability of a flight time of exactly 130 minutes.

Continuous: Normal probability distribution $N(\mu, \sigma)$

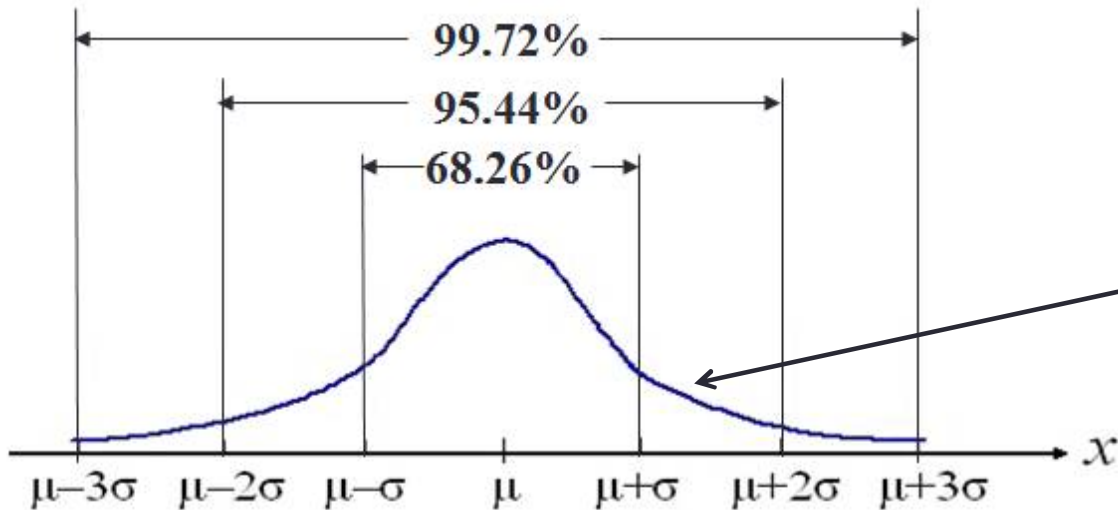
Continuous probability distribution with symmetrical, bell-shaped histogram defined by mean (μ) and standard deviation (σ)



Question: If all normal distributions have the same shape, how do they differ?

Probability of the normal random variable

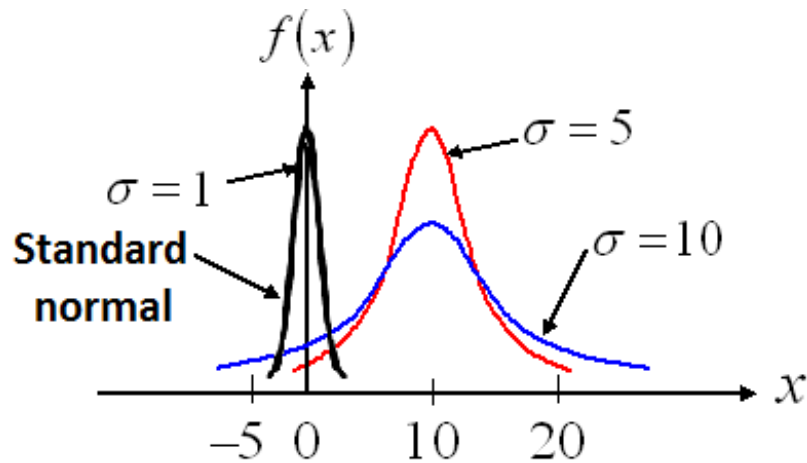
- Probabilities of normal random variable are given by areas under the normal curve.
- Some commonly used intervals are:
 - 1) $P(\mu - \sigma \leq x \leq \mu + \sigma) = 68.26\%$
 - 2) $P(\mu - 2\sigma \leq x \leq \mu + 2\sigma) = 95.44\%$
 - 3) $P(\mu - 3\sigma \leq x \leq \mu + 3\sigma) = 99.72\%$
- The total area under the normal curve is 1.



