

EXAMPLE OF STATISTICAL TECHNIQUES APPLIED TO ANALYSIS OF PAVED RUNWAY SIZES (BIVARIATE NORMAL DISTRIBUTION)

1. NOTATION AND UNITS

a	constants in Equations (5.4) and (5.5)	
b	constants in Equation (6.2)	
F	cumulative frequency	
f	frequency	
L	runway length	ft
r	sample correlation coefficient	
W	runway width	ft
w	transform defined in Equation (8.2)	
Δ	non-dimensional width, $W/25$	
$\zeta[\]$	transform of quantity in square brackets	
$\overline{\zeta[\Delta]}_{\Delta}$	mean value of transform at given value of Δ	
Λ	non-dimensional length, $L/\sigma[L]$	
$\sigma[\]$	standard deviation of quantity in square brackets	
$\sigma^2[\]$	variance of quantity in square brackets	
Φ	cumulative relative frequency or probability	
$\Phi_N^{-1}[\]$	Normally distributed variable whose distribution is Φ	
ϕ	probability density function	

Subscripts

i	numerical value
N	Normal distribution

Issued September 1996
With Amendment A, December 1996

2. INTRODUCTION

This Data Item presents an analysis of data for the length and width of a sample of paved runways to illustrate the use of the Bivariate Normal Distribution. The distribution of runways in terms of their lengths and widths is of relevance to the operations of aircraft, for example in assessing the validity of airworthiness assumptions with respect to the probability of finding any particular width of runway. This can be relevant to the definition of minimum control speeds on the ground or the acceptable cross-wind limit. An introduction to statistical techniques for use in analysis of aircraft performance data is given in Data Item No. 91017 (Reference 1).

3. DATA AND ANALYSIS METHOD

The questions that may be asked about the length and width of paved runways can be summarised as follows.

- (i) What is the probability that a paved runway is less than L in length?
- (ii) What is the probability that a paved runway is less than W in width?
- (iii) What is the probability that a paved runway of width W is less than L in length?
- (iv) What is the probability that a paved runway is less than width W and less than L in length?

Questions (i) and (ii) are answerable by reference to the *marginal probability distributions* of paved runways by length or width. Question (iii) is answerable by reference to the *conditional probability distribution* of paved runways by length given a width. Question (iv) is answerable by reference to the *joint probability distribution* of paved runways by length and width.

Furthermore, in order to make any answer to questions (iii) and (iv) a correlation must exist between the length and width of paved runways. This is the effective precondition that the variables form a *bivariate distribution*. In addition, to facilitate the study of any bivariate distribution it is advisable to transform the variables so that each one is described by a Univariate Normal Distribution; this eases the practical difficulties that can arise in establishing the *cumulative distribution functions* and obviates the need for often complicated algebraic manipulations in order to solve the joint integrals that arise in the process.

It is shown in Reference 1 that, if two statistically independent Normal variables are correlated, they form a Bivariate Normal Distribution and there is a relatively simple relationship between the marginal, conditional and joint distributions. The present Data Item illustrates the process of answering questions (i) through (iv) by finding transforms of the distributions of the length and width in a large sample of paved runways in two regions of the United States of America, showing that these transforms are approximately Normally distributed and simply correlated and then treating the sample as a Bivariate Normal Distribution. A simple numerical example is worked to consolidate the demonstration.

4. SAMPLE

The data for length and width of paved runways are taken from the volumes EC and SC (East Central and South Central) of Reference 2. Because some sites have up to six runways, the sample includes both main and secondary paved runways. A total of 1649 runways are listed. The frequencies in relation to length are given in Table 12.1 and those in relation to width are given in Table 12.2. A convenient choice of interval for listing the length data was made; for the width data, approximations to the runway width classes used by ICAO were chosen to accommodate an apparent tendency in the data for quite wide classifications to

emerge at the greater widths. It may be arguable, formally, that such broad groupings could invalidate the underlying assumption that the distribution is continuous rather than discrete. However, for practical purposes, the sample is acceptably large and the methods sufficiently robust to justify the presumption.

5. MARGINAL PROBABILITY DISTRIBUTION BY LENGTH

As a first step in identifying the probability density function for any sample it is natural to plot the data in the form of an histogram. Figure 1 shows that plot for the sample distribution by length. It is immediately clear that, although the distribution is “uni-modal”, the data are not Normally distributed – the distribution is definitely skew.

