

```

1  ---
2  ---
3  ---
4  ---
5  ---
6  ---
7  ---
8  ---
9  ---
10 ---
11 ---
12 ---
13 ---
14 ---
15 return require"lib".settings[[
16
17 brknbad: explore the world better, explore the world for good.
18 (c) 2022, Tim Menzies
19
20
21
22
23
24
25
26
27 USAGE:
28 ./bnb [OPTIONS]
29
30 OPTIONS:
31 -bins -b max. number of bins = 16
32 -best -B best set = .5
33 -cohen -c cohen = .35
34 -far -F how far to go for far = .9
35 -goal -g goal = recurrence-events
36 -K -K manage low class counts = 1
37 -leaves -l number of items in leaves = .5
38 -M -M manage low evidence counts = 2
39 -p -p coefficient on distance = 2
40 -rule -r rule for assessing bins; = optimize
41 one of: (optimize,monitor,tabu)
42 -rest -R rest is -Rbest = 3
43 -some -s sample size for distances = 512
44 -seed -S seed = 10019
45 -beam -T how many things to call top = 10
46 -wait -w wait = 10
47
48 OPTIONS (other):
49 -dump -d dump stack on error then quit = false
50 -file -f file name = ../etc/data/breastcancer.csv
51 -help -h show help = false
52 -todo -t start up action = nothing
53 ]]
54 ---
55 ---
56 ---
57 ---
58 ---
59 -- Copyright (c) 2022 Tim Menzies
60 -- All rights reserved.
61
62 -- Redistribution and use in source and binary forms, with or without
63 -- modification, are permitted provided that the following conditions are met:
64
65 -- 1. Redistributions of source code must retain the above copyright notice, thi
66 -- s
67 -- list of conditions and the following disclaimer.
68
69 -- 2. Redistributions in binary form must reproduce the above copyright notice,
70 -- this list of conditions and the following disclaimer in the documentation
71 -- and/or other materials provided with the distribution.
72
73 -- THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS"
74 -- AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE
75 -- IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE AR
76 -- E
77 -- DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE
78 -- FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL
79 -- DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR
80 -- SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER
81 -- CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY,
82 -- OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE
83 -- OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.
84
85 local b4={}; for k,_ in pairs(_ENV) do b4[k]=k end
86 local the, lib, go = require"the", require"lib", require"go"
87 lib.main(the, go, b4)
88
89
90
91
92
93
94
95
96
97
98
99
100
101 local ako={
102
103 ako.num = function(x) return x:find("[A-Z]" end
104 ako.goal = function(x) return x:find("[+]" end
105 ako.klass = function(x) return x:find["$" end
106 ako.ignore = function(x) return x:find["$" end
107 ako.weight = function(x) return x:find["$" and -1 or 1 end
108 ako.xnum = function(x) return ako.num(x) and not ako.goal(x) end
109
110 return ako
111

```

```

230 ---
231 ---
232 ---
233 ---
234 ---
235 local _, the, SYM = require"lib", require"the", require"sym"
236 local fmt, per, upx, push, sort = _fmt, _per, _upx, _push, _sort
237 local ent, o, oo = _ent, _o, _oo
238 local class, OBJ = _class, _OBJ
239
240 local BIN=class("BIN",OBJ)
241 function BIN:new(at,name, lo,hi,ys)
242   self.at, self.name = at or 0, name or ""
243   self.lo, self.hi, self.ys = lo, hi or lo, ys or SYM() end
244
245 function BIN:__tostring()
246   local x,lo,hi,big = self.name, self.lo, self.hi, math.huge
247   if lo == hi then return fmt("%s==%s",x, lo)
248   elseif hi == big then return fmt("%s>=%s",x, lo)
249   elseif lo == -big then return fmt("%s<=%s",x, hi)
250   else return fmt("%s<=%s<%s",lo,x,hi) end end
251
252 function BIN:select(row)
253   local x, lo, hi = row[self.at], self.lo, self.hi
254   return x=="?" or lo == hi and lo == x or lo <= x and x < hi end
255
256 function BIN:add(x,y)
257   if x<self.lo then self.lo = x end
258   if x>self.hi then self.hi = x end
259   self.ys:add(y) end
260
261
262 function BIN.mergeSameDivs(b4,after)
