4 Continuous Random Variables and Probability Distributions

Part II

4.3 Normal distribution

Normal Distribution

- Normal distribution or Gaussian distribution is the most important one in all of probability and statistics.
- Many numerical populations have distributions that can be fit very closely by an appropriate normal curve.

Normal Distribution

- Heights, weights, and other physical characteristics,
- Measurement errors in scientific experiments,

การวัดร่างกายมนุษย์ตามหลักวิทยาศาสตร์
- Anthropometric measurements on fossils,
- Reaction times in psychological experiments,
- Measurements of intelligence and aptitude, scores on various tests, and
- Numerous economic measures and indicators.

Normal Distribution

Definition

Continuous random variable X is said to have **Normal distribution** with parameters μ and σ (or μ and σ^2), where $-\infty < x < \infty$ and $\sigma > 0$, if the pdf of X is

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

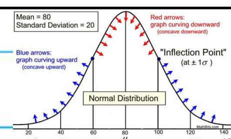
mathematical constant :3.14159

base of natural logarithm system and equals approximately 2.71828

- Statement that X is normally distributed with parameters μ and σ^2 is often abbreviated $X \sim N(\mu, \sigma^2)$.
- Clearly $f(x; \mu, \sigma) \geq 0$, but a somewhat complicated calculus argument must be used to verify that $\int_{-\infty}^{\infty} f(x:\mu,\sigma) = 1$.
- It can be shown that $E(X) = \mu$ and $V(X) = \sigma^2$, so the parameters are mean and standard deviation of X.

92

Normal Distribution



- Each density curve is symmetric about μ and bell-shaped, so center of bell (point of symmetry) is both mean of distribution and median.
- Value of σ is distance from μ to inflection points of curve (points at which curve changes from turning downward to turning upward).

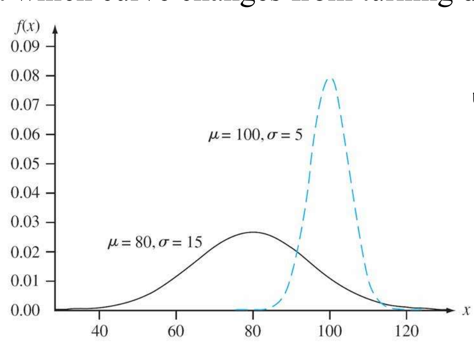


Figure 4.13(a) Two different normal density curves

ure 4.13 presents graphs of $f(x; \mu, \sigma)$ several different (μ, σ) pairs.

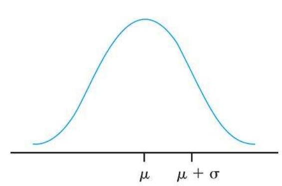
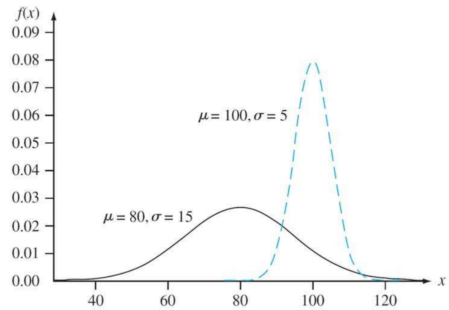


Figure 4.13(b) Visualizing μ and σ for a normal distribution

93

Normal Distribution

- Large values of σ yield graphs that are quite spread out about μ, whereas
- Small values of σ yield graphs with a high peak above μ and most of the area under the graph quite close to μ.
- Thus large σ implies that a value of X far from μ may well be observed, whereas such value is quite unlikely when σ is small.



95

Standard Normal Distribution

96

Standard Normal Distribution

- Computation of $P(a \le X \le b)$ when X is a normal rv with parameters μ and σ requires evaluating

$$P(a \le X \le b) = \int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx \qquad (4.4)$$

- None of standard integration techniques can be used to accomplish this.
- Instead, for $\mu = 0$ and $\sigma = 1$, Expression (4.4) has been calculated using numerical techniques and tabulated for certain values of a and b.
- This table can also be used to compute probabilities for any other values of μ and σ under consideration.

