

```

1  /*
2
3  PHD PROJECT: The role of depressive symptoms and cardiometabolic risk factors in the prediction
4  of dementia: a cross-country comparison in England, the United States and China
5
6  STUDY 3: Trajectories of depressive symptoms and their relationship with dementia
7
8  Method of analysis:
9  Group-based trajectory modelling (GBTM) approach
10 Latent Class Growth Analysis (LCGA)
11
12 TIMELINE
13
14 DEPRESS TRAJECTORIES: WV2 - WV5 (4 TIME POINTS)
15 DEMENTIA INCIDENCE AT YEAR 6 (WV5)
16 DEMENTIA INCIDENCE: W6 - WV9 (4 TIME POINTS)
17
18 */
19
20
21
22
23 * importing data (.dta)
24
25 use "S:\Research\pkhec\Study1_traj_depression\ELSA\1. traj depression and dementia
26 risk\elsa_depress_traj_recode_cesd_time_wv2_5.dta"
27
28
29
30 * KEEP NECESSARY VARIABLES
31
32 keep idauniq w2wtnur w2wtbld ///
33 E_sex E_age E_education_yrs E_education E_maritalstatus_3cat E_maritalstatus_4cat ///
34 E_wealthquintiles E_smoking_3cat E_physicalactivity E_alcohol_freq E_alcohol_status ///
35 E_cvd_comorbidity E_cognitive_index ///
36 Ewv2_cesd_happy_rand Ewv2_cesd_enlife_rand Ewv2_cesd_depressed_rand Ewv2_cesd_effort_rand
37 Ewv2_cesd_sleep_rand Ewv2_cesd_lonely_rand Ewv2_cesd_sad_rand Ewv2_cesd_going_rand ///
38 Ewv2_cesd_sumscore_raw Ewv2_cesd_sumitems cesd_0 Ewv2_cesd_score Ewv2_depressive_symptoms ///
39 Ewv3_cesd_happy_rand Ewv3_cesd_enlife_rand Ewv3_cesd_depressed_rand Ewv3_cesd_effort_rand
40 Ewv3_cesd_sleep_rand Ewv3_cesd_lonely_rand Ewv3_cesd_sad_rand Ewv3_cesd_going_rand cesd_1
41 Ewv3_depressive_symptoms ///
42 Ewv4_cesd_happy_rand Ewv4_cesd_enlife_rand Ewv4_cesd_depressed_rand Ewv4_cesd_effort_rand
43 Ewv4_cesd_sleep_rand Ewv4_cesd_lonely_rand Ewv4_cesd_sad_rand Ewv4_cesd_going_rand cesd_2
44 Ewv4_depressive_symptoms ///
45 Ewv5_cesd_happy_rand Ewv5_cesd_enlife_rand Ewv5_cesd_depressed_rand Ewv5_cesd_effort_rand
46 Ewv5_cesd_sleep_rand Ewv5_cesd_lonely_rand Ewv5_cesd_sad_rand Ewv5_cesd_going_rand cesd_3
47 Ewv5_depressive_symptoms ///
48 Ewv6_cesd_happy_rand Ewv6_cesd_enlife_rand Ewv6_cesd_depressed_rand Ewv6_cesd_effort_rand
49 Ewv6_cesd_sleep_rand Ewv6_cesd_lonely_rand Ewv6_cesd_sad_rand Ewv6_cesd_going_rand
50 Ewv6_cesd_sumscore_rand Ewv6_depressive_symptoms ///
51 Ewv7_cesd_happy_rand Ewv7_cesd_enlife_rand Ewv7_cesd_depressed_rand Ewv7_cesd_effort_rand
52 Ewv7_cesd_sleep_rand Ewv7_cesd_lonely_rand Ewv7_cesd_sad_rand Ewv7_cesd_going_rand
53 Ewv7_cesd_sumscore_rand Ewv7_depressive_symptoms ///
54 Ewv8_cesd_happy Ewv8_cesd_enlife Ewv8_cesd_depressed Ewv8_cesd_effort Ewv8_cesd_sleep
55 Ewv8_cesd_lonely Ewv8_cesd_sad Ewv8_cesd_going Ewv8_cesd_sumscore Ewv8_depressive_symptoms ///
56 Ewv9_cesd_happy Ewv9_cesd_enlife Ewv9_cesd_depressed Ewv9_cesd_effort Ewv9_cesd_sleep
57 Ewv9_cesd_lonely Ewv9_cesd_sad Ewv9_cesd_going Ewv9_cesd_sumscore Ewv9_depressive_symptoms ///
58 Ewv6to9_dementia_sum Ewv2_anydementia_iqcode_report Ewv3_anydementia_iqcode_report
59 Ewv4_anydementia_iqcode_report Ewv6to9_dementia_event Ewv5_anydementia_iqcode_report
60 Ewv6_anydementia_iqcode_report Ewv7_anydementia_iqcode_report Ewv8_anydementia_iqcode_report
61 Ewv9_anydementia_iqcode_report ///
62 Ewv2_interview_date Ewv3_interview_date Ewv4_interview_date Ewv5_interview_date
63 Ewv6_interview_date Ewv7_interview_date Ewv8_interview_date Ewv9_interview_date ///
64 Ewv6to9_newdementia_or_lastinter Ewv6to9_time_dementia_months Ewv6to9_dementia_free_date
65 Ewv6to9_time_dementia_midpoint Ewv6to9_time_dementia_midpoint_f Ewv6to9_time_event_dementia ///
66 t_0 t_1 t_2 t_3 nmisfollowup_cesd nmisfollowup_dementia_wv6to9

```

```

49
50
51
52
53
54 /* Latent class growth analysis (LCGA) of depressive symptoms */
55
56
57 * installing traj command
58
59 net from https://www.andrew.cmu.edu/user/bjones/traj
60 net install traj
61 help traj
62
63
64 * Generate a set of time variables to pass to traj, from wave 2 to 9 -> 8 time points (t0-t7)
65 forval i = 0/3 {
66     generate t_`i' = `i'
67 }
68
69 *recode time in months
70
71 recode t_1 (1=24)
72 recode t_2 (2=48)
73 recode t_3 (3=72)
74
75
76 *rename cesd score across the waves - discrete var min=0 max=8
77
78 rename Ewv2_cesd_sumscore_rand cesd_0
79 rename Ewv3_cesd_sumscore_rand cesd_1
80 rename Ewv4_cesd_sumscore_rand cesd_2
81 rename Ewv5_cesd_sumscore_rand cesd_3
82
83
84
85
86
87
88 *** Descriptive stats of depression and dementia
89
90 tabulate cesd_0
91 summarize cesd_0 , detail
92 histogram cesd_0, discrete frequency normal
93
94 misstable summarize cesd_0
95 misstable patterns cesd_0
96
97 tabulate Ewv2_depressive_symptoms
98 summarize Ewv2_depressive_symptoms
99
100 misstable summarize Ewv2_depressive_symptoms
101 misstable patterns Ewv2_depressive_symptoms
102
103 tabulate Ewv2_anydementia_iqcode_report
104 summarize Ewv2_anydementia_iqcode_report
105
106 misstable summarize Ewv2_anydementia_iqcode_report
107 misstable patterns Ewv2_anydementia_iqcode_report
108
109
110 tabulate Ewv3_anydementia_iqcode_report
111 summarize Ewv3_anydementia_iqcode_report
112
113 misstable summarize Ewv3_anydementia_iqcode_report
114 misstable patterns Ewv3_anydementia_iqcode_report
115
116

```

```

117 tabulate Ewv4_anydementia_iqcode_report
118 summarize Ewv4_anydementia_iqcode_report
119
120 misstable summarize Ewv4_anydementia_iqcode_report
121 misstable patterns Ewv4_anydementia_iqcode_report
122
123 tabulate Ewv5_anydementia_iqcode_report
124 summarize Ewv5_anydementia_iqcode_report
125
126 misstable summarize Ewv5_anydementia_iqcode_report
127 misstable patterns Ewv5_anydementia_iqcode_report
128
129
130 tabulate Ewv6_anydementia_iqcode_report
131 summarize Ewv6_anydementia_iqcode_report
132
133 misstable summarize Ewv6_anydementia_iqcode_report
134 misstable patterns Ewv6_anydementia_iqcode_report
135
136
137 tabulate Ewv7_anydementia_iqcode_report
138 summarize Ewv7_anydementia_iqcode_report
139
140 misstable summarize Ewv7_anydementia_iqcode_report
141 misstable patterns Ewv7_anydementia_iqcode_report
142
143 tabulate Ewv8_anydementia_iqcode_report
144 summarize Ewv8_anydementia_iqcode_report
145
146 misstable summarize Ewv8_anydementia_iqcode_report
147 misstable patterns Ewv8_anydementia_iqcode_report
148
149
150 tabulate Ewv9_anydementia_iqcode_report
151 summarize Ewv9_anydementia_iqcode_report
152
153 misstable summarize Ewv9_anydementia_iqcode_report
154 misstable patterns Ewv9_anydementia_iqcode_report
155
156
157
158
159
160
161
162 *** CLEANING DATA
163
164 * 1. drop missing data depression and dementia at baseline
165 * drop 79 depression missing data
166
167 drop if cesd_0== .
168 * (79 observations deleted)
169
170
171 * 2. drop dementia cases between wv2 and wv5 (total: 235 cases)
172
173 drop if Ewv2_anydementia_iqcode_report!=
174 * (44 observations deleted)
175
176 drop if Ewv3_anydementia_iqcode_report!=
177 * (50 observations deleted)
178
179 drop if Ewv4_anydementia_iqcode_report!=
180 * (65 observations deleted)
181
182
183
184 * 3. process to drop missing data depression in at least 2 follow-up waves

