

```

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: WV1 - WV4 (4 TIME POINTS)
15 DEMENTIA INCIDENCE AT YEAR 6: WV4
16
17 */
18
19
20
21
22 * importing data (.dta)
23
24 use
25 "S:\Research\pkhec\Study1_traj_depression\CHARLS\charls_depress_traj_final_4class_model_with_all_va
26 r complete data"
27
28
29 * KEEP NECESSARY VARIABLES
30
31 keep ID id_12char bloodweight ///
32 C_sex C_age C_education_10level C_education_harmon_3cat ///
33 C_education C_maritalstatus_8cat C_maritalstatus_3cat C_maritalstatus_4cat ///
34 Cwv1_smoking_2cat Cwv1_smoking_3cat Cwv1_physicalactivity Cwv1_alcohol_freq Cwv1_alcohol_status ///
35 C_cvd_comorbidity Cwv1_antidepressant Cwv1_psycholog_treat Cwv1_anytreat_psyche ///
36 Cwv1_memory_wordrecall Cwv1_concentration_serial7 Cwv1_orientation_time ///
37 Cwv1_executive_drawpicture Cwv1_cognition Cwv1_wealthquintiles Cwv1_netwealth_quintiles ///
38 Cwv1_cesd_depressed Cwv1_cesd_effort Cwv1_cesd_sleep Cwv1_cesd_lonely ///
39 Cwv1_cesd_bother Cwv1_cesd_going Cwv1_cesd_mind Cwv1_cesd_fear Cwv1_cesd_happy ///
40 Cwv1_cesd_hope Cwv1_cesd_sumscore Cwv1_cesd_score Cwv1_depressive_symptoms ///
41 Cwv2_cesd_depressed Cwv2_cesd_effort Cwv2_cesd_sleep Cwv2_cesd_lonely ///
42 Cwv2_cesd_bother Cwv2_cesd_going Cwv2_cesd_mind Cwv2_cesd_fear Cwv2_cesd_happy ///
43 Cwv2_cesd_hope Cwv2_cesd_sumscore Cwv2_cesd_score Cwv2_depressive_symptoms ///
44 Cwv3_cesd_depressed Cwv3_cesd_effort Cwv3_cesd_sleep Cwv3_cesd_lonely ///
45 Cwv3_cesd_bother Cwv3_cesd_going Cwv3_cesd_mind Cwv3_cesd_fear Cwv3_cesd_happy ///
46 Cwv3_cesd_hope Cwv3_cesd_sumscore Cwv3_cesd_score Cwv3_depressive_symptoms ///
47 Cwv4_cesd_bother Cwv4_cesd_mind Cwv4_cesd_depressed Cwv4_cesd_effort ///
48 Cwv4_cesd_fear Cwv4_cesd_sleep Cwv4_cesd_lonely Cwv4_cesd_going Cwv4_cesd_happy ///
49 Cwv4_cesd_hope Cwv4_cesd_sumscore Cwv4_depressive_symptoms ///
50 Cwv2to4_dementia_sum Cwv2to4_dementia_event ///
51 Cwv1_crp_level Cwv1_crp Cwv1_hdl_level Cwv1_male_hdl Cwv1_female_hdl ///
52 Cwv1_meds_dyslipid Cwv1_anymeds_dyslipid Cwv1_dyslipid_evr Cwv1_dyslipid_diagnosed ///
53 Cwv1_dyslipid_report_sum Cwv1_dyslipid_report Cwv1_hdl_sum Cwv1_hdl_cholesterol ///
54 Cwv1_waist Cwv1_malewaist_ao Cwv1_femalewaist_ao Cwv1_obesity_waist_sum ///
55 Cwv1_obesity_waist Cwv1_bmi_score Cwv1_obesity_bmi Cwv1_waist_bmi_sum ///
56 Cwv1_obesity Cwv1_tg_level Cwv1_tg Cwv1_triglyc_sum Cwv1_triglyc Cwv1_1systolic_bp ///
57 Cwv1_1diastolic_bp Cwv1_2systolic_bp Cwv1_2diastolic_bp Cwv1_3systolic_bp Cwv1_3diastolic_bp ///
58 Cwv1_systolic_mean Cwv1_diastolic_mean Cwv1_systolic_bp Cwv1_diastolic_bp Cwv1_meds_bp ///
59 Cwv1_anymeds_bp Cwv1_bp_evr Cwv1_bp_diagnosed Cwv1_bp_report_sum Cwv1_bp_report Cwv1_bp_sum ///
60 Cwv1_bp Cwv1_glucose_level Cwv1_glucose Cwv1_HbA1c_level Cwv1_HbA1c Cwv1_diabetes_evr ///
61 Cwv1_diabetes_diagnosed Cwv1_diabetes_report_sum Cwv1_diabetes_report Cwv1_meds_diabetes ///
62 Cwv1_anymeds_diabetes Cwv1_glucose_diabetes_sum Cwv1_glycemia ///
63 Cwv3_crp_level Cwv3_crp Cwv3_hdl_level Cwv3_male_hdl Cwv3_female_hdl ///
64 Cwv3_meds_dyslipid Cwv3_anymeds_dyslipid Cwv3_dyslipid_evr Cwv3_dyslipid_diagnosed ///
65 Cwv3_dyslipid_report_sum Cwv3_dyslipid_report Cwv3_hdl_sum Cwv3_hdl_cholesterol ///

