Package 'R4HCR'

May 7, 2026

Version 0.1

Date 2024-09-09

Title R for Health Care Research

Description
A collection of datasets that accompany the forthcoming book ``R for Health Care Research".

Depends R (>= 4.4.0)

License MIT + file LICENSE

LazyData true

Encoding UTF-8

Imports irr, mada, meta, metafor, survival

NeedsCompilation no

Author Jason L. Oke [aut, cre, cph] (ORCID:
<https://orcid.org/0000-0003-3467-6677>)

Maintainer Jason L. Oke <jasonoke98@gmail.com>

Repository CRAN

Date/Publication 2024-09-16 07:50:07 UTC

Contents

`Acupuncture` *Acupuncture for Chronic Headache.*

Description

Data from a randomised control trial (RCT) of acupuncture therapy for chronic headaches. The primary outcome was headache severity score measured using a 6-item Likert-type scale at the one-year follow-up.

Usage

```
Acupuncture
```

Format

A data frame with 301 observations on the following 4 variables.

`group` Randomisation group (`0` = Usual care, `1` = Acupuncture treatment).

`pk1` Headache severity score at baseline.

`pk5` Headache severity score at 1 year.

`change` Change score (`pk5 - pk1`).

Details

These are data from a randomised controlled trial comparing acupuncture therapy to usual care (no acupuncture therapy) on headache severity scores in patients with chronic headaches. 401 patients with chronic headache (predominantly migraine) were recruited from general practices in England and Wales. Patients were randomly allocated to receive up to 12 acupuncture treatments over three months or to a control intervention offering usual care. The primary outcome measure was headache score at the one-year follow-up.

Source

Teaching of Statistics in the Health Sciences Resources Portal Community https://www.causeweb.org/tshs/?s=Acupuncture

References

Vickers, A.J., Rees, R.W., Zollman, C.E., McCarney, R., Smith, C.M., Ellis, N., Fisher, P. and Van Haselen, R., 2004. Acupuncture for chronic headache in primary care: large, pragmatic, randomised trial. BMJ, 328(7442), p.744.

Examples

```
data(Acupuncture, package = "R4HCR")

# Checking baseline balance
with(Acupuncture,
  tapply(pk1,group,mean))

# Correlation between change scores and baseline scores
with(Acupuncture,
  cor(I(pk5-pk1),pk1))

# ANCOVA model
lm(pk5 ~ group + pk1, data  = Acupuncture)
```

BCG *Trials of BCG Vaccine against Tuberculosis.*

Description

Data from a meta-analysis of 13 studies of the efficacy of BCG vaccine against Tuberculosis (TB).

Usage

```
BCG
```

Format

A data frame with 13 observations on the following 8 variables.

`trialnam` Name of the trial.

`authors` Authors of the paper.

`startyr` Start year.

`latitude` Latitude in degrees from the equator.

`cases1` Number of TB cases in intervention group.

`tot1` Total number in intervention group.

`cases0` Number of TB cases in control group.

`tot0` Total number in control group.

Source

https://www.biostat.jhsph.edu/~fdominic/teaching/bio656/software/meta.analysis.pdf

References

Colditz GA, Brewer TF, Berkey CS, et al. Efficacy of BCG Vaccine in the Prevention of Tuberculosis: Meta-analysis of the Published Literature. JAMA. 1994;271(9):698–702. doi:10.1001/jama.1994.03510330076038.

Examples

```
require(meta)

data(BCG, package = "R4HCR")

# Meta-analysis using relative risk summary measure
ma5 <- metabin(
sm = "RR",
event.e = cases1,
n.e = tot1,
event.c = cases0,
n.c = tot0,
```

```
studlab = trialnam,
data  = BCG)
```

BMT *Bone Marrow Transplantation.*

Description

A simplified version of the data set printed in Klein and Moeschberger, 2003. Briefly, these data are from a study of 137 patients with acute myelocytic leukemia (AML) or acute lymphoblastic leukemia (ALL) aged 7 to 52 from four centres. Failure time is defined as the time (in days) to relapse or death.

Usage

```
BMT
```

Format

A data frame with 137 observations on the following 3 variables.

`group` Categorisation of the patients' Leukemia (`ALL` = Acute Lymphoblastic Leukemia, `AML-High Risk` = High risk Acute Myelocytic Leukemia, `AML-Low Risk` = Low risk Acute Myelocytic Leukemia).

`time` Failure time, defined as time (in days) to relapse or death.

`status` Disease-free survival indicator (`1` = Dead or Relapsed, `0` = Alive Disease Free).

Details

Bone marrow transplants are a standard treatment for acute leukemia.Recovery following bone marrow transplantation is a complex process and prognosis may depend on a number of different risk factors. Transplantation can be considered a failure when a patient's leukemia returns (relapse) or when he or she dies while in remission (treatment related death).

Source

Klein, J.P. and Moeschberger, M.L., 2003. Survival analysis: techniques for censored and truncated data (Vol. 1230). New York: Springer.

References

see also

Copelan,Biggs, Thompson, et al, Treatment for Acute Myelocytic Leukemia With Allogeneic Bone Marrow Transplantation Following Preparation With BuCy2, Blood, Volume 78, Issue 3, 1991, Pages 838-843, ISSN 0006-4971,

and

https://www.medicine.mcgill.ca/epidemiology/hanley/bios602/Klein-Moeschberger/Klein-Moeschberger-ch
pdf

Examples

```
data(BMT, package = "R4HCR")
```

CA19 *Diagnosis of Pancreatic Cancer with CA19-9 Biomarker.*

Description

Data from a diagnostic accuracy review of imaging techniques and tumor markers for the diagnosis of pancreatic carcinoma.

Usage

```
CA19
```

Format

A data frame with 22 observations on the following 5 variables.

`study` Name of study.

`TP` The number of true positive test results.

`FP` The number of false positive test results.

`FN` The number of false negative test results.

`TN` The number of true negative test results.

Details

Protein cancer antigen 19-9 (CA 19-9) is a test used to monitor response to treatment for cancers such as pancreatic, Bile duct, Colorectal, Stomach, Ovarian and Bladder cancer.

References

Niederau C, Grendell JH. Diagnosis of pancreatic carcinoma. Imaging techniques and tumor markers. Pancreas. 1992;7(1):66-86. doi: 10.1097/00006676199201000000011. PMID: 1557348.

Examples

```
require(mada)

data(CA19, package = "R4HCR")

# Bivariate Reitsma model/HSROC analysis.
reitsma(CA19, method = "ml")
```

CBF *Ciliary Beat Frequency Measurement Using Two Methods.*

Description

These data are a subset of a larger set of data collected by Low et al and reprinted in Hollander et al. The data correspond to two methods for measuring ciliary activity (ciliary beat frequency (CBF)); 1) nasal brushing and 2) the more invasive but accepted method of endobronchial forceps biopsy. The subjects in the study were all men undergoing bronchoscopies for diagnoses of various lung problems. The CBF values are averages of 10 consecutive measurements on each subject.

