Chapter 9

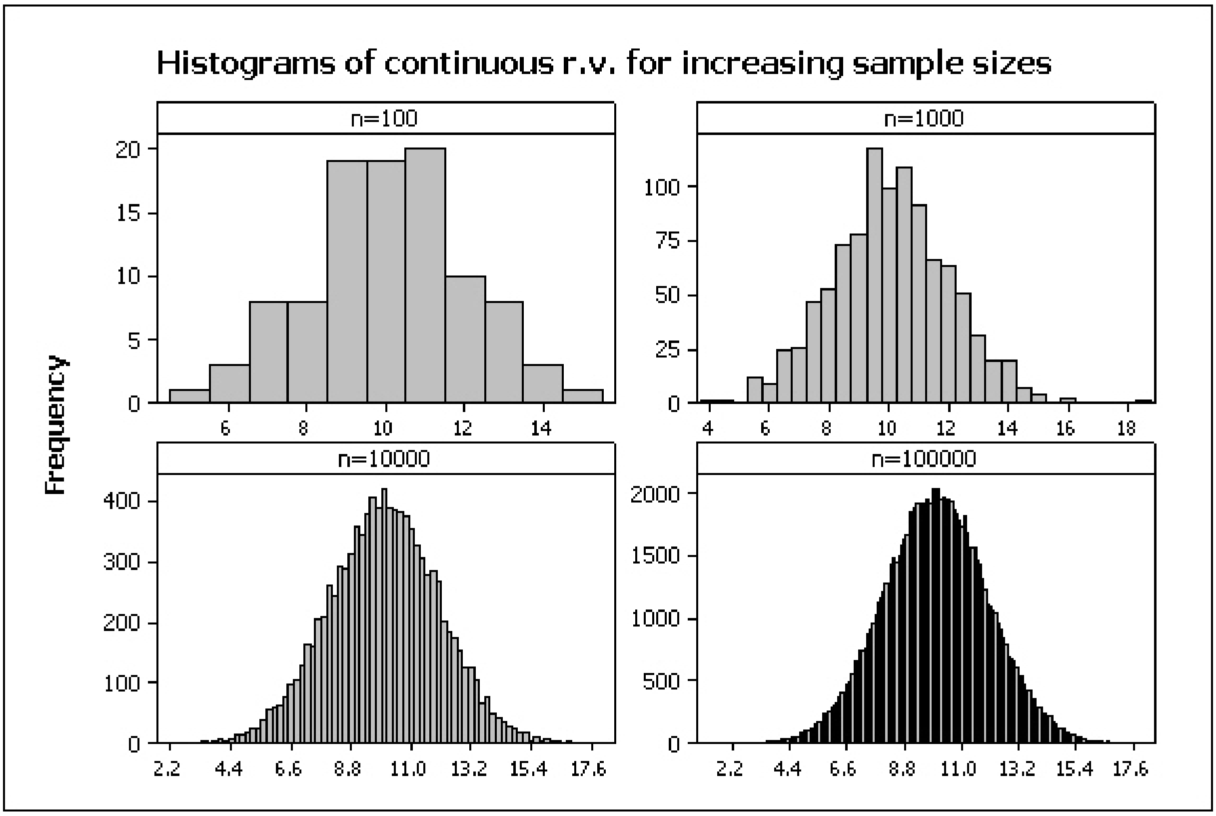
Continuous Probability Models

9.1 Introduction

We have seen how *discrete* random variables can be modelled by discrete probability distributions such as the binomial and Poisson distributions. We now consider how to model *continuous* random variables.

A variable is discrete if it takes a *countable* number of values, for example, $r = 0, 1, 2, \ldots, n$ or $r = 0, 1, 2, \ldots$ or $r = 0, 0.1, 0.2, \ldots, 0.9, 1.0$. In contrast, the values which a continuous variable can take form a continuous scale. One simple example of a continuous variable is height. Although in practice we might only record height to the nearest cm, if we could measure height exactly (to billions of decimal places) we would find that everyone had a different height. This is the essential difference between discrete and continuous variables. Therefore, if we could measure the exact height of every one of the n people on the planet, we would find that, for any height x, the proportion of people of height x is either $1/n$ or 0. And if we imagine the number of people on the planet growing over time ($n \to \infty$), this proportion tends to zero. This feature poses a problem for modelling continuous random variables as we can no longer use the methods we have seen work for discrete random variables.