```

```

66 Cwv3_waist Cwv3_malewaist_ao Cwv3_femalewaist_ao Cwv3_obesity_waist_sum ///
67 Cwv3_obesity_waist Cwv3_bmi_score Cwv3_obesity_bmi Cwv3_waist_bmi_sum Cwv3_obesity ///
68 Cwv3_tg_level Cwv3_tg Cwv3_triglyc_sum Cwv3_triglyc Cwv3_1systolic_bp Cwv3_1diastolic_bp ///
69 Cwv3_2systolic_bp Cwv3_2diastolic_bp Cwv3_3systolic_bp Cwv3_3diastolic_bp Cwv3_systolic_mean ///
70 Cwv3_diastolic_mean Cwv3_systolic_bp Cwv3_diastolic_bp Cwv3_meds_bp Cwv3_anymeds_bp ///
71 Cwv3_bp_evr Cwv3_bp_diagnosed Cwv3_bp_report_sum Cwv3_bp_report Cwv3_bp_sum Cwv3_bp
Cwv3_glucose_level ///
72 Cwv3_glucose Cwv3_HbA1c_level Cwv3_HbA1c Cwv3_diabetes_evr Cwv3_diabetes_diagnosed ///
73 Cwv3_diabetes_report_sum Cwv3_diabetes_report Cwv3_meds_diabetes Cwv3_anymeds_diabetes ///
74 Cwv3_glucose_diabetes_sum Cwv3_glycemia ///
75 Cwv4_smoking_2cat Cwv4_smoking_3cat Cwv4_physicalactivity ///
76 Cwv4_alcohol_freq Cwv4_alcohol_status Cwv4_cvd_comorbidity ///
77 Cwv4_memory_wordrecall Cwv4_concentration_serial7 Cwv4_orientation_time ///
78 Cwv4_executive_drawpicture Cwv4_cognition ///
79 Cwv1_dementia_report Cwv2_dementia_report Cwv3_dementia_report Cwv4_self_info_dementia ///
80 Cwv1_interview_date Cwv2_interview_date Cwv3_interview_date Cwv4_interview_date ///
81 Cwv2to4_newdementia_or_lastinter Cwv2to4_dementia_free_date C_time_dementia_months ///
82 C_time_dementia_midpoint C_time_dementia_midpoint_final C_time_of_event_dementia
83
84
85
86
87 /* Latent class growth analysis (LCGA) of depressive symptoms */
88
89
90 * installing traj command
91
92 net from http://www.andrew.cmu.edu/user/bjones/traj
93 net install traj, force
94 help traj
95
96
97
98 * Generate a set of time variables to pass to traj, from wave 1 to 4(t0-t3)
99 forval i = 0/3 {
100     generate t_`i' = `i'
101 }
102
103 *recode time in months
104
105 recode t_1 (1=24)
106 recode t_2 (2=48)
107 recode t_3 (3=84)
108
109
110 *rename cesd score across the waves - discrete var min=0 max=8
111
112 rename Cwv1_cesd_score cesd_0
113 rename Cwv2_cesd_score cesd_1
114 rename Cwv3_cesd_score cesd_2
115 rename Cwv4_cesd_sumscore cesd_3
116
117
118
119
120
121
122 *** Descriptive stats of depression and dementia
123
124 tabulate cesd_0
125 summarize cesd_0 , detail
126 histogram cesd_0, discrete frequency normal
127
128 misstable summarize cesd_0
129 misstable patterns cesd_0
130
131 tabulate Cwv1_depressive_symptoms
132 summarize Cwv1_depressive_symptoms

```

```

133
134 misstable summarize Cwv1_depressive_symptoms
135 misstable patterns Cwv1_depressive_symptoms
136
137 tabulate Cwv1_dementia_report
138 summarize Cwv1_dementia_report
139
140 misstable summarize Cwv1_dementia_report
141 misstable patterns Cwv1_dementia_report
142
143
144 tabulate Cwv2_dementia_report
145 summarize Cwv2_dementia_report
146
147 misstable summarize Cwv2_dementia_report
148 misstable patterns Cwv2_dementia_report
149
150
151 tabulate Cwv3_dementia_report
152 summarize Cwv3_dementia_report
153
154 misstable summarize Cwv3_dementia_report
155 misstable patterns Cwv3_dementia_report
156
157 tabulate Cwv4_self_info_dementia
158 summarize Cwv4_self_info_dementia
159
160 misstable summarize Cwv4_self_info_dementia
161 misstable patterns Cwv4_self_info_dementia
162
163
164
165
166
167
168
169
170 *** CLEANING DATA
171
172 * 1. drop missing data depression and dementia at baseline
173 * drop 663 depression missing data
174
175 drop if cesd_0== .
176 * (625 observations deleted)
177 drop if Cwv1_dementia_report== .
178 * (38 observations deleted)
179
180 * 2. drop dementia cases between wv1 and wv3 (total: 407 cases)
181
182 drop if Cwv1_dementia_report==1
183 * (234 observations deleted)
184
185 drop if Cwv2_dementia_report==1
186 * (73 observations deleted)
187
188 drop if Cwv3_dementia_report==1
189 * (100 observations deleted)
190
191
192
193
194 * 3. process to drop missing data depression in at least 2 follow-up waves
195
196 /*
197
198 check below how to see number of missing values in an observation (case) and patterns of missing
values
199 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/
200  install packages:
201  * install mdesc
202  * install tabmiss
203  * insatll dm31
204  * insall mvpatterna
205
206  */
207
208  search mdesc
209  search rmiss2
210  search mvpatterns
211
212
213  * see number of missing values vs non-missing in each variable
214  mdesc cesd_0 cesd_1 cesd_2 cesd_3
215  *or
216  mdesc cesd_*
217
218
219  * number of missing values per observation
220  * the code below creates a variable called nmisfollowup that gives the number of missing values
    for each observation in the variables of interest
221  egen nmisfollowup_cesd=rmiss2(cesd_1 cesd_2 cesd_3)
222
223  tab nmisfollowup_cesd
224
225  * drop observations "nmisfollowup_cesd" > 1 (those with 2 or 3 missing data)
226  drop if nmisfollowup>1
227  *(1549 observations deleted)
228
229
230
231  * 4. drop obs with no records on dementia at wave 4
232
233  drop if Cwv4_self_info_dementia== .
234  *(3823 observations deleted)
235
236
237
238
239  *descriptive stats of depressive symptoms cesd
240
241  tabulate cesd_0
242  summarize cesd_0, detail
243  histogram cesd_0, discrete frequency normal
244
245  tabulate cesd_1
246  summarize cesd_1 , detail
247  histogram cesd_1, discrete frequency normal
248
249  tabulate cesd_2
250  summarize cesd_2, detail
251  histogram cesd_2, discrete frequency normal
252
253  tabulate cesd_3
254  summarize cesd_3, detail
255  histogram cesd_3, discrete frequency normal
256
257
258  ta cesd_0, miss
259  ta cesd_1, miss
260  ta cesd_2, miss
261  ta cesd_3, miss
262
263  tabstat cesd_0, by(C_sex)stats (mean v n)
264  tabstat cesd_1, by(C_sex)stats (mean v n)
265  tabstat cesd_2, by(C_sex)stats (mean v n)

```

```

266 tabstat cesd_3, by(C_sex)stats (mean v n)
267
268 tabstat cesd_*,s(n me sk) by(C_sex)
269
270
271 bysort C_sex: tab cesd_0
272 bysort C_sex: tab cesd_1
273 bysort C_sex: tab cesd_2
274 bysort C_sex: tab cesd_3
275
276
277 tabstat cesd_0 cesd_1 cesd_2 cesd_3, s(sk kur)
278 sktest cesd_0 cesd_1 cesd_2 cesd_3
279
280
281 * missingness pateterns
282 misstable patterns cesd_*
283 * "1" means that the variable is observed and a "0" represents missing
284
285
286 * box plots of the observations at each occasion
287 graph box cesd_0 cesd_1 cesd_2 cesd_3, ascategory intensity (0) medtype (line)
288
289
290
291
292
293 /*
294 LCGA
295 longitudinal nalysis of trajectories
296 GBTM model
297
298 traj [if], var(varlist) indep(varlist) model(modeltype)
299         order(numlist) [additional options]
300
301
302 order(numlist)      0=intercept, 1=linear, 2=quadratic, 3=cubic -
303                     polynomial type for each group trajectory
304
305
306 ci                  parametric bootstrap confidence intervals of
307                     individual distal outcome and probability of
308                     group memberships.
309
310
311
312
313
314
315 Available Models -> command traj
316
317
318 Censored normal (CNORM) model distribution
319
320 traj, var(qcp*op) indep(age*) model(cnorm) min(0) max(999) order(1 3 2)
321
322 trajplot, xtitle(Age) ytitle(Opposition) xlabel(6(1)15)
323         ylabel(0(1)6)
324
325 /* Shows the assigned group and probabilities of group membership */
326 list _traj_Group - _traj_ProbG3 if _n < 3, ab(12)
327
328 /* trajT = x-axis, Avg# = data averages, Est# = model estimates */
329 matrix list e(plot1), format(%9.2f) noheader
330
331 /* Including time-stable covariates (risk) associated with group membership */
332
333 traj, var(qcp*op) indep(age*) model(cnorm) min(0) max(10) order(1 3

