

Package ‘CASdatasets’

May 5, 2026

Type Package

Title Insurance Datasets

Version 1.2-1

Description A collection of insurance datasets, originally for the book 'Computational Actuarial Science with R' <<https://www.routledge.com/Computational-Actuarial-Science-with-R/Charpentier/p/book/9781138033788>> edited by Arthur Charpentier. Now, the package contains a large variety of insurance datasets. Vignettes provide typical use-cases in actuarial science.

Depends R (>= 3.6.0), xts, survival

Imports lattice

Suggests sp, sf, knitr, kableExtra, quarto, rmarkdown, AER, boot, broom, caret, ChainLadder, demography, dplyr, forecast, ggplot2, glmnet, locfit, MASS, mgcv, pscl, rainbow, ranger, RColorBrewer, splines, tidyr, tidyverse, wesanderson

License GPL (>= 2)

URL <https://dutangc.github.io/CASdatasets/>,
<http://dutangc.free.fr/pub/RRepos/>,
<https://freakonometrics.github.io/CASdatasets/>,
<http://dutangc.perso.math.cnrs.fr/RRepository/>,
<https://cas.uqam.ca/>

BugReports <https://github.com/dutangc/CASdatasets/issues>

NeedsCompilation no

VignetteBuilder quarto

BuildVignettes true

BuildResaveData no

LazyData no

Encoding UTF-8

Classification/MSC-2010 62P05, 91B30, 97M30

Author Christophe Dutang [aut, cre] (ORCID:
 <<https://orcid.org/0000-0001-6732-1501>>),
 Arthur Charpentier [aut] (ORCID:
 <<https://orcid.org/0000-0003-3654-6286>>),
 Ewen Gallic [ctb] (ORCID: <<https://orcid.org/0000-0003-3740-2620>>),
 Julien Siharath [ctb]

Maintainer Christophe Dutang <dutangc@gmail.com>

Contents

asiacomrisk	4
ausautoBI8999	5
auscathist	6
ausNLHYby	7
ausNLHYglossary	10
ausNLHYlloyd	12
ausNLHYtotal	14
ausNSW	16
ausprivauto	17
austriLoB	19
beaonre	21
beMTPL16	21
beMTPL97	23
beseкура	24
bragg	25
brautocoll	26
brgeomunicins	27
brvehins	29
canlifins	31
CASdatasets	32
catelematic13	35
credit	37
danish	39
Davis	40
ECBYieldCurve	41
eqlist	41
eudirectlapse	43
euhealthinsurance	45
euMTPL	46
eusavingsurrender	47
eutelematic13	48
FedYieldCurve	49
forexUSUK	50
fre4LoBtriangles	51
freagnumber	52
frebiloss	53
freclaimset	54
freclaimset9207	55
frecomfire	57
freDemography	58
freDisTables	59

freIRSA	62
fremarine	63
freMortTables	64
fremotorclaim	66
freMPL	70
freMTPL	73
freportfolio	76
fretelematic	78
fretplclaimnumber	79
hurricanehist	80
ICB	81
itamtplcost	84
linearmodelfactor	85
lossalae	86
norauto	87
Norberg	88
norfire	88
nortrip18800	89
nzcathist	90
PnCdemand	91
pricingame	93
sgautonb	97
sgtriangles	99
SOAGMI	100
spacedata	101
swautoins	103
swbusscase	104
swmotorcycle	105
swtriangles	106
ukaggclaim	107
ukautocoll	108
usautoBI	109
usautotriangles	110
usexpense	111
usGLtriangles	112
ushurricane	113
ushustormloss4980	116
uslapseagent	116
usmassBI	118
usmedclaim	119
usMSHA1316	120
usMVTA	121
usprivautoclaim	124
usquakeLR	125
ustermlife	126
uswarrantagnum	127
usworkcomp	128

Description

A completed project by the Insurance Risk and Finance Research Centre (www.IRFRC.com) has assembled a unique dataset from Large Commercial Risk losses in Asia-Pacific (APAC) covering the period 2000-2013. The data was generously contributed by one global reinsurance company and two large Lloyd's syndicates in London. This dataset is the result of the project co-lead by Dr Milidonis (IRFRC and University of Cyprus) and Enrico Biffis (Imperial College Business School), which can be referred to as the IRFRC LCR Dataset.

As expected, the dataset is fully anonymised, as the LCR losses are aggregated along a few dimensions. First, data is categorised based on the World Bank's economic development classification. This means that losses either come from developed or developing countries. The second dimension used to aggregate the data is the time period covered. Data is grouped into (at least) two time-periods: the period before and after the 2008 crisis.

A large commercial risk (LCR) is defined as a loss caused by man-made risks (e.g. fire, explosion, etc.). We exclude natural catastrophe events, and started by focusing on claims that made the data provider incur a loss amount of at least EUR 1 million. We then extended our dataset to include claims leading to loss amounts smaller than EUR 1 million. Given time constraints, we only partially extended loss data by obtaining FGU losses larger than EUR 140k. One should note that any selection bias arising from the data collection exercise is driven by both data quality and reliability. Based on our experience, the latter two attributes are homogeneous across developed and developing countries APAC claims.

For further details, see the technical report: Benedetti, Biffis and Milidonis (2015a).

Usage

```
data(asiacomrisk)
```

Format

asiacomrisk contains 7 columns:

Period A character string for the period: "2000-2003", "2004-2008", "2009-2010", "2011-2013".

FGU From the Ground Up Loss (USD).

TIV Total Insurable Value (TIV) replaced with Total Sum Insured (TSI) when the TIV is not available (USD).

CountryStatus A character string for the country status: "Developped", "Emerging".

Usage A character string for the type of exposure hit by the loss: "Commercial", "Energy", "Manufacturing", "Misc.", "Residential".

SubUsage A character string for a precised type of exposure hit by the loss: "Commercial", "Energy", "General industry", "Metals/Mines/Chemicals", "Misc.", "Residential", "Utility".

DR A numeric for the destruction rate (FGU divided TIV capped to 1).

References

- Benedetti, D., Biffis, E., and Milidonis, A. (2015a). *Large Commercial Risks (LCR) in Insurance: Focus on Asia-Pacific*, Insurance Risk and Finance Research Centre Technical report.
- Benedetti, D., Biffis, E., and Milidonis, A. (2015b). *Large Commercial Exposures and Tail Risk: Evidence from the Asia-Pacific Property and Casualty Insurance Market*, Working paper.
- Chavez-Demoulin, V., Embrechts, P., and Hofert, M. (2015). *An extreme value approach for modeling operational risk losses depending on covariates*. The Journal of Risk and Insurance.

Examples

```
# (1) load of data
#
data(asiacomrisk)
dim(asiacomrisk)

# (2) basic boxplots
#

asiacomrisk
boxplot(DR ~ Usage, data=asiacomrisk)
boxplot(DR ~ SubUsage, data=asiacomrisk)
boxplot(DR ~ Period, data=asiacomrisk)
boxplot(DR ~ CountryStatus, data=asiacomrisk)
```

ausautoBI8999

Automobile bodily injury claim dataset in Australia

Description

This data set contains information on 22036 settled personal injury insurance claims in Australia. These claims arose from accidents occurring from July 1989 through to January 1999. Claims settled with zero payment are not included.

Usage

```
data(ausautoBI8999)
```

Format

ausautoBI8999 is a data frame of 8 columns and 1,340 rows:

AccDate, ReportDate, FinDate The accident date, the reporting date, the finalization date, note that the day is always set to the first day of the month.

AccMth, ReportMth, FinMth The accident month, the reporting month, the finalization month: 1 = July 1989, ..., 120 = June 1999).

OpTime The operational time.

InjType1, InjType2, InjType3, InjType4, InjType5 The injury code for the people injured (up to five).

InjNb Number of injured people.

Legal A character string for: Has the policyholder a legal representation?
 AggClaim Aggregate settled amount of claims.

Source

Formerly on a website dedicated to P. De Jong and G.Z. Heller (2008).

References

P. De Jong and G.Z. Heller (2008), *Generalized linear models for insurance data*, Cambridge University Press, doi:[10.1017/CBO9780511755408](https://doi.org/10.1017/CBO9780511755408).

Examples

```
# (1) load of data
#
data(ausautoBI8999)
dim(ausautoBI8999)
head(ausautoBI8999)
```

auscathist

Australian catastrophe historic

Description

Historical disaster statistics in Australia from 1967 to 2014.

Usage

```
data(auscathist)
```

Format

auscathist is a data frame of 9 columns:

Year a numeric for the Year.

Quarter a numeric for the quarter of the year.

Date a character string for the date.

FirstDay a Date object for the first day of natural catastrophe.

LastDay a Date object for the last day of natural catastrophe, when available.

Event a character string describing the event.

Type a factor describing the event type among the list: "Cyclone", "Earthquake", "Flood", "Flood, Storm", "Hailstorm", "Other", "Power outage", "Storm", "Tornado", "Weather", "Bushfire".

Location a character string describing the location.

OriginalCost Original cost in million of Australian dollars (AUD).

NormCost2011 Normed cost in million of 2011 Australian dollars (AUD) taking into account inflation, change in wealth and population.

NormCost2014 Normed cost in million of 2014 Australian dollars (AUD) computed as the inflated cost NormCost2011 using CPI.

Source

<https://insurancecouncil.com.au/>

Examples

```
# (1) load of data
#
data(auscathist)

# (2) plot of data
#
plot(ecdf(auscathist$NormCost2014))
```

ausNLHYby

Australian Market - non-life insurance (company, state, public level)

Description

Financial performance and financial position of insurers operating in Australia between 2005 and 2010 (company, state, public level).

Usage

```
data(ausNLHYClaimByState)
data(ausNLHYPremByState)

data(ausNLHYCapAdeqByComp)
data(ausNLHYFinPerfByComp)
data(ausNLHYFinPosByComp)
data(ausNLHYPrivInsur)

data(ausNLHYFinPerfPublic)
data(ausNLHYFinPosPublic)
data(ausNLHYOpIncExpPublic)
data(ausNLHYPremClaimPublic)
data(ausNLHYPubInsur)
```

Format

ausNLHYPremByState (Table 10) and ausNLHYClaimByState (Table 11) are data frames of 6 columns (values are in million of Australian dollars (AUD)):

- Class: Class of business.
- NSWACTYYYYMM: New South Wales / Australian Capital Territory for year YYYY.
- VICYYYYMM: Victoria in year YYYY reported on DateYYYYMM.
- QLDYYYYMM: Queensland in year YYYY reported on DateYYYYMM.
- SAYYYYYMM: South Australia in year YYYY reported on DateYYYYMM.
- WAYYYYYMM: Western Australia in year YYYY reported on DateYYYYMM.

- TAYYYYMM: Tasmania in year YYYY reported on DateYYYYMM.
- NTYYYYMM: Northern Territory in year YYYY reported on DateYYYYMM.
- TotalYYYYMM: Total in year YYYY reported on DateYYYYMM.

where YYYYMM is the concatenation of the year YYYY and month MM, e.g. 200506.

ausNLHYPrivInsur (Classification private) is a data frame of 6 columns (values are in thousand of Australian dollars (AUD)):

- Company: Company short name.
- FullNameYYYYMM: Full name of the company for year YYYY.
- DateYYYYMM: Date in year YYYY reported on DateYYYYMM.
- ClassificationYYYYMM: Classification in year YYYY reported on DateYYYYMM either Direct or Reinsurer.
- BranchYYYYMM: non empty when branch insurer in year YYYY reported on DateYYYYMM.
- RestrictionYYYYMM: Restriction on underwriting in year YYYY reported on DateYYYYMM.

where YYYYMM is the concatenation of the year YYYY and month MM, e.g. 200506.

ausNLHYCapAdeqByComp (Table 14) is a data frame of 6 columns (values are in thousand of Australian dollars (AUD)):

- Company: Company short name.
- DateYYYYMM: Balance Date for year YYYY.
- MCRYYYYMM: Minimum capital requirement in year YYYY reported on DateYYYYMM.
- CapitalYYYYMM: Capital base in year YYYY reported on DateYYYYMM.
- SurplusYYYYMM: Capital surplus in year YYYY reported on DateYYYYMM.
- SolRatioYYYYMM: Solvency coverage ratio in year YYYY reported on DateYYYYMM.

where YYYYMM is the concatenation of the year YYYY and month MM, e.g. 200506.

ausNLHYFinPerfByComp (Table 12) is a data frame of 9 columns (values are in thousand of Australian dollars (AUD)):

- Company: Company short name.
- DateYYYYMM: Balance Date for year YYYY.
- GWPYYYYMM: Gross written premium revenue in year YYYY reported on DateYYYYMM.
- REYYYYMM: Outwards reinsurance expense in year YYYY reported on DateYYYYMM.
- NWPYYYYMM: Net written premium revenue in year YYYY reported on DateYYYYMM.
- GICYYYYMM: Gross incurred claims in year YYYY reported on DateYYYYMM.
- NRRYYYYMM: Non-reinsurance recoveries revenue in year YYYY reported on DateYYYYMM.
- RRYYYYYMM: Reinsurance recoveries revenue in year YYYY reported on DateYYYYMM.
- NICYYYYMM: Net incurred claims in year YYYY reported on DateYYYYMM.
- UWEYYYYMM: Underwriting expenses in year YYYY reported on DateYYYYMM.
- UWRYYYYMM: Underwriting result in year YYYY reported on DateYYYYMM.
- IYYYYMM: Investment income in year YYYY reported on DateYYYYMM.
- OIYYYYMM: Other items in year YYYY reported on DateYYYYMM.
- NPATYYYYMM: Net profit-loss after tax in year YYYY reported on DateYYYYMM.

where YYYYMM is the concatenation of the year YYYY and month MM, e.g. 200506.

ausNLHYFinPosByComp (Table 13) is a data frame of 7 columns (values are in thousand of Australian dollars (AUD)):

- Company: Company short name.
- InvestYYYYMM: Investments for year YYYY.
- TotalAssetYYYYMM: Total assets in year YYYY reported on DateYYYYMM.
- ClaimReservYYYYMM: Outstanding claims provision in year YYYY reported on DateYYYYMM.
- PremLiabYYYYMM: Premium liabilities in year YYYY reported on DateYYYYMM.
- ClaimReservYYYYMM: Total liabilities in year YYYY reported on DateYYYYMM.
- TotalLiabYYYYMM: Shareholders equity in year YYYY reported on DateYYYYMM.
- EquityYYYYMM: Shareholders equity in year YYYY reported on DateYYYYMM.

where YYYYMM is the concatenation of the year YYYY and month MM, e.g. 200506.

ausNLHYPubInsur (Classification public) is a data frame of 1 column:

- CompanyYYYYMM: Company name for year YYYY.

ausNLHYFinPerfPublic (Table 15), ausNLHYOpIncExpPublic (Table 16), are data frames of 2 columns (values are in million of Australian dollars (AUD)):

- Content: Content.
- TotalYYYYMM: Total for year YYYY.

ausNLHYFinPosPublic (Table 17) is a data frame of 3 columns (values are in million of Australian dollars (AUD)):

- Content: Content.
- TotalYYYYMM: Total for year YYYY.
- InsideAustraliaOnlyYYYYMM: Inside Australia Only for year YYYY.

ausNLHYPremClaimPublic (Table 18) is a data frame of 6 columns (values are in million of Australian dollars (AUD)):

- Class: Class of business.
- GWPYYYYMM: Gross written premium revenue in year YYYY reported on DateYYYYMM.
- PEYYYYMM: Premium revenue in year YYYY reported on DateYYYYMM.
- REYYYYMM: Reinsurance expense in year YYYY reported on DateYYYYMM.
- GICYYYYMM: Gross incurred claims in year YYYY reported on DateYYYYMM.
- RORYYYYYMM: Reinsurance recoveries revenue in year YYYY reported on DateYYYYMM.

where YYYYMM is the concatenation of the year YYYY and month MM, e.g. 200506.

Source

Data is copyrighted by Australian Prudential Regulation Authority (APRA) and is under the Creative Commons - By licence. Please refer to <https://www.apra.gov.au/>

See Also

[ausNLHYtotal](#) for aggregate level, [ausNLHYlloyd](#) for LLoyds and [ausNLHYglossary](#) for glossary notes.

Examples

```

# (1) by company data
#
data(ausNLHYCapAdeqByComp)
data(ausNLHYFinPerfByComp)
data(ausNLHYFinPosByComp)

# (2) by state data
#
data(ausNLHYClaimByState)
data(ausNLHYPremByState)

# (3) public sector data
#
data(ausNLHYFinPerfPublic)
data(ausNLHYFinPosPublic)
data(ausNLHYOpIncExpPublic)
data(ausNLHYPremClaimPublic)

```

ausNLHYglossary

Australian Market - non-life insurance (Glossary)

Description

Financial performance and financial position of insurers operating in Australia between 2005 and 2010 (Glossary).

Details

Glossary notes:

- Capital base is the amount of eligible capital held by an insurer to provide a buffer against losses that have not been anticipated and, in the event of problems, enable the insurer to continue operating while those problems are addressed or resolved. For locally incorporated insurers it is the sum of tier 1 capital (net of deductions) and tier 2 capital . Capital base for branch insurers is derived from net assets inside Australia.
- Captive insurer is a company within a group of related companies performing the function of insurer to that group.
- Classes of business in tables 7-11 are shown in order of risk capital factors as described in guidance note GGN 110.3.
- Direct insurers are those insurers who, excluding intra-group arrangements, predominantly undertake liability by way of direct insurance business.
- Earned premium (as defined in AASB 1023) is the amount of premium earned during the financial year and includes movements in the unearned premium provision.
- Gross claims expense (as per table 11) relates to: claims that are paid during a financial period; and recognised claims liabilities (i.e. movement in outstanding claims provision).
- Gross incurred claims comprises claims paid during the period, movements in the outstanding claims provision and movements in premium liabilities .

- Gross premium revenue is recognised fully when the business is written. The accounting concepts of earned and unearned premium are no longer recognised under the APRA prudential framework, hence this item is not consistent with AASB 1023 requirements. Instead, the potential claims liabilities arising from the uncovered term of written insurance business are recognised through the creation of premium liabilities .
- LMI (Lenders mortgage insurers) provide cover to protect lenders from default by borrowers on loans secured by mortgage. Mortgage insurers are substantially different to other insurers and are subject to special condition of authority.
- Lower tier 2 ratio is lower tier 2 capital divided by tier 1 capital (net of deductions) . The regulatory maximum for this ratio is 50 percent.
- Lloyd's is a London based insurance market in which business is underwritten by both individuals and corporate members who form syndicates to accept risk.
- Minimum capital requirement is the amount of risk-based capital APRA requires general insurers to hold to meet its insurance obligations under a wide range of circumstances.
- Net incurred claims is gross incurred claims net of reinsurance recoveries revenue and non-reinsurance recoveries revenue.
- Net loss ratio is net incurred claims divided by net premium revenue. Net premium revenue is gross premium revenue net of outwards reinsurance expense.
- Net profit/loss refers to profit or loss from ordinary activities after income tax, before extraordinary items.
- Non-reinsurance recoverables comprise recoverables from subrogation, salvage, sharing arrangements etc, net of provision for doubtful debts.
- Non-reinsurance recoveries revenue comprises amounts the insurer has recovered or is entitled to recover from subrogation, salvage and other non-reinsurance recoveries.
- Other assets comprises investment income receivable, other reinsurance assets receivable from reinsurers (i.e. other than reinsurance recoveries), GST receivable, other receivables, tax assets, plant and equipment (net of depreciation) and other assets.
- Other investments are strategic investments/acquisitions and other investments that do not constitute investments integral to insurance operations.
- Other items comprises other operating income, goodwill amortisation and income tax expense or benefit. Other liabilities comprises creditors and accruals, other provisions and other liabilities. Other operating expenses are all operating expenses not related to underwriting.
- Outstanding claims provision is the insurer's liability for outstanding claims. It recognises the potential cost to the insurer of settling claims which it has incurred at the reporting date (including estimates of claims that have not yet been notified to the insurer), but which have not been paid. The amount reported is after taking account of inflation and discounting, without deducting reinsurance and non- reinsurance recoverables .
- Outwards reinsurance expense is premium ceded to reinsurers, recognised as an expense fully when incurred or contracted.
- Payables on reinsurance contracts comprise amounts payable to reinsurers. This includes premiums payable but not yet due for payment, deposits withheld from reinsurers, commissions due to reinsurers and the reinsurers' portion of recoveries and salvage.
- Premium liabilities relate to the future claims arising from future events insured under existing policies accepted. This fully prospective determination is a more effective means of recognising potential risk than the accounting concept of unearned premium. The amount reported is after taking account of inflation and discounting, without deducting reinsurance and non-reinsurance recoveries.

- Premium receivables are premiums due, net of provision for doubtful debts, including unclosed business written close to the reporting date.
- Reinsurance recoverables comprise amounts recoverable under reinsurance contracts. Reinsurance and other recoverables is the aggregate of reinsurance recoverables and non-reinsurance recoverables.
- Reinsurance recoveries revenue comprises amounts the insurer has recovered or is entitled to recover from reinsurers on incurred claims during the reporting period.
- Reinsurers are those insurers who, excluding intra-group arrangements, predominantly undertake liability by way of reinsurance business.
- Return on assets is net profit/loss divided by the average on-balance sheet total assets for the period. Return on equity is net profit/loss divided by the average shareholders' equity for the period.
- Run-off insurers are restricted by APRA from writing new or renewal insurance business. However, the company may still be acting as an insurance agent, broker or underwriting agent for other general insurers.
- Solvency coverage is capital base divided by minimum capital requirement.
- Tier 1 capital (net of deductions) comprises the highest quality capital elements, including: paid-up ordinary shares, general reserves, retained earnings, current year earnings net of expected dividends and tax expenses, technical provisions in excess of those required by GPS 210, non-cumulative irredeemable preference shares and other "innovative" capital instruments. This amount is net of goodwill, other intangible assets and future income tax benefits.

Source

Data is copyrighted by Australian Prudential Regulation Authority (APRA) and is under the Creative Commons - By licence. Please refer to <https://www.apra.gov.au/>

See Also

[ausNLHYby](#) for company, state, public level, [ausNLHYlloyd](#) for LLoyds and [ausNLHYtotal](#) for aggregate level.

ausNLHYlloyd

Australian Market - non-life insurance (LLoyds insurance business)

Description

Financial performance and financial position of insurers operating in Australia between 2005 and 2010 (LLoyds insurance business).

Usage

```
data(ausNLHYlloydAsset)
data(ausNLHYlloydGPI)
data(ausNLHYlloydUWAcc)
data(ausNLHYlloydUWRes)
```

Format

ausNLHYLloydUWAcc (Table 15) and ausNLHYLloydUWAcc (Table 16) are data frames of 4 columns (values are in thousand of Australian dollars (AUD)):

- Content: Content.
- AccYear2YrAgoYYYYMM: value in the 2-year-ago accounting year in year YYYY reported in December.
- AccYear1YrAgoYYYYMM: value in the 1-year-ago accounting year in year YYYY reported in December.
- AccYear0YrAgoYYYYMM: value in the current accounting year in year YYYY reported in December.

where YYYYMM is the concatenation of the year YYYY and month MM=12, e.g. 200512.

ausNLHYLloydGPI (Table 17) is a data frame of 4 columns (values are in thousand of Australian dollars (AUD)):

- Content: Content.
- DirectYYYYMM: Direct premiums (gross) including inward facultative reinsurance in year YYYY reported in December.
- InwardYYYYMM: Inward treaty reinsurance premiums (gross) in year YYYY reported in December.
- TotalYYYYMM: Total premium income (gross) in year YYYY reported in December.

where YYYYMM is the concatenation of the year YYYY and month MM=12, e.g. 200512.

ausNLHYLloydAsset (Table 18) is a data frame of 4 columns (values are in thousand of Australian dollars (AUD)):

- Content: Content.
- TrustFundYYYYMM: Lloyds Australia trust fund in year YYYY reported in December.
- AssetFund1.YYYYYMM: Lloyds Australia joint asset fund No.1 in year YYYY reported in December.
- AssetFund2.YYYYYMM: Lloyds Australia joint asset fund No.2 in year YYYY reported in December.

where YYYYMM is the concatenation of the year YYYY and month MM=12, e.g. 200512.

Details

It is not possible to compare Lloyd's with authorised companies. Lloyd's operates a unique three year accounting system that differs substantially from normal practices. Different classes of business are also used.

The individual syndicates, which are members of the Lloyd's market, are independent entities which are supervised by the Financial Services Authority (FSA) in the UK not by APRA. However, for the protection of policy holders in Australia, Lloyd's is required to maintain trust funds in Australia (refer to Lloyd's Assets Table 18).

Source

Data is copyrighted by Australian Prudential Regulation Authority (APRA) and is under the Creative Commons - By licence. Please refer to <https://www.apra.gov.au/>

See Also

[ausNLHYby](#) for company, state, public level, [ausNLHYtotal](#) for aggregate level and [ausNLHYglossary](#) for glossary notes.

Examples

```
# (1) lloyds data
#
data(ausNLHYLloydAsset)
data(ausNLHYLloydGPI)
data(ausNLHYLloydUWAcc)
data(ausNLHYLloydUWRes)
```

ausNLHYtotal	<i>Australian Market - non-life insurance (aggregate level)</i>
--------------	---

Description

Financial performance and financial position of insurers operating in Australia between 2005 and 2010 (aggregate level).

Usage

```
data(ausNLHYCapAdeq)
data(ausNLHYFinPerf)
data(ausNLHYFinPos)
data(ausNLHYLiability)
data(ausNLHYOffProf)
data(ausNLHYOpIncExp)
data(ausNLHYPremClaim)
data(ausNLHYPrivInsur)
data(ausNLHYPubInsur)
data(ausNLHYRecAASB)
data(ausNLHYReserve)
```

Format

All values are in million of Australian dollars (AUD).

ausNLHYFinPerf (Table 1), ausNLHYCapAdeq (Table 5), ausNLHYOpIncExp (Table 2) are data frames of 4 columns:

- Content: Content.
- InsurersYYYYMM: Insurers for year YYYY.
- ReinsurersYYYYMM: Reinsurers in year YYYY reported on DateYYYYMM.
- TotalYYYYMM: Total in year YYYY reported on DateYYYYMM.

where YYYYMM is the concatenation of the year YYYY and month MM, e.g. 200506.

ausNLHYRecAASB (Table 6) is data frames of 4 columns:

- Content: Content.
- NBInsurersYYYYMM: Non-branch Insurers for year YYYY.
- NBReinsurersYYYYMM: Non-branch Reinsurers in year YYYY reported on DateYYYYMM.
- NBTotalsYYYYMM: Non-branch Total in year YYYY reported on DateYYYYMM.

where YYYYMM is the concatenation of the year YYYY and month MM, e.g. 200506.

ausNLHYFinPos (Table 3) is a data frame of 5 columns:

- Content: Content.
- InsurersYYYYMM: Insurers for year YYYY.
- ReinsurersYYYYMM: Reinsurers in year YYYY reported on DateYYYYMM.
- TotalYYYYMM: Total in year YYYY reported on DateYYYYMM.
- InsideAustraliaOnlyYYYYMM: InsideAustraliaOnly in year YYYY reported on DateYYYYMM.

where YYYYMM is the concatenation of the year YYYY and month MM, e.g. 200506.

ausNLHYPremClaim (Table 7) is a data frame of 6 columns:

- Class: Class of business.
- GWPYYYYMM: Gross written premium revenue in year YYYY reported on DateYYYYMM.
- REYYYYMM: Outwards reinsurance expense in year YYYY reported on DateYYYYMM.
- NWPYYYYMM: Net written premium revenue in year YYYY reported on DateYYYYMM.
- GICYYYYMM: Gross incurred claims in year YYYY reported on DateYYYYMM.
- RRYYYYYMM: Reinsurance recoveries revenue in year YYYY reported on DateYYYYMM.

where YYYYMM is the concatenation of the year YYYY and month MM, e.g. 200506.

ausNLHYReserve (Table 8) is a data frame of 5 columns:

- Class: Class of business.
- GORYYYYMM: Gross Outstanding Reserve in year YYYY reported on DateYYYYMM.
- RRYYYYYMM: Reinsurance Recoverables in year YYYY reported on DateYYYYMM.
- NRRYYYYMM: Non Reinsurance Recoverables in year YYYY reported on DateYYYYMM.
- NORYYYYMM: Net Outstanding Reserve in year YYYY reported on DateYYYYMM.

where YYYYMM is the concatenation of the year YYYY and month MM, e.g. 200506.

ausNLHYLiability (Table 9) is a data frame of 5 columns:

- Content: Content.
- GPLYYYYMM: Gross Premium Liability in year YYYY reported on DateYYYYMM.
- RRYYYYYMM: Reinsurance Recoverables in year YYYY reported on DateYYYYMM.
- NRRYYYYMM: Non Reinsurance Recoverables in year YYYY reported on DateYYYYMM.
- NPLYYYYYMM: Net Premium Liability in year YYYY reported on DateYYYYMM.

where YYYYMM is the concatenation of the year YYYY and month MM, e.g. 200506.

ausNLHYOffProf (Table 4) is a data frame of 7 columns:

- Content: Content.
- AusInsurersYYYYMM: Australian Insurers for year YYYY.
- AusReinsurersYYYYMM: Australian Reinsurers in year YYYY reported on DateYYYYMM.
- AusTotalYYYYMM: Australian Total level in year YYYY reported on DateYYYYMM.
- OffInsurersYYYYMM: Offshore Insurers for year YYYY.
- OffReinsurersYYYYMM: Offshore Reinsurers in year YYYY reported on DateYYYYMM.
- OffTotalYYYYMM: Offshore Total level in year YYYY reported on DateYYYYMM.

where YYYYMM is the concatenation of the year YYYY and month MM, e.g. 200506.

Source

Data is copyrighted by Australian Prudential Regulation Authority (APRA) and is under the Creative Commons - By licence. Please refer to <https://www.apra.gov.au/>

See Also

[ausNLHYby](#) for company, state, public level, [ausNLHYlloyd](#) for LLoyds and [ausNLHYglossary](#) for glossary notes.