263   local merged = b4.ys:merged(after.ys)
264   if merged then
265     return BIN(b4.at, b4.name, b4.lo, after.hi, merged) end end
266
267 function BIN.mergeNext(b4,after)
268   if b4.hi == after.lo and b4.lo ~= b4.hi then
269     return BIN(b4.at, b4.name, b4.lo, after.hi, b4.ys:merge(after.ys)) end end
270
271 return BIN
272 ---
273 ---
274 ---
275 ---
276
277 local R = require
278 local _, the, COLS, BIN, NUM = R"lib", R"the", R"cols", R"bin", R"num"
279 local O,o,oo,downl,map,push,sort,powerset = _o,_oo,_downl,_map,_push,_sort,_
280 powerset
281 local slice,merge,slots,fmt = _slice, _merge,_slots,_fmt
282 local class,OBJ = _class, _OBJ
283
284 local RULE = class("RULE",OBJ)
285
286 function RULE.best(bins,h)
287   local function score1(b1,b2) return RULE({b1},h).score > RULE({b2},h).score e
288   nd
289   return slice(sort(bins, score1), 1, the.beam) end
290
291 function RULE.fromBins(bins,h,bests,rests, n,out,rule,sizes,scores)
292   out={}
293   sizes=NUM()
294   scores=NUM()
295   for _,some in pairs(powerset(RULE.best(bins,h))) do
296     if #some>0 then
297       rule = RULE(some,h)
298       sizes:add(#some)
299       scores:add(rule.score)
300       push(out, {size=#some,score=rule.score,rule=rule}) end end
301   local function order(one)
302     return ((0 - sizes:norm(one.size))^2 + (1 - scores:norm(one.score))^2)^.5 end
303   d
304   local n = 0
305   for _,three in pairs(sort(out, function(a,b) return order(a) < order(b) end))
306   do
307     local selected1= three.rule:selects(bests)
308     local cover1 = 100*#selected1/#bests//1
309     local selected2= three.rule:selects(rests)
310     local cover2 = 100*#selected2/#rests//1
311     if cover1 < 100 or cover2 < 100 then
312       print(fmt("%5.3f%4u%4u%",three.score, cover1, cover2, three.rule))
313       n=n+1
314       if n > the.beam then return end end end
315   return out end
316
317 function RULE:new(bins,h, t)
318   self.seen={}
319   self.bins = {}
320   for _,bin in pairs(bins) do
321     self.bins[bin.at] = self.bins[bin.at] or {}
322     push(self.bins[bin.at], bin) end
323   for _,one in pairs(self.bins) do sort(one, function(a,b) return a.lo < b.lo en
324   d) end
325   self.score = self:scored(h)
326   end
327 function RULE:__tostring() return self:show(self.bins) end --return self:show(s
328 elf.bins) end
329
330 function RULE:like(klass,h) -- h={"true"=100, "false"=40} n=100+40
331 local n=0; for _,v in pairs(h) do n = n + v end
332 local fs = {}
333 for at,bins in pairs(self.bins) do
334   fs[at] = 0
335   for _,bin in pairs(bins) do
336     fs[at] = fs[at] + (bin.ys.has[klass] or 0) end end
337   self.seen[klass] = fs
338   local prior = (h[klass] or 0) + the.K) / (n + the.K * 2)
339   local out = math.log(prior)
340   for at,v in pairs(fs) do
341     local inc = (v+the.M*prior)/(h[klass]+the.M)
342     out=out + math.log( inc)
343     end
344   return out end
345
346 RULE.bias = {}
347 local bias = RULE.bias
348 function bias:optimize(b,r) return b+r==0 and 0 or b^2/(b+r) end
349 function bias:monitor( b,r) return b+r==0 and 0 or r^2/(b+r) end
350 function bias:tabu( b,r) return b+r==0 and 0 or 1/(b+r) end
351
352 function RULE:scored(h)
353   return self.bias[the.rule] (self:like("left",h), self:like("right",h)) end
354
355 function RULE:selects(rows)
356   return map(rows, function(row) if self:select(row) then return row end end) end
357 d
358
359 function RULE:select(row)
360   local function ors(bins)
361     for _,bin in pairs(bins) do if bin:select(row) then return true end end
362     return false end
363   for at,bins in pairs(self.bins) do if not ors(bins) then return false end end
364   return true end
365
366
367
368
369 function RULE:show(ands)
370   local cat, order, sortor, sortand
371   cat = function(t,sep) return table.concat(t,sep) end
372   sortand= function(t) return map(slots(t),function(k) return t[k] end) end
373   sortor = function(a,b) return a.lo < b.lo end
374   return cat(map(sortand(ands),
375     function(and1)
