

## Normal Distribution Worksheet

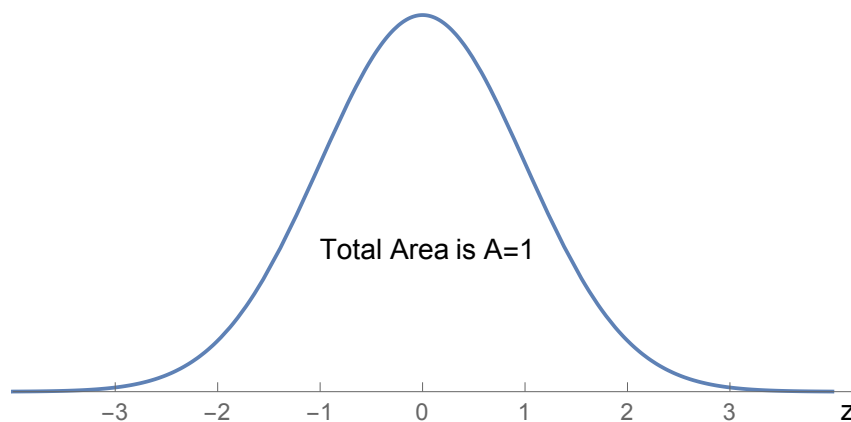
This worksheet provides coverage of **normally distributed data distributions**. These notes are self-contained and do not require you to have the textbook on hand.

### 1. INTRODUCTION

Normally distributed data sets are common in the real world. Real world observations that represent sums or averages of other numbers often have histograms that are shaped like a normal data distribution. Applications include but are not limited to business, finance, science and policy.

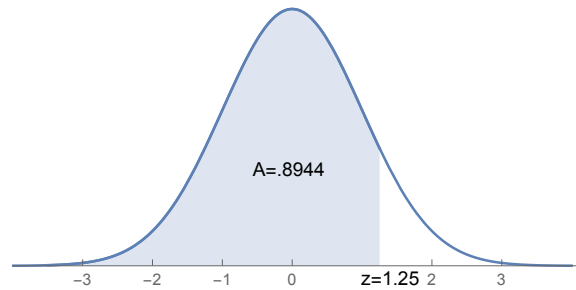
### 2. STANDARD NORMAL DISTRIBUTION ( $z$ -DISTRIBUTION)

A key tool for calculating proportions and percentages associated with normal distributions is the **standard normal distribution**. The standard normal distribution is a bell-shaped planar region bounded above by a bell-shaped curve and below by a horizontal line representing the  $z$ -values of the normal distribution.

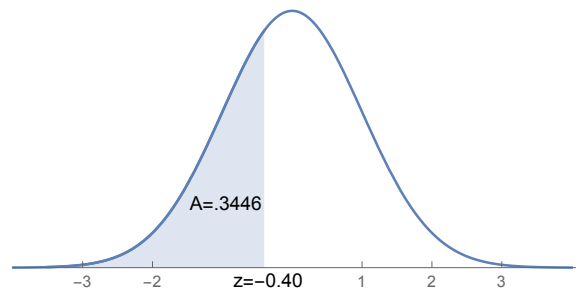


Possible data values of a standard normal random variable are indicated on the horizontal line of  $z$ -values (in class, I often say that the horizontal axis shows the “real world values”). The curve above it represents the likelihood of the values occurring in the sense that the area under the curve to the left of a specific  $z$ -value on the horizontal axis represents the percentage of real world values that are less than that particular  $z$ -value. To approximate these data percentages (e.g. areas under the curve), we use the  $z$  table provided on the last page of this worksheet.

