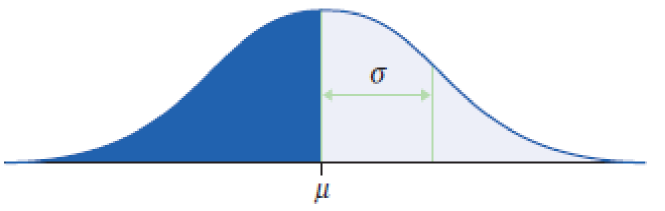
Chapter 5 - Normal Distribution



Lecture 13

Normal Distribution

Standard Normal Distribution

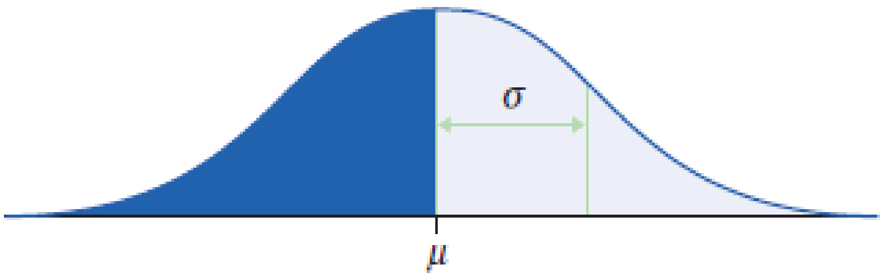
Dr. Lasantha Premarathna

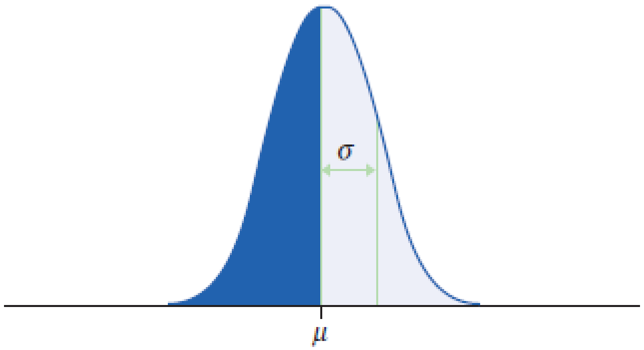
Chapter 5 - Normal Distribution
Outline

- Normal Distribution
- The 68-95-99.7 rule (Empirical rule)
- Z- Score
- The Standard Normal Distribution
- Finding Normal proportions
- Using the standard Normal table
- Finding a value given a proportion
- Important facts about the Normal Distribution

Normal Distributions

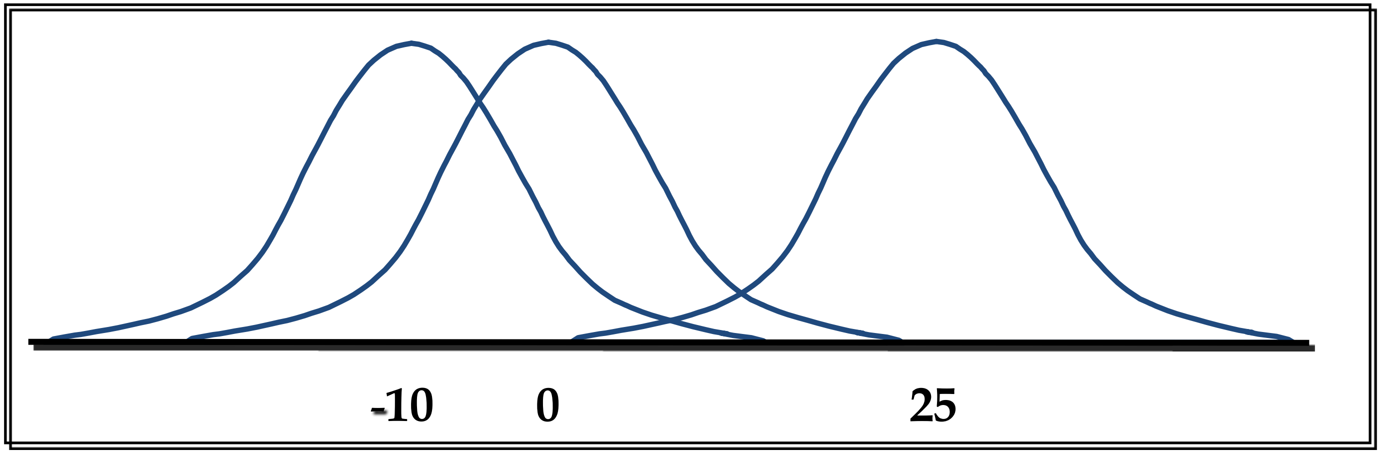
- The Normal distribution is the most important distribution in Statistics.
- All Normal curves are symmetric, single-peaked, and bell-shaped
- Any specific Normal curve is described by giving its mean *μ* (mu) and standard deviation *σ (sigma)* where *μ* and *σ* are “parameters” which control the central location and the dispersion