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

Standard normal probability distribution $N(\mu, \sigma) = N(0,1)$

- A normal distribution with a mean of 0 and a standard deviation of 1 is called a standard normal distribution.
- The letter z is commonly used to designate the normal random variable.
- Standard normal probability function: $f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$



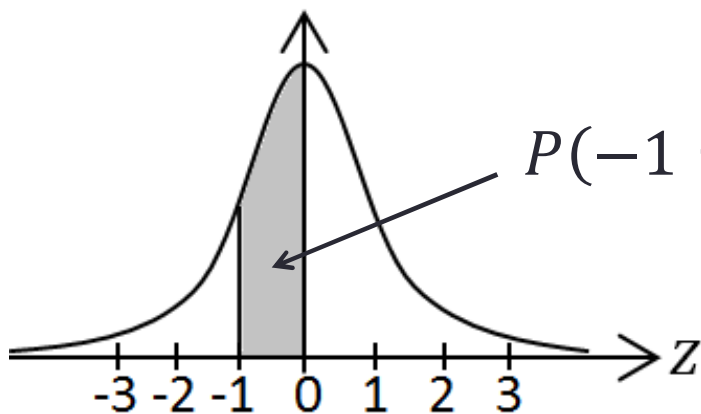
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \Rightarrow f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

Example

What is the probability that the z value for the standard normal random variable will be between -1 and 0? That is, $P(-1 \leq z \leq 0) = ?$

1-method: Using the definite integration formula:

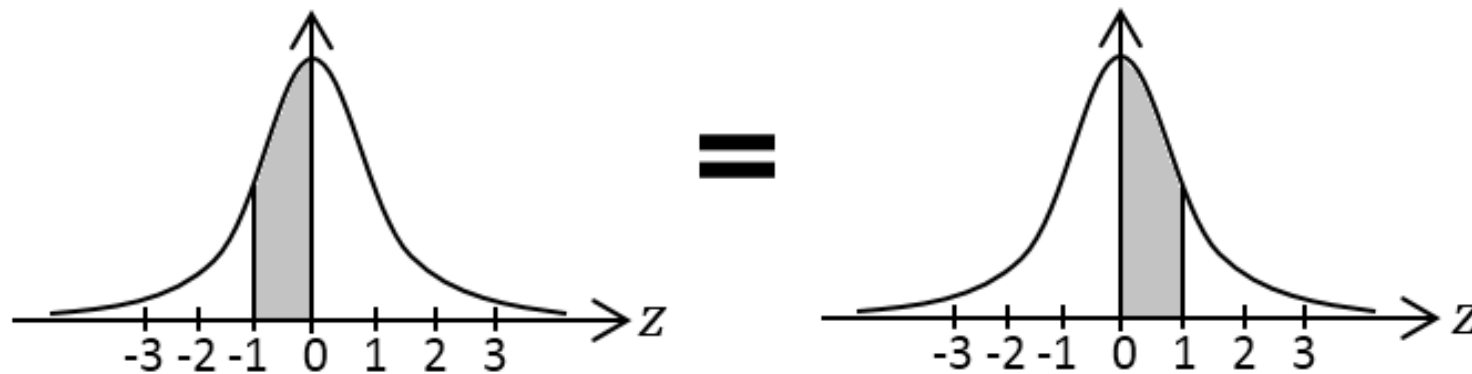
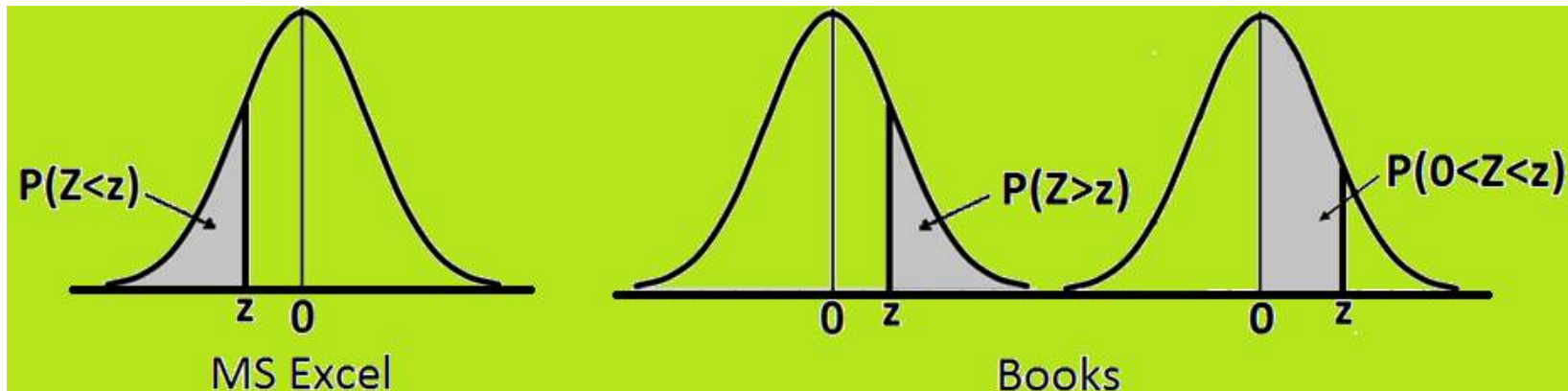
$$P(a \leq z \leq b) = \int_a^b f(z) dz = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$



