

```

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 4: Clustering of cardiometabolic risk factors and dementia incidence
7
8  Method of analysis:
9  Latent Class Analysis (LCA)
10
11 DATASET: ELSA
12 baseline: wave 2 (2004) follow-up waves 3-9 (2006-2018)
13
14 TIMELINE
15
16 LATENT CLASSES OF CARDIOMETABOLIC MARKERS: WV2 (BASELINE)
17 DEMENTIA INCIDENCE: W3 - WV9 (7 TIME POINTS)
18 COVARIATES ADJUSTMENT FOR HR MODELS: WV2
19
20
21 */
22
23
24
25 * KEEP NECESSARY VARIABLES
26
27 keep idauniq w2wtnur w2wtbld ///
28 E_sex E_age E_education_yrs E_education E_maritalstatus_3cat E_maritalstatus_4cat ///
29 E_wealthquintiles E_smoking_3cat E_physicalactivity E_alcohol_freq E_alcohol_status ///
30 E_cvd_comorbidity E_cognitive_index E_memory_wordrecall Ewv2_loneliness_quintiles ///
31 Ewv2_cesd_score Ewv2_depressive_symptoms ///
32 Ewv3_cesd_sumscore_rand Ewv3_depressive_symptoms ///
33 Ewv4_cesd_sumscore_rand Ewv4_depressive_symptoms ///
34 Ewv5_cesd_sumscore_rand Ewv5_depressive_symptoms ///
35 Ewv6_cesd_sumscore_rand Ewv6_depressive_symptoms ///
36 Ewv7_cesd_sumscore_rand Ewv7_depressive_symptoms ///
37 Ewv8_cesd_sumscore Ewv8_depressive_symptoms ///
38 Ewv9_cesd_sumscore Ewv9_depressive_symptoms ///
39 Ewv2_crp_level Ewv2_crp Ewv2_fibrinogen_level Ewv2_fibrinogen ///
40 Ewv2_hdl_level Ewv2_male_hdl Ewv2_female_hdl ///
41 Ewv2_meds_hdl Ewv2_cholesterol_evr Ewv2_hdl_sum Ewv2_hdl_cholesterol ///
42 Ewv2_waist Ewv2_malewaist_ao Ewv2_femalewaist_ao Ewv2_obesity_waist_sum Ewv2_obesity_waist ///
43 Ewv2_bmi_score Ewv2_obesity_bmi Ewv2_waist_bmi_sum Ewv2_obesity ///
44 Ewv2_tg_level Ewv2_tg ///
45 Ewv2_systolic_mean Ewv2_diastolic_mean Ewv2_systolic_bp Ewv2_diastolic_bp ///
46 Ewv2_meds_bp Ewv2_bp_reportevr Ewv2_bp_before Ewv2_bp_diagnosed_sum Ewv2_bp_diagnosed Ewv2_bp_sum
47 Ewv2_bp ///
48 Ewv2_diabetes_evr Ewv2_diabetes_before Ewv2_diabetes_diagnosed_sum Ewv2_diabetes_diagnosed ///
49 Ewv2_glucose_level Ewv2_glucose Ewv2_HbA1c_level Ewv2_HbA1c ///
50 Ewv2_meds1_diabetes Ewv2_meds2_diabetes Ewv2_insulin_diabetes Ewv2_diabetes_anymeds_sum
51 Ewv2_diabetes_anymeds ///
52 Ewv2_diabetes_glucose_sum Ewv2_glycemia ///
53 Ewv2_anydementia_iqcode_report Ewv3_anydementia_iqcode_report ///
54 Ewv4_anydementia_iqcode_report Ewv6to9_dementia_event ///
55 Ewv5_anydementia_iqcode_report Ewv6_anydementia_iqcode_report Ewv7_anydementia_iqcode_report ///
56 Ewv8_anydementia_iqcode_report Ewv9_anydementia_iqcode_report ///
57 Ewv2_interview_date Ewv3_interview_date Ewv4_interview_date ///
58 Ewv5_interview_date Ewv6_interview_date Ewv7_interview_date ///
59 Ewv8_interview_date Ewv9_interview_date ///
60 Ewv3to9_dementia_sum Ewv3to9_dementia_sum_no_iqcode ///
61 Ewv3to9_dementia_event Ewv3to9_dementia_event_no_iqcode ///
62 Ewv3to9_dementia_report_or_lasti Ewv3to9_dementia_report_free_dat ///
63 Ewv3to9_newdementia_or_lastinter Ewv3to9_dementia_free_date E_time_dementia_months ///
64 E_time_dementia_report_months_no E_time_dementia_midpoint ///
65 E_time_dementia_midpoint_final E_time_event_dementia E_time_dementia_report_midpoint_ ///
66 E_time_dementia_midpoint_no iqco E_time_event_dementia_report_no_

```

```

66
67
68
69
70
71 /* Latent class analysis - LCA of cardiometabolic risk factors for dementia
72
73 Useful links:
74 https://www.stata.com/meeting/uk18/slides/uk18\_MacDonald.pdf
75
76 https://www.stata.com/meeting/mexico18/slides/5\_Mexico18\_Canette.pdf
77
78 https://www.bgsu.edu/content/dam/BGSU/college-of-arts-and-sciences/center-for-family-and-demographic-research/documents/Workshops/2020-latent-class-analysis.pdf
79
80 https://www.stata.com/features/overview/latent-class-analysis/
81
82 https://www.stata.com/manuals/semexample50g.pdf
83
84 https://www.stata.com/manuals/semexample51g.pdf
85
86 https://www.stata.com/manuals/semexample52g.pdf
87
88 https://www.ucl.ac.uk/population-health-sciences/sites/population\_health\_sciences/files/lca.pdf
89
90 https://www.stata.com/manuals/semgsemlclassoptions.pdf
91
92 https://www.stata.com/meeting/nordic-and-baltic17/slides/nordic-and-baltic17\_Pitblado.pdf
93
94 https://www.frontiersin.org/articles/10.3389/fpsyg.2014.00920/full
95
96 https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2774074
97
98 https://www.statalist.org/forums/forum/general-stata-discussion/general/1412686-calculating-entropy-for-lca-latent-class-analysis-in-stata-15
99
100 https://www.statalist.org/forums/forum/general-stata-discussion/general/1590174-how-to-calculate-entropy-for-lca-with-stata
101
102 https://www.statalist.org/forums/forum/general-stata-discussion/general/1390895-combine-marginsplot-problem-with-plot-options
103
104
105
106
107 * gsem command to fit a latent class model
108
109 gsem (var1 var2 var3 <-), logit lclass(C 3)
110
111 OR TRY
112
113 gsem (var1 var2 var3 <-), logit lclass(C 3) ///
114 startvalues(randompr, draws(20) seed(15) difficult) ///
115 emopts(iterate(30) difficult)
116
117
118
119
120 Binary variables of cardiometabolic markers measured at wave 2
121
122 CRP: Ewv2_crp
123
124 HDL cholesterol: Ewv2_hdl_cholesterol
125
126 Obesity by waist cir: Ewv2_obesity_waist
127
128 systolic Blood pressure: Ewv2_systolic_bp
129

```

```

130     diastolic Blood pressure: Ewv2_diastolic_bp
131
132     Diabetes: Ewv2_diabetes_diagnosed
133
134     HbA1c: Ewv2_HbA1c
135
136
137     */
138
139
140
141
142
143
144     *** Descriptive stats of cardiometabolic markers
145
146     tabulate Ewv2_crp
147     summarize Ewv2_crp
148
149     misstable summarize Ewv2_crp
150     misstable patterns Ewv2_crp
151
152     tabulate Ewv2_hdl_cholesterol
153     summarize Ewv2_hdl_cholesterol
154
155     misstable summarize Ewv2_hdl_cholesterol
156     misstable patterns Ewv2_hdl_cholesterol
157
158     tabulate Ewv2_obesity_waist
159     summarize Ewv2_obesity_waist
160
161     misstable summarize Ewv2_obesity_waist
162     misstable patterns Ewv2_obesity_waist
163
164     tabulate Ewv2_systolic_bp
165     summarize Ewv2_systolic_bp
166
167     misstable summarize Ewv2_systolic_bp
168     misstable patterns Ewv2_systolic_bp
169
170
171     tabulate Ewv2_diastolic_bp
172     summarize Ewv2_diastolic_bp
173
174     misstable summarize Ewv2_diastolic_bp
175     misstable patterns Ewv2_diastolic_bp
176
177
178     tabulate Ewv2_diabetes_diagnosed
179     summarize Ewv2_diabetes_diagnosed
180
181     misstable summarize Ewv2_diabetes_diagnosed
182     misstable patterns Ewv2_diabetes_diagnosed
183
184
185     tabulate Ewv2_HbA1c
186     summarize Ewv2_HbA1c
187
188     misstable summarize Ewv2_HbA1c
189     misstable patterns Ewv2_HbA1c
190
191
192     tabulate Ewv2_anydementia_iqcode_report
193     summarize Ewv2_anydementia_iqcode_report
194
195     misstable summarize Ewv2_anydementia_iqcode_report
196     misstable patterns Ewv2_anydementia_iqcode_report
197

```

```

198
199
200
201
202 *** CLEANING DATA
203
204
205 * 1. drop dementia cases and missing data at baseline
206
207 drop if Ewv2_anydementia_iqcode_report==1
208 * (50 observations deleted)
209
210 drop if Ewv2_anydementia_iqcode_report== .
211 * (0 observations deleted)
212
213
214 * 2. drop missing values of cardiometabolic markers
215
216 drop if Ewv2_crp== .
217 * (1,753 observations deleted)
218
219 drop if Ewv2_hdl_cholesterol== .
220 * (6 observations deleted)
221
222 drop if Ewv2_obesity_waist== .
223 * (133 observations deleted)
224
225 drop if Ewv2_systolic_bp== .
226 * (660 observations deleted)
227
228 drop if Ewv2_diastolic_bp== .
229 * (0 observations deleted)
230
231 drop if Ewv2_diabetes_diagnosed== .
232 * (0 observations deleted)
233
234 drop if Ewv2_HbA1c== .
235 * (102 observations deleted)
236
237
238
239 * 3. drop obs with no records on dementia at any wave from 3-9 follow-ups
240
241
242 search mdesc
243 search rmiss2
244 search mvpatterns
245
246 * see number of missing values vs non-missing in each variable
247 mdesc Ewv3_anydementia_iqcode_report Ewv4_anydementia_iqcode_report ///
248 Ewv5_anydementia_iqcode_report Ewv6_anydementia_iqcode_report Ewv7_anydementia_iqcode_report ///
249 Ewv8_anydementia_iqcode_report Ewv9_anydementia_iqcode_report
250
251
252
253 /* number of missing values per observation
254 * the code below creates a variable called nmisfollowup that gives the number of missing values
255 for each observation in the variables of interest */
256 egen nmisfollowup_dementia_wv3to9=rmiss2(Ewv3_anydementia_iqcode_report ///
257 Ewv4_anydementia_iqcode_report Ewv5_anydementia_iqcode_report ///
258 Ewv6_anydementia_iqcode_report Ewv7_anydementia_iqcode_report ///
259 Ewv8_anydementia_iqcode_report Ewv9_anydementia_iqcode_report)
260
261 tab nmisfollowup_dementia_wv3to9
262
263 * drop observations "nmisfollowup_dementia_wv3to9" > 6 (those with 7 missing data = no records at
any wave)
264 drop if nmisfollowup_dementia_wv3to9>6