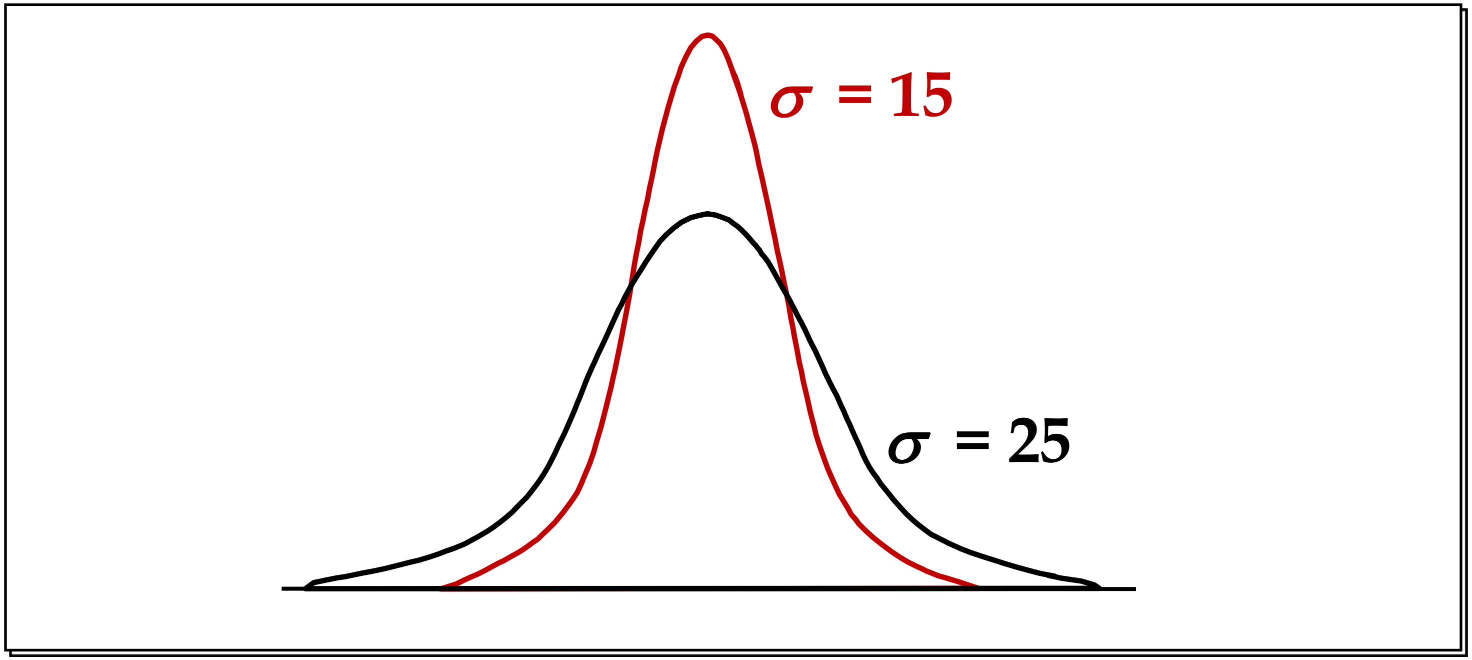
Normal Distributions

- Any particular Normal distribution is completely specified by two numbers: its mean μ and standard deviation σ
  - The mean is located at the center of the symmetric curve and is the same as the median. Changing μ without changing σ moves the Normal curve along the horizontal axis without changing its variability.