$$P(-1 \leq z \leq 0) = \int_{-1}^0 f(z) dz = \int_{-1}^0 \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = 0.3413$$

Example (continued)

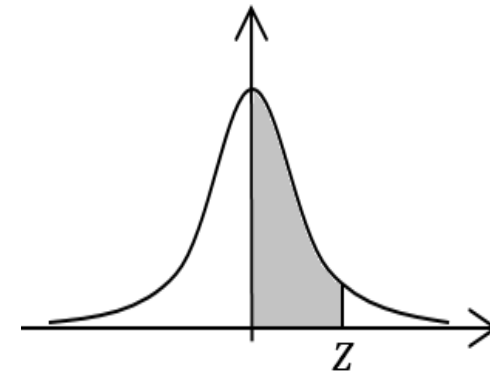
2-method: Using the Standard normal distribution table (see next 2 slides)



$$P(-1 \leq z \leq 0) = P(0 \leq z \leq 1) = 0.3413$$

Standard normal distribution table

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0	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	0.3304	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

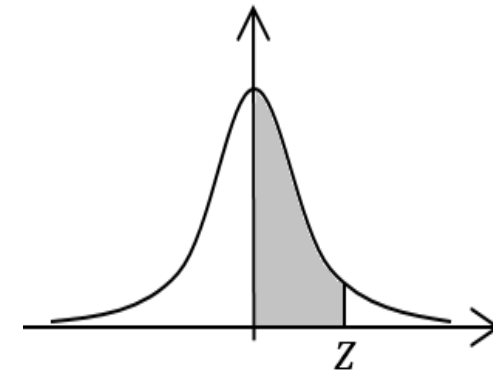


The entries in this table show the area under the standard normal curve between 0 and z .

$$P(0 \leq z \leq 1.64) = 0.4495; P(0 \leq z \leq 1.645) = \frac{0.4495 + 0.4505}{2} = 0.45; z_{0.05} = 1.645.$$

Standard normal distribution table (continued)

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
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	0.483	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
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.5	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998
3.6	0.4998	0.4998	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.7	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.8	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999



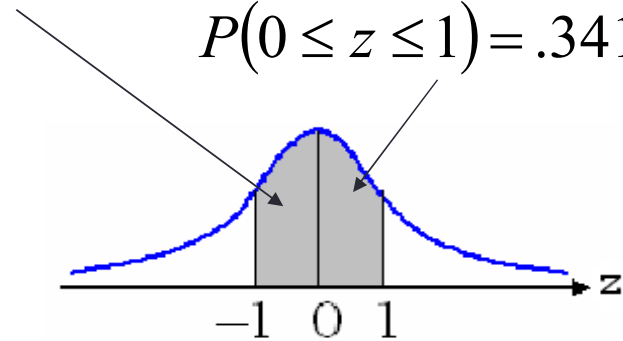
The entries in this table give the area under the standard normal curve to the left of z .