```

```

185
186  /*
187
188  check below how to see number of missing values in an observation (case) and patterns of missing
values
189  https://stats.idre.ucla.edu/stata/faq/how-can-i-see-the-number-of-missing-values-and-patterns-of-mi
ssing-values-in-my-data-file/
190  install packages:
191  * install mdesc
192  * install tabmiss
193  * insatll dm31
194  * insall mvpatterna
195
196  */
197
198  search mdesc
199  search rmiss2
200  search mvpatterns
201
202
203  * see number of missing values vs non-missing in each variable
204  mdesc cesd_0 cesd_1 cesd_2 cesd_3
205  *or
206  mdesc cesd_*
207
208
209  * number of missing values per observation
210  * the code below creates a variable called nmisfollowup that gives the number of missing values
for each observation in the variables of interest
211  egen nmisfollowup_cesd=rmiss2(cesd_1 cesd_2 cesd_3)
212
213  tab nmisfollowup_cesd
214
215  * drop observations "nmisfollowup_cesd" > 1 (those with 2 or 3 missing data)
216  drop if nmisfollowup>1
217  *(1804 observations deleted)
218
219
220
221
222
223  * 4. drop obs with no records on dementia at any wave from 6-9 follow-ups
224
225  * see number of missing values vs non-missing in each variable
226  mdesc Ewv6_anydementia_iqcode_report Ewv7_anydementia_iqcode_report ///
227  Ewv8_anydementia_iqcode_report Ewv9_anydementia_iqcode_report
228
229
230
231  /* number of missing values per observation
232  * the code below creates a variable called nmisfollowup that gives the number of missing values
for each observation in the variables of interest */
233
234  egen nmisfollowup_dementia_wv6to9=rmiss2(Ewv6_anydementia_iqcode_report
Ewv7_anydementia_iqcode_report ///
235  Ewv8_anydementia_iqcode_report Ewv9_anydementia_iqcode_report)
236
237  tab nmisfollowup_dementia_wv6to9
238
239  * drop observations "nmisfollowup_dementia_wv6to9" > 3 (those with 4 missing data = no records at
any wave)
240  drop if nmisfollowup_dementia_wv6to9>3
241  *(834 observations deleted)
242
243
244
245
246  *descriptive stats of depressive symptoms cesd
247

```

```

248
249
250 tabulate cesd_0
251 summarize cesd_0, detail
252 histogram cesd_0, discrete frequency normal
253
254 tabulate cesd_1
255 summarize cesd_1 , detail
256 histogram cesd_1, discrete frequency normal
257
258 tabulate cesd_2
259 summarize cesd_2, detail
260 histogram cesd_2, discrete frequency normal
261
262 tabulate cesd_3
263 summarize cesd_3, detail
264 histogram cesd_3, discrete frequency normal
265
266
267 ta cesd_0, miss
268 ta cesd_1, miss
269 ta cesd_2, miss
270 ta cesd_3, miss
271
272 tabstat cesd_0, by(E_sex)stats (mean v n)
273 tabstat cesd_1, by(E_sex)stats (mean v n)
274 tabstat cesd_2, by(E_sex)stats (mean v n)
275 tabstat cesd_3, by(E_sex)stats (mean v n)
276
277 tabstat cesd_*,s(n me sk) by(E_sex)
278
279
280 bysort E_sex: tab cesd_0
281 bysort E_sex: tab cesd_1
282 bysort E_sex: tab cesd_2
283 bysort E_sex: tab cesd_3
284
285
286 tabstat cesd_0 cesd_1 cesd_2 cesd_3, s(sk kur)
287 sktest cesd_0 cesd_1 cesd_2 cesd_3
288
289
290 * missingness pateterns
291 misstable patterns cesd_*
292 * "1" means that the variable is observed and a "0" represents missing
293
294
295 * box plots of the observations at each occasion
296 graph box cesd_0 cesd_1 cesd_2 cesd_3, ascategory intensity (0) medtype (line)
297
298
299
300
301
302
303
304
305 /*
306 LCGA
307 longitudinal analysis of trajectories
308 GBTM model
309
310 useful sources
311 use http://www.andrew.cmu.edu/user/bjones/traj/data/cambrdge.dta,clear
312 https://ssrc.indiana.edu/doc/wimdocs/2013-03-29\_nagin\_trajectory\_stata-plugin-info.pdf
313
314
315 traj [if], var(varlist) indep(varlist) model(modeltype)

```

```

316         order(numlist) [additional options]
317
318
319     order(numlist)      0=intercept, 1=linear, 2=quadratic, 3=cubic -
320                        polynomial type for each group trajectory
321
322
323     ci                  parametric bootstrap confidence intervals of
324                        individual distal outcome and probability of
325                        group memberships.
326
327
328
329
330
331
332     Available Models -> command traj
333
334
335     Censored normal (CNORM) model distribution
336
337     traj, var(qcp*op) indep(age*) model(cnorm) min(0) max(999) order(1 3 2)
338
339     trajplot, xtitle(Age) ytitle(Opposition) xlabel(6(1)15)
340                ylabel(0(1)6)
341
342     /* Shows the assigned group and probabilities of group membership */
343         list _traj_Group - _traj_ProbG3 if _n < 3, ab(12)
344
345     /* trajT = x-axis, Avg# = data averages, Est# = model estimates */
346         matrix list e(plot1), format(%9.2f) noheader
347
348     /* Including time-stable covariates (risk) associated with group membership */
349
350     traj, var(qcp*op) indep(age*) model(cnorm) min(0) max(10) order(1 3
351                2) risk(scolmer scolper)
352
353
354
355     Zero inflated poisson (ZIP) Model
356
357     It is an analysis of Poisson data with extra zeros
358
359
360     traj, model(zip) var(y*) indep(t*) order(2 1 3) iorder(1)
361
362     trajplot, xtitle(Age) ytitle(Opposition) ci
363
364
365
366     Time-Stable Covariates for Group Membership
367
368         traj, var(qcp*op) indep(age*) model(cnorm) min(0) max(10) order(1 3
369                2) risk(scolmer scolper)
370
371         trajplot, xtitle(Age) ytitle(Opposition)
372
373
374     Logistic (logit) model
375
376         use http://www.andrew.cmu.edu/user/bjones/traj/data/cambridge.dta,
377             clear
378
379         traj, var(p1-p23) indep(tt1-tt23) model(logit) order(0 3 3)
380
381         trajplot, xtitle(Scaled Age) ytitle(Prevalence)
382
383         /* Assigned group and probabilities of group membership */

```

```

384     list _traj_Group - _traj_ProbG3 if _n > 400, ab(12)
385
386
387
388 Model selection:
389
390 1. Type of model: The 'traj' can model normal, censored normal, zero-inflated Poisson and binary
logit models.
391 Capacity for incorporating effect of time-stable and time-varying covariates,
392 subsequent outcomes and joint trajectory models.
393
394 2. Number of groups/classes: determination of the optimal number of groups to compose the mixture
395
396 3. Shape of the trajectory: determination of the appropriate order of the
397 polynomial used to model each group's trajectory (linear, quadratic, cubic).
398
399
400
401 Model Fit Criteria to select the model with optimal class enumeration:
402
403 • Bayesian Information Criteria (BIC), where lower BIC or least negative BIC
404 (higher value closer to zero) represents a better fitting model.
405
406 • Bayes Factor greater than 10 indicates very strong evidence
407 to use the "more complex" model.
408
409 • Meaningful proportion of participants within each class
410 (smallest group percentage to be higher or equal to 5%).
411
412 • Average posterior probability (APP) to belong to each class higher than 0.70.
413
414 • Entropy to determine the accuracy of classification of individuals into the different latent
classes
415 If entropy is near 1.0, then classification of individuals is assumed to be adequate.
416 If entropy is near 0, then classification is assumed to be poor.
417
418
419
420
421
422
423 *****function to print out summary stats
424 program summary_table_procTraj
425     preserve
426     *look at the average posterior probability
427     gen Mp = 0
428     foreach i of varlist _traj_ProbG* {
429         replace Mp = `i' if `i' > Mp
430     }
431     sort _traj_Group
432     *and the odds of correct classification
433     by _traj_Group: gen countG = _N
434     by _traj_Group: egen groupAPP = mean(Mp)
435     by _traj_Group: gen counter = _n
436     gen n = groupAPP/(1 - groupAPP)
437     gen p = countG/_N
438     gen d = p/(1-p)
439     gen occ = n/d
440     *Estimated proportion for each group
441     scalar c = 0
442     gen TotProb = 0
443     foreach i of varlist _traj_ProbG* {
444         scalar c = c + 1
445         quietly summarize `i'
446         replace TotProb = r(sum)/ _N if _traj_Group == c
447     }
448     gen d_pp = TotProb/(1 - TotProb)
449     gen occ_pp = n/d_pp

