Package: smbdata (via r-universe)

May 21, 2026

Type Package

Title Data from ``Statistical Methods in Biology''

Version 0.2.0.9000

Description All data in the book ``Statistical Methods in Biology'' by Welham et al. (2015) <doi:10.1201/b17336> with a corresponding documentation and illustrative analysis of the data.

Encoding UTF-8

LazyData true

RoxygenNote 7.3.2

URL https://github.com/emitanaka/smbdata, http://emitanaka.org/smbdata/

BugReports https://github.com/emitanaka/smbdata/issues

Depends R (>= 3.5)

License GPL (>= 3)

Repository https://emitanaka.r-universe.dev

Date/Publication 2026-01-21 01:48:13 UTC

RemoteUrl https://github.com/emitanaka/smbdata

RemoteRef HEAD

RemoteSha 99d508e3c8689b3a4fd6f70751bce9e371560afc

Contents

`airtemp` *Air temperature*

Description

Air temperature measurements (°C) recorded at approximately 9 a.m. on 100 days during 2006 using two instruments: a standard glass mercury dry-bulb thermometer and a new electronic dry-bulb thermistor probe. For each day, the dataset includes the day number and paired temperature readings from both devices, enabling direct comparison between the established and new measurement methods.

Usage

```
airtemp
```

Format

A data frame with 100 rows and 4 variables:

Unit Factor. Unique identifier for each observation.

DayNo Integer. Day number on which the measurement was taken.

Mercury Numeric. Temperature (in degrees Celsius) measured using a mercury thermometer.

Thermistor Numeric. Temperature (in degrees Celsius) measured using a thermistor thermometer.

Source

Welham, S. J., Gezan, S. A., Clark, S. J., and Mead, A. (2015) *Statistical Methods in Biology: Design and analysis of experiments and regression*

Examples

```
# a simple linear regression
fit_ab <- lm(Thermistor ~ Mercury, data = airtemp)
# let intercept be 0
fit_b <- lm(Thermistor ~ 0 + Mercury, data = airtemp)
# test if intercept = 0
anova(fit_b, fit_ab)
# test if slope is equal to 1, given intercept = 0
fit_1 <- lm(Thermistor ~ 0 + offset(Mercury), data = airtemp)
anova(fit_1, fit_b)
```

---

aphids *Pea aphids survey*

Description

The dataset comes from an ecological survey of the pea aphid (Acyrthosiphon pisum). In three fields, 15 randomly selected triplets of adjacent bean plants were inspected, and the total number of pea aphids on each triplet was recorded. The dataset includes the explanatory factor Field (three levels), a Sample identifier for the 15 sampling locations within each field, and the response variable AphidCount, representing the total aphid count per sample, and is used to assess whether aphid infestation differs among fields.

Usage

```
aphids
```

Format

A data frame with 45 rows and 4 variables: `ID`, `Field`, `Sample`, `AphidCount`.

ID Factor. Unique identifier for each observation.

Field Factor. Identifier for the field in which the sample was collected.

Sample Factor. Sample point number within each field.

AphidCount Integer. Number of aphids counted in the sample.

Source

Welham, S. J., Gezan, S. A., Clark, S. J., and Mead, A. (2015) *Statistical Methods in Biology: Design and analysis of experiments and regression*

Examples

```
fit <- glm(AphidCount ~ Field, family = poisson(), data = aphids)
```

`beetles` *Beetle mating*

Description

The dataset comes from a completely randomized experiment investigating the viability of interspecies and intraspecies mating in two willow beetle species (Phratora vitellinae and Phratora vulgatissima). Females from each species were mated with males of either the same or the other species, giving four treatments with 10 replicates per treatment. The dataset includes the factor Treatment, identifying the mating combination, and the response variable Eggs, which records the number of eggs laid by each female, and is used to compare reproductive outcomes across mating types.

Usage

```
beetles
```

Format

A data frame with 5 variables: `DFemale`, `Treatment`, `Species`, `MateType`, `Eggs`.

DFemale Factor. Unique identifier for each female subject in the experiment.

Treatment Factor. Experimental treatment group assigned to each female.

Species Factor. Species designation for each female in the study.

MateType Factor. Type of mate provided for each female ("Inter" for interspecific or "Intra" for intraspecific, as relevant).

Eggs Integer. Number of eggs laid by each female during the observation period.

Source

Welham, S. J., Gezan, S. A., Clark, S. J., and Mead, A. (2015) *Statistical Methods in Biology: Design and analysis of experiments and regression*

Peacock, L., Batley, J., Dungait, J. A. J., Barker, J. H. A., Powers, S. J. & Karp, A. (2004). *A comparative study of interspecies mating of Phratora vulgatissima and P. vitellinae using behavioural tests and molecular markers*. Entomologia Experimentalis et Applicata, 110(3), 231–241.

Examples

```
fit <- lm(log10(Eggs) ~ Species * MateType, data = beetles)
anova(fit)
```

`biomassc` *Measuring soil microbial biomass*

Description

An experiment was conducted to investigate the effects of procedural modifications on measurements of soil microbial biomass carbon, expressed as mg C per kg of soil. The study employed a $2 \times 3 \times 2$ factorial design, testing two sieve sizes, three sample weights, and two shaking times, resulting in 12 distinct treatment combinations. Each combination was replicated four times in a completely randomized design. The response variable recorded was the amount of microbial carbon biomass (C). The purpose of the analysis is to quantify the main effects and possible interactions among sieve size, sample weight, and shaking time, as well as to determine whether any of the alternative procedures yield results within 10

Usage

```
biomassc
```

Format

A data frame with 5 variables: `DSample`, `Size`, `Weight`, `Time`, `C`.

DSample Factor. Experimental unit identifier, representing the replicate number.

Size Factor. Sieve size used for processing soil samples ("Small" or "Large").

Weight Integer. Sample weight used in the protocol.

Time Integer. Duration in minutes of shaking during sample processing.

C Integer. Microbial biomass carbon in soil, measured as mg C per kg soil.

Source

Welham, S. J., Gezan, S. A., Clark, S. J., and Mead, A. (2015) *Statistical Methods in Biology: Design and analysis of experiments and regression*

Examples

```
fit <- lm(C ~ Size * Weight * Time,
          data = biomassc |>
             transform(Weight = factor(Weight),
                       Time = factor(Time)))
anova(fit)
```

`brassica` *Sreening for pathogenicity*

Description

An experiment was conducted to assess the pathogenicity of various fungal isolates on oilseed rape seedlings. Fungal isolates were obtained from two Brassica species, labeled as A and B (Species), and included nine isolates from group A and four from group B (Isolate). The experimental design involved three replicate runs across time (Rep), with each replicate using trays of seedlings—either 22 or 23 seedlings per tray (Tray, with 13 levels; variate: Seedlings)—exposed toeach isolate. Five days after inoculation, the number of resistant seedlings showing no signs of infection (Resistant) was recorded for each tray, and the percentage of resistant seedlings was used as the response variable for analysis. This design allows for evaluation of differences in pathogenicity among isolates, considering variability due to species, replicates, and tray sizes.

Usage

```
brassica
```

Format

A data frame with 9 variables: `ID`, `Rep`, `Tray`, `Species`, `Isolate`, `TypeA`, `TypeB`, `Seedlings`, `Resistant`.

ID Factor. Unique identifier for each observation (row) in the dataset.