Because the Normal Distribution arises naturally in the consideration of random events, the function is readily available. It is reasonable then, to seek a transformation of the length distribution function that relates to the Standard Normal Variable tabulated in Table 11.1 of Reference 1. This variable has no dimensions and is defined with zero mean and unit variance. The first step is to non-dimensionalise the values of runway length, L , by dividing by some convenient length (taken here as the standard deviation of L) to reduce the information to something approaching a standardised form. That is

$$\Lambda = \frac{L}{\sigma[L]} \quad (5.1)$$

where, for these data,

$$\sigma[L] = 1800 \text{ ft.}$$

In addition, the cumulative relative frequency distribution, conventionally defined as

$$\Phi_i = \frac{\sum [f_i]}{\left(1 + \sum [f_n]\right)}, \quad (5.2)$$

is calculated in order to proceed towards the identification of a Normal transformation. Using these values to interpolate inversely in Table 11.1 of Reference 1, the function $\Phi_N^{-1}[\Phi_i]$ is found. This function is the Normally Distributed Variable whose distribution is Φ_i . The values are shown in Table 12.1 and are plotted against Λ in Figure 2. The desired transformation is a function that relates $\Phi_N^{-1}[\Phi_i]$ and Λ . This function is denoted by $\zeta[\Lambda]$ and $\zeta[\Lambda]$ is a Standard Normal Variable.

To find an appropriate function, it has to be recognised first that, unlike in many conventional circumstances, the condition of fit here is not the minimisation of the squares of the distances of the points from the function (least-squares); the criterion in this case is to find a function so that the value of

$$\chi^2 = \frac{(f_N - f)^2}{f_N} \quad (5.3)$$

is less than a specified value – or at the very least is a minimum and lies sufficiently close to a pre-specified value. This criterion effectively introduces the local slope of the function into what is, in practice, an optimising process that begins with the choice of function.

Furthermore, because of the convention used in the definition of the cumulative frequency as given in Equation (5.2), values in the upper tail (high values of L) must be treated with some caution and ought not to be given undue weight when a form of function is postulated and the optimisation process is started. It is also advisable to be wary of non-linear functions that do not tend to some limit; such functions can lead to optimistic predictions in the tails of fitted distributions.

Bearing these considerations in mind, a simple function that is easy to handle, and displays the desired characteristics, is the ratio of two linear expressions,

$$\zeta = \frac{a_0 + a_1 \Lambda}{1 + a_2 \Lambda} \quad (5.4)$$

By re-arranging this expression to

$$\zeta = a_0 + a_1 \Lambda - a_2 \Lambda \zeta, \quad (5.5)$$

the methods of Reference 3 can be used to find a useful first set of values for a_i to begin an iteration, using a simple spreadsheet, for a set of values that satisfies the criterion for χ^2 . In this case the transformation

$$\zeta[\Lambda] = \frac{(4.125\Lambda - 10)}{(1 + \Lambda)} \quad (5.6)$$

was found that gives a value of $\chi^2 = 15.46$, which is not significant at the 5% level for the eight degrees of freedom of the problem (see Reference 1, Section 5.1).

For reference, the probability density function, ϕ , is shown plotted against length in Figure 3 and the cumulative density function, Φ , is shown in Figure 4. These figures are complementary ways of displaying the marginal distribution by length of the sample of paved runways.

6. MARGINAL PROBABILITY DISTRIBUTION BY WIDTH

Following the same process as for the length data in Section 5, a histogram – Figure 5 – may be prepared. This histogram is not as helpful as that for the length data as there is the appearance of “bi-modality” (see Reference 1) in the distribution. However, the cluster plotted at $W = 150$ ft is the accumulation of all the information in the interval $100 \text{ ft} < W \leq 200 \text{ ft}$ so that the second mode (if the underlying distribution is assumed not to be discrete) is an illusion. The plot therefore illustrates that care must be taken in interpreting histograms, particularly when variable bandwidths are used in assembling the information. With this in mind, the histogram may be interpreted as representing a skewed distribution similar to that for the length data.

Proceeding in a manner similar to that followed for the sample of lengths, the runway width, W , is made non-dimensional by dividing by some convenient length (taken here as the smallest band-width),

$$\Delta = \frac{W}{25}. \quad (6.1)$$

The cumulative relative frequency may also be calculated as before. Those values can be used to interpolate linearly in Table 11.1 of Reference 1 to find $\Phi_N^{-1}[\Phi_i]$. The values are shown here in Table 12.2 and are plotted in Figure 6 against Δ . Again, a transformation $\zeta[\Delta]$ is sought so that $\zeta[\Delta]$ is a Standard Normal

Variable. Figure 6 is similar in form to Figure 2, so a ratio of two linear expressions,

$$\zeta = \frac{(b_0 + b_1 \Delta)}{(1 + b_2 \Delta)}, \quad (6.2)$$

is likely to be an appropriate fit. In this case, the transformation,

$$\zeta[\Delta] = \frac{(6.8\Delta - 23.9)}{(\Delta + 6)}, \quad (6.3)$$

gives a value of $\chi^2 = 7.47$ which is not significant at the 5% level for the three degrees of freedom for the problem (see Reference 1, Section 5.1).

The probability density function, ϕ , is shown in Figure 7 and the cumulative probability function, Φ , is shown in Figure 8. These plots are complementary ways of displaying the marginal distribution by width of the sample of paved runways.