```

```
2) risk(scolmer scolper)
```

Zero Inflated poisson (ZIP) Model

It is an analysis of Poisson data with extra zeros

```
traj, model(zip) var(y*) indep(t*) order(2 1 3) iorder(1)
```

```
trajplot, xtitle(Age) ytitle(Opposition) ci
```

Time-Stable Covariates for Group Membership

```
traj, var(qcp*op) indep(age*) model(cnorm) min(0) max(10) order(1 3
2) risk(scolmer scolper)
```

```
trajplot, xtitle(Age) ytitle(Opposition)
```

Logistic (logit) model

```
use http://www.andrew.cmu.edu/user/bjones/traj/data/cambridge.dta,
clear
```

```
traj, var(p1-p23) indep(tt1-tt23) model(logit) order(0 3 3)
```

```
trajplot, xtitle(Scaled Age) ytitle(Prevalence)
```

```
/* Assigned group and probabilities of group membership */
list _traj_Group - _traj_ProbG3 if _n > 400, ab(12)
```

Model selection:

1. Type of model: The 'traj' can model normal, censored normal, zero-inflated Poisson and binary logit models.
Capacity for incorporating effect of time-stable and time-varying covariates, subsequent outcomes and joint trajectory models.
2. Number of groups/classes: determination of the optimal number of groups to compose the mixture
3. Shape of the trajectory: determination of the appropriate order of the polynomial used to model each group's trajectory (linear, quadratic, cubic).

Model Fit Criteria to select the model with optimal class enumeration:

- Bayesian Information Criteria (BIC), where lower BIC or least negative BIC (higher value closer to zero) represents a better fitting model.
- Bayes Factor greater than 10 indicates very strong evidence to use the "more complex" model.
- Meaningful proportion of participants within each class (smallest group percentage to be higher or equal to 5%).
- Average posterior probability (APP) to belong to each class higher than 0.70.
- Entropy to determine the accuracy of classification of individuals into the different latent classes
If entropy is near 1.0, then classification of individuals is assumed to be adequate.

```

400 If entropy is near 0, then classification is assumed to be poor.
401
402
403
404
405
406
407 *****function to print out summary stats
408 program summary_table_procTraj
409     preserve
410     *look at the average posterior probability
411     gen Mp = 0
412     foreach i of varlist _traj_ProbG* {
413         replace Mp = `i' if `i' > Mp
414     }
415     sort _traj_Group
416     *and the odds of correct classification
417     by _traj_Group: gen countG = _N
418     by _traj_Group: egen groupAPP = mean(Mp)
419     by _traj_Group: gen counter = _n
420     gen n = groupAPP/(1 - groupAPP)
421     gen p = countG/ _N
422     gen d = p/(1-p)
423     gen occ = n/d
424     *Estimated proportion for each group
425     scalar c = 0
426     gen TotProb = 0
427     foreach i of varlist _traj_ProbG* {
428         scalar c = c + 1
429         quietly summarize `i'
430         replace TotProb = r(sum)/ _N if _traj_Group == c
431     }
432     gen d_pp = TotProb/(1 - TotProb)
433     gen occ_pp = n/d_pp
434     *This displays the group number [_traj_~p],
435     *the count per group (based on the max post prob), [countG]
436     *the average posterior probability for each group, [groupAPP]
437     *the odds of correct classification (based on the max post prob group assignment), [occ]
438     *the odds of correct classification (based on the weighted post. prob), [occ_pp]
439     *and the observed probability of groups versus the probability [p]
440     *based on the posterior probabilities [TotProb]
441     list _traj_Group countG groupAPP occ occ_pp p TotProb if counter == 1
442     restore
443 end
444
445 summary_table_procTraj
446
447
448
449 ***** to generate a plot of the individual trajectories
450
451 preserve
452 reshape long count_ t_, i(id)
453
454 gen count_jit = count_ + ( 0.2*runiform()-0.1 )
455 graph twoway scatter count_jit t_, c(L) by(_traj_Group) msize(tiny) mcolor(gray) lwidth(vthin)
456     lcolor(gray)
457
458 ***** to calculate the Bayes factor
459
460 log Bayes factor (2loge(B10) ≈ 2(ΔBIC)
461 This estimate approximately equals 2(BICcomplex model-BICnull model)
462
463 */
464
465
466