Rep Factor. Replicate/run number in the experiment (1–3), representing separate experimental runs across time.

Tray Factor. Tray identifier for the batch of seedlings tested in each replicate (13 levels).

Species Factor. Brassica species from which the fungal isolate was collected ("A" or "B").

Isolate Factor. Identifier for the fungal isolate tested (nine levels across groups; unique within each Species).

TypeA Integer. Isolate identifier re-coded for group A isolates (repeats the "Isolate" value for group A, NA or other coding for group B).

TypeB Integer. Isolate identifier re-coded for group B isolates (repeats the "Isolate" value for group B, NA or other coding for group A).

Seedlings Integer. Number of seedlings tested per tray (22 or 23).

Resistant Integer. Number of seedlings in the tray showing no signs of infection (i.e., counted as resistant) five days after inoculation.

Source

Welham, S. J., Gezan, S. A., Clark, S. J., and Mead, A. (2015) *Statistical Methods in Biology: Design and analysis of experiments and regression*

Examples

```
fit <- aov(log(P / (100 - P)) ~ Species / Isolate + Error(Rep / Tray),
           data = brassica |>
             transform(P = 100 * (Resistant + 1) / (Seedlings + 2)))
summary(fit)
```

`calcium` *Calcium pot trial*

Description

An experiment was conducted to assess the impact of four different calcium concentrations (levels A = 1, B = 5, C = 10, D = 20) on the root growth of plants. The study followed a completely randomized design, with each treatment assigned to five individual plants growing in separate pots, for a total of 20 pots. At the end of the experiment, the total root length (in cm) was measured for each pot. The dataset contains three columns: Pot, a unique identifier for each pot; Calcium, a factor indicating the assigned calcium treatment level; and Length, the measured total root length for each pot. This structure allows for comparison of root growth across the different calcium concentration treatments.

Usage

```
calcium
```

Format

A data frame with 3 variables: `Pot`, `Calcium`, `Length`. #' @format A data frame with the following variables:

Pot Factor. Unique identifier for each pot/experimental unit.

Calcium Factor. Treatment group indicating the relative concentration of calcium applied to each pot (levels: "A" = 1, "B" = 5, "C" = 10, "D" = 20).

Length Integer. Total root length (in centimeters) measured for each pot at the end of the experiment.

Source

Welham, S. J., Gezan, S. A., Clark, S. J., and Mead, A. (2015) *Statistical Methods in Biology: Design and analysis of experiments and regression*

Examples

```
anova(lm(Length ~ Calcium, data = calcium))
```

`calcium2` *Calcium pot trial with alternative representation*

Description

In this experiment, four relative concentrations of calcium (A = 1, B = 5, C = 10, D = 20) were each applied to five individual plants, with treatments assigned in a completely randomized design across 20 pots. After the experimental period, the total root length (in centimeters) of each plant was measured. The resulting dataset includes both the root length measurements and a set of dummy variables representing the levels of the Calcium treatment factor. This structure facilitates statistical analysis of the effects of different calcium concentrations on plant root growth.

Usage

```
calcium2
```

Format

A data frame with 7 variables: `Pot`, `Calcium`, `Length`, `d1`, `d2`, `d3`, `d4`.

Pot Factor. Unique identifier for each pot (experimental unit).

Calcium Factor. Calcium treatment group for each pot, with levels "A" = 1, "B" = 5, "C" = 10, "D" = 20.

Length Integer. Total root length (in centimeters) measured for each pot at the end of the experiment.

d1 Integer. Dummy variable indicating membership in calcium level "A" (1 if Calcium = "A", 0 otherwise).

d2 Integer. Dummy variable indicating membership in calcium level "B" (1 if Calcium = "B", 0 otherwise).

d3 Integer. Dummy variable indicating membership in calcium level "C" (1 if Calcium = "C", 0 otherwise).

d4 Integer. Dummy variable indicating membership in calcium level "D" (1 if Calcium = "D", 0 otherwise).

Source

Welham, S. J., Gezan, S. A., Clark, S. J., and Mead, A. (2015) *Statistical Methods in Biology: Design and analysis of experiments and regression*

Examples

```
anova(lm(Length ~ 1 + d2 + d3 + d4, data = calcium2))
```

`calibrate` *ELISA calibration*

Description

A calibration experiment was conducted to develop an appropriate protocol for an experimental procedure. The study tested three preparation methods (Prep) and four initial concentrations (Conc), combined in a completely randomized design with two replicates per combination. Absorbance values were measured for each solution after application to an ELISA plate and are recorded in the dataset. The data comprise the measured absorbances (Absorbance), the preparation method, and the initial concentration for each unit. One observation (unit 9) was excluded from analysis due to suspected contamination and was marked as missing.

Usage

```
calibrate
```

Format

A data frame with 4 variables: `Unit`, `Prep`, `Conc`, `Absorbance`.

Unit Factor. Unique identifier for each observation or experimental unit.

Prep Factor. Preparation method applied to the sample.

Conc Factor. Initial concentration applied to the sample.

Absorbance Numeric. Measured absorbance value for each sample unit (may contain missing values for invalid readings).

Source

Welham, S. J., Gezan, S. A., Clark, S. J., and Mead, A. (2015) *Statistical Methods in Biology: Design and analysis of experiments and regression*

Examples

```
fit <- lm(log(Absorbance) ~ Prep * Conc,
          data = calibrate |>
             subset(!is.na(Absorbance)))
anova(fit)
```

`competition` *Weed competition experiment*

Description

This dataset arises from a split-plot experiment designed to assess the competitive effects of different weed species and the impact of irrigation on the grain yield of winter wheat. The experimental design included four blocks (Block), with each block containing two whole plots assigned to either irrigation or no irrigation (Irrigation, WholePlot). Each whole plot was further divided into four subplots (Subplot), where different weed species treatments (none, Am, Ga, Sm; Species) were applied. The measured outcome was grain yield (Grain) in each subplot. The hierarchical, nested structure of the experiment (Block/WholePlot/Subplot) allows for analysis of main effects and interactions of weed species and irrigation, while accounting for variation between blocks and plots.

Usage

```
competition
```

Format

A data frame with 7 variables: `ID`, `Block`, `WholePlot`, `Subplot`, `Irrigation`, `Species`, `Grain`.

ID Factor. Unique identifier for each subplot (observation).

Block Factor. Block number in the experiment (four blocks in total).

WholePlot Factor. Identifier for each whole plot within a block (two per block, corresponding to irrigation treatments).

Subplot Factor. Identifier for each subplot within a whole plot (four per whole plot, corresponding to weed species treatments).

Irrigation Factor. Irrigation treatment applied to the whole plot ("yes" or "no").

Species Factor. Weed species sown in each subplot ("-", "Am", "Ga", "Sm"; "-" denotes no weeds).

Grain Numeric. Grain yield (in appropriate units) measured for each subplot of winter wheat.

