Package 'lcmodels'

July 7, 2023

Title Predictions from Lung Cancer Models

Version 4.1.1

Date 2023-07-07

Author Li C. Cheung [aut, cre],
Stephanie A. Kovalchik [aut],
Hormuzd A. Katki [aut]

Maintainer Bill Wheeler <wheelerb@imsweb.com>

Description The R package provides individual risks of lung cancer and lung cancer death based on various published papers: Bach et al., 2003; Spitz et al., 2007; Cassidy et al., 2008 (LLP); Hoggart et al., 2012; Tammemagi et al., 2013; Marcus et al., 2015 (LLPi); Wilson and Weissfeld, 2015 (Pittsburgh); Katki et al., 2016 (LCRAT, LCDRAT, and versions constrained to a few variables); Katki et al., 2018; Cheung et al., 2019 (LYFS-CT). This package also estimates the Life Years Gained From Screening-CT (LYFS-CT) as per Cheung et al., 2019. It requires the same variables as LCDRAT plus 12 additional comorbidities and the year of patient assessment.

Depends R (>= 3.6)

Imports survival, VGAM

License GPL-2

Encoding UTF-8

NeedsCompilation no

R topics documented:

`cuts` *Wrapper to the cut() function*

Description

A wrapper to the `cut` function, so that you can automatically break into quantiles as the default behavior, otherwise if the breakpoints are included, then just break on those. In all cases, include.lowest is set to True

Usage

```
cuts(data, npieces, simple.labels = TRUE, ...)
```

Arguments

data	A numeric vector
npieces	Number of cut points or numeric vector of cut points
simple.labels	TRUE or FALSE
...	Other arguments passed into cut

---

lcmodels *Risk Predictions from Lung Cancer Models*

Description

The R package provides individual risks of lung cancer and lung cancer death based on various published papers: Bach et al., 2003; Spitz et al., 2007; Cassidy et al., 2008 (LLP); Hoggart et al., 2012; Tammemagi et al., 2013 (PLCOm2012); Marcus et al., 2015 (LLPi); Wilson and Weissfeld, 2015 (Pittsburgh); Katki et al., 2016 (LCRAT, LCDRAT, and versions constrained to a few predictive variables); Katki et al., 2018. This package also estimates the Life Years Gained From Screening-CT (LYFS-CT) as per Cheung et al., 2019. It requires the same variables as LCDRAT plus 12 additional comorbidities and the year of patient assessment.

Usage

```
lcmodels(x, impute.missing=TRUE, counterfactual.race=0, nyears=NULL, nyears.mortality=5)
```

Arguments

x	A data frame or matrix containing individuals' covariate values. Columns 22-33 are needed only to estimate life gained from undergoing CT screening. Covariates should be in the following columns and numerical formats: • column 1 - current age (numeric); • column 2 - gender (1=Female, 0=Male); • column 3 - years smoked (numeric); • column 4 - years quit (numeric with 0 to indicate current smoker); • column 5 - cigarettes per day (numeric); • column 6 - race (0=Non-hispanic white, 1=Non-hispanic Black/African American, 2=Hispanic, 3=Other Ethnicity); • column 7 - lung disease (1=COPD or Emphysema, 0=No COPD or Emphysema); • column 8 - number of first degree relatives with lung cancer (0,1,2); • column 9 - bmi; • column 10 - highest education level (1=<12 grade, 2=HS graduate, 3=post hs, no college, 4=associate degree/some college, 5=bachelors degree, 6=graduate school); • column 11 - asbestos exposure (1=Yes,0=No); • column 12 - prior history of pneumonia (1=Yes,0=No);

- column 13 - prior history of cancer (1=Yes,0=No);
- column 14 - family history of lung cancer (0=none, 1=early onset, 2=late onset);
- column 15 - Dust exposure (1=Yes,0=No);
- column 16 - 2 or more first degree relatives with cancer (binary indicator);
- column 17 - 1 or more first degree relatives with smoking cancer (binary indicator);
- column 18 - no hay fever (1=No Hay Fever,0=Yes Hay Fever);
- column 19 - asian ethnicity (1=Yes,0=No);
- column 20 - islander ethnicity (1=Yes,0=No));
- column 21 - American indian ethnicity (1=Yes,0=No);
- column 22 - Hypertension (1=Yes,0=No);
- column 23 - Coronary Heart Disease (1=Yes,0=No);
- column 24 - Angina pectoris (1=Yes,0=No);
- column 25 - Heart Attack (1=Yes,0=No);
- column 26 - Other heart disease (1=Yes,0=No);
- column 27 - Stroke (1=Yes,0=No);
- column 28 - Diabetes (1=Yes,0=No);
- column 29 - Chronic bronchitis in past year (1=Yes,0=No);
- column 30 - Weak/failing kidneys in past year (1=Yes,0=No);
- column 31 - Liver condition in past year (1=Yes,0=No);
- column 32 - Health problem requiring special eqiupment (1=Yes,0=No);
- column 33 - Year of assessment.

`impute.missing` TRUE or FALSE to impute missing variables using the NHIS 2015 survey (default=TRUE). The columns that can be imputed are 7-10, 13, 22-32. The imputation assumes there is complete data for columns 1-6.

