5.1: Discrete versus Continuous Distributions

We can describe populations in terms of discrete variables ($x \in \mathbb{Z}$) or continuous variables ($x \in \mathbb{R}$). In the last chapter we saw how to describe discrete probability distributions with the example of the binomial distributions. Discrete probabilities need to be added in inferential statistics and this can lead to complicated formulae. Calculus turns sums into integrals[1] which generally lead to simpler formulae. In the following table we compare, and show the relationship between, discrete and continuous variables and their associated probability distributions.

Discrete	Continuous
• We have a finite number of values between the high and low values • A histogram plot of the random variables x may be interpreted as a **probability distribution**.	• We have an infinite number of values between the high and low values. • With continuous random variables we have a **probability density**.
By increasing the number of values in an appropriate limiting way you make $\longrightarrow$ the discrete probability distribution $\longrightarrow$ approach a probability density.	
• The units of $P(x)$ are probability.	• The units of $P(x)$ are probability density. Probabilities are given by areas under the curve only.

We will be slurring our language and call a probability density, a probability distribution. So we'll say normal distribution instead of normal density. Continuing the comparison, probability distributions and densities have means, moments, skewness, etc. :

- Means and variances of a discrete probability distribution, $P(x)$, are given by the application of the grouped data formulae we saw in Chapter 4 :

$$\mu=\sum_x x\cdot P(x) \qquad \sigma^2=\sum_x[(x-\mu)^2\cdot P(x)] \tag{5.1.1}$$

- Means and variances of a continuous probability density, $P(x)$ are given by the integrals :

$$\mu=\int x\cdot P(x)\,dx \qquad \sigma^2=\int(x-\mu)^2\cdot P(x)\,dx \tag{5.1.2}$$

  Recall that the variance is the second moment of x about the mean μ.

We don't have to stop at the second moment about the mean. The third and fourth moments about the mean are called skewness and kurtosis respectively :

	Discrete	Continuous
Skewness	$\mu_3=\frac{1}{\sigma^3}\sum(x-\mu)^3P(x)$	$\mu_3=\frac{1}{\sigma^3}\int(x-\mu)^3P(x)dx$
Kurtosis	$\mu_4=\frac{1}{\sigma^4}\sum(x-\mu)^4P(x)$	$\mu_4=\frac{1}{\sigma^4}\int(x-\mu)^4P(x)\,dx$

SPSS will easily compute skewness and kurtosis. μ_3 is positive for a positively skewed distribution, negative for a negative skewed distribution. The σ^3 and σ^4 are "normalization" factors; they make the moments of the normal distribution simple.

The moments of a probability distribution are important. In fact, if you specify all the moments of a distribution then you have completely specified the distribution. Let's say that in another way. The specify a probability distribution you can either give its formula (as generally derived from counting) or you can give all its moments. The normal distribution with a mean of μ and a variance of σ^2 is specified by the formula

(5.1) Rendered by QuickLaTeX.com

or by its moments. The normal distribution with a mean of μ and a variance of σ^2 is the only continuous probability distribution with moments (from first to second an on up) of: μ, σ^2, 0, 1, 0, 1, 0, $\ldots$. The normal distribution is special that way among probability distributions.

1. If you have no calculus background, an integral is a way of calculating areas under curves. ↵