What R we doing in Cancer Research? Examples for cancer prevention

Dr Hannah Lennon

@HannahLennon_

lennonh@iarc.fr

Lyon Data Science

28th February 2019

Why R we attending Lyon Data Science?

I spend my evening with you at Lyon Data Science

1. To see which projects other people are working on
2. To learn new programming hacks/ideas

Why R we attending Lyon Data Science?

I spend my evening with you at Lyon Data Science

1. See which projects other people are working on
   I will give an overview of a current cancer data project
2. Learn new programming hacks/ideas
   I love code so I'll present some of my most used R functions/work flows

About me

2007 Masters in Mathematics

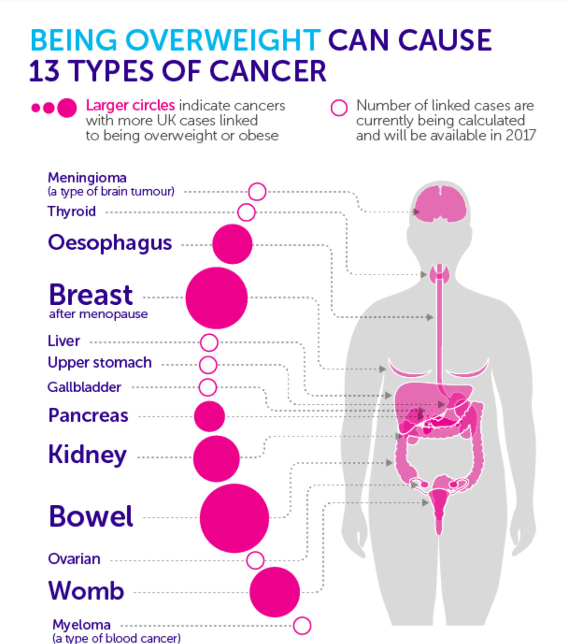
2011 PhD in Mathematics (Computational Statistics)

2015 (2+1) year post-doc in Statistical methodology -
Theme: obesity and **cancer risk/survival**

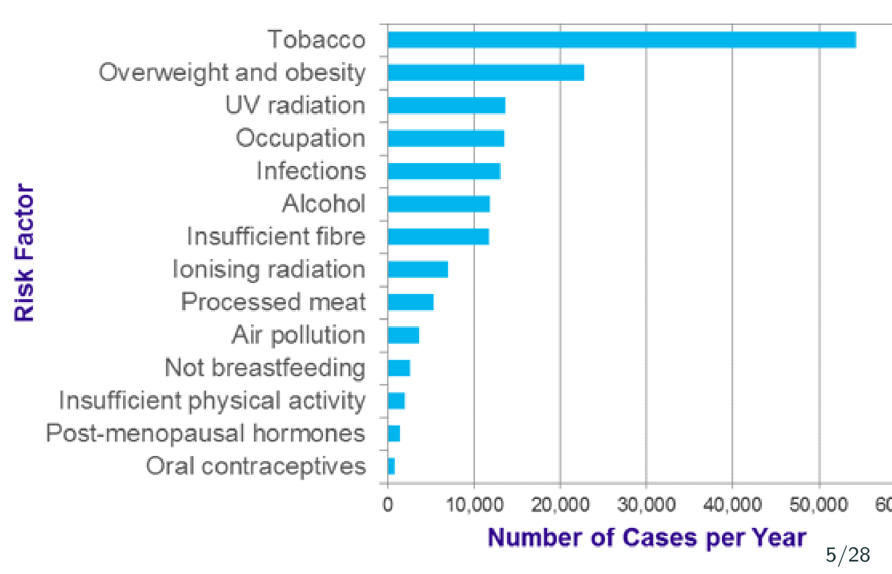
2018 2 year post-doc at IARC in Statistical methodology -
Theme: obesity and cancer & **comorbidities**

Why obesity?

