MDM4U: Mathematics of Data Management

What Is the Probability That. . . ?

Probabilities for Normally Distributed Data

J. Garvin

Properties of the Normal Distribution

Recap

For a normal distribution with $\mu = 17$ and $\sigma = 2$, what is the likelihood that a datum has a value between 15 and 21?

Since 15 is one standard deviation below the mean, and 21 is two standard deviations above it, we can use the empirical rule to determine the chance of falling in the specified range.

$68/2 = 34\%$ of the data fall between 15 and 17.

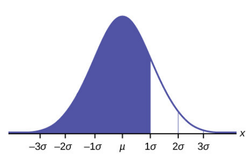
$95/2 = 47.5\%$ of the data between 17 and 21.

Therefore, the likelihood of a datum having a value between 15 and 21 is $34 + 47.5 = 81.5\%$.

Probabilities Using the Normal Distribution

For non-integral z-scores, tables have been created that associate a specific z-score with a probability.

These tables typically measure cumulative probabilities *from the left*. That is, $P(x \leq k)$ for some value k.



To use a table, first determine a datum's z-score, then look up its corresponding probability.

Probabilities Using the Normal Distribution

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051

The z-scores in the table are given to two decimals of accuracy, with the row being the first digit and the column the second.

For example, a z-score of 0.32 has a corresponding probability of 0.6255 in the table above, while a z-score of 0.56 has a corresponding probability of 0.7123.

Probabilities Using the Normal Distribution

Example

For a normal distribution with $\mu = 75$ and $\sigma = 8$, determine $P(x \leq 79)$.

The z-score for $x = 79$ is $z = \dfrac{79 - 75}{8} = 0.50$.

Using the table of probabilities, a z-score of 0.50 corresponds to a probability of approximately 0.6915.

Therefore, $P(x \leq 79) \approx 0.6915$.

Probabilities Using the Normal Distribution

Example

For a normal distribution with $\mu = 18$ and $\sigma = 1.4$, determine $P(x \leq 11.5)$.

The z-score for $x = 11.5$ is $z = \dfrac{11.5 - 18}{1.4} \approx -4.64$.

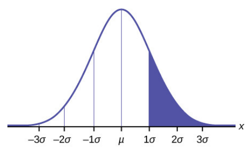
Since this datum is approximately 4.64 standard deviations below the mean, the probability is essentially zero.

Thus, $P(x \leq 11.5) \approx 0$.

Probabilities Using the Normal Distribution

Most tables do not include probabilities for areas *to the right* of the curve. That is, they do not explicitly state $P(x \geq k)$ for some k.

Recall, however, that the area under the normal curve is 1. Thus, $P(x \geq k) = 1 - P(x \leq k)$.



Probabilities Using the Normal Distribution

Example

For a normal distribution with $\mu = 37.8$ and $\sigma = 2.4$, determine $P(x \geq 33.5)$.

The z-score for $x = 33.5$ is $z = \frac{33.5 - 37.8}{2.4} \approx -1.79$.

The corresponding probability for a z-score of approximately -1.79 is 0.0367.

So, $P(x \geq 33.5) = 1 - P(x \leq 33.5) \approx 1 - 0.0367 \approx 0.9633$.

Probabilities Using the Normal Distribution

Example

For a normal distribution with $\mu = 93.8$ and $\sigma = 7.1$, determine $P(x \geq 67.9)$.

The z-score for $x = 67.9$ is $z = \frac{67.9 - 93.8}{7.1} \approx -3.65$.

Since this datum is approximately 3.65 standard deviations below the mean, $P(x \leq 67.9) \approx 0$.

Therefore, $P(x \geq 67.9) = 1 - P(x \leq 67.9) \approx 1 - 0 \approx 1$.

Thus, $P(x \geq 67.9) \approx 1$, or essentially one.

Probabilities Using the Normal Distribution

Calculating the probability that a random variable falls between two values, a and b, is a multi-step process.

A table will give both $P(x \leq a)$ and $P(x \leq b)$, representing the areas under the curve to a and b respectively.

The area between a and b is the area to the left of b, minus the area to the left of a.

Therefore, $P(a \leq x \leq b) = P(x \leq b) - P(x \leq a)$.

To calculate the probability of falling between a and b, determine the z-scores for both a and b, look up their probabilities, then subtract the smaller value from the larger.

Probabilities Using the Normal Distribution

Example

For a normal distribution with $\mu = 64$ and $\sigma = 5$, determine $P(63 \leq x \leq 72)$.

The z-score for 63 is $z = \frac{63 - 64}{5} = -0.20$, and for 72 is $z = \frac{72 - 64}{5} = 1.60$.

Thus, $P(x \leq 63) \approx 0.4207$, and $P(x \leq 72) \approx 0.9452$.

Therefore, $P(63 \leq x \leq 72) \approx 0.9452 - 0.4207 \approx 0.5245$.

Probabilities Using the Normal Distribution

Example

For a normal distribution with $\mu = 73$ and $\sigma = 5$, determine $P(x = 71)$.

To determine the probability of the random variable being *exactly* 71, we must use a small interval to represent the value. But how small of an interval should be used?

If we use the interval $70.9 - 71.1$, then $P(x = 71) \approx 0.0147$.

Using the interval $70.99 - 71.01$, $P(x = 71) \approx 0.0015$.

The value quickly converges toward zero. This is true for *all* specific probabilities $P(x = k)$.

Probabilities Using the Normal Distribution

Example

The average height of a 13 year old male is 156 cm, with a standard deviation of 4.2 cm. What is the probability that a randomly-selected 13 year old male will have a height between 150 cm and 160 cm?

The z-score for 150 is $z = \frac{150 - 156}{4.2} \approx -1.43$, and for 160 is $z = \frac{160 - 156}{4.2} \approx 0.95$.

Thus, $P(x \leq 150) \approx 0.0764$, and $P(x \leq 160) \approx 0.8289$.

Therefore, $P(150 \leq x \leq 160) \approx 0.8289 - 0.0764 \approx 0.7525$.

Reverse-Lookups

Example

What range of values corresponds to the lower 25% of all data for a normal distribution with $\mu = 10$ and $\sigma = 1.5$?

To answer this, we can use a *reverse-lookup* where we determine the z-score, based on the given probability, 0.25.

Since 0.25 does not appear exactly in the table, we use the closest value, 0.2514, when $z = -0.67$.

Thus, the range will be any value less than or equal to the value of the datum with a z-score of -0.67.

So, $-0.67 = \frac{x - 10}{1.5}$ or $x = -0.67 \times 1.5 + 10 \approx 8.995$.

The range of values is $x \leq 8.995$.

Reverse-Lookups

Example

What range of values corresponds to the central 10% of all data for a normal distribution with $\mu = 7.3$ and $\sigma = 0.4$?

The central 10% is the area between the lower 45% and the lower 55%.

For 0.45, the closest z-score is 0.4483 for $z = -0.13$, and for 0.55, the closest z-score is 0.5517 for $z = 0.13$. Note the symmetry.

Thus, $-0.13 = \frac{x - 7.3}{0.4}$ or $x = -0.13 \times 0.4 + 7.3 \approx 7.248$.

Similarly, $0.13 = \frac{x - 7.3}{0.4}$ or $x = 0.13 \times 0.4 + 7.3 \approx 7.352$.

The range of values is $7.248 \leq x \leq 7.352$.

Questions?