```



```

467  /*
468
469  Depressive symptoms (CES-D 10 item)
470  The trajectory groups of the CES-D scores (as a discrete variable) are tested
471  alone with time as the only independent variable, with no covariates added that could influence
472  class membership.
473  The censored normal distribution ('cnorm') is applied since the depressive symptom scores were
474  negatively skewed.
475
476  Initially, for each model, the linear, quadratic, and cubic functions of each trajectory can be
477  tested,
478  depending on the number of time points.
479  To ensure parsimony, consistent with the recommendations of Helgeson, Snyder, and Seltman (2004),
480  non-significant cubic and quadratic terms are removed from trajectories in a given model,
481  but linear parameters are retained irrespective of significance.
482
483  I tested the best fitting model with two, three, four five trajectories following the same
484  process.
485  The models were compared (in a table of comparison) using BIC statistics,
486  Bayes factor, entropy, percentage of each class and average posterior probabilities.
487
488  PROCESS TO SELECT THE BEST-FITTING MODEL
489
490  Shape and Classes
491
492  1. run one traj with quadratic (order 2)
493     - If quadratic is not significant run with linear parameter (order 1)
494
495  2. model with 2 traj with quadratic (order 2 2)
496     - If neither traj is significant rerun with linear (order 1 1)
497     - If one not significant adapt accordingly (e.g. order 1 2 OR order 2 1)
498
499  3. Compare models (complex-simple) with statistic criteria
500
501  4. Repeat the process with an increasing number of traj
502
503  */
504
505  *** CNORM MODEL
506
507  * 1 class - cnorm model - quadratic polynomial (2)
508  traj, var(cesd_*) indep(t_*) model(cnorm) min(0) max(30) order(2)
509  trajplot, xtitle(Time in Months) ytitle(Depressive symptoms CES-D) ci
510
511  /* Shows the assigned group and probabilities of group membership */
512  list _traj_Group - _traj_ProbG1 if _n < 3, ab(12)
513
514  /* trajT = x-axis, Avg# = data averages, Est# = model estimates */
515  matrix list e(plot1), format(%9.2f) noheader
516
517
518  /*
519  Trajectory shape
520  2 - p-value sig 0.0000
521  */
522
523
524
525
526  * 2 classes - cnorm model - quadratic polynomial (2 2)
527  traj, var(cesd_*) indep(t_*) model(cnorm) min(0) max(30) order(2 2)
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_ProbG2 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     * 3 classes - cnorm model - quadratic polynomial (2 2 2)
540     traj, var(cesd_*) indep(t_*) model(cnorm) min(0) max(30) order(2 2 2)
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_ProbG3 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
554
555     * 4 classes - cnorm model - quadratic polynomial (2 2 2 2)
556     traj, var(cesd_*) indep(t_*) model(cnorm) min(0) max(30) order(2 2 2 2)
557     trajplot, xtitle(Time in Months) ytitle(Depressive symptoms CES-D) ci
558
559
560
561     /* Shows the assigned group and probabilities of group membership */
562     list _traj_Group - _traj_ProbG4 if _n < 3, ab(12)
563
564     /* trajT = x-axis, Avg# = data averages, Est# = model estimates */
565     matrix list e(plot1), format(%9.2f) noheader
566
567
568
569
570
571
572
573
574     * 5 classes - cnorm model - quadratic polynomial (2 2 2 2 2)
575     traj, var(cesd_*) indep(t_*) model(cnorm) min(0) max(30) order(2 2 2 2 2)
576     trajplot, xtitle(Time in Months) ytitle(Depressive symptoms CES-D) ci
577
578
579
580     /* Shows the assigned group and probabilities of group membership */
581     list _traj_Group - _traj_ProbG4 if _n < 3, ab(12)
582
583     /* trajT = x-axis, Avg# = data averages, Est# = model estimates */
584     matrix list e(plot1), format(%9.2f) noheader
585
586
587
588     * 5 classes - cnorm model - quadratic and cubic polynomial (2 2 2 2 2)
589     traj, var(cesd_*) indep(t_*) model(cnorm) min(0) max(30) order(2 2 3 3 3)
590     trajplot, xtitle(Time in Months) ytitle(Depressive symptoms CES-D) ci
591
592
593     /* Shows the assigned group and probabilities of group membership */
594     list _traj_Group - _traj_ProbG5 if _n < 3, ab(12)
595
596     /* trajT = x-axis, Avg# = data averages, Est# = model estimates */
597     matrix list e(plot1), format(%9.2f) noheader
598

```

```

599
600
601 * 6 classes - cnorm model - quadratic and cubic polynomial (2 2 2 2 2 2)
602 traj, var(cesd_*) indep(t_*) model(cnorm) min(0) max(30) order(2 2 2 2 2 2)
603 trajplot, xtitle(Time in Months) ytitle(Depressive symptoms CES-D) ci
604
605
606 /* Shows the assigned group and probabilities of group membership */
607 list _traj_Group - _traj_ProbG5 if _n < 3, ab(12)
608
609 /* trajT = x-axis, Avg# = data averages, Est# = model estimates */
610 matrix list e(plot1), format(%9.2f) noheader
611
612
613
614 * The 5-model depressive traj is selected to be tested in different shapes.
615
616
617
618
619 * 5 classes - cnorm model - cubic polynomial (3 3 3 3 3)
620 traj, var(cesd_*) indep(t_*) model(cnorm) min(0) max(30) order(3 3 3 3 3)
621 trajplot, xtitle(Time in Months) ytitle(Depressive symptoms CES-D) ci
622
623
624
625 /* Shows the assigned group and probabilities of group membership */
626 list _traj_Group - _traj_ProbG5 if _n < 3, ab(12)
627
628 /* trajT = x-axis, Avg# = data averages, Est# = model estimates */
629 matrix list e(plot1), format(%9.2f) noheader
630
631
632
633
634 * OPTIMAL MODEL
635
636
637 * 5 classes - cnorm model - cubic polynomial (2 2 3 3 3)
638 traj, var(cesd_*) indep(t_*) model(cnorm) min(0) max(30) order(2 2 3 3 3)
639 trajplot, xtitle(Time in Months) ytitle(Depressive symptoms CES-D) ci
640
641
642
643 /* Shows the assigned group and probabilities of group membership */
644 list _traj_Group - _traj_ProbG5 if _n < 3, ab(12)
645
646 /* trajT = x-axis, Avg# = data averages, Est# = model estimates */
647 matrix list e(plot1), format(%9.2f) noheader
648
649
650
651
652
653
654 ** run after each traj model to estimate the average posterior probability (APP) for each group
655
656 program summary_table_procTraj
657 preserve
658 *look at the average posterior probability
659 gen Mp = 0
660 foreach i of varlist _traj_ProbG* {
661     replace Mp = `i' if `i' > Mp
662 }
663 sort _traj_Group
664 *and the odds of correct classification
665 by _traj_Group: gen cesdG = _N
666 by _traj_Group: egen groupAPP = mean(Mp)

```

```

667 by _traj_Group: gen counter = _n
668 gen n = groupAPP/(1 - groupAPP)
669 gen p = cesdG/_N
670 gen d = p/(1-p)
671 gen occ = n/d
672 *Estimated proportion for each group
673 scalar c = 0
674 gen TotProb = 0
675 foreach i of varlist _traj_ProbG* {
676     scalar c = c + 1
677     quietly summarize `i'
678     replace TotProb = r(sum)/_N if _traj_Group == c
679 }
680 gen d_pp = TotProb/(1 - TotProb)
681 gen occ_pp = n/d_pp
682 *This displays the group number [_traj_~p],
683 *the cesd per group (based on the max post prob), [countG]
684 *the average posterior probability for each group, [groupAPP]
685 *the odds of correct classification (based on the max post prob group assignment), [occ]
686 *the odds of correct classification (based on the weighted post. prob), [occ_pp]
687 *and the observed probability of groups versus the probability [p]
688 *based on the posterior probabilities [TotProb]
689 list _traj_Group cesdG groupAPP occ occ_pp p TotProb if counter == 1
690 restore
691 end
692
693 summary_table_procTraj
694
695
696
697
698 /*
699 ---- MODEL SELECTION ----
700 Best-fitting model to try survival analysis is the 5 class - order (2 2 3 3 3)
701 */
702
703
704
705
706 /*
707 Data and variable manipulation
708 */
709
710 * 5-class model: rename _traj_Group to C_traj_group5
711
712 rename _traj_Group C_traj_group5
713 recode C_traj_group5 (3=4) (4=3)
714 ta C_traj_group5
715 rename _traj_ProbG1 C_depres_traj_1
716 rename _traj_ProbG2 C_depres_traj_2
717 rename _traj_ProbG3 C_depres_traj_3
718 rename _traj_ProbG4 C_depres_traj_4
719 rename _traj_ProbG5 C_depres_traj_5
720
721
722
723 * labelling variable of C_traj_group4
724
725 label var C_traj_group5 "Traj 5 groups of depressive symptoms"
726
727 * labelling values
728 lab def traj_depres 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 C_traj_group5 traj_depres
733
734 ta C_traj_group5