The solution can be found by considering a (relative frequency) histogram of a sample of values taken by the continuous random variable, and thinking about what happens to the histogram as the sample size increases. For example, consider the following graphs which show histograms for samples of 100, 1000, 10000 and 100000 observations made on a continuous random variable which can take values between 0 and 20. The final graph shows what happens when the sample size becomes infinitely big. This final graph is called the *probability density function*.



As the population size gets larger, the histogram intervals get smaller and the jagged profile of the histogram smooths out to become a curve. We call this curve the *probability density function (pdf)* and it is usually written as $f(x)$. Note that probabilities such as $P(X < x)$ can be determined using the pdf as they equate to areas under the curve.

The key features of pdfs are

1. pdfs never take negative values
2. the area under a pdf is one: $P(-\infty < X < \infty) = 1$
3. areas under the curve correspond to probabilities
4. $P(X \leq x) = P(X < x)$ since $P(X = x) = 0$.

Over the next two weeks we will consider some particular probability distributions that are often used to describe continuous random variables. We start with the most important, most widely–used statistical distribution of all time...

9.2 The Normal Distribution

9.2.1 Introduction

The *normal* distribution is possibly the best know and most used continuous probability distribution. It provides a good model for data in so many different applications – for example, the level of rainfall on a particular day, the height of people in a class, the IQ levels of the population as a whole. The outcomes of many production processes also follow normal distributions and hence it is used widely in industry.

Recall the "parameters" of the binomial and Poisson distributions: the binomial distribution has two parameters, n and p, and the Poisson distribution has one parameter, λ. The normal distribution has two parameters: the mean, μ, and the standard deviation, σ. Its probability density function (pdf) has a "bell shaped" profile:

The formula for the pdf is

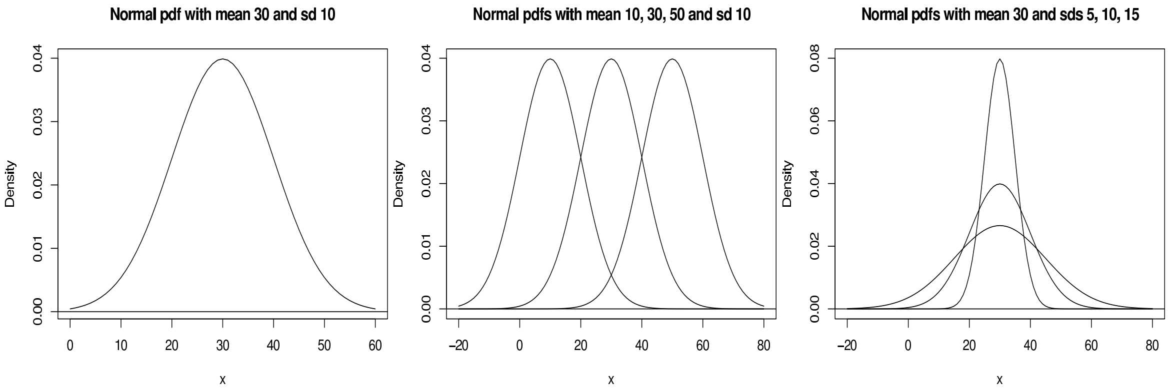
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}.$$

Unlike the binomial and Poisson distributions, there is no simple formula for calculating probabilities. However, they can be determined using tables (see tables at the end of this chapter) or statistical packages such as `Minitab`.

There are four important characteristics of the normal distribution:

1. It is symmetrical about its mean, μ.
2. The mean, median and mode all coincide.
3. The area under the curve is equal to 1.
4. The curve extends in both directions to infinity (∞).

Below are plots of the pdf for normal distributions with different values of μ and σ:



Note that the mean μ locates the distribution on the x–axis and the standard deviation σ affects the spread of the distribution, with larger values giving flatter and wider curves.