Source

Welham, S. J., Gezan, S. A., Clark, S. J., and Mead, A. (2015) *Statistical Methods in Biology: Design and analysis of experiments and regression*

Examples

```
fit <- aov(Grain ~ Irrigation * Species + Error(Block/WholePlot/Subplot),
           data = competition)
summary(fit)
```

`conidia` *Conidial release experiment*

Description

This experiment was designed to measure aphid infection rates in response to varying fungal conidia doses, as delivered by sporulating cadavers of two different sources: a clone or a standard source (Source). Inoculation chambers containing aphids were exposed to conidial showers for one of eight time periods, ranging from 0 to 80 minutes (Time). The actual conidial dose received in each treatment was estimated by counting spores deposited on slides (Conidia) placed in the chambers. Each combination of time period and source was replicated across two experimental runs (Run), with separate sources used for each replicate. The time zero category served as a negative control and, as no conidia should be present at this time, resulting zero counts confirm the absence of slide contamination; this category is excluded from analysis. The resulting dataset supports investigation of the relationship between exposure time, conidial dose, and source type under replicated experimental conditions.

Usage

```
conidia
```

Format

A data frame with 7 variables: `ID`, `Run`, `DUnit`, `Source`, `Time`, `Period`, `Conidia`.

ID Factor. Unique identifier for each observation.

Run Factor. Experimental run indicator (each time period and source combination is repeated in two separate runs).

DUnit Factor. Identifier for each experimental unit within a run.

Source Factor. Source of sporulating cadaver ("Clone" or "Standard") providing the fungal conidia.

Time Integer. Duration of exposure (in minutes) to the conidia shower (excluding zero-time controls).

Period Integer. Index for the time period (e.g., 1 for the first non-zero time, 2 for the second, etc.).

Conidia Integer. Number of conidia (spores) counted on slides for the corresponding experimental unit.

Source

Welham, S. J., Gezan, S. A., Clark, S. J., and Mead, A. (2015) *Statistical Methods in Biology: Design and analysis of experiments and regression*

Examples

```
glm(Conidia ~ Run + (log(Time) + Period) * Source,
    data = conidia |>
      transform(Period = factor(Period)),
    family = poisson())
```

`cotton` *Cotton response to herbicide and insecticide*

Description

This experiment investigated the combined effects of five herbicide doses (0.0, 0.5, 1.0, 1.5, and 2.0 lb/acre) and five insecticide doses (0, 20, 40, 60, and 80 lb/acre) on the root growth of cotton plants grown in containers in a glasshouse. The study employed a completely randomized design with four replicates for each of the 25 possible treatment combinations. Three weeks after treatment, the dry root biomass (grams per plant) was measured for each container. The dataset allows for the assessment of the main and interactive effects of herbicide and insecticide doses on cotton root growth.

Usage

```
cotton
```

Format

A data frame with 4 variables: `ID`, `H`, `I`, `Weight`.

ID Factor. Unique identifier for each container or experimental unit.

H Numeric. Herbicide dose applied (lb/acre; one of 0.0, 0.5, 1.0, 1.5, 2.0).

I Integer. Insecticide dose applied (lb/acre; one of 0, 20, 40, 60, 80).

Weight Numeric. Dry root biomass of cotton plants (grams per plant) measured after three weeks.

Source

Welham, S. J., Gezan, S. A., Clark, S. J., and Mead, A. (2015) *Statistical Methods in Biology: Design and analysis of experiments and regression*

Kuehl, R.O. 2000. Design of Experiments: Statistical Principles of Research Design and Analysis (2nd edition). Thomson Learning (Duxbury Press), Pacific Grove, California. 666 pp.

Examples

```
anova(lm(Weight ~ H * I, data = cotton))
```

cross *Genetics of root growth*

Description

This experiment aimed to investigate the genetic basis of root growth in manipulated lines by crossing two male parents (Male: M1, M2) with five female parents (Female: F1–F5), resulting in ten cross combinations. Up to eight seeds per cross were to be grown in a completely randomized design, and the maximum root length (in mm) was measured for each plant three weeks after sowing (Root). However, due to genetic incompatibilities affecting seed viability, several treatment combinations failed to produce the intended number of replicates, resulting in a total of only 30 observations. The dataset enables analysis of genetic effects on root growth while accounting for variable replication.

Usage

```
cross
```

Format

A data frame with 4 variables: `Seed`, `Female`, `Male`, `Root`.

Seed Factor. Unique identifier for each individual seed/planted observation.

Female Factor. Code for the female parent in the cross (levels: "F1" to "F5").

Male Factor. Code for the male parent in the cross (levels: "M1", "M2").

Root Integer. Maximum root length (in millimeters) measured for each seedling three weeks after planting.

Source

Welham, S. J., Gezan, S. A., Clark, S. J., and Mead, A. (2015) *Statistical Methods in Biology: Design and analysis of experiments and regression*

cuttings *Effect of type and size of cutting on willow yield*

Description

This field experiment was conducted to assess whether the type of cutting planted influences the growth and yield of willows, while also considering the effect of initial cutting size. Five different cutting types (Type: A–E) and three cutting sizes (Size: S, M, L) were included, though not all type × size combinations were available. The study employed a randomized complete block design with five blocks (Block) based on cutting type, and the 25 plots were arranged to balance cutting sizes as much as possible across blocks and types. The yield (Yield) measured at the end of the first year served as the response variable, enabling analysis of both main and combined effects of cutting type and size on willow growth.

Usage

```
cuttings
```

Format

A data frame with 6 variables: `ID`, `Block`, `Plot`, `Type`, `Size`, `Yield`.

ID Integer. Unique identifier for each plot/observation.

Block Factor. Block number in the randomized complete block design.

Plot Factor. Plot number within each block.

Type Factor. Cutting type (A, B, C, D, E) planted in each plot.

Size Factor. Cutting size category: S (small), M (medium), or L (large).

Yield Numeric. Willow yield measured at the end of the first year (units as recorded, e.g., g/plot or kg/plot).

Source

Welham, S. J., Gezan, S. A., Clark, S. J., and Mead, A. (2015) *Statistical Methods in Biology: Design and analysis of experiments and regression*

Examples

```
anova(lm(Yield ~ Block + Size * Type, data = cuttings))
```

---

demethylation *Demethylation experiment*

Description

This pilot study aimed to calibrate a scientific protocol by assessing the effect of a demethylation agent on plant phenotype. Seeds were treated with six different doses of the agent, including a zero-dose control, and then sown in trays, with each tray containing seeds treated at the same dose. Each dose was replicated across four trays: two containing 60 plants and two containing 100 plants. The experiment was arranged in a completely randomized design. For each tray, the number of plants exhibiting a normal phenotype (Normal) and the total number of plants (Total) were recorded, with each tray identified by a unique index (DTray). The dataset allows investigation of the relationship between agent dose and the binomially distributed probability of plants showing a normal phenotype.

Usage

```
demethylation
```

Format

A data frame with 4 variables: `DTray`, `Dose`, `Total`, `Normal`.

DTray Factor. Unique identifier for each tray in the experiment.

Dose Numeric. Dose of demethylation agent applied to the seeds (including zero for controls).

Total Integer. Total number of plants in each tray.

Normal Integer. Number of plants in each tray exhibiting a normal phenotype.

Source

Welham, S. J., Gezan, S. A., Clark, S. J., and Mead, A. (2015) *Statistical Methods in Biology: Design and analysis of experiments and regression*

Examples

```
glm(cbind(Normal, Total - Normal) ~ Dose,
    family = binomial(),
    data = demethylation)

glm(cbind(Normal, Total - Normal) ~ log(Dose + 0.1),
    family = binomial(),
    data = demethylation)
```

---

elisa *Elisa absorbance readings*

Description

A dataset was collected consisting of eight ELISA absorbance readings, each corresponding to a different, increasing concentration of a substrate. The primary focus of the analysis is to characterize the relationship between substrate concentration and measured absorbance, facilitating calibration or interpretation of ELISA response as a function of substrate level.