Examples

```
# (1) private sector data
#
data(ausNLHYCapAdeq)
data(ausNLHYFinPerf)
data(ausNLHYFinPos)
data(ausNLHYLiability)
data(ausNLHYOffProf)
data(ausNLHYOpIncExp)
data(ausNLHYPremClaim)
data(ausNLHYPrivInsur)
data(ausNLHYPubInsur)
data(ausNLHYRecAASB)
data(ausNLHYReserve)
```

ausNSW

Australian Statistics - New South Wales in 2004

Description

General statistics of Australian drivers in New South Wales in 2004.

Usage

```
data(ausNSWdriver04)
data(ausNSWdeath02)
```

Format

ausNSWdriver04 is 2-element list containing the following dataframes.

ausNSWdriver04\$injury ausNSWdriver04\$injury consists of all drivers involved in a crash in 2004 in New South Wales, Australia. There are a total of 82659 drivers in the data set. Drivers with unknown age, age less than 17 years, or road user class "Other" are omitted, leaving 76341 cases. It contains the driver age, the gender, the vehicle class, the crash degree, and the observed number of crashes.

ausNSWdriver04\$alcohol ausNSWdriver04\$alcohol consists of drivers involved in a crash in 2004 in New South Wales, Australia, in which the involvement of blood alcohol concentration (BAC) was known. Drivers with unknown age, age less than 17 years, or unknown BAC are omitted, leaving 58890 cases. It contains the driver age, the gender, the blood alcohol concentration (BAC), the crash degree, and the observed number of crashes.

ausNSWdeath02 is 2-element list containing the following dataframes.

ausNSWdeath02\$allcause ausNSWdeath02\$allcause contains all-cause mortality data for New South Wales, Australia in 2002, by age band and gender.

ausNSWdeath02\$diabete ausNSWdeath02\$diabete contains the number of deaths due to diabetes in New South Wales, Australia in 2002, provided by the Australian Institute of Health and Welfare, from their mortality database.

Source

Formerly on a website dedicated to P. De Jong and G.Z. Heller (2008).

References

P. De Jong and G.Z. Heller (2008), *Generalized linear models for insurance data*, Cambridge University Press, doi:[10.1017/CBO9780511755408](https://doi.org/10.1017/CBO9780511755408).

Examples

```
# (1) data
#
data(ausNSWdriver04)
data(ausNSWdeath02)
```

ausprivauto

Automobile claim datasets in Australia

Description

Third party insurance is a compulsory insurance for vehicle owners in Australia. It insures vehicle owners against injury caused to other drivers, passengers or pedestrians, as a result of an accident.

The ausprivauto0405 dataset is based on one-year vehicle insurance policies taken out in 2004 or 2005. There are 67856 policies, of which 4624 had at least one claim.

The ausMTP8486 dataset records the number of third party claims in a twelve-month period between 1984 and 1986 in each of 176 geographical areas (local government areas) in New South Wales, Australia.

The `ausprivautolong` is a simulated dataset containing counts of claims for 40 000 policies, for three periods (years). The simulation is based on a true non-life portfolio. The risk factors are driver's age and vehicle value. Each policy is regarded as a cluster, and hence there are $3 \times 40\,000 = 120\,000$ records.

Usage

```
data(ausprivautolong)
data(ausMTPPL8486)
data(ausprivauto0405)
```

Format

`ausprivauto0405` is a data frame of 9 columns and 67,856 rows:

`Exposure` The number of policy years.
`VehValue` The vehicle value in thousand of AUD.
`VehAge` The vehicle age group.
`VehBody` The vehicle body group.
`Gender` The gender of the policyholder.
`DrivAge` The age of the policyholder.
`ClaimOcc` Indicates occurrence of a claim.
`ClaimNb` The number of claims.
`ClaimAmount` The sum of claim payments.

`ausMTPPL8486` is a data frame of 7 columns and 176 rows:

`LocalGov` The local government area.
`StatDiv` The vehicle value in thousand of AUD.
`ClaimNb` The number of third-party claims.
`AccNb` The number of accidents.
`KillInjNb` The number of killed or injured.
`Pop` The population size.
`PopDens` The population density.

`ausprivauto0405` is a data frame of 6 columns and 120,000 rows:

`IDpol` The policy identification number.
`DrivAge` The age of the policyholder.
`VehValue` The vehicle value in thousand of AUD.
`Periode` The period number.
`ClaimNb` The number of claims.
`ClaimOcc` Indicates occurrence of a claim.

Source

Formerly on a website dedicated to P. De Jong and G.Z. Heller (2008).

References

P. De Jong and G.Z. Heller (2008), *Generalized linear models for insurance data*, Cambridge University Press, [doi:10.1017/CBO9780511755408](https://doi.org/10.1017/CBO9780511755408).

Examples

```
# (1) load of data
#
data(ausprivautolong)
data(ausMTPL8486)
data(ausprivauto0405)
```

austriLoB

Australian private motor triangles

Description

Dataset `austri1autoBI7895` contains claim triangles from an Australian non-life insurer between 1978 and 1995 for bodily injuries. `austri1autoBI7895` is a list of 5 elements : a triangle of paid amounts, a triangle of incurred amounts, a triangle of notified claim number, a vector of exposure (in number of vehicle) and a vector of claim inflation indices. This corresponds respectively to Tables 3.3 (incr) and 3.2 (cumul); Table 3.12 (cumul); Tables 2.2 (incr) and 2.6 (cumul); Table B.1; Table B.2 of Taylor (2000). Note that claim amounts of `austri1autoBI7895` are incremental.

Dataset `austri2auto` contains claim triangles from an Australian non-life insurer in run-off. Note that claim amounts are incremental.

Usage

```
#1st Line of Business
data(austri1autoBI7895)
```

```
#2nd Line of Business
data(austri2auto)
```

Format

`austri1autoBI7895$paid`, `austri1autoBI7895$incur`, `austri1autoBI7895$nb` contain the insurance triangle, respectively for paid, incurred claims and claim number. `austri1autoBI7895$expo` contains the vector of exposure, `austri1autoBI7895$infl` contains the vector of inflation indexes.

`austri2auto` contains the run-off insurance triangle.

Source

Formerly on a website dedicated to P. De Jong and G.Z. Heller (2008).

References

G. Taylor (2000), *Loss reserving: an actuarial perspective*, Springer Science + Business Media.

P. De Jong and G.Z. Heller (2008), *Generalized linear models for insurance data*, Cambridge University Press, doi:[10.1017/CBO9780511755408](https://doi.org/10.1017/CBO9780511755408).

Examples

```
# (1) load of data
#

#1st Line of Business
data(austri1autoBI7895)

#2nd Line of Business
data(austri2auto)

# (2) graph
#
i <- 2
matplot(cbind(cumsum(austri1autoBI7895$paid[i,]), cumsum(austri1autoBI7895$incur[i,])),
  type="l", ylab="Claim Amount (orig. USD)", xlab="Development Year",
  main="Incurred vs. paid claim")

#claim number per 100 000 vehicles
fexpo <- 100000 / austri1autoBI7895$expo[rownames(austri1autoBI7895$nb)]
round(austri1autoBI7895$nb * fexpo, 1) #Table 2.1 of Taylor (2000)
matplot(apply(austri1autoBI7895$nb * fexpo, 1, cumsum), type="l",
  main="number of claim notified per 100,000 veh", ylab="Cumulative number",
  xlab="Development Year")

#paid claim inflated
fclaim <- austri1autoBI7895$infl["1995", "EndYear"] / austri1autoBI7895$infl[, "EndYear"]

g <- function(x, n) c(tail(x, -n), rep(NA, n))
infl <- cbind(fclaim, sapply(1:17, function(n) g(fclaim, n)))

matplot(apply(austri1autoBI7895$paid * infl, 1, cumsum), type="l", ylab="Claim Amount (1995 AUD)",
  xlab="Development Year", main="Paid claim", col=1:10)
legend("bottomright", leg=rownames(austri1autoBI7895$paid), lty=1:5, col=1:10, cex=.5)

#incurred claim inflated
matplot(apply(austri1autoBI7895$incur * infl, 1, cumsum), type="l", ylab="Claim Amount (1995 AUD)",
  xlab="Development Year", main="Incurred claim", col=1:10)
legend("bottomright", leg=rownames(austri1autoBI7895$incur), lty=1:5, col=1:10, cex=.5)
```

beaonre	<i>AON Re Belgian dataset</i>
---------	-------------------------------

Description

The dataset was collected by the reinsurance broker AON Re Belgium and comprise 1,823 fire losses for which the building type and the sum insured are available.

Usage

```
data(beaonre)
```

Format

beaonre contains three columns and 1823 rows:

BuildType The building type either A, B, C, D, E or F.

ClaimCost The loss amount in thousand of Danish Krone (DKK).

SumInsured The sum insured in thousand of Danish Krone (DKK).

References

Dataset used in Beirlant, Dierckx, Goegebeur and Matthys (1999), *Tail index estimation and an exponential regression model*, Extremes 2, 177-200, doi:[10.1023/A:1009975020370](https://doi.org/10.1023/A:1009975020370).

Examples

```
# (1) load of data
#
data(beaonre)

# (2) plot and description of data
#

boxplot(ClaimCost ~ BuildType, data=beaonre, log="y",
        xlab="Building type", ylab="Claim size", main="AON Re Belgium data")
```

beMTPL16	<i>Belgian motor third-part liability dataset</i>
----------	---

Description

The dataset beMTPL was collected by an unknown Belgium insurer. It consists of 70 791 claims for private motor insurance.

Usage

```
data(beMTPL16)
```

Format

beMTPL16 contains:

insurance_contract a numeric for Unique identifier for the contract
 policy_year a numeric for Year of study or observation for the insured person
 exposure a numeric for Exposure duration in years
 insured_year_birth a numeric for insured's year of birth
 vehicle_age a numeric for Age of the vehicle in years
 policy_holder_age a numeric for Seniority of the insured at the insurance agency
 driver_license_age a numeric for Age of the driver's licence
 vehicle_brand a factor for Brand of the vehicle
 vehicle_model a factor for Model of the vehicle
 mileage a numeric for Mileage of the vehicle
 vehicle_power a numeric for Power value of the vehicle
 catalog_value a numeric for Catalog value of the vehicle
 claim_value a numeric for Value of the claim
 number_of_liability_claims a numeric for Number of liability claims
 number_of_bodily_injury_liability_claims a numeric for Number of bodily injury liability claims
 claim_time a factor for Time (within a day) of the accident
 claim_responsibility_rate a numeric for Rate of responsibility for the claim (100% full responsibility, 0% no responsibility)
 driving_training_label a factor indicating driving training program
 signal a numeric indicating 1 = warning, 0 = no warning

Source

Unknown insurer

Examples

```
# (1) load of data
#
data(beMTPL16)
```

Description

The portfolio contains 163,212 unique policyholders, each observed during a period of exposure-to-risk expressed as the fraction of the year during which the policyholder was exposed to the risk of filing a claim. Claim information is known in the form of the number of claims filed and the total amount claimed (in euro) by a policyholder during the period of exposure. The data set lists five categorical, four continuous and two spatial risk factors, each of them informing about specific characteristics of the policy or the policyholder. A detailed discussion on the distribution of all variables is available in Henckaerts et al. (2018) and some code examples is available at <https://github.com/henckr/treeML>.

Usage

```
data(beMTPL97)
```

Format

beMTPL97 contains:

`id` a numeric for the policy number.

`expo` a numeric for the exposure.

`claim` a factor indicating a claim.

`nclaims` a numeric for the claim number.

`amount` a numeric for the aggregate claim amount.

`average` a numeric for the average claim amount.

`coverage` a factor for the insurance coverage level: "TPL" only third party liability, "TPL+" TPL + limited material damage, "TPL++" TPL + comprehensive material damage.

`ageph` a numeric for the policyholder age.

`sex` a factor for the policyholder gender: "female", "male".

`bm` an integer for the level occupied in the former compulsory Belgian bonus-malus scale. From 0 to 22, a higher level indicates a worse claim history, see Lemaire (1995).

`power` a numeric for Horsepower of the vehicle in kilowatt.

`agec` a numeric for Age of the vehicle in years.

`fuel` a factor for Type of fuel of the vehicle: "gasoline" or "diesel".

`use` a factor for the use of the vehicle: "private" or "work".

`fleet` an integer indicating if the vehicle is part of a fleet: 1 or 0.

`postcode` the postal code of the policyholder.

`long` a numeric for the longitude coordinate of the center of the municipality where the policyholder resides.

`lat` a numeric for the latitude coordinate of the center of the municipality where the policyholder resides.

Source

Unknown insurer

References

Lemaire (1995). *Bonus-malus systems in automobile insurance*, Springer, New York, 1995, doi:[10.1007/9789401106313](https://doi.org/10.1007/9789401106313)

Denuit and Lang (2004), *Non-life rate-making with Bayesian GAMs*, Insurance: Mathematics and Economics, 35(3):627–647, doi:[10.1016/j.insmatheco.2004.08.001](https://doi.org/10.1016/j.insmatheco.2004.08.001)

Denuit et al. (2007) *Actuarial modelling of claim counts: Risk classification, credibility and bonus-malus systems*, John Wiley and Sons Ltd, West Sussex, doi:[10.1002/9780470517420.fmatter](https://doi.org/10.1002/9780470517420.fmatter)

Klein et al. (2014) *Nonlife ratemaking and risk management with Bayesian generalized additive models for location, scale, and shape*, Insurance: Mathematics and Economics, 55:225–249 doi:[10.1016/j.insmatheco.2014.02.001](https://doi.org/10.1016/j.insmatheco.2014.02.001)

Henckaerts et al. (2018). *A data driven binning strategy for the construction of insurance tariff classes*, Scandinavian Actuarial Journal, 2018(8):681–705, doi:[10.1080/03461238.2018.1429300](https://doi.org/10.1080/03461238.2018.1429300)

Frees, Carriere and Valdez (1995), *Annuity valuation with dependent mortality*, Actuarial Research Clearing House 1995, Vol. 2, doi:[10.2307/253744](https://doi.org/10.2307/253744).

Examples

```
# (1) load of data
#
data(beMTPL97)
```

besecura

Secura Re Belgian dataset

Description

The dataset was collected by the reinsurer Secura Re Belgium and comprises of 371 automobile claims from 1988 until 2001. The original claim numbers were corrected, among others, for inflation to reflect 2002 euros.

Usage

```
data(besecura)
```

Format

besecura contains two columns and 371 rows:

Year The year of claim occurrence.

Loss The loss amount in euros (EUR).

Source

<https://lstat.kuleuven.be/>

References

Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. L. (2004) *Statistics of Extremes: Theory and Applications.*, Chichester, England: John Wiley and Sons, [doi:10.1002/0470012382](https://doi.org/10.1002/0470012382).

Examples

```
# (1) load of data
#
data(besecura)

# (2) plot and description of data
#

plot(Loss ~ Year, data= besecura, log="y", xlab="Year",
      ylab="Claim size", main="Secura Re Belgian dataset")
```

bragg

Descriptive statistics of aggregate claims and premiums for the 41 Brazilian regions

Description

The datasets `braggclaim` and `braggprem` are descriptive statistics of the premium/claim per region and type of insurance coverage. Therefore, for each region, there are five rows, one for each type of insurance coverage, i.e. 405 row in total.

Usage

```
data(braggclaim)
data(braggprem)
```

Format

`braggprem` contains 7 columns:

- RegionNb** A numeric for the region number.
- RegionName** A character for the region name
- Guarantee** A character string for the guarantee.
- ExpoAvg** A numeric for the average of total exposures.
- PremAvg** A numeric for the average of gross written premium.
- SumInsAvg** A numeric for the average of sum insured.
- StateAb** A character string for the abbreviated state name.

`braggclaim` contains 6 columns:

- RegionNb** A numeric for the region number.
- RegionName** A character for the region name

Guarantee A character string for the guarantee.

ClaimNb A numeric for the claim number.

AggClaim A numeric for the aggregate claim amount.

StateAb A character string for the abbreviated state name.

Source

The original dataset was provided in Chapter 5 of Charpentier (2014).

References

Charpentier, A. (2014). *Computational Actuarial Science with R*. CRC Press.

Examples

```
# (1) load of data
#
data(braggclaim)
data(braggprem)
```

brautocoll

Brazilian Automobile Collision Claims

Description

Dataset of car traffic collisions that occurred in February 2011, in Belo Horizonte, a Brazilian city. A record consists of date, day, hour, locations (long, lat) and severity for a given collision.

Usage

```
data(brautocoll)
```

Format

brautocoll contains 5 columns:

Date The date of the traffic collision, see [Date](#).

Day A character string for the weekday.

Hour Hour on the format hh:mm.

Lat Latitude of the location.

Long Longitude of the location.

Type A character string for the claim type.

Severity A character string for the severity.

Source

The original dataset was provided in Chapter 5 of Charpentier (2014).

References

Charpentier, A. (2014). *Computational Actuarial Science with R*. CRC Press.

Examples

```
# (1) load of data
#
data(brautocoll)
dim(brautocoll)
```

brgeomunicins

Brazilian geospatial dataset for municipalities

Description

brgeomunic is a spatial database containing geospatial information of Brazilian municipalities provided by IBGE, the Brazilian governmental agency in charge of geographical issues and official statistics (ibge.gov.br, accessed in February, 2013). brgeomunic is a geospatial dataframe of class `sp` based on three files: one containing the geographical coordinates of the polygons, lines or dots (55mu2500gsd.shp); another with attribute data (55mu2500gsd.dbf); a third file with the index that allows the connection between the .shp and .dbf files (55mu2500gsd.shx). brgeomunic is provided in two versions `sp` and `sf` in the directory `geodata` on the github directory, please visit the repository at <https://github.com/dutangc/CASdatasets>. You may also consider the package `geobr` at <https://cran.r-project.org/package=geobr>.

The final database is restricted to the municipalities from only four Brazilian states (Sao Paulo (SP), Santa Catarina (SC), Parana (PR), and Rio Grande do Sul (RS)). These states are located in the southern region of Brazil and contain almost 70 million inhabitants (around 36 percent of the Brazilian population) and constitute one of the richest regions of the country (approximately 60 percent of the Brazilian gross product).

brgeomunicins is a dataframe with insurance statistic information. The insurance information comes from one large actuarial database provided by SUSEP, the agency responsible for the regulation and supervision of the Brazilian insurance, private pension, annuity, and reinsurance markets. SUSEP releases biannually a car insurance database composed of the aggregation of all insurance companies' information. Due to confidentiality concerns, there is no individual-level information, the data being aggregated into zip code areas. Originally, both SUSEP and IBGE databases did not present a unique identification column that provides a forward merge of the two databases. The joint information is the name and the state of each municipality.

Insurance information have been selected to compare premiums, claims, and reported damages for two specific groups: popular vehicles and luxury vehicles. The basic difference between the groups is the power of the engine and the materials and finishing quality. Popular cars have a power of 1,000 cc (cylinders), whereas luxury cars usually have a power of 2,000 cc or greater. Popular cars are thus affordable to most customers.

The Pop group contains the following selected popular vehicles: Celta 1.0 (Chevrolet), Corsa 1.0 (Chevrolet), Prisma 1.0 (Chevrolet), Uno 1.0 (Fiat), Palio 1.0 (Fiat), Gol 1. (Volkswagen), Fox 1.0 (Volkswagen), Fiesta 1.0 (Ford), and Ka 1.0 (Ford).

The Lux group contains the following selected luxury vehicles: Vectra (Chevrolet), Omega (Chevrolet), Linea (Fiat), Bravo (Fiat), Passat (Volkswagen), Polo (Volkswagen), Fusion (Ford), Focus (Ford), Corolla (Toyota), Civic (Honda), and Audi.

In summary, `brgeomunicins` is a dataframe with detailed information of region, city code, yearly exposure, premium, and frequency of claims for the following categories: robbery or theft (`Rob`), partial collision and total loss (`Coll`), fire (`Fire`), or others (`Other`).

In addition to insurance statistics, the final dataframe `brgeomunicins` also includes the municipality population (`CityDens10`) based on the 2010 Census, and the 2000 municipality Human Development Index (`HDIcity00`). The Human Development Index (HDI) is a summary measure of long-term progress in three basic dimensions of human development: income, education, and health. The HDI provides a counterpoint to another widely used indicator, the Gross Domestic Product (GDP) per capita, which only considers economic dimensions. Both `CityDens10` and `HDIcity00` columns were generated from the IBGE site (ibge.gov.br, accessed February 2013).

Usage

```
data(brgeomunicins)
```

Format

`brgeomunicins` contains 18 columns:

CityCode A character string for the severity.

State, StateAb Character string ("factor") for the full state name and the two-letter abbreviated state name.

City A character string ("factor") for the cityname.

PopExpo, LuxExpo The sum of exposure periods for policies in the Pop and the Lux groups, in years.

PopPrem, LuxPrem Gross written premium for the Pop and the Lux groups, respectively.

PopClaimRob, LuxClaimRob Aggregate robbery claim number for the Pop and the Lux groups, respectively.

PopClaimColl, LuxClaimColl Aggregate collision claim number for the Pop and the Lux groups, respectively.

PopClaimFire, LuxClaimFire Aggregate fire claim number for the Pop and the Lux groups, respectively.

PopClaimOther, LuxClaimOther Aggregate other claim number for the Pop and the Lux groups, respectively.

HDIcity00 A numeric for the HDI index of the city.

CityDens10 A numeric for the population density.

Source

The original dataset was provided in Chapter 5 of Charpentier (2014).

References

Charpentier, A. (2014). *Computational Actuarial Science with R*. CRC Press.

Examples

```
# (1) load of data
#
data(brgeomunicins)
```

brvehins

*Two Brazilian datasets for vehicle insurance***Description**

brvehins1's , brvehins2's are dataframes containing policy data based on the AUTOSEG (an acronym for Statistical System for Automobiles) and can be accessed online (www2.susep.gov.br/menuestatistica/Autoseg, accessed February 2013). Each record includes risk features, claim amount and claim history for year 2011. The dataset brvehins1 of 1,965,355 vehicle insurance policies has been splitted (randomly) in five datasets of 393,071 policies : brvehins1a, brvehins1b, brvehins1c, brvehins1d, brvehins1e. The dataset brvehins2 of 2,667,752 policies has also been splitted (randomly) in four datasets of 666,938 policies : brvehins2a, brvehins2b, brvehins2c, brvehins2d.

Usage

```
data(brvehins1a)
data(brvehins1b)
data(brvehins1c)
data(brvehins1d)
data(brvehins1e)
```

```
data(brvehins2a)
data(brvehins2b)
data(brvehins2c)
data(brvehins2d)
```

Format

brvehins1's contains 23 columns:

Gender A character string ("factor") for the gender (also indicate corporate policies).

DrivAge A character string ("factor") for the driver age group.

VehYear A numeric for the vehicle year.

FullVehCode A character string ("factor") for the full vehicle code.

VehCode A character string ("factor") for the vehicle group.

Area Local area name ("factor").

State A character string for the state name ("factor").

StateAb Abbreviated state name ("factor").

ExposTotal Total exposure

ExposFireRob Exposure for fire and robbery guarantees.

PremTotal Total premium.

PremFireRob Premium for fire and robbery guarantees.

SumInsAvg Average of sum insured.

ClaimNbRob,ClaimNbPartColl,ClaimNbTotColl,ClaimNbFire,ClaimNbOther Number of claims during the exposure period, respectively for robbery, partial collision, total collision, fire and other guarantees.

ClaimAmountRob,ClaimAmountPartColl,ClaimAmountTotColl,ClaimAmountFire,ClaimAmountOther Claim amounts during the exposure period, respectively for robbery, partial collision, total collision, fire and other guarantees.

brvehins2's contains 18 columns:

VehYear A numeric for the vehicle year.

FullVehCode A character string ("factor") for the full vehicle code.

VehCode A character string ("factor") for the vehicle group.

City A character string ("factor") for the city name.

CityCode A numeric for the city code.

Area Local area name ("factor").

State A character string ("factor") for the state name.

StateAb Abbreviated state name ("factor").

ExposTotal Total exposure

PremTotal Total premium.

ClaimNbRob,ClaimNbColl,ClaimNbFire,ClaimNbOther Number of claims during the exposure period, respectively for robbery, (partial and total) collision, fire and other guarantees.

ClaimAmountRob,ClaimAmountColl,ClaimAmountFire,ClaimAmountOther Claim amounts during the exposure period, respectively for robbery, (partial and total) collision, fire and other guarantees.

Source

www2.susep.gov.br/menuestatistica/Autoseg

Examples

```
## Not run:

# (1) load of data
#
data(brvehins1a)
dim(brvehins1a)
sapply(brvehins1a, class)
str(brvehins1a)

data(brvehins2a)
dim(brvehins2a)
sapply(brvehins2a, class)
str(brvehins2a)

## End(Not run)
```

canlifins

Canadian life insurance

Description

This dataset contains information of 14,889 contracts in force with a large Canadian insurer over the period December 29, 1988 through December 31, 1993. These contracts are joint and last-survivor annuities that were in the payout status over the observation period. For each contract, we have the date of birth, date of death (if applicable) and sex of each annuitant. Binary dummies for uncensored observations and exit times are also available.

Usage

```
data(canlifins)
data(canlifins2)
```

Format

canlifins is a data frame of 10 columns and 14,889 rows:

EntryAgeM Entry age of the male.

DeathTimeM Time of death of the male (zero if not applicable).

AnnuityExpiredM The date that the annuity guarantee expired (if applicable).

IsDeadM A binary indicating uncensored observation.

ExitAgeM Exit age of the male.

EntryAgeF Entry age of the female.

DeathTimeF Time of death of the female (zero if not applicable).

AnnuityExpiredF The date that the annuity guarantee expired (if applicable).

IsDeadF A binary indicating uncensored observation.

ExitAgeF Exit age of the female.

Originally in Frees et al. (1995), the dataset contains 22 contracts where both annuitants are male, 36 contracts where both annuitants are female, in addition to 14,889 contracts where one annuitant is male and the other female (so a total of 14,947 contracts).

canlifins2 is a data frame of 2 columns and 14,889 rows with either the observed death age in canlifins or simulated death age based on the residual survival time. Dependency between male and female is taken into account.

DeathAgeM Death age of the male.

DeathAgeF Death age of the female.

Source

Unknown private insurer.

References

Dataset used in Frees, Carriere and Valdez (1995), *Annuity valuation with dependent mortality*, Actuarial Research Clearing House 1995, Vol. 2, [doi:10.2307/253744](https://doi.org/10.2307/253744).

Examples

```

# (1) load of data
#
data(canlifins)
dim(canlifins)

# (2) Table 1 of Frees et al. (1995)
#

table(Age=cut(canlifins$EntryAgeM, c(0, 60, 70, 80, 120)),
      Status=cut(canlifins$DeathTimeM, c(-1, 0, 10)))
table(Age=cut(canlifins$EntryAgeF, c(0, 60, 70, 80, 120)),
      Status=cut(canlifins$DeathTimeF, c(-1, 0, 10)))

# (3) Figure 1 of Frees et al. (1995)
#

KMmale <- survfit(Surv(ExitAgeM, IsDeadM)~1, data=canlifins)
KMfemale <- survfit(Surv(ExitAgeF, IsDeadF)~1, data=canlifins)

plot(KMmale, xlim=c(40, 110), conf.int=FALSE)
lines(KMfemale, col="red", conf.int=FALSE)
grid()

```

CASdatasets

CASdatasets package

Description

Actuarial Datasets (originally for the 'Computational Actuarial Science with R' book)

Details

This package contains aggregated and policy-level datasets. Below a list by country or region is given.

- Australia:
 - [auscathist](#): Historical disaster statistics in Australia.
 - [ausNLHYCapAdeq](#), [ausNLHYFinPerf](#), [ausNLHYFinPos](#), [ausNLHYLiability](#), [ausNLHYOffProf](#), [ausNLHYOpIncExp](#), [ausNLHYPremClaim](#), [ausNLHYPrivInsur](#), [ausNLHYPubInsur](#), [ausNLHYRecAASB](#), [ausNLHYReserve](#): Australian Market - non-life insurance (aggregate level).
 - [ausNLHYCapAdeqByComp](#), [ausNLHYClaimByState](#), [ausNLHYFinPerfByComp](#), [ausNLHYFinPerfPublic](#), [ausNLHYFinPosByComp](#), [ausNLHYFinPosPublic](#), [ausNLHYOpIncExpPublic](#), [ausNLHYPremByState](#), [ausNLHYPremClaimPublic](#), [ausNLHYPrivInsur](#), [ausNLHYPubInsur](#): Australian Market - non-life insurance (company, state, public level).
 - [ausNLHYLloydAsset](#), [ausNLHYLloydGPI](#), [ausNLHYLloydUWAcc](#), [ausNLHYLloydUWRes](#): Lloyds Market in Australia.
 - [austri1autoBI7895](#), [austri2auto](#): Australian claim triangles.

