Printed by TIm Menzies

```
function Num:add(x)
     intion Num:add(x)
if x>self.hi then self.hi = x
elseif x<self.lo then self.lo = x end
push(self.has,x); self.n=self.n+1; self.ready=false end</pre>
    - Ensure that the returned list of numbers is sorted.
unction Num:all(x)
if not self.ready then table.sort(self.has) end
self.ready = true
return self.has end
function Num:dist(a,b)
if a=="?" then b=self:norm(b); a = b>.5 and 0 or 1
elseif b=="?" then a=self:norm(a); b = a>.5 and 0 or 1
else a,b = self:norm(a), self:norm(b) end
return abs(a-b) end
-- Combine two 'num's.
function Num:merge(other, new)
new = Num()
new.at, new.txt = self.at, self.txt
for _, x in pairs(self.has) do new:add(x) end
for _, x in pairs(other.has) do new:add(x) end
return new end
-- The 'mid' is the 50th percentile.
function Num:mid() return self:per(.5) end
-- Return 'x' normalized 0..1, lo..hi.
function Num:norm(x, lo,hi)
if x==""" then return x end
lo,hi = self.lo, self.hi
return abs(hi - lo) < 1E-32 and 0 or (x - lo)/(hi - lo) end
-- Return the 'p'-th percentile number.
function Num:per(p, t)
t = self:all()
p = p*#t//1
return #t<2 and t[1] or t[p < 1 and 1 or p>#t and #t or p] end
 -- The 10th to 90th percentile range is 2.56 times the standard deviation. function Num:sed() return (self:per(.9) - self:per(.1))/ 2.56 end function Num:spread() return self:sed() end
-- Create one span (each has the row indexes of the rows)
-- where each span has at least 'tiny' items and span is more than
-- 'tirvial'ly small.
-- trivially small.

local div -- defined below

function Num:spans(sample,tiny,trivial)

local xys = (airs (sample.egs)

for ... eg = eg(self.at)

if x -= """ then push(xys, (col=col, x=x, y=eg(sample.klass.at])) end end

return div(xys, tiny, trivial, self, getmetatable(sample.klass)) end
-- Stuff for tracking 'Sym'bol Counts.
-- Stym's track symbol counts and the 'mode' (most frequent symbol).
local Sym=obj"Sym"
function Sym.new(inits,at,txt, self)
self=has(Gym,(at-at or 0, txt=txt or "", has={}, n=0, mode=nil, most=0})
for __one in pairs(inits or {}) do self:add(one) end
return self end
function Sym:dist(a,b) return a==b and 0 or 1 end
function Sym:merge(other)
     unction Sym:merge(other)
new-Sym()
new.at, new,txt = self.at, self.txt
for k,n in pairs(self.has) do new:add(k,n) end
for k,n in pairs(other.has) do new:add(k,n) end
return new end
 function Sym:mid() return self.mode end
function Sym:spread()
  return sum(self.has,
  function(n1) return -n1/self.n * math.log(n1/self.n,2) end) end
-- ## Stuff for skipping all things sent to a column local Skip=obj"Skip" function Skip.new(_,at,txt) return has(Skip,(at=at or 0, txt=txt or"", n=0)) end function Skip:add(x) self.n = self.n + 1; return x end function Skip:mid() return "?" end
```

```
-- Samples store examples. Samples know about
-- (a) lo,hi ranges on the numerics
-- and (b) what are independent 'x' or dependent 'y' columns.
local Sample = obj"Sample"
 local Sample = ob;"Sample"
function Sample.new( src,self)
self = has(Sample, names=nil, klass=nil, all={}, ys={}, xs={}, egs={}})
if src then
if type(src) =="sring" then for x in csv(src) do self:add(x) end end
if type(src) =="table" then for _,x in pairs(src) do self:add(x) end end
return self end
function Sample:better(eq1,eq2, e,n,a,b,s1,s2)
n,s1,s2,e = #self.ys, 0, 0, 2.71828
for _,num in pairs (self.ys) do
    a = num:norm(eq2[num.at])
    b = num:norm(eq2[num.at])
    s1 = s1 - e^(num.w * (a-b)/n)
    s2 = s2 - e^(num.w * (b-a)/n) end
return s1/n < s2/n end</pre>
function Sample:betters(egs)
   return sort(egs or self.egs, function(a,b) return self:better(a,b) end) end
 function Sample:clone( inits,out)
  out = Sample.new():add(self.names)
  for _,eg in pairs(inits or ()) do out:add(eg) end
  return out end
     function Sample:mid(cols)
  return lap(cols or self.ys, function(col) return col:mid() end) end
-- Return spans of the column that most reduces variance function Sample:bestSplits(tiny, trivials) local function column1(col, total, xpect, spans, total, xpect) local function xpect1(span) return span.has.n/total * span.has:spread() end spans = col:spans(self, tiny,trivials[col.at]) total = sum(spans, function(span) return span.has.n end) xpect = sum(spans, xpect1)
         xpect = sum(spans, xpect1)
return {xpect, spans}
       return first(sort(lap(self.xs, column1), firsts))[2] end
    --print(span.col.at, x, span.lo, span.hi)
if x=="?" or (span.lo <= x and x <= span.hi) then new:add(eg) end end
            --end end return node end
   -- Find which leaf best matches an example 'eg'.:w
 function Sample:where(tree,eg, max,x,default)
  if #kid.has==0 then return tree end
     if #kid.has==U then artum...

max = 0

for _kid in pairs(tree.node) do

if #kid.has > max then default.max = kid.#kid.has end

x = eg[kid.at]

if x -= "" then

if and x > kid.lo then

return self:where(kid.has.eg) end end end

return self:where(default, eg) end
  -- Discrimination --
local function merge(b4) -- merge adjacent spans if combo simpler to he parts
           ocal function merge (b4) -- merge adjacent spans if combo simpler to he local j, tmp = 0, ()
while j < $b4 do
j = j + 1
local now, after = b4[j], b4[j+1]
if after then
local simpler = mergeable(now.has, after.has)
if simpler then
now = (col=col, lo=now.lo, hi= after.hi, has=simpler)
push(tmp,now)
and
return thup==$b4 and b4 or merge(tmp) -- recurse until nothing merged
now = 1 the person that th
      local spans,span
span = {col=col,lo=xys[1].x, hi=xys[1].x, has=yklass()}
```

