CHAPTER 7: THE NORMAL PROBABILITY DISTRIBUTION

[This chapter is based on Chapter 7 from the textbook]

In Chapter 6 we studied discrete random variables. Specifically, we focused on distributions given by a table and binomial distribution. In this chapter we study continuous random variables. Our goal is to study the normal distribution, its properties and applications.

When we were studying discrete random variables, we computed the probability of a specific outcome using a table or a function. For example, in a binomial experiment with 10 trials and probability of success 0.3, we were able to compute the probability of obtaining exactly 4 successes using the formula:

$$P(4) = \binom{10}{4} 0.3^4 0.7^{10-4}$$

We were able to do that because the probability of exactly 4 successes makes sense. However, in continuous random variables this is no longer true. Let's consider the following example.

Example 7.1. *You agreed to meet your friend John today at noon. You've been friends for a long time, and you know that John is almost always late. Further, you know that John can be on time or up to 30 minutes late with all intervals of time equally likely.*

(a) Model John's lateness as a random variable to answer the following questions.

(b) What is the probability that John is exactly 15 minutes late?

(c) What is the probability that John is between 15 and 16 minutes late?

Solution.

(a) Since the outcome of the experiment (lateness of John) is a number, we can assign a random variable. Specifically, suppose we use $X =$"lateness of John in minutes" and the possible values as $0 \leq x \leq 30$.

(b) Since X is a continuous random variable, we have $P(X = 15) = 0$ because there is only one possible outcome in the event '$X = 15$' and there are infinitely many possible outcomes in the state space.

(c) Now we have an interval of values, so the probability won't be zero. To compute it, observe that we want $P(15 \leq X \leq 16)$ and the interval is of length 1. There are 30 possible intervals of length 1 minute between 0 and 30 minutes, and each of them has the same probability of occurring. Also, 'between 15 and 16 minutes' is one of the possible outcomes. Hence,

$$P(15 \leq X \leq 16) = \frac{1}{30}$$

□

The easiest way to represent a continuous random variable is with a graph where the area under the curve in an interval, represents the probability of the outcomes in that interval. Then, the area under the curve represents the density of probabilities. We typically represent this density with a function, and we formally define it below.

Definition 7.1. *A probability density function (pdf) is an equation used to compute probabilities of continuous random variables. It must satisfy the following properties:*

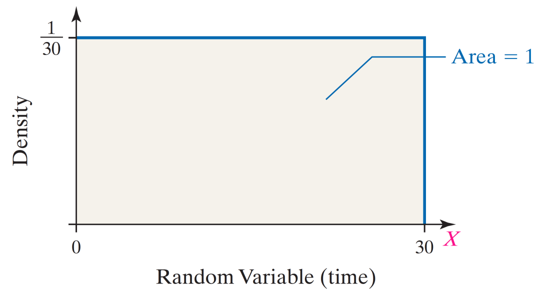
(i) The total area under the graph of the equation over all possible values of the random variable must equal 1

(ii) The height of the graph of the equation must be greater than or equal to zero for all possible values of the random variable.

Observe that the first property is similar to the property of discrete random variables saying that the probabilities must add up to 1; and the second property establishes that the probability of all possible intervals of values is nonnegative.

Example 7.2. *Graph the probability density function of the lateness of John in Example 7.1 and relate the* $P(15 \leq X \leq 16)$ *with the area under the curve.*

Solution. Since all the possible values of X have the same probability, the curve will be an horizontal line. The possible values are in the interval $[0, 30]$, so the density function will be a rectangle with possible values from 0 to 30. Since the area must be 1 and the area of a rectangle is length*height, we conclude that the height must be 1/30. Hence, we obtain the following graph:



Notice that the area under the curve in the interval $15 \leq T \leq 16$ is $\frac{1}{30}$. □

The relation between the area under the curve and the probability in the example is not a coincidence. Indeed, **the area under the graph of a density function over an interval represents the probability of observing a value of the random variable in that interval**.

