

MISC STIFF
maths

```

217 -----
218 --- CLASSES
219 ---
220 ---
221 ---
222 Num, Sym, Egs = obj"Num", obj"Sym", obj"Egs"
223 ---
224 --- create
225 ---
226 ---
227 function Sym:new(at,name)
228   return new({at=at, name=name, most=0,n=0,all={}}, Sym) end
229 ---
230 function Num:new(at,name)
231   return new({at=at, name=name, _all={}, w=(name or ""):find"-"$ and -1 or 1,
232             s=0, sd=0, mu=0, m2=0, lo=math.huge, hi=-math.huge), Num) end
233 ---
234 function Egs:new(names, i,col)
235   i = new({_all={}, cols={names=names, all={}, x={}, y={}}, Egs)
236   for at,name in pairs(names) do
237     col = push(i.cols.all, (name:find"^[A-Z]" and Num or Sym) (at,name) )
238     if not name:find"$" then
239       if name:find"$" then i.cols.class = col end
240       push(name:find"[+!]"$ and i.cols.y or i.cols.x, col) end end
241   return i end
242 ---
243 --- copy
244 ---
245 ---
246 function Sym.copy(i) return Sym(i.at, i.name) end
247 ---
248 function Num.copy(i) return Num(i.at, i.name) end
249 ---
250 function Egs.copy(i,rows, j)
251   j = Egs(i.cols.names)
252   for _,row in pairs(rows or {}) do j:add(row) end
253   return j end
254 ---
255 --- update
256 ---
257 ---
258 ---
259 function Egs.add(i,row)
260   push(i._all, row)
261   for at,col in pairs(i.cols.all) do col:add(row[col.at]) end end
262 ---
263 function Sym.add(i,x,inc)
264   if x ~= "" then
265     inc = inc or 1
266     i.n = i.n+inc
267     i.all[x] = inc + (i.all[x] or 0)
268     if i.all[x] > i.most then i.most, i.mode = i.all[x], x end end end
269 ---
270 function Sym.sub(i,x,inc)
271   if x ~= "" then
272     inc = inc or 1
273     i.n = i.n - inc
274     i.all[x] = i.all[x] - inc end end
275 ---
276 function Num.add(i,x,_, d,a)
277   if x ~= "" then
278     i.n = i.n + 1
279     d = x - i.mu
280     i.mu = i.mu + d/i.n
281     i.m2 = i.m2 + d*(x - i.mu)
282     i.sd = (i.m2<0 or i.n<2) and 0 or ((i.m2/(i.n - 1))^0.5)
283     i.lo = math.min(x, i.lo)
284     i.hi = math.max(x, i.hi)
285     a = i._all
286     if #a < the.keep then i.ok=false; push(a,x)
287     elseif r() < the.keep/i.n then i.ok=false; a[r(#a)]=x end end end
288 ---
289 function Num.sub(i,x,_, d)
290   if x ~= "" then
291     i.n = i.n - 1
292     d = x - i.mu
293     i.mu = i.mu - d/i.n
294     i.m2 = i.m2 - d*(x - i.mu)
295     i.sd = (i.m2<0 or i.n<2) and 0 or ((i.m2/(i.n - 1))^0.5) end end
296 ---
297 ---
298 ---
299 ---
300 ---
301 ---
302 ---
303 function Egs.better(i,row1,row2)
304   local s1, s2, n, a, b = 0, 0, #i.cols.y
305   for _,col in pairs(i.cols.y) do
306     a = col:norm( row1[col.at] )
307     b = col:norm( row2[col.at] )
308     s1 = s1 - 2.7183*(col.w * (a - b) / n)
309     s2 = s2 - 2.7183*(col.w * (b - a) / n) end
310   return s1 / n < s2 / n end
311 ---
312 function Egs.bettors(i,j)
313   return Egs.better(i:mid(i.cols.all), j:mid(j.cols.all)) end
314 ---
315 function Egs.mid(i,cols)
316   return map(cols or i.cols.y, function(col) return col:mid() end) end
317 ---
318 function Num.mid(i) return i.mu end
319 function Sym.mid(i) return i.mode end
320 ---
321 function Num.div(i) return i.sd end
322 function Sym.div(i, e)
323   e=0
324   for _,n in pairs(i.all) do
325     if n > 0 then e = e + n/i.n * math.log(n/i.n,2) end end
326   return -e end
327 ---
328 function Num.norm(i,x)
329   return i.hi - i.lo < 1E-32 and 0 or (x - i.lo)/(i.hi - i.lo) end

```

```

329 ---
330 function Num.all(i)
331   if not i.ok then table.sort(i._all); i.ok=true end
332   return i._all end
333 ---
334 --- cluster
335 ---
336 ---
337 function Num.dist(i,a,b)
338   if a=="?" and b=="?" then return 1 end
339   if a=="?" then b=i:norm(b); a=b<.5 and 1 or 0
340   elseif b=="?" then a=i:norm(a); b=a<.5 and 1 or 0