- [ausprivautolong](#), [ausMTPL8486](#), [ausprivauto0405](#): Australian private motor datasets (by policy).
- [ausNSWdriver04](#), [ausNSWdeath02](#): New South Wales statistics.
- Belgium:
 - [beaonre](#): AON Re Belgium dataset.
 - [beMTPL16](#): a MTPL Belgium dataset of 2016 (by policy).
 - [beMTPL97](#): a MTPL Belgium dataset of 1997 (by policy).
 - [besecura](#): Secura Re Belgium dataset.
- Brazil:
 - [brgeomunicins](#): a dataset of Brazilian cities with insurance indicators (exposure, claim frequency and premium). Geographical datasets [brgeomunic_sf](#), [brgeomunic_sp](#) should be downloaded from the github repository in extra directory [geodata](#).
 - [brautocoll](#): a Brazilian dataset of auto collision in Belo Horizonte during one month (categorical claim severity).
 - [braggprem](#), [braggclaim](#): aggregate Brazilian datasets per region.
 - [brvehins1](#): a Brazilian vehicle insurance datasets (by policy) with risk features except City code.
 - [brvehins2](#): a Brazilian vehicle insurance datasets (by policy) with risk features including City code.
- Canada:
 - [canlifins](#), [canlifins2](#): A portfolio of a Canadian life insurer.
 - [catelematic13](#): Canadian telematic insurance dataset.
- Denmark:
 - [danishuni](#), [danishmulti](#): Danish reinsurance claim dataset.
- European Union:
 - [eutelematic13](#): telematic insurance dataset.
 - [eudirectlapse](#): non-life insurance lapse dataset.
 - [euhealthinsurance](#): European health insurance dataset.
 - [euMTPL](#): European motor TPL insurance dataset.
- France:
 - [freaggnumber](#): a French aggregate claim number dataset.
 - [frebiloss](#): French business interruption losses.
 - [freclaimset](#), [freclaimset9207](#): French claim settlements.
 - [frecomfire](#): French commercial fire losses.
 - [freDemography](#): datasets of the French demography.
 - [freDisTables](#): French disability tables.
 - [freIRSA](#): French IRSA convention.
 - [fremarine](#): French marine claim dataset (by policy).
 - [freMortTables](#): French mortality tables.
 - [fremotorclaim](#): French private motor claim datasets (by policy).
 - [freMTPL](#): two French Motor-TPL claim datasets (by policy).
 - [freMPL](#): a collection of ten French Motor personal line datasets (by policy).
 - [freportfolio](#): fictive mortality tables and French nation-wide corresponding tables; two disability datasets from a French insurer.
 - [fretelematic](#): a French telematic Motor-TPL claim dataset (by policy).

- [fretplclaimnumber](#): a French Motor-TPL claim dataset (by policy).
- [fre4LoBtriangles](#): A collection of triangles line of business from a private insurer.
- [pricingame](#): Datasets used at Pricing Games of the French Institute of Actuaries.
- Germany:
 - [credit](#): A German Credit dataset.
- Italy:
 - [itamtplcost](#): Large losses of an Italian Motor-TPL company.
- New Zealand:
 - [nzcathist](#): Historical disaster statistics in New Zealand.
- Norway:
 - [norauto](#): Norwegian automobile dataset.
 - [norfire](#): Norwegian fire dataset.
 - [Norberg](#): Norberg’s credibility dataset.
 - [nortritpl8800](#): Norwegian claim triangle.
- Singapore:
 - [sgautonb](#): Singapore Automobile claim count dataset.
 - [sgtriangles](#): Singapore Property and Casualty triangles.
- Sweden:
 - [swautoins](#): Swedish Motor Insurance dataset
 - [swbusscase](#): Swedish Buss Insurance dataset
 - [swmotorcycle](#): Swedish Motorcycle Insurance dataset
- United Kingdom:
 - [ukaggclaim](#): United Kingdom Car Insurance Claims.
 - [ukautocoll](#): United Kingdom Car Collision Insurance Claims.
- United States of America:
 - [Davis](#): Davis height-weight dataset.
 - [ICB1](#), [ICB2](#): Insurance Company Benchmarks.
 - [lossalae](#), [lossalae-full](#): General Third Part-liability claims and expenses.
 - [SOAGMI](#): SOA Group Medical Insurance dataset.
 - [usautoBI](#): Automobile Bodily Injuries in US.
 - [usautotriangles](#): US automobile triangles.
 - [usexpense](#): US expense dataset.
 - [usGLtriangles](#): US Property and Casualty triangles.
 - [ushurricane](#), [ushustormloss4980](#): Historical hurricane statistics in United States of America.
 - [uslapseagent](#): US lapse dataset from tied-agent channel.
 - [usmassBI](#): US Massachusetts Automobile bodily injury claim datasets.
 - [usmedclaim](#): US medical claim triangle.
 - [usMSHA1316](#): US Mine Safety and Health Administration claim dataset.
 - [usMVTA](#): US motor vehicle traffic accident.
 - [usprivautoclaim](#): private automobile claims.
 - [usquakeLR](#): California earthquake loss ratios.
 - [ustermLife](#): Term life insurance survey.
 - [uswarrantagnum](#): US warranty automobile.

- [usworkcomp](#): US workers compensation datasets.
- Misc.:
 - [eqlist](#): Earthquake list.
 - [hurricanehist](#): Hurricane history.
 - [PnCdemand](#): Property and Casualty demand.
 - [spacedata](#): Space dataset.
 - [ECBYieldCurve](#), [FedYieldCurve](#) Yield curve for eurozone and US.
 - [forexUSUK](#) Foreign exchange rate between USD and GBP.
 - Use the HMD website <https://www.mortality.org/> for mortality tables.

Here is a list of datasets whose name has changed compared to the book ‘Computational Actuarial Science with R’:

- Chapter 1: `extreme2datasince1899` is [hurricanehist](#).
- Chapter 5: `accidents` and `accidents_data` are merged in [brautocoll](#); `55mu2500gsd` is `brgeomunic_sf` in extra directory `geodata` in the github repository; `sul_sp` is stored in [brgeomunicins](#).
- Chapter 9: `MyPortfolio` is [freprojqxINSEE](#).
- Chapter 10: `DataMortality` is [freptfpermdis](#).
- Chapter 11: `DEXUSUK` is [forexUSUK](#).
- Chapter 14: `CONTRACTS` is [freMTPLfreq](#); `CLAIMS` is [freMTPLsev](#).
- Chapter 15: `AutoClaimData` is [usmassBI2](#).

Author(s)

Arthur Charpentier, Christophe Dutang.

catelematic13

Canadian telematic insurance dataset

Description

This dataset is based on a real dataset acquired from a Canadian-based insurer, which offered a UBI program that was launched in 2013, to its automobile insurance policyholders. The observation period was for the years between 2013 and 2016, with over 70,000 policies being observed, for which the dataset drawn is pre-engineered for training a statistical model for predictive purposes.

Usage

```
data(catelematic13)
```

Format

catelematic13 is a data frame of 10 columns and 14,889 rows:

Insured.age Age of insured driver, in years.

Insured.sex Sex of insured driver (Male/Female).

Car.age Age of vehicle, in years.

Marital Marital status (Single/Married).

Car.use Use of vehicle: Private, Commute, Farmer, Commercial.

Credit.score Credit score of insured driver.

Region Type of region where driver lives: rural, urban.

Annual.miles.drive Annual miles expected to be driven declared by driver.

Years.noclaims Number of years without any claims.

Territory Territorial location of vehicle.

Annual.pct.driven Annualized percentage of time on the road.

Total.miles.driven Total distance driven in miles.

Pct.drive.xxx Percent of driving day xxx of the week: mon/tue/. . . /sun.

Pct.drive.xhrs Percent vehicle driven within x hrs: 2hrs/3hrs/4hrs.

Pct.drive.xxx Percent vehicle driven during xxx: wkday/wkend.

Pct.drive.rushxx Percent of driving during xx rush hours: am/pm.

Avgdays.week Mean number of days used per week.

Accel.xxmiles Number of sudden acceleration 6/8/9/. . . /14 mph/s per 1000 miles.

Brake.xxmiles Number of sudden brakes 6/8/9/. . . /14 mph/s per 1000 miles.

Left.turn.intensityxx Number of left turn per 1000 miles with intensity 08/09/10/11/12.

Right.turn.intensityxx Number of right turn per 1000 miles with intensity 08/09/10/11/12.

NB_Claim Number of claims during observation.

AMT_Claim Aggregated amount of claims during observation.

Source

<http://www2.math.uconn.edu/~valdez/data.html>

References

Banghee So, Jean-Philippe Boucher and Emiliano A. Valdez (2021), *Synthetic Dataset Generation of DriverTelematics*, Risks 9:58, doi:10.3390/risks9040058

Examples

```
# (1) load of data
#
data(catelematic13)
dim(catelematic13)
```

 credit

German Credit dataset

Description

This dataset contains information of 1,000 credit records. It is a consumer credit files, called the German Credit dataset in Tuff'ery (2011) and Nisbet et al. (2011). New applicants for credit and loans can be evaluated as good or bad payers using 21 explanatory variables.

Usage

```
data(credit)
```

Format

credit is a data frame of 21 columns and 1,000 rows:

checking_status Status of existing checking account, A11: less than 0, A12: from 0 to 200, A13: more than 200, and A14: no running account (or unknown).

duration credit duration in months.

credit_history credit history: A30: delay in paying off in the past, A31: critical account, A32: no credits taken or all credits paid back duly, A33: existing credits paid back duly till now, A34: all credits at this bank paid back duly.

purpose purpose of credit: A40: new car, A41: used car, A42: items of furniture/equipment, A43: radio/television, A44: domestic household appliances, A45: repairs, A46: education, A47: vacation, A48: retraining, A49: business, A410: others.

credit_amount credit amount in Deutsch marks.

savings saving account: A61: less than 100, A62: from 100 to 500, A63: from 500 to 1,000, A64: more than 1,000, A65: no savings account (or unknown).

employment Present employment since: A71: unemployed, A72: less than 1 year, A73: from 1 to 4 years, A74: from 4 to 7 years, A75: more than 7 years.

installment_rate Installment rate (in percentage of disposable income) A81: greater than 35, A82: between 25 and 35, A83: between 20 and 25, A84: less than 20.

personal_status Personal status and sex: A91: male: divorced/separated, A92: female: divorced/separated/married, A93: male: single, A94: male: married/widowed, A95: female: single.

other_parties Other debtors or guarantors: A101: none, A102: co-applicant, A103: guarantor.

residence_since Present residence since: A71: less than 1 year, A73: from 1 to 4 years, A74: from 4 to 7 years, A75: more than 7 years.

property_magnitude Property (most valuable): A121: real estate (ownership of house or land), A122: savings contract with a building society / Life insurance, A123: car or other, A124: unknown / no property.

age Age (in years).

other_payment_plans Other installment plans: A141: at other bank, A142: at department store or mail order house, A143: no further running credits.

housing Housing: A151: rented flat, A152: owner-occupied flat, A153: free apartment.

existing_credits Number of existing credits at this bank (including the running one) A161: one, A162: two or three, A163: four or five, A164: six or more.

job Job: A171: unemployed / unskilled with no permanent residence, A172: unskilled with permanent residence, A173: skilled worker / skilled employee / minor civil servant, A174: executive / self-employed / higher civil servant.

num_dependents Number of people being liable to provide maintenance for A181: zero to two, A182: three and more.

telephone Telephone: A191: none, A192: yes, registered under the customers name.

foreign_worker Foreign worker: A201: yes, A202: no.

class binary variable 0 stands for good and 1 bad (or credit-worthy against not credit-worthy, or no non-payments against existing non-payments).

Source

The original data was provided by:

Professor Dr. Hans Hofmann, Institut fuer Statistik und Oekonometrie,
Universitaet Hamburg, FB Wirtschaftswissenschaften, Von-Melle-Park 5, 2000 Hamburg 13

Professor Dr. Hans Hofmann, Institut fur Statistik und Oekonometrie,
Universitaet Hamburg, FB Wirtschaftswissenschaften, Von-Melle-Park 5, 2000 Hamburg 13

The dataset has been taken from the UCI Repository Of Machine Learning Databases at

[https://archive.ics.uci.edu/ml/datasets/Statlog+\(German+Credit+Data\)](https://archive.ics.uci.edu/ml/datasets/Statlog+(German+Credit+Data))

Formerly available at

<https://www.stat.lmu.de/en/>

References

Fahrmeir, L. and Tutz, G. (1994), *Multivariate Statistical Modelling Based on Generalized Linear Models*, Springer, doi:10.1007/9781489900104.

Nisbet, R., Elder, J. and Miner, G. (2011), *Handbook of Statistical Analysis and Data Mining Applications*, Academic Press, doi:10.1016/B9780123747655.X00010.

Tuff'ery, S. (2011), *Data Mining and Statistics for Decision Making*, Wiley, doi:10.1002/9780470979174.

See Also

For a good variable description, see also [https://archive.ics.uci.edu/ml/datasets/Statlog+\(German+Credit+Data\)](https://archive.ics.uci.edu/ml/datasets/Statlog+(German+Credit+Data)).

Examples

```
# (1) load of data
#
data(credit)
dim(credit)
head(credit)
```

danish

Danish reinsurance claim dataset

Description

The univariate dataset was collected at Copenhagen Reinsurance and comprise 2167 fire losses over the period 1980 to 1990. They have been adjusted for inflation to reflect 1985 values and are expressed in millions of Danish Krone.

The multivariate dataset is the same data as above but the total claim has been divided into a building loss, a loss of contents and a loss of profits.

Usage

```
data(danishuni)
data(danishmulti)
```

Format

danishuni contains two columns:

Date The day of claim occurrence.

Loss The total loss amount in millions of Danish Krone (DKK).

danishmulti contains five columns:

Date The day of claim occurrence.

Building The loss amount (mDKK) of the building coverage.

Contents The loss amount (mDKK) of the contents coverage.

Profits The loss amount (mDKK) of the profit coverage.

Total The total loss amount (mDKK).

All columns are numeric except Date columns of class Date.

Source

Embrechts, P., Kluppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance*. Berlin: Springer.

References

McNeil, A. (1996), *Estimating the Tails of Loss Severity Distributions using Extreme Value Theory*, ASTIN Bull, [doi:10.2143/AST.27.1.563210](https://doi.org/10.2143/AST.27.1.563210).

Davison, A. C. (2003) *Statistical Models*. Cambridge University Press, [doi:10.1017/CBO9780511815850](https://doi.org/10.1017/CBO9780511815850).

Examples

```

# (1) load of data
#
data(danishuni)

# (2) plot and description of data
#
plot(danishuni$Loss)

# (3) load of data
#
data(danishmulti)

# (4) plot and description of data
#
idx <- sample(1:NROW(danishmulti), 10)
barplot(danishmulti$Building[idx], col="grey25",
        ylim=c(0, max(danishmulti$Total[idx])), main="Some claims of danish dataset")
barplot(danishmulti$Content[idx], add=TRUE, col="grey50", axes=FALSE)
barplot(danishmulti$Profits[idx], add=TRUE, col="grey75", axes=FALSE)
legend("topleft", legend=c("Building", "Content", "Profits"), fill=c("grey25",
        "grey50", "grey75"))

```

 Davis

Davis dataset

Description

This dataset contains information of 200 individuals.

Usage

```
data(Davis)
```

Format

`data` is a data frame of 5 columns and 200 rows:

`sex` a factor: M for male and F for female.

`weight` a numeric for the weight in Kg.

`height` a numeric for the height in cm.

`reportedWeight` a numeric for the weight in Kg.

`reportedHeight` a numeric for the height in cm.

References

Davis (1990) *Body image and weight preoccupation: A comparison between exercising and non-exercising women*, *Appetite*, 15, 13-21, doi:[10.1016/01956663\(90\)90096q](https://doi.org/10.1016/01956663(90)90096q).

Examples

```
# (1) load of data
#
data(Davis)
dim(Davis)
head(Davis)
```

ECBYieldCurve	<i>Yield curve data spot rate, AAA-rated bonds, maturities from 3 months to 30 years</i>
---------------	--

Description

Government bond, nominal, all triple A issuer companies. The maturities are 3 and 6 months and from 1 year to 30 years with frequency business day, provided by European Central Bank. The range date is from 2006-12-29 to 2009-07-24.

Usage

```
data(ECBYieldCurve)
```

Format

It is an xts object with 32 interest rate at different maturities and 655 observations.

Source

ECB: https://www.ecb.europa.eu/stats/financial_markets_and_interest_rates/euro_area_yield_curves/html/index.en.html.

eqlist	<i>Earthquake list</i>
--------	------------------------

Description

This dataset contains a list of all earthquakes of magnitude greater than 6 between 1900 and 2024.

Usage

```
data(eqlist)
```

Format

eqlist is a data frame of 16 columns and 14,014 rows:

time A factor for the time the seismic event occurred.

latitude A numeric for the latitude of the event, in degrees (negative implies Southern Hemisphere).

longitude A numeric for the longitude of the event, in degrees (negative implies Western Hemisphere).

depth A numeric for the depth of the event, in kilometers.

mag A numeric for the magnitude of the event.

magType A factor for the method used to calculate the magnitude. For a full list of methods used, refer below.

nst An integer for the total number of seismic stations used to determine the location.

gap A numeric for the largest azimuthal gap between azimuthally adjacent stations, in degrees. In general, smaller gaps indicate better reliability in terms of the horizontal positioning of the event.

dmin A numeric for the horizontal distance between the epicenter of the event and the nearest station, in degrees. One degree is approximately 111.2 kilometers. In general, the smaller the distance, the more reliable is the calculated depth.

rms A numeric for root mean square travel time residual using all weights, in seconds. This measures the fit of the observed arrival times to the predicted arrival times for this location. Smaller numbers reflect a better fit of the data. The value is dependent on the accuracy of the method used to determine location, the quality weights assigned to the arrival time data, and the procedure used to locate the event.

net A factor for the identification number of the information source.

id A factor for the identification number of the event.

updated A factor for the last update.

place A factor for the location of the event, such as the name of the city or island.

type A factor for the type of seismic event: either "earthquake", "explosion", "nuclear explosion".

day A date for the day of the week.

Details

Below are the descriptions of the methods used to calculate the magnitude of seismic events, magType. See <https://www.usgs.gov/programs/earthquake-hazards/magnitude-types> for further technical details:

mb Short-period body wave magnitude.

mj Richter magnitude.

ms Surface wave magnitude.

Mt Duration magnitude.

mw, Mw, mww, Mww Moment magnitude.

mwb, Mwb Body wave magnitude.

mwc, Mwc Centroid magnitude.

mwp, Mwp Integrated p-wave magnitude.

mwr Regional magnitude.

uk Unknown method.

Source

Earthquake worldwide archive : <https://earthquake.usgs.gov/earthquakes/search/>.

Methods to calculate the magnitude of seismic events (magType) : <https://www.usgs.gov/programs/earthquake-hazards/magnitude-types>.

References

Young, J.B., Presgrave, B.W., Aichele, H., Wiens, D.A. and Flinn, E.A. (1996), *The Flinn-Engdahl Regionalisation Scheme: the 1995 revision*, Physics of the Earth and Planetary Interiors, v. 96, p. 223-297, doi:10.1016/00319201(96)03141X.

Flinn, E.A., Engdahl, E.R. and Hill, A.R. (1974), *Seismic and geographical regionalization*, Bulletin of the Seismological Society of America, vol. 64, p. 771-993, doi:10.1785/BSSA064320771.

Flinn, E.A., and Engdahl, E.R. (1965), *A proposed basis for geographical and seismic regionalization*, Reviews of Geophysics, vol. 3, p. 123-149, doi:10.1029/RG003i001p00123.

See Also

Northern California earthquake archive : <https://earthquaketrack.com/v/norcal/recent>

Examples

```
# (1) load of data
#
data(eqlist)
dim(eqlist)

# (2) simple plot
plot(eqlist$day[eqlist$mag > 6.5], eqlist$mag[eqlist$mag > 6.5], pch=".",
      xlab="Year", ylab="Magnitude", main="Earthquake above 6.5 mag (worldwide)")

# (3) most active place
table(eqlist$place)[table(eqlist$place) > 100]
```

eudirectlapse

European lapse dataset from the direct channel

Description

The eudirectlapse dataset is based on one-year vehicle insurance renewal quotes for an unknown year and an unknown insurer. There are 23,060 policies.

Usage

```
data(eudirectlapse)
```

Format

eudirectlapse is a data frame of 19 columns and 23,060 rows:

`lapse` A binary variable indicating the lapse of the customer.

`polholder_age` The age of the policyholder.

`polholder_BMCevol` The evolution of bonus/malus coefficient (BMC) of the policyholder: 3 categorical values ("down" when bonus increases, "stable" when coefficient does not change, "up" when malus increases).

`polholder_diffdriver` The difference status between the policyholder and the driver.

`polholder_gender` The gender of the policyholder.

`polholder_job` The job of the policyholder: either "medical" or "normal".

`policy_age` The age of the policy.

`policy_caruse` The car usage.

`policy_nbcontract` The number of policies given policyholder for this insurer.

`prem_final` The final renewal premium value proposed to policyholder.

`prem_freqperyear` The premium frequency per year.

`prem_last` The premium paid by the policyholder for the last insurance coverage.

`prem_market` A proxy of the market premium.

`prem_pure` The technical premium value.

`vehicl_age` The vehicle age.

`vehicl_agepurchase` The vehicle age at purchase.

`vehicl_garage` The garage type (categorical values).

`vehicl_powerkw` The horsepower of the car (categorical values).

`vehicl_region` The living region of policyholder (unknown category).

Source

Unknown non-life insurers from European Union.

Examples

```
# (1) load of data
#
data(eudirectlapse)
head(eudirectlapse)
```

euhealthinsurance	<i>European health insurance</i>
-------------------	----------------------------------

Description

The euhealthinsurance compiles data coming from a health group collective fund that covers different kind of health perils to the members. Available data are: gender, age at inception of coverage, role in the policy, number and aggregate amount.

Usage

```
data(euhealthinsurance)
```

Format

euhealthinsurance is a dataframe with 157221 observations and 21 columns

id_ano anonymized id.

relation role in the policy.

gender gender: M, F.

policy_years cumulated exposure.

age_at_inception attained age when the policy started.

num_analysis number of laboratory test covered.

num_dentistry number of dental health services covered.

num_diagnostics number of exams covered.

num_endoscopy number of endoscopies covered.

num_hospitalizations number of hospitalizations covered.

num_mammography number of mammographies covered.

num_operations number of surgeries covered.

num_visits number of specialist visits covered.

amt_analysis amount of laboratory test covered.

amt_dentistry amount of dental health services covered.

amt_diagnostics amount of exams covered.

amt_endoscopy amount of endoscopies covered.

amt_hospitalizations amount of hospitalizations covered.

amt_mammography amount of mammographies covered.

amt_operations amount of surgeries covered.

amt_visits amount of specialist visits covered.

Source

Unknown non-life insurers from European Union.

Examples

```
# (1) load of data
#
data(euhealthinsurance)
head(euhealthinsurance)
```

euMTPL

*European motor TPL insurance***Description**

The euMTPL compiles three years of experience from a European MTPL (Motor Third Party Liability) portfolio, including frequency and severity values for different types of losses. The data was collected during the first decade of the 21st century.

Usage

```
data(euMTPL)
```

Format

euMTPL is a data frame with 2,373,197 rows and 19 columns:

policy_id Unique identifier for each policy.
 year Calendar year of the policy.
 group Data split into training, validation, and test sets using a 70/10/20 ratio.
 fuel_type Fuel type of the insured vehicle.
 vehicle_category Category of the insured vehicle.
 vehicle_use Intended use of the vehicle (e.g., personal, commercial).
 province Province of residence of the policyholder.
 horsepower Power output of the insured vehicle, measured in horsepower.
 gender Gender of the policyholder.
 age Age of the policyholder at the start date of the policy.
 exposure Fraction of the year that the policy was in effect.
 cost_nc Total claim amount for No Card (NC) claims.
 num_nc Number of No Card (NC) claims.
 cost_cg Total claim amount for Card Gestionario (CG) claims.
 num_cg Number of Card Gestionario (CG) claims.
 cost_cd Total claim amount for Card Debitore (CD) claims.
 num_cd Number of Card Debitore (CD) claims.
 cost_fcd Total claim amount for Forfait Card Gestionario (FCD) claims.
 num_fcd Number of Forfait Card Gestionario (FCD) claims.

Source

Unknown non-life insurers from European Union.

Examples

```
# (1) load of data
#
data(euMTPL)
head(euMTPL)
```

eusavingsurrender	<i>European surrender dataset from the direct channel</i>
-------------------	---

Description

The eusavingULnoPS dataset is based on unit-linked saving products with no profit sharing sold in an unknown European country. Those insurance policies are observed between 1999 and 2008: entries and exits are possible. eusavingULnoPSperYr/perQtr/perMth are repeated version per year, per quarter or per month of eusavingULnoPS such that a policy is repeated per time interval as long as it stays in-force.

Usage

```
data(eusavingULnoPSperYr)
data(eusavingULnoPSperQtr)
data(eusavingULnoPSperMth)
data(eusavingULnoPS)
```

Format

eusavingULnoPS/perYr/perQtr/perMth are data frames of 30 columns:

policy.ID A character for the policy identification number.

issue.date, termination.date Issue and termination dates.

lapse.reason A character for the lapse reason.

premium.frequency A character for the premium frequency.

gender A character for the gender.

underwriting.age A character for the underwriting age.

face.amount A numeric for the underwriting face amount.

risk.premium A numeric for the underwriting risk premium.

saving.premium A numeric for the underwriting saving premium.

CPI.relvar1mth,CPI.relvar1qtr,CPI.relvar1yr,CPI.relvar2yr The relative variation of Consumer Price Index over a month, a quarter, a year or two years.

EUidx.relvar1mth,EUidx.relvar1qtr,EUidx.relvar1yr,EUidx.relvar2yr The relative variation of an European stock index over a month, a quarter, a year or two years.

rate1Y.relvar1mth,rate1Y.relvar1qtr The relative variation of one-year interest rate over a month, a quarter.

rate2Y.relvar1mth,rate2Y.relvar1qtr The relative variation of two-year interest rate over a month, a quarter.

rate10Y.relvar1mth,rate10Y.relvar1qtr The relative variation of ten-year interest rate over a month, a quarter.

unemploy.relvar1mth,unemploy.relvar1qtr The relative variation of an European unemployment rate over a month, a quarter.

industry.relvar1mth,industry.relvar1qtr The relative variation of an European industry index over a month, a quarter.

RTV.relvar1mth,RTV.relvar1qtr The relative variation of an European retail trade volume index over a month, a quarter.

Source

Unknown life insurers from European Union.

Examples

```
# (1) load of data
#
data(eusavingULnoPS)
head(eusavingULnoPS)
```

eutelematic13

European telematic insurance dataset

Description

This dataset is based on a real dataset acquired from a European-based insurer. The observation period was for the year 2013, with over 766,746 policies being observed, for which the dataset drawn is pre-engineered for training a statistical model for predictive purposes.

Usage

```
data(eutelematic13)
```

Format

eutelematic13 is a data frame of 17 columns and 1,999,028 rows. Column names have been anonymized. Continuous variables have been standardized and categorical variables have anonymized values.

identifier_int Identification code of the policyholder. Each policyholder may appear for one or more calendar period.

period Month of the year

dt_1 Device type code.

exposure_1, exposure_2 Two different exposure measures.

num_target_1, num_target_2 Two measures obtained from telematic events.

sev_target_1, sev_target_2 Corresponding median intentions.

vh1_1, vh1_2, vh1_3 Vehicle-related covariates.

trt_1, trt_2 Territory-related covariates.

ph1_1, ph1_2 Policyholder-related covariates.

group Data group for the application : 70% train, 20% test, 10% valid.

Source

Unknown

Examples

```
# (1) load of data
#
data(eutelematic13)
dim(eutelematic13)
```

FedYieldCurve

Federal Reserve interest rates

Description

The data-set contains the interest rates of the Federal Reserve, from January 1982 to December 2012. The interest rates are Market yield on U.S. Treasury securities constant maturity (CMT) (more information on the Treasury yield curve can be found at the following website <https://home.treasury.gov/>) at different maturities (3 months, 6 months, 1 year, 2 years, 3 years, 5 years, 7 years and 10 years), quoted on investment basis and have been gathered with monthly frequency.

Usage

```
data(FedYieldCurve)
```

Format

An object with class attributes xts.

Source

FED: <https://www.federalreserve.gov/datadownload/Build.aspx?rel=H15>.

forexUSUK

Foreign exchange rate between USD and GBP

Description

The dataset is the daily buying rates in New York City for cable transfers payable in foreign currencies from January 4, 1971 to March 1, 2013. The data can be downloaded from the FRED website. Access to this website was done on March 6, 2012.

Usage

```
data(forexUSUK)
```

Format

forexUSUK is a data frame of 2 columns and 10,583 rows:

Date Date.

Value The index value.

Source

FRED, Federal Reserve Economic Data, Federal Reserve Bank of St. Louis: U.S. - U.K. Foreign Exchange Rate (DEXUSUK): <https://fred.stlouisfed.org/series/DEXUSUK>.

References

Bollerslev (1987). *Regression Modeling with Actuarial and Financial Applications*, Cambridge University Press.

Examples

```
# (1) load of data
#
data(forexUSUK)
dim(forexUSUK)
head(forexUSUK)

# (2) plot of data
#

forexUSUK <- forexUSUK[forexUSUK$Date >= "2012-01-01", ]
plot(forexUSUK$Date, forexUSUK$Value, main = "US/UK FX Rate",
     xlab = "Year", ylab = "Index", type = "l")
```

fre4LoBtriangles

French private motor, third party liability or disability triangles

Description

Datasets `fretri1auto9605`, `fretri2auto9605` and `fretri3auto9605` contain claim triangles of motor policies from a French non-life insurer between 1996 and 2005. Dataset `fretri4auto9403` contains a claim triangle of motor policies from a French non-life insurer between 1994 and 2003. Note that the accident year 1994 corresponds to all of years before 1994 (included).

For each dataset, the variable `fretri1auto---` is a list of 3 elements for the damage guarantee, the body guarantee and the total. Each element is also a list of two elements with paid claims and incurred claim amounts. Note that claim amounts are cumulated.

Dataset `fretri1TPL9207` contains a claim triangle of third party liability policies from a French non-life insurer between 1992 and 2007. For `fretri1TPL9207`, only paid cumulative claim amounts are available.

Dataset `fretri1dis1418` contains a claim triangle of disability insurance policies from a French non-life insurer between 2014 and 2018. For `fretri1dis1418`, only cumulative claim numbers are available on a monthly basis.

Usage

```
#1st Line of Business  
data(fretri1auto9605)
```

```
#2nd Line of Business  
data(fretri2auto9605)
```

```
#3rd Line of Business  
data(fretri3auto9605)
```

```
#4th Line of Business  
data(fretri4auto9403)
```

```
#5th Line of Business  
data(fretri1TPL9207)
```

```
#6th Line of Business  
data(fretri1dis1418)
```

Format

`fretriX---YZZZ` contains the insurance triangle for Xth line of business from year YY to year ZZ.

