MATLAB Toolbox Envelope: Dataset Description

May 28, 2012

`fiberpaper` *Pulp and paper property*

Description

This data set contains measurements of properties of pulp fibers and the paper made from them.

Usage

```
load fiberpaper.dat
```

Format

A matrix with 62 observations on 8 variables.

Value

`Column 1`	Arithmetic fiber length.
`Column 2`	Long fiber fraction.
`Column 3`	Fine fiber fraction.
`Column 4`	Zero span tensile.
`Column 5`	Breaking length.
`Column 6`	Elastic modulus.
`Column 7`	Stress at failure.
`Column 8`	Burst strength.

References

Johnson, R.A. and Wichern, D.W. (2007). Applied Multivariate Statistical Analysis, 6th edition.

irisf — *Fisher's iris data*

Description

This data set contains measurements of 4 characteristics for 3 species of iris.

Usage

```
load irisf
```

Format

A matrix with 150 observation on 5 variables.

Value

`Column 1`	Indicator of the species: 1 for Iris Setaso, 2 for Iris Versicolor and 3 for Iris Verginica.
`Column 2`	Sepal length.
`Column 3`	Sepal width.
`Column 4`	Petal length.
`Column 5`	Petal width.

References

Johnson, R.A. and Wichern, D.W. (2007). Applied Multivariate Statistical Analysis, 6th edition.

Sales — *The quality of its sales staff*

Description

A firm is attempting to evaluate the quality of its sales staff and is trying to find an examination or series of tests that may reveal the potential for good performance in sales. The firm has selected a random sample of 50 sales people and has evaluated each on 3 measures of performance: growth of sales, profitability of sales, and new-account sales. These measures have been converted to a scale, on which 100 indicates "average" performance. Each of the 50 individuals took each of 4 tests, which purported to measure creativity, mechanical reasoning, abstract reasoning, and mathematical ability, respectively.

Usage

```
load sales.dat
```

Format

A matrix with 50 observations on 7 variables.

Value

Column 1	Index of sales growth.
Column 2	Index of sales profitability.
Column 3	Index of new-account sales.
Column 4	Score on creativity test.
Column 5	Score on mechanical reasoning test.
Column 6	Score on abstract reasoning test.
Column 7	Score on mathematics test.

References

Johnson, R.A. and Wichern, D.W. (2007). Applied Multivariate Statistical Analysis, 6th edition.

---

waterstrider *Water strider data*

Description

Water striders live on the surface of lakes and ponds. These insects grow in five distinct stages before reaching adulthood, these stages are called instars; at each transition thy shed their skins, which are also their skeletons. This data set contains measurements of 8 characteristics for 3 species of water striders: L. dissortis, L. rufoscutellatus and L. esakii. The measurements are taken for the first three instars, when the sex of the water striders are undetermined.

Usage

```
load waterstrider
```

Format

A matrix with 90 observations on 10 variables.

Value

Column 1	Binary indicator: 1 for L. esakii and 0 for other species.
Column 2	Binary indicator: 1 for L. dissortis and 0 for other species.
Column 3	Logarithm (natural base) of length of the first antennal segment. Raw measurements are in millimeters.
Column 4	Logarithm (natural base) of length of the second antennal segment. Raw measurements are in millimeters.
Column 5	Logarithm (natural base) of length of the third antennal segment. Raw measurements are in millimeters.
Column 6	Logarithm (natural base) of length of the fourth antennal segment. Raw measurements are in millimeters.
Column 7	Logarithm (natural base) of length of middle femora. Raw measurements are in millimeters.

Column 8	Logarithm (natural base) of length of middle tibiae. Raw measurements are in millimeters.
Column 9	Logarithm (natural base) of length of hind femora. Raw measurements are in millimeters.
Column 10	Logarithm (natural base) of length of hind tibiae. Raw measurements are in millimeters.

References

Klingenberg, C.R. and Spence, J.R. (1993). Heterochrony and Allometry: Lessons from the Water Strider Genus Limnoporus. Evolution 47, 1834 - 1853. datasets

---

Wheat Protein — *The protein content of ground wheat samples.*

Description

The data are the result of an experiment to calibrate a near infrared reflectance (NIR) instrument for measuring the protein content of ground wheat samples. The protein content of each sample (in percent) was measured by the standard Kjeldahl method. In Fearn (1983), the problem is to find a linear combination of the measurements that predicts protein content. The estimated coefficients can then be entered into the instrument allowing the protein content of future samples to be read directly. The first 24 cases were used for calibration and the last 26 samples were used for prediction.

Usage

```
load wheatprotein.dat
```

Format

A matrix with 50 observations on 8 variables.

Value

Column 1 to Column 6	Measurements of the reflectance of NIR radiation by the wheat samples at 6 wavelength in the range 1680-2310 nm. The measurements were made on the log(1/reflectance) scale.
Column 7	The protein content of each sample (in percent).
Column 8	Binary indicator, 0 for high protein content and 1 for low protein content. The cut off point is if the protein content is smaller than 9.75.

References

Fearn, T. (1983). A misuse of ridge regression in the calibration of a near infrared reflectance instrument.

Index