Usage

```
elisa
```

Format

A data frame with 3 variables: `DUnit`, `Concentration`, `Absorbance`.

DUnit Factor. Unique identifier for each ELISA reading.

Concentration Numeric. Substrate concentration used for each reading.

Absorbance Numeric. ELISA absorbance value measured at the given substrate concentration.

Source

Welham, S. J., Gezan, S. A., Clark, S. J., and Mead, A. (2015) *Statistical Methods in Biology: Design and analysis of experiments and regression*

Examples

```
summary(lm(Absorbance ~ log10(Concentration + 1), data = elisa))
```

`examine` *Aphid catch*

Description

The EXAMINE project collected data from 50 suction trap locations across Europe to explore environmental and landscape factors affecting aphid flight timing and abundance. This dataset specifically focuses on the Julian day of the first capture of the aphid species *Myzus persicae* at each site in 1995. Explanatory variables include geographical information (latitude, longitude, altitude), ten meteorologicalvariables (monthly rainfall from October 1994 to May 1995, mean temperature for the coldest 30-day period, and mean temperature for the subsequent 60-day period), and eight land-use variables representing the proportion of land within a 75 km radius used for different purposes (such as forest types, agricultural land, urban areas, and water bodies). These factors were selected for their potential influence on aphid migration patterns, with earlier flight dates expected in warmer and drier regions. The dataset enables analysis of how geography, climate, and landscape usage affect the seasonal timing of aphid arrival across Europe.

Usage

```
examine
```

Format

A data frame with 23 variables: `Trap`, `JDay`, `Latitude`, `Longitude`, `Altitude`, `OctRain`, `NovRain`, `DecRain`, `JanRain`, `FebRain`, `MarRain`, `AprRain`, `MayRain`, `C30Day`, `F60Day`, `ConForest`, `DecForest`, `MixForest`, `Grassland`, `Arable`, `InlandWater`, `Sea`, `Urban`.

Trap Factor. Unique identifier for each suction trap location.

JDay Integer. Julian day of first catch of *Myzus persicae* at the site in 1995.

Latitude Numeric. Latitude (in decimal degrees) of the trap site.

Longitude Numeric. Longitude (in decimal degrees) of the trap site.

Altitude Integer. Altitude (in meters above sea level) of the trap site.

OctRain Numeric. Rainfall (mm) at the trap site in October 1994.

NovRain Numeric. Rainfall (mm) at the trap site in November 1994.

DecRain Numeric. Rainfall (mm) at the trap site in December 1994.

JanRain Numeric. Rainfall (mm) at the trap site in January 1995.

FebRain Numeric. Rainfall (mm) at the trap site in February 1995.

MarRain Numeric. Rainfall (mm) at the trap site in March 1995.

AprRain Numeric. Rainfall (mm) at the trap site in April 1995.

MayRain Numeric. Rainfall (mm) at the trap site in May 1995.

C30Day Numeric. Mean temperature (°C) for the coldest consecutive 30-day period at the site.

F60Day Numeric. Mean temperature (°C) for the following 60-day period after the coldest period at the site.

ConForest Numeric. Proportion of land (within 75 km radius) under coniferous forest.

DecForest Numeric. Proportion of land (within 75 km radius) under deciduous forest.

MixForest Numeric. Proportion of land (within 75 km radius) under mixed forest.

Grassland Numeric. Proportion of land (within 75 km radius) as grassland.

Arable Numeric. Proportion of land (within 75 km radius) as arable land.

InlandWater Numeric. Proportion of land (within 75 km radius) as inland water.

Sea Numeric. Proportion of area (within 75 km radius) that is sea.

Urban Numeric. Proportion of land (within 75 km radius) classified as urban area.

Source

Welham, S. J., Gezan, S. A., Clark, S. J., and Mead, A. (2015) *Statistical Methods in Biology: Design and analysis of experiments and regression*

Examples

```
step(lm(JDay ~ . - Trap, data = examine), direction = "backward")
```

---

forage *Forage maize yields*

Description

An experiment conducted at Rothamsted Research in 1996 examined how varying rates of nitrogen fertilizer affect the yield of forage maize. The study used a randomized complete block design with three blocks, each containing four plots randomly assigned one of four nitrogen application rates: 0, 70, 140, or 210 kg N/ha. For each plot, whole crop forage yield (measured at 100 percent dry matter, in tonnes per hectare) was recorded. The resulting dataset enables analysis of the relationship between nitrogen fertilizer input and maize yield, with blocking incorporated to account for field heterogeneity.

Usage

```
forage
```

Format

A data frame with 5 variables: `ID`, `Block`, `Plot`, `N`, `Yield`.

ID Factor. Unique identifier for each observation/plot.

Block Factor. Block number in the randomized complete block design (three levels).

Plot Factor. Plot number within each block (four levels per block).

N Integer. Rate of nitrogen fertilizer applied to the plot (in kg N/ha; values: 0, 70, 140, 210).

Yield Numeric. Whole crop forage yield measured at 100 percent dry matter (in tonnes per hectare, t/ha).

Source

Welham, S. J., Gezan, S. A., Clark, S. J., and Mead, A. (2015) *Statistical Methods in Biology: Design and analysis of experiments and regression*

Examples

```
summary(aov(Yield ~ factor(N) + Error(Block/Plot), data = forage))
```

---

forest *Stand density of mixed nothofagus forest plots*

Description

A survey was conducted on 41 plots of pure or mixed Nothofagus forest located at the foot of the Andes. Each plot was classified into one of three stand types based on the dominant tree species: Coigue (type 1, 13 plots), Rauli (type 2, 9 plots), or Roble (type 3, 19 plots). For each plot, stand density (number of trees per hectare, SD) and mean quadratic diameter (in cm, QD) were recorded. The primary aim of the study was to model how stand density relates to quadratic diameter and to compare this relationship across the three forest stand types.

Usage

```
forest
```

Format

A data frame with 4 variables: `DPlot`, `Type`, `SD`, `QD`.

DPlot Factor. Unique identifier for each forest plot.

Type Factor. Forest stand type classified by dominant Nothofagus species in the plot: "Coigue", "Rauli", or "Roble".

SD Integer. Stand density, recorded as the number of trees per hectare in each plot.

QD Numeric. Mean quadratic diameter (in centimeters) of trees in the plot.

Source

Welham, S. J., Gezan, S. A., Clark, S. J., and Mead, A. (2015) *Statistical Methods in Biology: Design and analysis of experiments and regression*

Dra. Alicia Ortega Z., Universidad Austral de Chile

Examples

```
lm(log(SD) ~ log(QD) * Type, data = forest)
```

---

forest2 *Stand density of mixed nothofagus forest plots alternative representation*

Description

A survey of 41 plots in natural stands of pure or mixed Nothofagus forest at the foot of the Andes was conducted to investigate variation in stand structure. Each plot was classified by the dominant tree species into one of three stand types: Coigue (13 plots), Rauli (9 plots), or Roble (19 plots). For each plot, stand density (number of trees per hectare, SD) and mean quadratic diameter (in cm, QD) were recorded. To facilitate analysis, three dummy variables (d1, d2, d3) were created to represent the stand type factor. The main aim of the study was to model stand density as a function of quadratic diameter and to compare this relationship among the three types of Nothofagus stands.