Usage

```
CBF
```

Format

A data frame with 15 observations on the following 2 variables.

`Nasal` CBF (hertz) measured using nasal brushing method.

`Biopsy` CBF (hertz) measured using endobronchial forceps biopsy method.

Source

Originally from P. P. Low, C. K. Luk, M. J. Dulfano, and P. J. P. Finch (1984).

References

Hollander, M., Wolfe, D.A. and Chicken, E., 2013. Nonparametric statistical methods. John Wiley & Sons.

Examples

```
data(CBF, package = "R4HCR")

# Pearson's r
with(CBF,
cor(Nasal, Biopsy)
)
```

Cotinine *Salivary Cotinine Measurements on Scottish Schoolchildren.*

Description

Duplicate salivary cotinine measurements for 20 Scottish schoolchildren.

Usage

```
Cotinine
```

Format

A data frame with 20 observations on the following 3 variables.

`subject` Subject identifier

`cotinine1` First of two cotinine measurements (ng/ml).

`cotinine2` Second of two cotinine measurements (ng/ml).

Source

Cited as originating from D Strachan (by personal communication), first printed in Bland and Altman (1996).

References

Bland, J.M. and Altman, D.G., 1996. Measurement error proportional to the mean. BMJ: British Medical Journal, 313(7049), p.106.

Examples

```
data(Cotinine, package = "R4HCR")

mean <- rowMeans(Cotinine[,c(2,3)])

range <- abs(Cotinine[,2] - Cotinine[,3])

# error vs the mean.
plot(mean,range, pch=16, xlab = "Average of first and second measurement")
```

Doppler *Cardiac Output Measured by Doppler Echocardiography.*

Description

Cardiac output measured using Doppler echocardiography by two different observers.

Usage

```
Doppler
```

Format

A data frame with 23 observations on the following 2 variables.

`A` Cardiac ouput measured by observer A (litres/minute).

`B` Cardiac ouput measured by observer B (litres/minute).

Details

In a study to assess the inter-observer reproducibility of cardiac output. Twenty-three ventilated patients were measured non-invasively by Doppler echocardiography. From the four-chamber view of the heart, the readings were made by positioning the Doppler sample volume at the mitral anulus plane.

Source

Müller, R. and Büttner, P., 1994. A critical discussion of intraclass correlation coefficients. Statistics in Medicine, 13(23-24), pp.2465-2476.

Examples

```
require(irr)

data(Doppler, package = "R4HCR")

# Intra-class correlation.
icc(Doppler,
model = "twoway",
type = "agreement",
unit = "single")
```

`Duplex` *Duplex Ultrasonography for Detecting Peripheral Aterial Disease.*

Description

Diagnostic performance of duplex and color-guided duplex for detecting peripheral arterial disease (PAD) in 14 studies. PAD is defined as stenosis of 50-99% or an occlusion.

Usage

```
Duplex
```

Format

A data frame with 14 observations on the following 6 variables.

`study` Name of study

`test` Type of ultrasound (`Color` or `Duplex`)

`tp` The number of true positive test results.

`fn` The number of false negative test results.

`tn` The number of true negative test results.

`fp` The number of false positive test results.

Source

de Vries SO, Hunink MG, Polak JF. Summary receiver operating characteristic curves as a technique for meta-analysis of the diagnostic performance of duplex ultrasonography in peripheral arterial disease. Acad Radiol. 1996 Apr;3(4):361-9. doi:10.1016/s1076-6332(96)80257-1. PMID: 8796687.

Examples

```
require(metafor); require(meta)

data(Duplex, package = "R4HCR")

# Fitting the common effects model.

Duplex <- escalc(
  measure = "OR",
  add = 0.5,
  to = "all",
  ai = tp,
  bi = fp,
  ci = fn,
  di = tn,
  data = Duplex)
```

```
Duplex <- within(Duplex,
{
  S = log((fp + 0.5)/(tn + 0.5)) + log((tp + 0.5)/(fn + 0.5))
}
)

ma <- metagen(TE = yi, seTE = vi, data = Duplex,sm = "OR")

metareg(ma, formula = S,method = "FE")
```

`Earnings` *Gelman and Hill's Earnings and Height Data.*

Description

Data from a survey of adult Americans in 1994.

Usage

```
Earnings
```

Format

A data frame with 1192 observations on the following 4 variables.

`earn` Annual earnings (in dollars).

`sex` Sex (1 = men, 2 = women).

`yearbn` Year of birth.

`height` Height (in inches).

Details

This is a subset of the data was used in a number of regression examples in Data analysis using regression and multilevel/hierarchical models by Gelman and Hill (2006).

Source

http://www.stat.columbia.edu/~gelman/arm/software/

References

Gelman, Andrew, and Jennifer Hill. Data Analysis Using Regression and Multilevel/Hierarchical models. Cambridge university press, 2006.

Persico, Nicola, Andrew Postlewaite, and Dan Silverman. "The effect of adolescent experience on labor market outcomes: the case of height (No. w10522)." (2004).

Examples

```
data(Earnings, package = "R4HCR")

mod <- lm(earn ~ height, data = Earnings)

# % variation explained
summary(mod)$adj.r.squared

# regression coefficients.
coef(mod)

# log earnings model
logm <- lm(I(log(earn)) ~ height, data = Earnings)
coef(logm)
```

Endometrial *Exogenous Oestrogens and Endometrial Cancer.*

Description

This is a matched case control study investigated the effect of exogenous oestrogens on the risk of endometrial cancer.

Usage

```
Endometrial
```

Format

A data frame with 126 observations on the following 8 variables.

`set` Matched pair indicator (`1 - 63`).

`case` Indicator for case/control status (`0` = control, `1` = case).

`gallbladder` History of gallbladder disease (`0` = No, `1` = Yes).

`hypertension` History of hypertension (`0` = No, `1` = Yes).

`obesity` Obesity (`0` = No, `1` = Yes).

`estrogen` Any use of estrogen (`0` = No, `1` = Yes).

`age` Age of the women.

`dose` Conjugated estrogen dose (`1` = none, `2` = 0.1-0.299 mg, `3` = 0.3-0.625 mg and `4` = 0.626+ mg).

Details

Investigators matched 63 cases of endometrial cancer with four control women who were alive and living in the community at the time the case was diagnosed, who were born within one year of the case, who had the same marital status, and who had entered the community at approximately the same time. This data set includes all 63 cases and the first matched control, as per the results in Table 7.3 (page 255) of Breslow and Day (1980).

Source

Breslow, N.E., Day, N.E. and Heseltine, E., 1980. Statistical Methods in Cancer Research.

References

Mack, T.M., Pike, M.C., Henderson, B.E., Pfeffer, R.I., Gerkins, V.R., Arthur, M. and Brown, S.E., 1976. Estrogens and endometrial cancer in a retirement community. New England Journal of Medicine, 294(23), pp.1262-1267.