9.2.2 Notation

If a random variable X has a normal distribution with mean μ and variance σ^2, then we write

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

For example, a random variable X which follows a normal distribution with mean 10 and variance 25 is written as $X \sim N(10, 25)$ or $X \sim N(10, 5^2)$. It is important to note that the second parameter in this notation is the *variance* and not the *standard deviation.*

9.2.3 The standard normal distribution

For various reasons, all probabilities for the normal distribution can be expressed in terms of those for a normal distribution with mean 0 and variance 1. Usually, a random variable with this *standard normal distribution* is called Z, that is

$$Z \sim N(0, 1).$$

If our random variable follows a standard normal distribution, then we can obtain cumulative probabilities from statistical tables (see the table at the end of this chapter, which give "less than or equal to" probabilities). For example, if $Z \sim N(0, 1)$, then:

1. The probability that Z is less than -1.46 is $P(Z < -1.46)$. Therefore we look for the probability in tables corresponding to $z = -1.46$: row labelled -1.4, column headed -0.06. This gives $P(Z < -1.46) = \mathbf{0.0721}$.

2. The probability that Z is less than -0.01 is $P(Z < -0.01)$. Therefore we look for the probability in tables corresponding to $z = -0.01$: row labelled 0.0, column headed -0.01. This gives $P(Z < -0.01) = \mathbf{0.4960}$.

3. The probability that Z is less than 0.01 is $P(Z < 0.01)$. Therefore we look for the probability in tables corresponding to $z = 0.01$: row labelled 0.0, column headed 0.01. This gives $P(Z < 0.01) = \mathbf{0.5040}$.

4. The probability that Z is greater than 1.5 is $P(Z > 1.5)$. Now our tables give "less than" probabilities, and here we want a "greater than" probability. But! The area under the curve is 1:

So we find $P(Z < 1.5) = 0.9332$ and subtract this from 1 to give **0.0668**.

5. What about the probability that Z lies between -1.2 and 1.5? Graphically, this is:

And so

$$\begin{aligned} P(-1.2 < Z < 1.5) &= P(Z < 1.5) - P(Z \leq -1.2) \\ &= 0.9332 - 0.1151 \\ &= 0.8181. \end{aligned}$$

So how do we calculate probabilities for *any* normal distribution, not just the *standard* normal distribution for which we have tables? The easiest approach is to "make" the normal distribution that we have "look like" the standard normal distribution, and then we can just use the tables as before.

But how can we "make" *any* old normal distribution look like the *standard* normal distribution? We can use the "**slide–squash**" technique! This is best demonstrated through an example.

IQ Example

Suppose we are interested in the IQ of 18–19 year olds and that IQs follow a normal distribution with mean $\mu = 100$ and standard deviation $\sigma = 15$. Thus, we have:

$$X\text{: IQ of 18–19 year olds, and}$$

$$X \sim N\left(100, 15^2\right).$$

The distribution for IQs looks like this:

We don't have tables of probabilities for this distribution, but we *do* have tables for the *standard* normal distribution with mean 0 and standard deviation 1, which looks like:

So to make our distribution look like the *standard* normal distribution, we first need to **slide** it along to the left, and then **squash** it in so it has the same spread. For the *slide*, we subtract the mean from our distribution, i.e. subtract 100. Doing so centres the distribution on zero, just like the *standard* normal distribution. Then, for the *squash*, we divide by the standard deviation (in this case 15), which squashes our normal distribution so it has the same spread as the *standard* normal distribution.

The formula for slide–squash, where $X \sim N(\mu, \sigma^2)$ and $Z \sim N(0,1)$, is thus:

$$P(X \leq x) = P\left(Z \leq \frac{x-\mu}{\sigma}\right),$$

which transforms *any* normal distribution into the *standard* normal distribution. Thus, in the IQs example, let's suppose we wanted to find the probability that an 18–19 year old has an IQ less than 85, i.e. $P(X < 85)$. Using the slide–squash formula, this is transformed into a statement about the *standard* normal distribution as follows:

$$\begin{aligned} P(X < 85) &= P\left(Z < \frac{X-\mu}{\sigma}\right) \\ &= P\left(Z < \frac{85-100}{15}\right) \\ &= P\left(Z < -1\right) \\ &= 0.1587. \end{aligned}$$

What about:

(i) The probability that an 18–19 year old has an IQ less than 110?

(ii) The probability that an 18–19 year old has an IQ greater than 110?

(iii) The probability that an 18–19 year old has an IQ greater than 125?

(iv) The probability that an 18–19 year old has an IQ between 95 and 115?