Usage

```
forest2
```

Format

A data frame with 7 variables: `DPlot`, `Type`, `SD`, `QD`, `d1`, `d2`, `d3`.

DPlot Factor. Unique identifier for each forest plot.

Type Factor. Forest stand type classified by the dominant Nothofagus species: "Coigue", "Rauli", or "Roble".

SD Integer. Stand density, as the number of trees per hectare in each plot.

QD Numeric. Mean quadratic diameter (in centimeters) of trees in the plot.

d1 Integer. Dummy variable for stand type Coigue (1 if Type is "Coigue", 0 otherwise).

d2 Integer. Dummy variable for stand type Rauli (1 if Type is "Rauli", 0 otherwise).

d3 Integer. Dummy variable for stand type Roble (1 if Type is "Roble", 0 otherwise).

Source

Welham, S. J., Gezan, S. A., Clark, S. J., and Mead, A. (2015) *Statistical Methods in Biology: Design and analysis of experiments and regression*

Examples

```
anova(lm(log(SD) ~ log(QD) * (d2 + d3), data = forest2))
```

`heights` *Plant heights in glasshouse*

Description

A glasshouse experiment was conducted to examine how different doses of a growth regulator affect plant height under controlled conditions. Six increasing doses (Dose) were each applied to four replicate plants, with each plant grown in a separate pot. The pots were arranged in a completely randomized design on a bench grid comprising four rows (Row) and six columns (Column). After six weeks, the height of each plant (in cm; Height) was measured. This setup enables analysis of the effect of growth regulator dose on plant height, while accounting for any potential spatial variation across the grid layout.

Usage

```
heights
```

Format

A data frame with 5 variables: `Pot`, `Row`, `Column`, `Dose`, `Height`.

Pot Factor. Unique identifier for each pot (experimental unit).

Row Factor. Row position of the pot in the grid layout on the bench.

Column Factor. Column position of the pot in the grid layout on the bench.

Dose Factor. Applied dose of the growth regulator.

Height Numeric. Plant height (in centimeters) measured six weeks after treatment.

Source

Welham, S. J., Gezan, S. A., Clark, S. J., and Mead, A. (2015) *Statistical Methods in Biology: Design and analysis of experiments and regression*

Examples

```
anova(lm(Height ~ Column + Dose, data = heights))
```

herbicide *Herbicide efficacy*

Description

A factorial experiment was conducted to evaluate the efficacy of three herbicides (Herbicide) across nine different populations of black-grass (Population). Herbicides A and C belong to the same chemical group (Type 1), while herbicide B represents a different group (Type 2). The experiment followed a randomized complete block design with five replicates (Rep), each comprising 27 pots (dummy factor DPot). Six plants were grown per pot, and their combined fresh weight (Fwt, in grams) was measured at the conclusion of the study. The dataset allows investigation of differences in herbicide efficacy both between and within herbicide groups, as well as variation in response among different black-grass populations.

Usage

```
herbicide
```

Format

A data frame with 7 variables: `ID`, `Rep`, `DPot`, `Population`, `Type`, `Herbicide`, `Fwt`.

ID Factor. Unique identifier for each pot (experimental unit).

Rep Factor. Block number in the randomized complete block design corresponding to replicate (1–5).

DPot Factor. Dummy variable indicating the pot number within each block.

Population Factor. Identifier for the black-grass population (e.g., "P1"–"P9").

Type Factor. Chemical group of the applied herbicide (1: group for Herbicides A and C, 2: group for Herbicide B).

Herbicide Factor. Applied herbicide treatment ("A", "B", or "C").

Fwt Numeric. Total fresh weight (in grams) of six black-grass plants grown in the pot.

Source

Welham, S. J., Gezan, S. A., Clark, S. J., and Mead, A. (2015) *Statistical Methods in Biology: Design and analysis of experiments and regression*

R. Hull, Rothamsted Research.

Examples

```
aov(sqrt(Fwt) ~ Herbicide * Population + Error(Rep/DPot), data = herbicide)
```

`ladybird` *Ladybird transmission of fungus*

Description

An experiment was conducted to study the transmission of fungus from ladybirds to aphids on two host plant types (beans or birdsfoot trefoil). The study used containers, each holding one plant with 20 aphids, and varied the fungal load by distributing 5, 10, or 20 sporulating aphid cadavers per plant. For each host plant and fungal load combination, half of the containers were exposed to ladybird foraging for four hours, and half were not, creating a three-way factorial structure: Host (two levels), Cadaver (three levels), and Ladybird presence (two levels). This setup was blocked in two runs (Run), with six replicates per treatment combination per run, resulting in a total of 72 experimental units. Each unit (DPlant) was randomly assigned treatments, and after seven days, the numbers of live (Live) and infected (Infected) aphids were counted. Due to aphid predation by ladybirds, the number of live aphids varied, so the percentage of infected aphids was used to quantify transmission rates. This dataset enables analysis of the main and interactive effects of host plant, fungal load, and ladybird presence on aphid infection rates.

Usage

```
ladybird
```

Format

A data frame with 8 variables: `ID`, `Run`, `DPlant`, `Host`, `Ladybird`, `Cadaver`, `Live`, `Infected`.

ID Factor. Unique identifier for each experimental container (observation).

Run Factor. Experimental run (1 or 2), indicating replicate.

DPlant Factor. Unique identifier for each experimental plant within a run (1–36).

Host Factor. Type of host plant in the container ("beans" or "trefoil").

Ladybird Factor. Indicator for presence ("+") or absence ("-") of ladybird foraging in the container.

Cadaver Integer. Number of sporulating aphid cadavers distributed on each plant (5, 10, or 20).

Live Integer. Number of live aphids remaining in the container after seven days.

Infected Integer. Number of live aphids found to be infected after seven days.

Source

Welham, S. J., Gezan, S. A., Clark, S. J., and Mead, A. (2015) *Statistical Methods in Biology: Design and analysis of experiments and regression*

Examples

```
summary(aov(log(P / (100 - P)) ~ Host * Cadaver * Ladybird + Error(Run/DPlant),
            data = ladybird |>
              transform(P = 100 * (Infected + 1) / (Live + 2),
                        Cadaver = factor(Cadaver))))
```

latinsquare1 *Independent Latin squares*

Description

An experiment was designed to study the effect of petal colour on the influx of pollen beetles into oilseed rape crops. Five distinct shades of petal colour were tested using a Latin Square (LS) design to control for migration direction within the field. Due to high spatial variability in beetle counts observed in previous research, the LS experiment was replicated in two independent adjacent fields to improve the precision of treatment comparisons. The two squares had the same orientation, but row and column effects were considered independent between replicates, ensuring that comparisons reflected only the treatment and spatial trends within each field.

Usage

```
latinsquare1
```

Format

A data frame with 4 variables: `Field`, `Row`, `Column`, `Treatment`.

Field Factor. Identifier for each experimental field, corresponding to replicates of the Latin Square design.

Row Factor. Row position within the Latin Square in each field.

Column Factor. Column position within the Latin Square in each field.

Treatment Factor. Shade of petal colour assigned to the plot (five levels in total, coded as 1–5).