Source

Unknown private insurers

Examples

```
# (1) load of data

#1st Line of Business
data(fretri1auto9605)

#2nd Line of Business
data(fretri2auto9605)

#3rd Line of Business
data(fretri3auto9605)

#4th Line of Business
data(fretri4auto9403)

#5th Line of Business
data(fretri1TPL9207)

#6th Line of Business
data(fretri1dis1418)
```

freaggnumber

French aggregate claim numbers

Description

The dataset consists of 12513 classes for which we have the driver age, the age of driving licence, the vehicle age, the exposure and the claim number.

Usage

```
data(freaggnumber)
```

Format

danishuni contains 5 columns:

DriverAge The driver age.

LicenceAge The age at which the driver gets its driving licence.

VehAge The vehicle age.

Exposure The exposure (in policy-year).

ClaimNumber The claim number for that group.

Examples

```
# (1) load of data
#
data(freaggnumber)
dim(freaggnumber)

# (2) ecdf plot
```

```
#  
summary(freaggnumber$ClaimNumber / freaggnumber$Exposure)
```

fribiloss	<i>French business interruption losses</i>
-----------	--

Description

The univariate dataset was collected at FFSA and comprise 2387 business interruption losses over the period 1985 to 2000 (for losses above 100,000 French Francs).

Usage

```
data(fribiloss)
```

Format

danishuni contains 8 columns:

Year The year of claim occurrence.

OccurDate The day of claim occurrence.

PolicyID The policy identification number.

ClaimID The claim identification number.

ClaimCost Original claim cost in French Francs (FFR).

TotalCost Original total cost (claim+expense) in French Francs.

ClaimCost2007 Normed claim cost in thousand of 2007 euros (EUR).

TotalCost2007 Normed total cost in thousand of 2007 euros (EUR).

Source

FFSA

References

Dataset used in Zajdenweber (1996). *Extreme values in business interruption insurance*, Journal of Risk and Insurance, 1, 95-110, [doi:10.2307/253518](https://doi.org/10.2307/253518).

Examples

```
# (1) load of data  
#  
data(fribiloss)  
dim(fribiloss)  
  
# (2) ecdf plot  
#  
plot(ecdf(fribiloss$ClaimCost2007), log="x", xlim=c(10^1, 10^5))  
  
boxplot(ClaimCost2007~Year, data=fribiloss, log="y")
```

 freclaimset

Two datasets of French claim settlements

Description

The dataset `freclaimset` consists of 2306 claims settlements between 1996 and 2006.

The dataset `freclaimset2motor` consists of claims settlements of the damage guarantee of a French insurer for motor insurance between 1995 and 2014. 1,012,839 records for 735,079 claims are listed in the dataset in conjunction with some aggregated data (exposure, GWP, claim number) per occurrence year.

Usage

```
data(freclaimset)
data(freclaimset2motor)
```

Format

`freclaimset` contains 6 columns:

`PaymentDate` The payment date.
`Payment` The amount of money paid.
`FbFprov` The file-by-file provision.
`Risk` The risk category.
`Subrisk` The sub-category.
`Type` The risk type.

`freclaimset2motor` is a list of two components. `freclaimset2motor$claimset` contains 8 columns:

`ClaimID` The identification number of the claim, first four characters are the occurrence year.
`OccurYear` The occurrence year.
`ManagYear` The management year.
`ClaimStatus` A character string for the claim status.
`PaidAmount` The cumulative paid amount for the claim (euro).
`RecourseAmount` The cumulative paid recourse for the claim (euro).
`ExpectCharge` The expected amount for the claim (euro).
`ExpectRecourse` The expected recourse for the claim (euro).

`freclaimset2motor$claimset` contains 4 columns:

`Year` The management year.
`Exposure` The sum of insurance years of the portfolio.
`GWP` The gross written premium (in euro).
`ClaimNb` The Claim Number.

Source

Unknown private insurer

Examples

```

# (1) load of data
#
data(freclaimset)
dim(freclaimset)
data(freclaimset2motor)
dim(freclaimset2motor)

# (2) consistency check (should be the same)
#

somerow <- freclaimset2motor$claimset$OccurYear == freclaimset2motor$claimset$ManagYear

cbind(
  freclaimset2motor$aggdata$ClaimNb,
  table(freclaimset2motor$claimset[somerow, "OccurYear"])
)

# (3) some examples of claims
#

subset(freclaimset2motor$claimset, ClaimID == "1995-000127")

subset(freclaimset2motor$claimset, ClaimID == "1996-008979")

```

 freclaimset9207

French individual claim settlements

Description

freclaimset3multi9207, freclaimset3fire9207 and freclaimset3dam9207 comes from the same dataset of 282,000 claims of property and casualty policies of a French unknown insurer for commercial insurance between 1992 and 2007.

freclaimset3fire9207 and freclaimset3dam9207 consist of randomized claims settlements of the fire/damage guarantees only. 58,056 claims are listed in the dataset for which both paid and incurred (F/F) amounts (EUR) are available.

freclaimset3multi9207 contains aggregate claim amounts by guarantee type and period of some property-casualty commercial lines in France between 1992 and 2007. A 3-day period has been used to perform the aggregation process, see variable Occur, the first day of occurrence period. The guarantee type is structured as

- HSS=Hail, storm, snow: claims from natural disaster: hail, storm, snow, generally known as Tempete-Grele-Neige in France.
- TPL=Third-part liability: claims from third-part liabilities (both material and bodily injuries).
- Other=Other guarantees: other claims, e.g. legal protection, business interruption.
- Damage=Material damage: claims from material damages, e.g. machine breaks or waterleaks.
- Fire: claims related to fire guarantees, both building and vehicles.
- Thief: thieves of insured goods, mostly non-vehicle.

The resulted dataset contains 1,944 rows with claim variables named XY_Claim for guarantee XY. These guarantee groups are described by 5 categorical explanatory variables

- Employee: The aggregate employee number.
- Sites: The aggregate site number.
- Area: The insured area of buildings.
- Revenue: The aggregate revenue of companies.
- Goods: A proxy for the aggregate insured values of goods.

Explanatory variables are named on the same principle as claim amount. The resulted dataset contains 37 variables.

Usage

```
data(freclaimset3fire9207)
data(freclaimset3dam9207)
data(freclaimset3multi9207)
```

Format

freclaimset3fire9207 and freclaimset3dam9207 are data frames with 37 columns:

NbEmployee The category of employee number.

NbSite The category of site number.

Surface The insured surface.

RiskCateg An unknown risk category.

inc_Y15-inc_Y0 inc_Yj is the incurred amount of the claim at the end of year 2007-j, i.e. inc_Y0 is the latest estimate and inc_Y15 is the oldest estimate.

paid_Y15-paid_Y0 paid_Yj is the paid amount of the claim at the end of year 2007-j, i.e. paid_Y0 is the latest estimate and paid_Y15 is the oldest estimate.

OccurDate The occurrence date. Note that paid_Yj/inc_Yj is never empty (i.e. NA) even if the claim did occur after the year 2007-j.

freclaimset3multi9207 contains aggregate claim amounts by guarantee type and period of some property-casualty commercial lines in France between 1992 and 2007. A 3-day period has been used to perform the aggregation process, see variable Occur, the first day of occurrence period. The guarantee type is structured as

- HSS=Hail, storm, snow: claims from natural disaster: hail, storm, snow, generally known as Tempete-Grele-Neige in France.
- TPL=Third-part liability: claims from third-part liabilities (both material and bodily injuries).
- Other=Other guarantees: other claims, e.g. legal protection, business interruption.
- Damage=Material damage: claims from material damages, e.g. machine breaks or waterleaks.
- Fire: claims related to fire guarantees, both building and vehicles.
- Thief: thieves of insured goods, mostly non-vehicle.

The resulted dataset contains 1,944 rows with claim variables named XY_Claim for guarantee XY. These guarantee groups are described by 5 categorical explanatory variables

- Employee: The aggregate employee number.

- Sites: The aggregate site number.
- Area: The insured area of buildings.
- Revenue: The aggregate revenue of companies.
- Goods: A proxy for the aggregate insured values of goods.

Explanatory variables are named on the same principle as claim amount. The resulted dataset contains 37 variables.

Source

Unknown private insurer.

Examples

```
# (1) load of data
#
data(freclaimset3fire9207)
data(freclaimset3dam9207)
data(freclaimset3multi9207)

# (2) some examples of claims
#

head(freclaimset3fire9207)
tail(freclaimset3fire9207)

# (3) graph
#
par(mar=c(7,3,2,1))
boxplot(freclaimset3multi9207[, grep("Claim", colnames(freclaimset3multi9207))], log="y",
        las=3)
grid()

par(mar=c(4,4,2,1))
plot(freclaimset3multi9207$Occur, freclaimset3multi9207$HSS_Claim/1e6, type = "h",
     xlab="Occurrence date", ylab="Claim amount (million of euros)")
grid()
```

frecomfire

French commercial fire losses

Description

The univariate dataset was collected at FFSA and comprise 9613 commercial fire losses over the period 1982 to 1996.

Usage

```
data(frecomfire)
```

Format

frecomfire contains 4 columns:

Year The year of claim occurrence.

OccurDate The day of claim occurrence.

ClaimCost Original claim cost in French Francs (FFR).

ClaimCost2007 Normed claim cost in thousand of 2007 euros (EUR).

Source

Fédération Française des Sociétés d'Assurance

Examples

```
# (1) load of data
#
data(frecomfire)
dim(frecomfire)
```

freDemography

Some datasets relative to the French demography

Description

freRegion7526 and freRegion0ld7526 contains the total population estimate of France (mainland) by region between 1975 and 2026. One row contains estimates by year for a region. freRegion7526 uses the new region system from 2016 and freRegion0ld7526 uses the old region system used between 1982 and 2016, see https://en.wikipedia.org/wiki/Regions_of_France.

Usage

```
data(freRegion7526)
data(freRegion0ld7526)
```

Format

freRegion7526 contains 52 columns and 14 rows. freRegion0ld7526 contains 52 columns and 23 rows.

Source

<https://www.insee.fr/fr/statistiques/8721456>

Examples

```
# (1) load of data
#
data(freRegion7526)
matplot(1975:2026, t(freRegion7526[1:6, ]), type="l", log="y", xlab="year",
        ylab="population", main="French population by region")
legend("left", rownames(freRegion7526)[1:6], col=1:6, lty=1:5, bty="n")
```

 freDisTables

French Disability Tables and Probabilities

Description

Naming convention: X2Y stands for going from state X to state Y, where possible states are T (temporary disability), P (permanent disability), D (death). For instance, T2T stands for temporary to temporary disability.

Tables `freP2Pdis10`, `freT2Tdis10` and `freT2Pdis10` have been established by the French mutual (BCAC) under a mission mandated by the French association of insurance companies (FFSA) and imposed by the new retirement reglementation after an agreement of professional federations. These tables have been build in 1993 and extended to the age 62 in 2010 by the December 24 act in 2010, cf. JO (2010).

These tables have been entirely rebuilt in 2013 by BCAC: the new imposed tables are Tables `freP2Pdis13`, `freT2Tdis13` and `freT2Pdis13`, see Bagui (2013).

`freP2Pdis10/freP2Pdis13` contain the continuation table of permanent disability (so-called invalidity in France) based on a 10,000-person reference population for all age between 20 and 61 (resp. between 20 and 64). `freT2Tdis10/freT2Tdis13` contain the continuation table of temporary disability (so-called incapacity in France) based on a 10,000-person reference population for all age between 20 and 66. (resp. between 21 and 65). `freT2Pdis10/freT2Pdis13` contain the transition table (from temporary to permanent disability) based on a 10,000-person reference population for all age between 20 and 61 (resp. between 21 and 62). Note that in France temporary disability is limited to 36 months (irrespective of the entry age) and permanent disability age is capped at the age of retirement 62 for 2010 tables (resp. 65 for 2013 tables).

`freT2Pdisprob10/freT2Pdisprob13`, `freT2Tdisprob10/freT2Tdisprob13`, `freP2Pdisprob10/freP2Pdisprob13` are the corresponding probabilities deduced from the tables, respectively to go from temporary to permanent disability, to stay temporarily disabled and to stay permanently disabled, given the entry age and the number of month or years already disabled.

Tables `freT2Ddis10`, `freP2Ddis10` have been established by the French mutual (BCAC) under a mission mandated by the French association of insurance companies (FFSA) and imposed by the new retirement reglementation after an agreement of professional federations.

The `freP2Ddis10` contains the mortality table of permanent disability (so-called invalidity in France) based on a 10,000-person reference population for all age between 25 and 64. The `freT2Ddis10` contains the mortality table of temporary disability (so-called incapacity in France) based on a 10,000-person reference population for all age between 25 and 65.

`freP2Ddisprob10`, `freT2Ddisprob10` are the corresponding probabilities deduced from the tables, respectively to die from temporary disability, to die from permanent disability, given the entry age and the number of month or years already disabled.

Usage

```

data(freP2Pdis10)
data(freT2Tdis10)
data(freT2Pdis10)

data(freP2Pdisprob10)
data(freT2Tdisprob10)
data(freT2Pdisprob10)

data(freT2Ddis10)
data(freP2Ddis10)

data(freT2Ddisprob10)
data(freP2Ddisprob10)

data(freP2Pdis13)
data(freT2Tdis13)
data(freT2Pdis13)

data(freP2Pdisprob13)
data(freT2Tdisprob13)
data(freT2Pdisprob13)

```

Format

freP2Pdis10/freP2Pdis13 contains 44 (resp. 47) columns:

EntryAge The entry age in permanent disability.

NbYrSpent0,...,NbYrSpent42/NbYrSpent45 The number of people (among 10,000) who spent a certain number of years (0 to 42/45) in permanent disability.

freP2Pdisprob10/freP2Pdisprob13 contains the probabilities to stay permanently disabled given the number of years spent in such a state.

freT2Tdis10/freT2Tdis13 contains 38 columns:

EntryAge The entry age in permanent disability.

NbMthSpent0...NbMthSpent36 The number of people (among 10,000) who spent a certain number of months (0 to 36) in temporary disability.

freT2Tdisprob10/freT2Tdisprob13 contains in 36 columns from NbMthSpent0 to NbMthSpent35 the probabilities to stay temporarily disabled given the number of months spent in such a state.

freT2Pdis10/freT2Pdis13 contains 37 columns:

EntryAge The entry age in permanent disability.

NbMthSpent0...NbMthSpent35 Transition probably from temporary to permanent disability after a certain number of months (0 to 35) spent in temporary disability.

freT2Pdisprob10/freT2Pdisprob13 contains in 36 columns from NbMthSpent0 to NbMthSpent35 the probabilities to become permanently disabled given the number of months spent in temporary disability.

freT2Ddis10 contains 37 columns:

EntryAge The entry age in permanent disability.

NbMthSpent0...NbMthSpent35 The number of people (among 10,000) who spent a certain number of months (0 to 35) in temporary disability.

freT2Ddisprob10 contains in 36 columns the probabilities to die given the number of months spent in temporary disability.

freP2Ddis10 contains 37 columns:

EntryAge The entry age in permanent disability.

NbYrSpent0...NbYrSpent35 The number of people (among 10,000) who spent a certain number of years (0 to 35) in permanent disability.

freP2Ddisprob10 contains in 36 columns the probabilities to die given the number of years spent in permanent disability.

Source

[RessourcesActuarielles](#)

References

(all ref. in French)

Bagui (2013), *Refonte des loi de maintien en incapacite temporaire de travail*, ISFA actuary memoir.

JO (2010), *Arrete du 24 decembre 2010 fixant les regles de provisionnement des garanties d'incapacite de travail, d'invalidite et de deces*, Journal Officiel, Texte 55 sur 138, 30 decembre 2010.

FFSA (2005), *Demande de donnees relatives aux populations d'assures*, Document de travail FFSA.

Planchet (2005), *Tables de mortalite d'experience pour des portefeuilles de rentiers*, Note methodologique de l'Institut des Actuares.

Planchet (2006), *Construction des tables de mortalite d'experience pour les portefeuilles de rentiers - presentation de la methode de construction*, Note methodologique de l'Institut des Actuares.

Serant (2005), *Construction de tables prospectives de mortalite*, Document interne FFSA (confidential).

Tassin (2006), *Note qualitative sur les tables prospectives IA 2006 masculines et feminines*, Document interne de l'Institut des Actuares.

Examples

```
# (1) load of data
#
data(freP2Pdis10)
data(freT2Tdis10)
data(freT2Pdis10)

data(freP2Pdisprob10)
data(freT2Tdisprob10)
data(freT2Pdisprob10)

data(freT2Ddis10)
data(freP2Ddis10)
```

```
data(freT2Ddisprob10)
data(freP2Ddisprob10)
```

```
data(freP2Pdis13)
data(freT2Tdis13)
data(freT2Pdis13)
```

```
data(freP2Pdisprob13)
data(freT2Tdisprob13)
data(freT2Pdisprob13)
```

freIRSA

French IRSA claim amount convention

Description

freIRSA contains the fixed claim amount used in the French convention called convention d'indemnisation directe de l'assure et de recours entre societes d'assurance automobile which sets the claim amount used between insurers for small claims (under the claim cap).

Usage

```
data(freIRSA)
```

Format

freIRSA contains 3 columns:

year The year.

amount The fixed claim amount (without taxes).

cap The claim cap (without taxes).

Source

<https://www.flotauto.com/forfait-irsa-2022-20211122.html>

Examples

```
# (1) load of data
#
data(freIRSA)
freIRSA
```

fremarine

Some French marine losses

Description

The univariate dataset was collected by a French private insurer and comprise 1,274 marine losses between the January 2003 and June 2006. The status of the claim (settled or opened) is determined at the end of June 2006.

Usage

```
data(fremarine)
```

Format

fremarine contains 20 columns:

OccurDate The day of claim occurrence.

ReporDate The day of claim reporting.

ShipCateg The category of the insured ship (factor).

ShipBrand The brand of the insured ship (factor) (resampled).

ShipPower The power of the insured ship (factor).

ShipEngNb The engine number of the insured ship (factor).

ShipEngYear The engine year of the insured ship (factor).

ShipBuildYear The building year of the insured ship (factor).

ShipHull The hull of the insured ship (factor).

ShipLength The length of the insured ship (factor).

ShipTonnage The tonnage of the insured ship (factor).

InsuredValue The insured value of the insured ship (factor).

ClaimPaid The paid amount (thousands of EUR) of the claim (numeric) (rescaled and noisy).

ClaimCharge The charge amount (thousands of EUR) of the claim (numeric) (rescaled and noisy).

ClaimRecourse The recourse amount (thousands of EUR) of the claim (numeric) (rescaled and noisy).

ClaimStatus The status of the claim (factor).

ClaimCateg The category of the claim (unknown factor).

Deductible The deductible value (numeric) (rescaled and noisy).

Departement The corresponding French departement of the cityname (factor).

Source

Unknown private insurer

Examples

```
# (1) load of data
#
data(fremarine)
dim(fremarine)
```

Description

The frePM6064 (resp. frePF6064) table has been established on INSEE observations collected between 1960 and 1964 in the French male population (resp. the French female population).

The freTD7377 (resp. freTV7377) table has been established on INSEE observations collected between 1973 and 1977 in the French male population (resp. the French female population). The table was officially approved by the August 22 act in 1986 and applies to life insurance.

The freTD8890 (resp. freTV8890) table has been established on INSEE observations collected between 1988 and 1990 in the French male population (resp. the French female population). The table was officially approved by the April 27 act in 1993 and applies to life insurance.

The freTPRV93 table is extracted from the floor table for pricing life annuities. The table was officially approved by the July 28 act in 1993 and is based on the prospective table tracking mortalities for generations between 1887 and 1993 (full table for generation 1950), JO (1993).

The freTH0002 (resp. freTF0002) table has been established on INSEE observations collected in the French male population (resp. the French female population). The table was officially approved by the December 20 act in 2005 and applies to life insurance other than life annuities in conjunction with the table of age shifts freAS0002, JO (2005, 2006a, 2006b, 2006c).

The freTGH05 (resp. freTGF05) table has been established based on 19 portfolios (16 from FFSA and 3 from CTIP) in the French male population (resp. the French female population) between 1993 and 2005. The underlying prospective INSEE table has been built on the basis of mortality tables between 1962 and 2000. The table was officially approved by the August 1 act in 2006. The freTPG93full table has been built for comparison with TGH05 and TGF05.

freHMD is a subset of the French mortality dataset available on <https://mortality.org/> in order to build the mortality vignette. This dataset is not intended to be used directly. freHMD is an object of class "demogdata" provided by the package demography by Hyndman (2023).

Usage

```
data(frePM6064)
data(frePF6064)
```

```
data(freTD7377)
data(freTV7377)
```

```
data(freTD8890)
data(freTV8890)
```

```
data(freTPRV93)
data(freTPG93full)
```

```
data(freTF0002)
data(freTH0002)
data(freAS0002)
```

```
data(freTGH05)
data(freTGF05)
```

```
data(freHMD)
```

Format

frePM6064, frePF6064, freTD7377, freTV7377, freTD8890, freTV8890, freTPRV93, freTF0002 and freTH0002 contain 2 columns:

x The age x.

1x The number of people still alive at x among the initial 100,000 referenced people.

freAS0002 contains 6 columns:

LowerAgeF, LowerAgeH The lower bound of the age class.

UpperAgeF, UpperAgeH The upper bound of the age class.

ShiftF, ShiftH The value to shift.

freTGH05 and freTGF05 contain 107 columns:

x The age x.

1x1900, ..., 1x2005 The number of people still alive at x among the referenced people in year 1900 (etc.. 2005).

freTPG93full contains 95 columns:

x The age.

1x1900, ..., 1x1993 The number of people still alive at x among the referenced people in year 1900 (etc.. 1993).

Source

INSEE, JO, [RessourcesActuarielles](#)

References

FFSA (2005), *Demande de donnees relatives aux populations d'assures*, Document de travail FFSA.

IA (2006), *Notice d'utilisation des tables de mortalite TH0002 and TF0002*, Note methodologique de l'Institut des Actuaire.

JO (1986), *Arrete du 8 aout 1986*, Journal Officiel num 174, Texte 30, 22 aout 1986.

JO (1993), *Arrete du 28 juillet 1993*, Journal Officiel num 174, Texte 30, 30 juillet 1993.

JO (2005), *Arrete du 20 decembre 2005*, Journal Officiel num 302, Texte 40, 29 decembre 2005.

JO (2006a), *Arrete du 1 aout 2006*, Journal Officiel num 197, Texte 11, 26 aout 2006.

JO (2006b), *Arrete du 8 decembre 2006*, Journal Officiel num 302, Texte 93, 30 decembre 2006.

JO (2006c), *Arrete du 21 decembre 2006*, Journal Officiel num 9, Texte 31, 11 janvier 2007.

Planchet (2005), *Tables de mortalite d'experience pour des portefeuilles de rentiers*, Note methodologique de l'Institut des Actuaire.

Planchet (2006), *Construction des tables de mortalite d'experience pour les portefeuilles de rentiers - presentation de la methode de construction*, Note methodologique de l'Institut des Actuaire.

Serant (2005), *Construction de tables prospectives de mortalite*, Document interne FFSA (confidentiel).

Tassin (2006), *Note qualitative sur les tables prospectives IA 2006 masculines et feminines*, Document interne de l'Institut des Actuares.

R. Hyndman (2023), *demography: Forecasting Mortality, Fertility, Migration and Population Data*, R package, doi:10.32614/CRAN.package.demography.

Examples

```
# (1) load of data
#
data(frePM6064)
data(frePF6064)

data(freTD7377)
data(freTV7377)

data(freTD8890)
head(freTD8890)

data(freTV8890)
head(freTV8890)

data(freTPRV93)
head(freTPRV93)

data(freTF0002)
head(freTF0002)

data(freTH0002)
head(freTH0002)

data(freAS0002)
head(freAS0002)

data(freTGH05)
head(freTGH05)

data(freTGF05)
head(freTGF05)

data(freTPG93full)
head(freTPG93full)
```

fremotorclaim

French claims for private motor

Description

Datasets fremotor1freq0304a/b/c, fremotor1sev0304a/b/c, fremotor1prem0304a/b/c are nine datasets from the same database of an unknown private motor portfolio observed between January 2003 and December 2004, respectively claim frequency databases, claim severity databases and

premium databases. The last letter a, b or c distinguishes the random sampling for a given dataset series. Note that some records are common between resampling versions.

Datasets fremotor1freq0304a/b/c consist of 64,234 records with explanatory variables for policies (possibly with multiple vehicles insured under the same policy number). Datasets fremotor1prem0304a/b/c consist of 51,949 records of claim numbers (by policy) in 2003 and 2004. Datasets fremotor1sev0304a/b/c consist of 9,246 records of ClaimAmount, their occurrence date, the corresponding guarantee, in 2003 and 2004.

Datasets fremotor2sev9907, fremotor3sev9907, fremotor4sev9907, and fremotor2freq9907u, fremotor3freq9907u, fremotor4freq9907u, fremotor2freq9907b, fremotor3freq9907b, fremotor4freq9907b are claim severities and claim frequencies coming from the same database for a private motor portfolio observed between 1999 and 2007. For size reason, the database has been splitted into three parts fremotor2***9907, fremotor3***9907, fremotor4***9907. Furthermore, the claim frequencies are available on two different formats : longitudinal unbalanced data and longitudinal balanced data, respectively fremotor2freq9907u and fremotor2freq9907b. The policy number is only available for claim frequencies: it is impossible to match claim severities and claim frequencies.

Usage

```
data(fremotor1prem0304a)
data(fremotor1prem0304b)
data(fremotor1prem0304c)
```

```
data(fremotor1freq0304a)
data(fremotor1freq0304b)
data(fremotor1freq0304c)
```

```
data(fremotor1sev0304a)
data(fremotor1sev0304b)
data(fremotor1sev0304c)
```

```
data(fremotor2sev9907)
data(fremotor3sev9907)
data(fremotor4sev9907)
```

```
data(fremotor2freq9907u)
data(fremotor3freq9907u)
data(fremotor4freq9907u)
data(fremotor2freq9907b)
data(fremotor3freq9907b)
data(fremotor4freq9907b)
```

Format

fremotor1prem0304a/b/c contain 30 columns:

IDpol The policy ID (used to link with the claims dataset).

DrivAge The driver age, in years (in France, people can drive a car at 18).

DrivGender The gender (as "factor").

MaritalStatus The marital status (as "factor").

BonusMalus Bonus/malus, between 50 and 350: <100 means bonus, >100 means malus in France.

LicenceNb The licence number (at least one).
 PayFreq The payment frequency (as "factor").
 JobCode The job code (as "factor").
 VehAge The vehicle age, in years.
 VehClass The vehicle class (as "factor").
 VehPower The vehicle power (as "factor") from least powerful "P2" to most powerful car "P15".
 VehGas The car gas, Diesel or regular (as "factor").
 VehUsage The vehicle usage (as "factor").
 Garage The type of garage (as "factor").
 Area The area code (as "factor"): unknown category.
 Region The policy regions in France (based on a standard French classification).
 Channel The channel distribution code (as "factor"): unknown category.
 Marketing The marketing code (as "factor"): unknown category.
 PremWindscreen The premium for windscreen guarantee (annual basis).
 PremDamAll The premium for damage all-accident guarantee (annual basis).
 PremFire The premium for fire guarantee (annual basis).
 PremAcc1 The premium for type-1 accident guarantee (annual basis).
 PremAcc2 The premium for type-2 accident guarantee (annual basis).
 PremLegal The premium for legal protection guarantee (annual basis).
 PremTPLM The premium for mandatory third-part liability guarantee (annual basis).
 PremTPLV The premium for voluntary third-part liability guarantee (annual basis).
 PremServ The premium for service guarantee (annual basis).
 PremTheft The premium for theft guarantee (annual basis).
 PremTot The total premium (annual basis).
 Year Numeric for the year.

fremotor1freq0304a/b/c contain 6 columns:

IDpol The policy ID.
 Year The underwriting year.
 Damage The claim number for the Damage guarantee.
 Fire The claim number for the Fire guarantee.
 Other The claim number for the Other guarantee.
 Theft The claim number for the Theft guarantee.
 TPL The claim number for the TPL guarantee.
 Windscreen The claim number for the Windscreen guarantee.

fremotor1sev0304a/b/c contain 6 columns:

IDpol The policy ID.
 OccurDate The occurrence date.
 Payment The amount of money paid.
 IDclaim The claim ID.

Guarantee The corresponding guarantee of the claim.

fremotor2sev9907, fremotor3sev9907, fremotor4sev9907 contains 3 columns:

Year The occurrence year.

NbClaim The number of claims aggregated, mostly 1.

ClaimAmount The aggregate charge (i.e. expected claim amount).

fremotor2freq9907u, fremotor3freq9907u, fremotor4freq9907u contains 23 columns:

IDpol The policy ID.

Usage The usage (unknown category).

VehType The vehicle type (unknown category).

VehPower The vehicle power (unknown category).

NbYear The number of years under exposure.

NbClaimXXXX The number of claims for year XXXX.

ExpoXXXX The exposure for year XXXX.

fremotor2freq9907b, fremotor3freq9907b, fremotor4freq9907b contains 7 columns:

IDpol The policy ID.

Year The year.

NbClaim The number of claims.

Expo The exposure.

Usage The usage (unknown category).

VehType The vehicle type (unknown category).

VehPower The vehicle power (unknown category).

Source

Unknown private insurer

Examples

```
# (1) load of data
```

```
#
```

```
data(fremotor1prem0304a)
```

```
data(fremotor1prem0304b)
```

```
data(fremotor1prem0304c)
```

```
data(fremotor1freq0304a)
```

```
data(fremotor1freq0304b)
```

```
data(fremotor1freq0304c)
```

```
data(fremotor1sev0304a)
```

```
data(fremotor1sev0304b)
```

```
data(fremotor1sev0304c)
```

```
#gross written premium
```

```
tapply(fremotor1prem0304a$PremTot, fremotor1prem0304a$Year, sum)
```

```
# (1) load of data
#
data(fremotor2sev9907)
data(fremotor3sev9907)
data(fremotor4sev9907)

data(fremotor2freq9907u)
data(fremotor3freq9907u)
data(fremotor4freq9907u)
data(fremotor2freq9907b)
data(fremotor3freq9907b)
data(fremotor4freq9907b)
```

 freMPL

French Motor Personal Line datasets

Description

This collection of ten datasets comes from a private motor French insurer. Each dataset includes risk features, claim amount and claim history of around 30,000 policies for year 2004.