```

```

735
736
737
738
739
740
741
742
743
744 * Frequencies of covariates
745
746 tabulate C_age
747 summarize C_age
748
749
750 tabulate C_sex
751 summarize C_sex
752
753
754 tabulate C_education
755 summarize C_education
756
757
758 tabulate C_maritalstatus_4cat
759 summarize C_maritalstatus_4cat
760
761
762 tabulate Cwv1_netwealth_quintiles
763 summarize Cwv1_netwealth_quintiles
764
765
766 tabulate Cwv1_smoking_3cat
767 summarize Cwv1_smoking_3cat
768
769
770 tabulate Cwv1_physicalactivity
771 summarize Cwv1_physicalactivity
772
773
774
775 tabulate Cwv1_alcohol_status
776 summarize Cwv1_alcohol_status
777
778
779 tabulate Cwv1_cvd_comorbidity
780 summarize Cwv1_cvd_comorbidity
781
782
783 tabulate Cwv1_memory_wordrecall
784 summarize Cwv1_memory_wordrecall
785
786
787
788
789
790
791
792
793
794 /* MULTIPLE IMPUTATION (MI)
795
796 To handle with missing values of baseline and time 3 covariates
797
798
799 useful sources for MI and MICE:
800
801 https://stats.idre.ucla.edu/stata/seminars/mi\_in\_stata\_pt1\_new/
802 https://www.stata.com/manuals/mi.pdf - see page 139

```

```

803 https://www.stata.com/meeting/switzerland16/slides/medeiros-switzerland16.pdf
804 https://www.youtube.com/watch?v=i6S0lq0mjuc&ab\_channel=StataCorpLLC
805 https://dss.princeton.edu/training/MIS stata.pdf
806
807
808
809 Preparing to conduct MI
810 1. examine the number and proportion of missing values among the variables of interest
811     use the mdesc command
812 2. examine missing data patterns
813     use commands mi set and mi misstable patterns
814 3. identify potential auxiliary variables
815
816
817 Run MI using chained equations (MICE)
818 using the commands
819 1. how (in what style) to store the imputations
820     mi set wide
821 2. which variables will be imputed
822     mi register imputed
823 3. optionally, which variables will not be imputed
824     mi register regular
825 4. what imputation method is implemented to impute each of var - MICE
826     mi impute chained
827
828 */
829
830
831
832
833
834 /*
835
836 1. examining missing values
837     install packages:
838     * install mdesc
839     * install tabmiss
840     * insatll dm31
841     * insall mvpatterna
842
843 */
844
845 search mdesc
846 search rmiss2
847 search mvpatterns
848
849
850
851
852 * examining number of missing values vs non-missing in each variable
853
854 mdesc C_age C_sex C_education C_maritalstatus_4cat Cwv1_netwealth_quintiles ///
855 Cwv4_smoking_3cat Cwv4_physicalactivity Cwv4_alcohol_status ///
856 Cwv4_cvd_comorbidity Cwv3_glycemia Cwv3_bp Cwv3_obesity Cwv3_hdl_cholesterol Cwv3_crp ///
857 Cwv4_memory_wordrecall
858
859
860 * examining missing data patterns
861
862 mi set wide
863
864 mi misstable summarize C_age C_sex C_education C_maritalstatus_4cat Cwv1_netwealth_quintiles ///
865 Cwv4_smoking_3cat Cwv4_physicalactivity Cwv4_alcohol_freq Cwv4_alcohol_status ///
866 Cwv4_cvd_comorbidity Cwv3_glycemia Cwv3_bp Cwv3_obesity Cwv3_hdl_cholesterol Cwv3_crp ///
867 Cwv4_memory_wordrecall Cwv1_antidepressant Cwv1_psycholog_treat Cwv1_anytreat_psyche
868
869
870

```

```

871 mi misstable patterns C_age C_sex C_education C_maritalstatus_4cat Cwv1_netwealth_quintiles ///
872 Cwv4_smoking_3cat Cwv4_physicalactivity Cwv4_alcohol_freq Cwv4_alcohol_status ///
873 Cwv4_cvd_comorbidity Cwv3_glycemia Cwv3_bp Cwv3_obesity Cwv3_hdl_cholesterol Cwv3_crp ///
874 Cwv4_memory_wordrecall Cwv1_antidepressant Cwv1_psycholog_treat Cwv1_anytreatPsyche
875
876
877
878 /*
879 identifying potential auxiliary var
880 * Auxiliary variables are either correlated with a missing variable(s)
881 (the recommendation is  $r > 0.4$ ) or are believed to be associated with missingness
882 - a priori knowledge of var that would make good auxiliary var
883 - identify potential candidates by examining associations between missing var and other var in
the dataset
884 running correlation using the command: pwcorr v1 v2 v3, obs
885 the recommendation for good correlation is  $r > 0.4$ 
886
887
888 Missing var to be imputed:
889
890 C_age Cwv1_netwealth_quintiles
891 Cwv4_smoking_3cat Cwv4_alcohol_status
892 Cwv4_cvd_comorbidity Cwv3_glycemia Cwv3_bp Cwv3_obesity Cwv3_hdl_cholesterol Cwv3_crp
893 Cwv4_memory_wordrecall
894
895
896
897 Potential auxiliary var:
898 DV: Cwv4_self_info_dementia
899 IV: C_traj_group4 cesd_0 cesd_1 cesd_2 cesd_3
900 other var: C_sex C_education C_maritalstatus_4cat
901
902 */
903
904
905 * correlation
906
907 pwcorr C_age Cwv1_netwealth_quintiles ///
908 Cwv4_smoking_3cat Cwv4_alcohol_status ///
909 Cwv4_cvd_comorbidity Cwv3_glycemia Cwv3_bp Cwv3_obesity Cwv3_hdl_cholesterol Cwv3_crp ///
910 Cwv4_memory_wordrecall ///
911 Cwv4_self_info_dementia C_traj_group4 cesd_0 cesd_1 cesd_2 cesd_3 ///
912 C_sex C_education C_maritalstatus_4cat, obs
913
914
915 * The correlation showed that all the above potential var are good auxiliary
916 * A good auxiliary does not have to be correlated with every variable to be useful
917 * And it's not problematic if it has missing info of it's own
918
919
920
921
922 /*
923 MI by chained equations (MICE)
924 see: https://stats.idre.ucla.edu/stata/seminars/mi\_in\_stata\_pt1\_new/
925
926 MICE is known as the fully conditional specification or sequential generalized regression
927 does not assume a joint MVN distribution
928 but instead uses a separate conditional distribution for each imputed variable.
929
930 The multivariate normal (MVN) model - mi imputed mvn -
931 assumes multivariate normality of all var
932
933 The multivariate imputation by chained equations (MICE) - mi imputed chained -
934 offers flexibility in how each var is modeled
935
936 mi impute chained allows to specify models for a
937 variety of variable types, including

