The Z-Transformation of the Normal Distribution

Similar to other probability distributions, the area under the normal curve represents the probability of occurrence of X.

To more quickly calculate the area under the normal distribution curve statisticians have given us the Z-transformation, along with the Z-tables.

To perform the Z-transformation, you can use the following equation. This will transform your random variable X, into a Z-value based on the distributions mean & standard deviation.

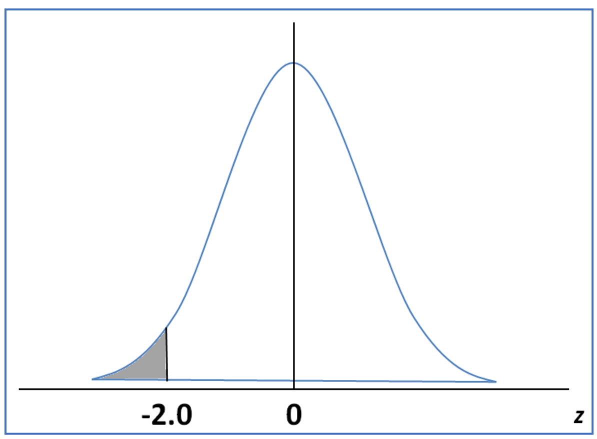
$$Z = \frac{X - \mu}{\sigma}$$

For example, let's say you've got a variable X (Grades on the CQE Exam) that follows the normal distribution with a mean value μ = 82 and a standard deviation σ = 6. The Z-score for an exam grade of 70 can be calculated as:

$$Z = \frac{70 - 82}{6} = -2.0$$

We can interpret this result by saying that the exam score of 70 is 2.0 standard deviations below the mean.

If you wanted to calculate the proportion of the population which scored less than 70% on the exam, it would look like the gray shaded area below on the distribution:

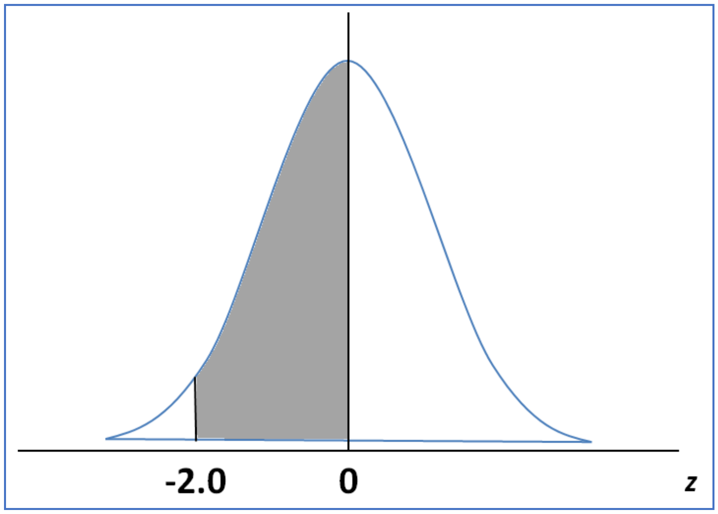


Notice this distribution is not a reflection of the exam score (centered at z=0), but it's a reflection of the transformed z-score associated with the exam.

We can then use the Z-score tables to answer any probability question associated with this value without having to use a calculator.

Z-Transformation Example

For example, the graph above shows all exam scores less than z = -2.0; however, you could also use the z-table to find the probability of a Z-score between -2 and 0, which graphically looks like this:



Once you've performed the Z-transformation, you can now calculate the probability associated with your Z-value using the table below.

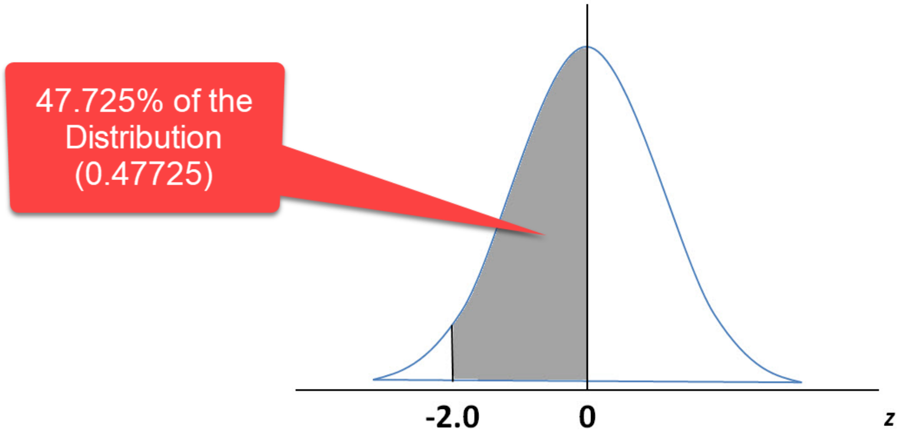
This Probability Table can be used to take your Z-value and convert it into the probability.