Normal Distributions

- The standard deviation σ controls the variability of a Normal curve. When the standard deviation is larger, the area under the normal curve is less concentrated about the mean.



The Normal Distribution: as a mathematical function (pdf)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \quad ; \quad -\infty < x < \infty \, , \quad -\infty < \mu < \infty \, , \quad \sigma > 0$$

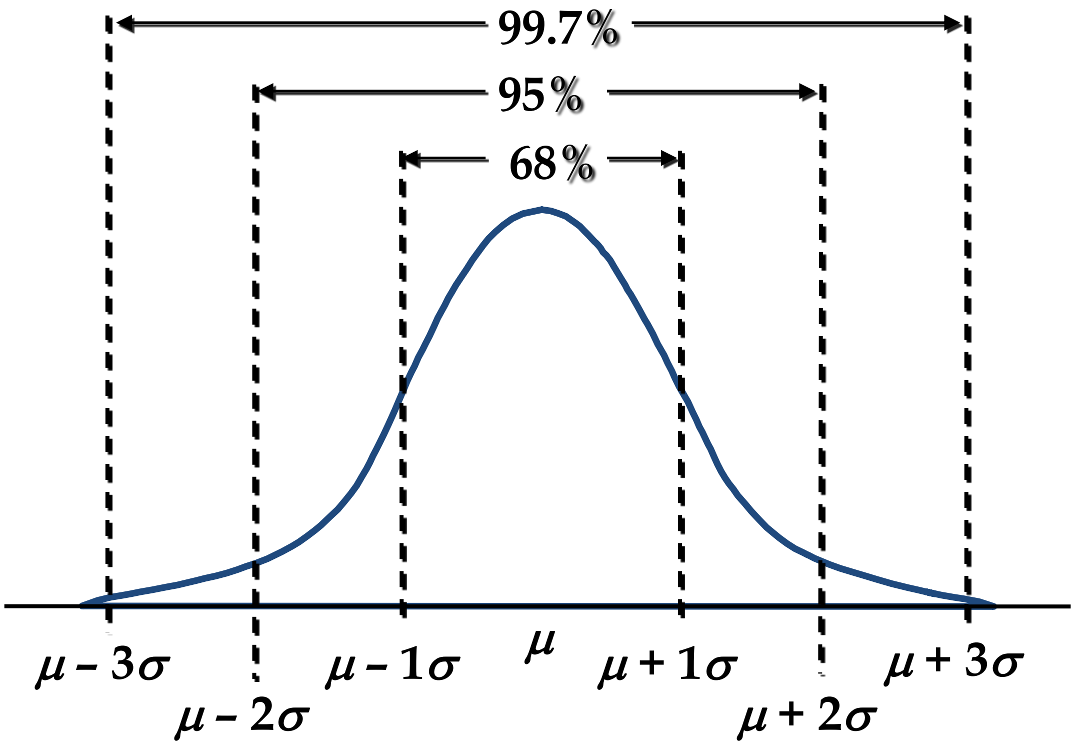
Notation:

$$X \sim N(\mu, \sigma^2)$$

The 68 - 95 - 99.7 Rule

In the Normal distribution with mean μ and standard deviation σ:

- Approximately 68% of the observations fall within σ of μ.
- Approximately 95% of the observations fall within 2σ of μ.
- Approximately 99.7% of the observations fall within 3σ of μ.