376       return "(..cat(map(sort(and1,sortor),
377         function(or1) return tostring(or1) end), "or ")..") end),"and
378 ")
379 end
380
381 -- print has to wipe out fullranges and print selected items
382 --sort(bins,order)
383 -- ors= function(bins)
384 --   return cat(map(merge(sort(bins,order),BIN.mergeNext)), " or ") end
385 -- return cat(map(bins, ors), " and ") end
386
387 return RULE
388

```

```

377 ---
378 ---
379 ---
380 ---
381 ---
382 local ako, _ = require"ako", require"lib"
383 local class, OBJ = _class, _OBJ
384 local o, oo = _o, _oo
385
386 local COL = class("COL", OBJ)
387 function COL:new(at, name)
388   self.at, self.name = at or 0, name or ""
389   self.n = 0
390   self.ignorep = ako.ignore(self.name)
391   self.indep = not ako.goal(self.name)
392   self.w = self.name:find"-" and -1 or 1 end
393
394 function COL:adds(t)
395   for _, x in pairs(t) do self:add(x) end; return self end
396
397 function COL:add(x, inc)
398   if x == "?" then
399     inc = inc or 1
400     self.n = self.n + inc
401     self:add1(x, inc) end
402   return x end
403
404 function COL:dist(x, y)
405   return x=="?" and y=="?" and 1 or self:dist1(x, y) end
406
407 function COL:merged(other, out)
408   out = self:merge(other)
409   if out:div()*0.95 <= (self.n*self:div() + other.n*other:div())/out.n then
410     return out end end
411
412 return COL
413 ---
414 ---
415 ---
416 ---
417 ---
418 local _, ako, COL = require"lib", require"ako", require"COL"
419 local map, slots, class, ent = _map, _slots, _class, _ent
420
421 local SYM = class("SYM", COL)
422 function SYM:new(at, name)
423   self:super(at, name)
424   self.has, self.most, self.mode = {}, 0, nil end
425
426 function SYM:add1(x, inc)
427   self.has[x] = inc + (self.has[x] or 0)
428   if self.has[x] > self.most then
429     self.mode, self.most = x, self.has[x] end end
430
431 function SYM:mid() return self.mode end
432 function SYM:div() return ent(self.has, self.n) end
433 function SYM:same(x, y) return x==y end
434 function SYM:dist1(x, y) return self:same(x, y) and 0 or 1 end
435
436 function SYM:like1(x, prior)
437   return ((i.has[x] or 0) + the.M*prior)/(self.n + the.M) end
438
439 function SYM:merge(other, out)
440   out = SYM(self.at, self.name)
441   for x, n in pairs(self.has) do out:add(x, n) end
442   for x, n in pairs(other.has) do out:add(x, n) end
443   return out end
444
445 function SYM:bins(other, BIN)
446   local out = {}
447   local function known(x) out[x] = out[x] or BIN(self.at, self.name, x, x) end
448   for x, n in pairs(self.has) do known(x); out[x].ys:add("left", n) end
449   for x, n in pairs(other.has) do known(x); out[x].ys:add("right", n) end
450   return map(slots(out), function(k) return out[k] end) end
451
452 return SYM
453
454 ---
455 ---
456 ---
457 ---
458 ---
459 local _, the, COL = require"lib", require"the", require"col"
460 local class, merge, per, push, sort, upx = _class, _merge, _per, _push, _sort, _upx
461 local sd = _sd
462 local norm, oo = _norm, _oo
463
464 local NUM = class("NUM", COL)
465 function NUM:new(at, name)
466   self:super(at, name)
467   self.has, self.ok = {}, false
468   self.lo, self.hi = math.huge, -math.huge end
469
470 local r=math.random
471 function NUM:add1(x, inc, pos)
472   for i=1, inc do
473     self.lo = math.min(x, self.lo)
474     self.hi = math.max(x, self.hi)
475     if self.has < the.some then pos = 1 + #self.has
476       elseif r() < the.some/self.n then pos = 1 + ((r()*#self.has)//1) end
477     if pos then
478       self.ok = false
479       self.has[pos] = x end end end
480
481 function NUM:div(a) a=self:all(); return (per(a,.9) - per(a,.1))/2.56 end
482 function NUM:mid() return per(self:all(), .5) end
483 function NUM:same(x, y) return math.abs(x - y) <= the.cohen * self:div() end
484
485 function NUM:norm(x) return norm(self.lo, self.hi, x) end
486
487 function NUM:dist1(x, y)
488   if x=="?" then y = self:norm(y); x=y<.5 and 1 or 0
489   elseif y=="?" then x = self:norm(x); y=x<.5 and 1 or 0