97

Standard Normal Distribution

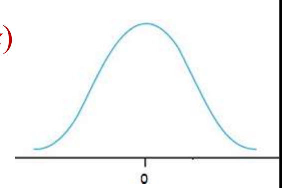
Definition

- Normal distribution with parameter values $\mu = 0$ and $\sigma = 1$ is called **Standard normal distribution.**
- Random variable having standard normal distribution is called **Standard normal random variable** and will be denoted by Z.
- The pdf of Z is

$$f(z:0,1) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} \qquad -\infty < z < \infty$$

- Graph of $f(z;\, 0,\, 1)$ is called *Standard normal* (or z)
- Its inflection points are at 1 and –1.
- The cdf of Z is

$$P(Z \le z) = \int_{-\infty}^{z} f(y:o,1)\,dy$$



which we will denote by $\Phi(z)$

98

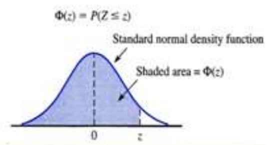
Standard Normal Distribution

○ Standard normal distribution almost never serves as model for naturally arising population.

○ Instead, it is reference distribution from which information about other normal distributions can be obtained.

○ Appendix Table A.3 gives $\Phi(z) = P(Z \le z)$, area under standard normal density curve to the left of z, for $z = -3.49, -3.48, \ldots, 3.48, 3.49$.

Table A.3 Standard Normal Curve Areas



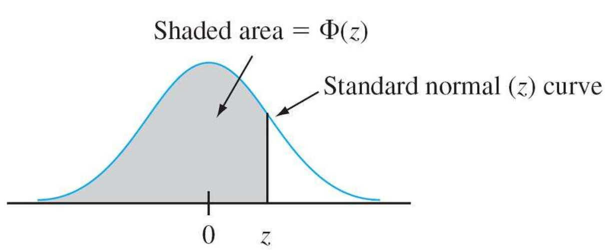
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007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

99

Standard Normal Distribution

○ Figure 4.14 illustrates type of cumulative area (probability) tabulated in Table A.3.

○ From this table, various other probabilities involving Z can be calculated.



Standard normal cumulative areas tabulated in Appendix Table A.3

Figure 4.14

100

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
−2.9	.0019	.0018	.0017	.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	.0038
−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	.0352	.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	.0722	.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
−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	.3482
−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

(continued)

101

Table A.3 Standard Normal Curve Areas *(cont.)*

$\Phi(z) = P(Z \le z)$

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	.9278	.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
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
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

102

Example 4.13

Let's determine the following standard normal probabilities:

$$P(Z \le z) = \int_{-\infty}^{z} f(y : o, 1) = \Phi(z)$$

(a) $P(Z \le 1.25)$,

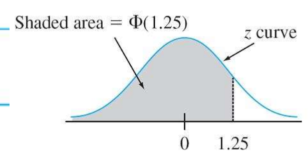
(b) $P(Z > 1.25)$,

(c) $P(Z \le -1.25)$, and

(d) $P(-0.38 \le Z \le 1.25)$.

103

Example 4.13



a. $P(Z \le 1.25) = \Phi\ (1.25)$,

probability that is tabulated in Appendix Table A.3 at the intersection of the row marked 1.2 and the column marked 0.05.

The number there is 0.8944, so $P(Z \le 1.25) = 0.8944$.

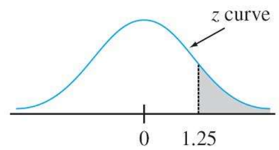
Table A.3 Standard Normal Curve Areas *(cont.)*

$\Phi(z) = P(Z \le z)$

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	[illegible]	[illegible]

104

Example 4.13



cont'd

b. $P(Z > 1.25) = 1 - P(Z \leq 1.25) = 1 - \Phi\,(1.25)$,

area under the z curve to the right of 1.25 (an upper-tail area).

Then $\Phi(1.25) = 0.8944$ implies that

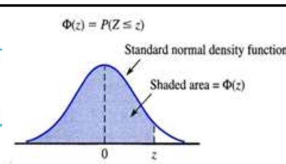
$$P(Z > 1.25) = 0.1056.$$

Table A.3 Standard Normal Curve Areas *(cont.)*

$\Phi(z) = P(Z \leq z)$

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	[illegible]	[illegible]	[illegible]	[illegible]

Example 4.13



ıt'd

c. $P(Z \leq -1.25) = \Phi\,(-1.25)$, a lower-tail area. Directly from Appendix Table A.3, $\Phi\,(-1.25) = 0.1056$.