```

```

265  *(451 observations deleted)
266
267
268  * FINAL SAMPLE -> 4511
269
270
271
272
273
274  /* Latent Class analysis - gsem
275
276  7 variables: Ewv2_crp Ewv2_hdl_cholesterol Ewv2_obesity_waist
277  Ewv2_systolic_bp Ewv2_diastolic_bp Ewv2_diabetes_diagnosed Ewv2_HbA1c
278
279  */
280
281
282  * change names to start with lowercase (STATA assumes variables starting with a capital letter
  are cont latent variables)
283
284  rename Ewv2_crp crp_lca
285  rename Ewv2_hdl_cholesterol hdl_lca
286  rename Ewv2_obesity_waist obesity_lca
287  rename Ewv2_systolic_bp systolic_lca
288  rename Ewv2_diastolic_bp diastolic_lca
289  rename Ewv2_diabetes_diagnosed diabetes_lca
290  rename Ewv2_HbA1c hba1c_lca
291
292
293
294  * Correlation matrix of the CM variables
295
296  corr crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca hba1c_lca
297
298  pwcorr crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca hba1c_lca, sig
299
300
301  * to create quality table in word - asdoc -
302  * https://www.youtube.com/watch?v=XHBl6PHf0zs&ab\_channel=StataProfessor
303
304  help asdoc
305
306  asdoc pwcorr crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca hba1c_lca, sig
307
308
309  asdoc pwcorr crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca hba1c_lca, nonum
  replace cor crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca hba1c_lca label
  replace star(.05) dec(2)
310
311
312
313  * LCA models
314
315
316  * one-class model
317
318  gsem (crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca ///
319  hba1c_lca <- _cons), family(bernoulli) link(logit) lclass(C 1)
320
321  estimates store oneclass_cm
322
323  * two-class model
324
325  gsem (crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca ///
326  hba1c_lca <- _cons), family(bernoulli) link(logit) lclass(C 2)
327
328
329  estimates store twoclass_cm

```

```

330
331
332 * three-class model
333
334 gsem (crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca ///
335 hba1c_lca <- _cons), family(bernoulli) link(logit) lclass(C 3)
336
337
338 gsem (crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca ///
339 hba1c_lca <- _cons), family(bernoulli) link(logit) lclass(C 3) ///
340 startvalues(randompr, draws(20) seed(15) difficult) ///
341 emopts(iterate(30) difficult)
342
343 estimates store threeclass_cm
344
345 * four-class model
346
347 gsem (crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca ///
348 hba1c_lca <- _cons), family(bernoulli) link(logit) lclass(C 4)
349
350 gsem (crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca ///
351 hba1c_lca <- _cons), family(bernoulli) link(logit) lclass(C 4) ///
352 startvalues(randompr, draws(20) seed(15) difficult) ///
353 emopts(iterate(30) difficult)
354
355
356
357
358 /*
359
360 ** Evaluating Fit to choose the number of classes **
361
362 1. a priori theory
363 2. Information Statistics
364     AIC, BIC, adjusted BIC
365 3. Chi-Square goodness of fit
366 4. Entropy
367
368 Others but not used here:
369 Lo-Mendell-Rubin (LMR)
370     Not recommended (designed for normal Y)
371 Bootstrapped Likelihood Ratio Test
372
373 */
374
375 * AIC and BIC to determine which of these models fits best
376
377 estimates stats oneclass_cm twoclass_cm threeclass_cm
378
379
380
381 * LCA postestimation
382 * Latent class marginal probabilities - lcprob -
383 * Latent class marginal means - lcmean -
384
385
386 estat lcprob
387
388 estat lcmean
389
390
391
392 * likelihood -ratio test (G2) to evaluate whether our model fits as well as the saturated model
393
394 estat lcgof
395
396 * if p value is sig means that we reject the null hypothesis and the model doesn't fit well
397 * if p value is not sig means that we fail to reject the null hypothesis that our model fits as

```

```

well as the saturated model.
398 * for a well-fitted model p value should be non-sig. A significant p-value indicates lack of
model fit in absolute terms.
399
400 * 3-class model:  $p < 0.001$ 
401
402
403
404 * Entropy
405
406 quietly predict classpost*, classposteriorpr
407 gen sum_p_lnp = 0
408 forvalues k = 1/2 {
409     replace sum_p_lnp = sum_p_lnp + classpost`k'*ln(classpost`k')
410 }
411 summ sum_p_lnp, meanonly
412 scalar E = 1+`r(sum)'/(`e(N)*ln(2))
413 drop classpost? sum_p_lnp
414 di E
415
416
417
418
419
420
421
422
423 /* We can use the predictions of the posterior probability of class membership to evaluate an
424 individual's probability of being in each class.
425
426 */
427
428 predict classpost1*, classposteriorpr
429 list in 1, abbrev(10)
430
431 /* We can determine the expected class for each individual based on whether the posterior
432 probability
433 is greater than 0.5
434 */
435 generate expclass1 = 1 + (classpost11>0.5)
436 tabulate expclass1
437
438
439 generate expclass2 = 1 + (classpost12>0.5)
440 tabulate expclass2
441
442
443 generate expclass3 = 1 + (classpost13>0.5)
444 tabulate expclass3
445
446
447
448 /* We can determine expected classification for each individual in the dataset based on the
449 predicted
450 posterior class probabilities.
451 */
452 predict cpost*, classposteriorpr
453 egen max = rowmax(cpost*)
454
455
456 * generate classes var
457
458 generate predclass = 1 if cpost1==max
459
460 replace predclass = 2 if cpost2==max
461

```

```

462  replace predclass = 3 if cpost3==max
463
464  tabulate predclass
465
466
467
468  *****function to print out summary stats
469  program summary_table_proclCla
470      preserve
471      *look at the average posterior probability
472      gen Mp = 0
473      foreach i of varlist cpost* {
474          replace Mp = `i' if `i' > Mp
475      }
476      sort predclass
477      *and the odds of correct classification
478      by predclass: gen countG = _N
479      by predclass: egen groupAPP = mean(Mp)
480      by predclass: gen counter = _n
481      gen n = groupAPP/(1 - groupAPP)
482      gen p = countG/ _N
483      gen d = p/(1-p)
484      gen occ = n/d
485      *Estimated proportion for each group
486      scalar c = 0
487      gen TotProb = 0
488      foreach i of varlist cpost* {
489          scalar c = c + 1
490          quietly summarize `i'
491          replace TotProb = r(sum)/ _N if predclass == c
492      }
493      gen d_pp = TotProb/(1 - TotProb)
494      gen occ_pp = n/d_pp
495      *This displays the group number [_traj_~p],
496      *the count per group (based on the max post prob), [countG]
497      *the average posterior probability for each group, [groupAPP]
498      *the odds of correct classification (based on the max post prob group assignment), [occ]
499      *the odds of correct classification (based on the weighted post. prob), [occ_pp]
500      *and the observed probability of groups versus the probability [p]
501      *based on the posterior probabilities [TotProb]
502      list predclass countG groupAPP occ occ_pp p TotProb if counter == 1
503      restore
504  end
505
506  summary_table_proclCla
507
508
509
510
511  /*
512  Class variable manipulation
513  */
514
515  * 3-class model: rename  predclass to E_lca_group3
516
517  rename  predclass E_lca_group3
518
519  * labelling variable of E_lca_group3
520
521  label var E_lca_group3 "Latent classes 3 groups of cardiometabolic markers"
522
523  * labelling values
524  lab def lca_group3 1 "Healthy with obesity" 2 "Obesity and Hypertension" 3 "Complex
cardiometabolic disorders"
525
526  * attach category labels to the variable through label value
527
528  lab val E_lca_group3 lca_group3

```



```

529
530 ta E_lca_group3
531
532
533
534 * margins and marginsplot
535
536 * use margins to calculate marginal predictions
537 * use marginsplot to graph marginal predictions
538
539
540 margins, predict(classpr class(1)) ///
541             predict(classpr class(2)) ///
542             predict(classpr class(3))
543
544 marginsplot, xtitle("") ytitle("") ///
545             xlabel(1 "Class 1" 2 "Class 2" 3 "Class 3") ///
546             title("Predicted Latent Class Probabilities with 95% CI")
547
548
549 margins, predict(classpr class(1)) ///
550             predict(classpr class(2)) ///
551             predict(classpr class(3))
552 marginsplot, recast(bar) xtitle("") ytitle("") ///
553             xlabel(1 "Class 1" 2 "Class 2" 3 "Class 3") ///
554             title("Predicted Latent Class Probabilities with 95% CI")
555
556
557 margins, predict(outcome(hba1c_lca) class(1)) ///
558             predict(outcome(hba1c_lca) class(2)) ///
559             predict(outcome(hba1c_lca) class(3))
560 marginsplot, recast(bar) xtitle("") ytitle("") ///
561             xlabel(1 "Class 1" 2 "Class 2" 3 "Class 3") ///
562             title("Predicted Pr(HbA1c=1) with 95% CI")
563
564 * repeat with all CM var
565
566
567 * class 1
568
569 margins, predict(outcome(crp_lca) class(1)) ///
570             predict(outcome(hdl_lca) class(1)) ///
571             predict(outcome(obesity_lca) class(1)) ///
572             predict(outcome(systolic_lca) class(1)) ///
573             predict(outcome(diastolic_lca) class(1)) ///
574             predict(outcome(diabetes_lca) class(1)) ///
575             predict(outcome(hba1c_lca) class(1)) ///
576
577
578
579 marginsplot, recast(bar) title("Class 1") xtitle("") ///
580             xlabel(1 "crp" 2 "hdl" 3 "obesity" 4 "systolic BP" ///
581             5 "diastolic BP" 6 "diabetes" 7 "hba1c", angle(45)) ///
582             ytitle("Predicted mean") ylabel(0(.20)1) name(class1)
583
584
585 * class 2
586
587 margins, predict(outcome(crp_lca) class(2)) ///
588             predict(outcome(hdl_lca) class(2)) ///
589             predict(outcome(obesity_lca) class(2)) ///
590             predict(outcome(systolic_lca) class(2)) ///
591             predict(outcome(diastolic_lca) class(2)) ///
592             predict(outcome(diabetes_lca) class(2)) ///
593             predict(outcome(hba1c_lca) class(2)) ///
594
595
596