Example 7.3. *The reaction time of a certain chemical process follows a uniform probability distribution between 5 and 10 minutes.*

(a) *Draw the graph of the density curve.*

(b) *What is the probability that the reaction time is between 6 and 8 minutes?*

(c) *What is the probability that the reaction time is between 5 and 8 minutes?*

(d) *What is the probability that the reaction time is less than 6 minutes?*

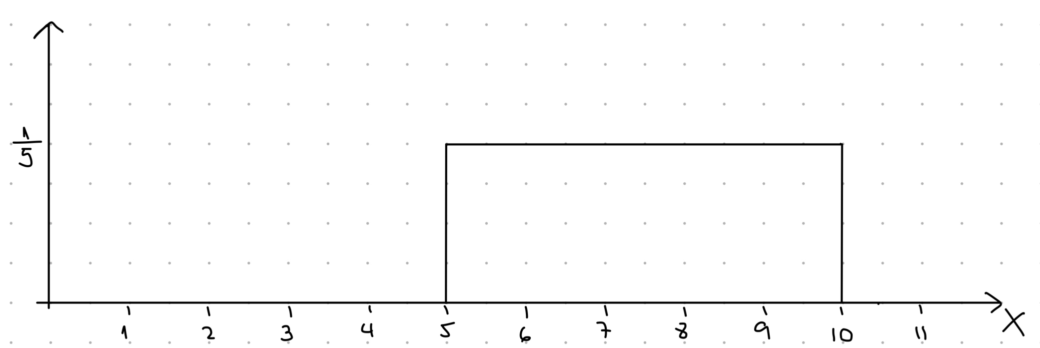
(e) *If you already waited 6 minutes, theres is a 50% probability that the reaction time will be at most* ________.

Solution. We use the random variable $X =$"reaction time of the chemical process"

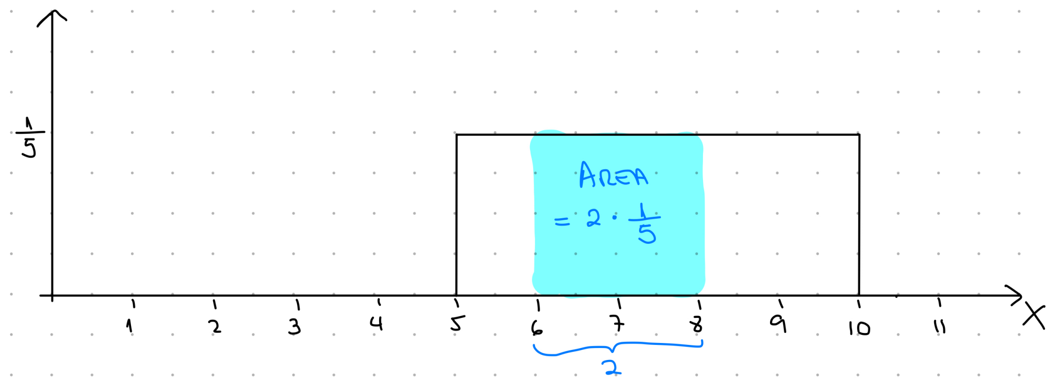
(a) We know that the chemical process takes between 5 and 10 minutes. Then, the possible values are $5 \leq x \leq 10$ and the length of the rectangle is $10 - 5 = 5$. Then, to compute the height we solve:

$$5 * \text{height} = 1 \qquad \Longrightarrow \qquad \text{height} = \frac{1}{5}$$

Hence, we obtain the following graph:



