# Data Annexe

**Data Set A (Policyholders.xlsx)**

This represents some data held by a motor insurance company on its policyholders.

**Variable Description**

ID policyholder identification code

Age policyholder’s age in years

Sex policyholder’s sex (Male or Female)

Type type of policy (Comprehensive, Third party)

License full years that policyholder has held a valid driving license

The data are as follows:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ID** | **Age** | **Sex** | **Type** | **License** |
| 1 | 17 | Female | Third Party | 0 |
| 2 | 29 | Male | Comprehensive | 5 |
| 3 | 34 | Male | Comprehensive | 12 |
| 4 | 19 | Female | Comprehensive | 2 |
| 5 | 54 | Female | Comprehensive | 31 |
| 6 | 47 | Female | Third Party | 28 |
| 7 | 55 | Male | Comprehensive | 15 |
| 8 | 18 | Male | Third Party | 1 |
| 9 | 23 | Female | Third Party | 2 |
| 10 | 48 | Male | Comprehensive | 29 |

**Data Set B (Diets.xlsx)**

These data relate to the weight losses achieved by two separate samples of 50 human subjects, each of whom undertook one of two different weight reducing diets (A or B).

**Variable Description**

Diet The diet undertaken (A or B)

Wtloss The individual’s weight loss (in kg) following a fixed period on the relevant diet

Note that a *negative* value of Wtloss indicates that the individual’s weight *increased* over the study period.

The data are as follows:

|  |  |
| --- | --- |
| **Diet** | **Wtloss** |
| A | 3.709 |
| A | 7.087 |
| A | 6.754 |
|  |  |
| B | -1.087 |
| B | 1.819 |
| B | 0.074 |
|  |  |

**Data Set C (Superplus.xlsx)**

This data set shows the incomes of 60 male and 60 female cardholders of a certain bank's "Superplus Diamond" card.

**Variable Description**

Sex The cardholder’s sex (F = female, M = male)

Income The cardholder’s gross annual income (in £’000’s)

The data are as follows:

|  |  |
| --- | --- |
| **Sex** | **Income** |
| M | 40.6 |
| M | 54.6 |
| M | 38.6 |
|  |  |
| F | 33.1 |
| F | 35.8 |
| F | 68.8 |
|  |  |

**Data Set D (Brandprefs.xlsx)**

As part of a marketing study, samples of individuals in each of two different demographic areas were asked to state their brand preferences for a certain type of breakfast cereal. Of particular interest were two brands (A and B) made by a certain manufacturer.

**Variable Description**

Area Demographic area (1 or 2)

Brand Preferred brand (A, B or Other)

*Notice that both these variables are nominal.*

The data are as follows:

|  |  |
| --- | --- |
| **Area** | **Brand** |
| 1 | B |
| 1 | Other |
| 1 | A |
|  |  |
| 2 | A |
| 2 | B |
| 2 | A |
|  |  |

**Data Set E (Heather.xlsx)**

In a study on the prevalence of a certain species of heather, two different heathland locations A and B were studied. A number of transepts (pre-defined areas of land) were examined at each location, and classified according to the prevalence of the species of interest.

**Variable Description**

Location Heathland Location (A or B)

Prevalence Prevalence of the target species (Absent, Sparse or Abundant)

Unlike the data in data set D, which were presented in "raw" form, these data have been collated into the following frequency table:

|  |  |  |
| --- | --- | --- |
| **Prevalence** | **Location A** | **Location B** |
| **Absent** | 8 | 20 |
| **Sparse** | 22 | 14 |
| **Abundant** | 26 | 10 |

The table entries are the relevant numbers of transepts (frequencies).

**Data Set F (Designs.xlsx)**

The market research staff at a detergent manufacturing company is considering two new, different container designs for a kitchen cleaning product. A pilot study was conducted by placing both containers of the product on sale at the same price in a sample of 10 retail stores for a fixed period of time. The numbers of items of the product sold were recorded for each container design.

**Variable Description**

Store store identification number (1 – 10)

Con1 number of items sold, Container Design 1

Con2 number of items sold, Container Design 2

The data are as follows

|  |  |  |
| --- | --- | --- |
| **Store** | **Con1** | **Con2** |
| 1 | 141 | 118 |
| 2 | 184 | 167 |
| 3 | 132 | 137 |
| 4 | 161 | 168 |
| 5 | 176 | 175 |
| 6 | 196 | 197 |
| 7 | 169 | 143 |
| 8 | 199 | 169 |
| 9 | 150 | 123 |
| 10 | 218 | 197 |

:

**Data Set G (Filtration.xlsx)**

The final process in the production of a chemical product involves filtration to remove impurities in the form of unwanted side products. The production manager wished to compare the effectiveness of two possible filter agents, Agent 1 and Agent 2. Each of 12 batches of the product were prepared, and half of each batch was filtered using Agent 1, and the other half using Agent 2. Following filtration, the amount of impurity (in parts per 1000 by weight) still present in the product was determined.

**Variable Description**

Batch Batch identification number (1 – 12)

Agent1 Impurities present after filtration with Agent 1 (parts per 1000)

Agent2 Impurities present after filtration with Agent 2 (parts per 1000)

The data are as follows:

|  |  |  |
| --- | --- | --- |
| **Batch** | **Agent1** | **Agent2** |
| 1 | 7.7 | 8.5 |
| 2 | 9.2 | 9.6 |
| 3 | 6.8 | 6.4 |
| 4 | 9.5 | 9.8 |
| 5 | 8.7 | 9.3 |
| 6 | 6.9 | 7.6 |
| 7 | 7.5 | 8.2 |
| 8 | 7.1 | 7.7 |
| 9 | 8.7 | 9.4 |
| 10 | 9.4 | 8.9 |
| 11 | 9.4 | 9.7 |
| 12 | 8.1 | 9.1 |

**Data Set H (Carloan.xlsx)**

In the sale of new cars, it is generally the case that the more expensive the car, the greater the profit to the dealer. The sales manager of a car dealership wishes to help the company’s salespersons to make the most profitable sale possible to each customer. One particular area of interest concerns the amount of credit successfully applied for by new car purchasers. The manager has obtained a sample of 30 recently completed new car sales where the customer successfully secured a credit loan.

**Variable Description**

**x**: the gross annual income of the purchaser (in £1000’s)

**y**: the amount of the credit loan obtained by the purchaser (in £1000’s)

The data are as follows:

|  |  |
| --- | --- |
| **X** | **Y** |
| 21.0 | 6.6 |
| 16.8 | 6.0 |
|  |  |
| 48.6 | 12.4 |

**Data Set I (Limb SBP.xlsx)**

During plastic operations to the head, neck or trunk it is inconvenient for the anaesthetist to record the patient's blood pressure from a cuff on the arm. A study was designed to investigate whether blood pressure as measured in the arm could usefully be predicted from a measurement made on the patient's leg. The data below are measurements of systolic blood pressure (SBP) simultaneously measured in the arm and leg of 23 patients anaesthetised with halothane.

**Variable Description**

ID Patient identification code

Arm Patient’s SBP as measured in the arm

Leg Patient’s SBP as measured in the leg

The data are as follows:

|  |  |  |
| --- | --- | --- |
| **ID** | **Leg** | **Arm** |
| 1 | 140 | 115 |
| 2 | 130 | 100 |
|  |  |  |
| 23 | 110 | 110 |