```

```

597 marginsplot, recast(bar) title ("Class 2") xtitle("") ///
598       xlabel(1 "crp" 2 "hdl" 3 "obesity" 4 "systolic BP" ///
599       5 "diastolic BP" 6 "diabetes" 7 "hba1c", angle(45)) ///
600       ytitle("Predicted mean") ylabel(0(.20)1)name (class2)
601
602
603
604
605 * class 3
606
607 margins, predict(outcome(crp_lca) class(3)) ///
608       predict(outcome(hdl_lca) class(3)) ///
609       predict(outcome(obesity_lca) class(3)) ///
610       predict(outcome(systolic_lca) class(3)) ///
611       predict(outcome(diastolic_lca) class(3)) ///
612       predict(outcome(diabetes_lca) class(3)) ///
613       predict(outcome(hba1c_lca) class(3)) ///
614
615
616
617 marginsplot, recast(bar) title ("Class 3") xtitle("") ///
618       xlabel(1 "crp" 2 "hdl" 3 "obesity" 4 "systolic BP" ///
619       5 "diastolic BP" 6 "diabetes" 7 "hba1c", angle(45)) ///
620       ytitle("Predicted mean") ylabel(0(.20)1) name (class3)
621
622
623
624 graph combine class1 class2 class3, cols(3)
625
626
627
628
629
630
631
632
633 * ----- *
634
635
636
637
638
639
640 /* MULTIPLE IMPUTATION (MI)
641
642 To handle with missing values of covariates
643
644
645 useful sources for MI and MICE:
646
647 https://stats.idre.ucla.edu/stata/seminars/mi\_in\_stata\_pt1\_new/
648 https://www.stata.com/manuals/mi.pdf - see page 139
649 https://www.stata.com/meeting/switzerland16/slides/medeiros-switzerland16.pdf
650 https://www.youtube.com/watch?v=i6S0lq0mjuc&ab\_channel=StataCorpLLC
651 https://dss.princeton.edu/training/MIStata.pdf
652
653
654
655 Preparing to conduct MI
656 1. examine the number and proportion of missing values among the variables of interest
657     use the mdesc command
658 2. examine missing data patterns
659     use commands mi set and mi misstable patterns
660 3. identify potential auxiliary variables
661
662
663 Run MI using chained equations (MICE)
664 using the commands

```

```

665 1. how (in what style) to store the imputations
666     mi set wide
667 2. which variables will be imputed
668     mi register imputed
669 3. optionally, which variables will not be imputed
670     mi register regular
671 4. what imputation method is implemented to impute each of var - MICE
672     mi impute chained
673
674 */
675
676
677
678
679
680 /*
681
682 1. examining missing values
683     install packages:
684     * install mdesc
685     * install tabmiss
686     * insatll dm31
687     * insall mvpatterna
688
689 */
690
691 search mdesc
692 search rmiss2
693 search mvpatterns
694
695
696 * examining number of missing values vs non-missing in each variable
697
698 mdesc E_age E_sex E_education E_maritalstatus_4cat E_wealthquintiles ///
699 E_smoking_3cat E_physicalactivity E_alcohol_status E_cvd_comorbidity ///
700 E_memory_wordrecall Ewv2_depressive_symptoms
701
702
703
704
705 * examining missing data patterns
706
707 mi set wide
708
709 mi misstable summarize E_age E_sex E_education E_maritalstatus_4cat E_wealthquintiles ///
710 E_smoking_3cat E_physicalactivity E_alcohol_status E_cvd_comorbidity ///
711 E_memory_wordrecall Ewv2_depressive_symptoms
712
713
714 mi misstable patterns E_age E_sex E_education E_maritalstatus_4cat E_wealthquintiles ///
715 E_smoking_3cat E_physicalactivity E_alcohol_status E_cvd_comorbidity ///
716 E_memory_wordrecall Ewv2_depressive_symptoms
717
718
719 /*
720     identifying potential auxiliary var
721     * Auxiliary variables are either correlated with a missing variable(s)
722     (the recommendation is  $r > 0.4$ ) or are believed to be associated with missingness
723     - a priori knowledge of var that would make good auxiliary var
724     - identify potential candidates by examining associations between missing var and other var in
       the dataset
725         running correlation using the command: pwcorr v1 v2 v3, obs
726         the recommnedation for good correlation is  $r > 0.4$ 
727
728
729 Missing var to be imputed:
730
731     E_education E_wealthquintiles

```

```

732     E_smoking_3cat E_physicalactivity E_alcohol_status
733     E_memory_wordrecall Ewv2_depressive_symptoms
734
735
736
737 Potential auxiliary var:
738 DV:  Ewv3to9_dementia_event
739 IV:  crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca hba1c_lca
740 other var:
741     E_age E_sex E_maritalstatus_4cat E_cvd_comorbidity
742
743 */
744
745
746 * correlation
747
748 pwcorr E_education E_wealthquintiles ///
749     E_smoking_3cat E_physicalactivity E_alcohol_status ///
750     E_memory_wordrecall Ewv2_depressive_symptoms ///
751     Ewv3to9_dementia_event crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca ///
752     diabetes_lca hba1c_lca ///
753     E_age E_sex E_maritalstatus_4cat E_cvd_comorbidity, obs
754
755
756 /* The correlation showed that all the following var are good auxiliary:
757 Ewv3to9_dementia_event obesity_lca diabetes_lca E_age E_sex E_maritalstatus_4cat E_cvd_comorbidity
758 A good auxiliary does not have to be correlated with every variable to be useful
759 And it's not problematic if it has missing info of it's own
760 */
761
762
763
764 /*
765 MI by chained equations (MICE)
766     see: https://stats.idre.ucla.edu/stata/seminars/mi\_in\_stata\_pt1\_new/
767
768 MICE is known as the fully conditional specification or sequential generalized regression
769 does not assume a joint MVN distribution
770 but instead uses a separate conditional distribution for each imputed variable.
771
772 The multivariate normal (MVN) model - mi imputed mvn -
773 assumes multivariate normality of all var
774
775 The multivariate imputation by chained equations (MICE) - mi imputed chained -
776 offers flexibility in how each var is modeled
777
778 mi impute chained allows to specify models for a
779 variety of variable types, including
780 continuous, binary, ordinal, nominal, truncated, and count variables
781
782
783 The MICE distributions available in Stata are:
784 binary, ordered and multinomial logistic regression for categorical variables,
785 linear regression and predictive mean matching (PMM)* for continuous variables,
786 and Poisson and negative binomial regression for count variables.
787
788
789
790 IMPUTATION PHASES
791
792 1. mi set wide
793     style to store imputations
794
795 2. mi register imputed
796     identifies which variables in the imputation model have missing information.
797
798 3. mi register regular (! optional)
799     which variables will not be imputed

```

```

800
801 4. mi impute chained
802     where the user specifies the imputation model to be used
803     and the number of imputed datasets to be created.
804     Example:
805         mi impute chained (regress) bmi age (logit) female ///
806         (mlogit) race = bpdiast i.region, add(20)
807
808 5. mi estimate
809     is used as a prefix to the standard regress command.
810     This executes the specified estimation model within each of the 20 imputed datasets
811     to obtain 20 sets of coefficients and standard errors.
812     Stata then combines these estimates to obtain one set of inferential statistics.
813     In the output from mi estimate you will see some metrics: Imputation Diagnostics
814     information for RVI (Relative Increase in Variance),
815     FMI (Fraction of Missing Information),
816     DF (Degrees of Freedom) ,
817     RE (Relative Efficiency),
818     and the between imputation and the within imputation variance estimates
819     to examine how the standard errors (SEs) are calculated.
820
821
822 -----
823
824 SELECTING MY IMPUTATION MODEL
825
826 - MICE -> mi impute chained
827
828 - var to be imputed:
829
830     linear regression for continuous var (regress) ->
831     E_memory_wordrecall
832
833     logistic for the binary var (logit) ->
834     Ewv2_depressive_symptoms
835
836
837     multinomial logistic for our nominal categorical var (mlogit) ->
838     E_education E_wealthquintiles
839     E_smoking_3cat E_physicalactivity E_alcohol_status
840
841
842
843 - auxiliary var:
844
845     DV -> Ewv3to9_dementia_event
846     IV -> obesity_lca diabetes_lca
847     other covariates -> E_age E_sex E_maritalstatus_4cat E_cvd_comorbidity
848
849
850
851 - imputation numbers (m) -> 10
852
853     White et al. (2010) recommendation: use the rule that m should equal the percentage of
854     incomplete cases
855
856 - rseed (53421) for reproducability reasons
857
858
859 - (! OPTIONAL) advance impute options -> force
860
861     proceed with imputation, even when missing imputed values (e.g. auxiliary have missing data)
862     are encountered
863
864 - impute options -> savetrace (trace1)
865
866     specifies Stata to save the means and standard deviations of imputed values from each

```

```

iteration to a Stata dataset named "trace1
*/
866
867
868
869 mi set wide
870
871
872 mi register imputed E_education E_wealthquintiles ///
873     E_smoking_3cat E_physicalactivity E_alcohol_status ///
874     E_memory_wordrecall Ewv2_depressive_symptoms
875
876
877
878 mi impute chained (logit) Ewv2_depressive_symptoms ///
879 (mlogit) E_education E_wealthquintiles ///
880     E_smoking_3cat E_physicalactivity E_alcohol_status ///
881 (regress) E_memory_wordrecall = Ewv3to9_dementia_event obesity_lca diabetes_lca ///
882     E_age E_sex E_maritalstatus_4cat E_cvd_comorbidity, add(10) rseed(53421) savetrace(trace1)
883
884
885
886
887 * save imputed data
888
889 * plot imputations
890
891
892 *it will open a file named trace1
893 use trace1, clear
894
895 describe
896
897
898 reshape wide *mean *sd, i(iter) j(m)
899
900 tsset iter
901
902
903
904
905 /*
906 The trace plot below graphs the predicted means value produced during the first imputation chain.
907 As before, the expectations is that the values would vary randomly to incorporate variation into
908 the predicted values for read.
909 */
910
911 tsline E_education_mean1, name(mice1,replace)legend(off) ytitle("Mean of education")
912 tsline E_wealthquintiles_mean1, name(mice1,replace)legend(off) ytitle("Mean of wealth")
913 tsline E_smoking_3cat_mean1, name(mice1,replace)legend(off) ytitle("Mean of smoking")
914 tsline E_physicalactivity_mean1, name(mice1,replace)legend(off) ytitle("Mean of physical activity")
915 tsline E_alcohol_status_mean1, name(mice1,replace)legend(off) ytitle("Mean of alcohol status")
916 tsline E_memory_wordrecall_mean1, name(mice1,replace)legend(off) ytitle("Mean of memory")
917
918 /*
919
920 All 10 imputation chains can also be graphed simultaneously to make sure that nothing unexpected
921 occurred in a single chain.
922 Every chain is obtained using a different set of initial values and this should be unique.
923 Each colored line represents a different imputation.
924 So all 10 imputation chains are overlaid on top of one another.
925 */
926
927
928 tsline E_memory_wordrecall_mean*, name(mice1,replace)legend(off) ytitle("Mean of memory")
929 tsline E_memory_wordrecall_sd*, name(mice2, replace) legend(off) ytitle("SD of memory")
930 graph combine mice1 mice2, xcommon cols(1) title(Trace plots of summaries of imputed values)