There are 13 obesity-related Cancers



Obesity is the SECOND preventable risk factor of cancer



The problem: For many cancers, development is over many years, even decades

One-off BMI doesn't capture lifetime exposure

Baseline BMI associated with increased risk of several adult cancers

BUT it does not capture the full 'obesity exposure' over life

The problem: For many cancers, development is over many years, even decades

One-off BMI doesn't capture lifetime exposure

Baseline BMI associated with increased risk of several adult cancers

BUT it does not capture the full 'obesity exposure' over life

The problem: For many cancers, development is over many years, even decades

One-off BMI doesn't capture lifetime exposure

Baseline BMI associated with increased risk of several adult cancers

BUT it does not capture the full 'obesity exposure' over life

The problem: For many cancers, development is over many years, even decades

One-off BMI doesn't capture lifetime exposure

Baseline BMI associated with increased risk of several adult cancers
BUT it does not capture the full 'obesity exposure' over life

Obesity Trajectories and cancer risk

Lifetime Trajectories

International Journa

Trajectory of body shape across the lifespan and cancer risk

Mingyang Song[1,2,3,4], Walter C. Willett[3,4,5], Frank B. Hu[3,4,5], Donna Spiegelman[3,4,5,6,7], Aviva Must[8], Kana Wu[3,5], Andrew T. Chan[1,2,5] and Edward L. Giovannucci[3,4,5]

Body shape throughout life and the risk for breast cancer at adulthood in the French E3N cohort

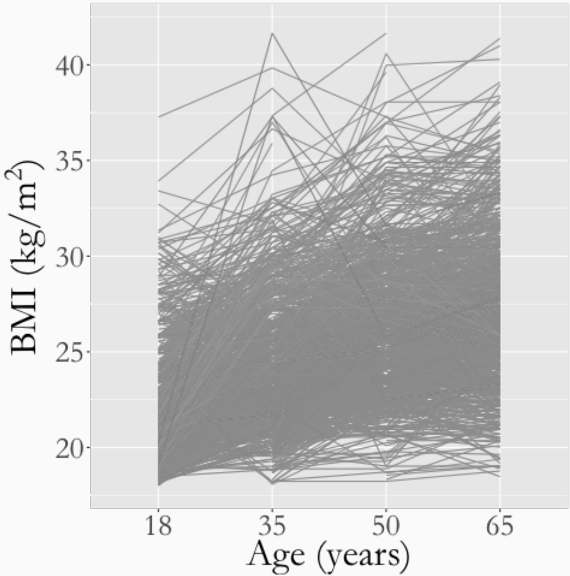
Guy Fagherazzi[a,b], Gwenaelle Guillas[a,b], Marie-Christine Boutron-Ruault[a,b], Françoise Clavel-Chapelon[a,b] and Sylvie Mesrine[a,b]

National Institute of Health AARP data

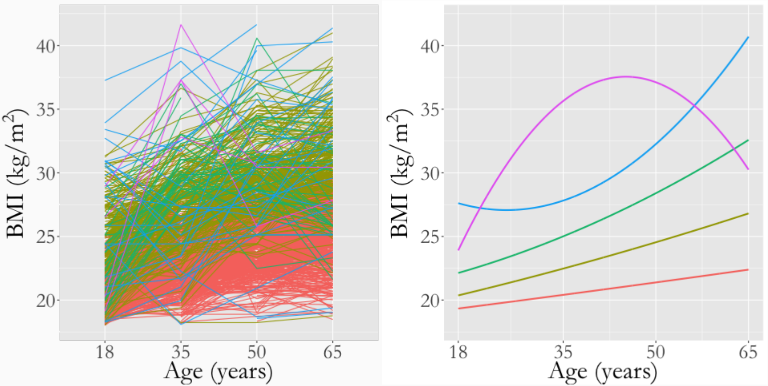
Over 500, 000 men & women, baseline age median 62.5 years, baseline collection- 1995 to 1996, followed to 2012 (last update), detailed lifestyle & dietary cohort

```
> head(dat[,1:12])
  WESTATID CANCER CANCER_DXDT CANCER_SITE CANCER_MORT SEX   DOD UNDERLYINGCOD_BROADGROUP
1 10000056      0          NA                       0   0 14178                       NA
2 10000125      1       15652        C649           1   0 15673                        8
3 10000261      2       17665                       0   0 17665                       10
4 10000352      0          NA                       0   0    NA                       NA
5 10000465      2       17299                       0   1 17299                        6
6 10000603      0          NA                       0   1    NA                       NA
  ENTRY_DT ENTRY_AGE EXIT_DT_ALLCAUSEMORT EXIT_AGE_ALLCAUSEMORT
1    13319    57.632                14178                59.984
2    13326    69.155                15673                75.581
3    13318    63.086                17665                74.987
4    13212    65.281                18262                79.107
5    13101    62.475                17299                73.969
6    13097    64.898                18262                79.039
>
>
```