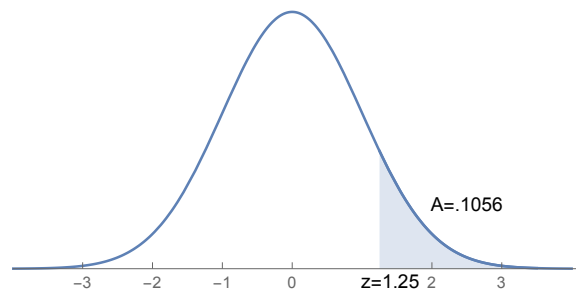
For example if we focus on the  $z$ -value given by  $z = 1.25$ , the area under the curve left of  $z = 1.25$ , pictured below, is the percentage of data less than  $z = 1.25$ ; using the table, we get that this area is  $A = .8944$ .



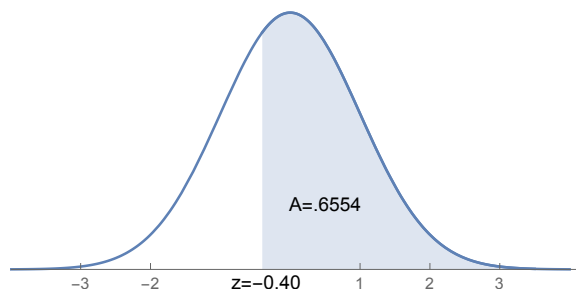
For another example if we focus on the  $z$ -value given by  $z = -0.40$ , the area under the curve left of  $z = -0.40$ , pictured below, is the percentage of data less than  $z = -0.40$ ; using the  $z$ -table, we get that this area is  $A = .3446$ .



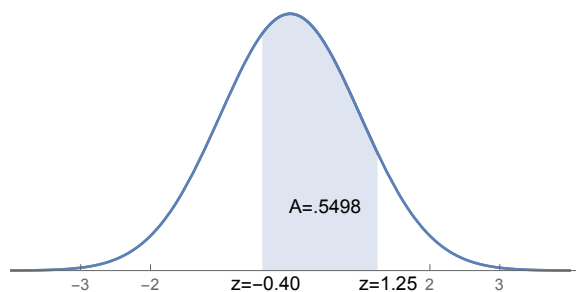
We can also use the complement rule with the  $z$ -table to get areas to the right of a specified  $z$ -value. Since the total area under the curve is 1, the area to the right of a specified  $z$ -value is the area to the left subtracted from 1. For example, the area to the right of  $z = 1.25$ , pictured below, is given by  $1 - .8944 = .1056$



For another example, the area to the right of  $z = -0.40$ , pictured below, is given by  $1 - .3466 = .6554$



Finally, we can determine areas under the normal curve between *two* specified  $z$ -values by subtracting the area to the left of the smaller  $z$ -value from the area to the left of the larger  $z$ -value. For example, if we want to find the area between  $z = -0.4$  and  $z = 1.25$ , we subtract the area to the left of  $z = -0.40$  from the area to the left of  $z = 1.25$ ; in this case, we find that the area between  $z = -0.4$  and  $z = 1.25$  is given by  $.8944 - .3446 = .5498$ , as pictured below.



### 3. NORMAL DISTRIBUTION WITH MEAN $\mu$ AND STANDARD DEVIATION $\sigma$

We now turn our attention to data sets with a normal distribution with a specified mean (e.g. average measurement) and a specified standard deviation; remember, the standard deviation of a data set measures the spread of a data set in the sense that *almost all observations in a data set are within 3 standard deviation of the mean*. For data sets with normal distributions, “almost all of the data” means “about 99% of the data.”