$$P(0 \leq z \leq 3.1) = \mathbf{0.4990}$$

Examples of using the Standard normal table

$$P(-1 \leq z \leq 0) = .3413$$

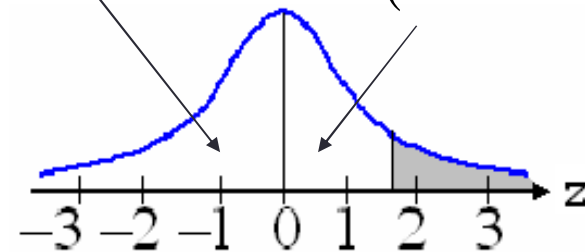
$$P(0 \leq z \leq 1) = .3413$$



$$P(-1 \leq z \leq 1) = .6826$$

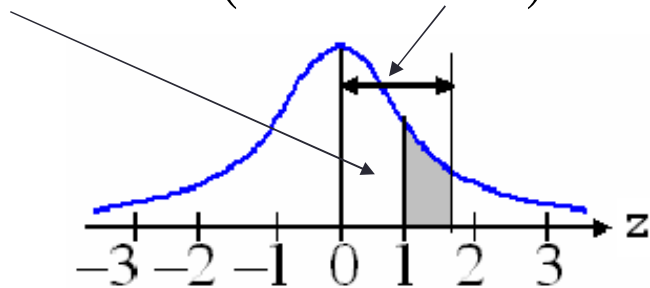
$$P(z \leq 0) = .5$$

$$P(0 \leq z \leq 1.58) = .4429$$



$$P(z \geq 1.58) = .5 - .4429 = .0571$$

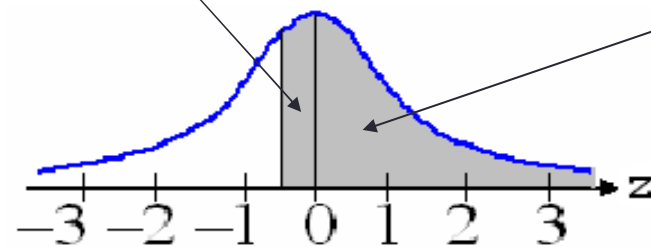
$$P(0 \leq z \leq 1) = .3413 \quad P(0 \leq z \leq 1.58) = .4429$$



$$P(1 \leq z \leq 1.58) = .4429 - .3414 = .1016$$

$$P(0 \leq z \leq -0.5) = .1915$$

$$P(z \geq 0) = .5$$



$$P(z \geq -0.5) = .5 + .1915 = .6915$$

Exercise

Find from the z -table:

1) $P(z < 0) = P(z \leq 0)$

2) $P(z < 1.645)$

3) $P(1.23 < z < 2.93)$

4) $P(-1.21 < z < 1.85)$

5) $P(z > 2)$

6) $P(z < -2.34)$

7) $P(z = 3)$

8) $z_{0.0505}$

9) $z_{0.0495}$

10) $z_{0.05}$

Example (continued)

3-method: MS Excel function:

NORM.DIST($X, \mu, \sigma, 1$) or NORM.S.DIST($Z, 1$)

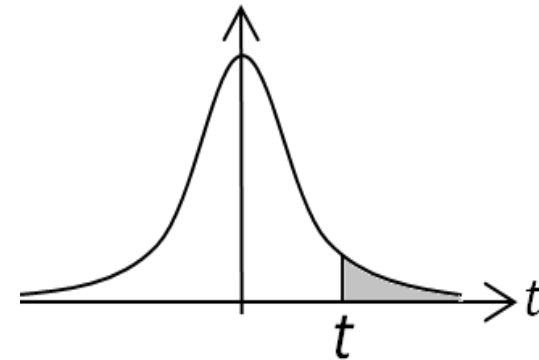
$$\begin{aligned} \text{NORM.DIST}(0, 0, 1, 1) - \text{NORM.DIST}(-1, 0, 1, 1) &= \\ 0.5 - 0.1587 &= 0.3413 \end{aligned}$$