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

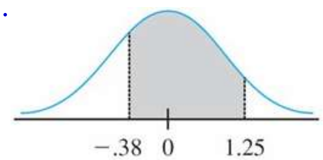
$$P(Z \leq -1.25) = 0.1056$$

By symmetry of the z curve, this is the same answer as in part (b).

Example 4.13

cont'd

d. $P(-0.38 \le Z \le 1.25)$ is area under the standard normal curve above the interval whose left endpoint is -0.38 and whose right endpoint is 1.25.



From Section 4.2, if X is a continuous rv with cdf $F(x)$, then

$$P(a \le X \le b) = F(b) - F(a)$$

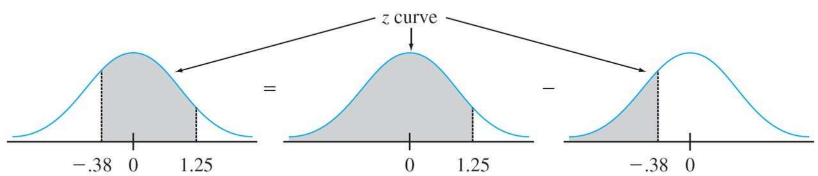


Figure 4.16 $P(-0.38 \le Z \le 1.25)$ as the difference between two cumulative areas

107

Example 4.13

cont'd

$$P(-0.38 \le Z \le 1.25) = P(Z \le 1.25) - P(Z \le -0.38)$$
$$= \Phi(1.25) - \Phi(-0.38)$$
$$= 0.8944 - 0.3520$$
$$= 0.5424$$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	[illegible]	[illegible]	[illegible]	[illegible]
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	[illegible]	[illegible]	[illegible]

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−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	.3482
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859

3

Percentiles of
Standard Normal Distribution

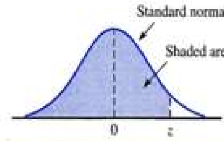
Percentiles of the Standard Normal Distribution

For any p between 0 and 1,

Appendix Table A.3 can be used to obtain the $(100p)^{\text{th}}$ percentile of standard normal distribution.

Table A.3 Standard Normal Curve Areas

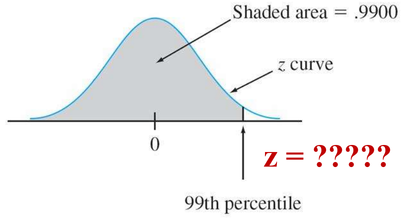
$\Phi(z) = P(Z \le z)$



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

Example 4.14

The 99^{th} percentile of standard normal distribution is that value on the horizontal axis such that the area under the z curve to the left of the value is 0.9900.



Appendix Table A.3 gives for fixed z the area under the standard normal curve to the left of z,

whereas here we have the area and want the value of z.

This is the "inverse" problem to $P(Z \le z) = ?$

Example 4.14

so the table is used in an inverse fashion:

○ Find in the middle of the table 0.9900;

○ Row and column in which it lies identify the 99^{th} z percentile.

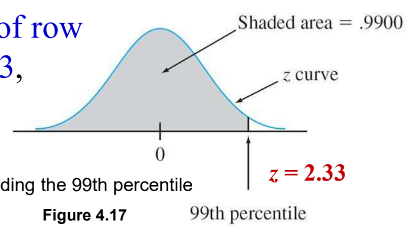
Table A.3 Standard Normal Curve Areas *(cont.)*

$\Phi(z) = P(Z \le z)$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	[illegible]	[illegible]	[illegible]	[illegible]

○ Here 0.9901 lies at intersection of row marked 2.3 and column marked .03,

○ so the 99th percentile is (approximately) $z = 2.33$.



Finding the 99th percentile

Figure 4.17

Example 4.14

cont'd

- By symmetry, the first percentile is as far below 0 as the 99th is above 0, so equals –2.33 (1% lies below the first and also above the 99th).

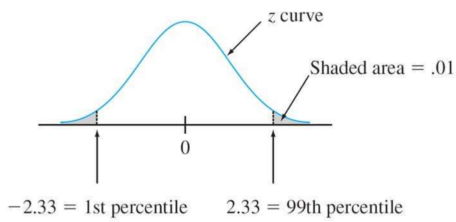


Figure 4.18 The relationship between the 1st and 99th percentiles

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

Percentiles of Standard Normal Distribution

- In general, the (100p)th percentile is identified by the row and column of Appendix Table A.3 in which entry ***p*** is found

- (e.g., the 67th percentile is obtained by finding 0.6700 in body of table, which gives $z = 0.44$).