Examples

```
require(survival)

data(Endometrial, package = "R4HCR")

# Conditional logistic regression.
mod2 <- clogit(case ~ estrogen + strata(set), data = Endometrial)

summary(mod2)
```

`Facemasks` *Face Masks while Exercising Trial (MERIT).*

Description

Data from a cross-over randomised controlled study on the effect of face-masks while taking exercise.

Usage

```
Facemasks
```

Format

A data frame with 216 observations on the following 3 variables.

`patid` Participant identifiction number.

`comparison` Variable indiciating which of the three comparisons the outcome corresponds to (Cloth vs None, Surgical vs None, FFP3 vs none).

`delta` Difference in oxygen saturation (SaO2) in percent (%).

Details

These data are from a cross-over randomised controlled study, completed between June 2021 and January 2022. Volunteers were aged 18–35 years, exercised regularly, and had no significant pre-existing health conditions. The primary outcome was change in oxygen saturation. Oxygen saturation levels were measured after exercise whilst wearing a cloth mask, a surgical mask,or filtering facepiece (FFP3) mask, and compared to oxygen saturation levels without any mask, during 4 15 min bouts of exercise. The exercise was running outdoors or indoor rowing at moderate-to-high intensity, with the consistency of distance traveled between bouts confirmed using a smartphone application (Strava). Each participant completed each bout in random order.

References

Jones N, Oke JL, Marsh S, et al. Face masks while exercising trial (MERIT): a cross-over randomised controlled study. BMJ Open 2023;13:e063014.

Examples

```
data(Facemasks, package = "R4HCR")

# focus on cloth - none comparison
t.test(delta ~ 1,
       data = Facemasks,
       subset = comparison == "Cloth - None")
```

FEV *Forced Expiratory Volume Data.*

Description

Pairs of measurements of Forced Expiratory Volume (FEV), taken a few weeks apart from 20 Scottish schoolchildren.

Usage

```
FEV
```

Format

A data frame with 20 observations on the following 3 variables.

`child` Child identification number

`fev1` First FEV measurement

`fev2` Second FEV measurement

Details

The data in table 1 of the original Bland and Altman paper does not correspond to the ANOVA analysis of Table 2. The corrected data does recreate the ANOVA analysis and so is given here.

Source

Corrected data can be found here https://www.bmj.com/content/suppl/1999/03/16/313.7048.41.DC1

References

Bland, JM. & Alman, DG. 1996. Measurement Error and Correlation Coefficients. Br Med J., 313, pp.41-42.

Examples

```
data(FEV, package="R4HCR")

# reshape to long
FEVl <- reshape(FEV,
                direction = "long",
                idvar = "child",
                varying =list(2:3),
                v.names = "fev")

# one-way ANOVA - as per table 2 of Bland and Altman.
anova(lm(fev ~ factor(child), data = FEVl))
```

Framingham *Framingham Heart Study Dataset*

Description

Many versions of the Framingham heart disease dataset exist, this one includes over 4,000 records and includes several cardiovascular disease risk factors such as blood pressure, blood chemistry, smoking history, markers of disease, and cardiovascular outcomes.

Usage

```
Framingham
```

Format

A data frame with 4240 observations on the following 16 variables.

`sex` Sex of participant (`0` = female, `1` = male).

`age` Age (in years).

`education` `1` = 0-11 years, `2` = High School Diploma, GED, `3` = Some College, Vocational School, `4` = College (BS, BA) degree or more.

`currentsmoker` Current cigarette smoking at exam, `0` = Not current smoker, `1` = Current smoker.

`cigsperday` Number of cigarettes smoked each day, `0` = Not current smoker. `1` = 1-90 cigarettes per day.

`bpmeds` Use of Anti-hypertensive medication at exam, `0` = Not currently used, `1` = Current Use.

`prevalentstroke` Prevalent Stroke (`0` = Free of disease `1` = Prevalent disease).

`prevalenthyp` Prevalent Hypertension (`0` = Free of disease `1` = Prevalent disease).

`diabetes` Diabetic according to criteria of first exam treated or first exam with casual glucose of 200 mg/dL or more (`0` = No diabetes, `1` = Diabetes).

`totchol` Serum Total Cholesterol (mg/dL).

`sysbp` Systolic Blood Pressure (mean of last two of three measurements) (mmHg).

`diabp` Diastolic Blood Pressure (mean of last two of three measurements) (mmHg).

`bmi` Body Mass Index, weight in kilograms/height meters squared.

`heartrate` Heart rate (Ventricular rate) in beats/min.

`glucose` Casual serum glucose (mg/dL).

`tenyearchd` Whether the invidividual developed Coronary Heart Disease within ten years (`0` = no, `1` = yes).

Details

The Framingham Heart Study is a long-term, ongoing cardiovascular cohort study of residents of the city of Framingham, Massachusetts. It began in 1948 and is now on its third generation of participants.

Source

https://www.kaggle.com/datasets/aasheesh200/framingham-heart-study-dataset?resource=download https://www.framinghamheartstudy.org

References

For a description of the full data set see here; https://biolincc.nhlbi.nih.gov/media/teachingstudies/FHS_Teaching_Longitudinal_Data_Documentation_2021a.pdf?link_time=2024-05-26_10:36:20.705109

For more details on the Heart study see for example: Mahmood SS, Levy D, Vasan RS, Wang TJ. The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. Lancet. 2014 Mar 15;383(9921):999-1008. PMID: 24084292; PMCID: PMC4159698.

Examples

```
data(Framingham, package = "R4HCR")
```

`Galton` *Galton's Height Data.*

Description

These data are from Galton's 1886 study of human height.

Usage

```
Galton
```

Format

A data frame with 898 observations on the following 9 variables.

`family` Indicator variable for family unit (or parentages).

`father` Height of the father in inches.

`mother` Height of the mother in inches.

`sex` Sex of the child (`M` = Male, `F` = Female).

`height` Height of the child.

`no.children` Number of children in family unit.

`mother.adj` Mother's height multiplied by 1.08.

`height.adj` Adjusted height of the children (see details).

`mid.parent` The "mid-parent" height (see details).

Details

Galton's data comprised 898 adult children from 197 family units (father-and-mother couples). Mid-parent is the mean of the height of the father and of his wife's height multiplied by 1.08. Similarly, adjusted height has the same correction with female children's height also multiplied by 1.08, and male child heights are left unchanged.

Source

Francis Galton, 2017, "Galton height data", Harvard Dataverse

References

Galton, Francis. "Regression towards mediocrity in hereditary stature." The Journal of the Anthropological Institute of Great Britain and Ireland 15 (1886): 246-263.

Stephen Senn, Francis Galton and Regression to the Mean, Significance, Volume 8, Issue 3, September 2011, Pages 124–126.