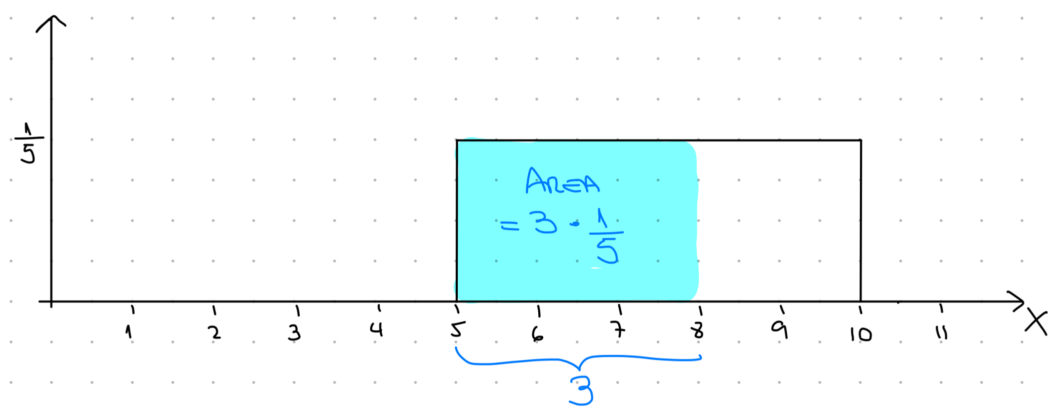
(b) We are interested in computing $P(6 \leq X \leq 8)$. That is, we compute the area of the rectangle of the following figure:



Since the length of the rectangle is 2 (and the height is always 1/5) we obtain:

$$P(6 \leq X \leq 8) = \frac{2}{5} = 0.4$$

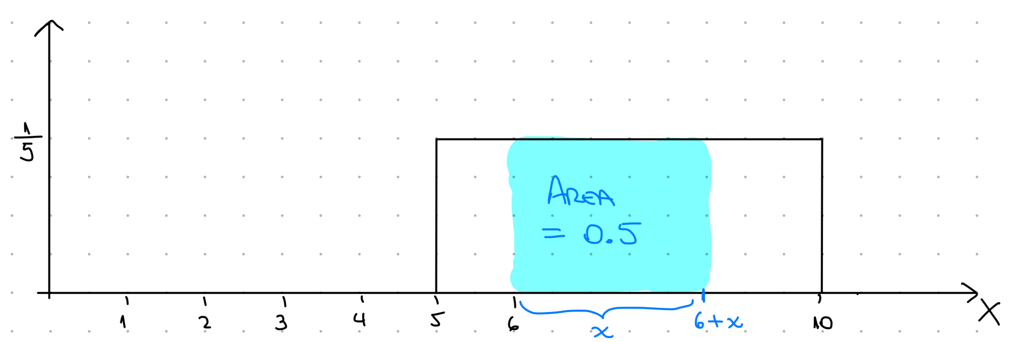
(c) This part is similar to part (b), but now the length of the interval is 3. We obtsain the following picture:



Therefore, the probability that the reaction takes between 5 and 8 minutes is:

$$P(5 \leq X \leq 8) = \frac{3}{5}$$

(d) In this part we know the are and the lower limit of the interval, and we need to compute the upper limit. In terms of the graph, we have the following:



Then, we solve the following equation

$$\frac{1}{5} * x = \frac{1}{2} \qquad \Longrightarrow \qquad x = \frac{5}{2} = 2.5$$

The interval starts at 6 and the length of the interval is 2.5. Then, the upper limit of the interval is 6+2.5 = 8.5. Hence, with 50% probability the reaction time will be at most 8.5 minutes if we are already in minute 6. Mathematically, we have