$\Phi(z) = P(Z \leq z)$

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

Percentiles of Standard Normal Distribution

○ If p does not appear, the number closest to it is often used, although linear interpolation gives a more accurate answer.

115

Percentiles of Standard Normal Distribution

For example, to find the 95th percentile, we look for 0.9500 inside the table.

$\Phi(z) = P(Z \le z)$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9278	.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

Although 0.9500 does not appear, both 0.9495 and 0.9505 do, corresponding to $z = 1.64$ and 1.65, respectively.

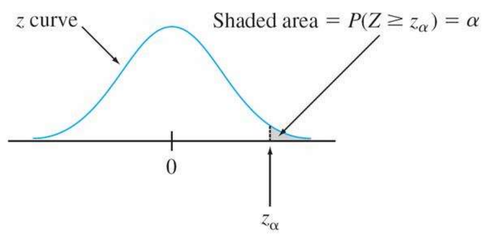
Since 0.9500 is halfway between the two probabilities that do appear, we will use **1.645** as the 95th percentile and –1.645 as the 5th percentile.

116

z_α Notation for z Critical Values

z_α Notation for z Critical Values

○ In statistical inference, we will need values on horizontal z axis that capture certain small tail areas under standard normal curve.



Notation

z_α will denote value on the z axis for which

α of area under z curve lies to the right of z_α.

z_α Notation for z Critical Values

For example, $z_{0.10}$ captures upper-tail area 0.10,

Since α of area under the z curve lies to the right of z_α, $1 - \alpha$ of the area lies to its left. Thus z_α *is the* $100(1 - \alpha)^{th}$ *percentile of standard normal distribution.*

By symmetry the area under the standard normal curve to the left of $-z_\alpha$ is also α.

The z_α's are usually referred to as ***z* critical values.**

z_α Notation for z Critical Values

Table 4.1 lists the most useful z percentiles and z_α values.

Percentile	90	95	97.5	99	99.5	99.9	99.95
α (tail area)	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
$z_\alpha = 100(1-\alpha)^{th}$ percentile	1.285	1.645	1.96	2.33	2.58	3.08	3.27

Table A.3 Standard Normal Curve Areas *(cont.)*

$\Phi(z) = P(Z \le z)$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
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	[illegible]	[illegible]	[illegible]	[illegible]				
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	[illegible]	[illegible]						[illegible]
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	[illegible]	[illegible]	[illegible]	[illegible]	.9985	.9986	.9986
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	[illegible]	[illegible]	[illegible]	[illegible]			

Example 4.15

$z_{0.05}$ is the $100(1 - 0.05)^{th} = 95^{th}$ percentile of the standard normal distribution, so $z_{0.05} = 1.645$.

Table A.3 Standard Normal Curve Areas *(cont.)*

$\Phi(z) = P(Z \le z)$

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

Area under standard normal curve to the left of $-z_{0.05}$ is also 0.05.

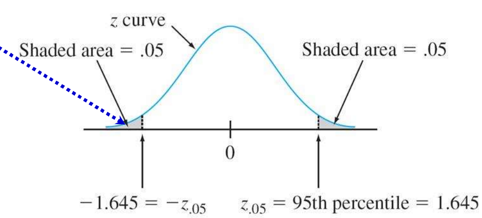


Figure 4.20 Finding $z_{.05}$

121

Nonstandard Normal Distributions

122

Nonstandard Normal Distributions

- When $X \sim N(\mu, \sigma^2)$, probabilities involving X are computed by "standardizing."
- **Standardized variable** is $(X - \mu)/\sigma$.

- Subtracting μ shifts mean from μ to zero, and then dividing by σ scales the variable so that standard deviation is 1 rather than σ.

Proposition
If X has Normal distribution with mean μ and standard deviation σ, then

$$Z = \frac{X - \mu}{\sigma}$$

has a standard normal distribution. Thus

123

Nonstandard Normal Distributions

Thus

$$P(a \le X \le b) = P\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right)$$

$$= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

$$P(X \le a) = \Phi\left(\frac{a-\mu}{\sigma}\right)$$

$$P(X \ge b) = 1 - \Phi\left(\frac{b-\mu}{\sigma}\right)$$

124

Nonstandard Normal Distributions

○ Key idea of proposition is that by standardizing, any probability involving X can be expressed as a probability involving a standard normal random variable Z, so that Appendix Table A.3 can be used.

