TRAINING COURSES ON APPLIED ECONOMETRIC ANALYSIS (SUMMER SCHOOL) FOR YOUNG ECONOMISTS / RESEARCHERS ORGANIZED BY WIUT AND IFPRI SEPTEMBER 13-23, 2016

Basics of Probability theory

Farrukh Ataev

Lecturer, WIUT

14 June 2017

Outline

- **Session 1 Basic probability concepts**
  - (Tue 13 June at 13:30-15:00)
- **Session 2 Basic probability concepts (continued)**
  - (Tue 13 June at 15:30-17:00)
- **Session 3 Probability distributions (Discrete)**
  - (Wed 14 June at 9:00-10:30)
- **Session 4 Probability distributions (Continuous)**
  - (Wed 14 June at 11:00-12:30)

Session 4 Continuous probability distribution

- ❖ Continuous probability distribution
  - ❖ Uniform probability distribution
  - ❖ Normal probability distribution
  - ❖ Standard normal probability distribution

Types of probability distributions

Two types of a probability distribution:
1) Discrete probability distribution:
   o Binomial
   o Poisson
   o Hypergeometric

2) Continuous probability distribution:
   o Uniform
   o Normal
   o Exponential

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: $f(x) = \begin{cases} \frac{1}{b-a}, & for\ a \leq x \leq b \\ 0, & elsewhere \end{cases}$
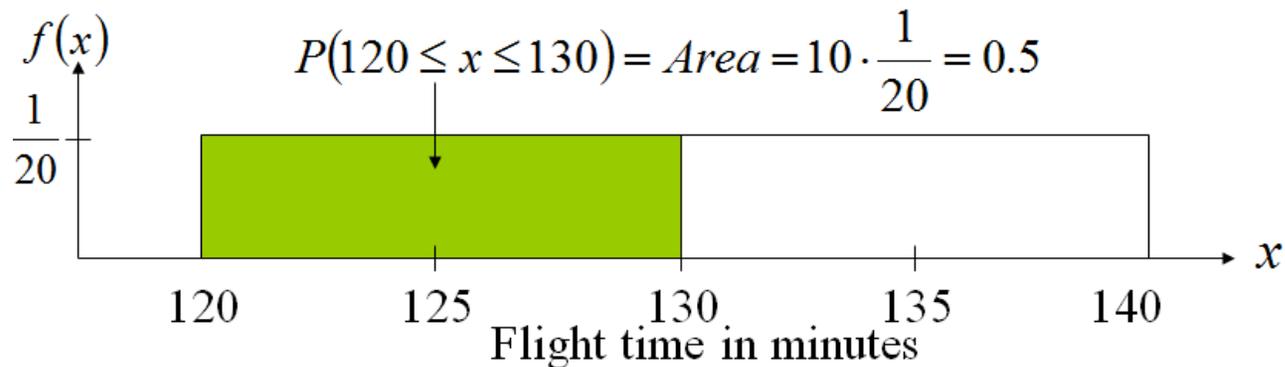
Example: Consider the random variable x that represents the total flight time of an airplane traveling from Chicago to New York. Suppose that the flight time can be any value in the interval from 120 minutes to 140 minutes.
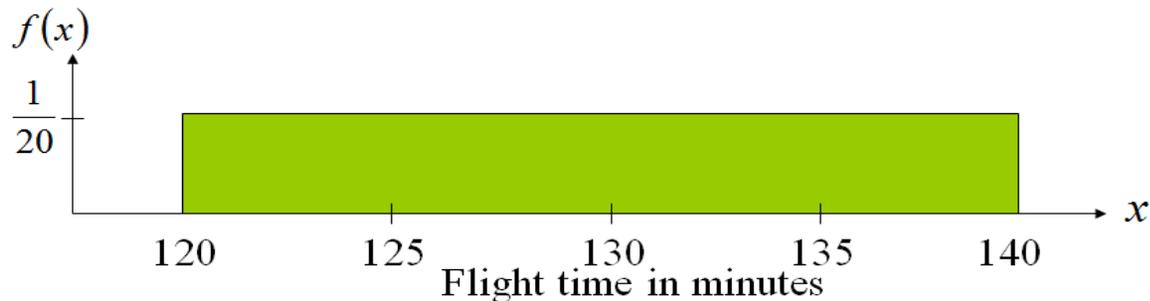
Probability versus probability function

