# Figure 5.7. PROBABILITY MODELLING: Normal Calculations

Program 5 in: *Against All Odds: Inside Statistics*

Although the 68–95–99.7 rule allows us to make quick approximate calculations about normal distributions, we also need more exact methods. This program shows in detail how to answer two types of questions for any normal distribution.

1. First, given an interval of possible outcomes, what is its proportion (or relative frequency)?
2. Second, given a proportion such as "the highest 5% of the outcomes", what is the value with that proportion above it or below it?

The video asks and answers the following questions to illustrate normal distribution calculations of type 1:

- The heights of young women can be modelled by a normal distribution with mean $\mu = 65.5$ inches and standard deviation $\sigma = 2.5$ inches; what percentage of these women are taller than 61.7 inches?
- The heights of seven year-old girls can be modelled by a normal distribution with $\mu = 48.7$ inches and $\sigma = 1.9$ inches; what percentage are shorter than 48.5 inches?
- The distribution of nitrogen oxide emissions from a prototype engine is estimated to be normal with $\mu = 0.78$ grams per mile and $\sigma = 0.525$ grams per mile; what percentage of these cars will exceed the regulatory limit of 1.0 grams per mile?
- Blood cholesterol levels (in mg per decilitre) among adults aged 20 to 74 years can be modelled by a normal distribution with $\mu = 213$ and $\sigma = 48.4$; what percentage of the population fall in the borderline risk group with levels between 200 and 250?

Calculations of type 2 are illustrated by a question asked by the U.S. army in deciding which soldiers need custom-made helmets:

- ○ The head circumferences of soldiers are approximately normal with $\mu = 22.8$ inches and $\sigma = 1.1$ inches; what head circumference is exceeded by just 5% of all soldiers?

Both types of calculation use the fact that all normal distributions are the *same* when a standardized scale is used; we now state this fact more exactly:

if $Y$ has a $N(\mu, \sigma)$ distribution, then the *standardized variable* $Z = (Y - \mu)/\sigma$ has the *standard normal* distribution $N(0, 1)$.

Table A on pages T2 and T3 (and on the front flyleaf) of the text, and Figure 5.4 of the Course Materials, give probabilities (or proportions) of the event $Z \leq z$ for many values of $z$ at intervals of 0.01. Normal distribution calculations are done in two steps:

- ∗ First, restate the problem in terms of a standard normal variable $Z$ by standardizing the original problem.
- ∗ Then use Table A of the text, or Figure 5.4 of the Course Materials, to obtain the answer.

Because normal distributions are important in statistics, we need a way to *assess* whether the distribution of a set of data *is* approximately normal. This is best done by looking at a special graph called a *normal quantile plot*, which is available on most statistical software systems. Deviations of the points on such a plot from a straight line show deviations of the data from normality. You do *not* need to be able to construct a normal quantile plot (unless you have software that does it automatically), only to *interpret* the plot. The video shows normal quantile plots for heights of girls (*close to* normal), nitrogen oxide emissions from vehicles (*roughly* normal), and carbon monoxide emissions from the same vehicles (skewed to the right, so *not* normal).



1. A little under *six* minutes into the video (*i.e.*, close to *four* minutes into its actual *contents*), Dr. Amabile measures the height of her daughter Christine. In point form, assess critically, from a *measuring* perspective, the process as we see it.
   - Suggest the *constraint(s)* on the producers of the video that might have been responsible for any *un*desirable features of the process.

2. A little under *eleven* minutes into the video (*i.e.*, close to *nine* minutes into its actual *contents*), there is a description of how General Motors checks whether a prototype of one of its car models is meeting U.S. Federal emission standards (*viz.*, 1 gram per mile) for nitrogen oxides ($NO_X$), constituents of car exhaust gases which have, at high concentrations, been linked to a variety of health problems and to the formation of acid rain. The video (and the third bullet of the summary above) shows the two expressions (in grams per mile): $\mu = 0.78$ and $\sigma = 0.525$. Explain briefly *in words* what (statistical) processes are indicated by these expressions; your explanation will need to include careful use of the words *model*, *population* and *sample*.

3. In the discussion on the video of cholesterol levels as a risk factor for heart disease, Dr. William Castelli (head of the Framingham heart study) describes the borderline risk group as people with levels of 200 to 260, whereas the video calculation, based on the normal model, of a population proportion of 38% applies to levels of 200 to 250 (see the fourth bullet of the summary overleaf). Calculate how much the proportion would *in*crease if it were to apply to Dr. Castelli's somewhat wider range of cholesterol levels.

**Blank page**