Source

Welham, S. J., Gezan, S. A., Clark, S. J., and Mead, A. (2015) *Statistical Methods in Biology: Design and analysis of experiments and regression*

latinsquare2 *Linked Latin squares*

Description

An experiment was conducted to assess the growth of five different strains of fungus on a novel substrate. Each strain was inoculated onto 10 dishes, which were arranged in vertical stacks of five within a controlled environment (CE) cabinet. The experimental design comprised two replicates of a 5×5 Latin Square (LS), with both replicates placed on the same shelf. This design allowed the investigator to control for potential effects of dish position within each stack (Position) and stack location on the shelf (Stack), both considered independent sources of variation in the study.

Usage

```
latinsquare2
```

Format

A data frame with 4 variables: `Rep`, `Stack`, `Position`, `Treatment`.

Rep Factor. Replicate indicator for each Latin Square (two replicates in total).

Stack Factor. Stack identifier (1–5), indicating the location of each vertical stack on the shelf within each replicate.

Position Factor. Position of the dish within the stack (1–5, from bottom to top or as defined in the experiment).

Treatment Factor. Fungus strain assigned to the dish (coded as 1–5, corresponding to the five fungal strains).

Source

Welham, S. J., Gezan, S. A., Clark, S. J., and Mead, A. (2015) *Statistical Methods in Biology: Design and analysis of experiments and regression*

---

lupin *Lupin trial*

Description

An experiment was conducted to examine the effects of soil type and water availability on the growth of lupin plants grown individually in pots. To control for possible gradients due to temperature and light, the pots were arranged in a square grid and a Latin square design was used, treating rows and columnsas crossed blocking factors. Treatments followed a 2×2 factorial structure, combining two soil types (Soil: clay [C] or sand [S]) and two water supply levels (Water: low [L] or high [H]). Each treatment combination was applied to a group of pots. Plant height (cm) was recorded for each pot at the end of the experiment, allowing assessment of both main and interaction effects of soil type and water availability on lupin growth.

Usage

```
lupin
```

Format

A data frame with 7 variables: `ID`, `Row`, `Column`, `Treatment`, `Water`, `Soil`, `Height`.

ID Factor. Unique identifier for each pot (experimental unit).

Row Factor. Row position of the pot in the Latin Square grid.

Column Factor. Column position of the pot in the Latin Square grid.

Treatment Factor. Combined treatment label for soil type and water supply (e.g., "CH", "CL", "SH", "SL").

Water Factor. Water supply level applied to the pot: "L" (low) or "H" (high).

Soil Factor. Soil type: "C" (clay) or "S" (sand).

Height Numeric. Height (in centimeters) of the lupin plant at the end of the experiment.

Source

Welham, S. J., Gezan, S. A., Clark, S. J., and Mead, A. (2015) *Statistical Methods in Biology: Design and analysis of experiments and regression*

I. Shield, Rothamsted Research.

Examples

```
anova(lm(Height ~ Row + Column + Treatment, data = lupin))
```

---

lupintrial *Lupin variety trial*

Description

A field trial was conducted to compare the performance of fourteen lupin breeding lines, including twelve dwarf (DTN) lines and two non-dwarf (CH-304) lines, with the candidate variety DTN20. The trial followed a randomized complete block design with three blocks, each containing fourteen plots. Among the multiple assessed traits, this dataset focuses on oil yield (t/ha; variate OilYield) measured in each plot. Each experimental unit is identified by its block and plot number (Block and Plot), and the line identity (Line). The data enable analysis of oil yield variation among breeding lines, with comparisons made to the candidate variety under controlled field conditions.

Usage

```
lupintrial
```

Format

A data frame with 6 variables: `ID`, `Block`, `Plot`, `Line`, `NPlant`, `OilYield`.

ID Factor. Unique identifier for each plot (observation).

Block Factor. Block number in the randomized complete block design.

Plot Factor. Plot number within each block.

Line Factor. Identity code of the lupin breeding line (e.g., "DTN84", "CH304-73").

NPlant Numeric. Average number of plants per square metre in the plot.

OilYield Numeric. Oil yield from the plot, measured in tonnes per hectare (t/ha).

Source

Welham, S. J., Gezan, S. A., Clark, S. J., and Mead, A. (2015) *Statistical Methods in Biology: Design and analysis of experiments and regression*

Examples

```
aov(OilYield ~ Line + Error(Block/Plot), data = lupintrial)
```

`phosphorus` *Olsen Phosphorus*

Description

The data originate from the long-term Exhaustion Land field trial at Rothamsted Research, which examines the relationship between crop yield and soil fertilizer inputs. In 1986, the yields of spring barley (variate Yield) were recorded from 20 plots. For each plot, the available soil phosphorus content was also measured using the Olsen P method (variate OlsenP). This dataset allows for the analysis of how variations in soil phosphorus levels influence barley yield under field conditions.

Usage

```
phosphorus
```

Format

A data frame with 3 variables: `DPlot`, `OlsenP`, `Yield`.

DPlot Factor. Unique identifier for each field plot.

OlsenP Numeric. Available soil phosphorus content measured as Olsen P (mg/kg or ppm).

Yield Numeric. Yield of spring barley from each plot (typically in tonnes per hectare, t/ha).

Source

Welham, S. J., Gezan, S. A., Clark, S. J., and Mead, A. (2015) *Statistical Methods in Biology: Design and analysis of experiments and regression*

Examples

```
lm(Yield ~ log10(OlsenP), data = phosphorus)
```

`potato` *Potato yields*

Description

A field experiment was conducted using a randomized complete block design to evaluate the effects of four different fungicides (F1–F4) on potato yield, compared with untreated control plots. The trial consisted of four blocks, each containing five plots, resulting in a total of 20 experimental units. Treatments (control and four fungicides) were randomly assigned to plots within each block. For each plot, yield was measured and recorded. The dataset includes the blocking factors Block (four levels) and Plot (five levels within each block), the treatment factor Fungicide (five levels: control, F1, F2, F3, F4), and the response variable Yield, allowing for the comparison of fungicide efficacy under controlled field conditions.

Usage

```
potato
```

Format

A data frame with 6 variables: `ID`, `Block`, `Plot`, `Type`, `Fungicide`, `Yield`.

ID Factor. Unique identifier for each plot (experimental unit).

Block Factor. Block number in the randomized complete block design.

Plot Factor. Plot number within each block (1–5).

Type Factor. Indicates if the plot is a "Control" or "Treated" (with fungicide).

Fungicide Factor. Fungicide treatment applied in the plot ("Control", "F1", "F2", "F3", or "F4").

Yield Integer. Potato yield for the plot (units as recorded, e.g., kg/plot).

Source

Welham, S. J., Gezan, S. A., Clark, S. J., and Mead, A. (2015) *Statistical Methods in Biology: Design and analysis of experiments and regression*

Examples

```
lm(Yield ~ Block + Fungicide, data = potato)
```

`potatorow` *Potato yields with row data*

Description

This dataset presents individual row yields from a field experiment designed as a randomized complete block design to evaluate the effects of four different fungicides (F1, F2, F3, and F4) compared with untreated control plots on potato yield. The trial was arranged as four blocks, each containing five plots, for a total of 20 units. For this analysis, yield measurements are broken down to the level of individual rows within each plot. The dataset includes classifying factors for Block, Plot, and Row, enabling detailed investigation of within-plot and within-block variation, as well as the overall impact of fungicide treatments on potato yield.

