

```

1
2  cls
3  clear all
4
5  *****
6  **#CLEANING**
7  *****
8  cap use "...", replace
9  drop hba1c_date fgt fgt_date smk_date weight w_date index_hes yob cause uts_cprd index_cprd hes_end pracid year agein_grp bmi_grp dob
10 rename smk_status smk
11
12 sum bmi
13 egen float bmig = cut(bmi), at(0 18.5 25 30 35 40 100) icodes label
14 tabstat bmi, statistics(N min max) by(bmig)
15 tab bmig, nolab
16 label define bmig 0 ` "<18.5"', modify
17 label define bmig 1 ` "≥18.5 to <25"', modify
18 label define bmig 2 ` "≥25 to <30"', modify
19 label define bmig 3 ` "≥30 to <35"', modify
20 label define bmig 4 ` "≥35 to <40"', modify
21 label define bmig 5 ` "≥40"', modify
22 tabstat bmi, statistics(min max) by(bmig)
23
24 *All_cancer*
25 gen all = 1 if allcancer_site != ""
26 replace all = 0 if allcancer_site == ""
27 distinct patid
28 mdesc
29 order index allcancer_site allcancer_end dodeath, after(gender)
30 gen dif = allcancer_end - dodeath
31 sum dif
32 mdesc dif
33 drop dif
34 gen t_all = (allcancer_end - index)/365.24
35 drop allcancer_site allcancer_end
36
37 *Four commonest_cancer and BMI_related*
38 foreach cancer in bmirel lung breast prostate colorectal {
39     gen t_`cancer' = (`cancer'_end - index)/365.24, after(`cancer')
40     drop `cancer'_end
41 }
42 drop liver_end liver pancreas_end pancreas gallbladder_end gallbladder endometrial_end endometrial
43 compress
44 order lung t_lung breast t_breast prostate t_prostate colorectal t_colorectal bmirel t_bmirel all t_all, last
45 cap save "...", replace
46
47 *****
48 **#CHARACTERISTICS**
49 *****
50 cap use "...", replace
51
52 distplot agein, over(gender) xlabel(35(5)105, grid) ylabel(#10, format(%5.1f) angle(h))
53 graph close _all
54 sum agein, d
55 egen float ageg = cut(agein), at(0 30 40 50 60 70 80 110) icodes label
56 tabstat agein, statistics(N min max) by(ageg)
57 tab ageg, nolab
58 label define ageg 1 ` "<40"', modify
59 label define ageg 2 ` "≥40 to <50"', modify
60 label define ageg 3 ` "≥50 to <60"', modify
61 label define ageg 4 ` "≥60 to <70"', modify
62 label define ageg 5 ` "≥70 to <80"', modify
63 label define ageg 6 ` "≥80"', modify
64 tabstat agein, statistics(N min max) by(ageg)
65
66 label variable agein ` "Age at diabetes diagnosis (years)""
67 label variable ageg ` "Age at diabetes diagnosis group (years)""

```

```

68 label variable bmi      ` "Body mass index (kg/m2)"`
69 label variable bmig     ` "Body mass index category (kg/m2)"`
70
71 ****Table1, including age at diagnosis****
72 preserve
73 drop if gender == 2
74 baselinetable           /*
75 */   agein(cts tab("p50 (p25, p75)")) /*
76 */   bmi(cts tab("p50 (p25, p75)"))   /*
77 */   bmig(cat)                       /*
78 */   , by(ageg, total missing) reportmissing countformat(%10.0fc) notable /*
79 */   exportexcel("...", replace)
80 restore
81
82 preserve
83 drop if gender == 1
84 baselinetable           /*
85 */   agein(cts tab("p50 (p25, p75)")) /*
86 */   bmi(cts tab("p50 (p25, p75)"))   /*
87 */   bmig(cat)                       /*
88 */   , by(ageg, total missing) reportmissing countformat(%10.0fc) notable /*
89 */   exportexcel("...", replace)
90 restore
91
92 ****TableS1 [ICD-10 codes for BMI-related cancers]****
93
94 ****TableS2, rates****
95 stset t_all, f(all == 1)
96 sum t_all, d
97 strate, per(1000)
98 tab all gender, m col
99 strate gender, per(1000)
100
101 foreach var of varlist lung breast prostate colorectal bmirel all {