This page has been left blank for your solutions to the last example

Vitamin C example
Suppose that the vitamin C content per 100g tin of tomato juice is normally distributed with mean $\mu = 20$mg and standard deviation $\sigma = 4$mg. Let X be the vitamin C content of a randomly chosen tin. Then

$$X \sim N\left(20, 4^2\right).$$

Find

(i) The probability that a tin has less than 25mg of vitamin C, and

(ii) The probability that the tin has between 18mg and 25mg of vitamin C.

For (i), we have

$$\begin{aligned}
P(X < 25) &= P\left(Z < \frac{25-\mu}{\sigma}\right) \\
&= P\left(Z < \frac{25-20}{4}\right) \\
&= P(Z < 1.25) \\
&= 0.8944 \qquad \text{(from tables).}
\end{aligned}$$

For (ii), we have $P(18 < X < 25)$. This can more easily be seen on a picture:

Thus,

$$\begin{aligned}
P(18 < X < 25) &= P(X < 25) - P(X < 18) \\
&= P\left(Z < \frac{25-20}{4}\right) - P\left(Z < \frac{18-20}{4}\right) \\
&= P(Z < 1.25) - P(Z < -0.5) \\
&= 0.8944 - 0.3085 \\
&= 0.5859.
\end{aligned}$$

9.2.4 Using tables in reverse

We can also use the tables in reverse. For example, we might want to know below what value are 95% of the population. This is equivalent to determining the value of z that satisfies $P(Z < z) = 0.95$. From tables, we can see that

$$\begin{aligned} P(Z < 1.64) &= 0.9495 \qquad \text{and} \\ P(Z < 1.65) &= 0.9505. \end{aligned}$$

Therefore, the value we want for z lies between 1.64 and 1.65. If a more accurate value is needed we can *interpolate* between these values: 0.95 is half-way between 0.9495 and 0.9505 and so we take $z = 1.645$. This is a more accurate answer and sufficient in most cases. However, the exact value for z can be found from more detailed tables or via a computer package such as `Minitab`. Here are some more examples.

1. Below what value does 10% of the standard normal population fall? From tables we get

$$P(Z < -1.28) = 0.1003 \quad \text{and} \quad P(Z < -1.29) = 0.0985.$$

   0.1003 is closer to 10% (0.1) than 0.0985, and so, roughly, 10% of the standard normal population falls below –1.28. We will see how to get a more accurate answer for this problem in the tutorials later on this week.

2. A similar calculation can be used to calculate the IQ that identifies the bottom 10% of 18–19 year olds. We need the value of x, where $P(X < x) = 0.1$. Now this population has $\mu = 100$ and $\sigma = 15$. Also

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

   and so we need x so that

$$P\left(Z \le \frac{x-100}{15}\right) = 0.1.$$

   We know (from earlier) that $P(Z < -1.2817) = 0.1$ and therefore we solve

$$\frac{x-100}{15} = -1.2817,$$

   that is

$$\begin{aligned} x &= 100 - 1.2817 \times 15 \\ &= 100 - 19.2255 \\ &= 80.7745. \end{aligned}$$

   Notice that the calculation that transforms the z–value onto the x–scale is

$$x = \mu + z\sigma.$$

3. What is the IQ that identifies the top 1% of 18-19 year olds? Again, we first determine the value z that identifies the top 1% of a standard normal population and then translate this into an IQ. So we need the value z that satisfies $P(Z > z) = 0.01$. This is the same value as satisfies $P(Z < z) = 0.99$. A quick examination of tables gives the two key probabilities as

$$P(Z < 2.32) = 0.9898 \quad \text{and} \quad P(Z < 2.33) = 0.9901;$$

0.9901 is closer (just!) to 0.99 than 0.9898, and so we take the Z value to be 2.33. Moving back to the IQ scale, we need the value x such that $P(X > x) = 0.01$ and so we take

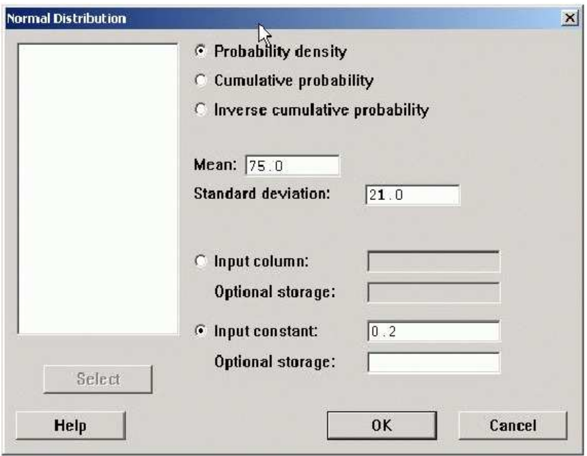
$$\begin{aligned} x &= \mu + z\sigma \\ &= 100 + 2.3267 \times 15 \\ &= 134.9. \end{aligned}$$