341   else a,b = i:norm(a), i:norm(b) end
342   return math.abs(a - b) end
343 ---
344 function Sym.dist(i,a,b)
345   return a=="?" and b=="?" and 1 or a==b and 0 or 1 end
346 ---
347 function Egs.dist(i,row1,row2, d)
348   d = sum(i.cols.x, function(c) return c:dist(row1[c.at], row2[c.at])^the.p end)
349   return (d/#i.cols.x)^(1/the.p) end
350 ---
351 function Egs.dists(i,r1,rows)
352   return sort(map(rows,function(r2) return {i:dist(r1,r2),r2} end),firsts) end
353 ---
354 function Egs.half(i, rows)
355   local project,far,some,left,right,c,lefts,rights
356   far = function(r,t) return per(i:dists(r,t), the.far)[2] end
357   project = function(r1, a,b)
358     a,b = i:dist(left,r1), i:dist(right,r1)
359     return ((a^2 + c^2 - b^2)/(2*c), r1) end
360   some = many(rows, the.some)
361   left = far(any(some), some)
362   right = far(left, some)
363   c = i:dist(left,right)
364   lefts,rights = i:copy(), i:copy()
365   for n, projection in pairs(sort(map(rows,project),firsts)) do
366     if n==#rows//2 then mid=row end
367     (n <= #rows//2 and lefts or rights):add( projection[2] ) end
368   return lefts, rights, left, right, mid, c end
369 ---
370 --- discretize
371 ---
372 ---
373 function Num.spans(i,j, cuts)
374   local xys,all = {}, Num
375   for _,n in pairs(i._all) do all:add(n); push(xys,{x=n,y="left"}) end
376   for _,n in pairs(j._all) do all:add(n); push(xys,{x=n,y="right"}) end
377   return bins(i,cuts,
378     bins1(sort(xys,first), (#xys)^the.minItems,all.sd*the.cohen,Sym,{})) end
379 ---
380 function bins1(col, old,new)
381   if #new>1 then
382     new[1].lo = -math.huge
383     new[#new].hi = math.huge
384     for _,cut in pairs(new) do cut.col= col; push(old,cut) end end end
385 ---
386 function bins1(xys, minItems, cohen, yclass, cuts, b4)
387   local lhs, rhs, b4, cut, div, xpect = yclass(), yclass(), b4 or xys[1].x
388   function xpect(i,j) return (i.n*i:div() + j.n*j:div()) / (i.n + j.n) end
389   for _,xy in pairs(xys) do rhs:add(xy.y) end
390   div = rhs:div()
391   for j,xy in pairs(xys) do
392     lhs:add(xy.y)
393     rhs:sub(xy.y)
394     if lhs.n >= minItems and rhs.n >= minItems then
395       if xy.x ~= xys[j+1].x then
396         if xy.x - xys[j+1].x >= cohen and xys[#xys].x - xy.x >= cohen then
397           if xpect(lhs,rhs) < div then
398             cut, div = j, xpect(lhs,rhs) end end end end end
399   if cut
400   then local l,r = {},{}
401     for n,xy in pairs(xys) do push(n<=cut and l or r, xy) end
402     bins1(l, minItems, cohen, yclass, cuts, b4)
403     bins1(r, minItems, cohen, yclass, cuts, xys[cut].x)
404     else push(cuts, {lo=b4, hi=xys[#xys].x, n=#xys, div=div}) end end
405 ---
406 ---
407 ---
408 ---
409 ---
410 ---
411 local xplain,xplains,selects,spanShow
412 function Egs.xplain(i,rows)
413   local stop,here,left,right,lefs0,rights0,lefs1,rights1
414   rows = rows or i._all
415   here = {all=rows}
416   stop = (#i._all)^the.minItems
417   if #rows >= 2*stop then
418     lefs0, rights0, here.left, here.right, here.mid, here.c = half(i, rows)
419     if #lefs0._all < #rows then
420       cuts = {}
421       for j,col in pairs(lefs0.col.x) do col:spans(rights0.col.x[j],cuts) end
422       lefs1,rights1 = {},{}
423       for _,row in pairs(rows) do
424         push(selects(here.selector, row) and lefs1 or rights1, row) end
425       if #lefs1 > stop then here.lefs0, lefs1 = xplain(i,lefs1) end
426       if #rights1 > stop then here.rights = xplain(i,rights1) end end end
427   return here end
428 ---
429 function xbestSpan(spans)
430   local divs,ns,n,div,stats,dist2heaven = Num(), Num()
431   function dist2heaven(s) return ((1 - n(s))^2 + (0 - div(s))^2)^.5,s) end
432   function div(s) return divs:norm( s.all:div() ) end