Usage

```
potatorow
```

Format

A data frame with 6 variables: `ID`, `Block`, `Plot`, `Row`, `Fungicide`, `RowYield`.

ID Factor. Unique identifier for each row yield observation.

Block Factor. Block number in the randomized complete block design with four levels.

Plot Factor. Plot number within each block (1–5).

Row Factor. Row number within each plot, corresponding to individual row yields.

Fungicide Factor. Fungicide treatment applied to the plot ("Control", "F1", "F2", "F3", or "F4").

RowYield Integer. Potato yield measured for the individual row (units as recorded, e.g., grams or kilograms).

Source

Welham, S. J., Gezan, S. A., Clark, S. J., and Mead, A. (2015) *Statistical Methods in Biology: Design and analysis of experiments and regression*

Examples

```
summary(aov(RowYield ~ Fungicide + Error(Block/Plot/Row), data = potatorow))
```

prey *Ladybird predation*

Description

This dataset comes from an experiment investigating factors affecting predation by the Harlequin ladybird. Individual ladybirds of known sex (Sex: Female, Male) were introduced into Petri dishes, each containing six prey items—either pea aphids or lacewing larvae (Prey: Aphid, Lacewing). The experiment followed a randomized complete block design with four treatment combinations arranged across 15 rows (blocks), resulting in 60 observations. On each occasion, one ladybird was observed for the number of prey (Eaten) consumed in 60 minutes. Rows 1–15 combine replicates from both multiple occasions and multiple blocks, enabling the analysis of treatment effects while accounting for replicate variation. The data allow for examination of the influence of ladybird sex and prey type on short-term predation rates, with the outcome variable assumed to follow a Binomial distribution.

Usage

```
prey
```

Format

A data frame with 7 variables: `ID`, `Row`, `Dish`, `Sex`, `Prey`, `Eaten`, `Total`.

ID Factor. Unique identifier for each observation (Petri dish).

Row Factor. Block or replicate identifier (1 to 15), combining experimental occasions and spatial replicates.

Dish Factor. Dish number within each row/block (1 to 4, one for each treatment combination).

Sex Factor. Sex of the Harlequin ladybird ("Female" or "Male") in the dish.

Prey Factor. Prey type offered: "Aphid" or "Lacewing".

Eaten Integer. Number of prey consumed by the ladybird in 60 minutes.

Total Integer. Total number of prey items offered in the dish (always 6).

Source

Welham, S. J., Gezan, S. A., Clark, S. J., and Mead, A. (2015) *Statistical Methods in Biology: Design and analysis of experiments and regression*

Examples

```
glm(cbind(Eaten, Total) ~ Row + Sex * Prey, family = binomial(), data = prey)
```

sosr *Weed abundance*

Description

This dataset comes from the UK-wide Farm Scale Evaluations (FSEs) conducted between 2000 and 2003 to assess the ecological effects of genetically modified (GM) herbicide-resistant versus conventional crop management in spring oilseed rape. Each field was divided into two half-fields that received either the GM or conventional treatment (Treatment), with a total of 62 fields sampled over three years (Year) on 37 different farms (Farm). Each field within a farm was uniquely numbered (Field), and half-fields were labelled (DHalf) according to their treatment allocation. For each half-field, the total abundance of weeds (variate Weeds) was recorded after the last GM herbicide application ("post-herbicide"), and baseline seedbank density (variate Seedbank) was measured before sowing. After excluding fields with missing or suspect seedbank data, the dataset comprises 118 half-field observations from 59 fields. This structure enables analysis of how GM and conventional management regimes impact weed abundance, controlling for initial differences in seedbank densities across a spatially and temporally replicated field trial.

Usage

```
sosr
```

Format

A data frame with 8 variables: `ID`, `Farm`, `Field`, `DHalf`, `Year`, `Treatment`, `Weeds`, `Seedbank`.

ID Factor. Unique identifier for each half-field observation.

Farm Factor. Identifier for each farm (37 farms in total).

Field Factor. Field number within each farm (usually 1–3, since different fields were used across years within farms).

DHalf Factor. Half-field number within each field (1 or 2), corresponding to experimental treatment allocation.

Year Integer. Year of the trial, coded chronologically as 1 (2000), 2 (2001), or 3 (2002).

Treatment Factor. Management regime applied to the half-field: "C" (conventional) or "GM" (genetically modified herbicide-resistant crop).

Weeds Integer. Total weed abundance recorded in the half-field after the last GM herbicide application.

Seedbank Integer. Seedbank density (initial seed count) measured in the half-field before sowing.

Source

Welham, S. J., Gezan, S. A., Clark, S. J., and Mead, A. (2015) *Statistical Methods in Biology: Design and analysis of experiments and regression*

Examples

```
summary(aov(log10(Weeds) ~ Year * Treatment + Error(Farm/Field/DHalf),
            data = sosr |>
               transform(Year = factor(Year))))
```

`temperature` *Rothamsted monthly mean temperature*

Description

This dataset contains the monthly mean temperatures recorded at Rothamsted Experimental Station from 1891 to 1990. Each observation represents the mean temperature (variate Temperature) for a specific month and year within this 100-year period. The data are suitable for time series analysis and are expected to exhibit an annual cyclic pattern, making them appropriate for modeling using trigonometric regression methods.

Usage

```
temperature
```

Format

A data frame with 3 variables: `MonthName`, `Month`, `Temperature`.

MonthName Factor. Name of the month (e.g., "January", "February").

Month Integer. Numeric code for the month (1 = January, ..., 12 = December).

Temperature Numeric. Mean temperature for the month (in degrees Celsius).

Source

Welham, S. J., Gezan, S. A., Clark, S. J., and Mead, A. (2015) *Statistical Methods in Biology: Design and analysis of experiments and regression*

Examples

```
lm(Temperature ~ sin(2 * pi * Month / 12) + cos(2 * pi * Month / 12),
   data = temperature)
```

tgw *Thousand grain weights*

Description

A field experiment was conducted using a completely randomized design to evaluate the effect of growth regulator application and oilseed rape variety (B or N) on seed production, with six replicates per treatment combination. During the trial, some plots were grazed by pigeons, raising concerns that this damage could influence plant growth and seed development. The percentage of each plot's area affected by bird grazing (variate Damage, recorded to the nearest 10 percent) was measured to allow for adjustment in the analysis. The main response variable was thousand grain weight (TGW), and treatment combinations reflect all four factorial combinations of growth regulator presence/absence and variety, coded as a single factor (Trt: +B, +N, -B, -N). The dataset enables investigation of the effects of variety and growth regulator on seed weight, while controlling for the confounding influence of bird grazing damage.

Usage

```
tgw
```

Format

A data frame with 6 variables: `Plot`, `GR`, `Variety`, `Trt`, `Damage`, `TGW`.

Plot Factor. Unique identifier for each experimental plot.

GR Factor. Growth regulator application: "+" (with growth regulator) or "-" (without growth regulator).

Variety Factor. Oilseed rape variety: "B" or "N".

Trt Factor. Treatment combination label indicating both growth regulator and variety (one of "+B", "+N", "-B", "-N").

Damage Integer. Percentage of plot area grazed by pigeons, recorded to the nearest 10 percent.

TGW Numeric. Thousand grain weight (TGW) response, measuring the average weight (in grams) of 1000 seeds from each plot.