Examples

```
data(Galton, package = "R4HCR")

# Regression to the mean
lm.mod <- lm(height.adj ~ mid.parent, data = Galton)

su <- summary(lm.mod)

coef(lm.mod)
```

Glucose *Comparison of impedance to insulin-mediated glucose uptake*

Description

Data from the study by Shen et al 'Comparison of impedance to insulin-mediated glucose uptake in normal subjects and in subjects with latent diabetes.

Usage

```
Glucose
```

Format

A data frame with 14 observations on the following 3 variables.

`diabetes` Indicator of whether the person had diabetes (`1`) or not (`0`).

`glucose` Weighted glucose response to an oral glucose tolerance test (mg/100ml).

`impedance` Glucose Impedance (ohms).

Details

These data are originally from Shen et al (1970) and reprinted in Hollander et al (2013). Glucose impedance represents the tissues' insensitivity or resistance to insulin-mediated glucose uptake. It was hypothesised that the newly developed technique of estimating impedance would allow the detection of a difference in glucose uptake efficiency between normal and mildly diabetic subjects. Two groups of normal-weight subjects were studied, one had maturity onset latent diabetes, and the other (matched for age, weight, and percent adiposity) were 'normal'. Impedance data is taken from Table II 'Results of Standard Infusion Studies', whereas the glucose response data is shown in Table 1.

Source

Shen SW, Reaven GM, Farquhar JW. Comparison of impedance to insulin-mediated glucose uptake in normal subjects and in subjects with latent diabetes. J Clin Invest. 1970 Dec;49(12):2151-60. doi: 10.1172/JCI106433. PMID: 5480843; PMCID: PMC322715.

References

Shen SW, Reaven GM, Farquhar JW. Comparison of impedance to insulin-mediated glucose uptake in normal subjects and in subjects with latent diabetes. J Clin Invest. 1970 Dec;49(12):2151-60. doi: 10.1172/JCI106433. PMID: 5480843; PMCID: PMC322715.

Hollander, M., Wolfe, D.A. and Chicken, E., 2013. Nonparametric statistical methods. John Wiley & Sons.

Examples

```
data(Glucose, package = "R4HCR")

# Kendall's Tau.
with(
subset(Glucose, diabetes==0),
cor.test(glucose, impedance,
exact = TRUE,
method = "kendall")
)
```

`Innova` *Rapid Antigen Detection for SARS-CoV-2 by Lateral Flow Assay.*

Description

The number of false positives in negative samples in each evaluation stage of the Innova lateral flow device.

Usage

```
Innova
```

Format

A data frame with 8 observations on the following 3 variables.

`phase` Evalution phase

`fp` Number of false positives

`total` Total number of tests conducted

Details

The Innova LFD was a first-generation Lateral Flow Device (LFD) for rapid point-of-care (POC) SARS-CoV-2 testing. Peto at al conducted a phased evaluation of available SARS-CoV-2 antigen LFDs from 15th August to December 2020 and reported the diagnostic performance of the Innova LFD.

References

Peto, T., Affron, D., Afrough, B., Agasu, A., Ainsworth, M., Allanson, A., Allen, K., Allen, C., Archer, L., Ashbridge, N. and Aurfan, I., 2021. COVID-19: Rapid antigen detection for SARS-CoV-2 by lateral flow assay: A national systematic evaluation of sensitivity and specificity for mass-testing. EClinicalMedicine, 36.

Examples

```
require(meta)

data(Innova, package = "R4HCR")

# Meta-analysis of false-positive fraction
ma1 <- metaprop(event = fp,
n = total,
studlab = phase,
backtransf=TRUE,
data = Innova)
```

`IPNs` *Artificial intelligence for Assessment of Indeterminate Pulmonary Nodules.*

Description

The performance of an artifical intelligence (AI) risk stratification tool for Indeterminate Pulmonary Nodules (IPN's) on chest CT scans.

Usage

```
IPNs
```

Format

A data frame with 200 observations on the following 2 variables.

`cancer` Indicator for an cancerous IPN (`1`) or non-cancerous IPN (`0`).

`rating` AI algorithm score for the likelihod of cancer.

Details

This data set is taken from a retrospective multireader multicase study performed in June and July 2020 on chest CT studies of Indeterminate Pulmonary Nodules (IPNs). An artificial intelligence tool was used to evaluate CT images and provide an estimated probability of cancer (from 0 to 100).

Source

This data set represents a subset of the orginal data.

References

Kim, R.Y., Oke, J.L., Pickup, L.C., Munden, R.F., Dotson, T.L., Bellinger, C.R., Cohen, A., Simoff, M.J., Massion, P.P., Filippini, C. and Gleeson, F.V., 2022. Artificial intelligence tool for assessment of indeterminate pulmonary nodules detected with CT. Radiology, 304(3), pp.683-691.

Examples

```
data(IPNs, package = "R4HCR")
```

LungCa *Years of Smoking and Lung Cancer Deaths in Men.*

Description

Data on man-years of risk and observed number of lung cancer deaths.

Usage

```
LungCa
```

Format

A data frame with 63 observations on the following 4 variables.

`yrs_smk` Years of smoking (`15-19`, `20-24`, `25-29`, `30-34`, `35-39`,`40-44`, `45-49`, `50-54`, `55-59`).

`pys` Person-years of follow-up.

`num_cigs` Number of cigarettes smoked per day (`0`, `1-9`, `10-14`, `15-19`, `20-24`, `25-34`, `35+`).

`deaths` Number of lung cancer deaths.

Source

These data come from Table 24-4, page 702 of Kleinbaum et al (1988).

References

Kleinbaum, D.G., Kupper, L.L., Muller, K.E. and Nizam, A., 1988. Applied regression analysis and other multivariable methods (Vol. 601). Belmont, CA: Duxbury press

Examples

```
data(LungCa, package = "R4HCR")
```

LVD *Left Ventricular Diastolic Diameter (LVD).*

Description

Transoesophageal measurements of left ventricular length (cm).

Usage

```
LVD
```

Format

Four matrices, each representing a block of 36 LVD measurements.

`block1` a 6x6 matrix, representing indices 1 - 36

`block2` a 6x6 matrix, representing indices 37 - 72

`block3` a 6x6 matrix, representing indices 73 - 108

`block4` a 6x6 matrix, representing indices 109 - 144

Details

These data were used to teach confidence intervals to undergraduate 1st year medical students in Oxford. Each student (from classes of between 20-25 students) draws a set of 12 numbers from a much larger list (the 'population') from which the mean is known to us, but not revealed to them. We instruct the students to use dice to select 12 numbers from the list in order to mimic a random sample. Each student then calculates a sample mean and a 95% confidence interval and they are invited to come up to the front and write their confidence intervals up on the board at the front of the class and the concept of confidence intervals demonstrated.

References

With thanks to Dr Thomas Fanshawe, Prof Richard Stevens and Prof Rafael Perera.

