

APPENDIX 18.B: CONSTRUCTING A NORMAL PROBABILITY PLOT

This appendix discusses the details of constructing the normal probability plot for the data in Example 18.9. The idea of the normal probability plot is to compare the quantiles of the data with the corresponding quantiles of the normal distribution. Because the shape of the normal distribution is the same whatever the mean and variance, we know that the *relative* distance of normal quantiles will be the same for any normal distribution.¹⁰ For a $\mathcal{N}(0, 1)$ distribution, the cumulative inverse distribution function tells us which x value corresponds to a particular quantile. For example $N^{-1}(0.1) = -1.282$ and $N^{-1}(0.3) = -0.524$. The distance between the x values that give rise to the data quantiles is

$$N^{-1}(0.3) - N^{-1}(0.1) = 0.757$$

$$N^{-1}(0.5) - N^{-1}(0.3) = 0.524$$

$$N^{-1}(0.7) - N^{-1}(0.5) = 0.524$$

$$N^{-1}(0.9) - N^{-1}(0.7) = 0.757$$

The lesser distance between quantiles closer to the median reflects the shape of the normal distribution. If the data come from a normal distribution, the relative distance between data points at these quantiles will have the same *relative* distances.

The procedure in creating Figure 18.6 is to plot the data points on the x -axis against the values from a $\mathcal{N}(0, 1)$ distribution with the same quantiles. Thus, we plot the data point with quantile q against $N^{-1}(q)$. In the case of the five sample data points, we plot the points $(3, -1.282)$, $(4, -0.524)$, $(5, 0)$, $(7, 0.524)$ and $(11, 1.282)$. These points are plotted in the top left panel of Figure 18.6. The y -axis is labeled “ z -value” since these are values from a standard normal distribution. The top right panel is exactly the same, except that the y -axis is labeled with probabilities corresponding to the z -values. The data do not appear normal, though with only five points there is a large possibility for error.

.....
¹⁰For distributions with variable shapes, the relative distance between quantiles can vary with the parameters of the distribution. For this reason, probability plots work best for distributions where the shape is always the same but location (mean) and scale (variance) can vary.