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:

$$P(120 \leq x \leq 130) = Area = 10 \cdot \frac{1}{20} = 0.5$$

$f(x)$

$\frac{1}{20}$

120   125   130   135   140 x

Flight time in minutes

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

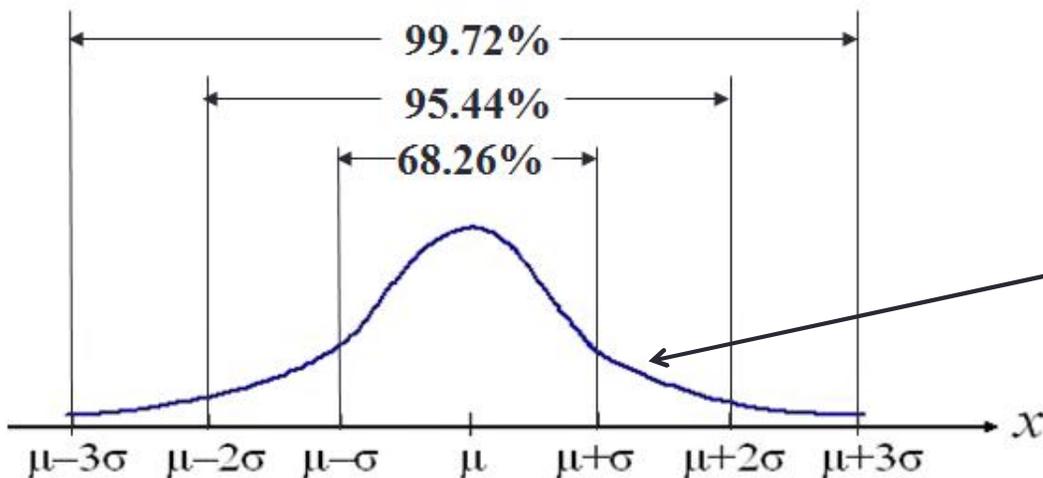# Normal probability distribution $N(\mu, \sigma)$

Continuous probability distribution with symmetrical, bell-shaped histogram defined by mean (μ) and standard deviation (σ)
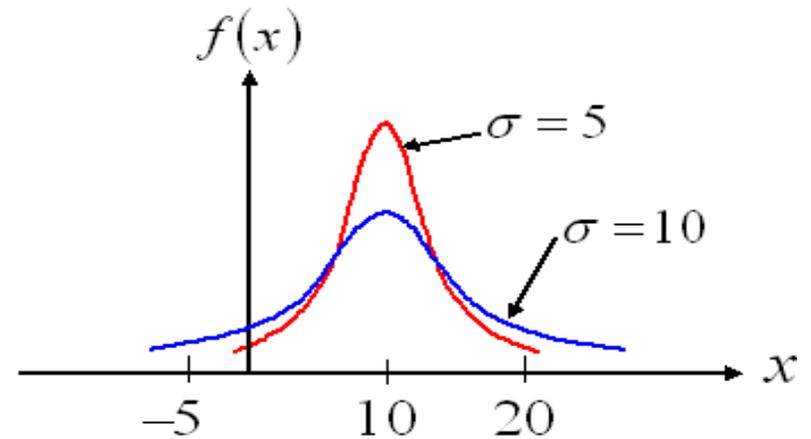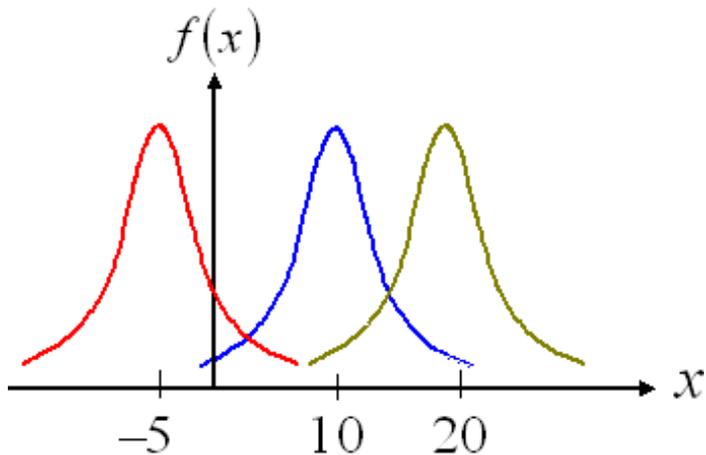
Exercise: If all normal distributions have the same shape, how do they differ?