Examples

```
data(LVD, package = "R4HCR")


# population is 144 individuals arranged in 4 blocks
# sampling is done with two dice -
# scores indicate which row and column to select
# sample, three from each of the four blocks
# sample size n = 12

# simulate 12 throws of 2 dice
die1 <- sample(x = 1:6, 12, TRUE)
die2 <- sample(x = 1:6, 12, TRUE)
```

```
# drawing the numbers from the blocks
smp <- c(
LVD[[1]][cbind(die1[1:3],die2[1:3])],
LVD[[2]][cbind(die1[4:6],die2[4:6])],
LVD[[3]][cbind(die1[7:9],die2[7:9])],
LVD[[4]][cbind(die1[10:12],die2[10:12])]
)

# the first four numbers of our sample
smp[1:4]
```

`Malformation` *Infant Malformation and Mother's Alcohol Consumption Data.*

Description

Data from a prospective study of maternal drinking and congenital malformation. Alcohol consumption was measured using a questionnaire (3 months after pregnancy). The presence or absence of congenital sex organ malformation was recorded following childbirth.

Usage

```
Malformation
```

Format

A data frame with 5 observations on the following four variables.

`Alcohol_consumption` Alcohol consumption measured as average numebr of drinks per day.

`Absent` Absence of any congential malformation

`Present` Congenital malformation present

`Midpoints` Midpoints of the alcohol consumption categories

Details

This data set appears in An Introduction to Categorical Data Analysis by Agresti (section 2.5.2, page 35). The original source is cited as B.I.Graubard and E.L.Korn, Biometrics 43: 471-476 (1987).

Source

Agresti, A., 2012. Categorical data analysis (Vol. 792). John Wiley & Sons.

Examples

```
data(Malformation, package = "R4HCR")

# Chi-square test.
with(Malformation,
     chisq.test(cbind(Absent,Present),
                simulate.p.value = TRUE))
```

MedSchools *Medical Humanities Teaching and World Ranking.*

Description

Medical humanities courses and average world ranking in 109 in US medical schools. Two rankings were used for medical schools: the Times Higher Education in the 'clinical, pre-clinical, and health' category and the U.S. News and World Report (USNWR) ranking.

Usage

```
MedSchools
```

Format

A data frame with 109 observations on the following 4 variables.

`School` Name of the medical school.

`Ranking` Average world ranking for the medical school.

`Humanities` The number of medical humanities courses offered to students.

`Compulsory` Whether at least one humanities course was offered.

Details

Medical humanities are believed to positively impact medical education and medical practice, yet the extent of medical humanities teaching in medical schools is largely unknown. As part of a larger study, Howick et al explored whether there was a relationship between the number (mandatory or not) of medical humanities topics offered and the average world ranking in 109 accredited medical schools in the US.

References

Howick, J., Zhao, L., McKaig, B., Rosa, A., Campaner, R., Oke, J.L. and Ho, D., 2022. Do medical schools teach medical humanities? Review of curricula in the United States, Canada and the United Kingdom. Journal of Evaluation in Clinical Practice, 28(1), pp.86-92.

Examples

```
data(MedSchools, package = "R4HCR")
```

Milk *Fat Content of Human Milk by Two Methods.*

Description

Fat content of human milk determined by enzymic procedure for the determination of triglycerides and measured by the standard Gerber method (g/100 ml).

Usage

```
Milk
```

Format

A data frame with 45 observations on the following 2 variables.

`Gerber` Fat content measured by the standard gerber method (g/100 ml).

`Trig` Fat content measured by determination of triglycerides (g/100 ml).

Details

Fat content of human milk determined by enzymic procedure for the determination of triglycerides (standard Gerber method) and determined by the measurement of glycerol released by enzymic hydrolysis of triglycerides.

References

Bland, J.M. and Altman, D.G., 1999. Measuring agreement in method comparison studies. Statistical methods in medical research, 8(2), pp.135-160.

Examples

```
data(Milk, package = "R4HCR")

d <- with(Milk, Trig - Gerber)
a <- with(Milk, (Trig + Gerber)/2)

# regression approach for nonuniform differences
M <- lm(d ~ a)

# as per Bland and Altman (1999) page 147.
coef(M)
```

Nodules *Incidental or Screen-Detected Lung Nodules.*

Description

A subset of retrospectively collected data from patients with pulmonary nodule(s) of up to 15mm detected on routinely performed CT chest scans aged 18 years old or older from 3 academic centres in the UK.

Usage

```
Nodules
```

Format

A data frame with 999 observations on the following 8 variables.

`sex` Sex of the patient (`F` = female, `M` = male)

`age` Age of the patient at CT scan (years)

`num.annotated` Number of nodules annotated

`location` Location of the nodule within the lung (`Lingular Segment` `Left Lower Lobe` `Left Upper Lobe` `Right Lower Lobe` `Right Middle Lobe` `Right Upper Lobe`)

`spiculate` Is the nodule spiculated (`No` or `Yes`)

`smoke.status` Smoking status (with levels `current`, `exsmoke`, `never`, `unknown`, `NR` - not recorded)

`diameter` Maximum diameter measured on a 2D axial CT slice (mm)

`malignant` Ground truth of the nodule `0` = benign, `1` = malignant,

Details

Small pulmonary nodules are a common finding on computed tomographic (CT) scans of the chest. Up to 75% of smokers scanned either as part of their clinical care or in lung cancer screening trials have sub-centimeter pulmonary nodules detected. Most nodules detected on CT scans of the chest are not malignant and detection of nodules is expensive and time-consuming with potential associated patient morbidity and mortality. The outcome or ground truth for each nodule was established routinely in clinical care using the accepted published standards of Histology, 1 year for volume stability or 2 year for diameter stability (for benign nodules only), Expert opinion (for subpleural or perifissural lymph nodes only), or Nodule resolution (i.e. infection clears up). Benign nodules are coded as zero, malignant nodules as 1.

References

Oke, J.L., Pickup, L.C., Declerck, J., Callister, M.E., Baldwin, D., Gustafson, J., Peschl, H., Ather, S., Tsakok, M., Exell, A. and Gleeson, F., 2018. Development and validation of clinical prediction models to risk stratify patients presenting with small pulmonary nodules: a research protocol. Diagnostic and prognostic research, 2, pp.1-6.

Examples

```
data(Nodules, package = "R4HCR")
```

NPguided *NP Guided Monitoring of Heart Failure.*

Description

Data from a meta-analysis of natriuretic peptide-guided (NP-guided) treatment for heart failure.

Usage

```
NPguided
```

Format

A data frame with 18 observations on the following 7 variables.

`studyid` Name and year of study.

`year` Year of publication.

`eventsnp` Number of events (all-cause mortality) in NP-guided monitoring group.

`totalnp` Total number of participants in NP-guided monitoring group.

`eventscntrl` Number of events (all-cause mortality) with treatment guided by clinical assessment alone.

`totalcntrl` Total number of participants with treatment guided by clinical assessment alone.

`comparator` Indicator for type of comparator arm in study (`0` = usual care, `1` = clinical assessment).