```

```

450      *This displays the group number [_traj_~p],
451      *the count per group (based on the max post prob), [countG]
452      *the average posterior probability for each group, [groupAPP]
453      *the odds of correct classification (based on the max post prob group assignment), [occ]
454      *the odds of correct classification (based on the weighted post. prob), [occ_pp]
455      *and the observed probability of groups versus the probability [p]
456      *based on the posterior probabilities [TotProb]
457      list _traj_Group countG groupAPP occ occ_pp p TotProb if counter == 1
458      restore
459  end
460
461  summary_table_procTraj
462
463
464
465  ***** to generate a plot of the individual trajectories
466
467  preserve
468  reshape long count_ t_, i(id)
469
470  gen count_jit = count_ + ( 0.2*runiform()-0.1 )
471  graph twoway scatter count_jit t_, c(L) by(_traj_Group) msize(tiny) mcolor(gray) lwidth(vthin)
472  lcolor(gray)
473
474  ***** to calculate the Bayes factor
475
476  log Bayes factor (2loge(B10) ≈ 2(ΔBIC)
477  This estimate approximately equals 2(BICcomplex model-BICnull model)
478
479  */
480
481
482
483  /*
484
485  Depressive symptoms (CES-D 8 item)
486  The trajectory groups of the CES-D scores (as a discrete variable) are tested
487  alone with time as the only independent variable, with no covariates added that could influence
488  class membership.
489  The Zero inflated poisson (ZIP) Model ('zip') is applied, given that the CES-D 8-item was a
490  count of symptoms and the majority of individuals scored 0 at each time point.
491
492  Initially, for each model, the linear, quadratic, and cubic functions of each trajectory can be
493  tested,
494  depending on the number of time points.
495  To ensure parsimony, consistent with the recommendations of Helgeson, Snyder, and Seltman (2004),
496  non-significant cubic and quadratic terms are removed from trajectories in a given model,
497  but linear parameters are retained irrespective of significance.
498
499  I tested the best fitting model with two, three, four five and then six trajectories following
500  the same process.
501  The models were compared (in a table of comparison) using BIC statistics,
502  Bayes factor, entropy, percentage of each class and average posterior probabilities.
503
504  PROCESS TO SELECT THE BEST-FITTING MODEL
505
506  Shape and Classes
507
508  1. run one traj with quadratic (order 2)
509  - If quadratic is not significant run with linear parameter (order 1)
510
511  2. model with 2 traj with quadratic (order 2 2)
512  - If neither traj is significant rerun with linear (order 1 1)
513  - If one not significant adapt accordingly (e.g. order 1 2 OR order 2 1)

```



```

513 3. Compare models (complex-simple) with statistic criteria
514
515 4. Repeat the process with an increasing number of traj
516
517 */
518
519
520
521
522
523 *** ZIP MODEL
524
525
526 * 1 class
527 traj, var(cesd_*) indep(t_*) model(zip) order(2) iorder(1)
528 trajplot, xtitle(Time in Months) ytitle(Depressive symptoms CES-D) ci
529
530 /* Shows the assigned group and probabilities of group membership */
531 list _traj_Group - _traj_ProbG1 if _n < 3, ab(12)
532
533 /* trajT = x-axis, Avg# = data averages, Est# = model estimates */
534 matrix list e(plot1), format(%9.2f) noheader
535
536
537
538
539 * 2 classes
540 traj, var(cesd_*) indep(t_*) model(zip) order(2 2) iorder(1)
541 trajplot, xtitle(Time in Months) ytitle(Depressive symptoms CES-D) ci
542
543 /* Shows the assigned group and probabilities of group membership */
544 list _traj_Group - _traj_ProbG2 if _n < 3, ab(12)
545
546 /* trajT = x-axis, Avg# = data averages, Est# = model estimates */
547 matrix list e(plot1), format(%9.2f) noheader
548
549
550
551
552
553 * 3 classes
554 traj, var(cesd_*) indep(t_*) model(zip) order(2 2 2) iorder(1)
555 trajplot, xtitle(Time in Months) ytitle(Depressive symptoms CES-D) ci
556
557 /* Shows the assigned group and probabilities of group membership */
558 list _traj_Group - _traj_ProbG3 if _n < 3, ab(12)
559
560 /* trajT = x-axis, Avg# = data averages, Est# = model estimates */
561 matrix list e(plot1), format(%9.2f) noheader
562
563
564
565
566
567
568 * 4 classes
569 traj, var(cesd_*) indep(t_*) model(zip) order(2 2 2 2) iorder(1)
570 trajplot, xtitle(Time in Months) ytitle(Depressive symptoms CES-D) ci
571
572
573 /* Shows the assigned group and probabilities of group membership */
574 list _traj_Group - _traj_ProbG4 if _n < 3, ab(12)
575
576 /* trajT = x-axis, Avg# = data averages, Est# = model estimates */
577 matrix list e(plot1), format(%9.2f) noheader
578
579
580

```

```

581 * 5 classes
582 traj, var(cesd_*) indep(t_*) model(zip) order(2 2 2 2 2) iorder(1)
583 trajplot, xtitle(Time in Months) ytitle(Depressive symptoms CES-D) ci
584
585
586
587 /* Shows the assigned group and probabilities of group membership */
588 list _traj_Group - _traj_ProbG4 if _n < 3, ab(12)
589
590 /* trajT = x-axis, Avg# = data averages, Est# = model estimates */
591 matrix list e(plot1), format(%9.2f) noheader
592
593
594
595
596
597
598 * 6 classes
599 traj, var(cesd_*) indep(t_*) model(zip) order(2 2 2 2 2 2) iorder(1)
600 trajplot, xtitle(Time in Months) ytitle(Depressive symptoms CES-D) ci
601
602
603
604 /* Shows the assigned group and probabilities of group membership */
605 list _traj_Group - _traj_ProbG4 if _n < 3, ab(12)
606
607 /* trajT = x-axis, Avg# = data averages, Est# = model estimates */
608 matrix list e(plot1), format(%9.2f) noheader
609
610
611
612 ** The 5-class model is the best fit
613
614
615
616 * 5 classes
617 traj, var(cesd_*) indep(t_*) model(zip) order(3 3 3 3 3) iorder(1)
618 trajplot, xtitle(Time in Months) ytitle(Depressive symptoms CES-D) ci
619
620
621
622 /* Shows the assigned group and probabilities of group membership */
623 list _traj_Group - _traj_ProbG4 if _n < 3, ab(12)
624
625 /* trajT = x-axis, Avg# = data averages, Est# = model estimates */
626 matrix list e(plot1), format(%9.2f) noheader
627
628
629
630
631
632 *** OPTIMAL ZIP MODEL
633
634 * 5 groups - zip model - cubic and linear polynomial (3 3 3 3 1)
635 traj, var(cesd_*) indep(t_*) model(zip) order(3 3 3 3 1) iorder(1)
636 trajplot, xtitle(Time in Months) ytitle(Depressive symptoms CES-D) ci
637
638
639 /* Shows the assigned group and probabilities of group membership */
640 list _traj_Group - _traj_ProbG5 if _n < 3, ab(12)
641
642 /* trajT = x-axis, Avg# = data averages, Est# = model estimates */
643 matrix list e(plot1), format(%9.2f) noheader
644
645
646
647
648