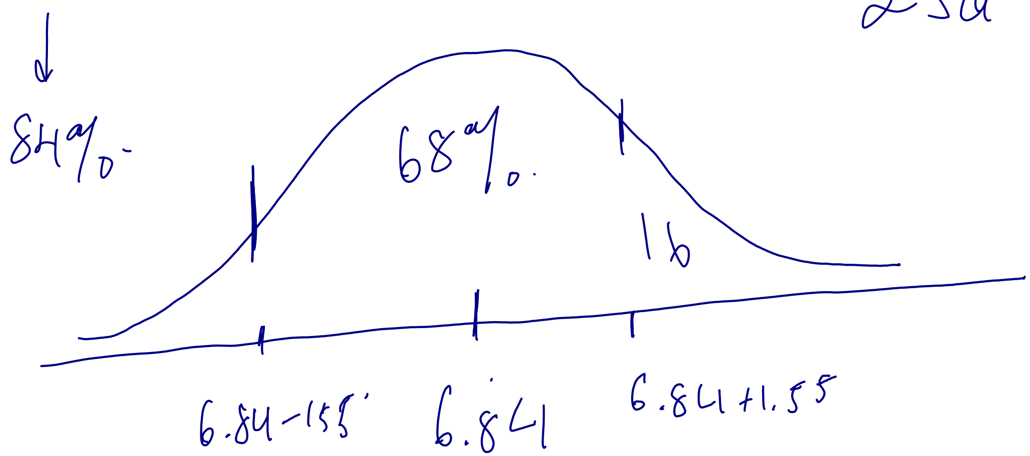


The 68 - 95 - 99.7 Rule - Example

- The distribution of Iowa Test of Basic Skills (ITBS) vocabulary scores for seventh-grade students in Gary, Indiana, is close to Normal. Suppose the distribution is $N(6.84, 1.55^2)$.
- Sketch the Normal density curve for this distribution.
- What percent of ITBS vocabulary scores is between 3.74 and 9.94?
- What percent of the scores is above 5.29?

2 sd - 95%

3.74, 9.94



Z-Score

- If x is an observation from a distribution that has mean μ and standard deviation σ, the standardized value of x

$$z = \frac{x - \mu}{\sigma}$$

- A standardized value is often called **Z-score**
- The Z-score for a value x of a random variable is the number of standard deviations that x falls from the mean
- A negative (positive) z-score indicates that the value is below (above) the mean.

Standardizing - Example

The heights of women aged 20 to 29 in the United States are approximately Normal with $\mu = 64.2$ and $\sigma = 2.8$ inches

- A woman 70 inches tall has standardized height

  $$z = (70 - 64.2) / 2.8 = 2.07$$

  or 2.07 standard deviations above the mean.

- Similarly, a woman 5 feet (60 inches) tall has standardized height

  $$z = (60 - 64.2) / 2.8 = -1.50$$

  or 1.5 standard deviations *less than* the mean height.

Example: Comparing Test Scores That Use Different Scales

- Z-scores can be used to compare observations from different normal distributions.

Picture the Scenario:

There are two primary standardized tests used by college admissions, the SAT and the ACT.

Consider that the SAT scores and ACT scores are normally distributed.

A students scored 650 on the SAT which has μ =500 and σ =100 and 30 on the ACT which has μ =21 and σ = 4.7.

How can we compare these scores to tell which score is relatively higher?

$$z_{SAT} = \frac{650-500}{100} = 1.5$$

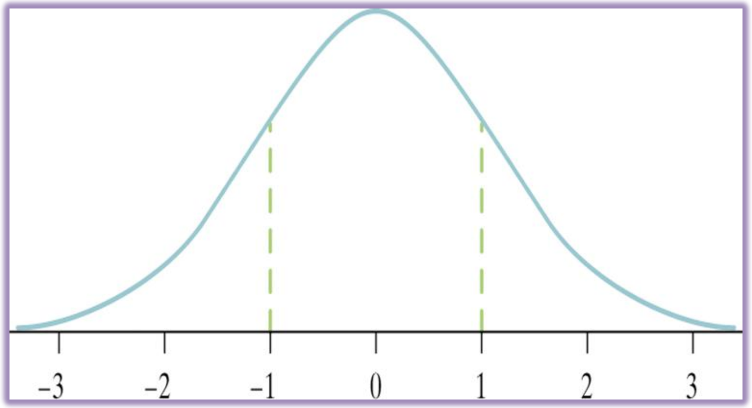
$$z_{ACT} = \frac{30-21}{4.7} = 1.91$$

The Standard Normal Distribution

- The **standard Normal distribution** is the Normal distribution with mean 0 and standard deviation 1.
- If a variable x has any Normal distribution $N(\mu,\sigma^2)$ with mean μ and standard deviation σ, then the standardized variable