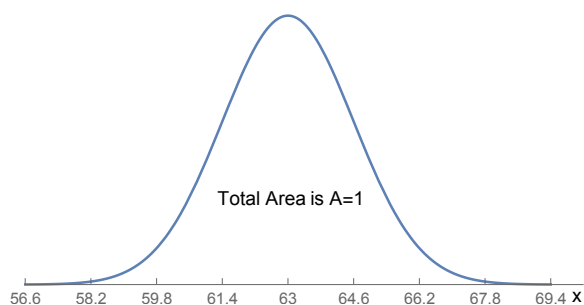
To calculate areas associated with non-standard normal distributions, we will need to use the standardization formula for  $z$ -scores,

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

This formula will be provided on the front of Exam 4. The formula calculates how many standard deviations a real world measurement,  $x$ , is from the mean of a data distribution. The *standard* normal distribution has mean  $\mu = 0$  and standard deviation  $\sigma = 1$ ; for real-world normal distributions, the mean, symbolized by  $\mu$ , can be any number and the standard deviation, symbolized by  $\sigma$ , can be any positive number.

Focusing on a specific example will help us understand the concept. Consider the length of gestation (i.e. pregnancy duration) of a mother dog of the Boxer breed that is carrying puppies. It is known that lengths of gestation for Boxers are approximately normally distributed with a mean of 63 days and a standard deviation of 1.6 days.

The first step in a problem covering a real-world normal distribution is to **draw a picture**. A sketch of the normal distribution that is centered at 63 and spread out with standard deviations that go in “steps” of 1.6 is given below. In this picture of the boxer lengths of gestations normal curve, the mean and the three standard deviations above and below the mean are labelled on the  $x$ -axis.



We will refer to this picture when we work on the specific problems about Boxer gestation lengths in our first example.

**Example 1:** Gestation lengths (in days) of Boxers form a normally distributed data set with mean  $\mu = 63$  and standard deviation  $\sigma = 1.6$ .

- (1) Determine the probability that a pregnant Boxer’s gestation length is at least 65 days long.
- (2) Determine the probability that a pregnant Boxer’s gestation length is at most 60.5 days long.
- (3) Determine the probability that a pregnant Boxer’s gestation length is between 63 and 65 days long.
- (4) Determine the probability that a pregnant Boxer’s gestation length is between 59 and 63.5 days long.

*Solution to Example 1:* First, the mean of this normal distribution is  $\mu = 63$  and the standard deviation is  $\sigma = 1.6$ , so the  $z$ -score standardization formula is

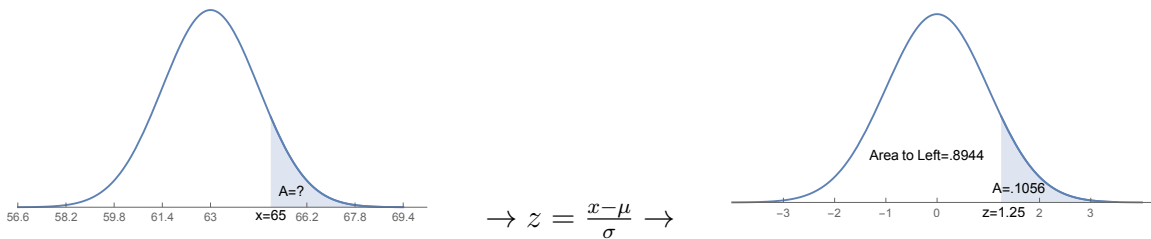
$$z = \frac{x - \mu}{\sigma} = \frac{x - 63}{1.6}$$

The solutions of each part are itemized below

- (1) For the possible value  $x = 65$ , we find that

$$z = \frac{x - 63}{1.6} = \frac{(65 - 63)}{1.6} = 1.25$$

Looking up the value  $z = 1.25$  in our normal distribution table we find that the area *to the left* of  $z = 1.25$  is given by .8944. Thus, the area to the right that we are trying to identify (see picture on next page) is given by .1056.



This completes the problem! We completed two main steps: (i) Change  $x$ -value to a  $z$ -value and (ii) Use the  $z$ -table to determine the area.