```

```

649
650 ** run after each traj model to estimate the average posterior probability (APP) for each group
651
652 program summary_table_procTraj
653     preserve
654     *look at the average posterior probability
655     gen Mp = 0
656     foreach i of varlist _traj_ProbG* {
657         replace Mp = `i' if `i' > Mp
658     }
659     sort _traj_Group
660     *and the odds of correct classification
661     by _traj_Group: gen cesdG = _N
662     by _traj_Group: egen groupAPP = mean(Mp)
663     by _traj_Group: gen counter = _n
664     gen n = groupAPP/(1 - groupAPP)
665     gen p = cesdG/_N
666     gen d = p/(1-p)
667     gen occ = n/d
668     *Estimated proportion for each group
669     scalar c = 0
670     gen TotProb = 0
671     foreach i of varlist _traj_ProbG* {
672         scalar c = c + 1
673         quietly summarize `i'
674         replace TotProb = r(sum)/ _N if _traj_Group == c
675     }
676     gen d_pp = TotProb/(1 - TotProb)
677     gen occ_pp = n/d_pp
678     *This displays the group number [_traj_~p],
679     *the cesd per group (based on the max post prob), [countG]
680     *the average posterior probability for each group, [groupAPP]
681     *the odds of correct classification (based on the max post prob group assignment), [occ]
682     *the odds of correct classification (based on the weighted post. prob), [occ_pp]
683     *and the observed probability of groups versus the probability [p]
684     *based on the posterior probabilities [TotProb]
685     list _traj_Group cesdG groupAPP occ occ_pp p TotProb if counter == 1
686     restore
687 end
688
689 summary_table_procTraj
690
691
692
693
694 /*
695 ---- MODEL SELECTION ----
696 Best-fitting model to try survival analysis is the 5 class order(3 3 3 3 1)
697 */
698
699
700
701
702
703
704 /*
705 Data and variable manipulation
706 */
707
708 * 5-class model: rename _traj_Group to C_traj_group5
709
710 rename _traj_Group E_traj_group5
711 recode E_traj_group5 (1=2) (2=1) (4=5) (5=4)
712 ta E_traj_group5
713 rename _traj_ProbG1 E_depres_traj_1
714 rename _traj_ProbG2 E_depres_traj_2
715 rename _traj_ProbG3 E_depres_traj_3
716 rename _traj_ProbG4 E_depres_traj_4

```

```

717  rename _traj_ProbG5 E_depres_traj_5
718
719
720
721
722
723  * labelling variable of E_traj_group4
724
725  label var E_traj_group5 "Traj 5 groups of depressive symptoms"
726
727  * labelling values
728  lab def traj_depre 1 "minimal" 2 "mild" 3 "increasing" 4 "decreasing" 5 "high"
729
730  * attach category labels to the variable through label value
731
732  lab val E_traj_group5 traj_depre
733
734  ta E_traj_group5
735
736
737
738
739
740
741
742  * Frequencies of covariates
743
744  tabulate E_age
745  summarize E_age
746
747
748  tabulate E_sex
749  summarize E_sex
750
751
752  tabulate E_education
753  summarize E_education
754
755
756  tabulate E_maritalstatus_4cat
757  summarize E_maritalstatus_4cat
758
759
760  tabulate E_wealthquintiles
761  summarize E_wealthquintiles
762
763
764  tabulate E_smoking_3cat
765  summarize Ewv2_smoking_3cat
766
767
768  tabulate E_physicalactivity
769  summarize E_physicalactivity
770
771
772  tabulate E_alcohol_status
773  summarize E_alcohol_status
774
775
776  tabulate E_cvd_comorbidity
777  summarize E_cvd_comorbidity
778
779
780  tabulate E_memory_wordrecall
781  summarize E_memory_wordrecall
782
783
784

```

```

785
786
787
788
789  /* MULTIPLE IMPUTATION (MI)
790
791  To handle with missing values of covariates
792
793
794  useful sources for MI and MICE:
795
796  https://stats.idre.ucla.edu/stata/seminars/mi_in_stata_pt1_new/
797  https://www.stata.com/manuals/mi.pdf - see page 139
798  https://www.stata.com/meeting/switzerland16/slides/medeiros-switzerland16.pdf
799  https://www.youtube.com/watch?v=i6S0lq0mjuc&ab_channel=StataCorpLLC
800  https://dss.princeton.edu/training/MIStata.pdf
801
802
803
804  Preparing to conduct MI
805  1. examine the number and proportion of missing values among the variables of interest
806      use the mdesc command
807  2. examine missing data patterns
808      use commands mi set and mi misstable patterns
809  3. identify potential auxiliary variables
810
811
812  Run MI using chained equations (MICE)
813  using the commands
814  1. how (in what style) to store the imputations
815      mi set wide
816  2. which variables will be imputed
817      mi register imputed
818  3. optionally, which variables will not be imputed
819      mi register regular
820  4. what imputation method is implemented to impute each of var - MICE
821      mi impute chained
822
823  */
824
825
826
827
828
829  /*
830
831  1. examining missing values
832      install packages:
833      * install mdesc
834      * install tabmiss
835      * insatll dm31
836      * insall mvpatterna
837
838  */
839
840  search mdesc
841  search rmiss2
842  search mvpatterns
843
844
845  * examining number of missing values vs non-missing in each variable
846
847  mdesc E_age E_sex E_education E_maritalstatus_4cat E_wealthquintiles ///
848  E_smoking_3cat E_physicalactivity E_alcohol_status ///
849  E_cvd_comorbidity E_memory_wordrecall
850
851
852

```

```

853 * examining missing data patterns
854
855 mi set wide
856
857 mi misstable summarize E_age E_sex E_education E_maritalstatus_4cat E_wealthquintiles ///
858 E_smoking_3cat E_physicalactivity E_alcohol_status ///
859 E_cvd_comorbidity
860
861
862 mi misstable patterns E_age E_sex E_education E_maritalstatus_4cat E_wealthquintiles ///
863 E_smoking_3cat E_physicalactivity E_alcohol_status ///
864 E_cvd_comorbidity
865
866
867 /*
868 identifying potential auxiliary var
869 * Auxiliary variables are either correlated with a missing variable(s)
870 (the recommendation is  $r > 0.4$ ) or are believed to be associated with missingness
871 - a priori knowledge of var that would make good auxiliary var
872 - identify potential candidates by examining associations between missing var and other var in
the dataset
873 running correlation using the command: pwcorr v1 v2 v3, obs
874 the recommendation for good correlation is  $r > 0.4$ 
875
876
877 Missing var to be imputed:
878
879 E_education E_wealthquintiles
880 Ewv6_smoking_3cat Ewv6_physicalactivity Ewv6_alcohol_status
881 Ewv6_cvd_comorbidity Ewv6_glycemia Ewv6_bp Ewv6_obesity Ewv6_hdl_cholesterol
882 Ewv6_memory_wordrecall
883
884
885
886 Potential auxiliary var:
887 DV: Ewv6to9_dementia_event
888 IV: E_traj_group4 cesd_0 cesd_1 cesd_2 cesd_3
889 other var:
890 E_age E_sex E_maritalstatus_4cat Ewv2_depressive_symptoms
891
892 */
893
894
895 * correlation
896
897 pwcorr E_education E_wealthquintiles ///
898 E_smoking_3cat E_physicalactivity E_alcohol_status ///
899 E_cvd_comorbidity ///
900 Ewv6to9_dementia_event E_traj_group4 cesd_0 cesd_1 cesd_2 cesd_3 ///
901 E_age E_sex E_maritalstatus_4cat Ewv2_depressive_symptoms, obs
902
903
904 * The correlation showed that all the above potential var are good auxiliary
905 * A good auxiliary does not have to be correlated with every variable to be useful
906 * And it's not problematic if it has missing info of it's own
907
908
909
910
911 /*
912 MI by chained equations (MICE)
913 see: https://stats.idre.ucla.edu/stata/seminars/mi\_in\_stata\_pt1\_new/
914
915 MICE is known as the fully conditional specification or sequential generalized regression
916 does not assume a joint MVN distribution
917 but instead uses a separate conditional distribution for each imputed variable.
918
919 The multivariate normal (MVN) model - mi imputed mvn -

```

```

920 assumes multivariate normality of all var
921
922 The multivariate imputation by chained equations (MICE) - mi imputed chained -
923 offers flexibility in how each var is modeled
924
925 mi impute chained allows to specify models for a
926 variety of variable types, including
927 continuous, binary, ordinal, nominal, truncated, and count variables
928
929
930 The MICE distributions available in Stata are:
931 binary, ordered and multinomial logistic regression for categorical variables,
932 linear regression and predictive mean matching (PMM)* for continuous variables,
933 and Poisson and negative binomial regression for count variables.
934
935
936 IMPUTATION PHASES
937
938 1. mi set wide
939     style to store imputations
940
941
942 2. mi register imputed
943     identifies which variables in the imputation model have missing information.
944
945 3. mi register regular (! optional)
946     which variables will not be imputed
947
948 4. mi impute chained
949     where the user specifies the imputation model to be used
950     and the number of imputed datasets to be created.
951     Example:
952         mi impute chained (regress) bmi age (logit) female ///
953         (mlogit) race = bpdiast i.region, add(20)
954
955 5. mi estimate
956     is used as a prefix to the standard regress command.
957     This executes the specified estimation model within each of the 20 imputed datasets
958     to obtain 20 sets of coefficients and standard errors.
959     Stata then combines these estimates to obtain one set of inferential statistics.
960     In the output from mi estimate you will see some metrics: Imputation Diagnostics
961     information for RVI (Relative Increase in Variance),
962     FMI (Fraction of Missing Information),
963     DF (Degrees of Freedom) ,
964     RE (Relative Efficiency),
965     and the between imputation and the within imputation variance estimates
966     to examine how the standard errors (SEs) are calculated.
967
968 -----
969
970
971 SELECTING MY IMPUTATION MODEL
972
973 - MICE -> mi impute chained
974
975 - var to be imputed:
976
977     linear regression for continuous var (regress) ->
978     Ewv6_memory_wordrecall
979
980     logistic for the binary var (logit) ->
981     Ewv6_cvd_comorbidity Ewv6_glycemia Ewv6_bp Ewv6_obesity Ewv6_hdl_cholesterol
982
983
984     multinomial logistic for our nominal categorical var (mlogit) ->
985     E_education E_wealthquintiles
986     Ewv6_smoking_3cat Ewv6_physicalactivity Ewv6_alcohol_status
987