Usage

```
data(freMPL1)
data(freMPL1sub)
data(freMPL2)
data(freMPL3)
data(freMPL4)
data(freMPL5)
data(freMPL6)
data(freMPL7)
data(freMPL8)
data(freMPL9)
data(freMPL10)
```

Format

For this collection of dataset, possible variables are given below. freMPL1–10 contains claim severity and frequency information. The following tabular gives the list of variables by file. freMPL1sub is a subset of freMPL1 with exposure closed to 1: rownames of freMPL1sub are extracted rownames of freMPL1.

	freMPL1	freMPL2	freMPL3	freMPL4	freMPL5	freMPL6	freMPL7	freMPL8	f
Exposure	1	1	1	1	1	1	1	1	
LicAge	1	1	1	1	1	1	1	1	
RecordBeg	1	1	1	1	1	1	1	1	
RecordEnd	1	1	1	1	1	1	1	1	
VehAge	1	1	1	1	0	0	0	0	

Gender	1	1	1	1	1	1	1	1
MariStat	1	1	1	1	1	1	1	1
SocioCateg	1	1	1	1	1	1	1	1
VehUsage	1	1	1	1	1	1	1	1
DrivAge	1	1	1	1	1	1	1	1
HasKmLimit	1	1	1	1	1	1	1	1
BonusMalus	1	1	1	1	1	1	1	1
VehBody	1	1	1	1	0	0	0	0
VehPrice	1	1	1	1	0	0	0	0
VehEngine	1	1	1	1	0	0	0	0
VehEnergy	1	1	1	1	0	0	0	0
VehMaxSpeed	1	1	1	1	0	0	0	0
VehClass	1	1	1	1	0	0	0	0
ClaimAmount	1	1	1	1	1	1	1	1
RiskVar	1	1	1	1	0	0	0	0
Garage	1	1	1	1	0	0	0	0
ClaimInd	1	1	1	1	1	1	1	1
DeducType	0	0	1	1	0	0	0	0
ClaimNbResp	0	0	0	0	1	1	1	1
ClaimNbNonResp	0	0	0	0	1	1	1	1
ClaimNbParking	0	0	0	0	1	1	1	1
ClaimNbFireTheft	0	0	0	0	1	1	1	1
ClaimNbWindscreen	0	0	0	0	1	1	1	1
OutUseNb	0	0	0	0	1	1	1	1
RiskArea	0	0	0	0	1	1	1	1

The comprehensive list of the variables (over all datasets) is given below, yet no dataset contains all these variables.

Exposure The exposure, in years.

RecordBeg Beginning date of record.

RecordEnd End date of record.

DrivAge The driver age, in years (in France, people can drive a car at 18).

LicAge The driving licence age, in months.

Gender The gender, either "Male" or "Female".

MariStat The marital status, either "Alone" or "Other".

SocioCateg The social category known as CSP in France, between "CSP1" and "CSP99".

Garage The garage, if any, among "Collective garage", "None", "Private garage".

HasKmLimit A numeric, 1 if there is a km limit for the policy, 0 otherwise.

BonusMalus A numeric for the bonus/malus, between 50 and 350: <100 means bonus, >100 means malus in France.

VehAge The vehicle age, in years.

VehUsage The vehicle usage among "Private", "Private+trip to office" "Professional", "Professional run".

VehBody The vehicle body, among "bus", "cabriolet", "coupe", "microvan", "other microvan", "sedan", "sport utility vehicle", "station wagon", "van".

VehPrice The category of the vehicle price from "A" (cheapest) to "Z" (most expensive).

VehEngine The vehicle engine, among "carburation", "direct injection overpowered", "electric", "GPL", "injection", "injection overpowered".

VehEnergy The vehicle energy, among "diesel", "electric", "GPL", "regular".

VehMaxSpeed The VehMaxSpeed, among "1-130 km/h", "130-140 km/h", "140-150 km/h", "150-160 km/h", "160-170 km/h", "170-180 km/h", "180-190 km/h", "190-200 km/h", "200-220 km/h", "220+ km/h".

VehClass The vehicle class (unknown categories), among "0", "A", "B", "H", "M1", "M2".

RiskVar Unknown risk variable between 1 and 20, possibly ordered.

DeducType Deductible type, among "Majorized", "Normal", "Partially refunded", "Proportional", "Refunded".

RiskArea Unknown risk area between 1 and 13, possibly ordered.

ClaimNbResp Number of responsible claims in the 4 preceding years.

ClaimNbNonResp Number of non-responsible claims in the 4 preceding years.

ClaimNbParking Number of parking claims in the 4 preceding years.

ClaimNbFireTheft Number of fire-theft claims in the 4 preceding years.

ClaimNbWindscreen Number of windscreen claims in the 4 preceding years.

OutUseNb Number of out-of-use in the 4 preceding years.

ClaimAmount Total claim amount of the guarantee.

ClaimInd Claim indicator of the guarantee. (this is not the claim number)

Source

Unknown French private insurer.

See Also

For the vehicle body variable, see https://en.wikipedia.org/wiki/Car_classification

For the French bonus/malus, see <https://en.wikipedia.org/wiki/Bonus-malus>

For the French career categories, see https://fr.wikipedia.org/wiki/Professions_et_cat%C3%A9gories_socioprofessionnelles_en_France

Examples

```
# (1) load of data
#
data(freMPL1)

# (2) some graphics about the risk category
#
barplot(tapply(freMPL1$ClaimInd, freMPL1$RiskVar, mean), xlab="RiskVar", ylab="ClaimInd")
abline(h=mean(freMPL1$ClaimInd))

freMPL1pos <- subset(freMPL1, ClaimAmount>0)
barplot(tapply(freMPL1pos$ClaimAmount, freMPL1pos$RiskVar, mean),
        xlab="RiskVar", ylab="ClaimAmount")
abline(h=mean(freMPL1pos$ClaimAmount))

barplot(proportions(table(5*round(freMPL1$LicAge/60), freMPL1$RiskVar), 2), beside = FALSE,
        col=rainbow(17), xlab="RiskVar", main="LicAge (year)", xlim=c(1, 30))
```

```

legend("right", fill=rainbow(17), legend=sort(unique(5*round(freMPL1$LicAge/60))), cex=0.8)

barplot(proportions(table(round(freMPL1$BonusMalus, -1), freMPL1$RiskVar), 2), beside = FALSE,
        col=rainbow(16), xlab="RiskVar", main="BonusMalus", xlim=c(1, 30))
legend("right", fill=rainbow(16), legend=sort(unique(round(freMPL1$BonusMalus, -1))), cex=0.8)

barplot(proportions(table(substr(freMPL1$SocioCateg, 1, 4), freMPL1$RiskVar), 2), beside = FALSE,
        col=rainbow(8), xlab="RiskVar", main="Social professional category", xlim=c(1, 30))
legend("right", fill=rainbow(8), legend=sort(unique(substr(freMPL1$SocioCateg, 1, 4))))

barplot(proportions(table(freMPL1$VehAge, freMPL1$RiskVar), 2), beside = FALSE, col=rainbow(9),
        xlab="RiskVar", main="VehAge", ylim=c(0, 1), xlim=c(1, 30))
legend("right", fill=rainbow(9), legend=sort(unique(freMPL1$VehAge)))

barplot(proportions(table(freMPL1$VehBody, freMPL1$RiskVar), 2), beside = FALSE,
        col=rainbow(9), xlab="RiskVar", main="VehBody", ylim=0:1, xlim=c(1, 30))
legend("right", fill=rainbow(9), legend=sort(unique(freMPL1$VehBody)), cex=0.8)

barplot(proportions(table(freMPL1$VehPrice, freMPL1$RiskVar), 2), beside = FALSE,
        col=rainbow(27), xlab="RiskVar", main="VehPrice", xlim=c(1, 30))
legend("right", fill=rainbow(27), legend=sort(unique(freMPL1$VehPrice)), cex=0.8)

barplot(proportions(table(freMPL1$VehMaxSpeed, freMPL1$RiskVar), 2), beside = FALSE,
        col=rainbow(10), xlab="RiskVar", main="VehMaxSpeed", xlim=c(1, 30))
legend("right", fill=rainbow(10), legend=sort(unique(freMPL1$VehMaxSpeed)), cex=0.8)

barplot(proportions(table(freMPL1$VehUsage, freMPL1$RiskVar), 2), beside = FALSE,
        col=rainbow(4), xlab="RiskVar", main="VehUsage", xlim=c(1, 30))
legend("right", fill=rainbow(4), legend=sort(unique(freMPL1$VehUsage)), cex=0.8)

barplot(proportions(table(freMPL1$VehClass, freMPL1$RiskVar), 2), beside = FALSE,
        col=rainbow(6), xlab="RiskVar", main="VehClass", xlim=c(1, 27))
legend("right", fill=rainbow(6), legend=sort(unique(freMPL1$VehClass)), cex=0.8)

```

freMTPL

French Motor Third-Part Liability datasets

Description

In the two datasets `freMTPLfreq`, `freMTPLsev`, risk features are collected for 413,169 motor third-part liability policies (observed mostly on one year). In addition, we have claim numbers by policy as well as the corresponding claim amounts. `freMTPLfreq` contains the risk features and the claim number while `freMTPLsev` contains the claim amount and the corresponding policy ID. Some claim amounts of `freMTPLsev` are fixed claim amounts based on the French IRSA-IDA claim convention, see, e.g., [freIRSA](#) and <https://www.index-assurance.fr/pratique/sinistre/convention-irsa>.

In the two datasets `freMTPL2freq`, `freMTPL2sev`, risk features are collected for 677,991 motor third-part liability policies (observed mostly on one year). The period is approximately 2011-2013. In addition, we have claim numbers by policy as well as the corresponding claim amounts.

freMTPL2freq contains the risk features and the claim number while freMTPL2sev contains the claim amount and the corresponding policy ID. Some claim amounts of freMTPL2sev are fixed claim amounts based on the French IRSA-IDA claim convention, see, e.g., [freIRSA](#) and <https://www.index-assurance.fr/pratique/sinistre/convention-irsa>.

Usage

```
data(freMTPLfreq)
data(freMTPLsev)
```

```
data(freMTPL2freq)
data(freMTPL2sev)
```

Format

freMTPLfreq contains 10 columns:

PolicyID The policy ID (used to link with the claims dataset).

ClaimNb Number of claims during the exposure period.

Exposure The period of exposure for a policy, in years.

Power The power of the car (ordered categorical).

CarAge The vehicle age, in years.

DriverAge The driver age, in years (in France, people can drive a car at 18).

Brand The car brand divided in the following groups: A- Renault Nissan and Citroen, B- Volkswagen, Audi, Skoda and Seat, C- Opel, General Motors and Ford, D- Fiat, E- Mercedes Chrysler and BMW, F- Japanese (except Nissan) and Korean, G- other.

Gas The car gas, Diesel or regular.

Region The policy region in France (based on the 1970-2015 classification).

Density The density of inhabitants (number of inhabitants per km²) in the city the driver of the car lives in.

freMTPLsev contains 2 columns:

PolicyID The occurrence date (used to link with the contract dataset).

ClaimAmount The cost of the claim, seen as at a recent date.

freMTPL2freq contains 11 columns:

IDpol The policy ID (used to link with the claims dataset).

ClaimNb Number of claims during the exposure period.

Exposure The period of exposure for a policy, in years.

VehPower The power of the car (ordered values).

VehAge The vehicle age, in years.

DriverAge The driver age, in years (in France, people can drive a car at 18).

BonusMalus Bonus/malus, between 50 and 350: <100 means bonus, >100 means malus in France.

VehBrand The car brand (unknown categories).

VehGas The car gas, Diesel or regular.

Area The density value of the city community where the car driver lives in: from "A" for rural area to "F" for urban centre.

Density The density of inhabitants (number of inhabitants per square-kilometer) of the city where the car driver lives in.

Region The policy region in France (based on the 1970-2015 classification).

freMTPL2sev contains 2 columns:

IDpol The occurrence date (used to link with the contract dataset).

ClaimAmount The cost of the claim, seen as at a recent date.

Source

Unknown private insurer.

References

M. Denuit, A. Charpentier, and J. Trufin (2021), Autocalibration and tweedie-dominance for insurance pricing with machine learning, *Insurance: Mathematics and Economics*, **101**, 485–497, doi:[10.1016/j.insmatheco.2021.09.001](https://doi.org/10.1016/j.insmatheco.2021.09.001).

T. Miljkovic and D. Fernández (2018). On Two Mixture-Based Clustering Approaches Used in Modeling an Insurance Portfolio, *Risks*, **6**(2), 1–18, doi:[10.3390/risks6020057](https://doi.org/10.3390/risks6020057).

A. Noll, R. Salzmänn, and M. V. Wuthrich (2020), Case study: French motor third-party liability claims, *Innovation Practice eJournal*, doi:[10.2139/ssrn.3164764](https://doi.org/10.2139/ssrn.3164764).

N. Pocuca, P. Jevtic, P. D. McNicholas, and T. Miljkovic (2020), Modeling frequency and severity of claims with the zero-inflated generalized cluster-weighted models, *Insurance: Mathematics and Economics*, **94**, 79–93, doi:[10.1016/j.insmatheco.2020.06.004](https://doi.org/10.1016/j.insmatheco.2020.06.004).

Examples

```
# (1) load of data
#
data(freMTPLfreq)
dim(freMTPLfreq)

data(freMTPLsev)
dim(freMTPLsev)

# (2) check
#should be equal
sum(freMTPLsev$PolicyID %in% freMTPLfreq$PolicyID)
sum(freMTPLfreq$ClaimNb)

# (1) load of data
#
data(freMTPL2freq)
dim(freMTPL2freq)

data(freMTPL2sev)
dim(freMTPL2sev)
```

 freportfolio

Portfolio Mortality and Disability Tables

Description

The freprojqxINSEE table has been established on INSEE projection for the period 2007-2060 based a median scenario, cf. Blanpain and Chardon (2010), adjusted and selected for the purpose of the book.

The refictivetable represents a fictive portfolio of 87,090 individuals that enter in a healthy condition and have been observed between 1996-01-01 and 2007-12-31. The exit (that may occur before December 2007) is either "deceased" or "other".

The refictivetable2, refictivetable3 represents a fictive portfolio of 100,000 individuals that enter in a healthy condition and have been observed between December 1988 and December 1998. The exit is either "deceased" or "other" for censored observation.

The freptfpermdis and freptftempdis datasets comes from two portfolio of two French private companies (insurer or institute), respectively for permanent disability insurance and temporary disability insurance.

Usage

```
data(freprojqxINSEE)
data(refictivetable)
data(refictivetable2)
data(refictivetable3)
data(freptfpermdis)
data(freptftempdis)
```

Format

freprojqxINSEE is a data frame of 109 columns and 66 rows:

Age The age.

F2007,..., F2060 The 1-year female death probabilities

M2007,..., M2060 The 1-year male death probabilities

refictivetable is a data frame of 6 columns and 87,090 rows:

Id the identification number.

Gender the gender as "factor".

DateOfBirth the date of birth as "Date".

DateIn the entry date as "Date".

DateOut the exit date as "Date".

Status the status at exit : "deceased" (i.e. non-censored observation) or "other" (i.e. censored observation) as "factor".

refictivetable2, refictivetable3 are data frames of 5 columns and 100,000 rows:

DateIn the entry date as "Date".

DateOut the exit date as "Date".

Status the status at exit : "deceased" (i.e. non-censored observation) or "other" (i.e. censored observation) as "factor".

DateOfBirth the date of birth as "Date".

Gender the gender as "factor".

freptfpermdis is a data frame of 6 columns and 1,048,575 rows:

PolicyID the policy identification number.

BirthDate the date of birth.

Gender the sex: M for male and F for female.

EntryDate the entry date.

ExitDate the exit date.

ExitStatus the status at exit: "deceased" (i.e. non-censored observation) or "other" (i.e. censored observation).

freptftempdis is a data frame of 9 columns and 560,725 rows:

Gender the sex: M for male and F for female.

JobType the job category: "employee", "managers, engineers, sales responsables", "non-manager employee" "other 1", "other 2", "other 3", "other 4", "other 5", "technician", "unemployed workers".

UWType the underwriting type: either "specific policy in a collective agreement", "specific policy not linked to a collective agreement", "standard policy in a collective agreement" or "standard policy not linked to a collective agreement".

JobStopType the reason for disability: "illness", "work accident", "pregnancy" (for women only).

Birthdate the date of birth.

OccurDate the date of occurrence.

EntryDate the entry date.

ExitDate the exit date.

JobComebackType the status at exit: "recovered" (i.e. non-censored observation: the person goes back to work), "disabled" (i.e. non-censored observation: the person is permanently disabled) or "on-going" (i.e. censored observation).

Source

For freprojqxINSEE, Blanpain and Chardon (2010).

For refictivetable, Chapter 9 of *Computational Actuarial Science with R*, Ed. Arthur Charpentier, Chapman and Hall/CRC The R Series, 2014.

For freptfpermdis, freptftempdis, [RessourcesActuarielles](#)

References

Blanpain, N. and Chardon, O. (2010). *Projections de populations 2007-2060 pour la France metropolitaine: methode et principaux resultats*. Serie des Documents de Travail de la direction des statistiques Demographiques et Sociales F1008, INSEE.

Examples

```
# (1) load of data
#
data(freprojqxINSEE)
data(frefictivetable)

head(freprojqxINSEE)
head(frefictivetable)

# (2) load of data
#

data(frefictivetable2)
range(frefictivetable2$DateIn)
range(frefictivetable2$DateOut)

# (3) other
#

## Not run:
data(freptfpermdis)
data(freptftempdis)

head(freptfpermdis)
head(freptftempdis)

## End(Not run)
```

fretlematic

A French telematic dataset

Description

A real Telematic dataset of French motor TPL policies.

Usage

```
data(fretlematic)
```

Format

fretlematic contains 1177 observations with the following variables

- Policy_ID : a character for the policy ID.
- Total_Distance : a numeric for the total distance.
- Drive_Score : a numeric for the driving score.
- Time_Day : a numeric for the time day.
- Style_Score : a numeric for the styling score.
- Corner_Score : a numeric for the corner score.

- Acceleration_Score: a numeric for the acceleration score.
- Braking_Score : a numeric for the braking score.
- Total_Night_Time : a numeric for the total night driving time.
- Total_Time : a numeric for the total driving time.
- Acceleration : Factor w/ 2 levels "High", "Low".
- Brake : Factor w/ 2 levels "High", "Low".
- Corner : Factor w/ 2 levels "High", "Low".
- Insured_Gender : Factor w/ 2 levels "F", "M".
- Insured_Age : a numeric for the policyholder age.
- Claim : Factor w/ 2 levels indicating claim "no", "yes".

Examples

```
data(fretelematic)
```

fretplclaimnumber	<i>TPL claim number dataset</i>
-------------------	---------------------------------

Description

The univariate dataset was collected in the French motor market and comprise 678 013 one-year policies for which the claim number is recorded.

Usage

```
data(fretplclaimnumber)
```

Format

fretplclaimnumber contains three columns:

policy.id The policy identification number.

claim.number The claim number.

driver.age The driver age (given in the insurance contract).

Examples

```
# (1) load of data
#
data(fretplclaimnumber)

# (2) plot and description of data
#
table(fretplclaimnumber$claim.number)
```

`hurricanehist`*Hurricane history: Per Storm Maximum Wind Speeds (North Atlantic)*

Description

The dataset consists of 2010 observations for all tropical cyclones in the NHC best track record over the period 1899-2006. Each observation contains per cyclone maximum wind speeds and other relevant information.

Usage

```
data(hurricanehist)
```

Format

hurricanehist contains 7 columns:

Year The Year.

Region The region among "Basin", "East Florida", "Gulf", "US".

Windmax The maximum windspeed in knot (1kt = 0.51 m/s).

NAO the North Atlantic Oscillation (NAO) index as an indicator of storm steering.

SOI the Southern Oscillation Index (SOI) as an indicator of El Nino-Southern Oscillation.

SST the Atlantic sea-surface temperature (SST) as an indicator of cyclone energy.

SSTmda the SST mda.

Source

See http://myweb.fsu.edu/jelsner/_site/.

References

Dataset used in Jagger and Elsner (2008), *Modelling tropical cyclone intensity with quantile regression*, International Journal of Climatology 29, 1351 - 1361.

Examples

```
# (1) load of data
#
data(hurricanehist)
dim(hurricanehist)

# (2) box plot
#
boxplot(Windmax ~ Year, data=hurricanehist,
        ylim=c(35,175), subset=Year > 1939)
```

ICB

*Insurance Company Benchmark***Description**

This data set used in the CoIL 2000 Challenge contains information on customers of an insurance company. The data consists of 86 variables and includes product usage data and socio-demographic data derived from zip area codes.

The data was collected to answer the following question: Can you predict who would be interested in buying a caravan insurance policy and give an explanation why?

Usage

```
data(ICB1)
data(ICB2)
```

Format

ICB1 (resp. ICB2) is a data frame of 86 columns (resp. 85) and 5,822 rows (resp. 4,000). Each record consists of 86 (resp 85) variables, containing sociodemographic data (variables 1-43) and product ownership (variables 44-86). The sociodemographic data is derived from zip codes. All customers living in areas with the same zip code have the same sociodemographic attributes. Variable 86 (Purchase) indicates whether the customer purchased a caravan insurance policy. As ICB2 does not have the 86th column, ICB1 should be used for training purposes and ICB2 for testing purposes.

Columns are detailed below

MOSTYPE Customer Subtype see L0
 MAANTHUI Number of houses 1 - 10
 MGEMOMV Avg size household 1 - 6
 MGEMLEEF Avg age see L1
 MOSHOOFD Customer main type see L2
 MGODRK Roman catholic see L3
 MGODPR Protestant ...
 MGODOV Other religion
 MGODGE No religion
 MRELGE Married
 MRELSA Living together
 MRELOV Other relation
 MFALLEEN Singles
 MFGEKIND Household without children
 MFWEKIND Household with children
 MOPLHOOG High level education
 MOPLMIDD Medium level education
 MOPLLAAG Lower level education

MBERHOOG High status
MBERZELF Entrepreneur
MBERBOER Farmer
MBERMIDD Middle management
MBERARBG Skilled labourers
MBERARBO Unskilled labourers
MSKA Social class A
MSKB1 Social class B1
MSKB2 Social class B2
MSKC Social class C
MSKD Social class D
MHHUUR Rented house
MHK00P Home owners
MAUT1 1 car
MAUT2 2 cars
MAUT0 No car
MZFONDS National Health Service
MZPART Private health insurance
MINKM30 Income < 30.000
MINK3045 Income 30-45.000
MINK4575 Income 45-75.000
MINK7512 Income 75-122.000
MINK123M Income >123.000
MINKGEM Average income
MKOOPKLA Purchasing power class
PWAPART Contribution private third party insurance see L4
PWABEDR Contribution third party insurance (firms) ...
PWALAND Contribution third party insurance (agriculture)
PPERSAUT Contribution car policies
PBESAUT Contribution delivery van policies
PMOTSCO Contribution motorcycle/scooter policies
PVRAAUT Contribution lorry policies
PAANHANG Contribution trailer policies
PTRACTOR Contribution tractor policies
PWERKT Contribution agricultural machines policies
PBROM Contribution moped policies
PLEVEN Contribution life insurances
PPERSONG Contribution private accident insurance policies
PGEZONG Contribution family accidents insurance policies
PWAOREG Contribution disability insurance policies

PBRAND Contribution fire policies
 PZEILPL Contribution surfboard policies
 PPLEZIER Contribution boat policies
 PFIETS Contribution bicycle policies
 PINBOED Contribution property insurance policies
 PBYSTAND Contribution social security insurance policies
 AWAPART Number of private third party insurance 1 - 12
 AWABEDR Number of third party insurance (firms) ...
 AWALAND Number of third party insurance (agriculture)
 APERSAUT Number of car policies
 ABESAUT Number of delivery van policies
 AMOTSCO Number of motorcycle/scooter policies
 AVRAAUT Number of lorry policies
 AAANHANG Number of trailer policies
 ATRACTOR Number of tractor policies
 AWERKT Number of agricultural machines policies
 ABROM Number of moped policies
 ALEVEN Number of life insurances
 APERSONG Number of private accident insurance policies
 AGEZONG Number of family accidents insurance policies
 AWAOREG Number of disability insurance policies
 ABRAND Number of fire policies
 AZEILPL Number of surfboard policies
 APLEZIER Number of boat policies
 AFIETS Number of bicycle policies
 AINBOED Number of property insurance policies
 ABYSTAND Number of social security insurance policies
 CARAVAN Number of mobile home policies 0 - 1

L0 information: 1 High Income, expensive child, 2 Very Important Provincials, 3 High status seniors, 4 Affluent senior apartments, 5 Mixed seniors, 6 Career and childcare, 7 Dinkies (double income no kids), 8 Middle class families, 9 Modern, complete families, 10 Stable family, 11 Family starters, 12 Affluent young families, 13 Young all american family, 14 Junior cosmopolitan, 15 Senior cosmopolitans, 16 Students in apartments, 17 Fresh masters in the city, 18 Single youth, 19 Suburban youth, 20 Ethnically diverse, 21 Young urban have-nots, 22 Mixed apartment dwellers, 23 Young and rising, 24 Young, low educated, 25 Young seniors in the city, 26 Own home elderly, 27 Seniors in apartments, 28 Residential elderly, 29 Porchless seniors: no front yard, 30 Religious elderly singles, 31 Low income catholics, 32 Mixed seniors, 33 Lower class large families, 34 Large family, employed child, 35 Village families, 36 Couples with teens (Married with children), 37 Mixed small town dwellers, 38 Traditional families, 39 Large religious families, 40 Large family farms, 41 Mixed rurals.

L1 information: 1 20-30 years, 2 30-40 years, 3 40-50 years, 4 50-60 years, 5 60-70 years, 6 70-80 years.

L2 information: 1 Successful hedonists, 2 Driven Growers, 3 Average Family, 4 Career Loners, 5 Living well, 6 Cruising Seniors, 7 Retired and Religeous, 8 Family with grown ups, 9 Conservative families, 10 Farmers.

L3 information: 0 0%, 1 1 - 10%, 2 11 - 23%, 3 24 - 36%, 4 37 - 49%, 5 50 - 62%, 6 63 - 75%, 7 76 - 88%.

L4 information: 0 0, 1 1 - 49, 2 50 - 99, 3 100 - 199, 4 200 - 499, 5 500 - 999, 6 1000 - 4999, 7 5000 - 9999, 8 10.000 - 19.999, 9 20.000 - Inf.

Source

Data is (c) Sentient Machine Research 2000

This dataset is owned and supplied by the Dutch datamining company Sentient Machine Research, and is based on real world business data. You are allowed to use this dataset and accompanying information for NON commercial research and education purposes only. It is explicitly NOT allowed to use this dataset for commercial education or demonstration purposes.

<http://kdd.ics.uci.edu/databases/tic/tic.data.html>.

References

P. van der Putten and M. van Someren (eds) . CoIL Challenge 2000: The Insurance Company Case. Published by Sentient Machine Research, Amsterdam. Also a Leiden Institute of Advanced Computer Science Technical Report 2000-09. June 22, 2000.

See Also

<http://kdd.ics.uci.edu/databases/tic/tic.html>

There is a special website for this benchmark at <http://www.liacs.nl/~putten/library/cc2000/>. On this website, you can find an online report featuring 29 papers written by participants in the CoIL Challenge 2000 and further background information.

Examples

```
# (1) load of data
#
data(ICB1)
dim(ICB1)
head(ICB1)

summary(ICB1)

data(ICB2)
```

itamtplcost

Italian MTPL cost

Description

This dataset contains large losses (in excess of 500 Keuro) of an Italian Motor-TPL company since 1997.

Usage

```
data(itamtplcost)
```

Format

itamtplcost is a data frame of 2 columns and 457 rows:

Date Date of loss (accident date).

UltimateCost Ultimate cost trended to 2013 and developed to ultimate losses.

Source

Unknown private insurer.

Examples

```
# (1) load of data
#
data(itamtplcost)
```

linearmodelfactor	<i>A simulated with linear model factor</i>
-------------------	---

Description

A simulated with linear model factor

Usage

```
data(linearmodelfactor)
```

Format

The variables for linearmodelfactor are

X A numeric.

Y A numeric.

Z A factor.

Examples

```
# (1) load of data
#
data(linearmodelfactor)
```

```
head(linearmodelfactor)
```

lossalae

*General Liability Claims***Description**

The `lossalae` is a data frame of 1500 rows and 2 columns containing 1,500 general liability claims randomly chosen from late settlement lags and were provided by Insurance Services Office, Inc. Each claim consists of an indemnity payment (the loss, X_1) and an allocated loss adjustment expense (ALAE). ALAE are types of insurance company expenses that are specifically attributable to the settlement of individual claims such as lawyers' fees and claims investigation expenses. The dataset also has an attribute called `capped`, which gives the row names of the indemnity payments that were capped at their policy limit. This dataset comes from the `evd` package.

The `lossalae.full` is a data frame of 1500 rows and 4 columns containing additional information compared to `lossalae`: the limit of the policy is available.

Usage

```
data(lossalae)
data(lossalae.full)
```

Format

`lossalae` contains two columns:

`Loss` A numeric vector containing the indemnity payments (USD).

`ALAE` A numeric vector containing the allocated loss adjustment expenses (USD).

`lossalae.full` contains four columns:

`Loss` A numeric vector containing the indemnity payments (USD).

`ALAE` A numeric vector containing the allocated loss adjustment expenses (USD).

`Limit` A numeric vector containing the policy limit (USD).

`Censored` A binary indicating that the payments are capped to their policy limit (USD).

Source

Frees, E. W. and Valdez, E. A. (1998) Understanding relationships using copulas. *North American Actuarial Journal*, **2**, 1–15, doi:[10.1080/10920277.1998.10595749](https://doi.org/10.1080/10920277.1998.10595749).