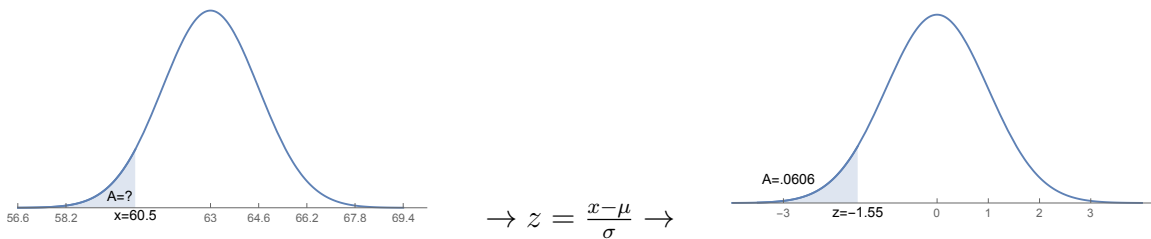
Another way of expressing our solution is “10.56% of Boxer gestation lengths are at least 65 days long.”

- (2) For the possible value  $x = 60.5$ , we find that

$$z = \frac{x - 63}{1.6} = \frac{(60.5 - 63)}{1.6} = -1.563$$

This exact  $z$ -value does not show up in our chart, so, in order to use our chart, we round this number to  $z \approx -1.55$ .

Looking up the value  $z = -1.55$  in our normal distribution table we find that the area *to the left* of  $z = -1.55$  is given by .0606. Thus, the area to the left that we are trying to identify (see picture below) is approximately .0606.



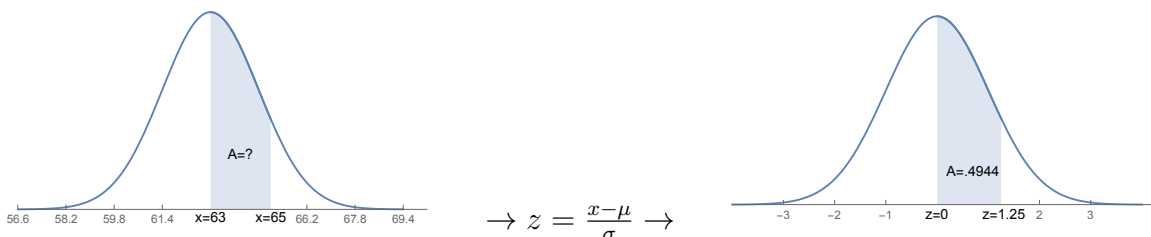
Another way of expressing our solution is “6.06% of Boxer gestation lengths are at least 65 days long.”

- (3) For the possible value  $x = 65$ , we already found in part (a) that the corresponding  $z$ -value is 1.25 and that the area *to the left* of  $z = 1.25$  is given by .8944.

For the possible value  $x = 63$ , we find that

$$z = \frac{x - 63}{1.6} = \frac{(63 - 63)}{1.6} = 0$$

As you might guess from the picture and the  $z$ -table confirms, the area to the left of  $z = 0$  is 0.5, subtracting the area to the left of  $z = 1.25$  from the area left of  $z = 0$  gives us the solution,  $A = .8944 - .5 = .4944$ , pictured below.



(4) For the possible value  $x = 59$ , we find that

$$z = \frac{x - 63}{1.6} = \frac{(59 - 63)}{1.6} = -2.5$$

Looking up the value  $z = -2.5$  in our normal distribution table we find that the area *to the left* of  $z = -2.5$  is given by .0062.

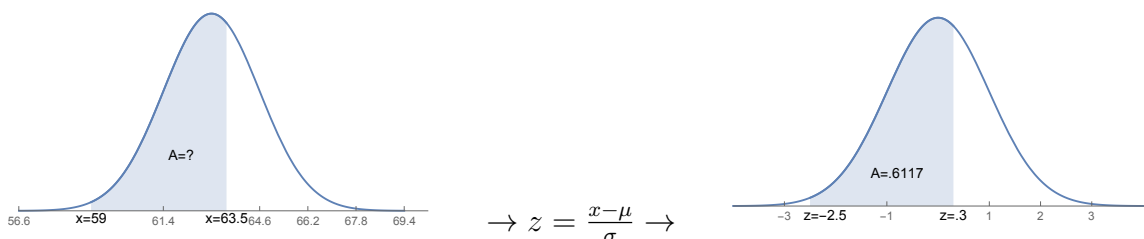
For the possible value  $x = 63$ , we find that