```

```

988
989
990 - auxiliary var:
991
992     DV -> Ewv6to9_dementia_event
993     IV -> E_traj_group4 Ewv2_depressive_symptoms
994     other covariates -> E_age E_sex E_maritalstatus_4cat
995
996
997
998 - imputation numbers (m) -> 20
999
1000     White et al. (2010) recommendation: use the rule that m should equal the percentage of
incomplete cases
1001
1002
1003 - rseed (53421) for reproducability reasons
1004
1005
1006 - (! OPTIONAL) advance impute options -> force
1007
1008     proceed with imputation, even when missing imputed values (e.g. auxiliary have missing data)
are encountered
1009
1010 - impute options -> savetrace (trace1)
1011
1012     specifies Stata to save the means and standard deviations of imputed values from each
iteration to a Stata dataset named "trace1
1013 */
1014
1015
1016 mi set wide
1017
1018
1019 mi register imputed E_education E_wealthquintiles ///
1020     E_smoking_3cat E_physicalactivity E_alcohol_status
1021
1022
1023
1024 mi impute chained (mlogit) E_education E_wealthquintiles E_smoking_3cat E_physicalactivity
E_alcohol_status ///
1025 = Ewv6to9_dementia_event E_traj_group4 Ewv2_depressive_symptoms ///
1026 E_age E_sex E_maritalstatus_4cat, add(10) rseed(53421) savetrace(trace1)
1027
1028
1029 * save imputed data
1030
1031 * plot imputations
1032
1033
1034 *it will open a file named trace1
1035 use trace1, clear
1036
1037 describe
1038
1039
1040 reshape wide *mean *sd, i(iter) j(m)
1041
1042 tsset iter
1043
1044
1045
1046
1047 /*
1048 The trace plot below graphs the predicted means value produced during the first imputation chain.
1049 As before, the expectations is that the values would vary randomly to incorporate variation into
the predicted values for read.
1050 */

```



```

1051
1052     tsline E_eduaction_mean1, name(mice1,replace)legend(off) ytitle("Mean of education")
1053     tsline E_wealthquintiles_mean1, name(mice1,replace)legend(off) ytitle("Mean of wealth")
1054     tsline E_smoking_3cat_mean1, name(mice1,replace)legend(off) ytitle("Mean of smoking")
1055     tsline E_physicalactivity_mean1, name(mice1,replace)legend(off) ytitle("Mean of physical activity")
1056     tsline E_alcohol_status_mean1, name(mice1,replace)legend(off) ytitle("Mean of alcohol status")
1057
1058
1059     /*
1060
1061     All 10 imputation chains can also be graphed simultaneously to make sure that nothing unexpected
1062     occurred in a single chain.
1063     Every chain is obtained using a different set of initial values and this should be unique.
1064     Each colored line represents a different imputation.
1065     So all 10 imputation chains are overlaid on top of one another.
1066
1067     */
1068
1069     tsline E_alcohol_status_mean*, name(mice1,replace)legend(off) ytitle("Mean of alcohol status")
1070     tsline E_alcohol_status_sd*, name(mice2, replace) legend(off) ytitle("SD of alcohol status")
1071     graph combine mice1 mice2, xcommon cols(1) title(Trace plots of summaries of imputed values)
1072
1073     * repeat for each imputed var
1074
1075
1076
1077
1078
1079
1080     /*
1081     ---- DESCRIPTIVE STATISTICS ----
1082
1083     General characteristics of participants
1084
1085     Participant characteristics by depressive symptom trajectory group
1086
1087     CHI-SQUARE (chi2) for categorical var (crosstabulation)
1088         Frequency tables -> two-way tables
1089         using the command tabulate, chi2
1090         reporting observations, column percentage (N, %) and p-value of Pearson's r
1091
1092     one-way ANOVA for continuous var
1093         check box plot
1094         using the command oneway
1095         reporting mean, sd (summary tables) and p-value of F
1096
1097     */
1098
1099     * General characteristics of ELSA participants
1100
1101     * Demographics
1102     sum E_age
1103     ta E_sex
1104     ta E_eduaction
1105     ta E_maritalstatus_4cat
1106     ta E_wealthquintiles
1107     * Lifestyle and health indicators
1108     ta E_smoking_3cat
1109     ta E_physicalactivity
1110     ta E_alcohol_status
1111     ta E_cvd_comorbidity
1112     * Cardiometabolic risk factors
1113     ta Ewv2_crp
1114     ta Ewv2_hdl_cholesterol
1115     ta Ewv2_obesity_waist
1116     ta Ewv2_systolic_bp
1117     ta Ewv2_diastolic_bp

```

```

1118 ta Ewv2_diabetes_diagnosed
1119 ta Ewv2_HbA1c
1120 * Depressive symptoms t1-t3 (cont and categ)
1121 sum cesd_0
1122 sum cesd_1
1123 sum cesd_2
1124 sum cesd_3
1125 ta depress_0
1126 ta depress_1
1127 ta depress_2
1128 ta depress_3
1129 * Memory score at baseline
1130 sum E_memory_wordrecall
1131
1132
1133
1134
1135
1136 * Sample characteristics by depressive symptom trajectories
1137 * crosstabs categ var (frequencies and chi2) !report column percentage!
1138 * oneway ANOVA cont var (mean, sd)
1139
1140
1141 * Demographics
1142 oneway E_age E_traj_group5, tabulate
1143 ta E_sex E_traj_group5, chi2 column row
1144 ta E_education E_traj_group5, chi2 column row
1145 ta E_maritalstatus_4cat E_traj_group5, chi2 column row
1146 ta E_wealthquintiles E_traj_group5, chi2 column row
1147 * Lifestyle factors
1148 ta E_smoking_3cat E_traj_group5, chi2 column row
1149 ta Ewv6_physicalactivity E_traj_group5, chi2 column row
1150 ta E_alcohol_status E_traj_group5, chi2 column row
1151 ta E_cvd_comorbidity E_traj_group5, chi2 column row
1152 * Cardiometabolic risk factors
1153 ta Ewv2_crp E_traj_group5, chi2 column row
1154 ta Ewv2_hdl_cholesterol E_traj_group5, chi2 column row
1155 ta Ewv2_obesity_waist E_traj_group5, chi2 column row
1156 ta Ewv2_systolic_bp E_traj_group5, chi2 column row
1157 ta Ewv2_diastolic_bp E_traj_group5, chi2 column row
1158 ta Ewv2_diabetes_diagnosed E_traj_group5, chi2 column row
1159 ta Ewv2_HbA1c E_traj_group5, chi2 column row
1160 * Memory score at baseline
1161 oneway E_memory_wordrecall E_traj_group5, tabulate
1162
1163
1164 ta E_traj_group5 Ewv6to9_dementia_event, chi2 column row
1165 ta E_traj_group5 Ewv6_anydementia_iqcode_report, chi2 column row
1166
1167
1168
1169
1170 /*
1171 ---- BINOMIAL LOGISTIC REGRESSION ON COMPLETE DATA ----
1172
1173 Command is:
1174 logistic DV IVs
1175 OR
1176 logit DV IVs, or
1177
1178
1179 --- Building the model using baseline covariates ---
1180
1181 Model 1: unadjusted - single predictor of depressive symptom trajectories C_traj_group5
1182 Model 2: model 1 + sociodemographics: age sex education marital status and wealth
1183 Model 3: model 2 + health behaviours: smoking, alcohol consumption
1184
1185 */

```

```

1186
1187
1188 * Unadjusted model - model 1 - single predictor
1189
1190 logistic Ewv5_anydementia_iqcode_report E_traj_group5
1191
1192 *OR
1193
1194 logit Ewv5_anydementia_iqcode_report E_traj_group5, or
1195
1196
1197
1198 * define design var by using i.
1199
1200 logistic Ewv5_anydementia_iqcode_report i.E_traj_group5
1201
1202 *OR
1203
1204 logit Ewv5_anydementia_iqcode_report i.E_traj_group5, or
1205
1206
1207 * Adjusted models - multivariable logistic regression
1208 * controlling for covariates
1209
1210 * model 2: model 1 + adjust for demographics: age sex education marital status and wealth
1211
1212 logistic Ewv5_anydementia_iqcode_report i.E_traj_group5 ///
1213 E_age i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles
1214
1215 * model 3: model 2 + adjust for lifestyle and health indicators
1216
1217 logistic Ewv5_anydementia_iqcode_report i.E_traj_group5 ///
1218 E_age i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles ///
1219 i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity
1220
1221
1222
1223
1224
1225
1226 /*
1227 ---- SURVIVAL ANALYSIS ----
1228
1229 Tests of proportional-hazards assumption
1230 Kaplan Meier survival curves
1231 Person-time
1232 Cox proportional regression - Hazard ratios - stcox
1233 Postestimation tools for stcox
1234 Test of Goodness of Fit
1235
1236 *** Cox regression in full data, complete data (listwise deletion of missing data) and imputed data
1237 Cox PH regression in complete data
1238 Cox PH regression model in imputed dataset - mi estimate
1239
1240
1241 */
1242
1243
1244
1245 * check dataset variables of interest only
1246
1247 codebook Ewv6to9_time_event_dementia Ewv6to9_dementia_event E_traj_group5 ///
1248 E_age E_sex E_education E_maritalstatus_4cat E_wealthquintiles ///
1249 E_smoking_3cat E_alcohol_status E_cvd_comorbidity
1250
1251
1252 * Declare Data to be Survival Data
1253 * Time to event: Ewv6to9_time_event_dementia (months)