490   else
491     x, y = self:norm(x); self:norm(y) end
492   return math.abs(x-y) end
493
494 function NUM:like1(i, x)
495   local sd= self:div()
496   if x < self.mu - 4*sd then return 0 end
497   if x > self.mu + 4*sd then return 0 end
498   local denom = (math.pi*2*sd^2)^.5
499   local nom = math.exp(1)^(-(x-self.mu)^2/(2*sd^2+1E-32))
500   return nom/(denom + 1E-32) end
501
502 function NUM:merge(other, out)
503   out = NUM(self.at, self.name)
504   for _, x in self(self.has) do out:add(x) end
505   for _, x in self(other.has) do out:add(x) end
506   return out end
507
508 function NUM:all()
509   if not self.ok then table.sort(self.has) end
510   self.ok=true
511   return self.has end
512
513 function NUM:bins(other, BIN)
514   local tmp, out = {}, {}
515   for _, x in pairs(self.has) do push(tmp, {x=x, y="left"}) end
516   for _, x in pairs(other.has) do push(tmp, {x=x, y="right"}) end
517   tmp = sort(tmp, upx) -- ascending on x
518   local now = push(out, BIN(self.at, self.name, tmp[1].x))
519   local epsilon = sd(tmp, function(z) return z.x end) * the.cohen
520   local minSize = (#tmp)^the.leaves
521   for j, xy in pairs(tmp) do
522     if j > minSize and j + minSize < #tmp then -- leave enough for other bins
523       if now.ys.n > minSize then -- enough in this bins
524         if xy.x ~= tmp[j+1].x then -- there is a break in the data
525           if now.hi - now.lo > epsilon then -- "now" not trivially small
526             now = push(out, BIN(self.at, self.name, now.hi)) end end end
527         now:add(xy.x, xy.y) end
528         out[j].lo = -math.huge
529         out[j].hi = math.huge
530         return merge(out, BIN.mergeSameDivs) end
531       end
532     end
533   end
534   return NUM
535

```

```

530 ---
531 ---
532 ---
533 ---
534 ---
535 local R=require
536 local _, ako, SYM, NUM = R"lib", R"ako", R"sym", R"num"
537 local class, OBJ, push = _class, _OBJ, _push
538
539 local COLS = class("COLS",OBJ)
540 function COLS:new(names)
541   self.names, self.class = names, nil
542   self.all, self.x, self.y = {}, {}, {}
543   for at,name in pairs(names) do
544     local now = push(self.all, (ako.num(name) and NUM or SYM) (at,name))
545     if not ako.ignore(name) then
546       if ako.klass(name) then self.klass=now end
547       push(now.indep and self.x or self.y, now) end end end
548
549 function COLS:add(row)
550   for _,col in pairs(self.all) do col:add(row[col.at]) end
551   return row end
552
553 return COLS
554
555 ---
556 ---
557 ---
558 ---
559 local R = require
560 local _, the, COLS, BIN
561 local map, sort, upl, items, push, norm = _map, _sort, _upl, _items, _push, _norm
562 local items, slice, o, oo, sort, many = _items, _slice, _o, _oo, _sort, _many
563 local class, OBJ
564
565 local EGS = class("EGS",OBJ)
566 function EGS:new() self.rows, self.cols = {}, nil end
567
568 function EGS:adds(y) for x in items(y) do self:add(x) end; return self end
569
570 function EGS:add(row)
571   if not self.cols then self.cols = COLS(row)
572   else push(self.rows, self.cols:add(row)) end end
573
574 function EGS:mid(cols)
575   return map(cols or self.cols.y, function(col) return col:mid() end) end
576
577 function EGS:div(cols)
578   return map(cols or self.cols.y, function(col) return col:div() end) end
579
580 function EGS:clone(rows)
581   local out = EGS()
582   out:add(self.cols.names)
583   for _,row in pairs(rows or {}) do out:add(row) end
584   return out end
585
586 function EGS:dist(row1,row2)
587   local d, n = 0, 0
588   for _,col in pairs(self.cols.x) do
589     n = n + 1
590     d = d + col:dist(row1[col.at], row2[col.at])^the.p end
591   return (d/n) ^ (1/the.p) end
592
593 function EGS:better(row1,row2)
594   local s1, s2, n, e = 0, 0, #self.cols.y, math.exp(1)
595   for _,col in pairs(self.cols.y) do
596     local a = norm(col.lo, col.hi, row1[col.at] )
597     local b = norm(col.lo, col.hi, row2[col.at] )
598     s1 = s1 - e^(col.w * (a - b) / n)
599     s2 = s2 - e^(col.w * (b - a) / n) end