$$z = \frac{x - \mu}{\sigma}$$

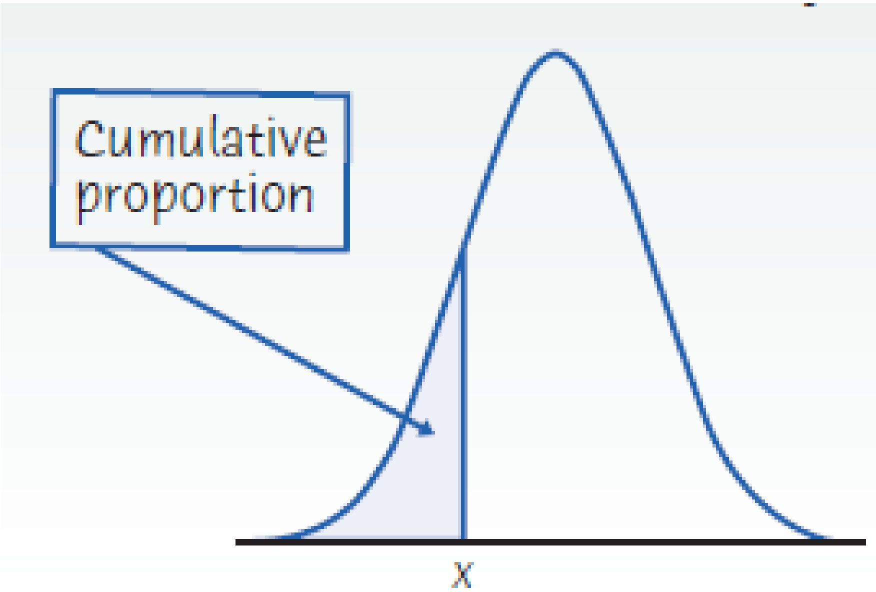
has the standard Normal distribution, $N(0,1)$.



Because all Normal distributions are the same when we standardize, we can find areas under any Normal curve from a single table

Cumulative Proportions

- The cumulative proportion for a value *x* in a distribution is the proportion of observations in the distribution that are less than or equal to *x*.



Standard Normal Table

Table entry for z is the area under the standard Normal curve to the left of z.

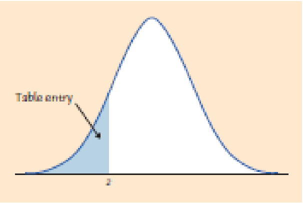


TABLE A Standard Normal cumulative proportions

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	.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
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Table entry for z is the area under the standard Normal curve to the left of z.

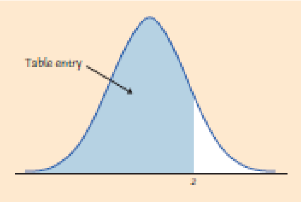


TABLE A Standard Normal cumulative proportions *(continued)*

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
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

Standard Normal Distribution is symmetric about zero. Therefore one table is enough to calculate any probability using normal distributions.

Table entry for *z* is the area under the standard normal curve to the left of *z*.

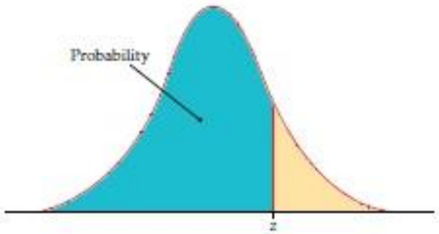


TABLE A

Standard normal probabilities (continued)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
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