```

```

938 continuous, binary, ordinal, nominal, truncated, and count variables
939
940
941 The MICE distributions available in Stata are:
942 binary, ordered and multinomial logistic regression for categorical variables,
943 linear regression and predictive mean matching (PMM)* for continuous variables,
944 and Poisson and negative binomial regression for count variables.
945
946
947 IMPUTATION PHASES
948
949
950 1. mi set wide
951     style to store imputations
952
953 2. mi register imputed
954     identifies which variables in the imputation model have missing information.
955
956 3. mi register regular (! optional)
957     which variables will not be imputed
958
959 4. mi impute chained
960     where the user specifies the imputation model to be used
961     and the number of imputed datasets to be created.
962     Example:
963         mi impute chained (regress) bmi age (logit) female ///
964         (mlogit) race = bpdiastr i.region, add(20)
965
966 5. mi estimate
967     is used as a prefix to the standard regress command.
968     This executes the specified estimation model within each of the 20 imputed datasets
969     to obtain 20 sets of coefficients and standard errors.
970     Stata then combines these estimates to obtain one set of inferential statistics.
971     In the output from mi estimate you will see some metrics: Imputation Diagnostics
972     information for RVI (Relative Increase in Variance),
973     FMI (Fraction of Missing Information),
974     DF (Degrees of Freedom) ,
975     RE (Relative Efficiency),
976     and the between imputation and the within imputation variance estimates
977     to examine how the standard errors (SEs) are calculated.
978
979
980
981 -----
982
983 SELECTING MY IMPUTATION MODEL
984
985 - MICE -> mi impute chained
986
987 - var to be imputed:
988
989     linear regression for continuous var (regress) ->
990     C_age Cwv1_memory_wordrecall
991
992     logistic for the binary var (logit) ->
993     Cwv1_cvd_comorbidity
994
995     multinomial logistic for our nominal categorical var (mlogit) ->
996     Cwv1_netwealth_quintiles
997     Cwv1_smoking_3cat Cwv1_alcohol_status
998
999
1000
1001 - auxiliary var:
1002
1003     DV -> Cwv4_self_info_dementia
1004     IV -> C_traj_group4
1005     other covariates -> C_sex C_education C_maritalstatus_4cat

```



```

1006
1007
1008
1009 - imputation numbers (m) -> 20
1010
1011     White et al. (2010) recommendation: use the rule that m should equal the percentage of
incomplete cases
1012
1013
1014 - rseed (53421) for reproducability reasons
1015
1016
1017 - (! OPTIONAL) advance impute options -> force
1018
1019     proceed with imputation, even when missing imputed values (e.g. auxiliary have missing data)
are encountered
1020
1021 - impute options -> savetrace (trace1)
1022
1023     specifies Stata to save the means and standard deviations of imputed values from each
iteration to a Stata dataset named "trace1
1024 */
1025
1026
1027 mi set wide
1028
1029
1030 mi register imputed C_age Cwv1_netwealth_quintiles ///
1031     Cwv1_smoking_3cat Cwv1_alcohol_status ///
1032     Cwv1_cvd_comorbidity ///
1033     Cwv1_memory_wordrecall
1034
1035
1036
1037
1038
1039 mi impute chained (logit) Cwv1_cvd_comorbidity ///
1040 (mlogit) Cwv1_netwealth_quintiles Cwv1_smoking_3cat Cwv1_alcohol_status ///
1041 (regress) C_age Cwv1_memory_wordrecall = Cwv4_self_info_dementia C_traj_group4 ///
1042 C_sex C_eduaction C_maritalstatus_4cat, add(20) rseed(53421) savetrace(trace1)
1043
1044
1045 * save imputed data
1046
1047 * plot imputations
1048
1049
1050 *it will open a file named trace1
1051 use trace1, clear
1052
1053 describe
1054
1055
1056 reshape wide *mean *sd, i(iter) j(m)
1057
1058 tsset iter
1059
1060
1061
1062
1063 /*
1064 The trace plot below graphs the predicted means value produced during the first imputation chain.
1065 As before, the expectations is that the values would vary randomly to incorporate variation into
the predicted values for read.
1066 */
1067
1068 tsline C_age_mean1, name(mice1,replace) legend(off) ytitle("Mean of age")
1069 tsline Cwv1_netwealth_quintiles_mean1, name(mice1,replace) legend(off) ytitle("Mean of wealth")

```

```

1070     tsline Cwv1_smoking_3cat_mean1, name(mice1,replace)legend(off) ytitle("Mean of smoking")
1071     tsline Cwv1_alcohol_status_mean1, name(mice1,replace)legend(off) ytitle("Mean of alcohol")
1072     tsline Cwv1_cvd_comorbidity_mean1, name(mice1,replace)legend(off) ytitle("Mean of cvd")
1073     tsline Cwv1_memory_wordrecall_mean1, name(mice1,replace)legend(off) ytitle("Mean of memory")
1074
1075
1076     /*
1077
1078     All 10 imputation chains can also be graphed simultaneously to make sure that nothing unexpected
1079     occurred in a single chain.
1080     Every chain is obtained using a different set of initial values and this should be unique.
1081     Each colored line represents a different imputation.
1082     So all 10 imputation chains are overlaid on top of one another.
1083
1084     */
1085
1086     tsline Cwv1_memory_wordrecall_mean*, name(mice1,replace)legend(off) ytitle("Mean of memory")
1087     tsline Cwv1_memory_wordrecall_sd*, name(mice2, replace) legend(off) ytitle("SD of memory")
1088     graph combine mice1 mice2, xcommon cols(1) title(Trace plots of summaries of imputed values)
1089
1090     * repeat for each imputed var
1091
1092
1093
1094
1095
1096
1097     /*
1098     ---- DESCRIPTIVE STATISTICS ----
1099
1100     General characteristics of participants
1101
1102     General characteristics of participnats stratified for study inclusion
1103
1104     General characteristics of participants stratified for dementia occurence
1105
1106     Participant characteristics by depressive symptom trajectory group
1107
1108     1. CHI-SQUARE (chi2) for categorical var (crosstabulation)
1109         Frequency tables -> two-way tables
1110         using the command tabulate, chi2
1111         reporting observations, column percentage (N, %) and p-value of Pearson's r
1112
1113
1114     2. one-way ANOVA for continuous var
1115         check box plot
1116         using the command oneway
1117         reporting mean, sd (summary tables) and p-value of F
1118     */
1119
1120
1121
1122
1123     * General characteristics of CHARLS participants
1124
1125     * Demographics
1126     sum C_age
1127     ta C_sex
1128     ta C_education
1129     ta C_maritalstatus_4cat
1130     ta Cwv1_netwealth_quintiles
1131     * Lifestyle factors
1132     ta Cwv1_smoking_3cat
1133     ta Cwv1_alcohol_status
1134     ta Cwv1_physicalactivity
1135     * Cardiometabolic health
1136     ta Cwv1_cvd_comorbidity