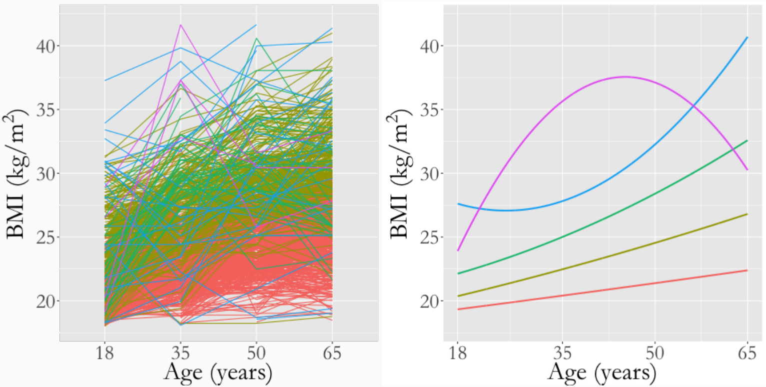
Latent classes derived depend on underlying assumptions



Latent classes derived depend on underlying assumptions



Latent classes derived depend on underlying assumptions



- Framework to construct and interpret latent class trajectory modelling (Lennon et al., 2018)
- The GRoLTS Checklist: Guidelines for Reporting on Latent Trajectory Studies (van de Schoot et al., 2017)

The latent class linear mixed model

The model is an extension of the linear mixed model of Laird & Ware (1982) for individuals $i = 1, \ldots, N$ at times $j = 1, \ldots, T$, for $k = 1, \ldots, K$ classes

$$Y_{ij|c_i=k} = X_i(t_{ij})\beta_k + Z_{ik}(t_{ij})u_{ik} + w_k(t_{ij}) + \varepsilon_{ijk}, \qquad (1)$$
$$\varepsilon_{ij} \overset{\text{iid}}{\sim} N(0, R).$$

where both the fixed effects and the distribution of the random effects can be class-specific.

The latent class linear mixed model

The model is an extension of the linear mixed model of Laird & Ware (1982) for individuals $i = 1, \ldots, N$ at times $j = 1, \ldots, T$, for $k = 1, \ldots, K$ classes

$$Y_{ij|c_i=k} = X_i(t_{ij})\beta_k + Z_{ik}(t_{ij})u_{ik} + w_k(t_{ij}) + \varepsilon_{ijk}, \tag{1}$$
$$\varepsilon_{ij} \overset{\text{iid}}{\sim} N(0, R).$$

where both the fixed effects and the distribution of the random effects can be class-specific.

The latent class linear mixed model

The model is an extension of the linear mixed model of Laird & Ware (1982) for individuals $i = 1, \ldots, N$ at times $j = 1, \ldots, T$, for $k = 1, \ldots, K$ classes

$$Y_{ij|c_i=k} = X_i(t_{ij})\beta_k + Z_{ik}(t_{ij})u_{ik} + w_k(t_{ij}) + \varepsilon_{ijk}, \qquad (1)$$
$$\varepsilon_{ij} \overset{\text{iid}}{\sim} N(0, R).$$

where both the fixed effects and the distribution of the random effects can be class-specific.