Or

$$\begin{aligned} \text{NORM.S.DIST}(0, 1) - \text{NORM.S.DIST}(-1, 1) &= \\ 0.5 - 0.1587 &= 0.3413 \end{aligned}$$

t-distribution table

df	0.25	0.2	0.15	0.1	0.05	0.025	0.02	0.01	0.005
1	1.000	1.376	1.963	3.078	6.31	12.70	15.90	31.82	63.65
2	0.817	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925
3	0.765	0.979	1.250	1.638	2.353	3.182	3.482	4.541	5.841
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106
12	0.696	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947



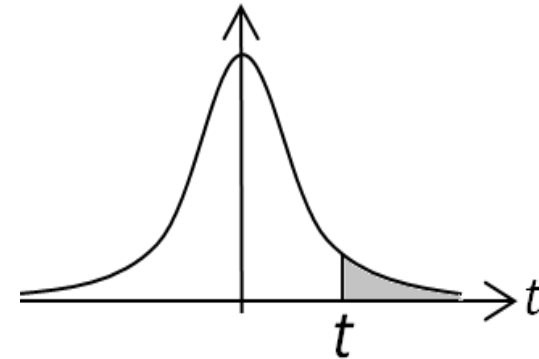
The entries in this table give t_α values, where α is the area (or probability) in the upper tail.

$$t_{0.05;2} = 2.92$$

$$P(t_{(2)} > 2.92) = 0.05$$

t-distribution table (continued)

df	0.25	0.2	0.15	0.1	0.05	0.025	0.02	0.01	0.005
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750



The entries in this table give t_α values, where α is the area (or probability) in the upper tail.

$$t_{0.05;17} = 1.74$$

$$P(t_{(17)} > 1.74) = 0.05$$

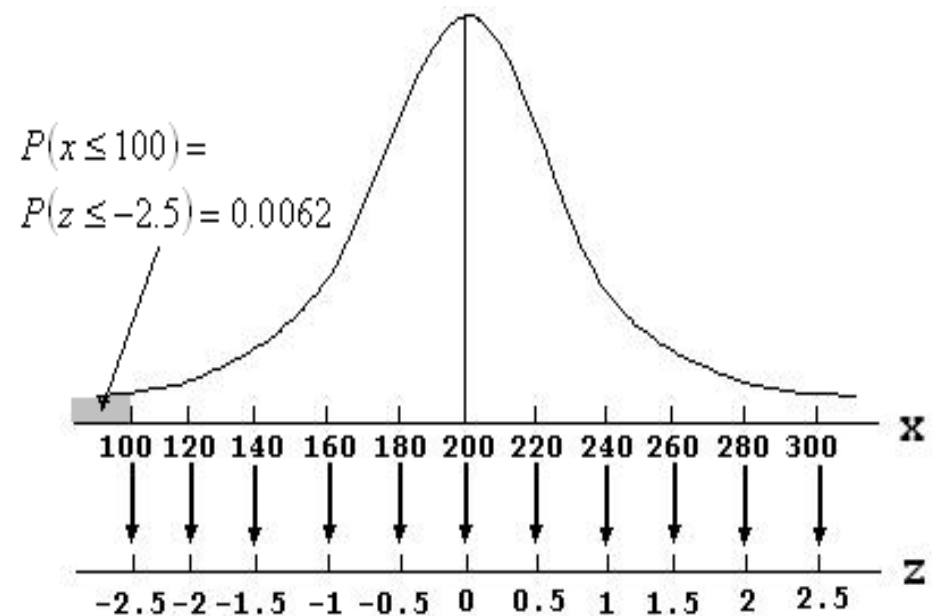
Converting Normal to Standard normal distribution

The formula used to convert any normal random variable x with mean μ and standard deviation σ to the standard normal random variable z is: $z = \frac{x-\mu}{\sigma}$.