This is illustrated in Figure 4.21.

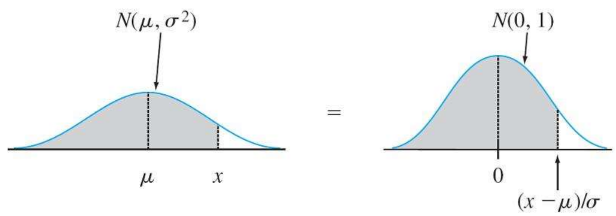


Figure 4.21 : Equality of nonstandard and standard normal curve areas

125

Example 4.16

○ The time that it takes a driver to react to the brake lights on a decelerating vehicle is critical in helping to avoid rear-end collisions.

○ The article "Fast-Rise Brake Lamp as a Collision-Prevention Device" (*Ergonomics*, 1993: 391–395) suggests that reaction time for an in-traffic response to a brake signal from standard brake lights can be modeled with a normal distribution having mean value 1.25 sec and standard deviation of 0.46 sec.

127

Example 4.16

cont'd

- What is the probability that reaction time is between 1.00 sec and 1.75 sec?
- If we let X denote reaction time, then standardizing gives

Example 4.16

cont'd

$$P(1.00 \le X \le 1.75) = P\left(\frac{1.00-1.25}{0.46} \le Z \le \frac{1.75-1.25}{0.46}\right) = P(-0.54 \le Z \le 1.09)$$

$$= \Phi(1.09) - \Phi(-0.54)$$

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

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776

$$= 0.8621 - 0.2946 \quad = 0.5675$$

This is illustrated in Figure 4.22

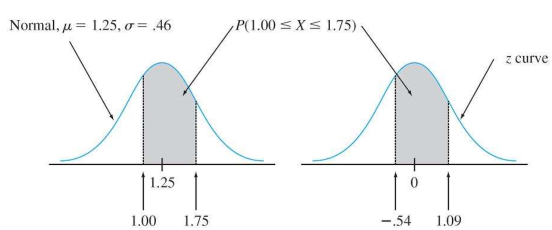


Figure 4.22 Normal curves for Example 4.16

Example 4.16

cont'd

○ Similarly, if we view 2 sec as a critically long reaction time, probability that actual reaction time will exceed this value is

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633

Percentiles of Arbitrary Normal Distribution

133

Percentiles of Arbitrary Normal Distribution

- The $(100p)^{th}$ percentile of normal distribution with mean μ and standard deviation σ is easily related to the $(100p)^{th}$ percentile of standard normal distribution.

Proposition

$$\begin{matrix}(100p)\text{th percentile} \\ \text{for normal } (\mu, \sigma)\end{matrix} = \mu + \begin{bmatrix}(100p)\text{th for} \\ \text{standard normal}\end{bmatrix} \cdot \sigma$$

- Another way of saying this is that if z is desired percentile for standard normal distribution, then desired percentile for normal (μ, σ) distribution is **z standard deviations from μ.**

134

Example 4.18

- The amount of distilled water dispensed by a certain machine is normally distributed with mean value 64 oz and standard deviation 0.78 oz.
- What container size c will ensure that overflow occurs only 0.5% of the time?

- If X denotes the amount dispensed, the desired condition is that $\boldsymbol{P(X > c) = 0.005}$, or, equivalently, that $\boldsymbol{P(X \le c) = 0.995}$.
- Thus c is the 99.5^{th} percentile of normal distribution with $\mu = 64$ and $\sigma = 0.78$.

135

Example 4.18

$$\begin{matrix}(100p)\text{th percentile} \\ \text{for normal } (\mu, \sigma)\end{matrix} = \mu + \begin{bmatrix}(100p)\text{th for} \\ \text{standard normal}\end{bmatrix} \cdot \sigma$$

cont'd

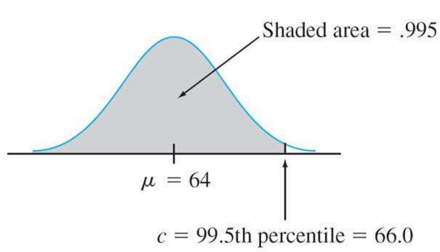
137

Example 4.18

cont'd

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

The 99.5th percentile of the standard normal distribution is 2.58, so



This is illustrated in Figure 4.23.