Probability of the normal random variable

➢ Probabilities for the normal 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 curve for the normal distribution 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: $f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$
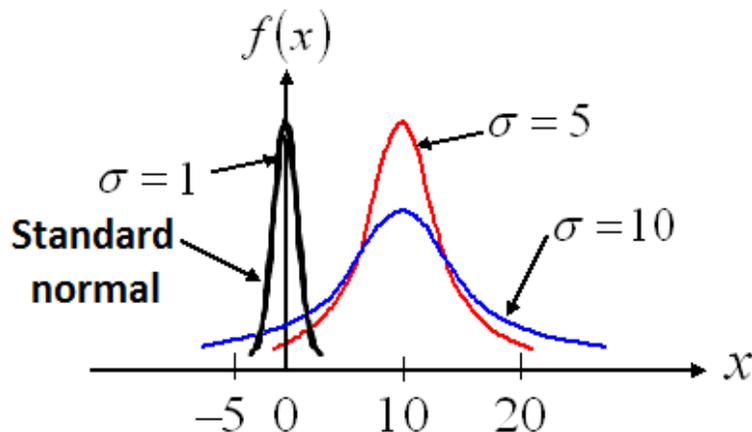
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \Rightarrow f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

Example

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

1-method: Using the definite integration formula:

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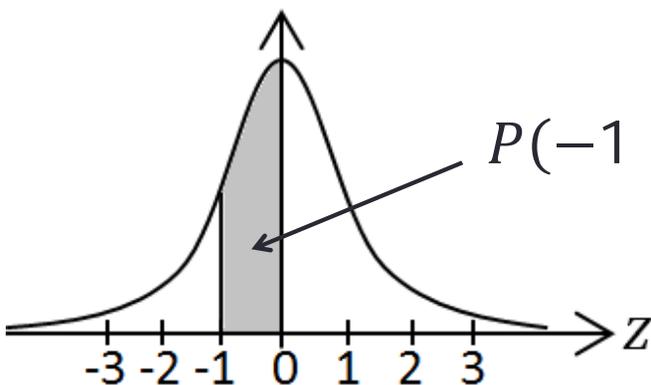
$$P(-1 \leq z \leq 0) = \int_{-1}^{0} f(z)dz = \int_{-1}^{0} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = 0.3413$$

Example (continued)

