Chapter 6 - Normal Distribution

The normal distribution is the most commonly used continuous probability density function. It has the largest set of statistical tools. Representing data as normal provides the researcher with a large set of tools.

Properties - continuous probability distribution

1. The events described by these distributions are continuous (not countable). What this means is that for at two events in the distribution, there will exist a third event lying between these two events. This implies that between any two events, there exists an uncountable set of events.

2. $P(x = k) = 0$ for any k in the set of events. This sounds strange but recall that within a tiny interval around K, there is an uncountable set of values. The probability that x is k becomes zero.

3. Shape of the function can best be described as a continuous probability function f(x) and best represented graphically using a line chart. The end points are $-\infty$ and ∞. The probability (f(x)) approaches zero as x approaches $-\infty$ or

∞. Because the probability is a measurement of the area under this function, we refer to this area as the density of the function (or the probability density function).

4. $P(x \leq k) = \int_{-\infty}^{k} f(x)dx$
   The area (or density) under the curve from its left tail to k measures the cumulative probability (F(x)) that $x \leq k$.

5. $P(x \geq k) = \int_{k}^{\infty} f(x)dx$
   The area (or density) under the curve from k to its right tail measures the cumulative probability (1-F(x)) that $x \geq k$.

6. $\int_{-\infty}^{\infty} f(x)dx = 1.0$
   The area under the curve from its left tails to its right tail is 1.0

7. $P(a < x < k) = \int_{a}^{b} f(x)dx =$
   $(\int_{-\infty}^{b} f(x)dx - \int_{-\infty}^{a} f(x)dx)$
   The area under the curve from a to b measures the probability that x is between a and b.

Normal distribution pdf - f(x)

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Normal distribution cdf

$$P(x \leq k) = \int_{-\infty}^{k} f(x)dx$$

Properties of the normal distribution

1. Center of the distribution is μ.
2. Distribution is symmetric
3. Symmetry implies that the probability that $x > \mu$ or $x < \mu$ is 0.50.
4. The shape is determined by σ^2.

Mathematica example

Normal Distribution Tables????

One would need to prepare a table for each value of μ and σ^2.

Solution: Standard Normal Distribution

1. We know from before

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

   This means that

   $$x = z\sigma + \mu$$

   Replace x in the normal distribution function with $z\sigma + \mu$ and after some algebraic steps, the probability density function (pdf) of the standard normal distribution is

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}z^2}$$

Properties of the standard normal distribution

1. Center of the distribution is $\mu = 0$.
2. Distribution is symmetric.
3. Symmetry implies that the probability that $x > \mu$ or $x < \mu$ is 0.50.
4. The shape is determined by σ^2 =1 $\rightarrow$ the shape is constant for all random variables.

Calculating probabilities ($P(a < x < b)$) which is equivalent to $P(a <= x <= b)$

1. Compute z scores for a and b - let's assume that $z(a) > z(b)$.
2. Table behind the front cover provides you with the probability values from z score from μ (0) to the absolute value of a or b.
3. Drawing the graph and determining the location of $z(a)$ and $z(b)$ relative to the mean of 0 is very helpful.

4. Suppose $z(a) > 0$ and $z(b) > 0$

$$P(z(a) < z < z(b)) =$$

$$P(z(\mu) < z < z(b)) - P(z(\mu) < z < z(a))$$

5. Suppose $z(a) < 0$ and $z(b) > 0$

$$P(z(a) < z < z(b)) =$$

$$P(z(\mu) < z < z(b)) + P(z(\mu) < z < z(a))$$

Normal approximation to a Binomial Distribution

If the following condition holds for a binomial distribution.

np > 5 and nq >5

The standard normal distribution approximates the binomial probabilities.

Correction factor

The binomial distribution is discrete and doesn't fit a continuous distribution. To correct for this, an adjustment, called a continuity correction is done. The value of the the binomial distribution are corrected in the following way.

1. Suppose the question ask you to compute the probability for a range of values and a lower bound (a) is given. In using the standard normal distribution, find a z value for (a-0.5).

2. Suppose the question ask you to compute the probability for a range of values and an upper bound (b) is given. In using the standard normal distribution, find a z value for (b+0.5).

Steps to compute approximation

1. Confirm that $np > 5$ and $nq > 5$ - **Important**
2. Correct the binomial value using the method described above.
3. Convert binomial value into a z score.
4. Use the z score tables to calculate approximate probability.