# 10 Chapter 6 - Normal Probability Distribution

Oh lord: Please make the world normal, linear, and independent.

A statistician's (and an applied economist's) dream

1. In this section, we will examine the normal and standard normal distribution functions. They are probability functions in the set of **continuous** distribution functions. Here are the properties of continuous probability functions.
   (a) 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.
   (b) $P(x = k) = 0$ 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.
   (c) 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 $\infty$. The probability (f(x)) approaches zero as x approaches $-\infty$ or $\infty$. 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 ).
   (d) $P(x \leq k) = \int_{-\infty}^{k} f(x)dx$ - The area (or density) under the curve from its left tail to k measures the cumulative probability (F(x)) that $x \leq k$.
   (e) $P(x \geq k) = \int_{k}^{\infty} f(x)dx$ - The area (or density) under the curve from k to its right tail measures the cumulative probability (1-F(x)) that $x \geq k$.
   (f) $\int_{-\infty}^{\infty} f(x)dx = 1.0$ - The area under the curve from its left tails to its right tail is 1.0
   (g) $P(a < x < b) = \int_{a}^{b} f(x)dx = (\int_{-\infty}^{b} f(x)dx - \int_{-\infty}^{a} f(x)dx)$ - The area under the curve from a to b measures the probability that x is between a and b.

2. **Normal Distribution**
   (a) Probability density function (f(x))

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

   where $\pi \approx 3.1416$ and e is the exponential operator. $\mu$ and $\sigma^2$ are the mean and variance of the population.

(b) Center of the distribution is $\mu$.

(c) Distribution is symmetric

(d) Symmetry implies that the probability that $x > \mu$ or $x < \mu$ is 0.50.

(e) The shape is determined by $\sigma^2$.

3. **Standard Normal Distribution**

   Given that the shape of the normal distribution varies by $\sigma^2$, computing the probabilities of different combinations of $\mu$ and $\sigma^2$ is very cumbersome.

   Statisticians found that if you convert random variables into *standardized* random variables (or Z scores), one can define a single set of probabilities that works for any normal distributions. This set makes up the standard normal distribution.

   (a) 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

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

   (b) Center of the distribution is $\mu = 0$.

   (c) Distribution is symmetric.

   (d) Symmetry implies that the probability that $x > \mu$ or $x < \mu$ is 0.50.

   (e) The shape is determined by $\sigma^2$ =1 $\rightarrow$ the shape is constant for all random variables.

4. Calculating probabilities for a normal distribution (using a table of probabilities from a standard normal distribution).

   (a) Recall: for continuous variables, we are interested in the probability of an interval. Using example 6.5, the value of $\mu$ is 10 and $\sigma$ is 2. The interval we are interested in lies between 11 and 13.6.

      i. Find Z scores for 11 and 13.6 $\rightarrow$ 0.5 and 1.8

      ii. Table in the back of the book (Table 3, Appendix 1) provides you with the probability values from $\mu$ (or zero) to k. Find the probability from 0 to 1.8 and 0 to 0.5.

iii. From before, you learned that $P(a \leq x \leq b) = P(k \leq x \leq b) - P(k \leq x \leq a)$ when $k < a$. Given this $P(0.5 \leq x \leq 1.8) = P(0 \leq x \leq 1.8) - P(0 \leq x \leq 0.5)$ $= 0.4641$ - $0.1915 = 0.2726$

5. Normal approximation of the binomial distribution

   (a) Previously, I mentioned that some variables can be represented by a discrete or normal distribution. Some can only be represented by a discrete distribution so a researcher working with this form of random variable may not be able to use the benefits offered by the normal distribution. One exception is a subset of distributions represented by the binomial distribution.

   (b) Rule of thumb - the normal approximation to the binomial distribution will be adequate if both $np > 5$ and $nq > 5$. If this hold, use the following steps.

      i. Calculate $\mu = np$ and $\sigma = \sqrt{npq}$

      ii. Write the probability in terms of x.

      iii. Correct x by $\pm 0.5$ to include the entire block. Now you have the interval $x_1(x - 0.5)$ and $x_2(x + 0.5)$.

      iv. Compute $z_1$ and $z_2$ based on $x_1$, $x_2$, $\mu$ and $\sigma$. Use the table to compute the probabilities.