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

$$P(-1 \leq z \leq 0) = P(z \leq 0) - P(z \leq -1) = 0.5 - 0.1587 = 0.3413$$

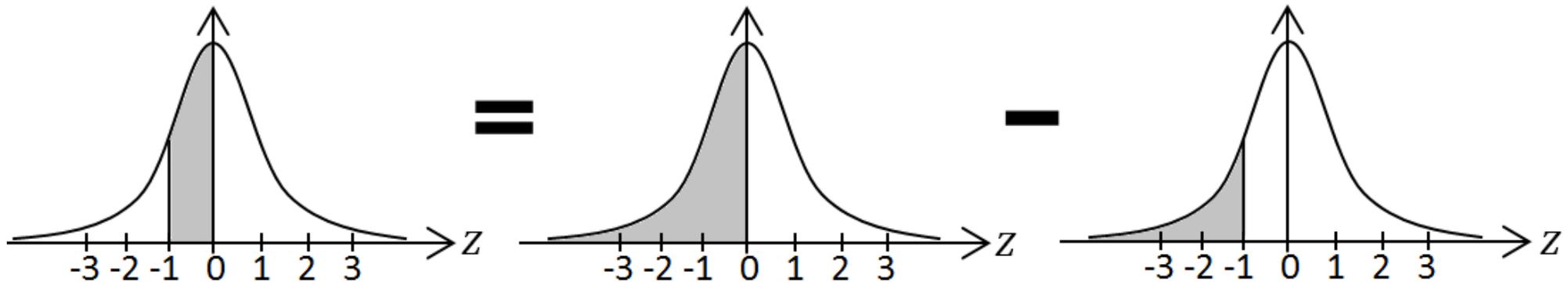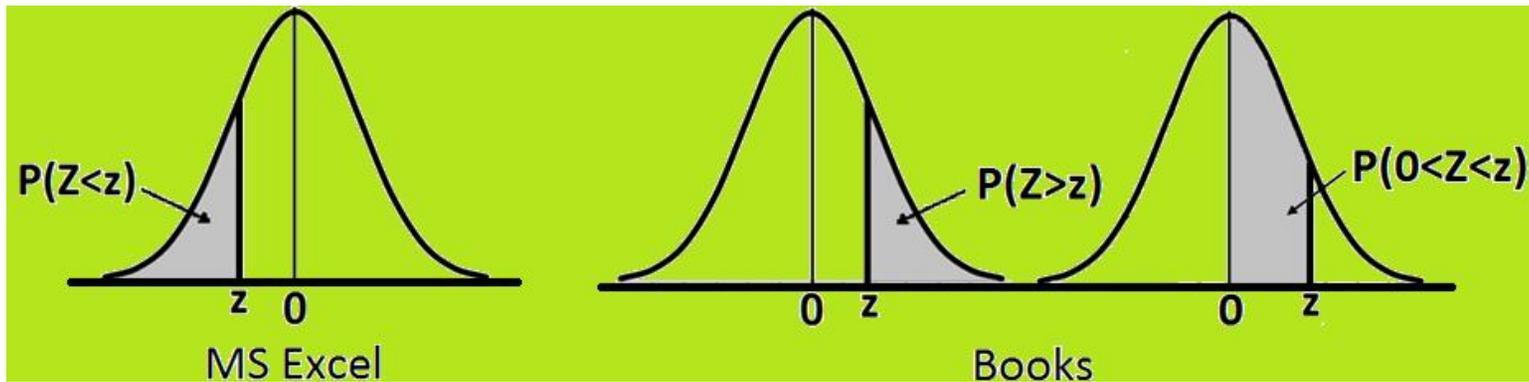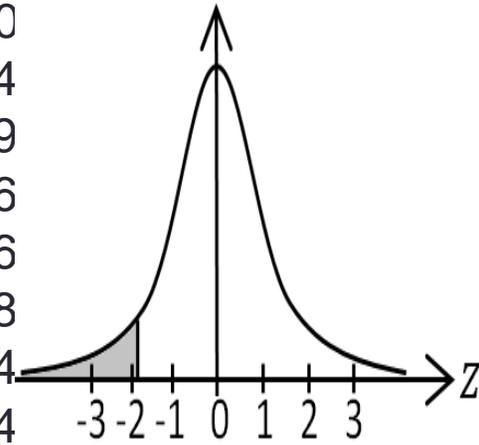# Standard normal distribution table

| z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.00	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
-2.90	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
-2.80	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
-2.70	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
-2.60	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
-2.50	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
-2.40	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
-2.30	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
-2.20	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
-2.10	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
-2.00	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
-1.90	.0287	.0281	.0274	.0268	.0262	.0256	**.0250**	.0244	.0239	.0233
-1.80	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
-1.70	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
-1.60	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455