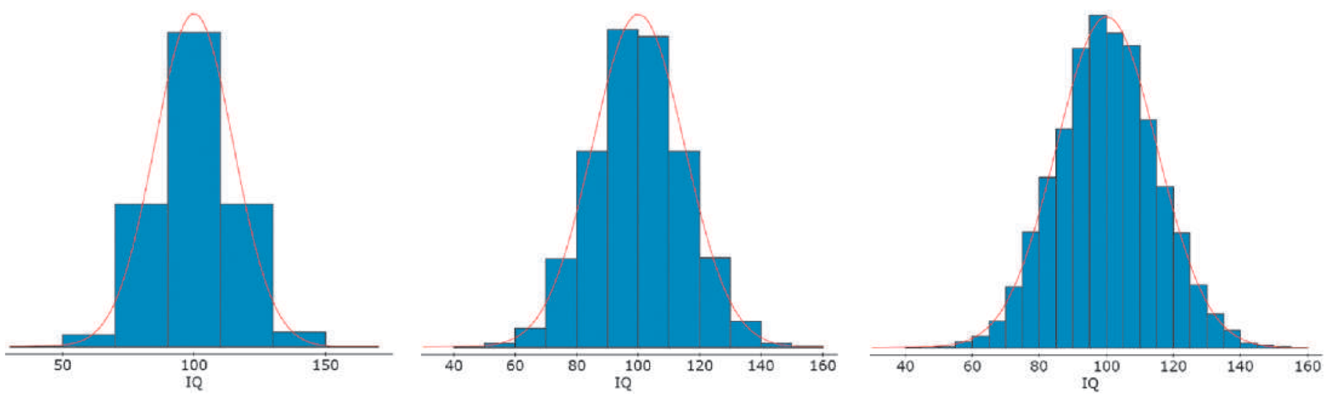
$$P(6 \leq X \leq 8.5) = 2.5$$

□

7.1 Properties of the Normal Distribution

The normal distribution is a continuous distribution where the shape of the curve is a bell. That means that if we draw a histogram of the data and we draw a bell-shaped curve on top of it, they look pretty similar.

Example 7.4. *The following histograms represent the IQ scores of 10,000 randomly selected adults. The three histograms represent the same data with different class widths, and the red curve is a normal density function.*



(a) **(b)** **(c)**

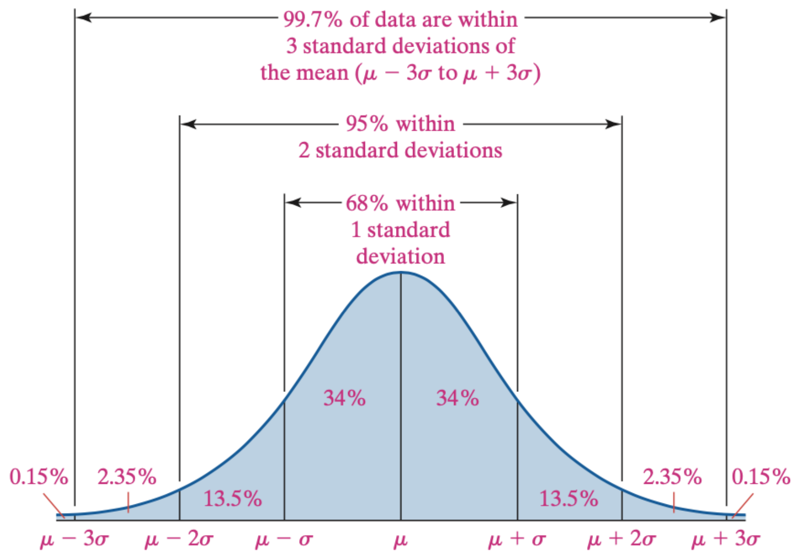
Notice that as the number of bins increases, the curve is more and more representative of the histogram. Then, we can use a normal model to study the IQ scores.

Definition 7.2. *In mathematics, a* <u>*model*</u> *is an equation, table or graph used to describe reality.*

The next question is how to know if we can use a normal distribution to model a random variable. To answer that question, we first need to know the normal distribution properties.