$$z = \frac{x - 63}{1.6} = \frac{(63 - 63)}{1.6} = .3125$$

This exact  $z$ -value does not show up in our chart, so, in order to use our chart, we round this number to  $z \approx .3$

Looking up the value  $z = .3$  in our normal distribution table we find that the area *to the left* of  $z = .3$  is given by .6179

Subtracting the area to the left of  $z = -2.5$  from the area left of  $z = .3$  gives us the solution,  $A = .6179 - .0062 = .6117$ , pictured below.



**Example 2:** A commercial music songwriter is paid commissions according to how many times the commercial they scored is aired on network television. Assume that the number of times the commercial using their music will air is normally distributed with a mean of 1200 and standard deviation of 300.

- (1) Compute the probability that the commercial airs at most 1200 times.
- (2) Compute the probability that the commercial airs between 1200 and 1500 times.
- (3) Compute the probability that the commercial airs more than 1500 times.
- (4) Compute the probability that the commercial airs between 1000 and 1500 times.

*Solution:* First, the mean of this normal distribution is  $\mu = 1200$  and the standard deviation is  $\sigma = 300$ , so the  $z$ -score standardization formula is

$$\frac{x - \mu}{\sigma} = \frac{x - 1200}{300}$$

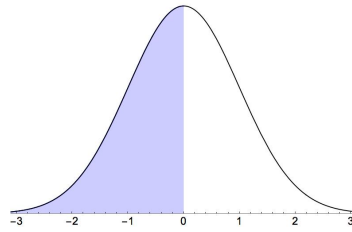
The solutions of each part are itemized below

- (1) For the possible value  $x = 1200$ , we find that

$$\frac{x - 1200}{300} = \frac{1200 - 1200}{300} = 0$$

This might seem confusing, but the mean of standard normal distribution *is* 0. The area under the part of the standard normal distribution curve associated with possible values less than or equal to  $z = 0$ , pictured below, is 0.5 since the geometry of the

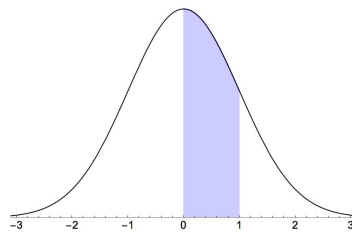
normal distribution is symmetric about  $z = 0$ . There is a 50% chance that the commercial airs at most 1200 times.



- (2) For the possible value  $x = 1500$ , we find that

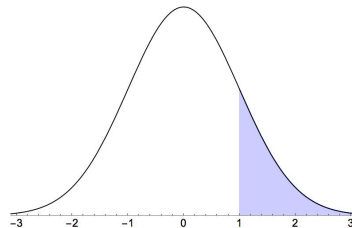
$$\frac{x - 1200}{300} = \frac{1500 - 1200}{300} = 1$$

The area under the part of the standard normal distribution curve between the possible values of  $z = 0$  and  $z = 1.00$ , pictured below, can be calculated to be 0.3413 using our  $z$ -table. There is a 34.13% chance that the commercial airs between 1200 and 1500 times.



- (3) We know from the previous part that the possible value  $x = 1500$  corresponds to the  $z$ -score,  $z = 1$ .

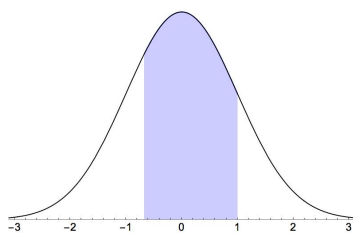
The area under the part of the standard normal distribution curve to the right of  $z = 1.00$ , pictured below, can be calculated to be 0.1587. There is a 15.87% chance the commercial airs more than 1500 times.