```

```

1254 * Censoring: Ewv6to9_dementia_event (1=dementia, 0=censored)
1255 * Command is stset TIMETOEVENT, failure(CENSORVARIABLE)
1256
1257
1258 stset Ewv6to9_time_event_dementia, failure (Ewv6to9_dementia_event==1) id(idauniq)
1259
1260
1261 *describe survival data using commnad stsum
1262
1263 stsum
1264
1265 stsum, by(E_traj_group)
1266
1267
1268
1269
1270 * Kaplan Meier Curve estimation
1271
1272 sts list
1273
1274 sts list, by(E_traj_group5)
1275
1276
1277
1278 * Kaplan Meier Curve Plot
1279
1280 * no frills plot
1281
1282 sts graph
1283
1284 * with frills
1285
1286 sts graph, xtitle("Time in Months") ytitle("Survival Prob") ///
1287 title("Kaplan Meier Curve") subtitle("n=4718, # events=263") ///
1288 caption("graph02.png", size(vsmall))
1289
1290
1291 * With Greenwood CI limits
1292
1293 sts graph, gwood legend(off) xtitle("Time in Months") ytitle("Survival Prob") ///
1294 title("Kaplan Meier Curve") subtitle("n=4718, # events=263") caption("graph03.png", size(vsmall))
1295
1296
1297
1298
1299 * Group Kaplan-Meier Curve Estimation
1300 * Command is sts graph, by(GROUPVAR) OPTION OPTION OPTION Note: Must have sorted by GROUPVAR first
1301
1302 sort E_traj_group5
1303
1304 sts list, by(E_traj_group5)
1305
1306 * graph with frills
1307
1308 sts graph, by(E_traj_group5) xlabel(0(20)120) ylabel(0.80(.05)1) xtitle("Time in Months") ///
1309 ytitle("Survival Prob") title("Kaplan Meier Curve") subtitle("n=4718, # events=263") ///
1310 caption("graph04.png", size(vsmall))
1311
1312
1313
1314
1315 * calculate person-time and incidence rates using command stptime
1316
1317 stptime, title(Person-years)
1318
1319 stptime, title(Person-years) per(1000)
1320
1321

```

```

1322 * calculate person-time by category of E_traj_group5
1323
1324 stptime, by(E_traj_group5)
1325
1326 stptime, by(E_traj_group5) per(1000)
1327
1328 * calculate the median of follow-up
1329 sum Ewv6to9_time_event_dementia, detail
1330
1331
1332 /* Log Rank Test of equality of survival distributions
1333 (NULL: equality of survival distributions among E_traj_group5 groups)
1334 We will consider including the predictor if the test has a p-value of 0.2 - 0.25 or less.
1335 If the predictor has a p-value greater than 0.25 in a univariate analysis
1336 it is highly unlikely that it will contribute anything to a model which includes other
predictors.
1337 Command is sts test GROUPVAR
1338 */
1339
1340
1341 sts test E_traj_group5, logrank
1342
1343 sts test E_age, logrank
1344
1345 sts test E_sex, logrank
1346
1347 sts test E_eduaction, logrank
1348
1349 sts test E_maritalstatus_4cat, logrank
1350
1351 sts test E_wealthquintiles, logrank
1352
1353 sts test E_smoking_3cat, logrank
1354
1355 sts test E_alcohol_status, logrank
1356
1357 sts test E_cvd_comorbidity, logrank
1358
1359
1360
1361
1362
1363 /* Cox PH regression model
1364
1365 using the command stcox
1366
1367 --- Building the model ---
1368
1369 Model 1: unadjusted - single predictor of depressive symptom trajectories E_traj_group5
1370 Model 2: model 1 + sociodemographics: age sex education marital status and wealth
1371 Model 3: model 2 + lifestyle and health indicators: smoking, alcohol consumption and cvd
1372
1373 */
1374
1375
1376 * Unadjusted model - model 1 - single predictor
1377
1378 stcox E_traj_group5
1379
1380 * define design var by using i.(low, moderate, high, ref: minimal)
1381
1382 stcox i.E_traj_group5
1383
1384
1385 * Adjusted models - multivariable Cox model
1386 * controlling for covariates
1387
1388 * model 2: model 1 + adjust for demographics: age sex education marital status and wealth

```

```

1389
1390 stcox i.E_traj_group5 E_age i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles
1391
1392
1393 * model 3: model 2 + adjust for lifestyle and health indicators
1394
1395 stcox i.E_traj_group5 E_age i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles ///
1396 i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity
1397
1398
1399 * Coefficients instead of hazard ratios by specifying the option nohr
1400
1401 stcox i.E_traj_group5, nohr
1402
1403
1404
1405 stcox i.E_traj_group5 E_age i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles ///
1406 i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity, nohr
1407
1408
1409
1410
1411
1412
1413 * Multivariable model development
1414 * Likelihood-ratio tests
1415
1416
1417
1418 *install eststo
1419 findit eststo
1420
1421
1422 * ---- rx controlling for age and sex ----*
1423 quietly: stcox E_age i.E_sex
1424 eststo modelagesex
1425
1426 quietly: stcox E_age i.E_sex i.E_traj_group5
1427 eststo modelagesex_4group
1428
1429 lrtest modelagesex modelagesex_4group
1430
1431
1432
1433 * ---- rx controlling for sociodemographics ----*
1434 quietly: stcox E_age i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles
1435 eststo modelsociodemo
1436
1437 quietly: stcox E_age i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles i.
1438 E_traj_group5
1439 eststo modelsociodemo_4group
1440
1441 lrtest modelsociodemo modelsociodemo_4group
1442
1443 * ---- rx controlling for lifestyle and health indicators----*
1444 quietly: stcox i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity
1445 eststo modellifestyle
1446
1447 quietly: stcox i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity i.E_traj_group5
1448 eststo modellifestyle_4group
1449
1450 lrtest modellifestyle modellifestyle_4group
1451
1452
1453
1454
1455 * side-by-side comparison of models

```

```

1456
1457
1458 quietly: stcox i.E_traj_group4
1459 eststo model1
1460
1461
1462 quietly: stcox E_age i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles i.
E_traj_group5
1463 eststo model2
1464
1465
1466 quietly: stcox E_age i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles ///
1467 i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity i.E_traj_group5
1468 eststo model3
1469
1470
1471
1472 * Display Betas and Summary Statistics
1473 estout model1 model2 model3, stats(n chi2 bic, star(chi2)) prehead("Betas")
1474
1475 /* Key Interpretation
1476 Chi2 = Value of LR test comparing the model fit ("full") to intercept only ("reduced")
1477 bic = Schwarz' Bayesian Information Criterion = It is a function of the log-likelihood.
1478 Smaller values indicate a better fit.
1479 */
1480
1481 * Display Hazard Ratios and Model Fit Statistics. Option eform produces hazard ratios
1482 estout model1 model2 model3, eform stats(n chi2 bic, star(chi2)) prehead("Hazard Ratios")
1483
1484
1485
1486
1487 * Postestimation tools for stcox
1488
1489 * Test of proportional hazards
1490
1491 estat phtest, detail
1492
1493
1494 /* Proportionality Assumption - method 1
1495 We will check proportionality by including time-dependent covariates in the model
1496 by using the tvc and the texp options in the stcox command.
1497 Time dependent covariates are interactions of the predictors and time.
1498 In this analysis we choose to use the interactions with log(time)
1499 because this is the most common function of time used in time-dependent covariates
1500 but any function of time could be used.
1501 If a time-dependent covariate is significant this indicates
1502 a violation of the proportionality assumption for that specific predictor.
1503 The conclusion is that all of the time-dependent variables are not significant
1504 either collectively or individually thus supporting the assumption of proportional hazard.
1505 */
1506
1507
1508
1509 stcox i.E_traj_group5 E_age i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles ///
1510 i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity, nohr ///
1511 tvc(E_traj_group4 E_age E_sex E_education E_maritalstatus_4cat E_wealthquintiles ///
1512 i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity) texp(ln(Ewv6to9_time_event_dementia))
1513
1514
1515
1516 /* Proportionality Assumption - method 2
1517 by using the Schoenfeld and scaled Schoenfeld residuals
1518 In the stphtest command we test the proportionality of the model as a whole
1519 and by using the detail option we get a test of proportionality for each predictor.
1520 By using the plot option we can also obtain a graph of the scaled Schoenfeld assumption.
1521 If the tests in the table are not significance (p-values over 0.05)
1522 then we can not reject proportionality and we assume

