Linear Combinations of Several Independent normal variables

The standard bivariate normal distribution was defined as the joint distribution of a particular pair of linear combinations of independent standard normal variables X and Z, namely, X and $\rho X + \sqrt{1-\rho^2}Z$. While this representation seems at first artificial, the examples show how it is the basis of all calculations involving the more general bivariate normal distribution, which is obtained by allowing arbitrary means and variances, but insisting that the two standardized variables are standard bivariate normal.

The rotational symmetry of the joint distribution of two independent standard normal variables Z_1 and Z_2 implies that the joint distribution of any two linear combinations of Z_1 and Z_2, say

$$V = a_1 Z_1 + a_2 Z_2 \quad \text{and} \quad W = a_1 Z_1 + a_2 Z_2$$

is bivariate normal. By reducing to this case by scaling, the same conclusion is obtained for any two independent normal variables Z_1 and Z_2 (not necessarily standard). It can be shown that this extends to linear combinations of any number of independent normal variables Z_i:

Two Linear Combinations of Independent Normal Variables

Let

$$V = \sum_i a_i Z_i \quad \text{and} \quad W = \sum_i b_i Z_i$$

be two linear combinations of independent normal (μ_i, σ_i^2) variables Z_i. Then the joint distribution of V and W is bivariate normal.

Granted this, the parameters of the bivariate normal distribution of V and W are easily computed:

$$\mu_V = \sum_i a_i \mu_i \quad \text{and} \quad \mu_W = \sum_i b_i \mu_i$$

$$\sigma_V^2 = \sum_i a_i^2 \sigma_i^2 \quad \text{and} \quad \sigma_W^2 = \sum_i b_i^2 \sigma_i^2$$

$$Cov(V, W) = \sum_i a_i b_i \sigma_i^2$$

$$\rho = Cov(V, W)/\sigma_V \sigma_W$$

Thus the bivariate normal distribution adequately describes the dependence between any two linear combinations of independent normal variables. In particular, this discussion implies the following result:

Independence of Linear Combinations

Two linear combinations $V = \sum_i a_i Z_i$ and $W = \sum_i b_i Z_i$ of independent normal(μ_i, σ_i^2) variables Z_i are independent if and only if they are uncorrelated, that is, if and only if $\sum_i a_i b_i {\sigma_i}^2 = 0$.

Just as the bivariate normal distribution is the joint distribution of two linear combinations of independent normal variables, the *multivariate normal distribution* is the joint distribution of several linear combinations of independent normal variables. It can be shown that several linear combinations of independent normal variables are mutually independent if and only if the covariance between every pair of them is zero. This is a special and important property of normally distributed random variables. It makes covariance and correlation perfectly suited to the analysis of linear combinations of such variables. Keep in mind however, that in general uncorrelated random variables are not necessarily independent.

Exercises 6.5

1. Here is a summary of Pre-SAT and SAT scores of a large group of students.

PSAT scores:	average: 1200	SD: 100
SAT scores:	average: 1300	SD: 90
	correlation: 0.6	

Assume the data are approximately bivariate normal in distribution.

a) Of the students who scored 1000 on the PSAT, about what percentage scored above average on the SAT?

b) Of the students who scored below average on the PSAT, about what percentage scored above average on the SAT?

c) About what percentage of students got at least 50 points more on the SAT than on the PSAT?

2. Data from a large population indicate that the heights of mothers and daughters in this population follow the bivariate normal distribution with correlation 0.5. Both variables have mean 5 feet 4 inches, and standard deviation 2 inches. Among the daughters of above average height, what percent were shorter than their mothers?