THE STANDARD NORMAL DISTRIBUTION

Although it is usually easy to use technology to find values associated with any normal distribution, it is often useful, especially when comparing distributions, to use what is called the **standard normal distribution**. This is a normal distribution that has a mean of 0, a variance of 1 and probability density function $f(x) = \frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}$ with domain $\mathbb{R}$. Due to its importance this has a special letter, Z, reserved for the random variable of the standard normal distribution.

The standard normal distribution is such that $Z \sim N(0, 1)$, i.e. it has a mean of 0 and a variance of 1.

In the previous section you saw that the probability density function for $X \sim N(\mu, \sigma^2)$ can be obtained from that of $Z \sim N(0, 1)$ by the transformation $x = \sigma z + \mu$. Thus, to transform a normal distribution into the standard normal distribution the transformation $z = \frac{x - \mu}{\sigma}$ is applied. That is, the mean μ is subtracted from the observed value, X, and the result is divided by the standard deviation σ. As stated above, the resulting standard normal distribution is usually referred to by the letter Z.

If $X \sim N(\mu, \sigma^2)$ then $\frac{x - \mu}{\sigma} = z$, $Z \sim N(0, 1)$.

Before digital technology was easily available, tables of values were used to calculate probabilities associated with normal distributions. This was one of the reasons it was so important to be able to change a distribution to the standard normal: it was the only distribution for which tables of values were easily available. In subjects such as this it is now used mainly for comparisons.

However, these standard z values (also called z-scores) have some important applications in science and elsewhere, such as in the field of paediatric health. Children grow at different rates, so it is difficult to use the more standard statistical tests when assessing particular children for serious health defects or conditions. For these applications to be useful, large statistical samples are needed for children of various ages, heights and weights. Then, comparisons can be made for a particular child against other children of the same age/height/weight, or some combination of these factors. For example, it may be suspected that a child has an inappropriately large dilation of a ventricle in their heart. If the dilation increases over time, this may not be a problem, because the dilation would always be expected to increase as the child grows older. However, if the z value increases over time, this is evidence that there may be a problem.

Calculating a *z* value

Example 7

X is a random variable following a normal distribution with mean 10 and variance 9 (i.e. $X \sim N(10, 9)$).
Find the z value that would be used to represent an x value of 14.

Solution

Identify σ and μ: $\mu = 10$, $\sigma = \sqrt{9} = 3$

Use the formula $z = \frac{x - \mu}{\sigma}$ to convert the x value to the equivalent z value: $z = \frac{14 - 10}{3}$

$= \frac{4}{3}$

≈ 1.33

This means that the observed value of X, 14, is 1.33 standard deviations greater than the mean.

Comparing *z* values

Example 8

Juan has been applying for scholarships. On one particular test he obtained 45 on a test that followed the distribution $X \sim N(40, 4)$ and on another he obtained 85 on a test that followed the distribution $Y \sim N(75, 25)$. On which test did Juan do better?

Solution

Find the z value for the first test: $\frac{X-\mu}{\sigma} = \frac{45-40}{2}$

$= \frac{5}{2} = 2.5$

Find the z value for the second test: $\frac{Y-\mu}{\sigma} = \frac{85-75}{5}$

$= \frac{10}{5}$

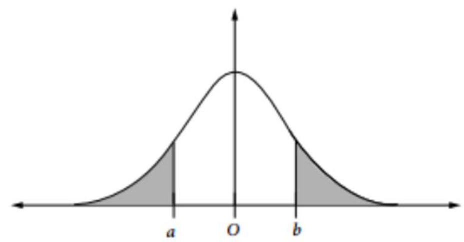
$= 2$

Juan did better on the first test, as his z value was further to the right.

The symmetry of the normal distribution helps with some calculations. This is most easily seen using the standard normal curve, where $\mu = 0$ and $\sigma^2 = 1$ (and therefore $\sigma = 1$).