References

Klugman, S. A. and Parsa, R. (1999) Fitting bivariate loss distributions with copulas. *Insurance: Mathematics and Economics*, **24**, 139–148, doi:[10.1016/S01676687\(98\)000390](https://doi.org/10.1016/S01676687(98)000390).

Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. L. (2004) *Statistics of Extremes: Theory and Applications*., Chichester, England: John Wiley and Sons, doi:[10.1002/0470012382](https://doi.org/10.1002/0470012382).

Cebrian, A.C., Denuit, M. and Lambert, P. (2003). *Analysis of bivariate tail dependence using extreme value copulas: An application to the SOA medical large claims database*, Belgian Actuarial Bulletin, Vol. 3, No. 1, <https://hdl.handle.net/2078.5/72404>.

Examples

```
# (1) load of data
#
data(lossalae)
data(lossalae$full)

# (2) plot of data
#
plot(lossalae$ALAE, lossalae$Loss, log="xy", pch=19)
```

norauto

Norwegian fire insurance dataset

Description

This dataset comprises 183,999 observations of automobile insurance policies losses over a one-year period.

Usage

```
data(norauto)
```

Format

norauto contains 7 columns (each row is a policy):

Male 1 if the policyholder is a male, 0 otherwise.

Young 1 if the policyholder age is below 26 years, 0 otherwise.

DistLimit The distance limit as stated in the insurance contract: "8000 km", "12000 km", "16000 km", "20000 km", "25000-30000 km", "no limit".

GeoRegion Density of the geographical region (from heaviest to lightest): "High+", "High-", "Medium+", "Medium-", "Low+", "Low-".

Expo Exposure as a fraction of year.

ClaimAmount 0 or the average claim amount if NbClaim > 0.

NbClaim The claim number.

Source

Unknown Norwegian insurer.

Downloaded from University of Oslo: <https://www.uio.no/studier/emner/matnat/math/STK4520/index-eng.html>

Examples

```
# (1) load of data
#
data(norauto)
summary(norauto)
```

Norberg

Ragnar Norberg's credibility dataset

Description

This univariate dataset was self-made by Norberg (1979) for pointing out the relevancy of credibility. It contains hypothetical records of binary claim of an insurance portfolio with 20 policies.

Usage

```
data(Norberg)
```

Format

Norberg contains 20 columns and 10 rows. Rows are the 10 years of experience, while columns are the 20 policies in the portfolio.

Source

Public.

References

Dataset used in Ragnar Norberg (1979), *The credibility approach to experience rating*, Scandinavian Actuarial Journal, 181-221, [doi:10.1080/03461238.1979.10413721](https://doi.org/10.1080/03461238.1979.10413721).

Examples

```
# (1) load of data
#
data(Norberg)

# (2) plot and description of data
#
matplot(0:9, apply(Norberg, 2, cumsum)/(1:10), type="l",
ylim=c(0, 1), main="Claim experience")
```

norfire

Norwegian fire insurance dataset

Description

This dataset comprises 9181 fire losses over the period 1972 to 1992 from an unknown Norwegian company. A priority of 500 thousands of Norwegian Krone (NKR) was applied to get this dataset.

Usage

```
data(norfire)
```

Format

norfire contains three columns:

Year The year of claim occurrence.

Loss The total loss amount NKR thousands.

Loss2012 The total loss amount in thousands of 2012 Norwegian Krone, inflated using the Norwegian CPI.

Source

<https://lstat.kuleuven.be/>

References

Beirlant, Teugels and Vynckier (1996), *Practical Analysis of Extreme Values*, Leuven University Press, <https://www.jstor.org/stable/2236602>.

Beirlant, Matthys and Diercks (2001), *Heavy-tailed distributions and rating*, ASTIN Bulletin, Vol. 31, Issue 1, doi:10.2143/AST.31.1.993.

Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. L. (2004) *Statistics of Extremes: Theory and Applications.*, Chichester, England: John Wiley and Sons, doi:10.1002/0470012382.

Examples

```
# (1) load of data
#
data(norfire)

# (2) plot and description of data
#

boxplot(Loss ~ Year, data= norfire, log="y", xlab="Year",
        ylab="Claim size", main="Norwegian fire dataset")
```

nortrip18800

Australian liability insurance triangles

Description

Dataset nortrip18800 contains claim triangles from a Norwegian non-life insurer between 1988 and 2000 for bodily injuries. nortrip18800 is a list of 5 elements : a triangle of claim counts by the sum of reporting and valuation delay, a triangle of claim payments by the sum of reporting and valuation delay, a triangle of reported incurred claims by the sum of reporting and valuation delay, a triangle of claim payments by valuation delay, a triangle of reported incurred claims by valuation delay. Values are cumulated amounts.

Usage

```
#1st Line of Business
data(nortrip18800)
```

Format

nortritpl8800\$countbyrepdel, nortritpl8800\$paidbyrepdel, nortritpl8800\$incurbyrepdel contain the insurance triangles by reporting+valuation delay. nortritpl8800\$paidbydel, nortritpl8800\$incurbydel contains the insurance triangles by valuation delay.

References

W. Neuhaus (2004), *On the Estimation of Outstanding Claims*, Australian Actuarial Journal, 10, 485-518.

Examples

```
# (1) load of data
#
```

```
#1st Line of Business
data(nortritpl8800)
```

nzcathist

New Zealand catastrophe historic

Description

Historical disaster statistics in Zealand from 1968 to 2014.

Usage

```
data(nzcathist)
```

Format

nzcathist is a data frame of 9 columns:

Year a numeric for the Year.

Quarter a numeric for the quarter of the year.

Date a character string for the date.

FirstDay a Date object for the first day of natural catastrophe.

Event a character string describing the event.

Type a factor describing the event type among the list: "Cyclone", "Earthquake", "Flood", "Flood, Storm", "Hailstorm", "Other", "Power outage", "Storm", "Tornado", "Weather".

Location a character string describing the location.

OriginalCost Original cost in million of Australian dollars (NZD).

NormCost2011 Normed cost in million of 2011 New Zealand dollars (NZD).

NormCost2014 Normed cost in million of 2014 New Zealand dollars (NZD).

Source

<https://www.icnz.org.nz/natural-disasters>

Examples

```
# (1) load of data
#
data(nzcathist)

# (2) plot of data
#
plot(ecdf(nzcathist$NormCost2014))
```

PnCdemand

Property and casualty insurance demand

Description

The PnCdemand contains indicators of the demand for property and liability insurance in terms of national economic and risk aversion characteristics. There are 22 countries over 7 years between 1987-1993.

Usage

```
data(PnCdemand)
```

Format

PnCdemand contains 22 columns:

"Name" A character for the country name.

"Country" A numeric for the country identifier.

"Time" A numeric for the time identifier.

"GNPCAP" A numeric for the Gross national product, in US dollars per capita..

"NewMEAS" A numeric for the new measure of wealth produced by the World Bank. It is a composite measure that includes human resources, produced or manufactured assets and natural resources. This variable is time-invariant. It is wealth per capita, in thousands of US dollars.

"RiskAversion" A numeric for the risk aversion, which is proxied by level of education. This is measured by the enrollment ratio of third-level education, that is, the ratio of total enrollment in third-level education institutions to the total population age 20 to 24. Education at the third level is provided by different types of institutions, including universities, teacher-training institutions and technical institutes.

"Protect" A numeric for the protective measures may reduce competition and thus raise prices. Trade barriers are proxied by the insurance market share of foreign firms. Specifically, this is the market share of branches or agencies of foreign undertakings in total domestic non-life insurance.

"PopDens" A numeric for the population density, the average number of people living within a square kilometer.

- "Urban" A numeric for the urbanization. The percentage of people living in urban areas.
- "LegalSyst" A numeric for the legal system. This is an indicator variable that is equal to one if the country has a common law system and is zero otherwise (statutory law system). This variable is time-invariant.
- "CPI" A numeric for the Consumer Price Index, as a percentage.
- "Auto" Automobile premium density, computed as total direct gross automobile insurance premiums divided by the country's population. It includes damage or loss to land vehicles as well as liability arising out of the use of motor vehicles. The measure is in US dollars per capita.
- "Transport" Transport premium density. Transport insurance includes railway loss, aircraft loss and liability and ship loss and liability.
- "Freight" Freight premium density. It includes all damage to or loss of goods in transit or baggage.
- "FireProp" Fire and other property damage premium density. It includes damage or loss of property due to fire, explosion, storm, other natural forces, nuclear energy and land subsidence as well as other damage to property.
- "PecLoss" Pecuniary loss premium density. It includes credit loss, surety loss and other miscellaneous financial losses.
- "GenLiab" General liability premium density. It includes all liability other than motor vehicle, aircraft and ship liability.
- "AccSick" Accident and sickness premium density.
- "OtherNL" Other non-life premium density. It includes legal expenses, assistance and other miscellaneous insurance.
- "MRATE" Motor vehicle ownership per capita.
- "NumAcc" ?
- "Population" Total population number.

Source

[FreesBook-LPD](#)

References

- Browne, M. J., Chung, J. and Frees, E. W. (2000). *International property-liability insurance consumption*. Journal of Risk and Insurance, 73-90, [doi:10.2307/253677](#).
- Frees, E. W. (2004). *Longitudinal and panel data: analysis and applications in the social sciences*. Cambridge University Press, [doi:10.1017/CBO9780511790928](#).

Examples

```
# (1) load of data
#
data(PnCdemand)
```

pricingame	<i>French Motor Third-Part Liability datasets used for 100 percent Data Science game</i>
------------	--

Description

pg15training, pg15pricing are the two datasets used for the 2015 pricing game of the French institute of Actuaries organized on November 5, 2015. pg15training contains 100,000 TPL policies for private motor insurance used to fit the models, whereas pg15pricing contains 36,311 policies of the same guarantee for which the premium is computed. Each record has been observed at most one year and contains risk features of the policyholder and the insured vehicle. For confidentiality reasons, most categorical levels have unknown meaning.

pg16trainpol, pg16trainclaim, pg16test are the three datasets used for the 2016 pricing game of the French institute of Actuaries organized on November 8, 2016. pg16trainpol contains 87,228 policies for private motor insurance and pg16trainclaim contains 4,568 claims of those 87,228 TPL policies. Policies are guaranteed for all kinds of material damages, but not bodily injuries. Both datasets are used to fit the models, whereas pg16test is used for training. For confidentiality reasons, most categorical levels have unknown meaning.

pg17trainpol, pg17trainclaim are the two training datasets used for the 2017 pricing game of the French institute of Actuaries organized on November 16, 2017. pg17trainpol contains 100,000 policies for private motor insurance and pg17trainclaim contains 14,243 claims of those 100,000 TPL policies. These training sets correspond to year $t = 0$. pg17testyear1, pg17testyear2, pg17testyear3, pg17testyear4 are the four test datasets used for the pricing game: each has 100,000 rows of new policies (drivers willing to purchase insurance for Year t with $t = 1, 2, 3, 4$).

Usage

```
data(pg15training)
data(pg15pricing)

data(pg16trainpol)
data(pg16trainclaim)
data(pg16test)

data(pg17trainpol)
data(pg17trainclaim)
data(pg17testyear1)
data(pg17testyear2)
data(pg17testyear3)
data(pg17testyear4)
```

Format

pg15training and pg15pricing are two dataframes with the same columns:

PolNum The policy number.
CalYear The underwriting year.
Gender The gender of the car driver.

Type The car type (a single letter).

Category The car category (a string character).

Occupation The occupation of the driver (a string character).

Age The driver age, in years (in France, people can drive a car at 18).

Group1 The group of the car.

Bonus The bonus-malus (French no-claim discount): -30 means a 30 percent bonus while +20 means a 20 percent malus; see details below.

PolDur The policy age (in year).

Value The car value (in euro).

Adind A dummy variable indicating a material cover.

SubGroup2 The subregion of the driver home (unknown category).

Group2 The region of the driver home (unknown category).

Density The density of inhabitants (number of inhabitants per km²) in the city the driver of the car lives in.

Expdays Exposure in days.

Numtppd The number of third-party material claims.

Numtpbi The number of third-party bodily injury claims.

Indtppd The total cost of third-party material claims (euro).

Indtpbi The total cost of third-party bodily injury claims (euro).

pg16trainpol, pg16trainclaim, pg16test are dataframes with the following columns:

Year The coverage year.

BeginDate,EndDate Beginning date and ending date of the coverage period (of class "Date").

Exposure The exposure as a fraction of year, computed as the difference between EndDate and BeginDate divided by 365.

PolicyID The identification number of the policy.

PolicyAgeCateg The category of the policy age.

PolicyCateg The category of the policy.

CompanyCreation A dummy indicating if the company has been created.

FleetMgt The fleet management category.

FleetSizeCateg The fleet size category

Area The geographical area.

PayFreq The payment frequency.

VehiclAge The vehicle age category.

VehiclNb The number of vehicles

VehiclCateg The vehicle category.

VehiclPower The vehicle power

LicNb The license number of the vehicle.

Deduc The deductible category

SumInsured The category of the sum insured.

BusinessType The business type.

ChannelDist The distribution channel.

ClaimNb The claim number.

ClaimCharge The claim charge.

DirectComp As claims correspond only to material damage, the French claim convention (IDA) was applied. So the insurer may directly refund the insured (when DirectComp=TRUE) even if the insurer will sue the third-party insurer to recover the indemnity afterwards.

CompRate The rate of compensation (in percent).

SettlYear The settlement year.

pg17*** are dataframes with the following columns:

id_client The client identification number: a string of the form Annnnnnnn (A followed by an 8-digit number). First client ID is A00000001 and last is A00091488.

id_vehicle The vehicle identification number: a string of the form Vnn (a V followed by a 2-digit number). First vehicle is always numbered V01. If a client has multiple vehicles, then the numeration increases by 1. There is no particular ordering in the vehicles, so their rank should not represent anything valuable.

id_policy The policy identification number, a string of the form Annnnnnnn-Vnn resulting from appending id_client and id_vehicle.

id_year The year of coverage, Year ID begins at "Year 0" and ends at "Year 4".

pol_bonus The policy bonus (French no-claim discount): 0.5 means a 30 percent bonus while 1.2 means a 20 percent malus; see details below.

pol_coverage The coverage category: The coverage are of 4 types : Mini, Median1, Median2 and Maxi, in this order. As you can guess, Mini policies covers only Third Party Liability claims, whereas Maxi policies covers all claims, including Damage, Theft, Windshield Breaking, Assistance, etc.

pol_duration The policy duration: Policy duration represents how old the policy is. It is expressed in year, accounted from the beginning of the current year i. Oldest policies in this portfolio can last since prehistoric ages of 45 years.

pol_sit_duration The policy current endorsement duration: Situation duration represent how old the current policy characteristics are. It can be different from pol duration, because the same insurance policy could have evolved in the past (e.g. by changing coverage, or vehicle, or drivers, ...).

pol_pay_freq The payment frequency: The price of the insurance coverage can be paid annually, bi-annually, quarterly or monthly.

pol_payd A dummy indicating pay as you drive: a string with Yes or No, which indicates whether our client has subscribed a mileage-based policy or not. In those early ages of Year 0, Pay As You Drive was not that current, so they represent a minority in the portfolio.

pol_usage The policy usage: it describes what usage the driver makes from his vehicle, most of time. There are 4 possible values : "WorkPrivate" which is the most common, "Retired" which is presumed to be aimed at retired people (who also are presumed driving less kilometers), "Professional" which denotes a professional usage of the vehicle, and "AllTrips" which is quite similar to Professional (including pro tours). As for the coverage, it would be very surprising that this variable had no effect on frequency.

pol_insee_code The INSEE code of the French city/municipality where the policyholder lives: it is a 5-digits alphanumeric code used by the French National Institute for Statistics and Economic Studies (hence INSEE) to identify "communes" and departments in France. There are about 36,000 "communes" in France, but not every one of them is present in the dataset

(there are only 18,000 of them). The first 2 digits of insee code identifies the department (they are 96, not including overseas departments). The insee code or department code can be used to possibly merge external data to the datasets: population density, OSM data, etc.

`drv_drv2` A character string indicating if there is a secondary driver: there is always a first driver, which characteristics (age, sex, licence) are provided, but a secondary driver is optional, and is present 1 time out of 3.

`drv_age1`, `drv_age2` The driver age of the *i*th driver: it is expressed in years counted from the beginning of the considered year. Then, `drv_age1` increases by 1 every year, like in real world... Legal age to drive is 18, so you shouldn't find any age below that limit. Due to the fact that the database is built on existing situations before Year 0, in fact the minimum age is 19 in Year 0 dataset. On the other side, you'll also find quite old drivers.

`drv_sex1`, `drv_sex2` The driver sex of the *i*th driver. European rules force insurers to charge the same price for women and men. But driver's gender can still be used in academic studies, and that's why `drv_sex1` is still available in the datasets, and can be used as discriminatory variable in this pricing game.

`drv_age_lic1`, `drv_age_lic2` The age of the driving license of the *i*th driver. As for the other ages, it is expressed in integer years from the beginning of the current year.

`vh_age` The vehicle age: This variable is the vehicle's age, the difference between the year of release and the current year.

`vh_cyl` The engine cylinder displacement is expressed in ml in a continuous scale. This variable should be highly correlated with `din` power of the vehicle.

`vh_din` The `vh_din` is a representation of the motor power. Highly correlated with `din` power, cylinder, speed and even value of the vehicle.

`vh_fuel` The vehicle fuel type: with mainly two values "Diesel" and "Gasoline". Very few Hybrid vehicles can also be found, but, 6 years ago, the hybrid market was still at its beginning.

`vh_make` The vehicle carmaker. As the database is built from a French insurance, the three major brands are Renault, Peugeot and Citroen.

`vh_model` The vehicle model. As a subdivision of the carmake, vehicle is identified by its model name.

`vh_sale_begin`, `vh_sale_end` `vh_sale_begin` and `vh_sale_end` are the dates (in fact: ages) from the beginning of the current year of the beginning and the end of marketing years of the vehicle. This could for instance identify policies that covers very new vehicles or second-hand ones.

`vh_speed` The vehicle maximum speed (km/h), as stated by the manufacturer.

`vh_type` The vehicle type, either "Tourism" or "Commercial". There are more "Commercial" types for "Professional" policy usage than for "WorkPrivate".

`vh_value` The vehicle's value (replacement value) is expressed in euros, without inflation so it should be stable from a year to another.

`vh_weight` The vehicle weight (kg).

`id_claim` The claim identification number: a string of the form CLnn (CL followed by a 2-digit number). Numbering of the claims begins at 1 for every policy and each year. Then, the last value of `id_claim` is the maximum number of claims for a vehicle in a year. Two-digits representation is sufficient : this maximum doesn't exceed 7 (but not on Year 0, where the maximum is 6).

`claim_nb` The claim number, as we are talking about individual claims, each `claim_nb` has a value of 1.

`claim_amount` The claim amount: amounts range from (approx.) -2,000 to +300,000. Yes, there are negative values, they come from claims where our driver's liability is not engaged, so there's a legal recourse.

The bonus/malus system is compulsory in France, but we will only use it here as a possible feature. The coefficient is attached to the driver. It starts at 1 for young drivers (i.e. first year of insurance). Then, every year without claim, the bonus decreases by 5 percent until it reaches its minimum of 0.5. Without any claim, the bonus evolution would then be : $1 \rightarrow 0.95 \rightarrow 0.9 \rightarrow 0.85 \rightarrow 0.8 \rightarrow 0.76 \rightarrow 0.72 \rightarrow 0.68 \rightarrow 0.64 \rightarrow 0.6 \rightarrow 0.57 \rightarrow 0.54 \rightarrow 0.51 \rightarrow 0.5$. Every time the driver causes a claim (only certain types of claims are taken into account), the coefficient increases by 25 percent, with a maximum of 3.5. Thus, the range of bonus/malus coefficient extends from 0.5 to 3.5 in the datasets.

Source

Datasets from unknown private insurers.

See <https://freakonometrics.hypotheses.org/20034> for the first pricing game.

See <https://actinfo.hypotheses.org/69> for the second pricing game.

See <https://actinfo.hypotheses.org/86> for the third pricing game.

Examples

```
# (1) load of data
#
data(pg15training)
data(pg15pricing)

data(pg16trainpol)
data(pg16trainclaim)
data(pg16test)

data(pg17trainpol)
data(pg17trainclaim)
data(pg17testyear1)

# (2) some check
# should be zero
sum(!pg16trainclaim$PolicyID %in% pg16trainpol$PolicyID)
# should be true
NROW(pg16trainclaim) == sum(pg16trainpol$ClaimNb)
```

Description

This dataset contains automobile injury claim number collected in 1993 in Singapore by the General Insurance Association of Singapore. Records contains individuals characteristics in addition to claim counts.

Usage

```
data(sgautonb)
```

Format

sgautonb is a data frame of 8 columns and 1,340 rows:

SexInsured Gender of insured, including male (M), female(F) and unspecified (U).

Female Numeric: 1 if female, 0 otherwise.

VehicleType The type of vehicle being insured, such as automobile (A), truck (T), and motorcycle (M).

PC Numeric: 1 if private vehicle, 0 otherwise.

Clm_Count Number of claims during the year.

Exp_weights Exposure weight or the fraction of the year that the policy is in effect.

LNWEIGHT Logarithm of exposure weight.

NCD No Claims Discount. This is based on the previous accident record of the policyholder. The higher the discount, the better is the prior accident record.

AgeCat The age of the policyholder, in years grouped into seven categories. 0-6 indicate age groups 21 and younger, 22-25, 26-35, 36-45, 46-55, 56-65, 66 and over, respectively.

VAgeCat The age of the vehicle, in years, grouped into seven categories. 0-6 indicate groups 0, 1, 2, 3-5, 6-10, 11-15, 16 and older, respectively.

AutoAge0 Numeric: 1 if private vehicle and VAgeCat = 0, 0 otherwise.

AutoAge1 Numeric: 1 if private vehicle and VAgeCat = 1, 0 otherwise.

AutoAge2 Numeric: 1 if private vehicle and VAgeCat = 2, 0 otherwise.

AutoAge Numeric: 1 if Private vehicle and VAgeCat = 0, 1 or 2, 0 otherwise.

VAgecat1 VAgeCat with categories 0, 1, and 2 combined.

Source

[FreesBook-RMAFA](#)

References

Frees, E.W. (2010), *Regression modelling with actuarial and financial applications*, Cambridge University Press, [doi:10.1017/CBO9780511814372](#).

Frees, E.W., and E. Valdez (2008). *Hierarchical insurance claims modeling*, Journal of the American Statistical Association 103, 1457-1469, [doi:10.1198/016214508000000823](#).

Examples

```
# (1) load of data
#
data(sgautonb)
dim(sgautonb)
head(sgautonb)
```

`sgtriangles`*Singapore general liability triangles*

Description

`sgautoprop9701` is a data report incremental payments from a portfolio of automobile policies for a Singapore property and casualty (general) insurer for years 1997-2001. Payments are for third party property damage from comprehensive insurance policies. All payments have been deflated using a Singaporean consumer price index, so they are in constant dollars.

`sgautoBI9301` contains incremental payments from a portfolio of automobile policies for a Singapore property and casualty (general) insurer for years 1993-2001. Payments, deflated for inflation, are for third party injury from comprehensive insurance policies.

Usage

```
data(sgautoprop9701)
data(sgautoBI9301)
```

Format

`sgautoprop9701` and `sgautoBI9301` are two matrices containing insurance triangles.

Source

[Freesbook-RMAFA](#)

References

Frees, E.W. (2010), *Regression modelling with actuarial and financial applications*, Cambridge University Press, [doi:10.1017/CBO9780511814372](#).

Frees, E.W., and E. Valdez (2008). *Hierarchical insurance claims modeling*, Journal of the American Statistical Association 103, 1457-1469, [doi:10.1198/016214508000000823](#).

Examples

```
# (1) load of data
#
data(sgautoprop9701)
data(sgautoBI9301)
```

SOAGMI

SOA Group Medical Insurance claim dataset

Description

The dataset was collected by SOA for a group medical insurance and contains records of all the claim amounts exceeding 25,000 USD over the period 1991 and is available at <https://www.soa.org>. There is no truncation due to maximum benefits.

Usage

```
data(SOAGMI)
```

Format

SOAGMI contains two columns and 371 rows:

Year The year of claim occurrence.

Loss The loss amount in euros (EUR).

Source

<https://lstat.kuleuven.be/>

References

Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. L. (2004) *Statistics of Extremes: Theory and Applications.*, Chichester, England: John Wiley and Sons, [doi:10.1002/0470012382](https://doi.org/10.1002/0470012382).

Grazier and G'Sell (1997), *Group Medical Insurance Large Claims Database and Collection*, SOA Monograph M-HB97-1, Society of Actuaries, Schaumburg.

Cebrian, A.C., Denuit, M. and Lambert, P. (2003). *Analysis of bivariate tail dependence using extreme value copulas: An application to the SOA medical large claims database*, Belgian Actuarial Bulletin, Vol. 3, No. 1, <https://hdl.handle.net/2078.5/72404>.

Examples

```
# (1) load of data
#
data(SOAGMI)
```

spacedata

*Space dataset***Description**

This dataset contains 1,698 observations of satellites between 1956 and 2013 where the study focuses failure and success once the satellite has reached its targeted orbit. Failures during the launching step or the testing step are not considered.

Usage

```
data(spacedata)
```

Format

spacedata is a data frame of 16 columns and 1,698 rows:

Event A character string describing the launch: always "LAUNCH: Satellite launched successfully".

EventDate The date of the launch.

MissionType A character string describing the mission goals.

InitOrbit A character string for the satellite orbit, see details.

OrbitRange A character string summarizing the satellite orbit.

Position A character for the position.

ContractLife The contractual life (in years).

Sector A character string: either "CIVIL" or "MILITARY".

IsCommercial When civil usage, 1 indicates private (commercial), 0 public (institution).

Mass Mass of satellite (Kg).

RetireDate Date of retirement, if any.

TotalFailDate Date of total failure, if any, see details.

PartialFailDate Date of partial failure, if any, see details.

AnyFailDate Date of first failure, in any.

OperLifeTime Life Length of the satellite (in years) when operating successfully.

Censored Indicator for censoring.

Details

The satellite orbit is an acronym given by

EO Elliptical Orbit.

G Geostationary.

GTO Geostationary Transfert Orbit.

HEL Heliocentric Orbit.

HEO Highly Elliptical Orbit.

LEO Low Earth Orbit.

MEO Medium Earth Orbit.

PEO Polar Elliptical Orbit.

PO Polar Orbit.

SSO Sun-Synchronous Orbit

Some details on earth orbit are given below:

LEO Low Earth orbits (LEO) are defined to be orbits with an average altitude that is less than 2,000 km. An important subset of LEO is the sun-synchronous orbit (SSO). These are circular orbits with an altitude between 500 km and 1200 km that provide an orbital period that result in passes over a point on the Earth's surface at the same time of day, a fixed number of days apart. This is ideal for Earth observation missions. LEO has predominantly been used by civil and military agencies for Earth observation, scientific missions, manned missions and intelligence or spy satellites.

MEO Medium Earth orbits (MEO) are defined to be orbits with an average altitude in the range of 5,000 to 20,000 km. The U.S. military were the first to exploit this orbit with the Global Positioning Satellites (GPS). The numerous satellites in the constellation appear to move slowly across the sky of an observer and several satellites are always visible at any point on the Earth's surface. A similar orbit is used by the Russia's equivalent Glonass system and the European Galileo.

GEO The Geostationary Earth Orbit GEO type orbit features an altitude of approximately 36,000 km. The matched orbital period means that the satellite will appear to be nearly stationary in the sky of an observer, allowing for simplified earth communications and a global coverage. The main use of this type of orbit has been for the telecommunications industry, point-to-point, mobile and direct broadcast. A significant secondary user has been for Earth observation, especially meteorological but also military missile launch and nuclear explosion detection satellites. Commercial use of space satellites has tended to concentrate on the GEO orbit with the market predominantly developing in the late 1970s and throughout the 1980s and 1990s. Total demand for launches to GEO again increased to 1997, mainly due to commercial interests, before a sharp decline in demand into the early 2000s.

Generally, a difference is made between partial losses and total losses with the following definitions:

Total Loss - Constructive Total Loss: (1) Total Loss means physical destruction of the spacecraft, no separation from the launch vehicle or injection in a useless orbit, loss of control of the spacecraft. (2) Constructive Total Loss means a partial loss where the loss ratio is equal or above 75 percent, assimilated to a Total Loss.

Partial Loss: loss of performance impacting the spacecraft intended mission, reduction of useful lifetime, permanently intermittent mission based on a predetermined loss formula.

Source

Data based on two actuarial memoirs and partially modified to fit package standards.

References

Guelou, S. (2013). *Risques spatiaux: modelisation de la fiabilite des satellites en orbite.*, EURO Institut d'Actuariat master thesis, University of Brest, France.

Gauche, J.F. (2012). *Space risks.*, Centre d'Etudes Actuarielles master thesis, Paris, France.

See Also

Castet, J.F. and Saleh, J.H. (2011). *Spacecraft reliability and multi-state failures : a statistical approach*, Wiley.

Castet, J.F., Dubos, G.F and Saleh, J.H. (2011). *Statistical reliability analysis of satellites by mass category : Does spacecraft size matter?*, Acta Astronautica, pages 584-595.

Examples

```
# (1) load of data
#
data(spacedata)
dim(spacedata)
```

swautoins

Swedish Motor Insurance dataset

Description

This dataset contains motor insurance data collected in 1977 in Sweden by the Swedish Committee on the Analysis of Risk Premium. Records contains individuals characteristics in addition to claim counts and severities.

Usage

```
data(swautoins)
```

Format

swautoins is a data frame of 7 columns and 2,182 rows:

Kilometres Distance driven by a vehicle, grouped into five categories.

Zone Graphic zone of a vehicle, grouped into 7 categories.

Bonus Driver claim experience, grouped into 7 categories.

Make The type of a vehicle

Insured The number of policyholder years. A policyholder year is the fraction of the year that the policyholder has a contract with the issuing company.

Claims Number of claims.

