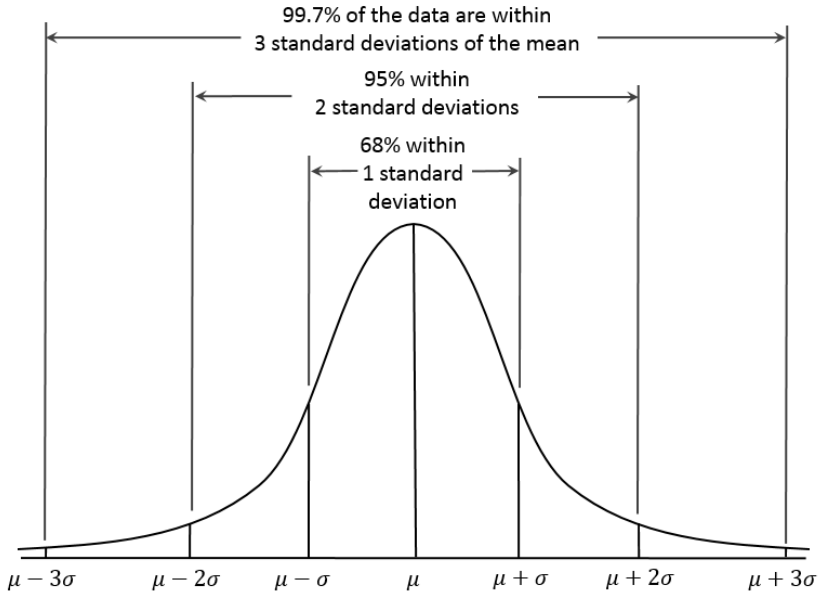
Normal distribution and Z-scores

1. Normal distribution

Each random variable has a probability distribution associated to it. One of the most common distributions is the **normal distribution**. The normal distribution is a probability density function represented by a bell shaped curve centred at the population mean. The total area under the curve is 1. If a random variable, X, is normally distributed we write $X \approx N(\mu, \sigma)$, where μ is a the population mean and σ is standard deviation.



- The area under the curve is 1. This is meant to reflect the probability of obtain any one value in my data at random.
- The curve is symmetric, meaning 50% of the population lie below the mean and 50% lie above.
- The area enclosed within one standard deviation of the mean represents 68% of the population.
- The area enclosed within two standard deviations of the mean represents 95% of the population.
- The area enclosed within three standard deviations of the mean represents 99.7% of the population.

2. Standardisation and Z-scores

Simply put, standardisation allows us to realise the data in a more convenient way. This is done by repositioning the data relative the to the standard normal distribution, which is a normal distribution with mean 0 and standard deviation 1, i.e., $X \approx N(0, 1)$. We can standardise any piece of data, x, using the **Z-score** formula:

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

where μ is the population mean and σ is the standard deviation.

Why do we standardise the data? Because it makes things easier! Once we have standardised the data, we can use us tables, known as z-tables to simply read off the probability we need. The number relevant number in the z-table (at the end of the document) gives the area to the left of the z-score under the standard normal curve.

Example 2.1. Question: Suppose that the average daily calorie intake of British men (in kcal) is normal distributed, with mean 3,100 and standard deviation 125. Calculate the probability that the average daily calorie intake of a randomly-chosen male will be less than 3, 190 kcal.

Solution:

Step 1 Calculate the z-score of 3,200:

$$z = \frac{3,190 - 3,100}{125} = 0.72$$

Step 2 Look up 0.72 in the z-table:

$$0.72 = 0.70 + 0.02.$$

So first we go to the 0.7 row. Then we go across that row until we get to the 0.02 column. That number is our probability. So the probability a randomly selected man's calorie intake is less than 3,190 is 0.7642. Or more mathematically:

$$\mathrm{P}\left(\mathrm{x} \leqslant 3,190\right) = 0.7642.$$

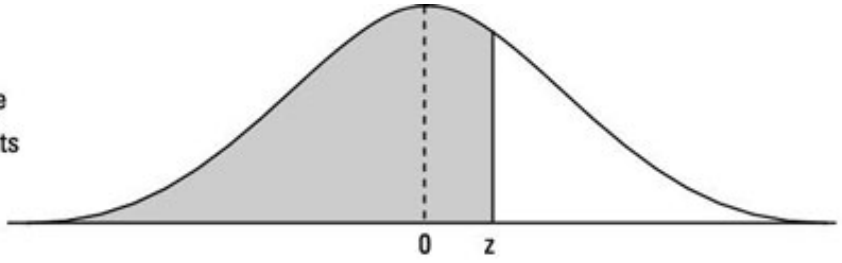
Alternatively if we wanted to know the probability of a randomly selected man's calorie intake is more than 3,190, we'd go through the same procedure except we would need to add one final step. For the probability of more than 3,190 is 1 minus the probability of less than 3,190! That is

$$\mathrm{P}(\mathrm{x} \geqslant 3,190) = 1 - 0.7642 = 0.2358.$$

Exercise 2.2. The average score for a class sitting a Physics test is 54, with standard deviation of 7. The average score for the same class in a Chemistry exam is 72, with standard deviation 10. Each set of scores is distributed normally. Which is more likely: that a student in the class

(1) scores over 77 on the Chemistry test; or
(2) scores over 58 on the Physics test?

Number in the table represents $P(Z \leq z)$



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998
3.6	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999