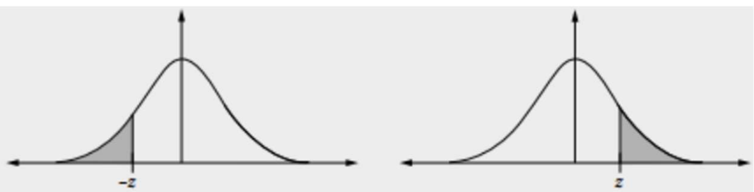
In the diagram the two shaded areas are equal. For this to be true, the distance between a and 0 (μ) must be the same as the distance between 0 and b. So $a = -b$.

This result is useful as it shows that $P(X < a) = P(X > b)$. Also worth noting from this diagram is $P(X > b) = 1 - P(X < b)$ as the total probability is 1.



$P(Z < -z) = P(Z > z) = 1 - P(Z < z)$

A negative z value indicates that the observed value is less than the mean.

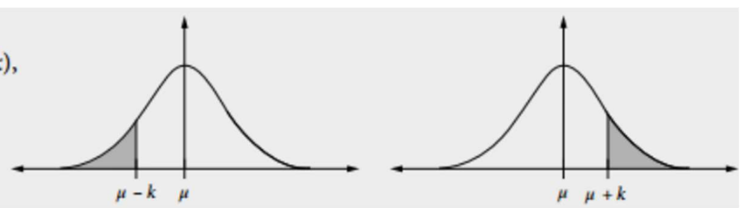


This result can be extended to any variable X that follows a normal distribution.

Assume $\mu = 20$. The value 5 units to the left of the mean is 15 and the value 5 units to the right of the mean is 25. This means, in this case, that $P(X < 15) = P(X > 25)$. This can also be expressed as $P(X < 15) = 1 - P(X < 25)$.

Expressing this result in general terms, for a variable X that follows a normal distribution with a mean of μ:

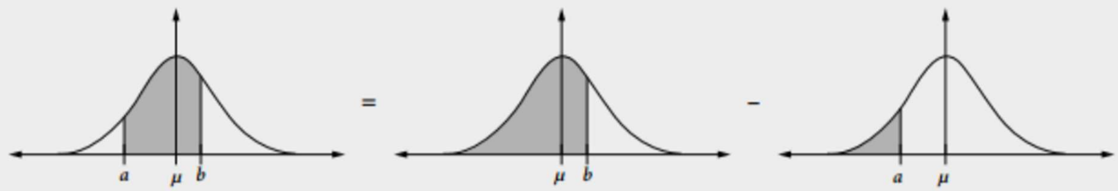
$P(X < \mu - k) = P(X > \mu + k)$ which leads to $P(X < \mu - k) = 1 - P(X < \mu + k)$, where k is the distance from the mean.



This result is useful when dealing with technology-free problems where the symmetry of the distribution is the only real information available to you.

For example, if you needed to find $P(10 < X < 15)$ where $\mu = 12$, you could find $P(X < 10)$ and subtract this from $P(X < 15)$.

Using general terms, to find $P(a < X < b)$:



Using the symmetry of the normal distribution

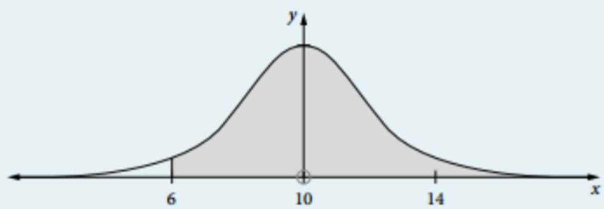
Example 9

A normal distribution graph is shown.

If $P(X > 14) = 0.22$ then $P(X > 6)$ is equal to:

A 0.22 **B** 0.44

C 0.56 **D** 0.78



Solution

Find $P(X > 6)$.

Identify the relevant symmetry aspects of the diagram in terms of probabilities: $P(X > 14) = P(X < 6)$

Write an expression using the required probability: $P(X > 6) = 1 - P(X < 6)$

Substitute known values and calculate the numerical answer: $P(X > 6) = 1 - 0.22 = 0.78$

Hence the answer is D.

Empirical rule

The results obtained earlier apply for the standard normal distribution:

For normally distributed random variables,

- approximately 68% of data will have z-scores between −1 and 1
- approximately 95% of data will have z-scores between −2 and 2
- approximately 99.7% of data will have z-scores between −3 and 3.