- (4) For the possible value  $x = 1000$ , we find that

$$\frac{x - 1200}{300} = \frac{1000 - 1200}{300} = -.67$$

Previously, we found that the  $z$ -score for  $x = 1500$  was  $z = 1$ . The area under the part of the standard normal distribution curve between the possible values of  $z \approx -.65$  and  $z = 1.00$ , pictured below, can be calculated to be  $.8413 - .2578 = 0.5835$  using our  $z$ -table. There is a 58.35% chance that the commercial airs between 1000 and 1500 times.



**Example 3:** A musical act expects the number of tickets sold for an upcoming concert to be normally distributed with a mean of 450 tickets and a standard deviation of 60 tickets

- (1) Compute the probability that between 400 and 450 tickets are sold.
- (2) Compute the probability that less than 300 tickets are sold.
- (3) Compute the probability that at least 550 tickets are sold
- (4) Compute the probability between 400 and 500 tickets are sold

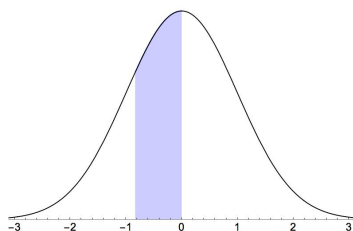
*Solution:* The mean of this normal distribution is  $\mu = 450$  and the standard deviation is  $\sigma = 60$ , so the  $z$ -score standardization formula is

$$\frac{x - \mu}{\sigma} = \frac{x - 450}{60}$$

- (1) For the possible values  $x = 400$  and  $x = 450$ , we find that

$$\frac{x - 450}{60} = \frac{400 - 450}{60} = -0.83 \quad \text{and} \quad \frac{x - 450}{60} = \frac{450 - 450}{60} = 0$$

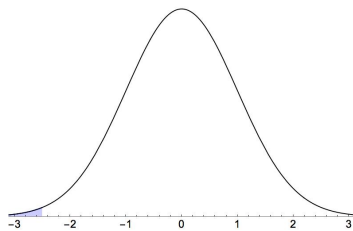
The area under the part of the standard normal distribution curve between the possible values of  $z \approx -0.85$  and  $z = 0$ , pictured below, are calculated as  $.5 - .1977 = .3023$  using our  $z$ -table. There is a 30.23% chance between 400 and 450 tickets are sold.



- (2) For the possible value  $x = 300$ , we find that

$$\frac{x - 450}{60} = \frac{300 - 450}{60} = -2.5$$

The area under the part of the standard normal distribution curve to the left of the possible value  $z = -2.50$ , pictured below, can be calculated to be 0.0062 using our  $z$ -table. There is a 0.62% chance that less than 300 tickets are sold.

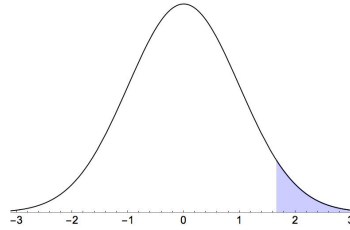




(3) For the possible value  $x = 550$ , we find that

$$\frac{x - 450}{60} = \frac{550 - 450}{60} = 1.67$$

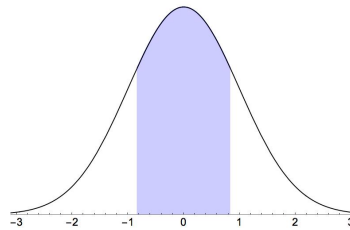
The area under the part of the standard normal distribution curve to the right of the possible value  $z \approx 1.65$  can be calculated to be 0.9505 using our  $z$ -table. There is a 4.95% chance that at least 550 tickets are sold.