600   return s1 / n < s2 / n end
601
602 function EGS:bins(other)
603   local out = {}
604   for n,left in pairs(self.cols.x) do
605     local right = other.cols.x[n]
606     local tmp = left:bins(right,BIN)
607     if #tmp > 1 then for _,bin in pairs(tmp) do push(out,bin) end end end
608   return out end
609
610 function EGS:bestRest()
611   self.rows = sort(self.rows, function(a,b) return self:better(a,b) end)
612   local n = (#self.rows)^the.best
613   return slice(self.rows, 1, n), -- top n things
614         many( self.rows, n*the.rest, n+1) end -- some sample of the rest
615
616 -- function egs.xplain(i)
617 --   best, rest = egs.bestRest(i)
618 --   return egs.contrasts(i, best,rest) end
619
620 return EGS
621

```

```

621 ---
622 ---
623 ---
624 ---
625 ---
626 --- 768
627 --- 384
628 --- 192
629 --- 96
630 --- 48 (positive)
631 --- 48 (positive)
632 --- 96
633 --- 48 (positive)
634 --- 48 (negative)
635 --- 192
636 --- 96
637 --- 48 (positive)
638 --- 48 (negative)
639 --- 96
640 --- 48 (positive)
641 --- 48 (positive)
642 --- 384
643 --- 192
644 --- 96
645 --- 48 (negative)
646 --- 48 (negative)
647 --- 96
648 --- 48 (negative)
649 --- 48 (negative)
650 --- 192
651 --- 96
652 --- 48 (negative)
653 --- 48 (negative)
654 --- 96
655 --- 48 (negative)
656 --- 48 (negative)
657 ---
658 local R = require
659 local the, egs, lib = R"the", R"egs", R"lib"
660 local per, cos, norm, o, fmt, rnds=lib.per, lib.cosine, lib.norm, lib.o, lib.fmt, lib.rnds
661 local map, any, many, sort, upl = lib.map, lib.any, lib.many, lib.sort, lib.upl
662 ---
663 local cluster={}
664 function cluster.new(top, egs1, i, lefts, rights)
665   egs1 = egs1 or top
666   i = {egs=egs1, top=top, rank=0}
667   lefts, rights, i.left, i.right, i.border, i.c = cluster.half(top, egs1.rows)
668   if #egs1.rows >= 2*(#top.rows)^the.leaves then
669     if #lefts.rows < #egs1.rows then
670       i.lefts = cluster.new(top, lefts)
671       i.rights = cluster.new(top, rights) end end
672   return i end
673 ---
674 ---
675 function cluster.show(i, pre, front)
676   pre = pre or ""
677   local front = fmt("%s", pre, #i.egs.rows)
678   if cluster.leaf(i)
679     then print(fmt("%-20s", front, o(rnds(egs.mid(i.egs, i.egs.cols.y))))
680     else print(front)
681       if i.lefts then cluster.show(i.lefts, " ".pre)
682       if i.rights then cluster.show(i.rights, " ".pre) end end end end
683 ---
684 function cluster.leaf(i) return not (i.lefts or i.rights) end
685 ---
686 ---
687 ---
688 function cluster.dist(egl, row1, row2)
689   local function sym(c, x, y) return x==y and 0 or 1 end
690   local function num(c, x, y)
691     if x=="?" then y = norm(c.lo, c.hi, y); x=y<.5 and 1 or 0
692     elseif y=="?" then x = norm(c.lo, c.hi, x); y=x<.5 and 1 or 0
693     else x, y = norm(c.lo, c.hi, x), norm(c.lo, c.hi, y) end
694     return math.abs(x-y) end
695   local function dist(c, x, y)
696     return x=="?" and y=="?" and 1 or (c.nump and num or sym)(c, x, y) end
697   local d, n = 0, #egl.cols.x
698   for key, c in pairs(egl.cols.x) do d=d+dist(c, row1[c.at], row2[c.at])^the.p end
699   d
700   return (d/n)^(1/the.p) end
701 ---
702 function cluster.neighbors(egl, r1, rows)
703   return sort(map(rows or egl.rows,
704     function(r2) return (cluster.dist(egl, r1, r2), r2) end), upl) end
705 ---
706 ---
707 ---
708 ---
709 function cluster.half(egl, rows)
710   local project, far, some, left, right, c, lefts, rights, border
711   rows = rows or egl.rows
712   far = function(r, t) return per(cluster.neighbors(egl, r, t), the.far)[2] end
713   project = function(r)
714     return {cos(cluster.dist(egl, left, r),
715       cluster.dist(egl, right, r),
716       c),
717     r} end
718   some = many(rows, the.some)
719   left = far(any(some), some)
720   right = far(left, some)
721   c = cluster.dist(egl, left, right)
722   lefts, rights = egs.clone(egl), egs.clone(egl)
723   for n, projection in pairs(sort(map(rows, project), upl)) do