The latent class linear mixed model

The model is an extension of the linear mixed model of Laird & Ware (1982) for individuals $i = 1, \ldots, N$ at times $j = 1, \ldots, T$, for $k = 1, \ldots, K$ classes

$$Y_{ij|c_i=k} = X_i(t_{ij})\beta_k + Z_{ik}(t_{ij})u_{ik} + w_k(t_{ij}) + \varepsilon_{ijk}, \qquad (1)$$
$$\varepsilon_{ij} \overset{\text{iid}}{\sim} N(0, R).$$

where both the fixed effects and the distribution of the random effects can be class-specific.

Each subject is allowed to vary about the mean trajectory of its class.

The latent class linear mixed model

The model is an extension of the linear mixed model of Laird & Ware (1982) for individuals $i = 1, \ldots, N$ at times $j = 1, \ldots, T$, for $k = 1, \ldots, K$ classes

$$Y_{ij|c_i=k} = X_i(t_{ij})\beta_k + Z_{ik}(t_{ij})u_{ik} + w_k(t_{ij}) + \varepsilon_{ijk}, \tag{1}$$
$$\varepsilon_{ij} \overset{\text{iid}}{\sim} N(0, R).$$

where both the fixed effects and the distribution of the random effects can be class-specific.

Each subject is allowed to vary about the mean trajectory of its class.

We let the discrete random variable c_i follow a multinomial logistic model according to covariates X_i,

$$Pr(c_i = k|X_i) = \frac{\exp(\xi_{0k} + X_i^\top \xi_k)}{\sum_{l=1}^{K} \exp(\xi_{0l} + X_i^\top \xi_l)}. \tag{2}$$

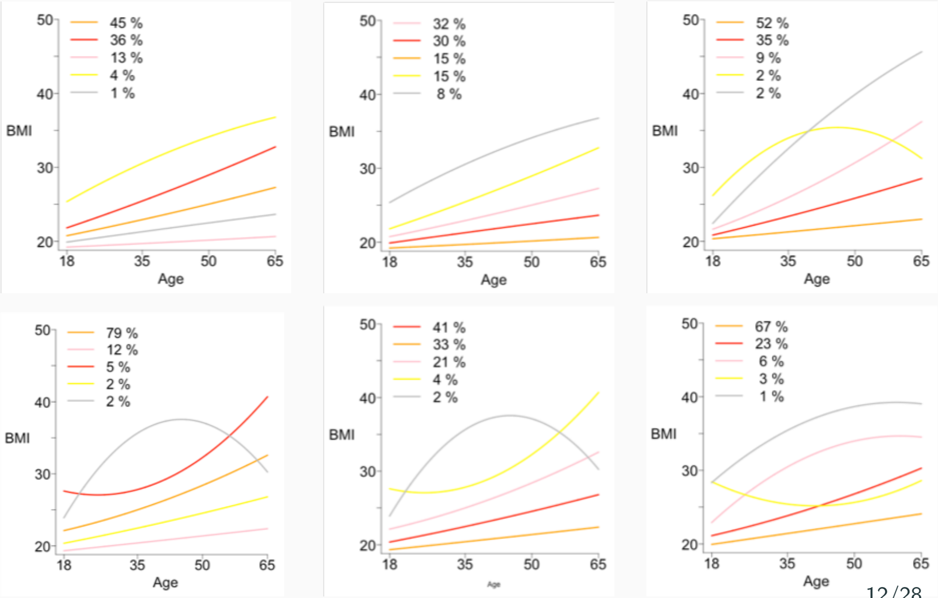
R or SAS software can be used to fit these models

Table S1 Different models with increasing complexity, associated software and referenced use in literature