```

```

931
932 * repeat for each imputed var
933
934
935
936
937
938
939 /*
940 ---- DESCRIPTIVE STATISTICS ----
941
942 General characteristics of participants
943
944 General characteristics of participnats stratified for study inclusion
945
946 General characteristics of participants stratified for dementia occurence
947
948 Participant characteristics by CM 3-class groups
949
950 CHI-SQUARE (chi2) for categorical var (crosstabulation)
951     Frequency tables -> two-way tables
952         using the command tabulate, chi2
953         reporting observations, column percentage (N, %) and p-value of Pearson's r
954
955 one-way ANOVA for continuous var
956     check box plot
957     using the command oneway
958     reporting mean, sd (summary tables) and p-value of F
959 */
960
961
962 * General characteristics of ELSA participants at baseline
963
964 * Socio-demographics
965 sum E_age
966 ta E_sex
967 ta E_education
968 ta E_maritalstatus_4cat
969 ta E_wealthquintiles
970 * Cardiometabolic disorders
971 ta crp_lca
972 ta hdl_lca
973 ta obesity_lca
974 ta systolic_lca
975 ta diastolic_lca
976 ta diabetes_lca
977 ta hba1c_lca
978 * Cardiovascular health factors
979 ta E_smoking_3cat
980 ta E_physicalactivity
981 ta E_alcohol_status
982 ta E_cvd_comorbidity
983 * Depressive symptoms (cont and categ)
984 sum Ewv2_cesd_score
985 ta Ewv2_depressive_symptoms
986 * Memory score
987 sum E_memory_wordrecall
988
989
990
991 * General baseline characteristics of ELSA participants by dementia status
992
993 * Socio-demographics
994 ttest E_age, by(Ewv3to9_dementia_event)
995 ta E_sex Ewv3to9_dementia_event, chi2 column row
996 ta E_education Ewv3to9_dementia_event, chi2 column row
997 ta E_maritalstatus_4cat Ewv3to9_dementia_event, chi2 column row
998 ta E_wealthquintiles Ewv3to9_dementia_event, chi2 column row

```



```

999  * Cardiometabolic disorders
1000  ta crp_lca Ewv3to9_dementia_event, chi2 column row
1001  ta hdl_lca Ewv3to9_dementia_event, chi2 column row
1002  ta obesity_lca Ewv3to9_dementia_event, chi2 column row
1003  ta systolic_lca Ewv3to9_dementia_event, chi2 column row
1004  ta diastolic_lca Ewv3to9_dementia_event, chi2 column row
1005  ta diabetes_lca Ewv3to9_dementia_event, chi2 column row
1006  ta hba1c_lca Ewv3to9_dementia_event, chi2 column row
1007  * Cardiovascular health factors
1008  ta E_smoking_3cat Ewv3to9_dementia_event, chi2 column row
1009  ta E_physicalactivity Ewv3to9_dementia_event, chi2 column row
1010  ta E_alcohol_status Ewv3to9_dementia_event, chi2 column row
1011  ta E_cvd_comorbidity Ewv3to9_dementia_event, chi2 column row
1012  * Depressive symptoms (cont and categ)
1013  ttest Ewv2_cesd_score, by(Ewv3to9_dementia_event)
1014  ta Ewv2_depressive_symptoms Ewv3to9_dementia_event, chi2 column row
1015  * Memory score
1016  ttest E_memory_wordrecall, by(Ewv3to9_dementia_event)
1017
1018
1019
1020  * Sample characteristics by CM 3-class groups
1021  * crosstabs categ var (frequencies and chi2) !report column percentage!
1022  * oneway ANOVA cont var (mean, sd)
1023
1024
1025  * Socio-demographics
1026  oneway E_age E_lca_group3, tabulate
1027  ta E_sex E_lca_group3, chi2 column row
1028  ta E_education E_lca_group3, chi2 column row
1029  ta E_maritalstatus_4cat E_lca_group3, chi2 column row
1030  ta E_wealthquintiles E_lca_group3, chi2 column row
1031  * Cardiovascular health factors
1032  ta E_smoking_3cat E_lca_group3, chi2 column row
1033  ta E_physicalactivity E_lca_group3, chi2 column row
1034  ta E_alcohol_status E_lca_group3, chi2 column row
1035  ta E_cvd_comorbidity E_lca_group3, chi2 column row
1036  * Depressive symptoms (cont and categ)
1037  oneway Ewv2_cesd_score E_lca_group3, tabulate
1038  ta Ewv2_depressive_symptoms E_lca_group3, chi2 column row
1039  * Memory score
1040  oneway E_memory_wordrecall E_lca_group3, tabulate
1041
1042
1043
1044
1045
1046  /*
1047  ---- SURVIVAL ANALYSIS IN COMPLETE DATA ----
1048
1049  Tests of proportional-hazards assumption
1050  Kaplan Meier survival curves
1051  Person-time
1052  Cox proportional regression - Hazard ratios - stcox
1053  Postestimation tools for stcox
1054  Test of Goodness of Fit
1055
1056  *** Cox regression in full data, complete data (listwise deletion of missing data) and imputed data
1057  Cox PH regression in complete data
1058  Cox PH regression model in imputed dataset - mi estimate
1059
1060
1061  */
1062
1063
1064
1065  * check dataset variables of interest only
1066

```



```

1067 codebook E_time_event_dementia Ewv3to9_dementia_event E_lca_group3 ///
1068 E_age E_sex E_education E_maritalstatus_4cat E_wealthquintiles ///
1069 E_smoking_3cat E_physicalactivity E_alcohol_status E_cvd_comorbidity ///
1070 Ewv2_depressive_symptoms E_memory_wordrecall,compact
1071
1072
1073
1074 * Declare Data to be Survival Data
1075 * Time to event: E_time_event_dementia (months)
1076 * Censoring: Ewv3to9_dementia_event (1=dementia, 0=censored)
1077 * Command is stset TIMETOEVENT, failure(CENSORVARIABLE)
1078
1079
1080 stset E_time_event_dementia, failure (Ewv3to9_dementia_event==1) id(idauniq)
1081
1082
1083 *describe survival data using commnad stsum
1084
1085 stsum
1086
1087 stsum, by(E_lca_group3)
1088
1089
1090
1091 * Kaplan Meier Curve estimation
1092
1093 sts list
1094
1095 sts list, by(E_lca_group3)
1096
1097
1098
1099 * Kaplan Meier Curve Plot
1100
1101 * no frills plot
1102
1103 sts graph
1104
1105 * with frills
1106
1107 sts graph, xtitle("Time in Months") ytitle("Survival Prob") ///
1108 title("Kaplan Meier Curve") subtitle("n=4511, # events=284") ///
1109 caption("graph02.png"), size(vsmall))
1110
1111
1112 * With Greenwood CI limits
1113
1114 sts graph, gwood legend(off) xtitle("Time in Months") ytitle("Survival Prob") ///
1115 title("Kaplan Meier Curve") subtitle("n=4511, # events=284") caption("graph03.png", size(vsmall))
1116
1117
1118
1119
1120 * Group Kaplan-Meier Curve Estimation
1121 * Command is sts graph, by(GROUPVAR) OPTION OPTION OPTION Note: Must have sorted by GROUPVAR first
1122
1123 sort E_lca_group3
1124
1125 sts list, by(E_lca_group3)
1126
1127 * graph with frills
1128
1129 sts graph, by(E_lca_group3) xlabel(0(20)180) ylabel(0.80(.05)1) xtitle("Time in Months") ///
1130 ytitle("Survival Prob") title("Kaplan Meier Curve") subtitle("n=4511, # events=284") ///
1131 caption("graph04.png", size(vsmall))
1132
1133
1134

```

```

1135
1136 * calculate person-time and incidence rates using command stptime
1137
1138 stptime, title(Person-years)
1139
1140 stptime, title(Person-years) per(1000)
1141
1142 stptime, title(Person-years) per(10000)
1143
1144
1145 * calculate person-time by category of E_lca_group3
1146
1147 stptime, by(E_lca_group3)
1148
1149 stptime, by(E_lca_group3) per(1000)
1150
1151
1152
1153
1154
1155 * mean and median of follow-up
1156 sum E_time_event_dementia
1157 sum E_time_event_dementia, detail
1158
1159
1160
1161 /* Log Rank Test of equality of survival distributions
1162 (NULL: equality of survival distributions among E_lca_group3 groups)
1163 We will consider including the predictor if the test has a p-value of 0.2 - 0.25 or less.
1164 If the predictor has a p-value greater than 0.25 in a univariate analysis
1165 it is highly unlikely that it will contribute anything to a model which includes other
predictors.
1166 Command is sts test GROUPVAR
1167 */
1168
1169
1170 sts test E_lca_group3, logrank
1171
1172 sts test E_age, logrank
1173
1174 sts test E_sex, logrank
1175
1176 sts test E_education, logrank
1177
1178 sts test E_maritalstatus_4cat, logrank
1179
1180 sts test E_wealthquintiles, logrank
1181
1182 sts test E_smoking_3cat, logrank
1183
1184 sts test E_physicalactivity, logrank
1185
1186 sts test E_alcohol_status, logrank
1187
1188 sts test E_cvd_comorbidity, logrank
1189
1190 sts test Ewv2_depressive_symptoms, logrank
1191
1192 sts test E_memory_wordrecall, logrank
1193
1194
1195
1196
1197
1198 /* Cox PH regression model
1199
1200 using the command stcox
1201