Details

Natriuretic peptides (NP) are released by the myocardium in response to pressure or fluid overload and are raised in patients with heart failure (HF). NP is a collective term for N-terminal pro-B-type natriuretic peptide (NT-proBNP) and B-type natriuretic peptide (BNP). Studies compared NP-guided treatment to treatment guided by clinical assessment alone. These data are from a study that aimed to determine whether NP-guided treatment of patients with HF reduces all-cause mortality, amongst other outcomes.

References

McLellan J, Bankhead CR, Oke JL, Hobbs FDR, Taylor CJ, Perera R. Natriuretic peptide-guided treatment for heart failure: a systematic review and meta-analysis. BMJ Evid Based Med. 2020 Feb;25(1):33-37. doi: 10.1136/bmjebm2019111208. Epub 2019 Jul 20. PMID: 31326896; PMCID: PMC7029248.

Examples

```
require(meta)

data(NPguided, package = "R4HCR")

metabin(
  sm = "RR",
  method = "MH",
  event.e = eventsnp,
  n.e = totalnp,
  event.c = eventscntrl,
  n.c = totalcntrl,
  studlab = studyid,
  data  = NPguided)
```

OXFIT *OXFIT data set*

Description

Faecal immunochemical testing for adults with symptoms of colorectal cancer attending English primary care.

Usage

```
OXFIT
```

Format

A data frame with 9.999 observations on the following 10 variables.

`sex` Sex of patient, coded 1 = male,2 = female

`fit_val` Faecal immunochemical test (FIT) micro grams per Hb/g faeces.

`albumin` Blood albumin in grams per decilitre (g/dL).

`alkphosphatase` Alkophosphatase (ALK) in units per litre (U/L).

`crp` C-reactive protein (CRP) in mg/dL.

`haemoglobin` Haemoglobin in grams per decilitre (g/dL).

`mean_cell_hgb` Mean cell haemoglobin in picograms per cell (pg).

`mean_cell_vol` Mean cell volume (MCV) in cubic microns (micrometre ^3).

`platelets` Platelets in millilitres per Kilogram (mL/Kg).

`cancer` Whether the patient had colorectal cancer (`0` = No, `1` = Yes)

Details

Faecal samples and other blood tests from routine primary care practice in Oxfordshire, UK between March 2017 and March 2020. FIT was analysed using the HM-JACKarc FIT method. Patients were followed for up to 36 months in linked hospital records for evidence of benign and serious colrectal disease (e.g. colorectal cancer, high-risk adenomas, and bowel inflammation).

Source

This is a synthetic data set generated from the original data set and therefore does not contain actual patient data, only data from simulated patients that share similar attributes to those of the original cohort.

References

Nicholson BD, James T, Paddon M, et al. Faecal immunochemical testing for adults with symptoms of colorectal cancer attending English primary care: a retrospective cohort study of 14 487 consecutive test requests. Aliment Pharmacol Ther. 2020; 52: 1031–1041.

Examples

```
data(OXFIT, package = "R4HCR")
```

PEFR *Peak Expiratory Flow Rate Measurement.*

Description

Repeated measurements of lung function (peak expiratory flow rate (PEFR)) in 20 schoolchildren (taken from a larger study).

Usage

```
PEFR
```

Format

A data frame with 20 observations on the following 7 variables.

`child` Child ID number.

`pefr1` First PEFR measurement (l/min).

`pefr2` Second PEFR measurement (l/min).

`pefr3` Third PEFR measurement (l/min).

`pefr4` Fourth PEFR measurement (l/min).

`mean` Row mean of the four PEFR measurements (l/min).

`sd` Row SD of the four PEFR measurements (l/min).

References

Bland JM, Altman DG. Measurement error. BMJ. 1996 Sep 21;313(7059):744.

Examples

```
data(PEFR, package = "R4HCR")
```

Peptides *Measurements of a Neurotoxic Bioactive Peptide in Brain Samples.*

Description

An amino acid bioactive peptide considered to be neurotoxic in the adult brain and a potential key driver of neurodegeneration is measured in samples from 17 men and 21 women.

Usage

```
Peptide
```

Format

A data frame with 38 observations on the following 2 variables.

`peptide` Peptide concentrations.

`sex` Sex of patient (`M` = male, `F` = female)

Examples

```
data(Peptides, package = "R4HCR")

# Compare levels in men and women.
t.test(peptide  ~ sex, data = Peptides)
```

PlasmaVolume *Measurements of Plasma Volume Using Two Sets of Normal Values.*

Description

Measurements of plasma volume expressed as a percentage of normal in 99 subjects, using two alternative sets of normal values due to Nadler and Hurley.

Usage

```
PlasmaVolume
```

Format

A data frame with 99 observations on the following 3 variables.

`Nadler` Plasma volume expressed as a percentage of normal using Nadler normal values.

`Hurley` Plasma volume expressed as a percentage of normal using Hurley normal values.

Source

Data originally supplied by C Dore, reprinted in Altman and Bland 1999.

References

Bland, J.M. and Altman, D.G., 1999. Measuring agreement in method comparison studies. Statistical methods in medical research, 8(2), pp.135-160.

Examples

```
data(PlasmaVolume, package = "R4HCR")
```

Potency *Potency of four cardiac substances.*

Description

Data from a study of the potencies of four cardiac substances (from Kleinbaum et al)

Usage

```
Potency
```

Format

A data frame with 40 observations on the following 2 variables.

`dosage` Dosage at which the guinea pig died.

`substance` The type of cardiac substance (`sub1-sub4`).

Details

In this experiment, a dilution of one of the substances was infused into an anaesthetized guinea pig, and the dosage at which the pig died was recorded. There were ten replicates in each group (cardiac substance).

Source

This data is featured in Kleinbaum et al (1988).

References

Kleinbaum, D.G., Kupper, L.L., Muller, K.E. and Nizam, A., 1988. Applied regression analysis and other multivariable methods (Vol. 601). Belmont, CA: Duxbury press.

Examples

```
data(Potency, package = "R4HCR")
```

PTX *Detecting Pneumothoraces.*

Description

A synthesised data set from a multicentre blinded fully-crossed multi-case multi-reader (MRMC) study conducted between October 2021 to January 2022.

Usage

```
PTX
```

Format

A data frame with 200 observations on the following 6 variables.

`PTX1` The judgment from one reader on whether a pneumothorax (PTX) is present(1) or absent (0) on an image.

`Conf1` The confidence score (1-4) from one reader on whether a pneumothorax is present.

`PTX2` The judgment from a second reader on whether a pneumothorax is present or absent on an image.

`Conf2` The confidence score (1-4) from a second reader on whether a pneumothorax is present.

`PTX3` The judgment from a third reader on whether a pneumothorax is present or absent on an image.

`Conf3` The confidence score (1-4) from third reader on whether a pneumothorax is present.