Model	Description	Interpretation	Software *Command*
A	Fixed effects homoscedastic (common residual variance across classes)	No random effects – with the interpretation that any deviation of an individuals trajectory from its mean class trajectory is due to random error only	SAS traj *PROC TRAJ*
B	Fixed effects heteroscedastic (class-specific residual variances)	The same interpretation as Model A with random errors that can be larger and smaller in different classes.	R mmlcr *mmlcr*
C	Random intercept	The interpretation is allowing individuals to vary in initial weight but each class member is assumed to follow the same shape and magnitude of the mean trajectory	SAS traj *PROC TRAJ*
D	Random slope	Allowing individuals to vary in initial weight and slope of the mean trajectory but same curvature as trajectory	SAS traj *PROC TRAJ*
E	Random quadratic – Common variance structure across classes	Additional freedom of allowing individuals to vary within classes by initial weight, shape and magnitude, however each class is assumed to have the same amount of variability	R lcmm *hlme/lcmm*
F	Random quadratic – Proportionality constraint to allow variance structures to vary across classes	Increasing flexibility of model E as variance structures are allowed to differ up to a multiplicative factor to allow some classes to have larger or smaller within-class variances. This model is can be thought of more parsimonious version of model G from (reducing the number of variance-covariance parameters to be estimated from 6xK parameters to 6+(K-1) parameters.	R lcmm *hlme/lcmm*
G	Random quadratic – Class-specific variance structure (unstructured)	The most flexible model in which each class has its own separate random quadratic variance structure to describe its own within-class variability. Statistically this permits the variance and covariance of the intercept, slope and quadratic term to vary freely across all classes.	SAS traj *PROC TRAJ*

[1]The SAS traj package has been converted for Stata users as the traj command in Stata (Collage Station, TX, USA).

Different modelling assumptions lead to different trajectories

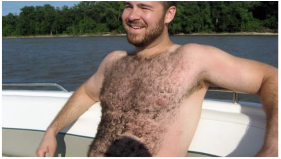


Different shapes & proportions SO PREFERRED MODEL IS?

Results were reported in the worldwide headlines

Beer belly increases cancer risk by 50%

HEALTH & WELLNESS / 8 NOVEMBER 2016, 07:00AM / DAILY MAIL



File Photo

MIDDLE-AGED SPREAD,AKA A BEER BELLY, IN MEN CAN RAISE THEIR RISK OF GETTING CANCER BY 50 PER CENT, A STUDY HAS FOUND.

Men need to gain only 2st 7lb from their teenage years to retirement to be in significantly greater danger. For women, the risk of cancer rises by almost 20 per cent over a lifetime if they put on 3st 7lb.

How middle-age spread increases cancer risk by 50%: Men need to gain only 2st 7lbs from teenage years to retirement to raise the risk

- **Study to share findings at National Cancer Research Institute conference**
- **Its findings have shed new light on how weight gain can cause cancer**
- **Men need to gain only 2st 7lb whilst women need to put on 3st 7lb**

By VICTORIA ALLEN SCIENCE CORRESPONDENT FOR THE DAILY MAIL
PUBLISHED: 00:15 GMT, 7 November 2016 | **UPDATED:** 19:48 GMT, 7 November 2016

Share 96 shares 89 View comments

Middle-age spread in men can raise their risk of getting cancer by 50 per cent, a study has found.

Men need to gain only 2st 7lb from their teenage years to retirement to be in significantly greater danger. For women, the risk of cancer rises by almost 20 per cent over a lifetime if they put on 3st 7lb.

The findings, to be presented at a National Cancer Research Institute conference in Liverpool today, shed new light on how weight gain causes cancer – most often breast and womb cancer in women and kidney and bowel cancer in both sexes.

medic was a junior doctor in specialist training.

At Pontefract Hospital, a GP was left in charge of A&E from midnight to 8am all week. The only hospitals shown to provide consultant cover throughout the period were Royal London Hospi-

an A&E consultant available by telephone.

Calderdale and Huddersfield NHS trust, which runs Huddersfield Hospital, said it could only provide seven-day cover if it centralised services on one site.

Women see the signs of killer illness

WOMEN are much better at identifying the signs of major illnesses than men, research has found.

In a poll of more than 8,000 over-50s, men were less likely than women to say they knew the signs of bowel cancer (54 per cent compared to 71 per cent of women), lung cancer (33 to 45 per cent) and skin cancer (56 to 73 per cent).

Four in 10 men did not know what to look for when it came to prostate cancer.

When asked what would prompt them to go to the GP, 63 per cent of men would go to their doctor if they discovered a new lump, compared

Obese men have doubled cancer risk

MEN who put on weight over several decades are more likely to get obesity-related cancers than women who do the same, research has shown.