```

```

1202 --- Building the model ---
1203
1204 Model 1: unadjusted - single predictor of CM classes
1205 Model 2: model 1 + sociodemographics: age sex education marital status and wealth
1206 Model 3: model 2 + cvd health: smoking, alcohol consumption, cvd comorbidity
1207 Model 4: model 3 + mental health: depressive symptoms
1208
1209
1210 !! I didn't adjust for physical activity because this variable can't be used in CHARLS (missing
values)
1211
1212 */
1213
1214
1215 * Unadjusted model - model 1 - single predictor
1216
1217 stcox E_lca_group3
1218
1219 * define design var by using i.(3 classes)
1220
1221 stcox i.E_lca_group3
1222
1223
1224 * Adjusted models - multivariable Cox model
1225 * controlling for covariates
1226
1227 * model 2: model 1 + adjust for sociodemographics: age sex education marital status and wealth
1228
1229 stcox i.E_lca_group3 E_age i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles
1230
1231
1232 * model 3: model 2 + adjust for cvd health
1233
1234 stcox i.E_lca_group3 E_age i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles ///
1235 i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity
1236
1237
1238 * model 4: model 3 + adjust for depression
1239
1240 stcox i.E_lca_group3 E_age i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles ///
1241 i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity ///
1242 i.Ewv2_depressive_symptoms
1243
1244
1245
1246
1247
1248
1249 * Coefficients instead of hazard ratios by specifying the option nohr
1250
1251 stcox i.E_lca_group3, nohr
1252
1253
1254 stcox i.E_lca_group3 E_age i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles ///
1255 i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity ///
1256 i.Ewv2_depressive_symptoms, nohr
1257
1258
1259
1260
1261
1262 * Multivariable model development
1263 * Likelihood-ratio tests
1264
1265
1266
1267 *install eststo
1268 findit eststo

```

```

1269
1270
1271 * ---- rx controlling for age and sex -----*
1272 quietly: stcox E_age i.E_sex
1273 eststo modelagesex
1274
1275 quietly: stcox E_age i.E_sex i.E_lca_group3
1276 eststo modelagesex_3group
1277
1278 lrtest modelagesex modelagesex_3group
1279
1280
1281
1282 * ---- rx controlling for sociodemographics -----*
1283 quietly: stcox E_age i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles
1284 eststo modelsociodemo
1285
1286 quietly: stcox E_age i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles i.
1287 E_lca_group3
1288 eststo modelsociodemo_3group
1289
1290 lrtest modelsociodemo modelsociodemo_3group
1291
1292
1293 * ---- rx controlling for cardiovascular health -----*
1294 quietly: stcox i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity
1295 eststo modelcardiovascular
1296
1297 quietly: stcox i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity i.E_lca_group3
1298 eststo modelcardiovascular_3group
1299
1300 lrtest modelcardiovascular modelcardiovascular_3group
1301
1302
1303 * ---- rx controlling for mental health-----*
1304 quietly: stcox i.Ewv2_depressive_symptoms
1305 eststo modelmentalcogn
1306
1307 quietly: stcox i.Ewv2_depressive_symptoms i.E_lca_group3
1308 eststo modelmentalcogn_3group
1309
1310 lrtest modelmentalcogn modelmentalcogn_3group
1311
1312
1313
1314
1315 * side-by-side comparison of models
1316
1317
1318 quietly: stcox i.E_lca_group3
1319 eststo model1
1320
1321
1322 quietly: stcox E_age i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles i.
1323 E_lca_group3
1324 eststo model2
1325
1326 quietly: stcox E_age i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles ///
1327 i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity i.E_lca_group3
1328 eststo model3
1329
1330
1331 quietly: stcox E_age i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles ///
1332 i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity ///
1333 i.Ewv2_depressive_symptoms i.E_lca_group3
1334 eststo model4

```

```

1335
1336
1337
1338
1339
1340 * Display Betas and Summary Statistics
1341 estout model1 model2 model3 model4, stats(n chi2 bic, star(chi2)) prehead("Betas")
1342
1343 /* Key Interpretation
1344 Chi2 = Value of LR test comparing the model fit ("full") to intercept only ("reduced")
1345 bic = Schwarz' Bayesian Information Criterion = It is a function of the log-likelihood.
1346 Smaller values indicate a better fit.
1347 */
1348
1349 * Display Hazard Ratios and Model Fit Statistics. Option eform produces hazard ratios
1350 estout model1 model2 model3 model4, eform stats(n chi2 bic, star(chi2)) prehead("Hazard Ratios")
1351
1352
1353
1354
1355 * Postestimation tools for stcox
1356
1357 * Test of proportional hazards
1358
1359 estat phtest, detail
1360
1361
1362 /* Proportionality Assumption - method 1
1363 We will check proportionality by including time-dependent covariates in the model
1364 by using the tvc and the texp options in the stcox command.
1365 Time dependent covariates are interactions of the predictors and time.
1366 In this analysis we choose to use the interactions with log(time)
1367 because this is the most common function of time used in time-dependent covariates
1368 but any function of time could be used.
1369 If a time-dependent covariate is significant this indicates
1370 a violation of the proportionality assumption for that specific predictor.
1371 The conclusion is that all of the time-dependent variables are not significant
1372 either collectively or individually thus supporting the assumption of proportional hazard.
1373 */
1374
1375
1376
1377 stcox i.E_lca_group3 E_age i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles ///
1378 i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity ///
1379 i.Ewv2_depressive_symptoms, nohr ///
1380 tvc(E_lca_group3 E_age E_sex E_education E_maritalstatus_4cat E_wealthquintiles ///
1381 E_smoking_3cat E_alcohol_status E_cvd_comorbidity ///
1382 Ewv2_depressive_symptoms) texp(ln(E_time_event_dementia))
1383
1384
1385
1386 /* Proportionality Assumption - method 2
1387 by using the Schoenfeld and scaled Schoenfeld residuals
1388 In the stptest command we test the proportionality of the model as a whole
1389 and by using the detail option we get a test of proportionality for each predictor.
1390 By using the plot option we can also obtain a graph of the scaled Schoenfeld assumption.
1391 If the tests in the table are not significant (p-values over 0.05)
1392 then we can not reject proportionality and we assume
1393 that we do not have a violation of the proportional assumption.
1394 The stphplot command uses log-log plots to test proportionality
1395 and if the lines in these plots are parallel then we have further indication
1396 that the predictors do not violate the proportionality assumption.
1397 */
1398
1399 quietly stcox E_lca_group3 E_age E_sex E_education E_maritalstatus_4cat E_wealthquintiles ///
1400 E_smoking_3cat E_alcohol_status E_cvd_comorbidity ///
1401 Ewv2_depressive_symptoms, schoenfeld(sch*) scaledsch(sca*)
1402 stptest, detail

```

```

1403 stptest, plot(E_lca_group3) msym(oh)
1404 stptest, plot(E_age) msym(oh)
1405 stptest, plot(E_sex) msym(oh)
1406 stptest, plot(E_education) msym(oh)
1407 stptest, plot(E_maritalstatus_4cat) msym(oh)
1408 stptest, plot(E_wealthquintiles) msym(oh)
1409 stptest, plot(E_smoking_3cat) msym(oh)
1410 stptest, plot(E_alcohol_status) msym(oh)
1411 stptest, plot(E_cvd_comorbidity) msym(oh)
1412 stptest, plot(Ewv2_depressive_symptoms) msym(oh)
1413
1414
1415
1416
1417
1418
1419 stphplot, by(E_lca_group3) plot1(msym(oh)) plot2(msym(th))
1420 stphplot, by(E_age) plot1(msym(oh)) plot2(msym(th))
1421 stphplot, by(E_sex) plot1(msym(oh)) plot2(msym(th))
1422 stphplot, by(E_education) plot1(msym(oh)) plot2(msym(th))
1423 stphplot, by(E_maritalstatus_4cat) plot1(msym(oh)) plot2(msym(th))
1424 stphplot, by(E_wealthquintiles) plot1(msym(oh)) plot2(msym(th))
1425 stphplot, by(E_smoking_3cat) plot1(msym(oh)) plot2(msym(th))
1426 stphplot, by(E_alcohol_status) plot1(msym(oh)) plot2(msym(th))
1427 stphplot, by(E_cvd_comorbidity) plot1(msym(oh)) plot2(msym(th))
1428 stphplot, by(Ewv2_depressive_symptoms) plot1(msym(oh)) plot2(msym(th))
1429
1430
1431
1432
1433 * Assessment of PH Assumption: adjust for age and sex
1434 stphplot, by(E_lca_group3) adjust(E_age E_sex) nolntime plot1opts(symbol(none) color(black)
lpattern(dash)) ///
1435 plot2opts(symbol(none) color(green)) plot3opts(symbol(none) color(red)) ///
1436 title("Assessment of PH Assumption") subtitle(" Predictor is E_lca_group3") xtitle("months")
1437
1438
1439
1440 * Assessment of PH Assumption: adjust for model 2
1441 stphplot, by(E_lca_group3) adjust(E_age E_sex E_education E_maritalstatus_4cat E_wealthquintiles)
///
1442 nolntime plot1opts(symbol(none) color(black) lpattern(dash)) ///
1443 plot2opts(symbol(none) color(green)) plot3opts(symbol(none) color(red)) ///
1444 title("Assessment of PH Assumption") subtitle(" Predictor is E_lca_group3") xtitle("months")
1445
1446
1447
1448 * Assessment of PH Assumption: adjust for model 3
1449 stphplot, by(E_lca_group3) adjust(E_age E_sex E_education E_maritalstatus_4cat E_wealthquintiles
///
1450 E_smoking_3cat E_alcohol_status E_cvd_comorbidity) ///
1451 nolntime plot1opts(symbol(none) color(black) lpattern(dash)) ///
1452 plot2opts(symbol(none) color(green)) plot3opts(symbol(none) color(red)) ///
1453 title("Assessment of PH Assumption") subtitle(" Predictor is E_lca_group3") xtitle("months")
1454
1455
1456
1457 * Assessment of PH Assumption: adjust for model 4
1458 stphplot, by(E_lca_group3) adjust(E_age E_sex E_education E_maritalstatus_4cat E_wealthquintiles
///
1459 E_smoking_3cat E_alcohol_status E_cvd_comorbidity Ewv2_depressive_symptoms) ///
1460 nolntime plot1opts(symbol(none) color(black) lpattern(dash)) ///
1461 plot2opts(symbol(none) color(green)) plot3opts(symbol(none) color(red)) ///
1462 title("Assessment of PH Assumption") subtitle(" Predictor is E_lca_group3") xtitle("months")
1463
1464
1465
1466