7. CONDITIONAL PROBABILITY DISTRIBUTION

The next step in the process of addressing questions (i) to (iv) in Section 3 is to establish whether the transformed lengths and widths are correlated. Figure 9 is a plot of $\zeta[\Lambda]$ against $\zeta[\Delta]$; Figure 10 is a plot of L against W which is shown both for reference and to illustrate the consequences of performing the transformations obtained in Section 5 and Section 6. From Figure 9 it can be seen that, over the majority of the range of the width parameter, the range of the spread of the length-parameter data about the (least-squares) line is relatively even. In other words, the conditional distribution by length parameter has a constant variance for all values of the width parameter*. This constant variance is calculated simply by assuming the correlation coefficient for the sample is applicable to the population. Thus

$$\sigma^2[\zeta[\Lambda]_{\Delta}] = (1 - r^2)\sigma^2[\zeta[\Lambda]] = 0.4868, \quad (7.1)$$

because $\sigma^2[\zeta[\Lambda]] = 1$ and $r = 0.7164$.

By contrast, Figure 10 does not have the same appearance. So, although a correlation could be accomplished directly, the further exercise of establishing confidence intervals and the conditional and joint distributions would not be so simple. In fact, by plotting two variables that are expected to be correlated, it is often possible to detect that the direct use of the Bivariate Normal Distribution will be successful. If the plot has the general appearance of Figure 9 it is highly likely that the data are Normally correlated. Of course, it is usually prudent to use the χ^2 test (see References 1 and 4) to confirm that the marginal distributions are Normal.

The conditional distribution is conveniently displayed as in Figure 11. For several values of paved-runway width, the probability distribution for length is shown. The interpretation of the plot is discussed in Section 9.

* The regression of $\zeta[\Lambda]$ on $\zeta[\Delta]$ is linear and homoscedastic (the dependent variable has a common variance at all values of the independent variable) and therefore satisfies the condition for a Bivariate Normal Distribution; the marginal distributions have already been shown to be Normal.

8. JOINT PROBABILITY DISTRIBUTION

In order to formulate the joint distribution of the width and length, the procedure outlined in Section 9 of Reference 1 can be followed. In this case the values of $\zeta[\Lambda]$ and $\zeta[\Delta]$ are already distributed about their respective means and are also in the required non-dimensional form by the nature of the formulation. The required joint probability distribution is therefore

$$\Phi[W, L] = \frac{1}{2\pi(1-r^2)^{1/2}} \exp \left[-\frac{(\zeta[\Lambda]^2 - 2r\zeta[\Lambda]\zeta[\Delta] + \zeta[\Delta]^2)}{2(1-r^2)} \right]. \quad (8.1)$$

Making the substitution

$$w = \frac{(\zeta[\Lambda] - r\zeta[\Delta])}{(1-r^2)^{1/2}}, \quad dw = \frac{d\zeta[\Lambda]}{(1-r^2)^{1/2}}, \quad (8.2)$$

the cumulative joint probability function can be found from the formulation

$$\Phi[W, L] = \Phi_N[w] \times \Phi_N[\zeta[\Delta]]. \quad (8.3)$$

This joint distribution is plotted in Figure 12. The interpretation of the Figure is discussed in Section 9.

9. DISCUSSION

The process followed in this Data Item establishes a regression (Figure 9) of transformed length $\zeta[\Lambda]$ on transformed width $\zeta[\Delta]$. Since both these variables have been shown to be Normally Distributed with zero mean and unit variance and there exists a linear relation (whose slope is equal to the correlation coefficient) then the variables are Normally Correlated. Thus, by definition, the variables form a Bivariate Normal Distribution and the answers to questions (i) through (iv) of Section 3 follow in a simple manner without complex double integration and algebra.

So, to answer questions (i) and (ii) it is necessary only to consult Figure 4 and Figure 8 for any length or width that is of interest. The same plots may also be used to answer the inverse question: what length (or width) of paved runway is likely to be exceeded at a given value of probability? To answer question (iii) it is necessary either to consult Figure 11 and interpolate "by eye" or to transform the width W and multiply this by the correlation coefficient to provide the mean value of the length parameter $\zeta[\Lambda]_{\Delta}$ at that value of $\zeta[\Delta]$. The variance of $\zeta[\Lambda]_{\Delta}$ is given by Equation (7.1): so the distribution of $\zeta[\Delta]_{\Delta}$ can be calculated. It is Normally Distributed about $\zeta[\Lambda]_{\Delta}$ with variance 0.4868. Table 11.1 of Reference 1 can be used to construct a plot if needed.

The answer to question (iv) may be taken from Figure 12 directly or by using Equations (8.2) and (8.3) for specific values of $\zeta[\Lambda]$ and $\zeta[\Delta]$. Furthermore, because the probabilities are effectively the areas under the curves of the underlying distributions, the proportion of runways with dimensions lying between two given widths and two given lengths can be found by a simple subtraction of the probability values. This procedure is illustrated by example in Section 10.

Using the process outlined in this Data Item provides a route to the interpretation of sometimes vague information. If, however, the conditions necessary for the exploitation of the Bivariate Normal Distribution do not exist or cannot be constructed, it is necessary to resort to the use of other less well-documented bivariate distributions. For guidance in such a case it is recommended that an advanced text in Applied