Weight gain increased the risk of obesity-related cancers by 50 per cent in men and almost 20 per cent in women, say researchers from the University of Manchester.

The study looked at the link between obesity and cancer in 300,000 Americans, monitoring their weight between the ages of 18 and 65, then recording cancer rates.

Men whose body mass index (BMI) rose from 22 to 27 had a 50 per cent increased risk of developing obesity-related cancer compared to

Friday, Mar 1st 2019

MailOnline health

Home | News | U.S. | Sport | TV&Showbiz | Australia | Femail | Health | Science | Money | Video | Travel | DailyMailTV

Latest Headlines | Health | Health Directory | Discounts

How middle-age spread increases cancer risk by 50%: Men need to gain only 2st 7lbs from teenage years to retirement to raise the risk

- Study to share findings at National Cancer Research Institute conference
- Its findings have shed new light on how weight gain can cause cancer
- Men need to gain only 2st 7lb whilst women need to put on 3st 7lb

By VICTORIA ALLEN SCIENCE CORRESPONDENT FOR THE DAILY MAIL

PUBLISHED: 00:15 GMT, 7 November 2016 | UPDATED: 19:48 GMT, 7 November 2016

Daily Express Monday November 7 2016

UK's soaring obesity crisis is creating a cancer time bomb

ALARM AT TOLL OF MEN WHO MISS SIGNS OF KILLER DISEASES

By Mark Reynolds

By John von Radowitz

Children whose mothers were undernourished during pregnancy may have prematurely ageing hearts, a new study has suggested.

Scientists made the discovery after studying pregnant baboons and their young. They were chosen because their hearts develop and age in much the same way as those of humans.

Offspring of baboons with a restricted diet showed signs of accelerated heart function with age, according to the research published in The Journal of Physiology. By five years old, equivalent to 20 human years, the structure and performance of their hearts were already impaired.

The same problem could affect children in developed countries as well as poorer developing countries, say the researchers. Food charity the Trussell Trust has found that more than one in five UK parents struggle to feed their children.

Dr Peter Nathanielsz, who co-led the study, said: "Society must pay attention to improving women's nutrition before and during pregnancy to prevent these adverse outcomes in babies."

The study may have implications for pregnant mothers PA

HEALTH

Men at increased risk of obesity-related cancers

By Jane Kirby

Piling on the pounds over several decades increases the risk of obesity-related cancers in men by 50 per cent and in women by almost 20 per cent, research has shown.

British scientists made the discovery after studying data from about 300,000 Americans, who were monitored between the ages of 18 and 65.

Being overweight or obese is linked to a wide range of cancers – including bowel, breast and pancreatic cancer.

A separate poll of more than 8,000 British over-50s found that women said they knew the signs for more than half of illnesses, while on average men knew only about a third.

NEWS MEDICAL LIFE SCIENCES

MEDICAL HOME | LIFE SCIENCES HOME | Become a Member | Search...

About | News | Health A-Z | Drugs | Medical Devices | Interviews | White Papers | MediKnowledge | More...

Long-term weight gain may increase risk of obesity-related cancers

Download PDF Copy

Nov 6 2016

Substantial weight gain over many years increases the risk of obesity-related cancers in men by 50 per cent and in women by almost 20 per cent, according to new research presented at the National Cancer Research Institute's (NCRI) Cancer Conference in Liverpool, today (Monday).

Researchers at The University of Manchester and The Health eResearch Centre, looked at weight gain over many years and assessed the risk of developing obesity-related cancers.

This is a new way of looking at the long-term impact of being obese throughout a person's life and the link to developing cancer.

In the study of approximately 300,000 people in America, including around 177,500 men and 111,500 women, researchers categorised the population into five different lifetime weight trajectories. They looked at changes in BMI between the ages of 18 and 65.

Some people gained a little weight between the ages of 18 and 65 years, while others became morbidly obese.