```

```

1467
1468
1469
1470 /* Test of overall goodness of fit
1471 Goodness of fit of the final model
1472 2 methods:
1473 - by using the commnad stcoxgof (good fit = non sig p-value)
1474 - by using the Cox-Snell residuals
1475     to create the Nelson-Aalen cumulative hazard function
1476     If the hazard function follows the 45 degree line then we know that it approximately
1477     has an exponential distribution with a hazard rate of one and that the model fits the data
1478 well.
1479     If the model fits the data, the plot of the cumulative hazard versus cs
1480     should approximate a straight line with slope 1.
1481 */
1482
1483 * by using the commnad stcoxgof
1484
1485 * install stcoxgof
1486
1487 findit stcoxgof
1488
1489
1490 stcox E_lca_group3 E_age E_sex E_education E_maritalstatus_4cat E_wealthquintiles ///
1491 E_smoking_3cat E_alcohol_status E_cvd_comorbidity ///
1492 Ewv2_depressive_symptoms, mgale(mgale)
1493
1494
1495 stcoxgof
1496
1497
1498
1499
1500 * by using the Cox-Snell residuals
1501
1502 quietly stcox E_lca_group3 E_age E_sex E_education E_maritalstatus_4cat E_wealthquintiles ///
1503 E_smoking_3cat E_alcohol_status E_cvd_comorbidity ///
1504 Ewv2_depressive_symptoms
1505 predict cs, csnell
1506
1507 * or
1508
1509 quietly stcox E_lca_group3
1510 predict cs, csnell
1511
1512
1513 stset cs, failure(Ewv3to9_dementia_event)
1514 sts generate km = s
1515 generate H = -ln(km)
1516 line H cs cs, sort ytitle("") clstyle(. refline)
1517
1518
1519
1520
1521
1522
1523
1524
1525 * ----- COX PH REGRESSION MODEL IN IMPUTED DATASET ----- *
1526
1527
1528 * Declare Data to be Survival Data by using mi
1529
1530 mi stset E_time_event_dementia, failure (Ewv3to9_dementia_event==1) id(idauniq)
1531
1532
1533 * Run Cox regression analysis in imputed dataset by using "mi estimate:"

```



```

1534 * Building the Model: Model 1 (unadjusted), Model 2, Model 3, Model 4
1535
1536
1537
1538 * Unadjusted model - model 1 - single predictor
1539
1540 * Model 1 (default coefficients)
1541 mi estimate: stcox E_lca_group3
1542
1543 * Model 1: define design var by using i.
1544 mi estimate: stcox i.E_lca_group3
1545
1546
1547 * Model 1 ask for hazard ratio by using the option eform("Haz.Ratio")
1548
1549 mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3
1550
1551
1552 * Adjusted models - multivariable Cox model
1553 * controlling for covariates
1554
1555 * Model 2: model 1 + adjust for demographics: age sex education marital status and wealth
1556
1557 mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
1558 E_age i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles
1559
1560 * Model 3: model 2 + adjust for cvd health
1561
1562 mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
1563 E_age i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles ///
1564 i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity
1565
1566
1567 * Model 4: model 3 + adjust for depression
1568 mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
1569 E_age i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles ///
1570 i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity ///
1571 i.Ewv2_depressive_symptoms
1572
1573
1574
1575
1576 /*
1577
1578 *** SENSITIVITY ANALYSES ***
1579
1580 1) multigroup latent class model by sex
1581
1582 2) interactions with age and gender
1583 survival analysis stratified by age
1584 two age groups: <70 and >=70
1585
1586 3) exclude participants with cvd
1587
1588 4) exclude dementia cases identified by IQCODE
1589
1590 5) survival analysis limiting to 5 year follow-up
1591
1592 6) Complete data
1593 Cox regression analysis on complete data (without imputed covariates)
1594 (see above)
1595
1596 */
1597
1598
1599 /*
1600 1) Multigroup latent class model by sex
1601

```



```

1602
1603 TWO STEP PROCESS
1604
1605 1) LCA by group (to build the model and get lcprob and lcmean and to get the marginplots for males)
1606
1607 gsem (crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca ///
1608 hba1c_lca <- _cons), family(bernoulli) link(logit) lclass(C 3) ///
1609 group(E_sex) ginvariant(coef)
1610
1611 estat lcprob
1612 estat lcmean
1613 estat lcgof
1614
1615 2) LCA sort sex (to get the marginplots for females)
1616
1617 sort E_sex
1618
1619 by E_sex: gsem (crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca ///
1620 hba1c_lca <- _cons), family(bernoulli) link(logit) lclass(C 3)
1621
1622
1623 */
1624
1625
1626 * LCA by group
1627 * three-class model
1628
1629
1630 gsem (crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca ///
1631 hba1c_lca <- _cons), family(bernoulli) link(logit) lclass(C 3) ///
1632 group(E_sex) ginvariant(coef)
1633
1634
1635 estimates store threeclass_cm
1636
1637
1638
1639
1640 * LCA postestimation
1641 * Latent class marginal probabilities - lcprob -
1642 * Latent class marginal means - lcmean -
1643
1644
1645 estat lcprob
1646
1647 estat lcmean
1648
1649 estat lcgof
1650
1651
1652
1653 /* We can use the predictions of the posterior probability of class membership to evaluate an
1654 individual's probability of being in each class.
1655
1656 */
1657
1658 predict m_classpost1*, classposteriorpr
1659 list in 1, abbrev(10)
1660
1661 /* We can determine the expected class for each individual based on whether the posterior
1662 probability
1663 is greater than 0.5
1664 */
1665 generate m_expclass1 = 1 + (m_classpost11>0.5)
1666 tabulate m_expclass1
1667
1668

```

```

1669 generate m_expclass2 = 1 + (m_classpost12>0.5)
1670 tabulate m_expclass2
1671
1672
1673 generate m_expclass3 = 1 + (m_classpost13>0.5)
1674 tabulate m_expclass3
1675
1676
1677
1678
1679 /* We can determine expected classification for each individual in the dataset based on the
1680 predicted
1681 posterior class probabilities.
1682 */
1683 predict m_cpost*, classposteriorpr
1684 egen m_max = rowmax(m_cpost*)
1685
1686
1687 * generate classes var
1688
1689 generate m_predclass = 1 if m_cpost1==m_max
1690
1691 replace m_predclass = 2 if m_cpost2==m_max
1692
1693 replace m_predclass = 3 if m_cpost3==m_max
1694
1695 tabulate m_predclass
1696
1697
1698
1699 * margins and marginsplot for MALES
1700
1701 * use margins to calculate marginal predictions
1702 * use marginsplot to graph marginal predictions
1703
1704
1705
1706 *Install/update combomarginsplot ado.
1707
1708 *https://www.statalist.org/forums/forum/general-stata-discussion/general/1425209-is-it-possible-to-do-multilevel-latent-class-analysis-with-stata-15-ic
1709
1710 ssc install combomarginsplot, replace
1711
1712
1713 margins, predict(classpr class(1)) ///
1714             predict(classpr class(2)) ///
1715             predict(classpr class(3)) subpop(if E_sex==0) saving(margin_male, replace)
1716 marginsplot, xtitle("") ytitle("") ///
1717             xlabel(1 "Class 1" 2 "Class 2" 3 "Class 3") ///
1718             title("Predicted Latent Class Probabilities with 95% CI") ///
1719             name(margin_male, replace)
1720
1721
1722 margins, predict(classpr class(1)) ///
1723             predict(classpr class(2)) ///
1724             predict(classpr class(3)) subpop(if E_sex==0) saving(margin_male, replace)
1725 marginsplot, recast(bar) xtitle("") ytitle("") ///
1726             xlabel(1 "Class 1" 2 "Class 2" 3 "Class 3") ///
1727             title("Predicted Latent Class Probabilities with 95% CI") ///
1728             name(margin_male, replace)
1729
1730
1731 * class 1
1732
1733 margins, predict(outcome(crp_lca) class(1)) ///
1734             predict(outcome(hdl_lca) class(1)) ///

```

```

1735         predict(outcome(obesity_lca) class(1)) ///
1736         predict(outcome(systolic_lca) class(1)) ///
1737         predict(outcome(diastolic_lca) class(1)) ///
1738         predict(outcome(diabetes_lca) class(1)) ///
1739         predict(outcome(hba1c_lca) class(1)) subpop(if E_sex==0) ///
1740         saving(class1_male, replace) ///
1741
1742     marginsplot, recast(bar) title ("Class 1") xtitle("") ///
1743         xlabel(1 "crp" 2 "hdl" 3 "obesity" 4 "systolic BP" ///
1744             5 "diastolic BP" 6 "diabetes" 7 "hba1c", angle(45)) ///
1745         ytitle("Predicted mean") ylabel(0(.20)1) name (class1_male, replace)
1746
1747
1748 * class 2
1749
1750     margins, predict(outcome(crp_lca) class(2)) ///
1751         predict(outcome(hdl_lca) class(2)) ///
1752         predict(outcome(obesity_lca) class(2)) ///
1753         predict(outcome(systolic_lca) class(2)) ///
1754         predict(outcome(diastolic_lca) class(2)) ///
1755         predict(outcome(diabetes_lca) class(2)) ///
1756         predict(outcome(hba1c_lca) class(2)) subpop(if E_sex==0) ///
1757         saving(class2_male, replace) ///
1758
1759     marginsplot, recast(bar) title ("Class 2") xtitle("") ///
1760         xlabel(1 "crp" 2 "hdl" 3 "obesity" 4 "systolic BP" ///
1761             5 "diastolic BP" 6 "diabetes" 7 "hba1c", angle(45)) ///
1762         ytitle("Predicted mean") ylabel(0(.20)1) name (class2_male, replace)
1763
1764
1765
1766
1767 * class 3
1768
1769     margins, predict(outcome(crp_lca) class(3)) ///
1770         predict(outcome(hdl_lca) class(3)) ///
1771         predict(outcome(obesity_lca) class(3)) ///
1772         predict(outcome(systolic_lca) class(3)) ///
1773         predict(outcome(diastolic_lca) class(3)) ///
1774         predict(outcome(diabetes_lca) class(3)) ///
1775         predict(outcome(hba1c_lca) class(3)) subpop(if E_sex==0) ///
1776         saving(class3_male, replace) ///
1777
1778     marginsplot, recast(bar) title ("Class 3") xtitle("") ///
1779         xlabel(1 "crp" 2 "hdl" 3 "obesity" 4 "systolic BP" ///
1780             5 "diastolic BP" 6 "diabetes" 7 "hba1c", angle(45)) ///
1781         ytitle("Predicted mean") ylabel(0(.20)1) name (class3_male, replace)
1782
1783
1784
1785     graph combine class1_male class2_male class3_male, cols(3)
1786
1787
1788
1789
1790
1791
1792
1793 * LCA sort by sex
1794 * three-class model
1795
1796 sort E_sex
1797
1798 by E_sex: gsem (crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca ///
1799     hba1c_lca <- _cons), family(bernoulli) link(logit) lclass(C 3) ///
1800     startvalues(randompr, draws(20) seed(15) difficult) ///
1801     emopts(iterate(30) difficult)
1802