Figure 4.23 Distribution of amount dispensed for Example 18

138

Normal Distribution and Discrete Populations

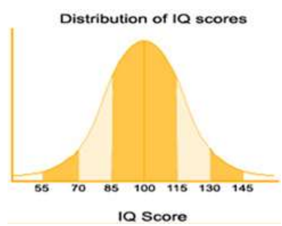
Normal Distribution and Discrete Populations

- **Normal distribution** is often used as an approximation to the distribution of values in **discrete population**.
- In such situations, extra care should be taken to ensure that probabilities are computed in an accurate manner.

Example 4.19

- Intelligence Quotient (IQ) in a particular population (as measured by a standard test) is known to be approximately normally distributed with $\mu = 100$ and $\sigma = 15$.

- What is the probability that a randomly selected individual has an **IQ of at least 125?**



ระดับ	ไอคิว	ร้อยละ
อัจฉริยะ	>144	0.13
ปัญญาเลิศ	130-144	2.14
เหนือค่าเฉลี่ย	115-129	13.59
สูงกว่าค่าเฉลี่ย	100-114	34.13
ค่อนข้างต่ำ	85-99	34.13
ต่ำกว่าค่าเฉลี่ย	70-84	13.59
คาบเส้น	55-69	2.14
ต่ำ	<55	0.13

141

Example 4.19

Solution

- Letting X = the IQ of a randomly chosen person, we wish

$$P(X \geq 125).$$

- The temptation here is to standardize $X \geq 125$ as in previous examples.

- However, the IQ population distribution is actually discrete, since IQs are integer-valued.

142

Example 4.19

cont'd

- So normal curve is approximation to a discrete probability histogram, as pictured in Figure 4.24.

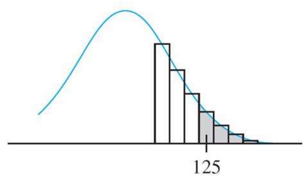


Figure 4.24 Normal approximation to a discrete distribution

- Rectangles of histogram are *centered* at integers, so IQs of at least 125 correspond to rectangles beginning at 124.5, as shaded in Figure 4.24.

- Thus we really want the area under the approximating normal curve to the right of 124.5.

143

Example 4.19

cont'd

$$Z=\frac{X-\mu}{\sigma} \qquad Z=\frac{124.5-100}{15}=\frac{24.5}{15}=1.63333 \qquad Z=\frac{125-100}{15}=\frac{25}{15}=1.66666$$

- Standardizing this value (124.5) gives $P(Z \geq 1.63) = 0.0516$, whereas standardizing 125 results in $P(Z \geq 1.67) = 0.0475$.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	[illegible]	[illegible]	[illegible]	[illegible]

- Difference is not great, but the answer 0.0516 is more accurate.
- Similarly, $P(X = 125)$ would be approximated by area between 124.5 and 125.5, since area under normal curve above the single value 125 is zero.

144

Example 4.19

cont'd

○ Correction for discreteness of the underlying distribution in Example 19 is often called a **continuity correction.**

○ It is useful in the following application of normal distribution to the computation of binomial probabilities.

145

Approximating Binomial Distribution

146

Approximating Binomial Distribution

- Recall that mean value and standard deviation of binomial random variable X are

$$\mu_X = np$$

$$\sigma_X = \sqrt{npq}$$

147

Approximating Binomial Distribution

Figure 4.25 displays a binomial probability histogram for binomial distribution with $n = 20$, $p = 0.6$, for which

$$\mu_X = np = 20(0.6) = 12 \quad \text{and}$$

$$\sigma_X = \sqrt{npq} = \sqrt{20(0.6)(0.4)} = 2.19$$

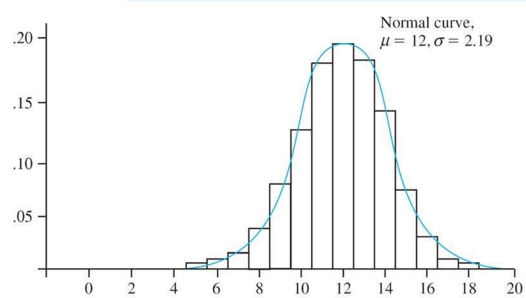


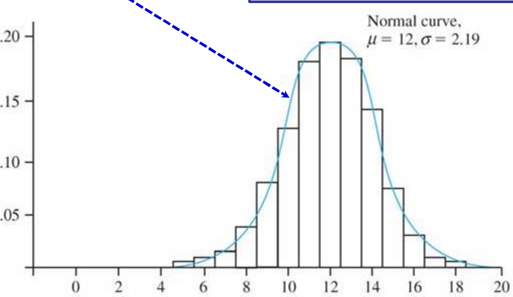
Figure 4.25 Binomial probability histogram for $n = 20$, $p = 0.6$ with normal approximation curve superimposed