The Standard Normal Table

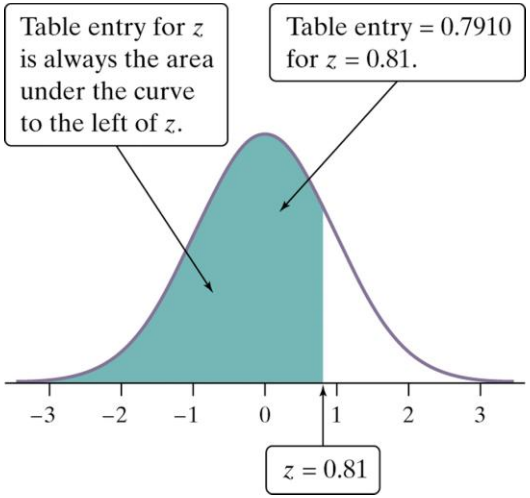
The Standard Normal Table

Table A is a table of areas under the standard Normal curve. The table entry for each value *z* is the area under the curve to the left of *z*.

Suppose we want to find the proportion of observations from the standard Normal distribution that are less than 0.81.
We can use Table A:

Z	.00	.01	.02
0.7	.7580	[illegible]	.7642
0.8	.7881	.7910	.7939
0.9	.8159	.8186	.8212

$P(z < 0.81) = .7910$



Normal Calculations

Find the proportion of observations from the standard Normal distribution that are between –1.25 and 0.81.

$P(X < 0.81) = 0.7910$

$P(X < -1.25) = 0.1056$

$P(-1.25 < X < 0.81) = 0.7910 - 0.1056$

$= 0.6854$

Example

SAT reading scores for a recent year are distributed according to a $N(500, 100^2)$ distribution. You scored 650 this particular year. What proportion of test takers in this year is better than you?

$$z = \frac{650 - 500}{100} = \frac{150}{100} = 1.5$$

$$= 0.9332$$

93.3%.

Normal Calculations

USING TABLE A TO FIND NORMAL PROPORTIONS

- Step 1. **State the problem** in terms of the observed variable *x*. **Draw a picture** that shows the proportion you want in terms of cumulative proportions.
- Step 2. **Standardize *x*** to restate the problem in terms of a standard Normal variable *z*.
- Step 3. **Use Table A** and the fact that the total area under the curve is 1 to find the required area under the standard Normal curve.

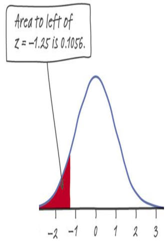
Normal Calculations using R

- $P(Z < 0.81) = ?$

  > pnorm(0.81, mean=0, sd=1)

  or simply use > pnorm(0.81)

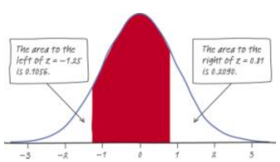
  [1] **0.7910299**



- $P(-1.25 < Z < 0.81) = ?$

  > pnorm(0.81) - pnorm(-1.25)

  [1] **0.6853801**

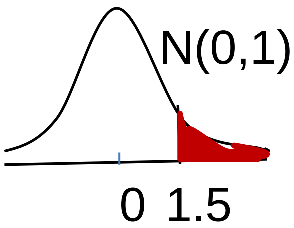


- $P(X > 650) = ?$ Where $X \sim N(500, 100^2)$

  $P(X > 650) = P(Z > 1.5)$

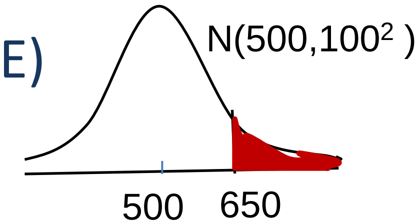
  > pnorm(1.5, lower.tail=FALSE)