724     if n==#rows//2 then border = projection[1] end
725     egs.add(n <= #rows//2 and lefts or rights, projection[2]) end
726   return lefts, rights, left, right, border, c end
727 ---
728 return cluster
729 ---

```

```

729 ---
730 ---
731 ---
732 ---
733 ---
734 local _, the = require"lib", require"the"
735 local fmt, inc, slots = _fmt, _inc, _slots
736 local class, OBJ = _class, _OBJ
737 ---
738 local ABCD = class("ABCD", OBJ)
739 ---
740 function ABCD:new(data, rx)
741   self.data, self.rx = data or "", rx or ""
742   self.yes, self.no = 0, 0
743   self.known, self.a, self.b, self.c, self.d = {}, {}, {}, {} end
744 ---
745 function ABCD:exists(x, new)
746   new = not self.known[x]
747   inc(self.known, x)
748   if new then
749     self.a[x]=self.yes + self.no; self.b[x]=0; self.c[x]=0; self.d[x]=0 end end
750 ---
751 function ABCD:report(p, out, a, b, c, d, pd, pf, pn, f, acc, g, prec)
752   p = function(z) return math.floor(100*z + 0.5) end
753   out = {}
754   for x, xx in pairs(self.known) do
755     pd, pf, pn, prec, g, f, acc = 0, 0, 0, 0, 0, 0
756     a = (self.a[x] or 0); b = (self.b[x] or 0);
757     c = (self.c[x] or 0); d = (self.d[x] or 0);
758     if b+d > 0 then pd = d / (b+d) end
759     if a+c > 0 then pf = c / (a+c) end
760     if a+c > 0 then pn = (b+d) / (a+c) end
761     if c+d > 0 then prec = d / (c+d) end
762     if 1-pf+pd > 0 then g=2*(1-pf) * pd / (1-pf+pd) end
763     if prec+pd > 0 then f=2*prec*pd / (prec + pd) end
764     if self.yes + self.no > 0 then
765       acc = self.yes / (self.yes + self.no) end
766     out[x] = {data=self.data, rx=self.rx, num=self.yes+self.no,
767       a=a, b=b, c=c, d=d, acc=p(acc),
768       prec=p(prec), pd=p(pd), pf=p(pf), f=p(f), g=p(g), class=x} end
769   return out end
770 ---
771 function ABCD:pretty(t)
772   print""
773   local s1 = "%10s| %10s| %4s| %4s| %4s| %4s"
774   local s2 = "| %3s| %3s| %4s| %3s| %3s|"
775   local d, s = "", (s1 .. s2)
776   print(fmt(s, "db", "x", "a", "b", "c", "d", "acc", "pd", "pf", "prec", "f", "g"))
777   print(fmt(s, d, d, d, d, d, d, d, d, d, d, d, d, d, d, d))
778   for key, x in pairs(slots(t)) do
779     local u = t[x]
780     print(fmt(s, "%s", u.data, u.rx, u.a, u.b, u.c, u.d,
781       u.acc, u.pd, u.pf, u.prec, u.f, u.g, x)) end end
782 ---
783 function ABCD:adds(gotwants, show)
784   for key, one in pairs(gotwants) do
785     self:exists(one.want)
786     self:exists(one.got)
787     if one.want == one.got then self.yes=self.yes+1 else self.no=self.no+1 end
788     for x, xx in pairs(self.known) do
789       if one.want == x
790         then inc(one.want == one.got and self.d or self.b, x)
791         else inc(one.got == x and self.c or self.a, x) end end end
792   return show and self:pretty(self:report()) or self:report() end
793 ---
794 return ABCD
795 ---

```

```

795 ---
796 ---
797 ---
798 ---
799 ---
800 local lib={}
801 ---
802 ---
803 ---
804 local r = math.random
805 function lib.normal(mu,sd)
806     mu, sd = (mu or 0), (sd or 1)
807     return mu + sd*math.sqrt(-2*math.log(r()))*math.cos(6.2831853*r()) end
808 ---
809 function lib.per(t,p) return t[ ((p or .5)*#t) // 1 ] end
810 ---
811 function lib.norm(lo,hi,x) return math.abs(hi-lo)<1E-9 and 0 or (x-lo)/(hi-lo) end
812 ---
813 function lib.ent(t, n)
814     if not n then n=0; for _,v in pairs(t) do n=n+v end end
815     local e=0; for _,v in pairs(t) do e=e-v/n*math.log(v/n,2) end
816     return e,n end
817 ---
818 function lib.sd(sorted, f)
819     f=f or function(x) return x end
820     local denom = 2.564 -- 2*(1.2 + 0.1*(0.9-0.88493)/(0.9032-0.88493))
821     return f(lib.per(sorted, .9)) - f(lib.per(sorted, 1))/denom end
822 ---
823 function lib.cosine(a,b,c)
824     return math.max(0,math.min(1, (a^2+c^2-b^2)/(2*c+1E-32))) end
825 ---
826 ---
827 ---
828 ---
829 function lib.ish(x,y,z) return math.abs(x-y) <= (z or 0.001) end
830 ---
831 ---
832 ---
833 ---
834 ---
835 function lib.inc(f,a,n) f=f or {}; f[a]=(f[a] or 0) + (n or 1) return f end
836 d
