MDM4U: Mathematics of Data Management

"Normal" Data

Properties of the Normal Distribution

J. Garvin

Discrete vs. Continuous Data

Recap

Identify the discrete probability distribution that is appropriate for each scenario.

- Tossing heads exactly four times in ten tosses of a fair coin. Binomial.
- Obtaining three hearts in a random seven-card hand. Hypergeometric.
- Counting the number of rolls of a die until a four is rolled. Geometric.

Discrete vs. Continuous Data

Each of the previous scenarios involves discrete data – there are "gaps" between values of the random variable involved.

Some data are continuous – they can assume *any* value within a specified range.

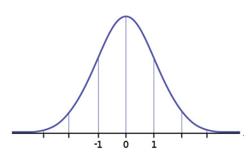
Common examples of continuous data are height, distance, mass, temperature, etc.

If data are continuous, a discrete probability distribution cannot be used as a model. Instead, we must choose a continuous probability distribution.

Properties of the Normal Distribution

The most common continuous probability distribution is the *Gaussian distribution*, or *normal distribution*.

It is a symmetric, bell-shaped distribution (thus commonly called the "bell curve").

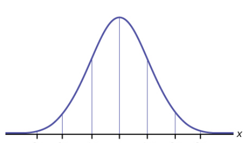


It has the equation $y = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, where $e \approx 2.71828$, and x is the value of the random variable.

Properties of the Normal Distribution

In the standard normal distribution, the mean has a value of $\mu = 0$, and a standard deviation of $\sigma = 1$.

A more general equation, $y = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2[(x-\mu)/\sigma]^2}$, describes a normal distribution with mean μ and standard deviation σ.

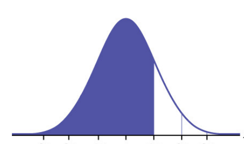


Such distributions can be transformed to the standard normal distribution using z-scores.

Properties of the Normal Distribution

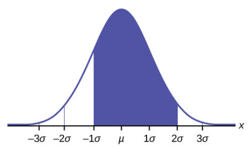
The total area under the curve is 1, and the distribution never touches the x-axis.

The area from the left of the curve up to some value $x = k$ represents the probability that x is less than or equal to k, or $P(x \leq k)$.



Properties of the Normal Distribution

More specifically, the area between any two values $x = a$ and $x = b$ represents the probability that x is between a and b, or $P(a \leq x \leq b)$.



Tables have been created to find these areas. We will look at these in the next lesson.

Properties of the Normal Distribution

Since every datum below the mean has a corresponding datum above the mean, the median is equal to the mean in a normal distribution.

Also since normally distributed data are symmetric about the mean, 50% of all data are above the mean, while 50% lie below it.

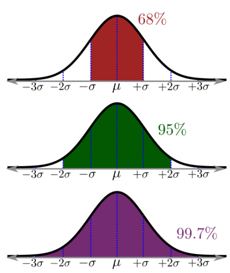
Virtually all data fall within three standard deviations of the mean, with the majority of the data falling within only two.

This is known as the *empirical rule* for the normal distribution.

Properties of the Normal Distribution

Empirical Rule for the Normal Distribution

In a normal distribution, 68% of the data lie within one standard deviation of the mean, 95% within two standard deviations, and 99.7% within three standard deviations.



Properties of the Normal Distribution

Example

A normal distribution has a mean of 150 and a standard deviation of 12. What range of values represents the central 68% of the data?

The central 68% fall within one standard deviation of the mean, so the values are $138 - 162$.

Example

A normal distribution has a mean of 41.7 and a standard deviation of 3.9. What range of values represents the central 95% of the data?

The central 95% fall within two standard deviations of the mean, so the values are $33.9 - 49.5$.

Properties of the Normal Distribution

Example

A normal distribution has a mean of 82 and a standard deviation of 3. Determine the z-score of a datum with a value of 91.

Using the formula for z-scores, $z = \frac{91 - 82}{3} = 3$.

Example

A normal distribution has a mean of 75 and a standard deviation of 14. How many standard deviations below the mean is a datum with a value of 47?

Using the formula for z-scores, $z = \frac{47 - 75}{14} = -2$. Thus, the datum is 2 standard deviations below the mean.

Properties of the Normal Distribution

Example

A normal distribution has a mean of 271. A datum located 4 standard deviations above the mean has a value of 285. What is the standard deviation?

$4 = \frac{285 - 271}{\sigma}$, or $\sigma = \frac{285 - 271}{4} = 3.5$.

Example

A normal distribution has a standard deviation of 6.2. A datum located 2 standard deviations below the mean has a value of 48.9. What is the mean?

$-2 = \frac{48.9 - \mu}{6.2}$, or $\mu = 48.9 + 2 \times 6.2 = 61.3$.

Properties of the Normal Distribution

Example

A new species of fish is discovered, with an average adult length of 13.7 cm and a standard deviation of 0.8 cm. What percentage of such fish are up to 14.5 cm long?

Since $z = \frac{14.5 - 13.7}{0.8} = 1$, a 14.5 cm fish is one standard deviation above the mean.

50% of the data are equal to or less than 13.7 cm.

An additional $68/2 = 34\%$ are between 13.7 cm and 14.5 cm.

Therefore, approximately $50 + 34 = 84\%$ of all such fish are up to 14.5 cm long.

Questions?