Payment Sum of payments.

Source

[FreesBook-RMAFA](#)

References

Frees, E.W. (2010), *Regression modelling with actuarial and financial applications*, Cambridge University Press, doi:10.1017/CBO9780511814372.

Hallin and Ingenbleek (1983), *The Swedish automobile portfolio in 1977. A statistical study*, Scandinavian Actuarial Journal, 49-64, doi:10.1080/03461238.1983.10408691.

Andrews and Herzberg (1985), *Data. A collection of problems from many fields for the student and research worker*, Springer-Vedag, New York, pp. 4t3-421, doi:10.1080/00401706.1987.10488305.

Examples

```
# (1) load of data
#
data(swautoins)
dim(swautoins)
head(swautoins)
```

swbuscase

Swedish Buss Insurance dataset

Description

This data comes from the former Swedish insurance company Wasa, before its 1999 fusion with Laensfoersaekringar Alliance. In Sweden, insurance involves three types of cover: TPL (third party liability), partial casco and hull. TPL covers any bodily injuries plus property damages caused to others in a traffic accident. Partial casco (may not be used in all countries) covers theft but also some other causes of loss such as fire. Hull covers damage on the policyholder's own vehicle. Note that The TPL insurance is mandatory, while the others are optional. The three types of cover are often sold in a package as a comprehensive insurance, but they are usually priced separately. This dataset contains information relative to partial casco only for buss in the commercial lines. Transportation companies own one or more buses which are insured for a shorter or longer period. It contains aggregated data on 670 companies that were policyholders at Wasa insurance company during the years 1990-1998.

Usage

```
data(swbuscase)
```

Format

swbuscase is a data frame of 7 columns and 1,542 rows:

IDpol The policy ID, recoded for confidentiality reasons.

Area The type of area.

BusAgeClass The bus age class with 5 unknown categories.

ObsNb The number of observations for the company in a given tariff cell based on area and age class. There may be more than one observation per record, since each renewal is counted as a new observation.

ClaimNb The number of claims.
AggClaim The sum of claim payments.
Exposure The number of policy years.

Source

[OhlsonBook](#)

References

E. Ohlsson and B. Johansson (2010), *Non-Life Insurance Pricing with Generalized Linear Models*, Springer, [doi:10.1007/9783642107917](https://doi.org/10.1007/9783642107917).

Examples

```
# (1) load of data
#
data(swbusscase)
dim(swbusscase)
head(swbusscase)
```

swmotorcycle

Swedish Motorcycle Insurance dataset

Description

This data comes from the former Swedish insurance company Wasa, before its 1999 fusion with Laensfoersaekringar Alliance. In Sweden, insurance involves three types of cover: TPL (third party liability), partial casco and hull. TPL covers any bodily injuries plus property damages caused to others in a traffic accident. Partial casco (may not be used in all countries) covers theft but also some other causes of loss such as fire. Hull covers damage on the policyholder's own vehicle. Note that The TPL insurance is mandatory, while the others are optional. The three types of cover are often sold in a package as a comprehensive insurance, but they are usually priced separately. This dataset contains information relative to partial casco only for motorcycles. It contains aggregated data on all insurance policies and claims during 1994-1998.

Usage

```
data(swmotorcycle)
```

Format

swmotorcycle is a data frame of 9 columns and 64,548 rows:

OwnerAge The owner age.
Gender The gender.
Area The type of area.

RiskClass The motorcycle class, a classification by the so called EV ratio, defined as (Engine power in kW x 100) / (Vehicle weight in kg + 75), rounded to the nearest lower integer. The 75 kg represent the average driver weight. The EV ratios are divided into seven classes.

VehAge The Vehicle age, between 0 and 99.

BonusClass The bonusclass,taking values from 1 to 7. A new driver starts with bonus class 1; for each claim-free year the bonus class is increased by 1. After the first claim the bonus is decreased by 2; the driver can not return to class 7 with less than 6 consecutive claim free years.

Exposure The number of policy years.

ClaimNb The number of claims.

ClaimAmount The sum of claim payments.

Source

[OhlsonBook](#)

References

E. Ohlsson and B. Johansson (2010), *Non-Life Insurance Pricing with Generalized Linear Models*, Springer, [doi:10.1007/9783642107917](https://doi.org/10.1007/9783642107917).

Examples

```
# (1) load of data
#
data(swmotorcycle)
dim(swmotorcycle)
head(swmotorcycle)
```

swtriangles

Switzerland general liability triangles

Description

swtrilauto is a named list of two triangles : the incurred (cumulative) amounts and the paid (cumulative) amounts.

Usage

```
data(swtrilauto)
```

Format

swtriangles is a named list of two matrices, respectively for incurred and paid amounts.

References

Dahms, R. (2008), *A Loss Reserving Method for Incomplete Claim Data*, Bulletin of the Swiss Association of Actuaries, pp. 127-148.

Dahms, R., Merz, M., Wuethrich, M.V. (2009), *Claims development result for combined claims incurred and claims paid data*. Bulletin Francais d'Actuariat 9 (18), 5-39.

Merz, M., and M. V. Wuethrich (2010), *Paid-Incurred Chain Claims Reserving Method*, Insurance: Mathematics and Economics 46, 2010, pp. 568-579, doi:10.1016/j.insmatheco.2010.02.004.

Merz, M., and M. V. Wuethrich (2013), *Estimation of Tail Development Factors in the Paid-Incurred Chain Reserving Method*, Variance 71, pp. 61-73.

Examples

```
# (1) load of data
#
data(swtr1auto)
```

ukaggclaim

UK Car Insurance Claims for 1975

Description

The data give the average claims for damage to the owner's car for privately owned and comprehensively insured vehicles in Britain in 1975. Averages are given in pounds sterling adjusted for inflation. The datasets contains 128 observations.

Usage

```
data(ukaggclaim)
```

Format

ukaggclaim contains 5 columns:

OwnerAge Policy-holder's age in years, categorized into 8 levels.

Model Type of car, in 4 groups.

CarAge Vehicle age in years, categorized into 4 levels.

NClaims Number of claims.

AveCost Average cost of each claim in pounds.

Source

The original dataset was provided by Baxter et al. (1980), then used in McCullagh and Nelder (1989).

References

Baxter, L. A., Coutts, S. M., and Ross, G. A. F. (1980). *Applications of linear models in motor insurance*. In Proceedings of the 21st International Congress of Actuaries, Zurich, Society of Actuaries, pages 11-29.

McCullagh, P., and Nelder, J. A. (1989). *Generalized linear models*. Chapman and Hall, London.

Atkinson, A.C. and Riani, M. (2000), *Robust Diagnostic Regression Analysis*, First Edition. New York: Springer, Table A.16

Examples

```
# (1) load of data
#
data(ukaggclaim)
dim(ukaggclaim)

# (2) summary
#
sapply(1:5, function(i) summary(ukaggclaim[,i]))
```

ukautocoll

UK Automobile Collision Claims

Description

The data give the average claims and claim counts for insured vehicles in UK. Averages are given in pounds sterling adjusted for inflation. The datasets contains 32 observations.

Usage

```
data(ukautocoll)
```

Format

ukautocoll contains 5 columns:

Age Policy-holder's age in years, categorized into 8 levels.

Model Type of car, in 4 groups.

CarAge Vehicle age in years, categorized into 4 levels.

NClaims Number of claims.

AveCost Average cost of each claim in pounds.

Source

The original dataset was provided by Baxter et al. (1980), then used in McCullagh and Nelder (1989) and Mildenhall (1999)

References

Baxter, L. A., Coutts, S. M., and Ross, G. A. F. (1980). *Applications of linear models in motor insurance*. In Proceedings of the 21st International Congress of Actuaries, Zurich, Society of Actuaries, pages 11-29.

McCullagh, P., and Nelder, J. A. (1989). *Generalized linear models*. Chapman and Hall, London.

Mildenhall, S. J. (1999). *A systematic relationship between minimum bias and generalized linear models*. Casualty Actuarial Society Proceedings 86, 393-487, Casualty Actuarial Society. Arlington, Virginia.

Examples

```
# (1) load of data
#
data(ukautocoll)
dim(ukautocoll)

# (2) summary
#
sapply(1:NCOL(ukautocoll), function(i) summary(ukautocoll[,i]))
```

usautoBI

Automobile bodily injury claim dataset

Description

This dataset contains automobile injury claims collected in 2002 by the Insurance Research Council (part of AICPCU and IIA). There are 1,340 records with demographic information, in addition to the claim amount.

Usage

```
data(usautoBI)
```

Format

usautoBI is a data frame of 8 columns and 1,340 rows:

CASENUM Case number to identify the claim.

ATTORNEY Whether the claimant is represented by an attorney: 1 is yes.

CLMSEX Claimant's gender: M for male and F for female.

MARITAL claimant's marital status : 1 if married, 2 if single, 3 if widowed, and 4 if divorced/separated.

CLMINSUR Whether or not the driver of the claimant's vehicle was uninsured: 1 if yes, 2 if no, and 3 if not applicable.

SEATBELT Whether or not the claimant was wearing a seatbelt/child restraint: 1 if yes, 2 if no, and 3 if not applicable.

CLMAGE Claimant's age.

LOSS The claimant's total economic loss (in thousands of USD).

Source

[FreesBook-RMAFA](#)

References

Frees, E.W. (2010), *Regression modelling with actuarial and financial applications*, Cambridge University Press, [doi:10.1017/CBO9780511814372](https://doi.org/10.1017/CBO9780511814372).

Examples

```
# (1) load of data
#
data(usautoBI)
dim(usautoBI)
head(usautoBI)
```

usautotriangles	<i>US Automobile triangles</i>
-----------------	--------------------------------

Description

usautotri9504 comes from Wacek (2007) and represent industry aggregates for private passenger auto liability/medical coverages. This dataset contains cumulative payments between 1995 and 2004 in millions of dollars. Amounts are based on insurance company annual statements from Schedule P (Part 3B). The elements of the triangle represent cumulative net payments, including defense and cost containment expenses.

usreauto8700 comes from the 2001 edition of the Historical Loss. This dataset has been used by Braun (2004). These data are from reinsurance business for automobile liability coverages for years 1987-2000 and contain cumulative incurred amounts in thousands of US dollars.

Usage

```
data(usautotri9504)
data(usreauto8700)
```

Format

usautotri9504, data(usreauto8700) are matrices containing insurance triangles.

Source

[FreesBook-RMAFA](#)

References

Frees, E.W. (2010), *Regression modelling with actuarial and financial applications*, Cambridge University Press, [doi:10.1017/CBO9780511814372](https://doi.org/10.1017/CBO9780511814372).

Wacek, M.G. (2007). *The path of the ultimate loss ratio estimate*, Variance 1, no. 2, 173-92.

Braun, C. (2004), *The prediction error of the chain ladder method applied to correlated run-off triangles*, ASTIN Bulletin 34, no. 2, 399-423, [doi:10.1017/S0515036100013751](https://doi.org/10.1017/S0515036100013751).

Examples

```
# (1) load of data
#
data(usautotri9504)
data(usreauto8700)
```

usexpense

US expense dataset

Description

This dataset is originally from the National Association of Insurance Commissioners and was examined by Frees (2011). This dataset contains financial statements based on 2005 annual reports for all the property and casualty insurance companies in United States. The annual reports are financial statements that use statutory accounting principles.

Usage

```
data(usexpense)
```

Format

usexpense is a data frame of 15 columns and 384 rows:

CompanyName Name of the company.

Group Indicates if the company is affiliated.

Mutual Indicates if the company is a mutual company.

Stock Indicates if the company is a stock company.

RBC Risk-Based Capital.

Expenses Total expenses incurred, in millions of dollars.

StaffWage Annual average wage of the insurer's administrative staff, in thousands of dollars.

AgentWage Annual average wage of the insurance agent, in thousands of dollars.

LongLoss Losses incurred for long tail lines, in millions of dollars.

ShortLoss Losses incurred for short tail lines, in millions of dollars.

GWPpersonal Gross written premium for personal lines, in millions of dollars.

GWPcommercial Gross written premium for commercial lines, in millions of dollars.

Assets Net admitted assets, in millions of dollars.

Cash Cash and invested assets, in millions of dollars.

LiqRatio The ratio of the liquid assets to the current liabilities level.

Source

[FreesBook-RMAFA](#)

References

Frees, E.W. (2011). *Regression Modeling with Actuarial and Financial Applications*, Cambridge University Press, doi:[10.1017/CBO9780511814372](https://doi.org/10.1017/CBO9780511814372).

Examples

```
# (1) load of data
#
data(usexpense)
```

usGLtriangles

US general liability triangles

Description

usreGL8190 comes from the 1991 edition of the Historical Loss Development Study published by the Reinsurance Association of American (page 91). This dataset has been used by Mack (1994) and by England and Verrall (2002). These data are from automatic facultative reinsurance business in general liability (excluding asbestos and environmental) coverages for years 1981-1990. Under a facultative basis, each risk is underwritten by the reinsurer on its own merits.

usreGL8700 comes from the 2001 edition of the Historical Loss. This dataset has been used by Braun (2004). These data are from reinsurance business for general liability coverages for years 1987-2000 and contain cumulative incurred amounts in thousands of US dollars.

ustri1fire is a list of two triangles for fire insurance (one for incurred amounts and the other for paid amounts) from Quard and Mack (2008).

ustri2GL is a list of three triangles for three line-of-business: commercial automobile businesses, homeowners, workers' compensation from Kirschner, Kerley and Isaacs (2002). These are cumulative paid amounts in thousands of dollars.

Usage

```
data(usreGL8700)
data(usreGL8190)
data(ustri1fire)
data(ustri2GL)
```

Format

usreGL8700 and usreGL8190 are two matrices containing insurance triangles. ustri1fire, ustri2GL are named lists.

Source

[FreesBook-RMAFA](#)

References

- Braun, C. (2004), *The prediction error of the chain ladder method applied to correlated run-off triangles*, ASTIN Bulletin 34, no. 2, 399-423, doi:10.1017/S0515036100013751.
- England, P.D., and R.J. Verrall (2002), *Stochastic claims reserving in general insurance*, British Actuarial Journal 8, 443-544, doi:10.1017/S1357321700003809.
- Frees, E.W. (2010), *Regression modelling with actuarial and financial applications*, Cambridge University Press, doi:10.1017/CBO9780511814372.
- Mack, T. (1994), *Measuring the variability of chain-ladder reserve estimates*, Casualty Actuarial Society, Spring Forum, Arlington, Virginia.
- Quard, G. and Mack, T. (2008), *Munich Chain Ladder: a reserving method that reduces the gap between IBNR projections based on paid losses and IBNR projections based on incurred losses*, Variance, Volume 2, Issue 2.
- Kirschner, G.S., Kerley C. and Isaacs B. (2002), *Two approaches to calculating correlated reserves indicators across multiple lines of business*, CAS forum fall.

Examples

```
# (1) load of data
#
data(usreGL8700)
data(usreGL8190)

data(ustrilfire)
data(ustril2GL)
```

ushurricane

Normalized Hurricane Damages

Description

Normalized Hurricane Damages in the United States: 1900-2005 was studied in Pielke et al. (2008). Weinkle et al. (2018) provides a major update to the leading dataset on normalized US hurricane losses in the continental United States from 1900 to 2017. Over this period, 197 hurricanes resulted in 206 landfalls with about US\$2 trillion in normalized (2018) damage, or just under US\$17 billion annually.

Grinsted et al. (2018) develop a record of normalized damage since 1900 based on an equivalent area of total destruction (ATD). Their record of normalized damage, framed in terms of an equivalent area of total destruction, is a more reliable measure for climate-related changes in extreme weather, and can be used for better risk assessments on hurricane disasters.

Usage

```
data(ushu17stormloss)
data(ushu17annualloss)
data(ushu17inflation)
data(ushu17population)
```

```
data(ushu18ICAT)
data(ushu18W)
data(ushu18NCEI)
```

Format

ushu17stormloss is a data frame of 7 columns and 207 rows:

Year Year of the Hurricane.

Storm ID ID

Storm Name Description of the Hurricane.

Catgeory Category of the Hurricane.

State States damaged by the Hurricane.

Base Economic Damage Economic damages (original USD).

Normalized PL 2018 Normalized PL18 damages (2018 USD).

Normalized CL 2018 Normalized CL18 damages (2018 USD).

ushu17annualloss is a data frame of 2 columns and 106 rows:

Year Year

PL18 Sum for Year Aggregate of PL18 over a year

CL18 Sum for Year Aggregate of CL18 over a year

ushu17inflation is a data frame of 9 columns and 106 rows:

Year Year.

Implicit.Price.Deflator Implicit price deflator.

Inflation.Multiplier Inflation multiplier.

Wealth Wealth.

Real.Wealth.2005.Base Real wealth (2005 base).

Real.Wealth.Per.Capita Real wealth per capita.

Real.Wealth.Per.Capita.Multiplier Real wealth per capita multiplier.

Real.Wealth.Per.Housing.Unit Real wealth per housing unit.

Real.Wealth.Per.Housing.Unit.Multiplier Real wealth per housing multiplier.

ushu17population is a data frame of 12 columns and 217 rows:

Storm ID Storm ID.

Year Year of the Storm.

Storm Name Name of the Storm.

Original Population Original population in counties affected by storm.

2018 Population 2018 population in counties affected by storm.

Population Multiplier County population multiplier.

Original Housing Units Original housing units in counties affected by storm.

2018 Housing Units 2018 housing units in counties affected by storm.

Housing Units Multiplier Housing units multiplier.

Population Total US population.

Housing Units Total US housing units.

ushu18ICAT, ushu18W, ushu18NCEI are data frames with most of the following columns:

ATCF_ID storm identifier.

Storm ID storm identifier.

name name of the storm.

basedamage Current refers to "at the time" in dollars.

ATD Area of Total Destruction.

CPI Consumer price index adjusted damages (2018 USD).

ND Normalized damage (conventional) 2018 USD.

lf_ISO_TIME time of landfall (as in IBTrACS).

lf_wind windspeed at landfall (from IBTrACS).

lf_pressure pressure near landfall (from IBTrACS).

lf_state US state dominating the damage record (from source dataset).

WPC primarily based on Current-Cost Net Stock of Fixed Assets and Consumer Durable Goods from the U.S. Bureau of Economic Analysis (current USD).

population population

lf_lat designated landfall location: latitude.

lf_lon designated landfall location: longitude.

References

Pielke, Gratz, Landsea, Collins, Saunders, and Musulin (2008), *Normalized Hurricane Damages in the United States: 1900-2005*, Natural Hazards Review, Volume 9, Issue 1, pp. 29-42. doi:[10.1061/\(ASCE\)15276988\(2008\)9:1\(29\)](https://doi.org/10.1061/(ASCE)15276988(2008)9:1(29))

Weinkle, J., Landsea, C., Collins, D., Musulin, R., Crompton, R. P., Klotzbach, P. J., Pielke Jr, R. (2018) *Normalized hurricane damage in the continental United States 1900-2017*, Nature sustainability, 1(12), 808-813. doi:[10.1038/s4189301801652](https://doi.org/10.1038/s4189301801652)

Grinsted, A., Ditlevsen, P., & Christensen, J. H. (2019). *Normalized US hurricane damage estimates using area of total destruction, 1900-2018* Proceedings of the National Academy of Sciences, 116(48), 23942-23946. doi:[10.1073/pnas.1912277116](https://doi.org/10.1073/pnas.1912277116)

Examples

```
# (1) load of data
#
```

```
data(ushu17stormloss)
data(ushu17annualloss)
data(ushu17inflation)
data(ushu17population)
```

```
data(ushu18ICAT)
data(ushu18W)
data(ushu18NCEI)
```

ushustormloss4980 *Normalized Hurricane Damages in US between 1949 and 1980*

Description

Normalized Hurricane Damages in the United States due to single hurricanes. They applied to the period from 1949 and 1980 and are adjusted for inflation. Originally, the dataset was compiled by the American Insurance Association and is also reported in Beirlant, Teugels and Vynckier (1996).

Usage

```
data(ushustormloss4980)
```

Format

ushustormloss4980 is a data frame of 7 columns and 207 rows:

NormLoss80 Normalized damages (million of 1980 USD).

References

Dataset used in Beirlant, Teugels and Vynckier (1996), *Practical Analysis of Extreme Values*, Leuven University Press.

Examples

```
# (1) load of data
#
data(ushustormloss4980)
```

uslapseagent *United States lapse dataset from tied-agent channel*

Description

The uslapseagent portfolio contains detailed information on the 29,317 Whole Life policies, all sold from the tied-agent channel between January 1995 and December 2008.

For each policy, we know the issuance date, the gender of the policyholder, the age category, etc...~Unfortunately, some variables are rather uninformative.

Usage

```
data(uslapseagent)
```

Format

uslapseagent is a data frame of 14 columns and 29,317 rows:

issue.date Issue date. For policies not terminated in December 2008, we have non information: fixed right censored.

duration Time duration in quarters, unknown if censored.

acc.death.rider Indicates if the policy has an accidental death rider (i.e. an option covering accidental death).

gender The gender of the policyholder.

premium.frequency The premium frequency: either infra-annual (monthly, quarterly, semi-annual); annual or supra-annual.

risk.state The risk state: either "Smoker" or "NonSmoker".

underwriting.age The underwriting age: either "Young" (between 0 and 34 years old), "Middle" (between 35 and 54 years old) or "Old" (between 55 and 84 years old).

living.place The living place (categorical value).

annual.premium The annual premium (standardized scale): mean 560.88 and standard deviation 526.58 in original USD scale.

DJIA the last observed quarterly variation of the DowJones Index (in standardized scale): mean 0.00178 and standard deviation 0.0494 in original scale.

termination.cause The type of termination.

surrender A binary variable indicating the surrender by policyholder.

death A binary variable indicating the death of policyholder.

other A binary variable indicating other termination such as term.

allcause A binary variable indicating all termination.

Source

Unknown non-life insurers from United States, used in Milhaud and Dutang (2018), preprint at <https://hal.science/hal-01985256>.

References

Milhaud, X., Dutang, C. (2018), *Lapse tables for lapse risk management in insurance: a competing risk approach*. European Actuarial Journal, 8 (1), 97-126, doi:10.1007/s1338501801657.

Examples

```
# (1) load of data
#
data(uslapseagent)
head(uslapseagent)
```

usmassBI

*Massachusetts Automobile bodily injury claim datasets***Description**

The dataset usmassBI contains automobile bodily injury claims collected in 2001 in Massachusetts, and studied in Frees (2010) and Rempala and Derrig (2005). There are 348 records with demographic information, in addition to the claim amount. Claims that are closed by year end are excluded. Potential fraudulent claims are from provider=A.

The dataset usmassBI2 contains automobile bodily injury claims collected between 1993 and 1998 in Massachusetts, and studied in Frees and Wang (2005). This is a sample of 29 Massachusetts towns described in Frees (2003). Claim amounts have been rescaled to adjust for the effects of inflation: all claims are in 1991 dollars, using the Consumer Price Index (CPI) for the rescaling factor.

Usage

```
data(usmassBI)
data(usmassBI2)
```

Format

usmassBI is a data frame of 8 columns and 1,340 rows:

claims Claim amount for bodily insurance coverage (in millions of USD).
 provider Health care provider is either "A" or "Other".
 providerA Binary variable indicating the presence of "Other" provider.
 logclaims Logarithm of claim amount.

usmassBI2 is a data frame of 5 columns and 174 rows:

TOWNCODE The index of Massachusetts towns.
 YEAR The calendar year of the observation.
 AC Average claims per unit of exposure.
 PCI Per-capita income of the town.
 PPSM Population per square mile of the town.

Source

[FreesBook-RMAFA](#)

References

- Frees, E.W. (2003), *Multivariate Credibility for Aggregate Loss Models*, North American Actuarial Journal 7(1), 13-37, [doi:10.1080/10920277.2003.10596074](#).
- Frees, E.W. (2010), *Regression modelling with actuarial and financial applications*, Cambridge University Press, [doi:10.1017/CBO9780511814372](#).
- Frees, E.W. and Wang, P. (2005), *Credibility using copulas*, North American Actuarial Journal, 9(2), 31-48, [doi:10.1080/10920277.2005.10596196](#).
- Rempala, G.A., and R.A. Derrig (2005), *Modeling hidden exposures in claim severity via the EM algorithm*, North American Actuarial Journal 9(2), 108-128, [doi:10.1080/10920277.2005.10596206](#).

Examples

```

# (1) load of data
#
data(usmassBI)
dim(usmassBI)
head(usmassBI)

# (1) load of data
#
data(usmassBI2)
dim(usmassBI2)
head(usmassBI2)

# summary tables
sapply(levels(usmassBI2$TOWNCODE), function(x) summary(subset(usmassBI2, TOWNCODE == x)$AC))
sapply(unique(usmassBI2$YEAR), function(x) summary(subset(usmassBI2, YEAR == x)$AC))

#plot average claims
plot(AC~YEAR, data=usmassBI2)
for(i in usmassBI2$TOWNCODE) lines(AC~YEAR, data=subset(usmassBI2, TOWNCODE== i), col=i)

```

usmedclaim

*US Medical claim incremental triangles***Description**

This dataset comes from Gamage et al. (2007) and contains medical-care payments by month between January 2001 and December 2003. Payments for medical-care coverage come from policies with no deductible or coinsurance. For a given month and a development year, payments are aggregated among members but are cumulated over development year. The payments exclude prescription drugs that typically have a shorter payment pattern than other medical claims.

Usage

```
data(usmedclaim)
```

Format

usmedclaim is a matrix containing two columns (with members count and month) and the insurance triangle.

Source

[FreesBook-RMAFA](#)

References

Frees (2010), *Regression modelling with actuarial and financial applications*, Cambridge University Press, doi:[10.1017/CBO9780511814372](https://doi.org/10.1017/CBO9780511814372).

Gamage, J., Linfield, J., Ostaszewski, K. and S. Siegel (2007). *Statistical methods for health actuaries - IBNR estimates: An introduction*, Society of Actuaries Working Paper, Schaumburg, Illinois.

Examples

```
# (1) load of data
#
data(usmedclaim)
head(usmedclaim, 10)

# (2) graph of data
#
matplot(t(as.matrix(usmedclaim[,-(1:2)])), type="b", main="Payment by accident month",
        xlab="Month", ylab="Amount (USD)")
```

usMSHA1316

*U.S. Mine Safety and Health Administration claim datasets***Description**

usMSHA1316 is a data set from the U.S. Mine Safety and Health Administration from 2013 to 2016. The data set was used in the Predictive Analytics exam administered by the Society of Actuaries in December 2018. This data set contains 53,746 observations described by 20 variables, including compositional variables.

Usage

```
data(usMSHA1316)
```

Format

usMSHA1316 is a data frame of 8 columns and 1,340 rows:

US_STATE U.S. state where mine is located.

COMMODITY Class of commodity mined.

PRIMARY Primary commodity mined.

SEAM_HEIGHT Coal seam height in inches (coal mines only).

TYPE_OF_MINE Type of mine.

MINE_STATUS Status of operation of mine.

AVG_EMP_TOTAL Average number of employees.

EMP_HRS_TOTAL Total number of employee hours.

PCT_HRS_UNDERGROUND Proportion of employee hours in underground operations hours.

PCT_HRS_SURFACE Proportion of employee at surface operations of underground mine hours.

PCT_HRS_STRIP Proportion of employee at strip mine hours.

PCT_HRS_AUGER Proportion of employee in auger mining hours.

PCT_HRS_CULM_BANK Proportion of employee in culm bank operations hours.

PCT_HRS_DREDGE Proportion of employee in dredge operations hours.

PCT_HRS_OTHER_SURFACE Proportion of employee in other surface mining operations hours.

PCT_HRS_SHOP_YARD Proportion of employee in independent shops and yards.

PCT_HRS_MILL_PREP Proportion of employee hours in mills or prep plants.

PCT_HRS_OFFICE Proportion of employee hours in offices.

NUM_INJURIES Total number of accidents reported.

Source

<https://www.soa.org/>

References

Gan, Guojun, and Emiliano A. Valdez. 2024. *Compositional Data Regression in Insurance with Exponential Family PCA*, *Variance* 17 (1), <https://variancejournal.org/article/116404-compositional-data-re> doi:10.48550/arXiv.2112.14865 of the arxiv version.

Examples

```
# (1) load of data
#
data(usMSHA1316)
dim(usMSHA1316)
head(usMSHA1316)
```

usMVTA

Motor Vehicle Traffic Accident: A consolidated database of police-reported motor vehicle traffic accidents in the United States for actuarial applications

Description

usMVTA dataset contains a sample of 1 583 520 people involved in 20 years of fatal and non-fatal accidents. The dataset is a representative sample of motor vehicle traffic accidents from the United States of America during the period 2001 to 2020. The dataset is derived from the publicly available data collected by an agency of the U.S. Department of Transportation called the National Highway Traffic Safety Administration (see NHTSA(2022)). There are 49 available variables in the dataset. All variables are denoted below, refer to Iturria et al.(2021a). This dataset is available on Zenodo, see Iturria et al.(2021b).

Usage

```
data(usMVTA)
```

Format

usMVTA is a data frame of 49 columns and 1 583 520 rows: (character strings are of class factor)

ST_CASE Unique case number assigned to each crash.

VEH_NO Assigned to each vehicle in the case.

PER_NO Consecutive number assigned to each person in the case.

AGE Discrete age categories. Due to historical coding practices, people aged 97 or older are coded as 97. The range is (0, 97).

GENDER A character string either 'Female' or 'Male'.

YEAR This data element records the year in which the crash occurred.

SOURCE Source of the element (CRSS = Crash Report Sampling System, FARS = Fatality Analysis Reporting System, GES = General Estimates System).

PER_TYP This variable describes the role of the individual. Stationary non-occupants (SNO) are people in a working vehicle, transport device or standing in buildings. A character string: 'Driver', 'Passenger', 'Pedalcyclists', 'Pedestrians', 'SNO'.

INJ_SEV The 9,325 and 2,648 records in GES/CRSS and FARS, respectively, that were reported as injured but their injury severity is unknown (historically coded with 5) are not useful to quantify insurance losses. Therefore, these records were randomly reassigned with equal probabilities to the categories of the severity of the injury. A character string: 'Fatal Injury', 'Minor Injury', 'No Injury', 'Possible Injury', 'Serious Injury'.