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

$P(z \leq -1.96) = 0.0250$

Standard normal distribution table (continued)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-1.50	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
-1.40	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
-1.30	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
-1.20	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
-1.10	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
-1.00	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
-0.90	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
-0.80	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
-0.70	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
-0.60	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
-0.50	.3085	**.3050**	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
-0.40	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
-0.30	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
-0.20	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
-0.10	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.00	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
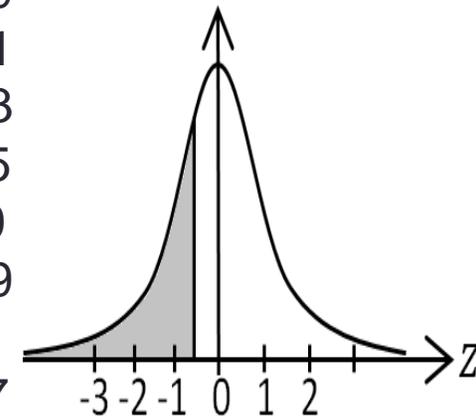
The entries in this table give the area under the standard normal curve to the left of z.

$P(z \leq -0.51) = 0.9750$

Exercise

Find from the z-table:

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

2) $P(z < -2.34)$

3) $P(-1 < z < 1)$

4) $P(z > 2)$

5) $P(z = 3)$

6) $P(-1.23 < z < 2.93)$

7) $P(z < -1.645)$

7) $z_{0.0505}$

8) $z_{0.0495}$

9) $z_{0.05}$

10) $z_{0.95}$

Example (continued)

3-method: MS Excel function:

NORM.DIST(X, μ, σ, 1) or NORM.S.DIST(Z, 1)

NORM.DIST(0, 0, 1, 1) − NORM.DIST(-1, 0, 1, 1) =

0.5 − 0.1587 = 0.3413

Or

NORM.S.DIST(0, 1) − NORM.S.DIST(-1, 1) =

0.5 − 0.1587 = 0.3413

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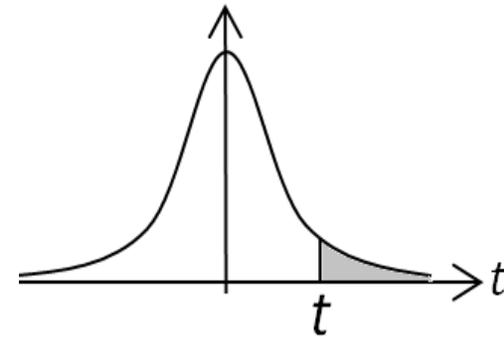
The entries in this table give t_α values, where α is the area (or probability) in the upper tail.

$t_{0.05}(2) = 2.92$

$P(t(2) > 2.92) = 0.05$

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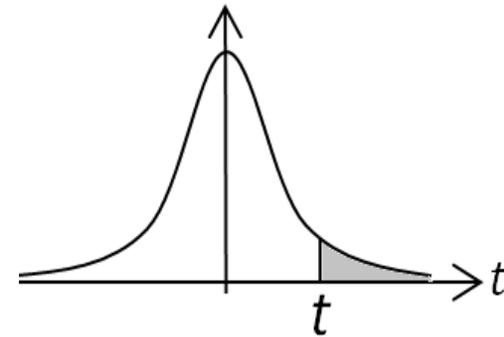
The entries in this table give t_α values, where α is the area (or probability) in the upper tail.

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

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 $\sigma = 40$. Letting x be the number of units demanded, find a) P $(x \leq 100)$; b) P $(180 \leq x \leq 220)$; c) P $(x \geq 250)$.

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

b) $P(180 \leq x \leq 220) = P\left(\frac{180-200}{40} \leq z \leq \frac{220-200}{40}\right) =$
$$= P(-0.5 \leq z \leq 0.5) = 0.3830$$

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

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

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

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

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.