This table is potentially different from other Z-Probability tables in that it only provides the probability of positive Z values.

Recall though that the Z-value is symmetric around the mean value, so if you were looking for the probability from -1.04 to 0, it would be the same probability as that from +1.04 to 0.

Also, if you wanted to look up the probability between -1 to +1, then you'd double probability of Z=1.0 (0.34134).

The Z-Transformation of the Normal Distribution



Area under the Normal Curve from 0 to X

X	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.00000	0.00399	0.00798	0.01197	0.01595	0.01994	0.02392	0.02790	0.03188	0.03586
0.1	0.03983	0.04380	0.04776	0.05172	0.05567	0.05962	0.06356	0.06749	0.07142	0.07535
0.2	0.07926	0.08317	0.08706	0.09095	0.09483	0.09871	0.10257	0.10642	0.11026	0.11409
0.3	0.11791	0.12172	0.12552	0.12930	0.13307	0.13683	0.14058	0.14431	0.14803	0.15173
0.4	0.15542	0.15910	0.16276	0.16640	0.17003	0.17364	0.17724	0.18082	0.18439	0.18793
0.5	0.19146	0.19497	0.19847	0.20194	0.20540	0.20884	0.21226	0.21566	0.21904	0.22240
0.6	0.22575	0.22907	0.23237	0.23565	0.23891	0.24215	0.24537	0.24857	0.25175	0.25490
0.7	0.25804	0.26115	0.26424	0.26730	0.27035	0.27337	0.27637	0.27935	0.28230	0.28524
0.8	0.28814	0.29103	0.29389	0.29673	0.29955	0.30234	0.30511	0.30785	0.31057	0.31327
0.9	0.31594	0.31859	0.32121	0.32381	0.32639	0.32894	0.33147	0.33398	0.33646	0.33891
1.0	0.34134	0.34375	0.34614	0.34849	0.35083	0.35314	0.35543	0.35769	0.35993	0.36214
1.1	0.36433	0.36650	0.36864	0.37076	0.37286	0.37493	0.37698	0.37900	0.38100	0.38298
1.2	0.38493	0.38686	0.38877	0.39065	0.39251	0.39435	0.39617	0.39796	0.39973	0.40147
1.3	0.40320	0.40490	0.40658	0.40824	0.40988	0.41149	0.41309	0.41466	0.41621	0.41774
1.4	0.41924	0.42073	0.42220	0.42364	0.42507	0.42647	0.42785	0.42922	0.43056	0.43189
1.5	0.43319	0.43448	0.43574	0.43699	0.43822	0.43943	0.44062	0.44179	0.44295	0.44408
1.6	0.44520	0.44630	0.44738	0.44845	0.44950	0.45053	0.45154	0.45254	0.45352	0.45449
1.7	0.45543	0.45637	0.45728	0.45818	0.45907	0.45994	0.46080	0.46164	0.46246	0.46327
1.8	0.46407	0.46485	0.46562	0.46638	0.46712	0.46784	0.46856	0.46926	0.46995	0.47062
1.9	0.47128	0.47193	0.47257	0.47320	0.47381	0.47441	0.47500	0.47558	0.47615	0.47670
2.0	0.47725	0.47778	0.47831	0.47882	0.47932	0.47982	0.48030	0.48077	0.48124	0.48169
2.1	0.48214	0.48257	0.48300	0.48341	0.48382	0.48422	0.48461	0.48500	0.48537	0.48574
2.2	0.48610	0.48645	0.48679	0.48713	0.48745	0.48778	0.48809	0.48840	0.48870	0.48899
2.3	0.48928	0.48956	0.48983	0.49010	0.49036	0.49061	0.49086	0.49111	0.49134	0.49158
2.4	0.49180	0.49202	0.49224	0.49245	0.49266	0.49286	0.49305	0.49324	0.49343	0.49361
2.5	0.49379	0.49396	0.49413	0.49430	0.49446	0.49461	0.49477	0.49492	0.49506	0.49520
2.6	0.49534	0.49547	0.49560	0.49573	0.49585	0.49598	0.49609	0.49621	0.49632	0.49643
2.7	0.49653	0.49664	0.49674	0.49683	0.49693	0.49702	0.49711	0.49720	0.49728	0.49736
2.8	0.49744	0.49752	0.49760	0.49767	0.49774	0.49781	0.49788	0.49795	0.49801	0.49807
2.9	0.49813	0.49819	0.49825	0.49831	0.49836	0.49841	0.49846	0.49851	0.49856	0.49861
3.0	0.49865	0.49869	0.49874	0.49878	0.49882	0.49886	0.49889	0.49893	0.49896	0.49900