`counterfactual.race`
0-3 to compute counterfactual estimates based on a different race, where (0=Non-hispanic white, 1=Non-hispanic Black/African American, 2=Hispanic, 3=Other Ethnicity). See details for more information. The default is 0.

`nyears` Time frame for prediction of the Bach, LCDRAT, and LCRAT models. The allowable values are 1-10. By default, the Bach model will use a time frame of 10 years, and a time frame of 5 years will be used for the other models.

`nyears.mortality`
Number of years to calculate risks for all-cause mortality. A maximum age of 98 is used for this calculation. If age is greater than 98, then the calculation is performed at age 98 and a warning is produced.

Details

The age range for estimation is from 40-98; results may not be accurate for ages outside of this range. The `counterfactual.race` option is used to compute additional life expectancy and life gained estimates. These estimates will appear as columns with the prefix `Counterfactual` in the returned object.

Risk factors used by each model

Covariate	Bach	Spitz	LLP	Hoggart	PLCOm2012	Pittsburgh	LLPi	LCRAT
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Gender	Yes	Yes	Yes	No	No	No	Yes	Yes
Race/ethnicity	No	No	No	No	Yes	No	No	Yes
Asian	No	No	No	No	Yes	No	No	No
Pacific islander	No	No	No	No	Yes	No	No	No
America Indian	No	No	No	No	Yes	No	No	No
Education	No	No	No	No	Yes	No	No	Yes
BMI	No	No	No	No	Yes	No	No	Yes
Smoking status	No	Yes	No	Yes	Yes	Yes	No	No
Years/Age quit	Yes	Yes	No	Yes	Yes	No	No	Yes
Years smoked	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Cigs per day	Yes	No	No	Yes	Yes	Yes	No	Yes
Pack-years	No	Yes	No	No	No	No	No	Yes
Prior cancer	No	No	Yes	No	Yes	No	Yes	No
Lung disease	No	Yes	No	No	Yes	No	Yes	Yes
Pneumonia	No	No	Yes	No	No	No	No	No
Hayfever	No	Yes	No	No	No	No	No	No
Asbestos exposure	Yes	Yes	Yes	No	No	No	No	No
Dust exposure	No	Yes	No	No	No	No	No	No
Any FDR w/ cancer	No	Yes	No	No	No	No	No	No
Any FDR w/ smoking-related cancer	No	Yes	No	No	No	No	No	No
Any FDR w/ LC	No	No	Yes	No	Yes	No	Yes	Yes
Num. FDR w/ LC	No	No	No	No	No	No	No	Yes
Early onset FDR w/ LC	No	No	Yes	No	No	No	Yes	No

Value

A numeric matrix containing individuals' predictions:

- column 1 - An indicator variable for whether the individual is eligible for CT lung screening according to US Preventive Services Task Force (USPSTF) recommendations.
- column 2 - This is the probability of dying from lung cancer within 5 years if not undergoing screening (LCDRAT from Katki, 2016).
- column 3 - This is the reduction in the probability of dying from lung cancer in 5 years
- column 4 - This is the probability of being diagnosed with lung cancer within 5 years if not undergoing screening (LCRAT from Katki, 2016).
- column 5 - This is the extra probability of lung cancer diagnosis in 5 years if undergoing 3 yearly CT lung screens as in the NLST (Katki, 2016).
- column 6 - This is the probability of having at least one false-positive CT screen out of 3 screens (Katki, 2016).
- column 7 - This is the expected number of false-positive CT screens after 3 screens (Katki, 2016).
- column This is the probability of being diagnosed with lung cancer within 10 years if not undergoing screening (Bach, 2003).
- column 9 - This is the probability of being diagnosed with lung cancer within 1 years if not undergoing screening (Hoggart, 2012).
- column 10 - This is the probability of being diagnosed with lung cancer within 5 years if not undergoing screening (LLP, 2008).
- column 11 - This is the probability of being diagnosed with lung cancer within 8.7 years if not undergoing screening (LLPi, 2015).

- column 12 - This is the probability of being diagnosed with lung cancer within 1 years if not undergoing screening (Spitz, 2007).
- column 13 - This is the probability of being diagnosed with lung cancer within 6 years if not undergoing screening (PLCOm2012 from Tammemagi, 2013).
- column 14 - This is the probability of being diagnosed with lung cancer within 6 years if not undergoing screening (Pittsburgh, 2015).
- column 15 - This is the probability of dying from lung cancer within 5 years if not undergoing screening (constrained LCDRAT).
- column 16 - This is the probability of being diagnosed with lung cancer within 5 years if not undergoing screening (constrained LCRAT).
- column 17 - Days of life gained from undergoing 3 rounds of CT screening (LYFS-CT from Cheung, 2019).
- column 18 - Life expectancy without CT screening.
- column 19 - Life expectancy with CT screening.
- column 20 - Years of life gained if lung cancer is found early due to screening.
- column 21 - Years of life gained if lung cancer death is averted due to screening
- column 22 - All-cause mortality risk
- column 23 - Counterfactual days of life gained from undergoing 3 rounds of CT screening (LYFS-CT from Cheung, 2019).
- column 24 - Counterfactual life expectancy without CT screening.
- column 25 - Counterfactual life expectancy with CT screening.
- column 26 - Counterfactual years of life gained if lung cancer is found early due to screening.
- column 27 - Counterfactual years of life gained if lung cancer death is averted due to screening

Model Objects in Package

- LCDRAT - model for lung cancer death in absence of screening;
- LCRAT - model for lung cancer incidence in absence of screening;
- cox.death - model for deaths from causes other than lung cancer;
- morat - model for overall mortality;
- polytmod - polytomous model for false positive CT lung screens.