```



```

1523 that we do not have a violation of the proportional assumption.
1524 The stphplot command uses log-log plots to test proportionality
1525 and if the lines in these plots are parallel then we have further indication
1526 that the predictors do not violate the proportionality assumption.
1527 */
1528
1529 quietly stcox E_traj_group5 E_age E_sex E_education E_maritalstatus_4cat E_wealthquintiles ///
1530 E_smoking_3cat E_alcohol_status E_cvd_comorbidity, schoenfeld(sch*) scaledsch(sca*)
1531 stphtest, detail
1532 stphtest, plot(E_traj_group5) msym(oh)
1533 stphtest, plot(E_age) msym(oh)
1534 stphtest, plot(E_sex) msym(oh)
1535 stphtest, plot(E_education) msym(oh)
1536 stphtest, plot(E_maritalstatus_4cat) msym(oh)
1537 stphtest, plot(E_wealthquintiles) msym(oh)
1538 stphtest, plot(E_smoking_3cat) msym(oh)
1539 stphtest, plot(E_alcohol_status) msym(oh)
1540 stphtest, plot(E_cvd_comorbidity) msym(oh)
1541
1542
1543
1544
1545
1546 stphplot, by(E_traj_group5) plot1(msym(oh)) plot2(msym(th))
1547 stphplot, by(E_age) plot1(msym(oh)) plot2(msym(th))
1548 stphplot, by(E_sex) plot1(msym(oh)) plot2(msym(th))
1549 stphplot, by(E_education) plot1(msym(oh)) plot2(msym(th))
1550 stphplot, by(E_maritalstatus_4cat) plot1(msym(oh)) plot2(msym(th))
1551 stphplot, by(E_wealthquintiles) plot1(msym(oh)) plot2(msym(th))
1552 stphplot, by(E_smoking_3cat) plot1(msym(oh)) plot2(msym(th))
1553 stphplot, by(E_alcohol_status) plot1(msym(oh)) plot2(msym(th))
1554 stphplot, by(E_cvd_comorbidity) plot1(msym(oh)) plot2(msym(th))
1555
1556
1557
1558
1559 * Assessment of PH Assumption: adjust for age and sex
1560 stphplot, by(E_traj_group5) adjust(E_age E_sex) nolntime plot1opts(symbol(none) color(black)
1561 lpattern(dash)) ///
1562 plot2opts(symbol(none) color(navy)) plot3opts(symbol(none) color(green)) plot4opts(symbol(none)
1563 color(red)) ///
1564 title("Assessment of PH Assumption") subtitle(" Predictor is E_tarj_group5") xtitle("months")
1565
1566
1567 * Assessment of PH Assumption: adjust for model 2
1568 stphplot, by(E_traj_group5) adjust(E_age E_sex E_education E_maritalstatus_4cat E_wealthquintiles)
1569 ///
1570 nolntime plot1opts(symbol(none) color(black) lpattern(dash)) ///
1571 plot2opts(symbol(none) color(navy)) plot3opts(symbol(none) color(green)) plot4opts(symbol(none)
1572 color(red)) ///
1573 title("Assessment of PH Assumption") subtitle(" Predictor is E_tarj_group5") xtitle("months")
1574
1575
1576 * Assessment of PH Assumption: adjust for model 3
1577 stphplot, by(E_traj_group5) adjust(E_age E_sex E_education E_maritalstatus_4cat E_wealthquintiles
1578 ///
1579 E_smoking_3cat E_alcohol_status E_cvd_comorbidity) ///
1580 nolntime plot1opts(symbol(none) color(black) lpattern(dash)) ///
1581 plot2opts(symbol(none) color(navy)) plot3opts(symbol(none) color(green)) plot4opts(symbol(none)
1582 color(red)) ///
1583 title("Assessment of PH Assumption") subtitle(" Predictor is E_tarj_group5") xtitle("months")
1584

```



```

1585
1586 /* Test of overall goodness of fit
1587 Goodness of fit of the final model
1588 2 methods:
1589 - by using the commnad stcoxgof (good fit = non sig p-value)
1590 - by using the Cox-Snell residuals
1591     to create the Nelson-Aalen cumulative hazard function
1592     If the hazard function follows the 45 degree line then we know that it approximately
1593     has an exponential distribution with a hazard rate of one and that the model fits the data
1594 well.
1595     If the model fits the data, the plot of the cumulative hazard versus cs
1596     should approximate a straight line with slope 1.
1597 */
1598
1599 * by using the commnad stcoxgof
1600
1601 * install stcoxgof
1602 findit stcoxgof
1603
1604
1605 stcox E_traj_group5 E_age E_sex E_education E_maritalstatus_4cat E_wealthquintiles ///
1606 E_smoking_3cat E_alcohol_status E_cvd_comorbidity, mgale(mgale)
1607
1608
1609 stcoxgof
1610
1611
1612
1613 * by using the Cox-Snell residuals
1614
1615 quietly stcox E_traj_group5 E_age E_sex E_education E_maritalstatus_4cat E_wealthquintiles ///
1616 E_smoking_3cat E_alcohol_status E_cvd_comorbidity
1617 predict cs, csnell
1618
1619 * or
1620
1621 quietly stcox E_traj_group5
1622 predict cs, csnell
1623
1624
1625 stset cs, failure(Ewv6to9_dementia_event)
1626 sts generate km = s
1627 generate H = -ln(km)
1628 line H cs cs, sort ytitle("") clstyle(. refline)
1629
1630
1631
1632
1633
1634
1635 **** ----- IMPUTED DATASET -----****
1636
1637
1638
1639 /* ----- BINOMIAL LOGISTIC REGRESSION IN IMPUTED DATASET using baseline covariates ----- */
1640
1641 Command is
1642
1643 mi estimate : logit DV IV, or
1644
1645     OR
1646
1647 mi estimate: logistic DV IV
1648
1649 */
1650
1651

```

```

1652
1653 * Unadjusted model - model 1 - single predictor
1654
1655 mi estimate, eform("Odds Ratio"): logistic Ewv5_anydementia_iqcode_report E_traj_group5
1656
1657 *OR
1658
1659 mi estimate, eform("Odds Ratio"): logit Ewv5_anydementia_iqcode_report E_traj_group5, or
1660
1661
1662
1663 * define design var by using i.
1664
1665 mi estimate, eform("Odds Ratio"): logistic Ewv5_anydementia_iqcode_report i.E_traj_group5
1666
1667 *OR
1668
1669 mi estimate, eform("Odds Ratio"): logit Ewv5_anydementia_iqcode_report i.E_traj_group5, or
1670
1671
1672 * Adjusted models - multivariable logistic regression
1673 * controlling for covariates
1674
1675 * model 2: model 1 + adjust for demographics: age sex education marital status and wealth
1676
1677 mi estimate, eform("Odds Ratio"): logistic Ewv5_anydementia_iqcode_report i.E_traj_group5 ///
1678 E_age i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles
1679
1680 * model 3: model 2 + adjust for lifestyle and health indicators
1681
1682 mi estimate, eform("Odds Ratio"): logistic Ewv5_anydementia_iqcode_report i.E_traj_group5 ///
1683 E_age i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles ///
1684 i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity
1685
1686
1687
1688
1689 * ----- COX PH REGRESSION MODEL IN IMPUTED DATASET ----- *
1690
1691
1692 * Declare Data to be Survival Data by using mi
1693
1694 mi stset Ewv6to9_time_event_dementia, failure(Ewv6to9_dementia_event==1) id(idauniq)
1695
1696
1697 * Run Cox regression analysis in imputed dataset by using "mi estimate:"
1698 * Building the Model: Model 1 (unadjusted), Model 2, Model 3
1699
1700
1701
1702 * Model 1 (default coefficients)
1703 mi estimate: stcox E_traj_group5
1704
1705 * Model 1: define design var by using i.(low, moderate, high, ref: minimal)
1706 mi estimate: stcox ib2.E_traj_group5
1707
1708
1709 * Model 1 ask for hazard ratio by using the option eform("Haz.Ratio")
1710
1711 mi estimate, eform("Haz. Ratio"): stcox i.E_traj_group5
1712
1713
1714 * Model 2: sociodemographics
1715
1716 mi estimate, eform("Haz. Ratio"): stcox i.E_traj_group5 ///
1717 E_age i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles
1718
1719