Reliability Example Using the Normal Distribution

Let's do another example of the Z-transformation in a real-life situation.

Within the world of **Reliability**, the normal distribution curve can be used to model the reliability of a system over time.

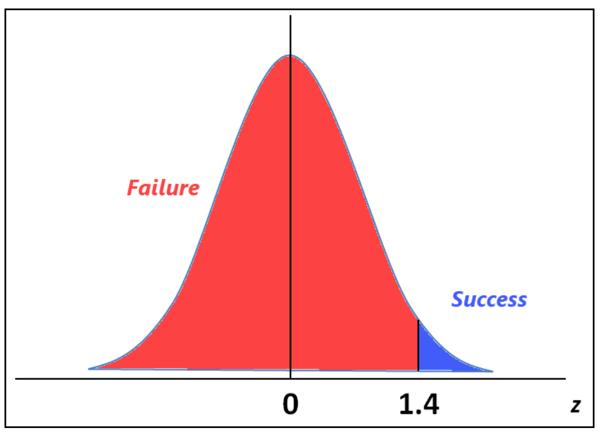
Let's say we're dealing with a motor and we've modeled the motors failure over time and it fits the normal distribution.

Your test data indicates that the mean and standard deviation associated with the motor is 6,500 hours and 500 hours respectively.

What is the **reliability** (*the probability that the motor is still operational*) of the motor at 7,200 hours?

$$\boldsymbol{Z = \frac{(X - \mu)}{\sigma} = \frac{(7,200 - 6500)}{500} = 1.4}$$

Graphically, this looks like:



Using the Z-Tables, *the area under the curve at Z = 1.4 is .4192*, and we add to that the 0.500 that represents the left half of the normal distribution curve which add up to 0.9192.

Remember that the Z-Score and the resulting probability represent the area to the left of the time value (7,200 hours).

So the **reliability** is the area to the **right of the curve**, which is 1 - .9192 = 0.0802.

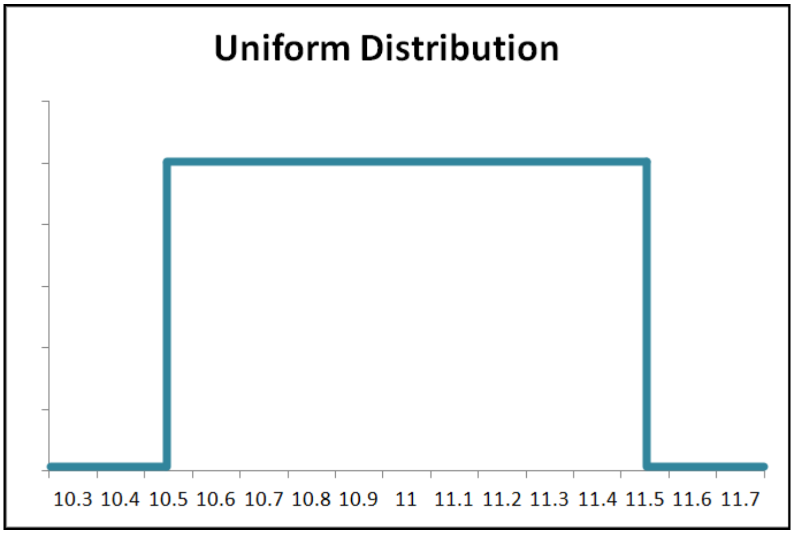
Therefore, there is an 8% probability that the motor has not yet failed after 7,200 hours.

Or, said differently, 8% of the original population of motors are likely still operational after this amount of time.

Uniform Distribution

The uniform distribution is probably the easiest one to spot, and the easiest one to analyze.

For example, let's say you've got a machining process that is producing part who's length is uniformly distributed between 10.5mm & 11.5mm - this distribution is shown below.