102     forvalues j = 1/2 {
103         qui stset t_`var', f(`var' == 1)
104         qui strate ageg if gender == `j', per(1000) output("...", replace)
105     }
106 }
107
108 clear
109 tempfile ratescan
110 save `ratescan', emptyok replace
111 foreach nm in lung breast prostate colorectal bmirel all {
112     forvalues j = 1/2 {
113         qui use "..." , clear
114         qui gen cancer = "`nm'"
115         qui gen sex = `j'
116         qui append using `ratescan'
117         save `ratescan', replace
118     }
119 }
120 use `ratescan', replace
121 replace cancer = proper(cancer)
122 replace cancer = "All cancers" if cancer == "All"
123 renames sex _D _Y _Rate _Lower _Upper cancer \ Sex Events PYRs1000 Rate1000 LB95 UB95 Type
124 order Sex Type, first
125 tostring Sex, replace
126 replace Sex = "Men" if Sex == "1"
127 replace Sex = "Women" if Sex == "2"
128 gsort -Sex Type ageg
129 foreach var of varlist PYRs1000-UB95 {
130     tostring `var', format(%5.1f) force replace
131 }
132 tostring Events, replace
133 drop PYRs1000
134 gen Rate = Rate1000 + " (" + LB95 + ", " + UB95 + ")"

```

```

135 drop Rate1000-UB95
136 reshape wide Events Rate, i(Sex Type) j(ageg)
137 replace Type = "zBmirel" if Type == "Bmirel" /*last in the table*/
138 gsort -Sex Type
139 compress
140 export excel using "...", firstrow(variables) replace
141
142 foreach nm in lung breast prostate colorectal bmirel all {
143     forvalues j = 1/2 {
144         qui erase "..."
145     }
146 }
147
148 *****
149 **REGRESSIONS**
150 *****
151
152 ****All and 4 most-common, by sex****
153 cap use "...", replace
154 rename agein ageatt
155
156 preserve
157 keep if gender == 1
158 tempfile db1_Men
159 save `db1_Men', replace
160 restore
161
162 preserve
163 keep if gender == 2
164 tempfile db1_Women
165 save `db1_Women', replace
166 restore
167
168
169 foreach nm in Men Women {
170
171     foreach type in lung breast prostate colorectal all {
172
173         di "-----"
174         di "SEX -- `nm' || TYPE -- `type' -- $S_TIME $S_DATE"
175         di "-----"
176
177         if ("`nm'" == "Men" & "`type'" == "breast") | ("`nm'" == "Women" & "`type'" == "prostate") {
178             di "Few events"
179         }
180
181         else {
182
183             qui use `db1_`nm'', clear
184
185             stset t_`type', f(`type' == 1) id(patid)
186             qui stmt, time1(df(4)) time2(df(4) start(ageatt)) timeint(t1:t2 4:4)
187
188             preserve
189             forvalues j = 0(0.1)20.1 { /*time since diagnosis/duration, i.e. timeinc*/
190                 qui clear
191                 qui set obs 1201
192                 qui generate time1 = `j'
193                 qui range time2 0 120 1201 /*attained age*/
194                 qui tempfile temppred`n'
195                 qui save `temppred`n''
196                 local datalist `datalist' `temppred`n''
197                 local n = `n' + 1
198             }
199             qui clear
200             qui set obs 0
201             qui append using `datalist'

```

```

202     qui tempfile timedata
203     qui save `timedata'
204     restore
205
206     qui merge 1:1 _n using `timedata', nogen
207
208     preserve
209     qui predict rate, hazard time1var(time1) time2var(time2) ci
210     qui gen agedd_p = time2 - time1
211     qui renames time1 time2 \ diabdur_p ageatt_p
212     qui order agedd_p, before(diabdur_p)
213     qui keep agedd_p-ageatt_p rate*
214     qui cap drop if time1 == .