```

```

1720 * Model 3: lifestyle and health indicators
1721
1722 mi estimate, eform("Haz. Ratio"): stcox i.E_traj_group5 ///
1723 E_age i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles ///
1724 i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity
1725
1726
1727
1728
1729
1730
1731
1732 /*
1733
1734 *** SENSITIVITY ANALYSES ***
1735
1736 1) single assessment of depressive symptoms and dementia risk at t0 and t3
1737 continuous var of CES-D 8 items (0-8)
1738 model 3 was further adjusted for cesd_0 and cesd_3
1739
1740 2) LCGA logit trajectories with dichotomous variable
1741
1742
1743 3) Complete data
1744
1745 */
1746
1747
1748
1749
1750 * 1) Single assessment of cesd_0 and cesd_3 at model 3
1751
1752
1753 * IMPUTED DATA
1754
1755
1756 * single assessment model 3 adjust for cesd_0 and cesd_3
1757
1758 mi estimate, eform("Haz. Ratio"): stcox i.E_traj_group5 ///
1759 E_age i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles ///
1760 i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity ///
1761 cesd_0
1762
1763
1764 mi estimate, eform("Haz. Ratio"): stcox i.E_traj_group5 ///
1765 E_age i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles ///
1766 i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity ///
1767 cesd_3
1768
1769
1770
1771
1772
1773
1774 /*
1775 2) Logistic model LCGA
1776
1777 use Ewv2_depressive_symptoms dichotomous variables (0-1)
1778
1779 Logistic (logit) model
1780
1781 use http://www.andrew.cmu.edu/user/bjones/traj/data/cambridge.dta,
1782 clear
1783
1784 traj, var(p1-p23) indep(tt1-tt23) model(logit) order(0 3 3)
1785
1786 trajplot, xtitle(Scaled Age) ytitle(Prevalence)
1787

```

```

1788      /* Assigned group and probabilities of group membership */
1789      list _traj_Group - _traj_ProbG3 if _n > 400, ab(12)
1790
1791  */
1792
1793
1794
1795  *rename Ewv2_depressive_symptoms score across the waves
1796
1797  rename Ewv2_depressive_symptoms depress_0
1798  rename Ewv3_depressive_symptoms depress_1
1799  rename Ewv4_depressive_symptoms depress_2
1800  rename Ewv5_depressive_symptoms depress_3
1801
1802
1803  net from http://www.andrew.cmu.edu/user/bjones/traj
1804  net install traj, force
1805  help traj
1806
1807
1808  *** LOGIT MODEL
1809
1810
1811  * 1 class - logit model - quadratic polynomial (2)
1812  traj, var(depress_*) indep(t_*) model(logit) order(2)
1813  trajplot, xtitle(Time in Months) ytitle(Depressive symptom caseness) ci
1814
1815  /* Assigned group and probabilities of group membership */
1816      list _traj_Group - _traj_ProbG1 if _n < 3, ab(12)
1817
1818
1819
1820
1821
1822  * 2 class - logit model - quadratic polynomial (2 2)
1823  traj, var(depress_*) indep(t_*) model(logit) order(2 2)
1824  trajplot, xtitle(Time in Months) ytitle(Depressive symptom caseness) ci
1825
1826  /* Assigned group and probabilities of group membership */
1827      list _traj_Group - _traj_ProbG2 if _n < 3, ab(12)
1828
1829
1830
1831  * 3 class - logit model - quadratic polynomial (2 2 2)
1832  traj, var(depress_*) indep(t_*) model(logit) order(2 2 2)
1833  trajplot, xtitle(Time in Months) ytitle(Depressive symptom caseness) ci
1834
1835  /* Assigned group and probabilities of group membership */
1836      list _traj_Group - _traj_ProbG3 if _n < 3, ab(12)
1837
1838
1839
1840
1841  * 4 class - logit model - quadratic polynomial (2 2 2 2)
1842  traj, var(depress_*) indep(t_*) model(logit) order(2 2 2 2)
1843  trajplot, xtitle(Time in Months) ytitle(Depressive symptom caseness)
1844
1845  /* Assigned group and probabilities of group membership */
1846      list _traj_Group - _traj_ProbG4 if _n < 3, ab(12)
1847
1848
1849
1850  * 5 class - logit model - quadratic polynomial (2 2 2 2 2)
1851  traj, var(depress_*) indep(t_*) model(logit) order(2 2 2 2 2)
1852  trajplot, xtitle(Time in Months) ytitle(Depressive symptom caseness)
1853
1854  /* Assigned group and probabilities of group membership */
1855      list _traj_Group - _traj_ProbG5 if _n < 3, ab(12)

```

```

1856
1857
1858
1859
1860 * The 4-model depressive traj is selected to be tested in different shapes.
1861
1862
1863 * 4 class - logit model - quadratic polynomial (2 2 3 3)
1864 traj, var(depress_*) indep(t_*) model(logit) order(2 2 3 3)
1865 trajplot, xtitle(Time in Months) ytitle(Depressive symptom caseness) ci
1866
1867 /* Assigned group and probabilities of group membership */
1868 list _traj_Group - _traj_ProbG3 if _n < 3, ab(12)
1869
1870
1871
1872
1873
1874 program summary_table_procTraj
1875 preserve
1876 *look at the average posterior probability
1877 gen Mp = 0
1878 foreach i of varlist _traj_ProbG* {
1879     replace Mp = `i' if `i' > Mp
1880 }
1881 sort _traj_Group
1882 *and the odds of correct classification
1883 by _traj_Group: gen cesdG = _N
1884 by _traj_Group: egen groupAPP = mean(Mp)
1885 by _traj_Group: gen counter = _n
1886 gen n = groupAPP/(1 - groupAPP)
1887 gen p = cesdG/_N
1888 gen d = p/(1-p)
1889 gen occ = n/d
1890 *Estimated proportion for each group
1891 scalar c = 0
1892 gen TotProb = 0
1893 foreach i of varlist _traj_ProbG* {
1894     scalar c = c + 1
1895     quietly summarize `i'
1896     replace TotProb = r(sum)/_N if _traj_Group == c
1897 }
1898 gen d_pp = TotProb/(1 - TotProb)
1899 gen occ_pp = n/d_pp
1900 *This displays the group number [_traj_~p],
1901 *the cesd per group (based on the max post prob), [countG]
1902 *the average posterior probability for each group, [groupAPP]
1903 *the odds of correct classification (based on the max post prob group assignment), [occ]
1904 *the odds of correct classification (based on the weighted post. prob), [occ_pp]
1905 *and the observed probability of groups versus the probability [p]
1906 *based on the posterior probabilities [TotProb]
1907 list _traj_Group cesdG groupAPP occ occ_pp p TotProb if counter == 1
1908 restore
1909 end
1910
1911 summary_table_procTraj
1912
1913
1914
1915 ta _traj_Group
1916
1917 recode _traj_Group 1=2 2=1 3=3 4=4
1918
1919
1920
1921
1922
1923

```

```

1924 * IMPUTED DATA: Logistic regression (Odds Ratio)
1925
1926 * Unadjusted model (model 1)
1927
1928 mi estimate, eform("Odds Ratio"): logistic Ewv5_anydementia_iqcode_report i._traj_Group
1929
1930
1931 * model 2: model 1 + adjust for demographics: age sex education marital status and wealth
1932
1933 mi estimate, eform("Odds Ratio"): logistic Ewv5_anydementia_iqcode_report i._traj_Group ///
1934 E_age i.E_sex i.E_eduaction i.E_maritalstatus_4cat i.E_wealthquintiles
1935
1936
1937 * model 3: model 2 + adjust for lifestyle health indicators
1938
1939 mi estimate, eform("Odds Ratio"): logistic Ewv5_anydementia_iqcode_report i._traj_Group ///
1940 E_age i.E_sex i.E_eduaction i.E_maritalstatus_4cat i.E_wealthquintiles ///
1941 i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity
1942
1943
1944
1945
1946
1947 * IMPUTED DATA: Cox regression (Hazard Ratio)
1948
1949
1950
1951 * Declare Data to be Survival Data by using mi
1952
1953 mi stset Ewv6to9_time_event_dementia, failure(Ewv6to9_dementia_event==1) id(idauniq)
1954
1955
1956
1957 * Unadjusted model (model 1)
1958
1959 mi estimate, eform("Haz. Ratio"): stcox i._traj_Group
1960
1961
1962 * Model 2: sociodemographics
1963
1964 mi estimate, eform("Haz. Ratio"): stcox i._traj_Group ///
1965 E_age i.E_sex i.E_eduaction i.E_maritalstatus_4cat i.E_wealthquintiles
1966
1967
1968 * Model 3: lifestyle and health indicators
1969
1970 mi estimate, eform("Haz. Ratio"): stcox i._traj_Group ///
1971 E_age i.E_sex i.E_eduaction i.E_maritalstatus_4cat i.E_wealthquintiles ///
1972 i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity
1973
1974
1975
1976
1977
1978 * 3) complete data analysis (see above)
1979
1980
1981
1982 * ----- *
1983
1984
1985

```