9.2.5 Using `Minitab`

`Minitab` is very helpful with calculating normal probabilities. The following commands will calculate probabilities $P(X < x)$ and also values of x that satisfy $P(X < x) = p$:

1. `Calc > Probability Distributions > Normal`

   opens up dialogue box



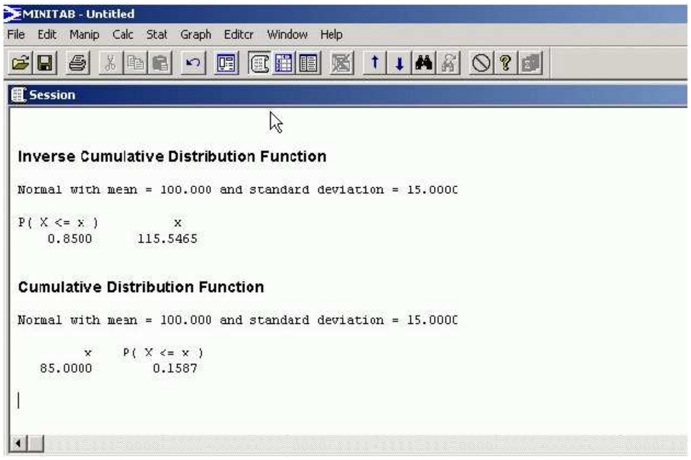
2. Select `Cumulative probability` for $P(X < x)$ or `Inverse cumulative probability` for the value of x satisfying $P(X < x) = p$

3. Enter the `Mean` (μ) and the `Standard Deviation` (σ)

4. Select `Input Constant` and enter the value for x or p (as appropriate)

5. Click `OK`

6. The answer is displayed in the Session Window:


MINITAB - Untitled
File Edit Manip Calc Stat Graph Editor Window Help
Session
Inverse Cumulative Distribution Function
Normal with mean = 100.000 and standard deviation = 15.0000
P(X <= x) x
0.8500 115.5465
Cumulative Distribution Function
Normal with mean = 100.000 and standard deviation = 15.0000
x P(X <= x)
85.0000 0.1587

9.3 Exercises

1. The lengths of steel beams made in a steel mill follow a normal distribution with mean $\mu = 8.25$m and standard deviation $\sigma = 0.07$m.
   (a) What is the probability that the length of a steel beam is less than 8.4m?
   (b) What is the probability that the length of a steel beam is between 8.2m and 8.4m?
   (c) A customer requires beams no larger then 8.05m. What percentage of the mill's output can be used to supply this customer?
   (d) The mill is trying to negotiate a new contract with this customer. It is in the mill's interests to be able to supply 98% of its output to the customer. What is the smallest length that achieves this requirement?

2. A drinks machine is regulated by its manufacturer so that it discharges an average of 200ml per cup. However, the machine is not particularly accurate and actually discharges an amount that has a normal distribution with standard deviation 15ml.
   (a) What percentage of cups contain below the minimum permissible volume of 170ml?
   (b) What percentage of cups contain over 225ml?
   (c) What is the probability that the cup contains between 175ml and 225ml?
   (d) How many cups would you expect to overflow if 240ml cups are used for the next 10000 drinks?

3. A company promises delivery within 20 working days of receipt of order. However, in reality, they deliver according to a normal distribution with a mean of 16 days and a standard deviation of 2.5 days.
   (a) What proportion of customers receive their order late?
   (b) What proportion of customers receive their orders between 10 and 15 days of placing their order?
   (c) How many days should the delivery promise be adjusted to if only 3% of orders are to be late?
   (d) A new order processing system promises to reduces the standard deviation of delivery times to 1.5 days. If this system is used, what proportion of customers will receive their deliveries within 20 days?

Probability Tables for the Standard Normal Distribution

The table contains values of $P(Z < z)$, where $Z \sim N(0, 1)$.

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

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