433   function n(s) return ns:norm( s.all.n ) end
434   for _,s in pairs(spans) do
435     add(divs, s.all:div())
436     add(ns, s.all.n) end
437   return sort(map(spans, dist2heaven), firsts)[1][2] end
438 ---
439 function selects(span,row, lo,hi,at,x)
440   lo, hi, at = span.lo, span.hi, span.col.at
441   x = row[at]
442   if x=="?" then return true end
443   if lo==hi then return x==lo else return lo <= x and x < hi end end
444 ---
445 function xplains(i,format,t,pre,how, sel,front)
446   pre, how = pre or "", how or ""
447   if t then
448     prepre or ""
449     front = fmt("%s%s%s",pre,how, #t.all, t.c and rnd(t.c) or "")
450     if t.lefs and t.rights then print(fmt("%-35s",front)) else
451       print(fmt("%-35s",front, o(rnds(mids(i,t.all),format))))
452     end
453     sel = t.selector
454     xplains(i,format,t.lefs, " | .. pre, spanShow(sel).." )
455     xplains(i,format,t.rights, " | .. pre, spanShow(sel,true) .." ) end end

```

```

455 ---
456 ---
457
458 function quintiles(ts,width,  nums,out,all,n,m)
459 width=width or 32
460 nums=Num(); for _,t in pairs(ts) do
461     for _,x in pairs(sort(t)) do add(nums,x) end end
462 all,out = nums.all, {}
463 for _,t in pairs(ts) do
464     local s, where = {}
465     where = function(n) return (width*nums:norm(n))/1 end
466     for j = 1, width do s[j]="" end
467     for j = where(per(t,.1)), where(per(t,.3)) do s[j]="-" end
468     for j = where(per(t,.7)), where(per(t,.9)) do s[j]="-" end
469     s[where(per(t,.5))] = "|"
470     push(out,{display=table.concat(s),
471         data = t,
472         pers = map({.1,.3,.5,.7,.9},
473             function(p) return rnd(per(t,p))end)}) end
474
475 return out end
476
477 function smallfx(xs,ys,      x,y,lt,gt,n)
478 lt,gt,n = 0,0,0
479 if #ys > #xs then xs,ys=ys,xs end
480 for _,x in pairs(xs) do
481     for j=1, math.min(64,#ys) do
482         y = any(ys)
483         if y<x then lt=lt+1 end
484         if y>x then gt=gt+1 end
485         n = n+1 end end
486 return math.abs(gt - lt) / n <= the.cliffs end
487
488 function bootstrap(y0,z0)
489 local x, y, z, b4, yhat, zhat, bigger
490 local function obs(a,b, c)
491     c = math.abs(a.mu - b.mu)
492     return (a.sd + b.sd) == 0 and c or c/((x.sd^2/x.n + y.sd^2/y.n)^.5) end
493 local function adds(t, num)
494     num = num or Num(); map(t, function(x) add(num,x) end); return num end
495 y,z = adds(y0), adds(z0)
496 x = adds(y0, adds(z0))
497 b4 = obs(y,z)
498 yhat = map(y._all, function(y1) return y1 - y.mu + x.mu end)
499 zhat = map(z._all, function(z1) return z1 - z.mu + x.mu end)
500 bigger = 0
501 for j=1,the.boot do
502     if obs( adds(many(yhat,#yhat)), adds(many(zhat,#zhat))) > b4
503     then bigger = bigger + 1/the.boot end end
504 return bigger >= the.conf end
505
506 --- xxx mid has to be per and
507 -- XXX implement same
508 -- XXX need tests for stats
509 function scottKnot(nums,      all,cohen)
510 local mid = function(z) return z.some:mid()
511 end
512 local function summary(i,j,      out)
513     out = copy(nums[i])
514     for k = i+1, j do out = out:merge(nums[k]) end
515     return out
516 end
517 local function div(lo,hi,rank,b4,      cut,best,l,l1,r,r1,nov)
518     best = 0
519     for j = lo,hi do
520         if j < hi then
521             l = summary(lo, j)
522             r = summary(j+1, hi)
523             now = (l.n*(mid(l) - mid(b4))^2 + r.n*(mid(r) - mid(b4))^2) / (l.n + r.n)
524             if now > best then
525                 if math.abs(mid(l) - mid(r)) >= cohen then
526                     cut, best, l1, r1 = j, now, copy(l), copy(r)
527             end end end
528             if cut and not l1:same(r1,the) then
529                 rank = div(lo,      cut, rank, l1) + 1