215     qui gen gender = "`nm'"
216     qui gen type = "`type'"
217     qui tempfile rates_`nm'`type'
218     qui save `rates_`nm'`type'
219     restore
220
221     qui drop time1 time2
222     qui egen float time2 = seq() in 1/6, from(5) to(10) block(1)
223     qui replace time2 = time2*10
224     qui gen time1_5 = 5 if time2 !=.
225     qui gen time1_20 = 20 if time2 !=.
226     qui predictnl lnhr = ln(predict(hazard time1var(time1_20) time2var(time2))) - ln(predict(hazard time1var(time1_5) time2var(time2))), ci(lnhr_lci lnhr_uci)
227     qui keep time2 time1_5 time1_20 lnhr*
228     qui cap drop if time2 == .
229     qui gen gender = "`nm'"
230     qui gen type = "`type'"
231     qui tempfile ratesratio_`nm'`type'
232     qui save `ratesratio_`nm'`type'
233 }
234 }
235 }
236
237 clear
238 foreach nm in Men Women {
239     foreach type in lung breast prostate colorectal all {
240         cap append using `rates_`nm'`type'
241     }
242 }
243 foreach var of varlist rate* {
244     replace `var' = `var'*1000
245 }
246 foreach var of varlist agedd_p diabdur_p ageatt_p {
247     tostring `var', replace force format(%5.1f)
248 }
249 foreach var of varlist agedd_p diabdur_p ageatt_p {
250     destring `var', replace
251 }
252 cap drop if rate == .
253 duplicates drop
254 cap save "...", replace
255
256 clear
257 foreach nm in Men Women {
258     foreach type in lung breast prostate colorectal all {
259         cap append using `ratesratio_`nm'`type'
260     }
261 }
262 cap save "...", replace
263
264
265 *****BMI-related, by sex and BMI*****
266 cap use "...", replace
267 rename agein ageatt
268 drop if bmi == .

```

```

269 drop if bmig == 0 /*removed very few subjects with BMI<18.5*/
270 tab bmig, gen(dbmi)
271 tabstat bmi if dbmi1 == 1, statistics(min max)
272
273 preserve
274 keep if gender == 1
275 tempfile db1_Men
276 save `db1_Men', replace
277 restore
278
279 preserve
280 keep if gender == 2
281 tempfile db1_Women
282 save `db1_Women', replace
283 restore
284
285 foreach nm in Men Women {
286
287     di "SEX -- `nm' -- $S_TIME $S_DATE"
288
289     qui use `db1_`nm'', clear
290
291     stset t_bmirel, f(bmirel == 1) id(patid)
292     stmt dbmi2 dbmi3 dbmi4 dbmi5, ///
293     time1(df(4) tvc(dbmi2 dbmi3 dbmi4 dbmi5) dftvc(4)) ///
294     time2(df(4) tvc(dbmi2 dbmi3 dbmi4 dbmi5) dftvc(4) start(ageatt)) ///
295     timeint(t1:t2 4:4)
296
297     preserve
298     forvalues j = 0(0.1)20.1 { /*time since diagnosis/duration, i.e. timeinc*/
299         qui clear
300         qui set obs 1201
301         qui generate time1 = `j'
302         qui range time2 0 120 1201 /*attained age*/
303         qui tempfile temppred`n'
304         qui save `temppred`n'
305         local datalist `datalist' `temppred`n'
306         local n = `n' + 1
307     }
308     qui clear
309     qui set obs 0
310     qui append using `datalist'
311     qui tempfile timedata
312     qui save `timedata'
313     restore
314
315     qui merge 1:1 _n using `timedata', nogen
316
317     preserve
318     qui predict rbmi1, hazard time1var(time1) time2var(time2) at(dbmi2 0 dbmi3 0 dbmi4 0 dbmi5 0) ci
319     qui predict rbmi2, hazard time1var(time1) time2var(time2) at(dbmi2 1 dbmi3 0 dbmi4 0 dbmi5 0) ci
320     qui predict rbmi3, hazard time1var(time1) time2var(time2) at(dbmi2 0 dbmi3 1 dbmi4 0 dbmi5 0) ci
321     qui predict rbmi4, hazard time1var(time1) time2var(time2) at(dbmi2 0 dbmi3 0 dbmi4 1 dbmi5 0) ci
322     qui predict rbmi5, hazard time1var(time1) time2var(time2) at(dbmi2 0 dbmi3 0 dbmi4 0 dbmi5 1) ci
323     qui gen agedd_p = time2 - time1
324     qui renames time1 time2 \ diabdur_p ageatt_p
325     qui order agedd_p, before(diabdur_p)
326     qui keep agedd_p-ageatt_p rbmi*
327     qui cap drop if time1 == .
