**Topic 12**

**Normal Distributions**



Normal Distributions are the most common distribution, and have been studied extensively. It is especially common wrt to measurement, which represents a large proportion of our data collection.

A normal distribution is concisely defined. Even if your data looks like a normal distribution, it may not be. However, this distribution occurs so often in nature, it probably is normal.

Normal distributions are:

* Bell Shaped
* Symmetric
* Single peaked
* has the following characteristics
	+ 68.2 % of the data falls within 1 standard deviation of the mean
	+ 95.6% of the data falls within 2 standard deviations of the mean
	+ 99.98% of the data falls within 3 standard deviations of the mean



The Normal distribution is often *standardized* to create the **Standard Normal Distribution**. When a distribution is standard normal, then an (arbitrary) area under the curve is the **probability** of a value occurring within that z-score range. [[1]](#footnote-1) The total area under a standard normal distribution is 1, which states that the probability of an event within the total range of values is one.

Note that this probability can easily be turned into a percentage, because it represents in what proportion of 1 (the total area under the curve), an event might occur. For example, if an area under a standard normal curve is equal to .4, this is the same as saying that when an event occurs, there is a 40% chance it will be in the range represented by that area.[[2]](#footnote-2)

When we standardize data, we “map” it to a standard normal distribution.

To standardize data we calculate z-scores.$ z\_{i}= \frac{x\_{i}-μ}{σ} or (z\_{i}=\frac{x\_{i}-\overbar{x}}{s}$ for a sample)

For a System with a Normal distribution with µ=12 and σ=4, we make some measurements

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| x values | 11 | 12 | 10 | 16 | 12.5 | 9 | 13 | 15 |
| z-scores | -0.25 | 0 | -0.5 | 1.0 | 0.125 | -0.75 | 0.25 | 0.75 |

Note that the mean is mapped to 0 and a value one standard deviation away is mapped to 1.

Lets look at a more contrived set for the same distribution with µ=12 and σ=4,

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| x values | 4 | 8 | 10 | 12 | 14 | 16 | 20 | 24 |
| z-scores | -2 | -1 | -0.5 | 0 | 0.5 | 1 | 2 | 3 |

Clearly, 12 is mapped to 0, and values that are one, two, and three standard deviations away are mapped to 1, 2, and 3 respectively.

Unfortunately, calculating the probability is *not so easy.* The best way to do it is to use a table, often in the back of the (or some) book.[[3]](#footnote-3) In the back of Ross and Chance, the number in the table represents the probability that and event with z-score at or less than that value will occur. It is not the probability of that value, but that value and all values to the left.

It is possible to calculate probabilities for everything greater than this value, simply subtract that value from 1, since the total area under the curve is 1.

There are actually 3 different possible Calculations:

1. Area to the left of a z-score
2. Area to the right of a z-score
3. Area between two z-scores
4. Areas to the left and right of two z-scores

A and B can be obtained from one another by subtracting the other one from the value 1. A=1-B and B=1-A. Easy.

C is a little trickier, but not much. To calculate the probability of an event with a value between two z-scores, perform the calculation for A twice (one for each z-score), then subtract the smaller from the larger. Done.

A B



C D



D? Perform the calculations for A and B, then add them up.

It is worth noting that not all tables at the back of the book are the same. Some calculate the area starting from the mean, 0, and not left z=-3.5.

1. When talking about continuous distributions, frequency and probability are discussed in terms of a range of values, never a single value. For example, in a normal distribution with mean 10 and standard deviation 5, we never ask, “what is the probability of getting a 7?” We only ask, “what is the probability of getting a value between, say, 6.5 and 7.5?” [↑](#footnote-ref-1)
2. Technically speaking here, range is domain, or x values. [↑](#footnote-ref-2)
3. There are other ways. Modern calculators can do it, and excel can do it. [↑](#footnote-ref-3)