148

Approximating Binomial Distribution

Normal curve with this μ and σ has been superimposed on probability histogram.

Although probability histogram is a bit skewed (because $p \neq 0.5$), normal curve gives a very good approximation, especially in the middle part of picture.



Area of any rectangle (probability of any particular X value) except those in the extreme tails can be accurately approximated by the corresponding normal curve area.

149

Approximating Binomial Distribution

For example, $P(X = 10) = B(10; 20, 0.6) - B(9; 20, 0.6)$

$= 0.245 - 0.128 = 0.117$,

n=20

x	0.01	0.05	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.75	0.80	0.90	0.95	0.99
9	1.000	1.000	1.000	.997	.986	.952	.755	.412	.128	.017	.004	.001	.000	.000	.000
10	1.000	1.000	1.000	.999	.996	.983	.872	.588	.245	.048	.014	.003	.000	.000	.000
11	1.000	1.000	1.000	1.000	.999	.995	.943	.748	.404	.113	.041	.010	.000	.000	.000

whereas area under the normal curve between 9.5 and 10.5 is

$P(-1.14 \leq Z \leq -0.68) = 0.1212.$ $\mu = 12$ $\sigma = 2.19$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170

150

Approximating Binomial Distribution

- More generally, as long as binomial probability histogram is not too skewed, binomial probabilities can be well approximated by normal curve areas.

- It is then customary to say that X has approximately a normal distribution.

151

Approximating Binomial Distribution

Proposition

- Let X be binomial random variable based on n trials with success probability p.
- Then if binomial probability histogram is not too skewed, X has approximately normal distribution with $\mu = np \quad and \quad \sigma = \sqrt{npq}$

In particular, for x = possible value of X,

$$P(X \le x) = B(x, n, p) \approx \begin{pmatrix} area\ under\ normal\ curve \\ to\ the\ left\ of\ \ x + 0.5 \end{pmatrix} \quad = \Phi\left(\frac{x + 0.5 - np}{\sqrt{npq}}\right)$$

- In practice, approximation is adequate provided that both ***np* ≥ 10** and ***nq* ≥ 10**, since there is enough symmetry in underlying binomial distribution.

152

Example 4.20

- Suppose that 25% of all students at a large public university receive financial aid.
- Let X be the number of students in a random sample of size 50 who receive financial aid, so that $p = 0.25$.

  Then μ **= 12.5** and σ **= 3.06**.

- Since $np = 50(0.25) = 12.5 \geq 10$ and $nq = 50(0.75) = 37.5 \geq 10$, the approximation can safely be applied.

Example 4.20

$P(X \leq x) = B(x,n,p) \approx \begin{pmatrix} \text{area under normal curve} \\ \text{to the left of } x+0.5 \end{pmatrix} = \Phi\left(\frac{x+0.5-np}{\sqrt{npq}}\right)$

cont'd

- Probability that at most 10 students receive aid is

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−0.6	.2743	.2709	.2676	.2643	.2611	$\boxed{.2578}$	.2546	.2514	.2483	.2451

Example 4.20

$P(X \le x) = B(x, n, p) \approx \begin{pmatrix} \text{area under normal curve} \\ \text{to the left of } x + 0.5 \end{pmatrix} = \Phi\left(\frac{x + 0.5 - np}{\sqrt{npq}}\right)$ cont'd

- Similarly, probability that between 5 and 15 (inclusive) of selected students receive aid is

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389

Example 4.20

$P(X \le x) = B(x, n, p) \approx \begin{pmatrix} \text{area under normal curve} \\ \text{to the left of } x + 0.5 \end{pmatrix} = \Phi\left(\frac{x + 0.5 - np}{\sqrt{npq}}\right)$ cont'd

- Exact probabilities are 0.2622 and 0.8348, respectively, so approximations are quite good.