NORMAL DISTRIBUTIONS

- We can sometimes describe the overall pattern of a distribution by a **density curve.** A density curve has total area 1 underneath it. An area under a density curve gives the proportion of observations that fall in a range of values.
- A density curve is an idealized description of the overall pattern of a distribution that smooths out the irregularities in the actual data. We write the **mean of a density curve** as μ and the **standard deviation of a density curve** as σ to distinguish them from the mean x and standard deviation s of the actual data.
- The mean, the median, and the quartiles of a density curve can be located by eye. The **mean** μ is the balance point of the curve. The **median** divides the area under the curve in half. The **quartiles** and the median divide the area under the curve into quarters. The **standard deviation** σ cannot be located by eye on most density curves.
- The mean and median are equal for symmetric density curves. The mean of a skewed curve is located farther toward the long tail than is the median.
- The **Normal distributions** are described by a special family of bell-shaped, symmetric density curves, called **Normal curves.** The mean μ and standard deviation σ completely specify a Normal distribution $N(\mu, \sigma)$. The mean is the center of the curve, and σ is the distance from μ to the change-of-curvature points on either side.
- To **standardize** any observation x, subtract the mean of the distribution and then divide by the standard deviation. The resulting ***z*-score**

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

  Says how many standard deviations x lies from the distribution mean.
- All Normal distributions are the same when measurements are transformed to the standardized scale. In particular, all Normal distributions satisfy the **68−95−99.7 rule,** which describes what percent of observations lie within one, two, and three standard deviations of the mean.
- If x has the $N(\mu, \sigma)$ distribution, then the **standardized variable** $z = (x - \mu)/\sigma$ has the **standard Normal distribution** $N(0, 1)$ with mean 0 and standard deviation 1. The normal distribution table below gives the **cumulative proportions** of standard Normal observations that are less than z for many values of z. By standardizing, we can use Table A for any Normal distribution.

Standard Normal Curve Probability Distribution (Z)

The table is based on the upper right 1/2 of the Normal Distribution; total area shown is .5
The Z-score values are represented by the column value + row value, up to two decimal places
The probabilities up to the Z-score are in the cells

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
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

Critical Values of t-Distribution (To be used later in the course)

The table shows the critical t-values for a given alpha level (one-tailed) and degrees of freedom
The degrees of freedom are the rows (denoted by v)
Note: The probability levels represent the whole of alpha (you must divide alpha by 2 if you want the t-value for a two-tailed test)

v	0.1000	0.0500	0.0250	0.0100	0.0050	0.0010	0.0005
1	3.078	6.314	12.706	31.821	63.657	318.309	636.619
2	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
50	1.299	1.676	2.009	2.403	2.678	3.261	3.496
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
100	1.290	1.660	1.984	2.364	2.626	3.174	3.390
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
Infinity	**1.282**	**1.645**	**1.960**	**2.326**	**2.576**	**3.090**	**3.291**

Normal is only approximate: IQ test scores. Here are the IQ test scores of 31 seventh-grade girls in a Nyahururu school.

114 100 104 89 102 91 114 114 103 105 108 130 120 132 111 128 118 119
86 72 111 103 74 112 107 103 98 96 112 112 93

(a) We expect IQ scores to be approximately Normal. Make a stem plot to check that there are no major departures from Normality.
(b) Nonetheless, proportions calculated from a Normal distribution are not always very accurate for small numbers of observations. Find the mean x and standard deviation s for these IQ scores. What proportions of the scores are within one standard deviation and within two standard deviations of the mean? What would these proportions be in an exactly Normal distribution?
c) Standardize the data.

Normal is only approximate: Tests. Scores on a test for the 2011 high school graduating class had mean 21.2 and standard deviation 5.0. In all, 1,300,599 students in this class took the test. Of these, 149,164 had scores higher than 27 and another 50,310 had scores exactly 27. The test scores are always whole numbers. The exactly Normal $N(21.2, 5.0)$ distribution can include any value, not just whole numbers. What is more, there is *no* area exactly above 27 under the smooth Normal curve. So the test scores can be only approximately Normal. To illustrate this fact, find
(a) The percent of 2007 ACT scores greater than 27.
(b) The percent of 2007 ACT scores greater than or equal to 27.
(c) The percent of observations from the $N(21.2, 5.0)$ distribution that are greater than 27. (The percent greater than or equal to 27 is the same, because there is no area exactly over 27.)

Are the data Normal? Rains. Here are the amounts of rainfall (millimeters) for Nyahururu in the 100 years 1901 to 2000:

722.4	792.2	861.3	750.6	716.8	885.5	777.9	897.5	889.6	935.4
736.8	806.4	784.8	898.5	781.0	951.1	1004.7	651.2	885.0	719.4
866.2	869.4	823.5	863.0	804.0	903.1	853.5	768.2	821.5	804.9
877.6	803.8	976.2	913.8	843.9	908.7	842.4	908.6	789.9	853.6
728.7	958.1	868.6	920.8	911.3	904.0	945.9	874.3	904.2	877.3
739.2	793	923.4	885.8	930.5	983.6	789.0	889.6	944.3	839.9
1020	810.0	858.1	922.8	709.6	740.2	860.3	754.8	831.3	940.0
887.0	653.1	913.6	748.3	963.0	857.0	883.4	909.5	708.0	882.9
852.4	735.6	955.9	836.9	760.0	743.2	697.4	961.7	866.9	908.8
784.7	785.0	896.6	938.4	826.4	857.3	870.5	873.8	827.0	770.2

(a) Construct a stem plot for the data (IGNORE THE DECIMALS).
(b) Make a histogram of these rainfall amounts. Find the mean and the median.
(c) Although the distribution is reasonably Normal, your work shows some departure from normality. In what way are the data not Normal?

Are the data Normal? Fruit fly thorax lengths. Here are the lengths in millimeters of the thorax for 49 male fruit flies:

0.64	0.64	0.64	0.68	0.68	0.68	0.72	0.72	0.72	0.72
0.74	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.78
0.80	0.80	0.80	0.80	0.80	0.82	0.82	0.84	0.84	0.84

0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.88	0.88	0.88
0.88	0.88	0.88	0.88	0.88	0.92	0.92	0.92	0.94	

(a) Make a stem plot and a histogram of the distribution. Although the result depends a bit on your choice of classes, the distribution appears roughly symmetric with no outliers.

(b) Find the mean, median, standard deviation, and quartiles for these data. Comparing the mean and the median and comparing the distances of the two quartiles from the median suggest that the distribution is quite symmetric. Why?

(c) If the distribution were exactly Normal with the mean and standard deviation you found in (b), what proportion of observations would lie between the two quartiles you found in (b)? What proportion of the actual observations lie between the quartiles (include observations equal to either quartile value). Despite the discrepancy, this distribution is "close enough to Normal" for statistical work in later chapters.