Trending Stories | Latest Interviews | Top Health Articles

Plant-based diet increases secretion of insulin, incretin hormones in people with type 2 diabetes

Virtual reality shows promise for reducing fears and phobia in autistic adults

Migrants' integration into host community involves trade-off between social status and fertility

The story featured news outlets worldwide including translated thai, french, german, australian, czech, e.t.c.

A video description at www.ecancer.org

Model selection tools for Latent Class Trajectory Models

Likelihood based measures such as Information Criteria
can be used among a finite set of models **such that a lower is preferred.**

Akaike Information Criteria, AIC

$$\mathrm{AIC} = 2k - 2\ln(\hat{L}),$$

Bayesian Information Criteria, BIC

$$\mathrm{BIC} = \ln(n)k - 2\ln(\hat{L}),$$

- Likelihood Ratio Tests (LRT),
- Lo-Mendel-Rubin (LMT-LRT),
- Bootstrap-LRT

Model selection tools for Latent Class Trajectory Models

APPA $> 70\%$ for each class

The Average Posterior Probability Assignment, $\hat{p}_k$, looks at whether individuals are assigned with high probability and the overall average probability of assignment to each class.

OCC > 5.0 for each class

The Odds of Correct Classification (OCC) is the ratio of the odds of a correct classification into each group on the basis of the maximum probability classification rule and the estimated class membership proportions,

$$\mathrm{OCC}_k = \frac{\hat{p}_k(1 - \hat{p}_k)}{\hat{\pi}_k(1 - \hat{\pi}_k)}. \tag{3}$$

Model selection tools for Latent Class Trajectory Models

Mismatch, Close to 0 for each class

The difference between the estimated class proportions and the class membership proportions once individuals have been assigned to a class, i.e.

$$\mathrm{Mismatch}_k = \hat{\pi}_k - \frac{N_k}{N},$$

where N_k is the number of individuals in a class and N is the total.

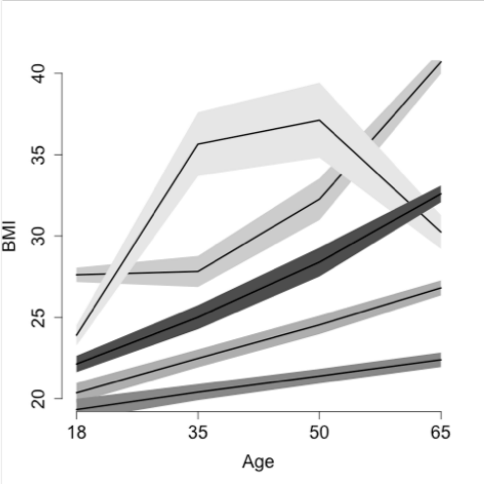
Entropy, Close to 0

Entropy is a global measure of classification uncertainty, which takes into account all $N \times K$ posterior probabilities, defined as

$$E = -\sum_{i=1}^{N}\sum_{k=1}^{K} p_{ik} \log p_{ik}$$

which takes values from $[0, \infty)$, with higher values indicating a larger amount of uncertainty. Entropy values closest to 0 correspond to models

Graphical plots of Elsensohn's Residuals

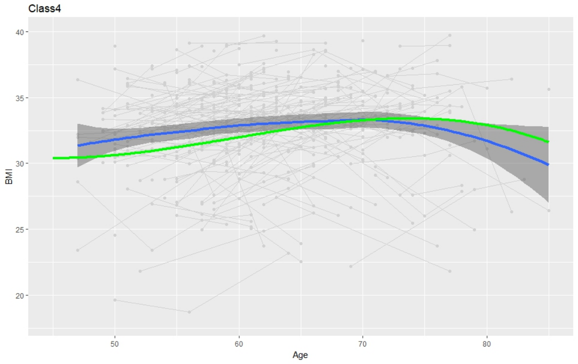


[1]

How do we know if these trajectories are model based or data based?

PhD student Charlotte Watson, University of Manchester

Adulthood BMI trajectories And Cancer using clusters [ABACus]



LCTMtools: an R package

Freely available on github to compute these commands.