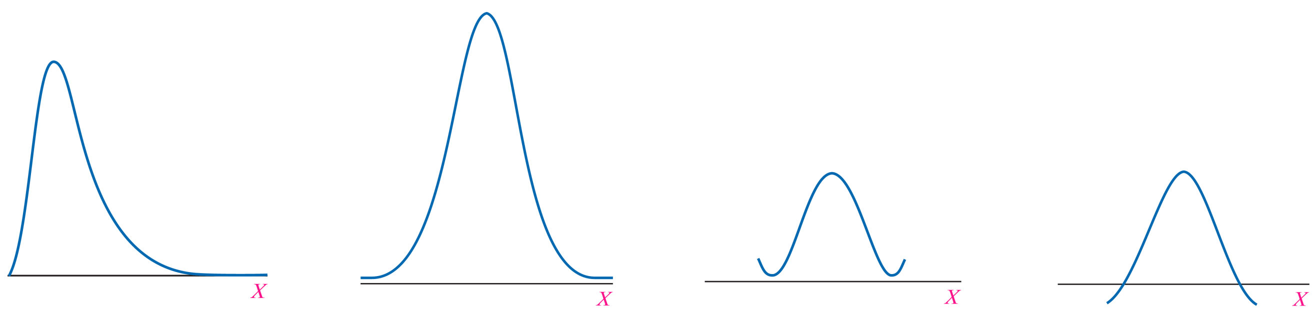
Theorem 7.1 (Properties of the normal distribution)**.** *The normal probability density function satisfies the following properties:*

1. *The normal curve is symmetric about its mean μ*
2. *The normal curve has a single peak at $x = \mu$ because the mean equals the mode and the median*
3. *The area under the curve is 1*
4. *The area under the curve to the right of μ equals the area under the curve to the left of μ, and both equal $\frac{1}{2}$.*
5. *As x increases, the graph approaches 0 (but neves reaches or crosses 0). Similarly as x decreases.*
6. *The normal curve follows the empirical rule represented in the following picture:*



Then, if the relative frequency histogram of our random variable satisfies these properties, we can say that the random variable follows a normal distribution.

Example 7.5. *Determine if the following histograms can be modeled as a random variable with normal distribution. In the cases where the model does not fit, indicate why.*



(a) (b) (c) (d)

Solution. We go one by one:

(a) No because it is not symmetric

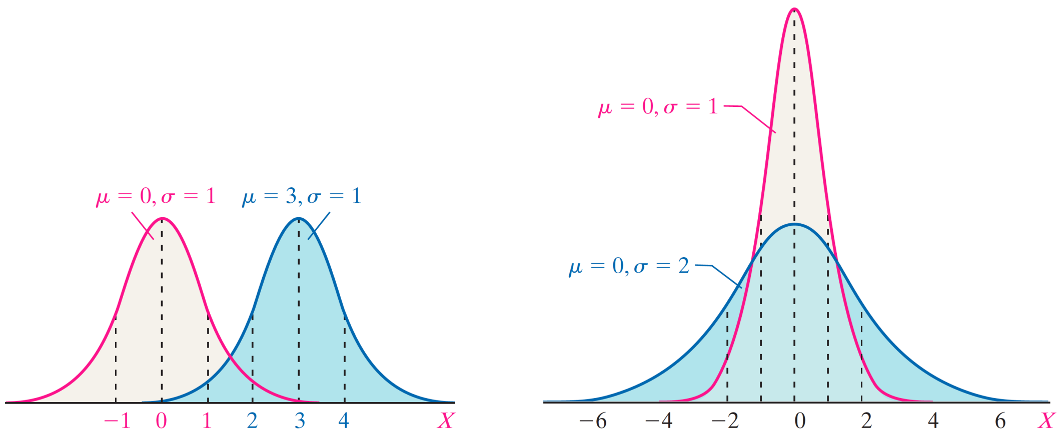
(b) Yes

(c) No because the tails to the right and left of the center do not approach zero (property 5)

(d) No because there are negative values

□

Changes in the mean μ and the standard deviation σ notoriously change the shape and location of a normal distribution, as indicated in the following picture.