As you can see from this distribution, all outcomes (10.5, 10.6, etc.) within this distribution share the same probability of occurrence.

A classic example of the uniform distribution is the dice roll. The probability of each outcome (1, 2, 3. . . .) is constant between the 6 possible outcomes.

Let's use the example of the machining process above to discuss the **Mean** & **Variance** associated with the uniform distribution.

For the Uniform distribution, the **mean value** is simply the middle value which can be calculated as the minimum value (a) plus the maximum value (b) divided by 2:

$$\boldsymbol{\mu = \frac{a+b}{2}} \qquad \textit{for this example} \rightarrow \qquad \mu = \frac{10.5+11.5}{2} = 11.0$$

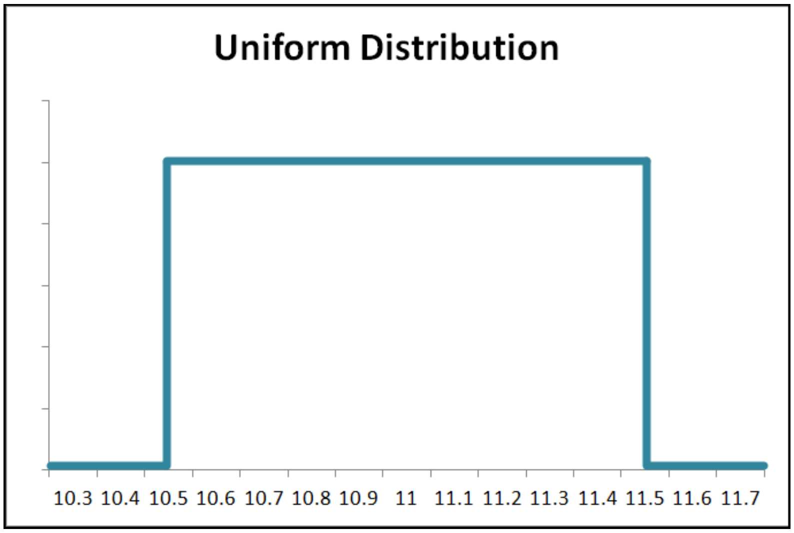
For the **Variance**, that calculation looks like this:

$$\boldsymbol{\sigma^2 = \frac{(b-a)^2}{12}} \qquad \textit{for this example} \rightarrow \qquad \sigma^2 = \frac{(11.5-10.5)^2}{12} = \frac{1}{12}$$

Probability Example #1 - Uniform Distribution

Now let's see how we'd calculate **probability with the uniform distribution**.

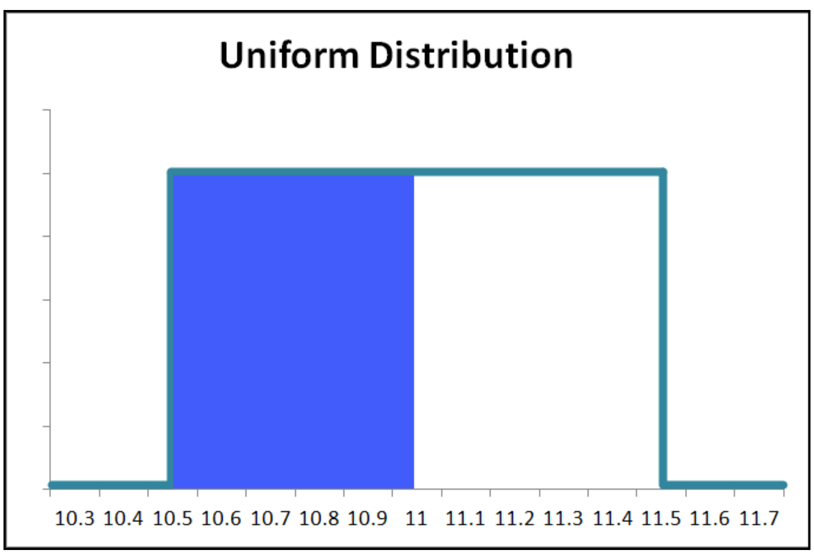
Using the example above of a machining process that is producing part who's length is uniformly distributed between 10.5mm & 11.5mm, ***what proportion of the overall population is produced within the lengths of 10.5mm - 11.0mm?***



This can be calculated using the following equation:

$$\boldsymbol{P}(\boldsymbol{X_1} < x < \boldsymbol{X_2}) = \frac{(\boldsymbol{X_1} - \boldsymbol{X_2})}{(\boldsymbol{b} - \boldsymbol{a})} \qquad Where \rightarrow \quad P(10.5 > x > 11.0) = \frac{(11.0 - 10.5)}{(11.5 - 10.5)} = \frac{.5}{1} = \frac{1}{2} = 0.50$$

This answer should make sense as 11.0 is the mean value of this distribution and we would expect 50% of the population to fall to the left of the mean value.