Source

Welham, S. J., Gezan, S. A., Clark, S. J., and Mead, A. (2015) *Statistical Methods in Biology: Design and analysis of experiments and regression*

Examples

```
lm(Damage ~ Trt, data = tgw)
lm(TGW ~ Damage * Trt, data = tgw)
```

`transect` *Crop transect beetle counts*

Description

A pilot study was carried out to examine the entry pattern of an insect pest (beetles) into a susceptible crop. Researchers hypothesized that beetles infiltrate the crop from the field edge, advancing toward the centre. Following initial detection of beetles in the field, a transect was established running from the field edge toward the centre, with beetle sampling conducted at 2-meter intervals. At each sampling point, beetles were counted on four randomly selected plants to provide replicate measurements for each distance. The resulting data consist of the distance from the crop edge (variate Distance) and the corresponding beetle count (variate Count) for each sampled plant. This structure allows for analysis of spatial trends in beetle infestation across the field transect.

Usage

```
transect
```

Format

A data frame with 4 variables: `DPlant`, `Distance`, `fDist`, `Count`.

DPlant Factor. Unique identifier for each sampled plant.

Distance Integer. Distance (in meters) from the edge of the field along the transect where the plant was sampled.

fDist Factor. Factor-level code for distance group along the transect.

Count Integer. Number of beetles counted on the sampled plant.

Source

Welham, S. J., Gezan, S. A., Clark, S. J., and Mead, A. (2015) *Statistical Methods in Biology: Design and analysis of experiments and regression*

Examples

```
anova(lm(log10(Count) ~ Distance + fDist, data = transect))
```

`triticum` *Diploid wheat*

Description

A dataset was collected to investigate morphological characteristics associated with seed weight in a line of diploid wheat (Triticum monococcum). Measurements were taken on 190 randomly selected seeds, recording five traits for each seed: weight (mg), diameter (mm), length (mm), moisture content (percentage), and endosperm hardness (index value from a single-kernel characterization system). Each seed is uniquely identified by the variable DSeed. The primary aim of the study is to identify which variables, particularly seed length, contribute to variation in seed weight. This dataset enables the analysis of relationships among seed traits within a genetically uniform wheat line.

Usage

```
triticum
```

Format

A data frame with 6 variables: `DSeed`, `Weight`, `Length`, `Diameter`, `Moisture`, `Hardness`.

DSeed Factor. Unique identifier for each seed.

Weight Numeric. Weight of the seed (in milligrams, mg).

Length Numeric. Length of the seed (in millimeters, mm).

Diameter Numeric. Diameter of the seed (in millimeters, mm).

Moisture Numeric. Moisture content of the seed (as a percentage).

Hardness Numeric. Endosperm hardness, measured as a single-kernel characterization system index value.

Source

Welham, S. J., Gezan, S. A., Clark, S. J., and Mead, A. (2015) *Statistical Methods in Biology: Design and analysis of experiments and regression*

Examples

```
lm(Weight ~ Length, data = triticum)
```

`voltage` *Voltage response*

Description

An experiment was carried out to assess the affinity of a sugar transporter protein in plant cells by measuring electric current (Km) in response to a range of substrate-associated voltages. Nine voltage levels, ranging from -160 to 0 mV, were tested (Voltage), and the experiment followed a randomized complete block design with two blocks corresponding to separate experimental occasions (Rep). For each combination, one observation was recorded per block (DUnit, reflecting individual measurement units within each replicate). This dataset enables analysis of how membrane voltage influences transporter activity, with blocking to account for any variation between experimental runs.

Usage

```
voltage
```

Format

A data frame with 5 variables: `ID`, `Rep`, `DUnit`, `Voltage`, `Km`.

ID Factor. Unique identifier for each observation.

Rep Factor. Block number, corresponding to the experimental occasion (replicate).

DUnit Factor. Measurement unit within each replicate (used in place of actual plot or plant randomization).

Voltage Integer. Membrane voltage applied (in millivolts, mV; values range from -160 to 0).

Km Numeric. Measured electric current for the respective voltage (unit appropriate to current, e.g., microamperes).

Source

Welham, S. J., Gezan, S. A., Clark, S. J., and Mead, A. (2015) *Statistical Methods in Biology: Design and analysis of experiments and regression*

Examples

```
lm(log(Km) ~ Rep + Voltage, data = voltage)
```

`weedseed` *Weed seed abundance*

Description

An observational study was conducted to explore the relationship between seed production and plant characteristics in rye-grass. Between 17 and 24 samples were collected at each of four study sites (Site: C, L, P, and W). For each sample point, the total number of seeds per square meter (TotalSeed), the average head length in millimeters (HLength), and the average number of spikelets per head (Spikelets) were measured. This dataset facilitates analysis of how head length and spikelet count may influence seed yield across multiple field sites.

Usage

```
weedseed
```

Format

A data frame with 5 variables: `Sample`, `Site`, `HLength`, `Spikelets`, `TotalSeed`.

Sample Factor. Unique identifier for each sample point.

Site Factor. Study site where the sample was collected (levels: "C", "L", "P", "W").

HLength Numeric. Average head length of rye-grass plants at the sampling point (in millimeters, mm).

Spikelets Numeric. Average number of spikelets per head at the sampling point.

TotalSeed Integer. Total number of seeds per square meter at the sampling point.

Source

Welham, S. J., Gezan, S. A., Clark, S. J., and Mead, A. (2015) *Statistical Methods in Biology: Design and analysis of experiments and regression*

Examples

```
anova(lm(log10(TotalSeed) ~ Site * (HLength + Spikelets) , data = weedseed))
```

wheat *Wheat yields*

Description

In this field trial, the yields of a standard commercial wheat variety and a new 'improved' variety were compared to assess differences in average performance. Seven small plots were planted with each variety, resulting in yield measurements from a total of 14 plots. Yields were determined for each plot and standardized to tonnes per hectare (t/ha). The primary objective of the study was to evaluate whether the new variety offers a significant advantage in yield over the standard commercial type.

Usage

```
wheat
```

Format

A data frame with 3 variables: `DPlot`, `Variety`, `Yield`.

DPlot Factor. Unique identifier for each plot in the field trial.

Variety Factor. Wheat variety grown in each plot ("Commercial" or "Improved").

Yield Numeric. Grain yield from the plot, measured in tonnes per hectare (t/ha).

Source

Welham, S. J., Gezan, S. A., Clark, S. J., and Mead, A. (2015) *Statistical Methods in Biology: Design and analysis of experiments and regression*

Examples

```
lm(Yield ~ Variety, data = wheat)
```

willow *Willow beetle measurements*

Description

A sample of 50 willow beetles (*Phratora vulgatissima*) was collected from a willow crop near Bristol, UK, to investigate morphological variation within the population. For each beetle, measurements were taken of total body length and width, among other characteristics. This dataset enables analysis of size variation and potential relationships between morphological traits in willow beetles from this location.

Usage

```
willow
```

Format

A data frame with 3 variables: `DBeetle`, `Length`, `Width`.

DBeetle Factor. Unique identifier for each beetle in the sample.

Length Numeric. Total body length of the beetle (in millimeters, mm).

Width Numeric. Maximum body width of the beetle (in millimeters, mm).

Source

Welham, S. J., Gezan, S. A., Clark, S. J., and Mead, A. (2015) *Statistical Methods in Biology: Design and analysis of experiments and regression*

Index