(a) (b)

In panel (a), we see that increasing the mean from $\mu = 0$ to $\mu = 3$ shifted the graph to the right, but the shape was maintained. In panel (b), we see that increasing the standard deviation from $\sigma = 1$ to $\sigma = 2$ caused the graph to become wider and shorter, but the location of the center is the same.

7.2 Applications of the Normal Distribution

In this section we learn how to compute probabilities associated to random variables with normal distribution. We already know that the probability that a random variable takes a value in an interval $[a, b]$ is the area under the probability density function (pdf) between a and b. With the uniform distribution, computing this number was very easy. However, computing the area under the normal pdf is more complicated. Hence, we will use a table.

But, how do we use a table if different random variables have different mean and standard deviation? We will only use a table for a normal random variable with mean $\mu = 0$ and $\sigma = 1$, and we will use z-scores to rescale any normally distributed random variable, as follows.

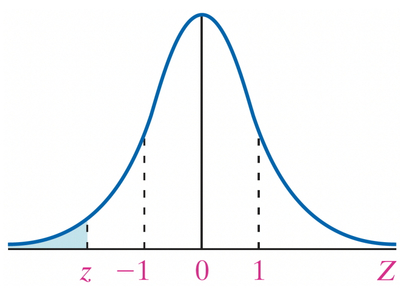
Definition 7.3. *A random variable Z, normally distributed with mean $\mu = 0$ and $\sigma = 1$ is said to have standard normal distribution.*

Theorem 7.2. *Suppose that the random variable X is normally distributed with mean μ and standard deviation σ. Define the random variable Z as:*

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

Then, Z has standard normal distribution.

The following tables (Table V from Appendix A in the textbook) shows the cumulative distribution function (cdf) of a standard normal distribution, that is, it shows $P(Z \leq z)$ for different values of z. Pictorially, the table shows the blue area in the following picture:



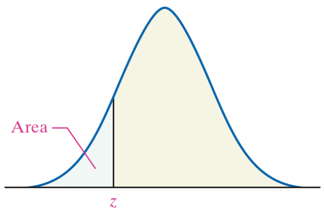


Table V

Standard Normal Distribution

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
−0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

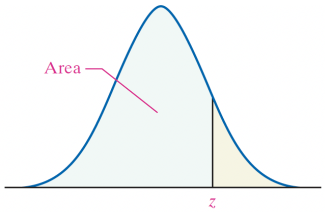


Table V (*continued*)

Standard Normal Distribution

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

The rows represent units and first decimal, and the columns represent the second decimal. Then, if we want to compute $P(Z \leq -1.53)$ we go to the first table and find the row corresponding to -1.50 and the column corresponding to 0.03 as shown in the figure below. We obtain $P(Z \leq -1.53) = 0.0630$.

z	0.00	0.01	0.02	0.03	0.04	0.05
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606

Similarly, if we want $P(Z \leq 0.95)$ we go to the second table and find the row corresponding to 0.90 and the column corresponding to 0.05 as shown in the figure below. We obtain $P(Z \leq 0.95) = 0.8289$.

z	0.00	0.01	0.02	0.03	0.04	0.05
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289

As we have learned throughout the last chapters, we can use the cdf to compute plenty of probabilities. In the following example we learn how to use the normal tables to compute the probability of “at most,” “at least,” “between” and to compute percentiles.

Example 7.6. *A pediatrician obtains the heights of her three-year-old female patients. The heights are approximately normally distributed, with mean 38.72 in and standard deviation 3.17 in.*

(a) What is the proportion of patients that have height less than 35 in?

(b) What is the proportion of patients that have height more than 41 in?

(c) What is the proportion of patients that have height between 30 in and 40 in?

(d) What is the height of a three-year-old female at the 20th percentile?

Solution. Let’s call X the height of the patients, which follow a normal distribution with mean $\mu = 38.72$ and standard deviation $\sigma = 3.17$. Then,

$$Z = \frac{X - 38.72}{3.17}$$

follows a standard normal distribution. We convert every value of X in the questions to a value of Z and use the tables.