837 function lib.inc2(f,a,b,n) f=f or {}; f[a]=lib.inc(f[a] or {},b,n); return f end
838 d
839 function lib.inc3(f,a,b,c,n) f=f or {}; f[a]=lib.inc2(f[a] or {},b,c,n); return f end
840 d
841 ---
842 ---
843 ---
844 ---
845 ---
846 lib.unpack = table.unpack
847 ---
848 function lib.push(t,x) t[1 + #t] = x; return x end
849 ---
850 function lib.powerset(s)
851     local function fun(s)
852         local t = {}
853         for i = 1, #s do
854             for j = 1, #t do
855                 t[#t+1] = {s[i], lib.unpack(t[j])} end end
856             return t end
857     return lib.sort(fun(s), function(a,b) return #a < #b end) end
858 ---
859 function lib.merge(b4, merge)
860     local j,n,tmp = 1,#b4,{}
861     while j<=n do
862         local a, b = b4[j], b4[j+1]
863         if b then
864             local c = merge(a, b) -- returns nil if merge fails
865             if c then
866                 a,j = c,j+1 end end
867             tmp[#tmp+1] = a
868             j = j+1 end
869     return #tmp==#b4 and tmp or lib.merge(tmp,merge) end
870 ---
871 ---
872 ---
873 ---
874 ---
875 ---
876 function lib.map(t, f, u)
877     u={} for k,v in pairs(t) do u[1+#u]=f(v) end; return u end
878 function lib.collect(t,f,u)
879     u={} for k,v in pairs(t) do u[k]=f(k,v) end; return u end
880 function lib.copy(t, u)
881     if type(t) ~= "table" then return t end
882     u={} for k,v in pairs(t) do u[lib.copy(k)] = lib.copy(v) end; return u end
883 ---
884 ---
885 ---
886 ---
887 ---
888 function lib.sort(t,f) table.sort(t,f); return t end
889 ---
890 function lib.upx(a,b) return a.x < b.x end
891 function lib.upl(a,b) return a[1] < b[1] end
892 function lib.downl(a,b) return a[1] > b[1] end
893 ---
894 function lib.slots(t, u)
895     local function public(k) return tostring(k):sub(1,1) ~= "." end
896     u={} for k,v in pairs(t) do if public(k) then u[1+#u]=k end end
897     return lib.sort(u) end
898 ---
899 ---
900 ---
901 ---
902 ---
903 function lib.settings(help)
904     local d,used = {},{}
905     help:gsub("\n ([^%s]+)(%s+)(-[^%s]+)[^\n]*%s([%s]+)",
906         -- e.g. " -bins -b max.number of bins = 16"
907         -- parses to (-)(bins) (-b) max.number of bins = (16)
908         -- i.e. (long (key)) (short) (x)
909         function(long,key,short,x)
910             assert(not used[short], "repeated short flag [".short.."]")
911             used[short]=short
912             for n,flag in ipairs(arg) do
913                 if flag==short or flag==long then
914                     x = x=="false" and true or x=="true" and "false" or arg[n+1] end end
915             d[key] = lib.coerce(x) end)
916     if d.help then os.exit(print(help)) end
917     return d end
918 ---
919 lib.go = {_fails=0}
920 lib.no = {}
921 function lib.ok(test,msg)
922     print("", test and "PASS" or "FAIL", msg or "")
923     if not test then
924         lib.go._fails= lib.go._fails+1
925         if the and the.dump then assert(test,msg) end end end
926 ---
927 ---
928 ---
929 ---
930 ---
931 ---
932 ---
933 ---
934 ---
935 ---
936 ---
937 ---
938 ---
939 ---
940 ---
941 ---
942 ---
943 ---
944 ---
945 ---
946 ---
947 ---
948 ---
949 ---
950 ---
951 ---
952 ---
953 ---
954 ---
955 ---
956 ---
957 ---
958 ---
959 ---
960 ---
961 ---
962 ---
963 ---
964 ---
965 ---
966 ---
967 ---
968 ---
969 ---
970 ---
971 ---
972 ---
973 ---
974 ---
975 ---
976 ---
977 ---
978 ---
979 ---
980 ---
981 ---
982 ---
983 ---
984 ---
985 ---
986 ---
987 ---
988 ---
989 ---
990 ---
991 ---
992 ---
993 ---
994 ---
995 ---
996 ---
997 ---
998 ---
999 ---

```

```

1056 ---
1057 ---
1058 ---
1059 ---
1060
1061 local R = require
1062 local _, the, ABCD = R"lib", R"the", R"ABCD"
1063 local NUM, SYM, BIN, EGS, COLS, RULE = R"num", R"sym", R"bin", R"egs", R"cols", R"rule"
1064 local per, map, dent = _, per, _, map, _, dent
1065
1066 local ish, copy, items, o, oo, powerset = _, ish, _, copy, _, items, _, o, _, oo, _, powerset
1067 local map, fmt, rnds, rnd, push = _, map, _, fmt, _, rnds, _, rnd, _, push