530                 rank = div(cut+1, hi, rank, r1)
531             else
532                 for i = lo,hi do nums[i].rank = rank end end
533                 return rank
534             end
535         end
536         table.sort(nums, function(x,y) return mid(x) < mid(y) end)
537         all = summary(1,#nums)
538         cohen = all.sd * the.cohen
539         div(1, #nums, 1, all)
540     return nums end

```

```

540 -----
541 ---
542 ---
543 ---
544 ---
545 function go.last()
546     ok( 30 == last({10,20,30}, "lasts") end
547
548 function go.per( t)
549     t={};for i=1,100 do push(t,i*1000) end
550     ok(70000 == per(t,.7), "per") end
551
552 function go.many( t)
553     t={};for i=1,100 do push(t,i) end; many(t,10) end
554
555 function go.sum( t)
556     t={};for i=1,100 do push(t,i) end; ok(5050==sum(t), "sum")end
557
558 function go.sample( m,n)
559     m,n = 10^5,Num(); for i=1,m do n:add(i) end
560     for j=.1,.9,.1 do do push(t,i*m*0.05) end end
561     print(j,per(n:all(),j),ish(per(n:all(),j),m*j,m*0.05)) end end
562
563 function go.sym( s)
564     s=Sym(); map({1,1,1,1,2,2,3}, function(x) s:add(x) end)
565     ok(ish(s:div(),1.378, 0.001), "cnt") end
566
567 function go.num( n)
568     n=Num(); map({10, 12, 23, 23, 16, 23, 21, 16}, function(x) n:add(x) end)
569     print(n:div())
570     ok(ish(n:div(),5.2373, .001), "div") end
571
572 function go.nums( num,t,b4)
573     b4,t,num={},{};Num()
574     for j=1,1000 do push(t,100*r(i)*j) end
575     for j=1,#t do
576         num:add(t[j])
577         if j%100==0 then b4[j] = fmt("%.5f",num:div()) end end
578     for j=#t,-1 do
579         if j%100==0 then ok(b4[j] == fmt("%.5f",num:div()),"div"..j) end
580         num:sub(t[j]) end end
581
582 function go.syms( t,b4,s,sym)
583     b4,t,sym, s={},{};Sym(), "I have gone to seek a great perhaps."
584     t={}; for j=1,20 do s:gsub('.',function(x) t[#t+1]=x end) end
585     for j=1,#t do
586         sym:add(t[j])
587         if j%100==0 then b4[j] = fmt("%.5f",sym:div()) end end
588     for j=#t,-1 do
589         if j%100==0 then ok(b4[j] == fmt("%.5f",sym:div()),"div"..j) end
590         sym:sub(t[j]) end
591     end
592
593 function go.loader( num)
594     for row in things(the.file) do
595         if num then num:add(row[1]) else num=Num() end end
596         ok(ish(num.mu, 5.455,0.001), "loadmu")
597         ok(ish(num.sd, 1.701,0.001), "loadsds") end
598
599 function go.egsShow( t)
600     oo(Egs{"name","Age","Weigh-"}) end
601
602 function go.egsHead( )
603     ok(Egs({"name","age","Weight!").cols.x, "Egs") end
604
605 function go.egs( eggs)
606     eggs = csv2egs(the.file)
607     ok(ish(egs.cols.x[1].mu, 5.455,0.001),"loadmu")
608     ok(ish(egs.cols.x[1].sd, 1.701,0.001),"loadsds") end
609
610 function go.dist( ds,egs,one,d1,d2,d3,r1,r2,r3)
611     eggs = csv2egs(the.file)
612     one = eggs._all[1]
613     ds={};for j=1,20 do
614         push(ds,egs:dist(any(egs._all), any(egs._all))) end
615     oo(rnds(sort(ds),"%5.3f"))
616     for j=1,10 do
617         r1,r2,r3 = any(egs._all), any(egs._all), any(egs._all)
618         d1=egs:dist(r1,r2)
619         d2=egs:dist(r2,r3)
620         d3=egs:dist(r1,r3)
621         ok(d1<= 1 and d2 <= 1 and d3 <= 1 and d1>=0 and d2>=0 and d3>=0 and
622             eggs:dist(r1,r2) == eggs:dist(r2,r1) and
623             eggs:dist(r1,r1) == 0
624             and
625             d3 <= d1+d2, "dist"..j) end end
626
627 function go.far( eggs,lefts,rights)
628     eggs = csv2egs(the.file)
629     lefts,rights = eggs:half(egs._all)
630     oo(rnds(egs:mid()))
631     oo(rnds(lefts:mid()))
632     oo(rnds(rights:mid())) end
633
634 the = settings(help)
635 go.main(the.todo, the.seed)
636 os.exit(go.fail)

```