  [1] **0.0668072**



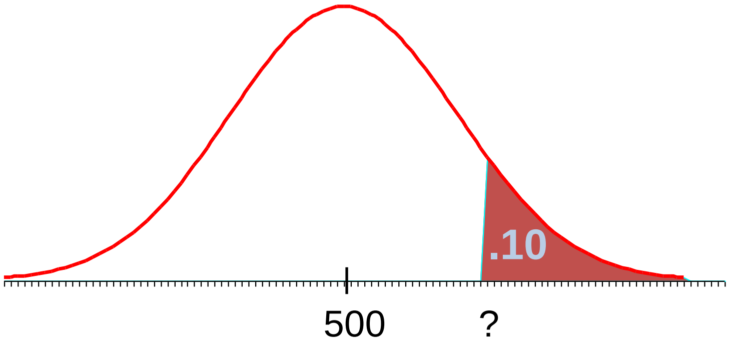
  > pnorm(650, mean=500, sd=100, lower.tail=FALSE)

  [1] 0.0668072



Normal Calculations

- How high must a student score in order to be in the top 10% of the distribution?



Look up the closest probability (closest to 0.10) in the table.

Find the corresponding **standardized score**.

The value you seek is that many standard deviations from the mean.

z	.07	.08	.09
1.1	.8790	.8[illegible]0	.8830
1.2	.8980	.8997	.9015
1.3	.9147	.9162	.9177

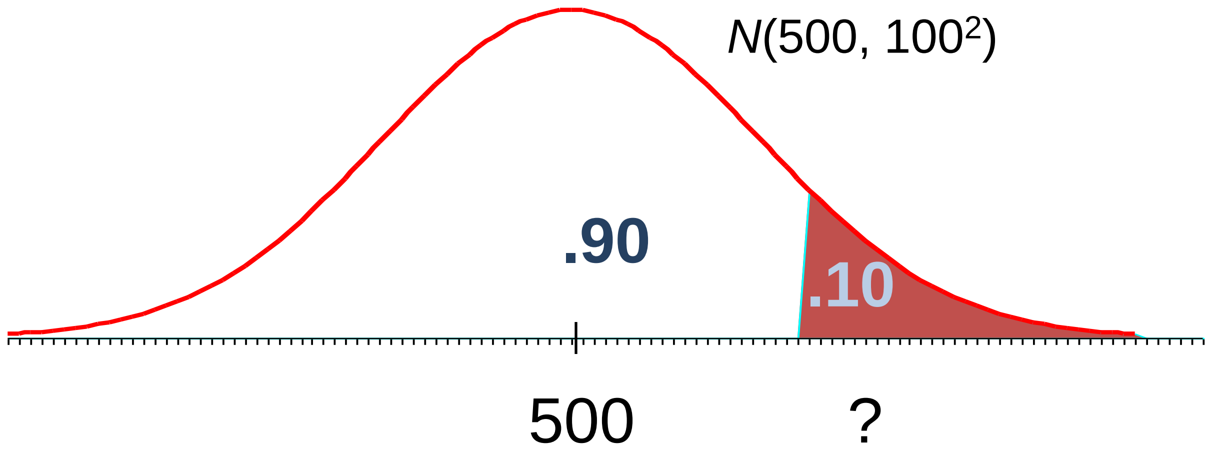
z = 1.28

Finding a Value Given a Proportion

- SAT reading scores for a recent year are distributed according to a $N(500, 100^2)$ distribution.
- How high must a student score in order to be in the top 10% of the distribution?

Normal Calculations

- SAT reading scores for a recent year are distributed according to a $N(500, 100^2)$ distribution.
- How high must a student score in order to be in the top 10% of the distribution?
- In order to use table A, equivalently, what score has cumulative proportion 0.90 *below* it?



“Backward” Normal Calculations

USING TABLE A GIVEN A NORMAL PROPORTION

- Step 1. **State the problem** in terms of the given proportion. **Draw a picture** that shows the Normal value, x, you want in relation to the cumulative proportion.
- Step 2. **Use Table A**, the fact that the total area under the curve is 1, and the given area under the standard Normal curve to find the corresponding z-value.
- Step 3. **Unstandardize** z to solve the problem in terms of a non-standard Normal variable x.

Summary

- Normal Distribution, The 68-95-99.7 rule (Empirical rule), Z- Score, The Standard Normal Distribution, bribability calculation using Normal distributions

Before the next class

- Review the lecture 13 and related sections in the text book

Next Class:

- Chapter 5 : Normal Distribution