Part I

Analysis of Variance

In this section

- introduction to the terminology of *designed experiments*
- extension of statistical testing theory to comparison of more than two population means
- THE ANALYSIS OF VARIANCE (ANOVA) F-TEST

1.1 DESIGNED EXPERIMENTS

Data collection studies typically fall into one of two categories:

(i) Observational studies: the experimenter has no control over the variables under study, and can only measure outcomes.

- The IQ of MAC and PC users
- The relationship between environmental exposure to toxins and health status.

i.e. The experimenter does not control the exposure to variables that may cause changes in the outcome of interest.

This type of study is common in medicine and epidemiology as it is relatively cheap to carry out.

Common type of observational study:

CASE-CONTROL STUDY

Example: Smoking and Lung Cancer.

A study (Doll and Hill, 1950) investigated 649 lung cancer cases and 649 matched healthy controls, both drawn from a population of men in the UK. They found out what proportion in each group were smokers.

Neither health status nor smoking status were controlled by the experimenter, but were merely observed.

	Smokers	Non-smokers	Total
Lung cancer	647	2	649
Controls	622	27	649

This type of study can be unreliable, and cannot uncover all the relationships of interest.

A preferred approach involves the experimenter controlling the variables that cause variation in the other variables.

Note that this may not be ethical in a smoking/lung cancer study.

(ii) Designed experiments: the experimenter can the levels of variables that may affect the variable of interest.

Example: Birthweight study.

GROUP A : 5 or fewer visits

GROUP B : More than 5 visits.

at the control of

(a) Mothers $\longrightarrow$ OBSERVATIONAL STUDY

(b) Doctors $\longrightarrow$ DESIGNED EXPERIMENT

- after each mother is recruited to take part in the study, they are RANDOMLY assigned to either GROUP A or GROUP B. This is termed a

RANDOMIZED EXPERIMENTAL STUDY

This type of study is preferable, but can be more difficult to implement.