Author(s)

Li C. Cheung, <li.cheung@nih.gov>, Stephanie A. Kovalchik, Hormuzd A. Katki

References

- Bach PB, Kattan MW, Thornquist MD, et al. Variations in lung cancer risk among smokers. J Natl Cancer Inst 2003;95:470-8.
- Spitz MR, Hong WK, Amos CI, et al. A risk model for prediction of lung cancer. J Natl Cancer Inst 2007;99:715-26.
- Cassidy A, Myles JP, van Tongeren M, et al. The LLP risk model: an individual risk prediction model for lung cancer. Br J Cancer 2008;98:270-6.
- Hoggart C, Brennan P, Tjonneland A, et al. A risk model for lung cancer incidence. Cancer Prev Res (Phila) 2012;5:834-46.

- Tammemagi MC, Katki HA, Hocking WG, et al. Selection criteria for lung-cancer screening. N Engl J Med 2013;368:728-36.
- Marcus MW, Chen Y, Raji OY, Duffy SW, Field JK. LLPi: Liverpool lung project risk prediction model for lung cancer incidence. Cancer Prev Res (Phila) 2015;8:570-5.
- Wilson DO, Weissfeld J. A simple model for predicting lung cancer occurrence in a lung cancer screening program: the Pittsburgh predictor. Lung Cancer 2015;89:31-37.
- Katki HA, Kovalchik SA, Berg CD, Cheung LC, Chaturvedi AK. Development and validation of risk models to select ever-smokers for CT lung cancer screening. JAMA. 2016;315:2300-11. doi: 10.1001/jama.2016.6255.
- Katki HA, Kovalchik SA, Petito LC, Cheung LC, Jacobs E, Jemal A, Berg CD, Chaturvedi AK. Implications of nine risk prediction models for selecting ever-smokers for computed tomography lung cancer screening. Ann Intern Med. 2018;169(1):10-19. doi: 10.7326/M17-2701.
- Cheung LC, Berg CD, Castle PE, Katki HA, Chaturvedi AK. Life-gained-based versus risk-based selection of smokers for lung cancer screening. Ann Intern Med. 2019;171:623-632. doi:107326/M19-1263.

Examples

```
age <- c(66,58,75,72,56)
bmi <- c(23,28,26,27,24)
cpd <- c(36,36,40,24,40)
emp <- c(0,1,1,0,1)
fam.lung.trend <- c(0,2,0,2,0)
female <- c(0,1,0,1,0)
smkyears <- c(43,37,45,42,29)
qtyears <- c(0,0,9,6,6)
race <- c(0,1,2,2,3)
edu6 <- c(3,5,4,5,5)
asb <- c(0,0,0,0,0)
pneu <- c(0,0,0,0,0)
prior.cancer <- c(0,0,0,0,0)
fam.cancer.onset <- c(0,1,0,2,0)
dust <- c(0,0,0,0,0)
fam.cancer <- c(0,1,0,1,0)
fam.smoke.cancer <- c(0,1,0,1,0)
no.hayfever <- c(1,1,1,1,1)
asian <- c(0,0,0,0,1)
islander <- c(0,0,0,0,0)
indian <- c(0,0,0,0,0)
hypertension <- c(0,0,1,0,1)
chd <- c(0,0,0,0,0)
angina <- c(0,0,0,0,0)
heartattack <- c(0,0,0,0,1)
heartdisease <- c(0,0,0,0,0)
stroke <- c(0,0,0,0,0)
diab <- c(1,0,0,0,0)
bron <- c(0,1,0,0,1)
kidney <- c(0,0,0,0,0)
liver <- c(0,0,0,0,0)
speceq <- c(0,1,0,0,0)
year <- rep(2019,5)

persons <- data.frame(age,
```

```
                        female,
                        smkyears,
                        qtyears,
                        cpd,
                        race,
                        emp,
                        fam.lung.trend,
                        bmi,
                        edu6,
                        asb,
                        pneu,
                        prior.cancer,
                        fam.cancer.onset,
                        dust,
                        fam.cancer,
                        fam.smoke.cancer,
                        no.hayfever,
                        asian,
                        islander,
                        indian,
                        hypertension,
                        chd,
                        angina,
                        heartattack,
                        heartdisease,
                        stroke,
                        diab,
                        bron,
                        kidney,
                        liver,
                        speceq,
                        year)

persons_predictions <- lcmodels(persons)
persons_predictions
```

Index