328     qui gen gender = "`nm'"
329     qui gen type = "bmirel"
330     qui tempfile rates_`nm'_bmirel
331     qui save `rates_`nm'_bmirel'
332     restore
333
334     qui drop time1 time2
335     qui egen float time2 = seq() in 1/6, from(5) to(10) block(1)

```

```

336     qui replace time2 = time2*10
337     qui gen time1_5 = 5 if time2 !=.
338     qui gen time1_20 = 20 if time2 !=.
339     qui predictnl lnhr_bmi1 = ln(predict(hazard time1var(time1_20) time2var(time2) at(dbmi2 0 dbmi3 0 dbmi4 0 dbmi5 0))) /*
340     */ - ln(predict(hazard time1var(time1_5) time2var(time2) at(dbmi2 0 dbmi3 0 dbmi4 0 dbmi5 0))), ci(lnhr_bmi1_lci lnhr_bmi1_uci)
341     qui predictnl lnhr_bmi2 = ln(predict(hazard time1var(time1_20) time2var(time2) at(dbmi2 1 dbmi3 0 dbmi4 0 dbmi5 0))) /*
342     */ - ln(predict(hazard time1var(time1_5) time2var(time2) at(dbmi2 1 dbmi3 0 dbmi4 0 dbmi5 0))), ci(lnhr_bmi2_lci lnhr_bmi2_uci)
343     qui predictnl lnhr_bmi3 = ln(predict(hazard time1var(time1_20) time2var(time2) at(dbmi2 0 dbmi3 1 dbmi4 0 dbmi5 0))) /*
344     */ - ln(predict(hazard time1var(time1_5) time2var(time2) at(dbmi2 0 dbmi3 1 dbmi4 0 dbmi5 0))), ci(lnhr_bmi3_lci lnhr_bmi3_uci)
345     qui predictnl lnhr_bmi4 = ln(predict(hazard time1var(time1_20) time2var(time2) at(dbmi2 0 dbmi3 0 dbmi4 1 dbmi5 0))) /*
346     */ - ln(predict(hazard time1var(time1_5) time2var(time2) at(dbmi2 0 dbmi3 0 dbmi4 1 dbmi5 0))), ci(lnhr_bmi4_lci lnhr_bmi4_uci)
347     qui predictnl lnhr_bmi5 = ln(predict(hazard time1var(time1_20) time2var(time2) at(dbmi2 0 dbmi3 0 dbmi4 0 dbmi5 1))) /*
348     */ - ln(predict(hazard time1var(time1_5) time2var(time2) at(dbmi2 0 dbmi3 0 dbmi4 0 dbmi5 1))), ci(lnhr_bmi5_lci lnhr_bmi5_uci)
349     qui keep time2 time1_5 time1_20 lnhr*
350     qui cap drop if time2 == .
351     qui gen gender = "`nm'"
352     qui gen type = "bmirel"
353     qui tempfile ratesratio_`nm'_bmirel
354     qui save `ratesratio_`nm'_bmirel'
355 }
356
357 clear
358 foreach nm in Men Women {
359     append using `rates_`nm'_bmirel'
360 }
361 foreach var of varlist rbmi* {
362     replace `var' = `var'*1000
363 }
364 foreach var of varlist agedd_p diabdur_p ageatt_p {
365     tostring `var', replace force format(%5.1f)
366 }
367 foreach var of varlist agedd_p diabdur_p ageatt_p {
368     destring `var', replace
369 }
370 cap drop if rbmi1 == .
371 duplicates drop
372 cap save "...", replace
373
374 clear
375 foreach nm in Men Women {
376     append using `ratesratio_`nm'_bmirel'
377 }
378 cap save "...", replace

```