```

```

1803 estat lcprob
1804
1805 estat lcmean
1806
1807 estat lcgof
1808
1809
1810
1811
1812 /* We can use the predictions of the posterior probability of class membership to evaluate an
1813 individual's probability of being in each class.
1814
1815 */
1816
1817 predict f_classpost1*, classposteriorpr
1818 list in 1, abbrev(10)
1819
1820 /* We can determine the expected class for each individual based on whether the posterior
1821 probability
1822 is greater than 0.5
1823 */
1824 generate f_expclass1 = 1 + (f_classpost11>0.5)
1825 tabulate f_expclass1
1826
1827
1828 generate f_expclass2 = 1 + (f_classpost12>0.5)
1829 tabulate f_expclass2
1830
1831
1832 generate f_expclass3 = 1 + (f_classpost13>0.5)
1833 tabulate f_expclass3
1834
1835
1836
1837 /* We can determine expected classification for each individual in the dataset based on the
1838 predicted
1839 posterior class probabilities.
1840 */
1841
1842 predict f_cpost*, classposteriorpr
1843 egen f_max = rowmax(f_cpost*)
1844
1845 * generate classes var
1846
1847 generate f_predclass = 1 if f_cpost1==f_max
1848
1849 replace f_predclass = 2 if f_cpost2==f_max
1850
1851 replace f_predclass = 3 if f_cpost3==f_max
1852
1853 tabulate f_predclass
1854
1855
1856
1857
1858 * margins and marginsplot for FEMALES
1859
1860 * use margins to calculate marginal predictions
1861 * use marginsplot to graph marginal predictions
1862
1863
1864
1865 margins, predict(classpr class(1)) ///
1866         predict(classpr class(2)) ///
1867         predict(classpr class(3)) subpop(if E_sex==1) saving(margin_female, replace)
1868 marginsplot, xtitle ("") ytitle ("") ///

```

```

1869         xlabel (1 "Class 1" 2 "Class 2" 3 "Class 3") ///
1870         title ("Predicted Latent Class Probabilities with 95% CI") ///
1871         name(margin_female, replace)
1872
1873
1874 margins, predict(classpr class(1)) ///
1875         predict(classpr class(2)) ///
1876         predict(classpr class(3)) subpop(if E_sex==1) saving(margin_female, replace)
1877 marginsplot, recast(bar) xtitle("") ytitle("") ///
1878         xlabel(1 "Class 1" 2 "Class 2" 3 "Class 3") ///
1879         title("Predicted Latent Class Probabilities with 95% CI") ///
1880         name(margin_female, replace)
1881
1882
1883 * class 1
1884
1885 margins, predict(outcome(crp_lca) class(1)) ///
1886         predict(outcome(hdl_lca) class(1)) ///
1887         predict(outcome(obesity_lca) class(1)) ///
1888         predict(outcome(systolic_lca) class(1)) ///
1889         predict(outcome(diastolic_lca) class(1)) ///
1890         predict(outcome(diabetes_lca) class(1)) ///
1891         predict(outcome(hba1c_lca) class(1)) subpop(if E_sex==1) ///
1892         saving(class1_female, replace) ///
1893
1894 marginsplot, recast(bar) title ("Class 1") xtitle("") ///
1895         xlabel(1 "crp" 2 "hdl" 3 "obesity" 4 "systolic BP" ///
1896         5 "diastolic BP" 6 "diabetes" 7 "hba1c", angle(45)) ///
1897         ytitle("Predicted mean") ylabel(0(.20)1) name (class1_female, replace)
1898
1899
1900 * class 2
1901
1902 margins, predict(outcome(crp_lca) class(2)) ///
1903         predict(outcome(hdl_lca) class(2)) ///
1904         predict(outcome(obesity_lca) class(2)) ///
1905         predict(outcome(systolic_lca) class(2)) ///
1906         predict(outcome(diastolic_lca) class(2)) ///
1907         predict(outcome(diabetes_lca) class(2)) ///
1908         predict(outcome(hba1c_lca) class(2)) subpop(if E_sex==1) ///
1909         saving(class2_female, replace) ///
1910
1911 marginsplot, recast(bar) title ("Class 2") xtitle("") ///
1912         xlabel(1 "crp" 2 "hdl" 3 "obesity" 4 "systolic BP" ///
1913         5 "diastolic BP" 6 "diabetes" 7 "hba1c", angle(45)) ///
1914         ytitle("Predicted mean") ylabel(0(.20)1) name (class2_female, replace)
1915
1916
1917
1918
1919 * class 3
1920
1921 margins, predict(outcome(crp_lca) class(3)) ///
1922         predict(outcome(hdl_lca) class(3)) ///
1923         predict(outcome(obesity_lca) class(3)) ///
1924         predict(outcome(systolic_lca) class(3)) ///
1925         predict(outcome(diastolic_lca) class(3)) ///
1926         predict(outcome(diabetes_lca) class(3)) ///
1927         predict(outcome(hba1c_lca) class(3)) subpop(if E_sex==1) ///
1928         saving(class3_female, replace) ///
1929
1930 marginsplot, recast(bar) title ("Class 3") xtitle("") ///
1931         xlabel(1 "crp" 2 "hdl" 3 "obesity" 4 "systolic BP" ///
1932         5 "diastolic BP" 6 "diabetes" 7 "hba1c", angle(45)) ///
1933         ytitle("Predicted mean") ylabel(0(.20)1) name (class3_female, replace)
1934
1935
1936

```

```

1937 graph combine class1_female class2_female class3_female, cols(3)
1938
1939
1940
1941
1942
1943
1944 *combine margins male and female class probabilities
1945
1946 graph combine margin_male margin_female, cols(3)
1947
1948 *combine margins male and female 3 classes mean
1949
1950 graph combine class1_male class2_male class3_male class1_female class2_female class3_female, cols(
3)
1951
1952
1953
1954
1955
1956
1957
1958 /* 2) Interaction with age and gender
1959 Survival analysis stratified by age
1960
1961 generate age_group variable
1962 Age groups: 1) young old (< 70) 2) old old (>= 70)
1963
1964 Kaplan Meier curves
1965 Cox regression models in imputed data
1966
1967 young old <70
1968 if E_age_group==1
1969
1970 old old >70
1971 if E_age_group==2
1972
1973
1974 */
1975
1976
1977
1978
1979 gen E_age_group=1 if E_age < 70
1980 replace E_age_group=2 if E_age >=70 & !missing(E_age)
1981
1982 label var E_age_group "Age groups <70 young-old / 70 old-old"
1983 lab def age_group 1 "young old <70" 2 "old old >70"
1984 lab val E_age_group age_group
1985
1986 tab E_age_group
1987
1988
1989 stset E_time_event_dementia, failure (Ewv3to9_dementia_event==1) id(idauniq)
1990
1991
1992 * <70 Kaplan Meier
1993
1994 * Group Kaplan-Meier Curve Estimation
1995 * Command is sts graph, by(GROUPVAR) OPTION OPTION OPTION Note: Must have sorted by GROUPVAR first
1996
1997 sort E_lca_group3
1998
1999 sts list if E_age_group==1, by(E_lca_group3)
2000
2001 * graph with frills
2002
2003 sts graph if E_age_group==1, by(E_lca_group3) xlabel(0(20)180) ylabel(0.80(.05)1) xtitle("Time in

```

```

Months") ///
2004 ytitle("Survival Prob") title("Kaplan Meier Curve <70 years") ///
2005 caption("graph04.png", size(vsmall))
2006
2007
2008 * calculate person-time and incidence rates using command ststtime
2009
2010 stptime if E_age_group==1,title(Person-years)
2011
2012 stptime if E_age_group==1, title(Person-years) per(1000)
2013
2014
2015 * calculate person-time by category of E_lca_group3
2016
2017 stptime if E_age_group==1, by(E_lca_group3)
2018
2019 stptime if E_age_group==1, by(E_lca_group3) per(1000)
2020
2021
2022
2023
2024 * >70 Kaplan Meier
2025
2026 sts list if E_age_group==2, by(E_lca_group3)
2027
2028 * graph with frills
2029
2030 sts graph if E_age_group==2, by(E_lca_group3) xlabel(0(20)180) ylabel(0.80(.05)1) xtitle("Time in
Months") ///
2031 ytitle("Survival Prob") title("Kaplan Meier Curve >=70 years") ///
2032 caption("graph04.png", size(vsmall))
2033
2034
2035 * calculate person-time and incidence rates using command ststtime
2036
2037 stptime if E_age_group==2,title(Person-years)
2038
2039 stptime if E_age_group==2, title(Person-years) per(1000)
2040
2041
2042 * calculate person-time by category of E_lca_group3
2043
2044 stptime if E_age_group==2, by(E_lca_group3)
2045
2046 stptime if E_age_group==2, by(E_lca_group3) per(1000)
2047
2048
2049
2050
2051 * COX PH REGRESSION MODEL IN IMPUTED DATASET
2052
2053
2054 * Declare Data to be Survival Data by using mi
2055
2056 mi stset E_time_event_dementia, failure (Ewv3to9_dementia_event==1) id(idauniq)
2057
2058
2059
2060
2061
2062 *** INTERACTION sex*cardiometabolic cluster ***
2063
2064 mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 c.E_sex#i.E_lca_group3
2065
2066
2067 mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
2068 E_age i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles ///
2069 i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity ///

```



```

2070 i.Ewv2_depressive_symptoms c.E_sex#i.E_lca_group3
2071
2072
2073
2074
2075 *** INTERACTION age*cardiometabolic cluster ***
2076
2077 mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 c.E_age#i.E_lca_group3
2078
2079
2080 mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
2081 E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles ///
2082 i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity ///
2083 i.Ewv2_depressive_symptoms c.E_age#i.E_lca_group3
2084
2085
2086
2087
2088 * YOUNG OLD <70 Cox regression models
2089
2090 * Model 1 ask for hazard ratio by using the option eform("Haz.Ratio")
2091
2092 mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 if E_age_group==1
2093
2094
2095 * Adjusted models - multivariable Cox model
2096 * controlling for covariates
2097
2098 * Model 2: model 1 + adjust for demographics: sex education marital status and wealth
2099
2100 mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
2101 i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles if E_age_group==1
2102
2103 * Model 3: model 2 + adjust for cvd health
2104
2105 mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
2106 i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles ///
2107 i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity if E_age_group==1
2108
2109
2110 * Model 4: model 3 + adjust for depression
2111
2112 mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
2113 i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles ///
2114 i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity ///
2115 i.Ewv2_depressive_symptoms if E_age_group==1
2116
2117
2118
2119
2120 * OLD OLD >70 Cox regression models
2121
2122
2123 * Model 1 ask for hazard ratio by using the option eform("Haz.Ratio")
2124
2125 mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 if E_age_group==2
2126
2127
2128 * Adjusted models - multivariable Cox model
2129 * controlling for covariates
2130
2131 * Model 2: model 1 + adjust for demographics: sex education marital status and wealth
2132
2133 mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
2134 i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles if E_age_group==2
2135
2136 * Model 3: model 2 + adjust for cvd health
2137