Details

The original data consisted of 400 retrospectively collected and de-identified chest X-ray images of patients aged 18 years or older, identified from the CRIS database in Oxford University Hospitals NHS Trust. The study included two reader phases. In the first phase (from which the data is synthesised) readers were asked to interpret the entire dataset over three weeks, recording the perceived presence/absence of a pneumothorax on each image and their degree of confidence on a Likert type scale. A second phase (not included here) repeated the exercise with readers re-interpreting the images with assistance from Artificial Intelligence (AI)

Source

This is a synthetic data set generated from the original data set and therefore does not contain actual patient data, only data from simulated patients that share similar attributes to those of the original cohort.

References

Novak, Alex, Ather, S, Gleeson, F, Espinosa, M, et al. Evaluation of the Impact of Artificial Intelligence-Assisted Image Interpretation on the Diagnostic Performance of Clinicians When Identifying Pneumothoraces on Plain Chest X-Ray: A Multi-Case Multi-Reader Study.

Examples

```
data(PTX, package = "R4HCR")
```

PTXII *Confidence in Detecting Pneumothoraces.*

Description

Subjective confidence rating in the presence of a pneumothorax (PTX) on X-ray.This dataset represents a subset of one reader's confidence scores, in one phase of the study.

Usage

```
PTXII
```

Format

A data frame with 300 observations on the following 2 variables.

`response` Indicator for presence `1` or absence `0` of a pneumothorax on X-ray

`predictor` Subjective connfidence score (1-8) in the absence or presence of a pneumothorax on a X-ray

Details

The original data consisted of 400 retrospectively collected and de-identified chest X-ray images of patients aged 18 years or older, identified from the CRIS database in Oxford University Hospitals NHS Trust. The study included two reader phases. In the first phase (from which the data is synthesised) readers were asked to interpret the entire dataset over three weeks, recording the perceived presence/absence of a pneumothorax on each image and their degree of confidence on a Likert type scale. A second phase (not included here) repeated the exercise with readers re-interpreting the images with assistance from Artificial Intelligence (AI)

Source

The dataset represents a subset of one reader, in one phase of the study.

References

Novak, Alex, Ather, S, Gleeson, F, Espinosa, M, et al. Evaluation of the Impact of Artificial Intelligence-Assisted Image Interpretation on the Diagnostic Performance of Clinicians When Identifying Pneumothoraces on Plain Chest X-Ray: A Multi-Case Multi-Reader Study.

Examples

```
data(PTXII, package = "R4HCR")
```

`Remission` *Effect of 6-mercaptopurine (6-MP) on the Duration of Remission in Acute Leukemia.*

Description

Duration of remission for acute leukemia patients on active treatment or placebo.

Usage

```
Remission
```

Format

A data frame with 42 observations on the following 5 variables.

`sex` Sex of the patient (`0` = male, `1` = female).

`wbc` log white-blood cell count (WBC).

`time` Time to event, where the event is either relapse or loss to follow up.

`event` Indicator of event type, either `Relapse` or `Censored`.

`grp` Treatment group (`6-MP` = allocated to active treament, or `Placebo`).

Details

In this study, patients in remission were randomly assigned to maintenance therapy with 6-MP, an active antileukemic compound 6-MP, or a placebo. White blood cell count was also recorded as this was considered a prognostic indicator of survival for leukemia patients, with the higher values being associated with a worse prognosis.

Source

Kleinbaum, D.G. and Klein, M., 1996. Survival Analysis: A Self-Learning Text. Springer.

References

Acute Leukemia Group B, Freireich, E.J., Gehan, E., Frei III, E.M.I.L., Schroeder, L.R., Wolman, I.J., Anbari, R., Burgert, E.O., Mills, S.D., Pinkel, D. and Selawry, O.S., 1963. The effect of 6-mercaptopurine on the duration of steroid-induced remissions in acute leukemia: A model for evaluation of other potentially useful therapy. Blood, 21(6), pp.699-716.

Examples

```
data(Remission, package = "R4HCR")

# Number of events/censored by group
aggregate(event ~ grp,
data = Remission,
FUN = table)

# median survival times, ignoring the censoring.
aggregate(time ~ grp,
data = Remission,
FUN = median)
```

SCAN *Suspected CANcer (SCAN) Pathway*

Description

Blood test results from people presenting to primary care with non-specific symptoms of cancer.

Usage

```
SCAN
```

Format

A data frame with 750 observations on the following 8 variables.

`age` Age of the patient (in years).

`comorbidity` Charlson comorbidity score.

`haemoglobin` Haemoglobin (g/dL)

`albumin` Blood Albumin (g/dL)

`alaninetrans` Alanine Transaminase (U/L)

`whitebloodcell` White blood cell count (per microlitre x 10^9/L)

`bilirubin` Bilirubin (umol/L)

`calcium` Calcium in milligrams (mg/dL)

Source

This is a synthetic data set generated from the original data set and therefore does not contain actual patient data, only data from simulated patients that share similar attributes to those of the original cohort.

References

Nicholson BD, Oke JL, Friedemann Smith C, et al. The Suspected CANcer (SCAN) pathway: protocol for evaluating a new standard of care for patients with non-specific symptoms of cancer. BMJ Open 2018;8:e018168.

Examples

```
data(SCAN, package = "R4HCR")
```

Scotland *Scottish Death Registration data for 2021.*

Description

The number of deaths registered in Scotland per week for the first 42 weeks of 2021, stratified by cause of death.

Usage

```
Scotland
```

Format

A matrix with five rows and 42 columns.

`rows` Cancer, Dementia, Respiratory, SARS-Cov2 and Other causes of death.

`columns` Regsitration Weeks (Wk1 - Wk42).

Source

Downloaded from https://www.nrscotland.gov.uk/research/guides/birth-death-and-marriage-records in Nov 2021.

Examples

```
data(Scotland, package = "R4HCR")

# A stacked barplot.
barplot(Scotland,
        legend.text = c("Cancer","Dementia/Alzheimers",
                        "Circulatory","Respiratory","Covid-19","Other"),
        beside = FALSE,
        cex.names = 0.8,
        angle = c(45,90,135,180,215),
        density = 45,
        args.legend = c(ncol = 3, cex = 0.65, x = 45))
```

Smartphone *Cervical cancer Screening with Smartphones.*

Description

The objective of this study was to evaluate the diagnostic accuracy of CIN2+ detection using a combined approach (naked-eye and digital VIA (visual inspection with acetic acid) using a Samsung Galaxy J5 smartphone) compared to a traditional naked-eye alone.

Usage

```
Smartphone
```

Format

A data frame with 181 observations on the following 10 variables.

`hpv16` `negative` or `positive` for HPV16.

`hpv1845` HPV18 and/or HPV45 (`present` or `absent`)

`hpvother` Other high-risk HPV types (`present` or `absent`).

`naked_via` Convential visual assessment using naked eye alone (`negative`, `positive`).

`smart_via` Digital VIA result (`negative` or `positive`).

`treatment` Decision to treat (`no` or `yes`).

`combined_via` Combined naked-eye and digital VIA diagnosis (`neither positive` or `either positive`).