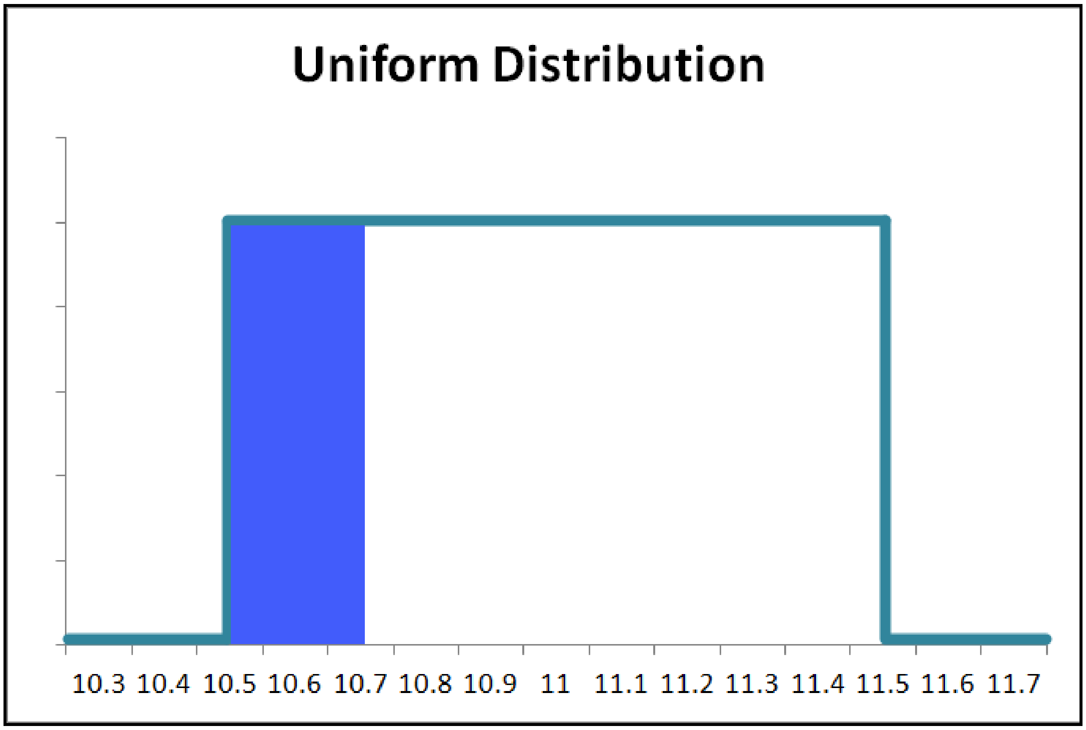
Probability Example #2 - Uniform Distribution

We could also ask the question, **what proportion of the population is less than 10.7mm**; let's see what that calculation would look like:

$$\boldsymbol{P}(\boldsymbol{x} < \boldsymbol{X_1}) = \frac{(\boldsymbol{X_1} - \boldsymbol{a})}{(\boldsymbol{b} - \boldsymbol{a})} \qquad Where \rightarrow \quad P(x > 11.0) = \frac{(10.7 - 10.5)}{(11.5 - 10.5)} = \frac{.2}{1} = \frac{1}{5} = 0.20$$

So we can interpret that to mean that 20% of the population is machined at less than 10.7mm in length.