```

```

1137 ta Cwv1_diabetes_report
1138 ta Cwv1_HbA1c
1139 ta Cwv1_crp
1140 ta Cwv1_hdl_cholesterol
1141 ta Cwv1_waist
1142 ta Cwv1_systolic_bp
1143 ta Cwv1_diastolic_bp
1144 * Depressive symptoms t1-t3 (cont and categ)
1145 sum cesd_0
1146 sum cesd_1
1147 sum cesd_2
1148 sum cesd_3
1149 ta depress_0
1150 ta depress_1
1151 ta depress_2
1152 ta depress_3
1153 * Memory score at baseline
1154 sum Cwv1_memory_wordrecall
1155
1156
1157
1158
1159
1160
1161 * Sample characteristics by depressive symptom trajectories
1162 * crosstabs categ var (frequencies and chi2) !report column percentage!
1163 * oneway ANOVA cont var (mean, sd)
1164
1165
1166 * Demographics
1167 oneway C_age C_traj_group5, tabulate
1168 ta C_sex C_traj_group5, chi2 column row
1169 ta C_education C_traj_group5, chi2 column row
1170 ta C_maritalstatus_4cat C_traj_group5, chi2 column row
1171 ta Cwv1_netwealth_quintiles C_traj_group5, chi2 column row
1172 * Lifestyle factors
1173 ta Cwv1_smoking_3cat C_traj_group5, chi2 column row
1174 ta Cwv1_alcohol_status C_traj_group5, chi2 column row
1175 ta Cwv1_physicalactivity C_traj_group5, chi2 column row
1176 * Cardiometabolic health
1177 ta Cwv1_cvd_comorbidity C_traj_group5, chi2 column row
1178 ta Cwv1_diabetes_report C_traj_group5, chi2 column row
1179 ta Cwv1_HbA1c C_traj_group5, chi2 column row
1180 ta Cwv1_crp C_traj_group5, chi2 column row
1181 ta Cwv1_hdl_cholesterol C_traj_group5, chi2 column row
1182 ta Cwv1_waist C_traj_group5, chi2 column row
1183 ta Cwv1_systolic_bp C_traj_group5, chi2 column row
1184 ta Cwv1_diastolic_bp C_traj_group5, chi2 column row
1185 * Memory score
1186 oneway Cwv1_memory_wordrecall C_traj_group5, tabulate
1187
1188
1189
1190
1191 /*
1192 ---- BINOMIAL LOGISTIC REGRESSION ON COMPLETE DATA ----
1193
1194 Command is:
1195 logistic DV IVs
1196 OR
1197 logit DV IVs, or
1198
1199
1200 --- Building the model using baseline covariates ---
1201
1202 Model 1: unadjusted - single predictor of depressive symptom trajectories C_traj_group5
1203 Model 2: model 1 + sociodemographics: age sex education marital status and wealth
1204 Model 3: model 2 + health behaviours: smoking, alcohol consumption

```

```

1205
1206
1207 */
1208
1209
1210
1211 * Unadjusted model - model 1 - single predictor
1212
1213 logistic Cwv4_self_info_dementia C_traj_group5
1214
1215 *OR
1216
1217 logit Cwv4_self_info_dementia C_traj_group5
1218
1219 *OR
1220
1221 logit Cwv4_self_info_dementia C_traj_group5, or
1222
1223
1224
1225 * define design var by using i.(decreasing, increasing, high ref: low)
1226
1227 logistic Cwv4_self_info_dementia i.C_traj_group5
1228
1229 *OR
1230
1231
1232 logit Cwv4_self_info_dementia i.C_traj_group5, or
1233
1234
1235 * Adjusted models - multivariable logistic regression
1236 * controlling for covariates
1237
1238 * model 2: model 1 + adjust for demographics: age sex education marital status and wealth
1239
1240 logistic Cwv4_self_info_dementia i.C_traj_group5 ///
1241 C_age C_sex i.C_education i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles
1242
1243 * model 3: model 2 + adjust for lifestyle and cardiovascular factors
1244
1245 logistic Cwv4_self_info_dementia i.C_traj_group5 ///
1246 C_age C_sex i.C_education i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles ///
1247 i.Cwv1_smoking_3cat i.Cwv1_alcohol_status i.Cwv1_cvd_comorbidity
1248
1249
1250
1251
1252
1253
1254
1255
1256 /* ----- BINOMIAL LOGISTIC REGRESSION IN IMPUTED DATASET using time 3 covariates ----- */
1257
1258 Command is
1259
1260 mi estimate : logit DV IV, or
1261
1262 OR
1263
1264 mi estimate: logistic DV IV
1265
1266 */
1267
1268
1269
1270 * to redefine reference group to be the trajectory of "minimal symptoms" then use
1271 ib2.exposure_var -- the ib and the nnumber of var make the change in reference groups

```

```

1272
1273
1274
1275 * Unadjusted model - model 1 - single predictor
1276
1277 mi estimate, eform("Odds Ratio"): logistic Cwv4_self_info_dementia C_traj_group5
1278
1279 *OR
1280
1281 mi estimate, eform("Odds Ratio"): logit Cwv4_self_info_dementia C_traj_group5, or
1282
1283
1284
1285 * define design var by using i.(decreasing, increasing, high, ref: low)
1286
1287 mi estimate, eform("Odds Ratio"): logistic Cwv4_self_info_dementia i.C_traj_group5
1288
1289 *OR
1290
1291 mi estimate, eform("Odds Ratio"): logit Cwv4_self_info_dementia i.C_traj_group5, or
1292
1293
1294 * Adjusted models - multivariable logistic regression
1295 * controlling for covariates
1296
1297 * model 2: model 1 + adjust for demographics: age sex education marital status and wealth
1298
1299 mi estimate, eform("Odds Ratio"): logistic Cwv4_self_info_dementia i.C_traj_group5 ///
1300 C_age C_sex i.C_education i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles
1301
1302 * model 3: model 2 + adjust for lifestyle factors
1303
1304 mi estimate, eform("Odds Ratio"): logistic Cwv4_self_info_dementia i.C_traj_group5 ///
1305 C_age C_sex i.C_education i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles ///
1306 i.Cwv1_smoking_3cat i.Cwv1_alcohol_status i.Cwv1_cvd_comorbidity
1307
1308
1309
1310
1311
1312
1313
1314 /*
1315
1316 *** SENSITIVITY ANALYSES ***
1317
1318
1319 1) LCGA logit trajectories with dichotomous variable
1320
1321
1322 2) Complete data
1323
1324 */
1325
1326
1327
1328
1329
1330 /*
1331 1) Logistic model LCGA
1332
1333 use Ewv2_depressive_symptoms dichotomous variables (0-1)
1334
1335 Logistic (logit) model
1336
1337 use http://www.andrew.cmu.edu/user/bjones/traj/data/cambridge.dta,
1338 clear
1339