DRINKING This variable records whether the individual was recorded as having been drinking. A character string: 'No', 'Yes'.

DRUGS This variable records whether the individual was under the influence of drugs. A character string: 'No', 'Yes'.

NUMOCCS Discrete number of occupants in the vehicle, an integer ranges in (1,80).

MAKE Discrete vehicle's make categories. Coding has been standard since 1988 and 1991 for GES/CRSS and FARS, respectively. In the FARS user's manual, code 77 corresponds to the make Victory which is omitted in both user's manual for GES/CRSS. Regardless, this code appears in 52 records for GES/CRSS, which we assume corresponds to Victory and therefore, omitted in the NHTSA notes. A character string converted from an integer in (1, 98).

MODEL Discrete vehicle's model categories. Models for non- standard cars are recoded as NaN. FARS and CRSS have the same coding practice. GES uses the same as FARS for the period 2011-2015 but there is a different coding standard during 2001-2010. To standardize, the Make-Model tables were checked for the records that make up 80 percent of the data. Differences were standardized with some models of: Volkswagen, KIA and Oldsmobile. A character string converted from an integer in (1, 63).

MOD_YEAR Discrete number for the vehicle's model year. Ranges in (1900, 2021).

HIT_RUN An indicator of a hit-and-run. A character string: 'No', 'Yes'

BODY_TYP Classification of the vehicle based on its configuration, shape, size and doors. A character string: '(2,3)-door hatchback', '(4,5)-door hatchback', '2-door sedan', '3-door coupe', '3-wheel automobile', '4-door sedan', 'auto-based panel', 'auto-based pickup', 'buses', 'convertible', 'hatchback (unknown door number)', 'large limousine', 'light trucks', 'medium/heavy trucks', 'motorcycles', 'other automobiles', 'other vehicles', 'sedan (unknown door number)', 'station wagon', 'utility vehicles', 'van-based trucks'.

DEFORMED This variable records the amount of damage sustained by the vehicle. A character string: 'minor damage', 'moderate damage', 'no damage', 'severe damage'.

SPEC_USE Example of a vehicle with a special use are taxi, military vehicle, police vehicle, ambulance, fire truck, among others. A character string: 'no special use', 'special use'.

TRAV_SP Discrete number for travel speeds in miles per hour. Values greater than 96 coded as 97. An integer ranges in (0, 97).

DR_ZIP Driver's address U.S. zip codes. An integer of the form XXXXX.

SPEEDREL This variable records whether the driver's speed was related to the crash. Different speed related categories in all datasets grouped to the 'Yes' classification. FARS data prior to 2009 did not include this variable and instead, the variable DR_SF1 had speeding categories with codes 43, 44 and 46. Thus, from 2001 to 2008, the aforementioned codes are standardized so that 'Yes' corresponds to 1. A character string: 'No', 'Yes'.

- DR_SF1 Factors related to the driver expressed in the case materials. Careless driving includes: improper driving, road rage or driving in an emotional state (fatigued, depressed, among others). Police related factors include: police pursuit, alcohol and or drug test refused and nontraffic violation charged (manslaughter, homicide, among others). A character string: 'Careless driving', 'Miscellaneous', 'None', 'Police related'.
- HARM_EV This field describes the first injury or damage producing event of the crash. MVT stands for motor vehicles in transport. Non-collision includes rollover, fire or explosion, gas inhalation, surface irregularities, among others. A character string: 'Collision with fixed object', 'Collision with MVT', 'Collision with object not fixed', 'Non-collision'.
- HOUR Discrete number denoting the hour of the accident. Accidents that occurred at 12:00 am standardized to 0 hours. An integer ranges in (0, 23).
- WEATHER Weather at the time of the accident. An 'atmospheric condition' includes rain, snow, cloudy, fog/smoke, sand, among others. A character string: 'Atmospheric condition', 'Clear'.
- STRATUM This data element identifies the number of the categories in which the police report was originally listed. An integer ranges from 1 to 10.
- REGION NHTSA Region. A character string: 'Midwest', 'Northeast', 'South', 'West'.
- PSU Primary sampling unit (PSU). 3117 counties in the country were grouped into 707 PSU.
- PJ This integer identifies the number of the police jurisdiction from which the police crash report was originally sampled.
- WEIGHT Case weight, this data element is used to produce national estimates from the data.
- NUM_VEH Denotes the number of vehicles involved in the MVTA.
- MAX_SEV The maximum severity variable is the highest injury severity of all the people involved in the same MVTA. A character string: 'Minor Injury', 'No Injury', 'Possible Injury', 'Serious Injury'.
- MAKEMODEL An integer is created as a concatenation of MODEL and MAKE of the vehicle.
- COUNTYNAME Reflects the location of the accident. Derived from driver's zip code if unavailable and possible. A character string from 'Abbeville', 'Acadia',... to 'Yuma', 'Zavala'.
- STATENAME Reflects the location of the accident. Derived from driver's zip code if unavailable. A character string from 'Alabama', 'Alaska',... to 'Wisconsin', 'Wyoming'.
- SEG Socio-economic groups. An integer ranges (1, 10).
- MARITAL The marital status, denoted by MARITAL, is randomly assigned using probabilities based on age, gender and zip code. For parsimony, we use only two mutually exclusive categories for marital status. A character string: 'Married', 'Single'.
- POP2018 Population count for the zip code. This allows to distinguish between rural and urban areas.
- RACE The so-called race of the individual by NHTSA. A character string: 'Asian', 'Black', 'Hispanic', 'White'.
- PREV Summary of the driving record variables (PREV_ACC, PREV_SUS, PREV_DWI and PREV_SPD). An integer: 1 if the person has had one or more accidents or driving offences in the last 5 years, and to 0 otherwise.
- DR_DRINK This field records whether the driver was drinking. A character string: 'No', 'Yes'.
- CDL_STAT This field indicates the status of the driver's commercial driver's license (CDL). A character string: 'Cancelled or Denied', 'Commercial Learner's Permit', 'Disqualified', 'Expired', 'No Driver Present/Unknown if Driver Present', 'No license', 'Not Reported', 'Other - Not Valid', 'Revoked', 'Suspended', 'Unknown CDL', 'Unknown License Status', 'Valid'.

PREV_ACC This field indicates if there was any previous crashes for this driver that occurred within 5 years of the crash date.

PREV_SUS This field indicates if there was any previous license suspensions or revocations for this driver that occurred within 5 years of the crash date.

PREV_DWI This field indicates if there was any previous DWI (driving while intoxicated) convictions for this driver that occurred within 5 years of the crash date.

PREV_SPD This field records any previous speeding convictions for this driver that occurred within 5 years of the crash date.

COUNTY This data element records the location of the unstabilized event with regard to the County. The codes are from the General Services Administration's (GSA) publication of worldwide Geographic Location Codes (GLC).

ZCTA U.S. Zip code of the crash. An integer of the form XXXXX.

Source

Iturria, A., Andres, C., Hardy, M. and Marriott, P., (2021a), see below.

References

Iturria, A., Andres, C., Hardy, M. and Marriott, P., (2021a) *A Consolidated Database of Police-Reported Motor Vehicle Traffic Accidents in the United States for Actuarial Applications*, 2021. Available at [doi:10.2139/ssrn.3977693](https://doi.org/10.2139/ssrn.3977693)

Iturria, A., Hardy, M. and Marriott, P., (2021b) *A consolidated database of police-reported motor vehicle traffic accidents in the United States for actuarial applications*, 2021 (3.1.0), Zenodo. [doi:10.5281/zenodo.7120835](https://doi.org/10.5281/zenodo.7120835)

NHTSA, *Crash Report Sampling System Analytical User's Manual, 2016-2020*, 2022. Available at <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813236>

Examples

```
# (1) load of data
#
data(usMVTA)
```

usprivautoclaim

US Private Auto Claims

Description

This dataset contains claim amounts for private motor insurance from a US property and casualty insurer. Claims that were not closed by the year end are excluded. A risk classification is available and is based on driver and vehicle characteristics.

Usage

```
data(usprivautoclaim)
```

Format

usprivautoclaim contains 5 columns:

STATE State in US.

CLASS Risk category.

GENDER Gender.

AGE Driver age.

PAID Claim amount.

Source

[FreesBook-RMAFA](#)

References

Frees, E.W. (2010), *Regression modelling with actuarial and financial applications*, Cambridge University Press, doi:10.1017/CBO9780511814372.

Hallin and Ingenbleek (1983), *The Swedish automobile portfolio in 1977. A statistical study*, Scandinavian Actuarial Journal, 49-64, doi:10.1080/03461238.1983.10408691.

Andrews and Herzberg (1985), *Data. A collection of problems from many fields for the student and research worker*, Springer-Vedag, New York, pp. 4t3-421, doi:10.1080/00401706.1987.10488305.

Examples

```
# (1) load of data
#
data(usprivautoclaim)
dim(usprivautoclaim)
```

usquakeLR

California earthquake loss ratios

Description

Loss ratios for earthquake insurance in California between 1971 and 1994.

Usage

```
data(usquakeLR)
```

Format

usquakeLR is a data frame of 2 columns and 24 rows:

Year Year of the earthquake.

LossRatio Loss ratio.

References

Jaffee, D.M. and Russell, T. (1996), *Catastrophe Insurance, Capital Markets and Uninsurable Risks*, Philadelphia: Financial Institutions Center, The Wharton School, p. 96-112, doi:10.2307/253729.

Embrechts, Resnick and Samorodnitsky (1999). *Extreme Value Theory as a Risk Management Tool*, North American Actuarial Journal, Volume 3, Number 2, doi:10.1080/10920277.1999.10595797.

Examples

```
# (1) load of data
#
data(usquakeLR)

# (2) plot log scale
#
plot(usquakeLR$Year, usquakeLR$LossRatio+1e-3,
      ylim=c(1e-3, 1e4), log="y", ylab="Loss Ratio", xlab="Year")
```

ustermLife

US Term Life insurance

Description

This dataset comes from Survey of Consumer Finances (SCF), a nationally representative sample that contains extensive information on assets, liabilities, income, and demographic characteristics of those sampled (potential U.S. customers). It contains a random sample of 500 households with positive incomes that were interviewed in the 2004 survey. For term life insurance, the quantity of insurance is measured by the policy face, the amount that the company will pay in the event of the death of the named insured. Characteristics include annual income, the number of years of education of the survey respondent and the number of household members.

Usage

```
data(ustermLife)
```

Format

ustermLife is a data frame of 15 columns and 384 rows:

Gender Gender of the survey respondent.

Age Age of the survey respondent.

MarStat Marital status of the survey respondent: 1 if married, 2 if living with partner, and 0 otherwise.

Education Number of years of education of the survey respondent.

Ethnicity Ethnicity.

SmarStat Marital status of the respondent's spouse.

Sgender Gender of the respondent's spouse.

Sage Age of the respondent's spouse.
 Seducation Education of the respondent's spouse.
 NumHH Number of household members.
 Income Annual income of the family.
 TotIncome Total income.
 Charity Charitable contributions.
 Face Amount that the company will pay in the event of the death of the named insured.
 FaceCVLifePol Face amount of life insurance policy with a cash value.
 CashCVLifePol Cash value of life insurance policy with a cash value.
 BorrowCVLifePol Amount borrowed on life insurance policy with a cash value.
 NetValue Net amount at risk on life insurance policy with a cash value.

Source

[FreesBook-RMAFA](#)

References

Frees, E.W. (2011). *Regression Modeling with Actuarial and Financial Applications*, Cambridge University Press, [doi:10.1017/CBO9780511814372](https://doi.org/10.1017/CBO9780511814372).

Examples

```
# (1) load of data
#
data(ustermlife)
```

uswarrantagnum

Warranty Automobile claims

Description

This dataset contains claims numbers for a sample of 15,775 automobiles that were sold and under warranty for 365 days. Warranties are guarantees of product reliability issued by the manufacturer. The warranty data are for one vehicle system (e.g., brakes or power train) and cover one year with a 12,000 mile limit on coverage.

Usage

```
data(uswarrantagnum)
```

Format

uswarrantagnum is a data frame of 8 columns and 1,340 rows:

PolicyNumber Policy number.
 ClaimNumber Claim number. 5 is actually 5 and more.

Source

[FreesBook-RMAFA](#)

References

Cook, R.J. and J.F. Lawless (2002), *The statistical analysis of recurrent events*, Springer, doi:10.1007/9780387698106.

Frees, E.W. (2010), *Regression modelling with actuarial and financial applications*, Cambridge University Press, doi:10.1017/CBO9780511814372.

Examples

```
# (1) load of data
#
data(uswarrantagnum)
uswarrantagnum
```

usworkcomp

US workers compensation datasets

Description

The dataset usworkcomp is originally from the National Council on Compensation Insurance and was examined by Klugman (1992), Frees et al. (2001) and Frees (2011). This database contains records of losses due to permanent or partial disability claims for workers compensation insurance in US. For each claim amount, the payroll is available as a measure of exposure units. A total of 847 data points is available coming from the observation of 121 risk classes over 7 years.

The dataset usworkcomptri8807 comes from an unknown US insurer: this reserve triangle was used in Lacoume (2007).

Usage

```
data(usworkcomp)
```

Format

usworkcomp is a data frame of 4 columns and 847 rows:

CL Occupation class identifier, 1-124.

YR Year identifier, 1-7.

PR Payroll, a measure of exposure to loss, in dollars.

LOSS Losses related to permanent partial disability, in dollars.

usworkcomptri8807 is a reserve triangle with 21 development years and 20 accident years.

Source

[FreesBook-RMAFA](#)

References

Klugman, S.A. (1992). *Bayesian Statistics in Actuarial Science*, Kluwer, Boston, doi:10.1007/978-9401708456.

Frees, E.W. and Young, V.R. and Luo, Y. (2001), *Case studies using panel data models*, North American Actuarial Journal, 5, 24-42, doi:10.1080/10920277.2001.10596010.

Lacoume, A. (2007), *Mesure du risque de reserve sur un horizon de un an*, Actuary memoir, ISFA.

Frees, E.W. (2011). *Regression Modeling with Actuarial and Financial Applications*, Cambridge University Press, doi:10.1017/CBO9780511814372.

Examples

```
# (1) load of data
#
data(usworkcomp)
```

```
# Table 3 of Fres et al. (2001)
# (in million USD)
```

```
t(sapply(unique(usworkcomp$YR),
function(y) summary( subset(usworkcomp, YR == y)[,"PR"] / 10^6 )))
```

Index

* datasets

asiacomrisk, 4
ausautoBI8999, 5
auscathist, 6
ausNLHYby, 7
ausNLHYglossary, 10
ausNLHYlloyd, 12
ausNLHYtotal, 14
ausNSW, 16
ausprivauto, 17
austriLoB, 19
beaonre, 21
beMTPL16, 21
beMTPL97, 23
besecura, 24
bragg, 25
brautocoll, 26
brgeomunicins, 27
brvehins, 29
canlifins, 31
CASdatasets, 32
catelematic13, 35
credit, 37
danish, 39
Davis, 40
ECBYieldCurve, 41
eqlist, 41
eudirectlapse, 43
euhealthinsurance, 45
euMTPL, 46
eusavingsurrender, 47
eutelematic13, 48
FedYieldCurve, 49
forexUSUK, 50
fre4LoBtriangles, 51
freaggnumber, 52
frebiloss, 53
freclaimset, 54
freclaimset9207, 55
frecomfire, 57
freDemography, 58
freDisTables, 59
freIRSA, 62
fremarine, 63
freMortTables, 64
fremotorclaim, 66
freMPL, 70
freMTPL, 73
freportfolio, 76
fretplclaimnumber, 79
hurricanehist, 80
ICB, 81
itamtplcost, 84
linearmodelfactor, 85
lossalae, 86
norauto, 87
Norberg, 88
norfire, 88
nortritpl8800, 89
nzcathist, 90
PnCdemand, 91
pricingame, 93
sgautonb, 97
sgtriangles, 99
SOAGMI, 100
spacedata, 101
swautoins, 103
swbusscase, 104
swmotorcycle, 105
swtriangles, 106
ukaggclaim, 107
ukautocoll, 108
usautoBI, 109
usautotriangles, 110
usexpense, 111
usGLtriangles, 112
ushurricane, 113
ushustormloss4980, 116
uslapseagent, 116
usmassBI, 118
usmedclaim, 119
usMSHA1316, 120
usMVTA, 121
usprivautoclaim, 124
usquakeLR, 125
ustermlife, 126

- uswarrantagnum, 127
- usworkcomp, 128
- * **dataset**
 - fretelematic, 78
- asiacomrisk, 4
- ausautoBI8999, 5
- auscathist, 6, 32
- ausMTP18486, 33
- ausMTP18486 (ausprivauto), 17
- ausNLHYby, 7, 12, 14, 16
- ausNLHYCapAdeq, 32
- ausNLHYCapAdeq (ausNLHYtotal), 14
- ausNLHYCapAdeqByComp, 32
- ausNLHYCapAdeqByComp (ausNLHYby), 7
- ausNLHYClaimByState, 32
- ausNLHYClaimByState (ausNLHYby), 7
- ausNLHYFinPerf, 32
- ausNLHYFinPerf (ausNLHYtotal), 14
- ausNLHYFinPerfByComp, 32
- ausNLHYFinPerfByComp (ausNLHYby), 7
- ausNLHYFinPerfPublic, 32
- ausNLHYFinPerfPublic (ausNLHYby), 7
- ausNLHYFinPos, 32
- ausNLHYFinPos (ausNLHYtotal), 14
- ausNLHYFinPosByComp, 32
- ausNLHYFinPosByComp (ausNLHYby), 7
- ausNLHYFinPosPublic, 32
- ausNLHYFinPosPublic (ausNLHYby), 7
- ausNLHYglossary, 9, 10, 14, 16
- ausNLHYLiability, 32
- ausNLHYLiability (ausNLHYtotal), 14
- ausNLHYlloyd, 9, 12, 12, 16
- ausNLHYlloydAsset, 32
- ausNLHYlloydAsset (ausNLHYlloyd), 12
- ausNLHYlloydGPI, 32
- ausNLHYlloydGPI (ausNLHYlloyd), 12
- ausNLHYlloydUWAcc, 32
- ausNLHYlloydUWAcc (ausNLHYlloyd), 12
- ausNLHYlloydUWRes, 32
- ausNLHYlloydUWRes (ausNLHYlloyd), 12
- ausNLHYOffProf, 32
- ausNLHYOffProf (ausNLHYtotal), 14
- ausNLHYOpIncExp, 32
- ausNLHYOpIncExp (ausNLHYtotal), 14
- ausNLHYOpIncExpPublic, 32
- ausNLHYOpIncExpPublic (ausNLHYby), 7
- ausNLHYPremByState, 32
- ausNLHYPremByState (ausNLHYby), 7
- ausNLHYPremClaim, 32
- ausNLHYPremClaim (ausNLHYtotal), 14
- ausNLHYPremClaimPublic, 32
- ausNLHYPremClaimPublic (ausNLHYby), 7
- ausNLHYPrivInsur, 32
- ausNLHYPrivInsur (ausNLHYby), 7
- ausNLHYPubInsur, 32
- ausNLHYPubInsur (ausNLHYby), 7
- ausNLHYRecAASB, 32
- ausNLHYRecAASB (ausNLHYtotal), 14
- ausNLHYReserve, 32
- ausNLHYReserve (ausNLHYtotal), 14
- ausNLHYtotal, 9, 12, 14, 14
- ausNSW, 16
- ausNSWdeath02, 33
- ausNSWdeath02 (ausNSW), 16
- ausNSWdriver04, 33
- ausNSWdriver04 (ausNSW), 16
- ausprivauto, 17
- ausprivauto0405, 33
- ausprivauto0405 (ausprivauto), 17
- ausprivauto1005, 33
- ausprivauto1005 (ausprivauto), 17
- austri1autoBI7895, 32
- austri1autoBI7895 (austriLoB), 19
- austri2auto, 32
- austri2auto (austriLoB), 19
- austriLoB, 19
- beaonre, 21, 33
- beMTP16, 21, 33
- beMTP197, 23, 33
- besequera, 24, 33
- bragg, 25
- braggclaim, 33
- braggclaim (bragg), 25
- braggprem, 33
- braggprem (bragg), 25
- brautocoll, 26, 33, 35
- brgeomunic (brgeomunicins), 27
- brgeomunicins, 27, 33, 35
- brvehins, 29
- brvehins1, 33
- brvehins1 (brvehins), 29
- brvehins1a (brvehins), 29
- brvehins1b (brvehins), 29
- brvehins1c (brvehins), 29
- brvehins1d (brvehins), 29
- brvehins1e (brvehins), 29
- brvehins2, 33
- brvehins2 (brvehins), 29
- brvehins2a (brvehins), 29
- brvehins2b (brvehins), 29
- brvehins2c (brvehins), 29
- brvehins2d (brvehins), 29
- canlifins, 31, 33

- canlifins2, [33](#)
- canlifins2 (canlifins), [31](#)
- CASdatasets, [32](#)
- casdatasets (CASdatasets), [32](#)
- catelematic13, [33](#), [35](#)
- credit, [34](#), [37](#)
- danish, [39](#)
- danishmulti, [33](#)
- danishmulti (danish), [39](#)
- danishuni, [33](#)
- danishuni (danish), [39](#)
- Date, [26](#)
- Davis, [34](#), [40](#)
- ECBYieldCurve, [35](#), [41](#)
- eqlist, [35](#), [41](#)
- eudirectlapse, [33](#), [43](#)
- euhealthinsurance, [33](#), [45](#)
- euMTPL, [33](#), [46](#)
- eusavingsurrender, [47](#)
- eusavingULnoPS (eusavingsurrender), [47](#)
- eusavingULnoPSperMth (eusavingsurrender), [47](#)
- eusavingULnoPSperQtr (eusavingsurrender), [47](#)
- eusavingULnoPSperYr (eusavingsurrender), [47](#)
- eutelematic13, [33](#), [48](#)
- FedYieldCurve, [35](#), [49](#)
- forexUSUK, [35](#), [50](#)
- fre4LoBtriangles, [34](#), [51](#)
- freaggnumber, [33](#), [52](#)
- freAS0002 (freMortTables), [64](#)
- frebiloss, [33](#), [53](#)
- freclaimset, [33](#), [54](#)
- freclaimset2motor (freclaimset), [54](#)
- freclaimset3dam9207 (freclaimset9207), [55](#)
- freclaimset3fire9207 (freclaimset9207), [55](#)
- freclaimset3multi9207 (freclaimset9207), [55](#)
- freclaimset9207, [33](#), [55](#)
- frecomfire, [33](#), [57](#)
- freDemography, [33](#), [58](#)
- freDisTables, [33](#), [59](#)
- frefictivetable (freportfolio), [76](#)
- frefictivetable2 (freportfolio), [76](#)
- frefictivetable3 (freportfolio), [76](#)
- freHMD (freMortTables), [64](#)
- freIRSA, [33](#), [62](#), [73](#), [74](#)
- fremarine, [33](#), [63](#)
- freMortTables, [33](#), [64](#)
- freMotor1freq0304a (freMotorclaim), [66](#)
- freMotor1freq0304b (freMotorclaim), [66](#)
- freMotor1freq0304c (freMotorclaim), [66](#)
- freMotor1prem0304a (freMotorclaim), [66](#)
- freMotor1prem0304b (freMotorclaim), [66](#)
- freMotor1prem0304c (freMotorclaim), [66](#)
- freMotor1sev0304a (freMotorclaim), [66](#)
- freMotor1sev0304b (freMotorclaim), [66](#)
- freMotor1sev0304c (freMotorclaim), [66](#)
- freMotor2freq9907b (freMotorclaim), [66](#)
- freMotor2freq9907u (freMotorclaim), [66](#)
- freMotor2sev9907 (freMotorclaim), [66](#)
- freMotor3freq9907b (freMotorclaim), [66](#)
- freMotor3freq9907u (freMotorclaim), [66](#)
- freMotor3sev9907 (freMotorclaim), [66](#)
- freMotor4freq9907b (freMotorclaim), [66](#)
- freMotor4freq9907u (freMotorclaim), [66](#)
- freMotor4sev9907 (freMotorclaim), [66](#)
- freMotorclaim, [33](#), [66](#)
- freMPL, [33](#), [70](#)
- freMPL1 (freMPL), [70](#)
- freMPL10 (freMPL), [70](#)
- freMPL1sub (freMPL), [70](#)
- freMPL2 (freMPL), [70](#)
- freMPL3 (freMPL), [70](#)
- freMPL4 (freMPL), [70](#)
- freMPL5 (freMPL), [70](#)
- freMPL6 (freMPL), [70](#)
- freMPL7 (freMPL), [70](#)
- freMPL8 (freMPL), [70](#)
- freMPL9 (freMPL), [70](#)
- freMTPL, [33](#), [73](#)
- freMTPL2freq (freMTPL), [73](#)
- freMTPL2sev (freMTPL), [73](#)
- freMTPLfreq, [35](#)
- freMTPLfreq (freMTPL), [73](#)
- freMTPLsev, [35](#)
- freMTPLsev (freMTPL), [73](#)
- freP2Ddis10 (freDisTables), [59](#)
- freP2Ddisprob10 (freDisTables), [59](#)
- freP2Pdis10 (freDisTables), [59](#)
- freP2Pdis13 (freDisTables), [59](#)
- freP2Pdisprob10 (freDisTables), [59](#)
- freP2Pdisprob13 (freDisTables), [59](#)
- frePF6064 (freMortTables), [64](#)
- frePM6064 (freMortTables), [64](#)
- freportfolio, [33](#), [76](#)
- freprojqxINSEE, [35](#)
- freprojqxINSEE (freportfolio), [76](#)
- freptfpermdis, [35](#)

- freptfpermdis (freportfolio), 76
- freptftempdis (freportfolio), 76
- freRegion7526 (freDemography), 58
- freRegionOld7526 (freDemography), 58
- freT2Ddis10 (freDisTables), 59
- freT2Ddisprob10 (freDisTables), 59
- freT2Pdis10 (freDisTables), 59
- freT2Pdis13 (freDisTables), 59
- freT2Pdisprob10 (freDisTables), 59
- freT2Pdisprob13 (freDisTables), 59
- freT2Tdis10 (freDisTables), 59
- freT2Tdis13 (freDisTables), 59
- freT2Tdisprob10 (freDisTables), 59
- freT2Tdisprob13 (freDisTables), 59
- freTD7377 (freMortTables), 64
- freTD8890 (freMortTables), 64
- fretelematic, 33, 78
- freTF0002 (freMortTables), 64
- freTGF05 (freMortTables), 64
- freTGH05 (freMortTables), 64
- freTH0002 (freMortTables), 64
- freTPG93full (freMortTables), 64
- fretplclaimnumber, 34, 79
- freTPRV93 (freMortTables), 64
- fretri1auto9605 (fre4LoBtriangles), 51
- fretri1dis1418 (fre4LoBtriangles), 51
- fretri1TPL9207 (fre4LoBtriangles), 51
- fretri2auto9605 (fre4LoBtriangles), 51
- fretri3auto9605 (fre4LoBtriangles), 51
- fretri4auto9403 (fre4LoBtriangles), 51
- freTV7377 (freMortTables), 64
- freTV8890 (freMortTables), 64

- hurricanehist, 35, 80

- ICB, 81
- ICB1, 34
- ICB1 (ICB), 81
- ICB2, 34
- ICB2 (ICB), 81
- itamtplcost, 34, 84

- linearmodelfactor, 85
- lossalae, 34, 86
- lossalae-full, 34
- lossalae-full (lossalae), 86

- norauto, 34, 87
- Norberg, 34, 88
- norfire, 34, 88
- nortritpl8800, 34, 89
- nzcathist, 34, 90

- pg15pricing (pricingame), 93
- pg15training (pricingame), 93
- pg16test (pricingame), 93
- pg16trainclaim (pricingame), 93
- pg16trainpol (pricingame), 93
- pg17testyear1 (pricingame), 93
- pg17testyear2 (pricingame), 93
- pg17testyear3 (pricingame), 93
- pg17testyear4 (pricingame), 93
- pg17trainclaim (pricingame), 93
- pg17trainpol (pricingame), 93
- PnCdemand, 35, 91
- pricingame, 34, 93

- sgautoBI9301 (sgtriangles), 99
- sgautonb, 34, 97
- sgautoprop9701 (sgtriangles), 99
- sgtriangles, 34, 99
- SOAGMI, 34, 100
- spacedata, 35, 101
- swautoins, 34, 103
- swbusscase, 34, 104
- swmotorcycle, 34, 105
- swtri1auto (swtriangles), 106
- swtriangles, 106

- ukaggclaim, 34, 107
- ukautocoll, 34, 108
- usautoBI, 34, 109
- usautotri9504 (usautotriangles), 110
- usautotriangles, 34, 110
- usexpense, 34, 111
- usGLtriangles, 34, 112
- ushu17annualloss (ushurricane), 113
- ushu17inflation (ushurricane), 113
- ushu17population (ushurricane), 113
- ushu17stormloss (ushurricane), 113
- ushu18ICAT (ushurricane), 113
- ushu18NCEI (ushurricane), 113
- ushu18W (ushurricane), 113
- ushurricane, 34, 113
- ushustormloss4980, 34, 116
- uslapseagent, 34, 116
- usmassBI, 34, 118
- usmassBI2, 35
- usmassBI2 (usmassBI), 118
- usmedclaim, 34, 119
- usMSHA1316, 34, 120
- usMVTA, 34, 121
- usprivautoclaim, 34, 124
- usquakeLR, 34, 125
- usreauto8700 (usautotriangles), 110
- usreGL8190 (usGLtriangles), 112
- usreGL8700 (usGLtriangles), 112

ustermlife, [34](#), [126](#)
ustri1fire (usGLtriangles), [112](#)
ustri2GL (usGLtriangles), [112](#)
uswarrantagnum, [34](#), [127](#)
usworkcomp, [35](#), [128](#)
usworkcomptri8807 (usworkcomp), [128](#)