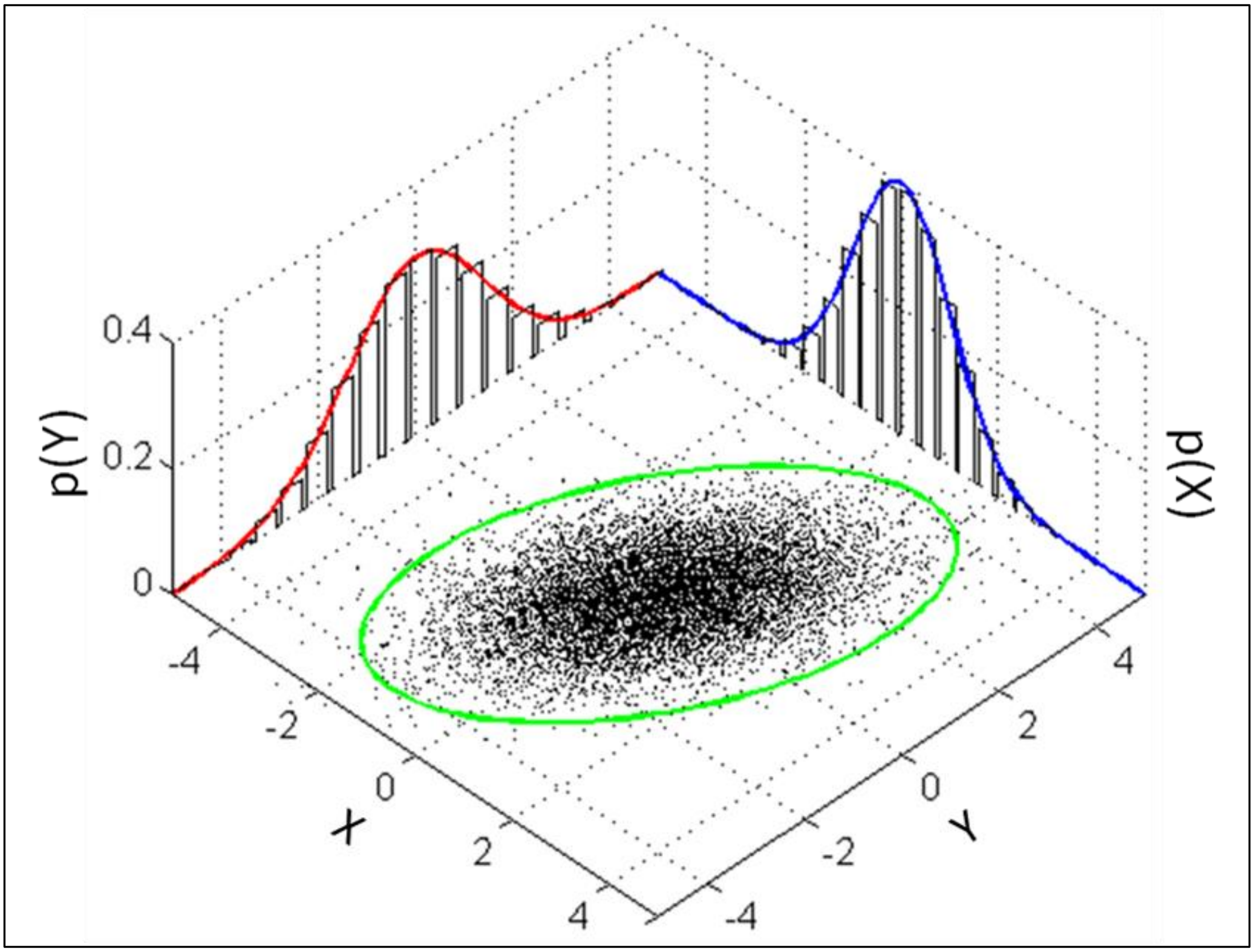
We can also interpret this result to mean that there's a 20% chance that any randomly sampling part will be measured to have a length less than 10.7mm.



Bivariate Normal Distribution

The Bivariate Normal Distribution is a spin-off of the normal distribution and it occurs when you have 2 variables that both are normally distributed. These variables can be completely independent, or a co-variance can exist between them.

Since this distribution is dealing with 2 variables - it's called the **bivariate** (2 variable) distribution.



This distribution can be characterized by its mean values, its standard deviations and it's correlation, specifically:

- the **Mean of X** (μ_x) & **Standard Deviation of X** (σ_x),
- The **Mean of Z** (μ_z) & **Standard Deviation of Z** (σ_z),
- ***p*** is the **correlation** between X & Z.

Due to the complexity of these calculations, it's unlikely that you'll be asked to use this distribution on the CQE Exam.