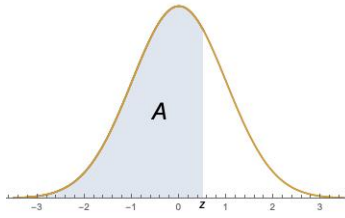


(4) For the possible value  $x = 500$ , we find that

$$\frac{x - 450}{60} = \frac{500 - 450}{60} = .83$$

Previously, we found that the  $z$ -score for  $x = 400$  was  $z \approx -.86$ . The area under the part of the standard normal distribution curve between the possible values of  $z \approx -.83$  and  $z \approx .85$ , pictured below, can be calculated to be  $.8023 - .1977 = .6046$  using our  $z$ -table. There is a 60.46% chance that between 400 and 500 tickets are sold.





**Key:**  $z$ -score=  $z$ -value       $A$ =Area of normal distribution to the left of  $z$

<b>z-score</b>	<b>A</b>	<b>z-score</b>	<b>A</b>	<b>z-score</b>	<b>A</b>	<b>z-score</b>	<b>A</b>
-3.	0.0013	-1.5	0.0668	0.	0.5	1.5	0.9332
-2.95	0.0016	-1.45	0.0735	0.05	0.5199	1.55	0.9394
-2.9	0.0019	-1.4	0.0808	0.1	0.5398	1.6	0.9452
-2.85	0.0022	-1.35	0.0885	0.15	0.5596	1.65	0.9505
-2.8	0.0026	-1.3	0.0968	0.2	0.5793	1.7	0.9554
-2.75	0.003	-1.25	0.1056	0.25	0.5987	1.75	0.9599
-2.7	0.0035	-1.2	0.1151	0.3	0.6179	1.8	0.9641
-2.65	0.004	-1.15	0.1251	0.35	0.6368	1.85	0.9678
-2.6	0.0047	-1.1	0.1357	0.4	0.6554	1.9	0.9713
-2.55	0.0054	-1.05	0.1469	0.45	0.6736	1.95	0.9744
-2.5	0.0062	-1.	0.1587	0.5	0.6915	2.	0.9772
-2.45	0.0071	-0.95	0.1711	0.55	0.7088	2.05	0.9798
-2.4	0.0082	-0.9	0.1841	0.6	0.7257	2.1	0.9821
-2.35	0.0094	-0.85	0.1977	0.65	0.7422	2.15	0.9842
-2.3	0.0107	-0.8	0.2119	0.7	0.758	2.2	0.9861
-2.25	0.0122	-0.75	0.2266	0.75	0.7734	2.25	0.9878
-2.2	0.0139	-0.7	0.242	0.8	0.7881	2.3	0.9893
-2.15	0.0158	-0.65	0.2578	0.85	0.8023	2.35	0.9906
-2.1	0.0179	-0.6	0.2743	0.9	0.8159	2.4	0.9918
-2.05	0.0202	-0.55	0.2912	0.95	0.8289	2.45	0.9929
-2.	0.0228	-0.5	0.3085	1.	0.8413	2.5	0.9938
-1.95	0.0256	-0.45	0.3264	1.05	0.8531	2.55	0.9946
-1.9	0.0287	-0.4	0.3446	1.1	0.8643	2.6	0.9953
-1.85	0.0322	-0.35	0.3632	1.15	0.8749	2.65	0.996
-1.8	0.0359	-0.3	0.3821	1.2	0.8849	2.7	0.9965
-1.75	0.0401	-0.25	0.4013	1.25	0.8944	2.75	0.997
-1.7	0.0446	-0.2	0.4207	1.3	0.9032	2.8	0.9974
-1.65	0.0495	-0.15	0.4404	1.35	0.9115	2.85	0.9978
-1.6	0.0548	-0.1	0.4602	1.4	0.9192	2.9	0.9981
-1.55	0.0606	-0.05	0.4801	1.45	0.9265	2.95	0.9984
						3.	0.9987