Example: The demand for a new product is assumed to be normally distributed with $\mu = 200$ and $\sigma = 40$. Let x be the number of units demanded. Find

- a) $P(x \leq 100)$;
- b) $P(180 \leq x \leq 220)$;
- c) $P(x \geq 250)$.

$$\begin{aligned} \text{a) } P(x \leq 100) &= P\left(z \leq \frac{100-200}{40}\right) = \\ &= P(z \leq -2.5) = 0.0062 \end{aligned}$$

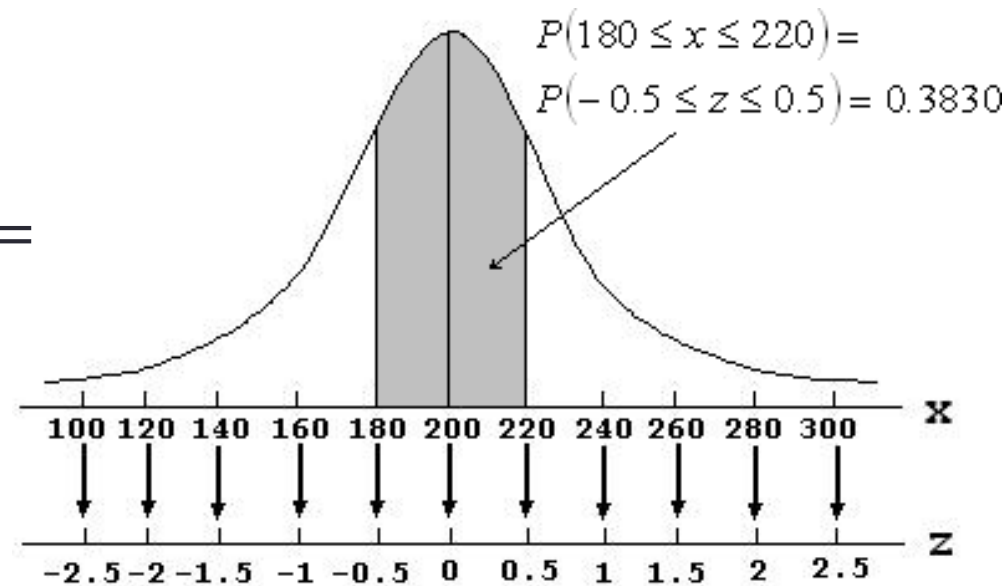


Converting Normal to Standard normal distribution (continued)

Example: The demand for a new product is assumed to be normally distributed with $\mu = 200$ and $\sigma = 40$. Let x be the number of units demanded. Find

a) $P(x \leq 100)$; b) $P(180 \leq x \leq 220)$; c) $P(x \geq 250)$.

$$\begin{aligned} \text{b) } P(180 \leq x \leq 220) &= \\ &= P\left(\frac{180 - 200}{40} \leq z \leq \frac{220 - 200}{40}\right) = \\ &= P(-0.5 \leq z \leq 0.5) = 0.3830. \end{aligned}$$



$$\text{c) } P(x \geq 250) = P\left(z \geq \frac{250 - 200}{40}\right) = P(z \geq 1.25) = 0.1056.$$

Exercise

From past experience, the management of a well-known fast-food restaurant estimates that the number of weekly customers at a particular location is normally distributed, with a mean of 5000 and a standard deviation of 800 customers.

- a) Find the probability that on a given week the number of customers will be between 4760 and 5800.
- b) What is the probability of a week with more than 6500 customers?
- c) For 90% of the weeks, the number of customers should exceed what amount?

Exercise solution

$$\text{a) } P(4760 \leq x \leq 5800) = P\left(\frac{4760 - 5000}{800} \leq z \leq \frac{5800 - 5000}{800}\right) = P(-0.3 \leq z \leq 1) = 0.4592$$

$$\text{b) } P(x \geq 6500) = P\left(z \geq \frac{6500 - 5000}{800}\right) = P(z \geq 1.875) = 0.5 - P(0 \leq z \leq 1.875) = 0.5 - 0.4696 = 0.0304$$

$$\text{c) } P(z \geq -1.29) = 0.9 \Rightarrow P\left(\frac{x - 5000}{800} \geq -1.29\right) = 0.9 \Rightarrow x \geq 3968$$

Reading

- 1) Murray R. Spiegel, *Schaum's outline of Theory and Problems of Probability and Statistics*, McGraw-Hill, 23 edition, 1998.
- 2) Nitis Mukhopadhyay, *Probability and Statistical Inference*, Marcel Dekker, Inc. 2000.