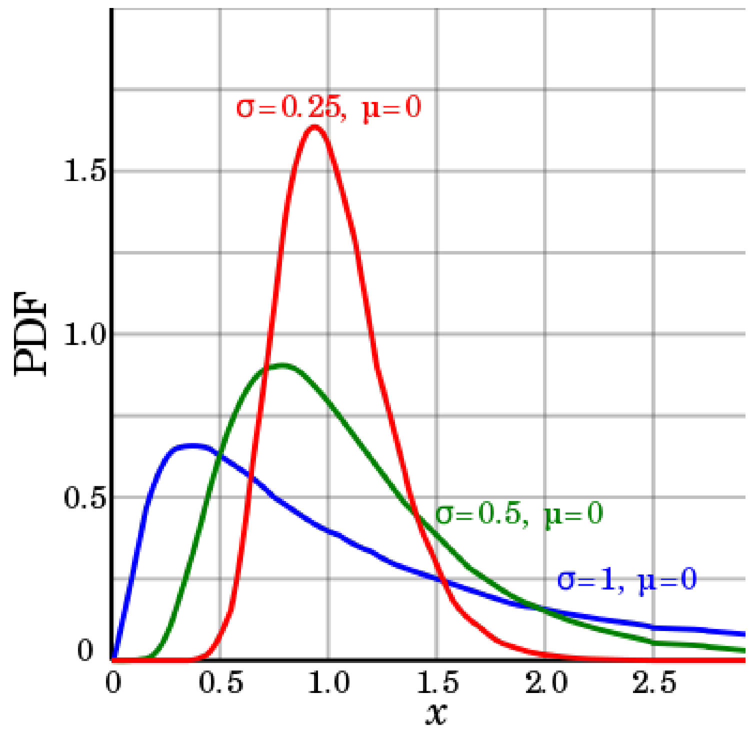
Lognormal Distribution

Another Distribution that you may run into is the Lognormal Distribution, and it also shares a relationship with the Normal Distribution.

So let's say you've got a data set for a random variable X that is log-normally distribution.

You can transform that data using the natural log (ln), and create a transformed random variable Y (Y = ln X), that is normally distributed.

This transformation is beneficial because normality is required to use certain statistical operations like constructing confidence intervals, and conducting tests of hypothesis, etc.



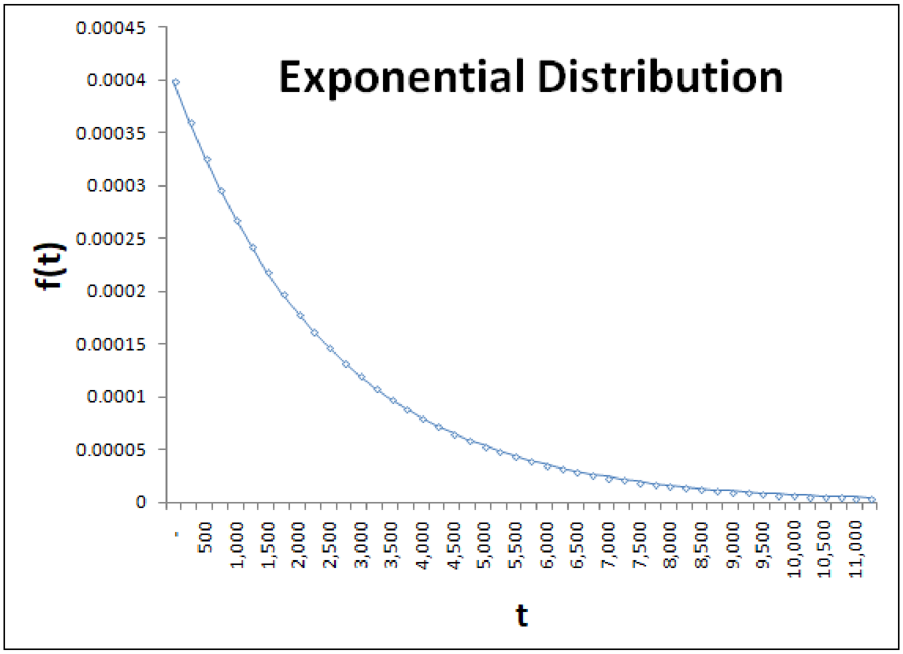
Similarly, if your data fits the normal distribution then a logarithmic transformation ($y = e^x$) can be use to transform that data to fit the lognormal distribution.

One common attribute about this distribution is that it only takes on positive real values and the distribution is always skewed to the right.

Exponential Distribution

The Exponential Distribution is used widely in the area of Reliability Engineering and usually models the length of time between the occurrence of events.

This includes items like the time between repairs, or the time between failures, etc.



The exponential distribution has a unique relationship the Poisson Distribution, which we will discuss below.