To install the R package, in the R console use the command

```
install.packages("devtools")
devtools::install_github("hlennon/LCTMtools")
```

LCTMtools: an R package

Freely available on github to compute these commands.

To install the R package, in the R console use the command

```
install.packages("devtools")
devtools::install_github("hlennon/LCTMtools")
```

LCTMtools: an R package

```
> library(LCTMtools)
> data(bmi_long, package = "LCTMtools" )
>
>
> # Use the hlme function from the 'lcmm' R package to fit a 2 class latent class trajectory model
> set.seed(002010800)
> library(lcmm)
Loading required package: survival
> model2classes <- hlme(fixed = BMI ~ Age + I(Age^2),
+                      mixture= ~ Age,
+                      random = ~ Age,
+                      ng = 2,
+                      nwg = TRUE,
+                      subject = "ID",
+                      data = bmi_long[1:500, ] )
Be patient, hlme is running ...
The program took 3.78 seconds
>
>
> # Compute model adequacy measures
> LCTMtoolkit(model2classes)
$`Class-specific`
         Class_1 Class_2   Recommendation
APPA       0.982   0.782 Greater than 0.7
OCC        8.430  23.107   Greater than 5
Mismatch  -0.018   0.018    Close to zero

$`Model-specific`
                    Model Recommendation
Entropy            13.620  Close to zero
Relative_entropy    0.843     Close to 1
BIC              2878.722              -
AIC              2847.610              -
```

LCTMtools: an R package

LCTMtools

Latent Class Trajectory Modelling Tools: an R Package

Maintainer: Hannah Lennon
Contact: lennonh@iarc.fr
Last Updated: 11th February 2019

To install the R package, in the R console use the command

```
devtools::install_github("hlennon/LCTMtools")
```

All statistical (R and SAS) codes used to implement Latent Class Trajectory Modelling and the tools described in the manuscript "A framework to construct and interpret Latent Class Trajectory Modelling", are available here and can be downloaded from www.github.com/hlennon/LCTMtools.

An example (simulated) dataset 'bmi' and 'bmi_long' (long format version) is provided to describe the steps throughout.

Reference

Lennon H, Kelly S, Sperrin M, et al., Framework to construct and interpret Latent Class Trajectory Modelling, BMJ Open 2018;8:e020683.

Available at https://bmjopen.bmj.com/content/8/7/e020683

Help Files

Link to the paper: Framework to construct and interpret latent class trajectory modelling

Framework to construct and interpret latent class trajectory modelling, BMJ Open

Lennon H, Kelly S, Sperrin M, et al. (2018) [2]

Link to Reporting Guidelines paper:

The GRoLTS-Checklist: Guidelines for Reporting on Latent Trajectory Studies

Rens van de Schoot et al., (2017) [3]

Our 1st R Ladies Lyon Event: COMING SOON

April 2019

meetup.com/RLadies-Lyon

twitter.com/RLadiesLyon

Thank you to

Miss Charlotte Watson

Dr Matthew Sperrin

Prof Andrew Renehan

National Institute of Health AARP

Cancer Research UK (Funding Body)

Questions?

M.-H. Elsensohn, A. Klich, R. Ecochard, M. Bastard, C. Genolini, J.-F. Etard, and M.-P. Gustin.
A graphical method to assess distribution assumption in group-based trajectory models.
Statistical Methods in Medical Research, 25(2):968–982, 2016.

PMID: 23427224.

H. Lennon, S. Kelly, M. Sperrin, I. Buchan, A. J. Cross, M. Leitzmann, M. B. Cook, and A. G. Renehan.
Framework to construct and interpret latent class trajectory modelling.
BMJ Open, 8(7), 2018.

R. van de Schoot, M. Sijbrandij, S. D. Winter, S. Depaoli, and J. K. Vermunt.
The grolts-checklist: Guidelines for reporting on latent trajectory studies.
Structural Equation Modeling: A Multidisciplinary Journal, 24(3):451–467, 2017.

What R we doing in Cancer Research?
Examples for cancer prevention

Dr Hannah Lennon

@HannahLennon_

lennonh@iarc.fr

Lyon Data Science

28th February 2019