```



```

2138 mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
2139 i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles ///
2140 i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity if E_age_group==2
2141
2142
2143 * Model 4: model 3 + adjust for depression
2144
2145 mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
2146 i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles ///
2147 i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity ///
2148 i.Ewv2_depressive_symptoms if E_age_group==2
2149
2150
2151
2152
2153
2154
2155
2156 /*
2157
2158 3) exclude participants with cvd
2159
2160 use the command if E_cvd_comorbidity==0
2161
2162 */
2163
2164
2165 * COX PH REGRESSION MODEL IN IMPUTED DATASET
2166
2167
2168 * Declare Data to be Survival Data by using mi
2169
2170 mi stset E_time_event_dementia, failure (Ewv3to9_dementia_event==1) id(idauniq)
2171
2172
2173 * Model 1 ask for hazard ratio by using the option eform("Haz.Ratio")
2174
2175 mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 if E_cvd_comorbidity==0
2176
2177
2178 * Adjusted models - multivariable Cox model
2179 * controlling for covariates
2180
2181 * Model 2: model 1 + adjust for demographics: sex education marital status and wealth
2182
2183 mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
2184 E_age i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles if E_cvd_comorbidity==0
2185
2186 * Model 3: model 2 + adjust for cvd health
2187
2188 mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
2189 E_age i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles ///
2190 i.E_smoking_3cat i.E_alcohol_status if E_cvd_comorbidity==0
2191
2192
2193 * Model 4: model 3 + adjust for depression
2194
2195 mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
2196 E_age i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles ///
2197 i.E_smoking_3cat i.E_alcohol_status ///
2198 i.Ewv2_depressive_symptoms if E_cvd_comorbidity==0
2199
2200
2201
2202
2203
2204 /*
2205 4) survival analysis excluding cases diagnosed with IQCODE

```

```

2206
2207 Variables
2208 Ewv3to9_dementia_event_no_iqcode E_time_event_dementia_report_no_
2209
2210 */
2211
2212
2213
2214
2215 * COX PH REGRESSION MODEL IN IMPUTED DATASET
2216
2217
2218
2219 * Declare Data to be Survival Data by using mi
2220
2221 mi stset E_time_event_dementia_report_no_, failure (Ewv3to9_dementia_event_no_iqcode==1) id(
idauniq)
2222
2223
2224 * Model 1 ask for hazard ratio by using the option eform("Haz.Ratio")
2225
2226 mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3
2227
2228
2229 * Adjusted models - multivariable Cox model
2230 * controlling for covariates
2231
2232 * Model 2: model 1 + adjust for demographics: age sex education marital status and wealth
2233
2234 mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
E_age i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles
2235
2236
2237 * Model 3: model 2 + adjust for cvd health
2238
2239 mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
E_age i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles ///
2240 i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity
2241
2242
2243
2244 * Model 4: model 3 + adjust for depression
2245
2246 mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
E_age i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles ///
2247 i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity ///
2248 i.Ewv2_depressive_symptoms
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258 /*
2259 5) survival analysis limiting to 5 year follow-up period
2260
2261 elsa follow-up wave 3-6
2262
2263 */
2264
2265
2266 merge 1:m idauniq using "S:\Research\pkstudies\Study3_cardio_lca\ELSA\elsa_lca data sensitivity
3to6followup.dta"
2267
2268
2269 * COX PH REGRESSION MODEL IN IMPUTED DATASET
2270
2271

```

```

2272
2273 * Declare Data to be Survival Data by using mi
2274
2275 mi stset Ewv3to6_time_event_dementia, failure (Ewv3to6_dementia_event==1) id(idauniq)
2276
2277
2278
2279 * Model 1 ask for hazard ratio by using the option eform("Haz.Ratio")
2280
2281 mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3
2282
2283
2284 * Adjusted models - multivariable Cox model
2285 * controlling for covariates
2286
2287 * Model 2: model 1 + adjust for demographics: age sex education marital status and wealth
2288
2289 mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
2290 E_age i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles
2291
2292 * Model 3: model 2 + adjust for cvd health
2293
2294 mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
2295 E_age i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles ///
2296 i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity
2297
2298
2299 * Model 4: model 3 + adjust for depression
2300
2301 mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
2302 E_age i.E_sex i.E_education i.E_maritalstatus_4cat i.E_wealthquintiles ///
2303 i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity ///
2304 i.Ewv2_depressive_symptoms
2305
2306
2307 * 6) complete data (see above)
2308
2309
2310 * ----- *
2311
2312
2313
2314
2315 *** EXTRA SENSITIVITY ANALYSES FOR PAPER ***
2316
2317 /*
2318
2319 compare baseline characteristics between complete sample (before exclusion) and sample with
2320 missing data (overall after exclusion)
2321
2322 */
2323
2324 * General characteristics of ELSA participants at baseline
2325
2326 * Socio-demographics
2327 sum E_age
2328 ta E_sex
2329 ta E_education
2330 ta E_maritalstatus_4cat
2331 ta E_wealthquintiles
2332 * Cardiometabolic disorders
2333 ta Ewv2_crp
2334 ta Ewv2_hdl_cholesterol
2335 ta Ewv2_obesity_waist
2336 ta Ewv2_systolic_bp
2337 ta Ewv2_diastolic_bp
2338 ta Ewv2_diabetes_diagnosed

```

```

2339 ta Ewv2_HbA1c
2340 * Cardiovascular health factors
2341 ta E_smoking_3cat
2342 ta E_physicalactivity
2343 ta E_alcohol_status
2344 ta E_cvd_comorbidity
2345 * Depressive symptoms
2346 ta Ewv2_depressive_symptoms
2347 * Memory score
2348 sum E_memory_wordrecall
2349
2350
2351
2352 * compare health characteristics between those survived and dropped out
2353
2354
2355
2356 *** CLEANING DATA to keep those who dropped out
2357
2358
2359 * 1. drop dementia cases and missing data at baseline
2360
2361 drop if Ewv2_anydementia_iqcode_report==1
2362 * (50 observations deleted)
2363
2364 drop if Ewv2_anydementia_iqcode_report== .
2365 * (0 observations deleted)
2366
2367
2368 * 2. drop missing values of cardiometabolic markers
2369
2370 drop if Ewv2_crp== .
2371 * (1,753 observations deleted)
2372
2373 drop if Ewv2_hdl_cholesterol== .
2374 * (6 observations deleted)
2375
2376 drop if Ewv2_obesity_waist== .
2377 * (133 observations deleted)
2378
2379 drop if Ewv2_systolic_bp== .
2380 * (660 observations deleted)
2381
2382 drop if Ewv2_diastolic_bp== .
2383 * (0 observations deleted)
2384
2385 drop if Ewv2_diabetes_diagnosed== .
2386 * (0 observations deleted)
2387
2388 drop if Ewv2_HbA1c== .
2389 * (102 observations deleted)
2390
2391
2392
2393 * 3. drop obs with no records on dementia at any wave from 3-9 follow-ups
2394
2395
2396 search mdesc
2397 search rmiss2
2398 search mvpatterns
2399
2400 * see number of missing values vs non-missing in each variable
2401 mdesc Ewv3_anydementia_iqcode_report Ewv4_anydementia_iqcode_report ///
2402 Ewv5_anydementia_iqcode_report Ewv6_anydementia_iqcode_report Ewv7_anydementia_iqcode_report ///
2403 Ewv8_anydementia_iqcode_report Ewv9_anydementia_iqcode_report
2404
2405
2406

```

```

2407 /* number of missing values per observation
2408 * the code below creates a variable called nmisfollowup that gives the number of missing values
2409 for each observation in the variables of interest */
2410 egen nmisfollowup_dementia_wv3to9=rmiss2(Ewv3_anydementia_iqcode_report ///
2411 Ewv4_anydementia_iqcode_report Ewv5_anydementia_iqcode_report ///
2412 Ewv6_anydementia_iqcode_report Ewv7_anydementia_iqcode_report ///
2413 Ewv8_anydementia_iqcode_report Ewv9_anydementia_iqcode_report)
2414
2415 tab nmisfollowup_dementia_wv3to9
2416
2417 * drop observations "nmisfollowup_dementia_wv3to9" < 7
2418 drop if nmisfollowup_dementia_wv3to9<7
2419
2420
2421 * Socio-demographics
2422 sum E_age
2423 ta E_sex
2424 ta E_education
2425 ta E_maritalstatus_4cat
2426 ta E_wealthquintiles
2427 * Cardiometabolic disorders
2428 ta Ewv2_crp
2429 ta Ewv2_hdl_cholesterol
2430 ta Ewv2_obesity_waist
2431 ta Ewv2_systolic_bp
2432 ta Ewv2_diastolic_bp
2433 ta Ewv2_diabetes_diagnosed
2434 ta Ewv2_HbA1c
2435 * Cardiovascular health factors
2436 ta E_smoking_3cat
2437 ta E_physicalactivity
2438 ta E_alcohol_status
2439 ta E_cvd_comorbidity
2440 * Depressive symptoms
2441 ta Ewv2_depressive_symptoms
2442 * Memory score
2443 sum E_memory_wordrecall
2444
2445
2446
2447 * compare health characteristics between <70 and >=70
2448
2449
2450
2451 * General baseline characteristics of ELSA participants by age group
2452
2453 * Socio-demographics
2454 ttest E_age, by(E_age_group)
2455 ta E_sex E_age_group, chi2 column row
2456 ta E_education E_age_group, chi2 column row
2457 ta E_maritalstatus_4cat Ewv3to9_dementia_event, chi2 column row
2458 ta E_wealthquintiles E_age_group, chi2 column row
2459 * Cardiometabolic disorders
2460 ta crp_lca E_age_group, chi2 column row
2461 ta hdl_lca E_age_group, chi2 column row
2462 ta obesity_lca E_age_group, chi2 column row
2463 ta systolic_lca E_age_group, chi2 column row
2464 ta diastolic_lca E_age_group, chi2 column row
2465 ta diabetes_lca E_age_group, chi2 column row
2466 ta hba1c_lca E_age_group, chi2 column row
2467 * Cardiovascular health factors
2468 ta E_smoking_3cat E_age_group, chi2 column row
2469 ta E_physicalactivity E_age_group, chi2 column row
2470 ta E_alcohol_status E_age_group, chi2 column row
2471 ta E_cvd_comorbidity E_age_group, chi2 column row
2472 * Depressive symptoms
2473 ta Ewv2_depressive_symptoms E_age_group, chi2 column row
2474 * Memory score

```

```
2475 ttest E_memory_wordrecall, by(E_age_group)
2476 ta E_lca_group3 E_age_group, chi2 column row
2477
2478
2479
2480
2481
2482 * ----- *
2483
2484
2485
2486
2487
```