The exponential distribution is commonly used to models the time between occurrences of events. This is the inverse of the Poisson distribution which models the number of occurrences in a given interval of time.

$$Exponential = \frac{Time}{Occurrences} \qquad\qquad Poisson = \frac{Occurrence}{Time}$$

These two distributions are related in that they are inverses of each other.

So if you've got a random variable, X, that follows the exponential distribution, then the inverse of x, which we will call y = 1/x will follow the Poisson Distribution.

The inverse of this statement is true; if a random variable follows the Poisson distribution then the reciprocal of that variable will follow an exponential distribution.

Probability Calculations & the Exponential Distribution

Below are the Probability Equations that you can use when dealing with data that fits the exponential distribution.

$$\boldsymbol{Probability}: \boldsymbol{P(X = x)}: \ \mathrm{f(x)} = \mathrm{xe}^{-\lambda \mathrm{x}}$$

$$\boldsymbol{Cumulative\ Probability}: \boldsymbol{P(X > x)}: \ \mathrm{f(x)} = \mathrm{e}^{-\lambda \mathrm{x}}$$

$$\boldsymbol{Cumulative\ Probability}: \boldsymbol{P(X < x)}: \ F(x) = 1 - \mathrm{e}^{-\lambda \mathrm{x}}$$

Where

$$\boldsymbol{Mean\ Value} = \boldsymbol{\theta} = \frac{\boldsymbol{1}}{\boldsymbol{\lambda}} \quad where \quad \boldsymbol{\lambda} = \boldsymbol{Occurrence\ Rate}$$

Reliability Example for the Exponential Distribution

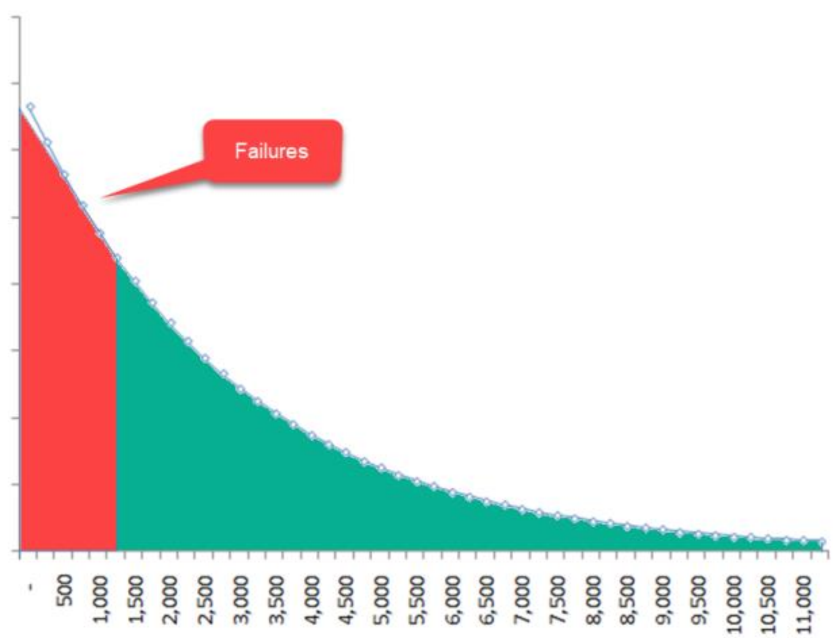
Let's say we're testing a motor who's failure rate follows the exponential distribution and we've found that our Mean time to failure (θ) is 2,996 Hours.

What is the probability that our motor is still operational after at 1,200 hours?

$$\boldsymbol{P(X} > 1{,}200) = \ \boldsymbol{e^{-\lambda x}} = \ \boldsymbol{e^{\frac{-x}{\theta}}} = \boldsymbol{e^{\frac{-1{,}200}{2{,}966}}} = \boldsymbol{.6673}$$

So, the probability that our product will perform successfully past the 1,200 hour mark is approximately 66%.

This result can also be applicable to the entire population of product; i.e. 66% of the population can be expected to surpass the 1,200 hour mark. The remaining units will have failed by that point.



Reliability Example #2

Or we could calculate the probability of a failure in the first 1,000 hours.

$$\boldsymbol{CDF}\text{:}\ \ F(t) = 1 - \mathrm{e}^{-\lambda \mathrm{t}}$$

$$\boldsymbol{F(1,000) = 1 - \mathrm{e}^{-\frac{1,000}{2,966}} = .286}$$

This result could be interpreted to mean that the 28.6% of motors will fail at or before the 1,000-hour mark.