1068 local class, Obj = _, class, _, Obj
1069 local no, go, ok = _, no, _, go, _, ok
1070
1071 function go.class()
1072   local EMP=class("EMP",Obj)
1073   function EMP:show() return {"name", "age", "_id"} end
1074   function EMP:new(name) self._id=1; self.name=name; self.age=0 end
1075   local fred = EMP("tim")
1076   local MANAGER=class("MANAGER",EMP)
1077   local jane = MANAGER("jane")
1078   print(jane) end
1079
1080 function go.copy( t,u)
1081   t={a={b={c=10},d={e=200}}, f=300}
1082   u= copy(t)
1083   t.a.b.c = 20
1084   ok(u.a.b.c ~= 20,"copy") end
1085
1086 function go.rnd()
1087   ok("23.11" == rnds({23.11111})[1],"rnds") end
1088
1089 function go.collect()
1090   local function aux(x,y) return x*y end
1091   oo(_.collect({10,20,30},aux)) end
1092
1093 function go.items()
1094   for x in items{10,20,30} do oo(x) end
1095   local n=0
1096   for x in items(the.file) do n=n+1; if n<=5 then oo(x) end end end
1097
1098 function go.powerset()
1099   for _,x in pairs(powerset{10,20,30,40,50}) do oo(x) end end
1100
1101 function go.many( t)
1102   local o,many=_,o,_,many
1103   t[1];for j = 1,1000 do t[#t+1] = j end
1104   print(900,"+", o(many(t, 10, 900)))
1105   print(1,100, o(many(t, 10, 1, 100)))
1106   print(300,700, o(many(t, 10, 300, 700))) end
1107
1108 function go.some( n)
1109   the.some=512
1110   n=NUM()
1111   for i=1,999 do n:add( i//100) end
1112   for k,v in pairs(SYM():adds(n:all()).has) do print(k,v) end end
1113
1114 function go.ent()
1115   local n = SYM()
1116   n:add("a",9)
1117   n:add("b",7)
1118   ok(ish(n:div(), .98886), "entropy") end
1119
1120 function go.normal( n)
1121   n=NUM()
1122   for i=1,10^3 do n:add( _.normal(10,2) //1) end
1123   for n,k in pairs(SYM():adds(n:all()).has) do print(n,k) end end
1124
1125 function go.nums( n)
1126   n=NUM()
1127   for i=1,10^2 do n:add( _.normal(8,1)) end
1128   oo(rnds(n:mid(), n:div())) end
1129
1130 function go.cols()
1131   _,dent(COLS{"Name","Age","gender","Weight-"}) end
1132
1133 function go.egs( i)
1134   i= EGS():adds(the.file)
1135   ok(7 == i.cols.x[2].has["l40"], "counts")
1136   ok(286 == #i.rows,"egs") end
1137
1138 function go.clone( i,j)
1139   i= EGS():adds("../etc/data/auto93.csv")
1140   j= i:clone(i.rows)
1141   local flag = true
1142   for k,n in pairs(i.cols.y[1]:all()) do
1143     flag=flag and n==j.cols.y[1]:all()[k] end
1144   ok(flag,"clone") end
1145
1146 function go.mid( all,best,rest)
1147   all = EGS():adds("../etc/data/auto93.csv")
1148   best,rest = all:bestRest()
1149   best = all:clone(best)
1150   rest = all:clone(rest)
1151   print("all",o(all:mid()))
1152   print("best",o(best:mid()))
1153   print("rest",o(rest:mid())) end
1154
1155 function go.bins( all,best,rest,b4)
1156   all = EGS():adds("../etc/data/auto93.csv")
1157   best,rest = all:bestRest()
1158   best = all:clone(best)
1159   rest = all:clone(rest)
1160   for _,bin in pairs(best:bins(rest)) do
1161     if bin.at ~= b4 then print("") end
1162     print(bin.name, bin.at,bin.lo,bin.hi,
1163           bin.ys.has["left"] or 0,
1164           bin.ys.has["right"] or 0)
1165     b4 = bin.at end end
1166
1167 local function _rules(file, all,bests,rests,left,right,b4,bins,rules,h)
1168   all = EGS():adds(file)
1169   bests,rests = all:bestRest()
1170   left = all:clone(bests)
1171   right = all:clone(rests)
1172   h = {left=#bests, right=#rests}
1173   rules = RULE.fromBins(left:bins(right),h,bests,rests)
1174   end
1175
1176 function go.rules1() _rules("../etc/data/auto93.csv") end
1177 function go.rules2() _rules("../etc/data/china.csv") end
1178 function go.rules3() _rules("../etc/data/nasa93dem.csv") end
1179
1180
1181 local function _dist(file, i,all)
1182   local any=_.any
1183   i= EGS():adds(file)
1184   local yes=true
1185   all=NUM()
1186   for j=1,1000 do
1187     if (j % 50)==0 then io.write(".") end
1188     local a,b,c = any(i.rows), any(i.rows), any(i.rows)
1189     local aa = i:dist(a,a)
1190     local ba = i:dist(b,a)
1191     local ab = i:dist(a,b)

```