(a) We are interested in computing $P(X \leq 35)$. The first step is to compute the z-score of $x = 35$. We obtain:

$$z = \frac{35 - 38.72}{3.17} = -1.17$$

Then,

$$P(X \leq 35) = P(Z \leq -1.17) = 0.1210$$

where we computed the last number using the table. To find $z = -1.174$ in the normal table we look at the first table because it is a negative number. Then, the probability is in the row corresponding to -1.1 and the column corresponding to 0.07, as shown below:

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379

Therefore, the probability that a patient's height is less than 35 in is 12.10%.

(b) We are now interested in $P(X \geq 41)$. Since we only know how to compute the cdf, we use the complement rule and compute:

$$P(X \geq 41) = 1 - P(X < 41)$$

Observe that the random variable is continuous, so $P(X < 41) = P(X \leq 41)$ because the probability of the individual value $X = 41$ is 0.

We now compute the z-score of $x = 41$:

$$z = \frac{41 - 38.72}{3.17} = 0.72$$

Then,

$$\begin{aligned} P(X \geq 41) &= 1 - P(X \leq 41) \qquad \text{(complement rule as stated above)} \\ &= 1 - P(Z \leq 0.72) \qquad (z\text{-score}) \\ &= 1 - 0.7642 \qquad \text{(table)} \qquad = 0.2358 \end{aligned}$$

where the step "table" comes from finding 0.72 in the second table because it is a positive number. From the table, we have:

z	0.00	0.01	0.02	0.03	0.04	0.05
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289

Therefore, the probability that a patient's height is more than 41 in is 23.58%.

(c) We now compute the probability $P(30 \leq X \leq 40)$. We know that the probability of an interval is the difference of the cdfs, as follows:

$$P(30 \leq X \leq 40) = P(X \leq 40) - P(X \leq 30)$$

We compute each of the cdfs separately.

- We start with $P(X \leq 40)$. Computing its z-score and finding the probability in the table, we obtain:

$$\begin{aligned} P(X \leq 40) &= P\left(Z \leq \frac{40 - 38.72}{3.17}\right) \\ &= P(Z \leq 0.40) \\ &= 0.6554 \end{aligned}$$

  The picture below shows how we found the value in the (second) table:

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772

- Similarly, we compute $P(X \leq 30)$ as follows:

$$\begin{aligned} P(X \leq 30) &= P\left(Z \leq \frac{30 - 38.72}{3.17}\right) \\ &= P(Z \leq -2.75) \\ &= 0.003 \end{aligned}$$

  The picture below shows how we found the value in the (first) table:

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051

Therefore, the probability that a patient's height is between 30 in and 40 in is:

$$\begin{aligned} P(30 \leq X \leq 40) &= P(X \leq 40) - P(X \leq 30) \\ &= P(Z \leq 0.40) - P(Z \leq -2.75) \\ &= 0.6554 - 0.003 \\ &= 0.6551 \end{aligned}$$

(d) In the last question, we need to compute x so that $P(X \leq x) = 0.2$. To do that, we search for the value 0.2 in the probabilities given by the tables (not the row and column as before). This will give us the 20th percentile of z. Then, we compute the corresponding value of x using the definition of z-score. Let's go step by step.

Observe that the first table has all the probabilities ≤ 0.5 and the second table, all the probabilities ≥ 0.5. Then, we search in the first table:

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
−0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

None of the probabilities equals exactly 0.2, but the closest is 0.2005, which corresponds to $z = -0.84$. Then, we compute x so that the z-score is -0.84, that is, we solve:

$$-0.84 = \frac{x - 38.72}{3.17} \qquad \Longrightarrow \qquad x = (-0.84) \cdot 3.17 + 38.72 = 36.06$$

Hence, the 20th percentile of height is 36.06 in, that is, 20% of the patients are smaller than 36.06 in.

□