```

```

1340     traj, var(p1-p23) indep(tt1-tt23) model(logit) order(0 3 3)
1341
1342     trajplot, xtitle(Scaled Age) ytitle(Prevalence)
1343
1344     /* Assigned group and probabilities of group membership */
1345     list _traj_Group - _traj_ProbG3 if _n > 400, ab(12)
1346
1347 */
1348
1349
1350
1351
1352
1353
1354 *rename Cwv1_depressive_symptoms score across the waves
1355
1356 rename Cwv1_depressive_symptoms depress_0
1357 rename Cwv2_depressive_symptoms depress_1
1358 rename Cwv3_depressive_symptoms depress_2
1359 rename Cwv4_depressive_symptoms depress_3
1360
1361
1362 net from http://www.andrew.cmu.edu/user/bjones/traj
1363 net install traj, force
1364 help traj
1365
1366
1367
1368
1369 *** LOGIT MODEL
1370
1371
1372 * 1 class - logit model - quadratic polynomial (2)
1373 traj, var(depress_*) indep(t_*) model(logit) order(2)
1374 trajplot, xtitle(Time in Months) ytitle(Depressive symptom caseness) ci
1375
1376 /* Assigned group and probabilities of group membership */
1377 list _traj_Group - _traj_ProbG1 if _n < 3, ab(12)
1378
1379
1380
1381
1382 * 2 class - logit model - quadratic polynomial (2 2)
1383 traj, var(depress_*) indep(t_*) model(logit) order(2 2)
1384 trajplot, xtitle(Time in Months) ytitle(Depressive symptom caseness) ci
1385
1386 /* Assigned group and probabilities of group membership */
1387 list _traj_Group - _traj_ProbG2 if _n < 3, ab(12)
1388
1389
1390
1391
1392
1393 * 3 class - logit model - quadratic polynomial (2 2 2)
1394 traj, var(depress_*) indep(t_*) model(logit) order(2 2 2)
1395 trajplot, xtitle(Time in Months) ytitle(Depressive symptom caseness) ci
1396
1397 /* Assigned group and probabilities of group membership */
1398 list _traj_Group - _traj_ProbG3 if _n < 3, ab(12)
1399
1400
1401
1402
1403 * 4 class - logit model - quadratic polynomial (2 2 2 2)
1404 traj, var(depress_*) indep(t_*) model(logit) order(2 2 2 2)
1405 trajplot, xtitle(Time in Months) ytitle(Depressive symptom caseness)
1406
1407 /* Assigned group and probabilities of group membership */

```

```

1408     list _traj_Group - _traj_ProbG4 if _n < 3, ab(12)
1409
1410
1411
1412 * 5 class - logit model - quadratic polynomial (2 2 2 2 2)
1413 traj, var(depress_*) indep(t_*) model(logit) order(2 2 2 2 2)
1414 trajplot, xtitle(Time in Months) ytitle(Depressive symptom caseness)
1415
1416 /* Assigned group and probabilities of group membership */
1417     list _traj_Group - _traj_ProbG5 if _n < 3, ab(12)
1418
1419
1420
1421
1422
1423
1424
1425 * The 4-model depressive traj is selected to be tested in different shapes.
1426
1427 * 4 class - logit model - quadratic polynomial (3 3 3 3)
1428 traj, var(depress_*) indep(t_*) model(logit) order(3 3 3 3)
1429 trajplot, xtitle(Time in Months) ytitle(Depressive symptom caseness) ci
1430
1431 /* Assigned group and probabilities of group membership */
1432     list _traj_Group - _traj_ProbG4 if _n < 3, ab(12)
1433
1434
1435
1436
1437
1438 * The 4-model depressive traj is selected to be tested in different shapes.
1439
1440 * 4 class - logit model - quadratic polynomial (3 2 3 3)
1441 traj, var(depress_*) indep(t_*) model(logit) order(3 2 3 3)
1442 trajplot, xtitle(Time in Months) ylabel(0(.20)1) ytitle(Depressive symptom caseness)
1443
1444 /* Assigned group and probabilities of group membership */
1445     list _traj_Group - _traj_ProbG4 if _n < 3, ab(12)
1446
1447
1448
1449
1450
1451 program summary_table_procTraj
1452     preserve
1453     *look at the average posterior probability
1454     gen Mp = 0
1455     foreach i of varlist _traj_ProbG* {
1456         replace Mp = `i' if `i' > Mp
1457     }
1458     sort _traj_Group
1459     *and the odds of correct classification
1460     by _traj_Group: gen cesdG = _N
1461     by _traj_Group: egen groupAPP = mean(Mp)
1462     by _traj_Group: gen counter = _n
1463     gen n = groupAPP/(1 - groupAPP)
1464     gen p = cesdG/_N
1465     gen d = p/(1-p)
1466     gen occ = n/d
1467     *Estimated proportion for each group
1468     scalar c = 0
1469     gen TotProb = 0
1470     foreach i of varlist _traj_ProbG* {
1471         scalar c = c + 1
1472         quietly summarize `i'
1473         replace TotProb = r(sum)/_N if _traj_Group == c
1474     }
1475     gen d_pp = TotProb/(1 - TotProb)

```



```

1476     gen occ_pp = n/d_pp
1477     *This displays the group number [_traj_~p],
1478     *the cesd per group (based on the max post prob), [countG]
1479     *the average posterior probability for each group, [groupAPP]
1480     *the odds of correct classification (based on the max post prob group assignment), [occ]
1481     *the odds of correct classification (based on the weighted post. prob), [occ_pp]
1482     *and the observed probability of groups versus the probability [p]
1483     *based on the posterior probabilities [TotProb]
1484     list _traj_Group cesdG groupAPP occ occ_pp p TotProb if counter == 1
1485     restore
1486 end
1487
1488 summary_table_procTraj
1489
1490
1491
1492 ta _traj_Group
1493
1494
1495
1496
1497
1498
1499 * IMPUTED DATA: Logistic regression (Odds Ratio)
1500
1501 * Unadjusted model (model 1)
1502
1503 mi estimate, eform("Odds Ratio"): logistic Cwv4_self_info_dementia i._traj_Group
1504
1505 * model 2: model 1 + adjust for demographics: age sex education marital status and wealth
1506
1507 mi estimate, eform("Odds Ratio"): logistic Cwv4_self_info_dementia i._traj_Group ///
1508 C_age C_sex i.C_education i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles
1509
1510 * model 3: model 2 + adjust for lifestyle and health factors
1511
1512 mi estimate, eform("Odds Ratio"): logistic Cwv4_self_info_dementia i._traj_Group ///
1513 C_age C_sex i.C_education i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles ///
1514 i.Cwv1_smoking_3cat i.Cwv1_alcohol_status i.Cwv1_cvd_comorbidity
1515
1516
1517
1518 * 2) complete data analysis (see above)
1519
1520
1521
1522 * ----- *
1523
1524
1525
1526

```