`histology` Histological result (`negative`,`CIN1`,`CIN2`, `CIN3`, `cancer`).

`cytology` Cytological result (`negative`, `LSIL`, `HSIL`, `ASC-US`, `AGC`, `ASC-H`, `cancer`, `non-interpretable`).

`CIN2plus` Histological result CIN2 or higher (`<CIN2`, `CIN2+`).

Details

These data are from a screening trial conducted in Dschang (West Cameroon) between February 2019 and March 2020. Women aged 30 to 49 were invited to participate in a free cervical cancer screening campaign. Primary HPV-based screening was followed by a pelvic exam for visual assessment (viewing the cervix with the naked eye to identify colour changes on the cervix) and then cervical biopsy and endocervical curettage. The study aimed to assess whether the use, in addition to normal visual inspection, of images captured using a smartphone could improve the detection of precancerous lesions or cancer.

Source

Data directly available from https://yareta.unige.ch/archives/ffbeb6d7-b390-4755-987e-8faf85f97c67

References

Dufeil, E., Kenfack, B., Tincho, E., Fouogue, J., Wisniak, A., Sormani, J., Vassilakos, P. and Petignat, P., 2022. Addition of digital VIA/VILI to conventional naked-eye examination for triage of HPV-positive women: A study conducted in a low-resource setting. Plos one, 17(5), p.e0268015.

Examples

```
data(Smartphone, package = "R4HCR")
```

`Systolic` *Systolic Blood Pressure Measured by Two Observers and a Machine.*

Description

Systolic blood pressure measurements made simultaneously by two observers (J and R) using a sphygmomanometer and an automatic blood pressure measuring machine (S), each making three observations in quick succession.

Usage

```
Systolic
```

Format

A data frame with 85 observations on the following 9 variables.

`J1` First (of three) measurements made by observer J.

`J2` Second (of three) measurements made by observer J.

`J3` Third (of three) measurements made by observer J.

`R1` First (of three) measurements made by observer R.

`R2` Second (of three) measurements made by observer R.

`R3` Third (of three) measurements made by observer R.

`S1` First (of three) measurements made using a machine.

`S2` Second (of three) measurements made using a machine.

`S3` Third (of three) measurements made using a machine.

Source

Data supplied originally by Dr E O'Brien, and reprinted in Altman and Bland (1999).

References

Bland, J.M. and Altman, D.G., 1999. Measuring agreement in method comparison studies. Statistical Methods in Medical Research, 8(2), pp.135-160.

Examples

```
data(Systolic, package = "R4HCR")
```

`Thrombosis` *Mortality from Coronary Thrombosis.*

Description

Data from the study of Hill and Doll (1966) on the mortality of British doctors in relation to smoking: observations on coronary thrombosis and used in Agresti (1996).

Usage

```
Thrombosis
```

Format

A data frame with 10 observations on the following 4 variables.

`age` Age band of strata (`35-44`, `45-54`, `55-64`, `65-74`).

`smoking` Smoking status (`Nonsmokers` or `Smokers`).

`deaths` Number of deaths from coronary thrombosis per strata.

`pyrs` Sum of person-years in strata.

Source

Agresti, A., 1996. An introduction to categorical data analysis.

References

Doll R, Hill AB. Mortality of British doctors in relation to smoking: observations on coronary thrombosis. Natl Cancer Inst Monogr. 1966 Jan;19:205-68. PMID: 5905669.

Examples

```
data(Thrombosis)

with(Thrombosis,
xtabs(cbind(deaths,pyrs) ~ age + smoking))
```

USCancerStats *Change in Cancer Incidence, Mortality and Survival Statistics.*

Description

US Incidence, mortality, and survival statistics for 20 solid tumor types.

Usage

```
USCancerStats
```

Format

A data frame with 20 observations on the following 4 variables.

`site` The site (or organ) of the cancer.

`survival` Absolute change in site-specific five-year survival.

`mortality` Percentage change in site-specific mortality.

`incidence` Percentage change in sit-specific incidence.

Details

Incidence, mortality, and survival statistics for 20 solid tumor types reported by the SEER program. For each tumor, the absolute difference in 5-year survival between 1989-1995 and 1950-1954 is reported, along with the percentage change in mortality and incidence for 1950 - 1996.

References

Welch, H.G., Schwartz, L.M. and Woloshin, S., 2000. Are increasing 5-year survival rates evidence of success against cancer?. JAMA, 283(22), pp.2975-2978.

Examples

```
data(USCancerStats, package = "R4HCR")

cor.test( ~ survival + mortality,
          data = USCancerStats,
          exact = FALSE,
          method = "sp")
```

`Vaccinated` *Vaccination Uptake Among European Countries.*

Description

Number of people with at least one vaccination against SARS-COV2 as of Nov 2021

Usage

```
Vaccinated
```

Format

A data frame with 15 observations on the following 3 variables.

`country` Name of European country.

`vaccinated` Percentage of people vaccinated against SARS-COV2.

`fully_vaccinated` Percentage of people fully vaccinated against SARS-COV2.

Details

These data are the number of people with at least one vaccination against SARS-COV2 (a.k.a Covid-19) as per the week ending the 12th November 2021, per hundred for countries in Europe with a population greater than 10 million. Fully vaccinated refers to having completed all vaccinations (including boosters) for that country.

Examples

```
data(Vaccinated, package = "R4HCR")

heights <- Vaccinated$vaccinated
names <- Vaccinated$country
bp <- barplot(height = heights,
col = "white",
ylim=c(0,100),
names.arg = names,
cex.names = 0.9,
las = 2,
ylab = "People vaccinated per 100")

# using round here to save space
labels <- round(Vaccinated$vaccinated,0)

text(x = bp, y = labels-2, labels = labels,
cex = 0.9, pos = 3)
```

VSA *Volatile Substance Abuse Mortality in Great Britain, 1971-83.*

Description

Mortaility associated with volatile substance abuse (VSA).This study collated all known death associated with VSA from 1971 to 1983 (inclusively).

Usage

```
VSA
```

Format

A data frame with 9 observations on the following 4 variables.

`age` Age band in nine categories `0-9`,`10-14`,`15-19`,`20-24`,`25-29`,`30-39`,`40-49`,`50-59`,`60+`.

`country` The country in which the deaths were recorded (`Great Britain` or `Scotland`).

`pop` Population size of the age band.

`deaths` The number of deaths associated with VSA per age band.

Details

The data was taken from Bland (2015), who cites Anderson et al (1985) as the source of the data. Note that Scotland is one of the three countries that make up Great Britain, along with England and Wales.

Source

Bland, M., 2015. An introduction to medical statistics. Oxford University Press.

References

Anderson, H.R., Macnair, R.S. and Ramsey, J.D., 1985. Deaths from abuse of volatile substances: a national epidemiological study. Br Med J (Clin Res Ed), 290(6464), pp.304-307.

Examples

```
data(VSA, package = "R4HCR")
```

Index