spans = {span}

Printed by TIm Menzies

```
Dec 13, 21 9:32
                                                                                                                                                                                                                                                                                                                                   Page 4/7
                 for j,xy in pairs(xys) do
local x, y = xy.x, xy.y

if j < ixys - tiny and -- enough items remaining after split
x -= xys[j+1].x and -- next item is different (so can split here)
span.has.n > tiny and -- span has enough items
span.hi - span.lo > trivial -- span is not trivially small
then span = push(spans, {col=col, lo=x, hi=x, has=yklass()}) -- then new span
end
span.hi = x
span.has.add(y) end
-return merge(spans) end
--return coverGaps(merge(spans)) end
```

15.lua

Printed by TIm Menzies

I5.lua Page 6/7

```
t=()
for i=1,32 do push(t,i) end
u = shuffle(copy(t))
v = shuffle(copy(t))
assert(#t == #u and u[1] ~= v[1]) end
   function it._eg.lap()
assert(3==lap({1,2},function(x) return x+1 end)[2]) end
   function it._eg.map()
  assert(3==map({1,2},function(_,x) return x+1 end)[2]) end
   function it._eg.tables()
  assert(20==sort(shuffle({{10,20},{30,40},{40,50}}),firsts)[1][2]) end
   function it._eg.csv( n,z)
     for eg in csv(it.FILE) do n=n+1; z=eg end assert(n==399 and z[#z]==50) end
   function it._eg.dump()
    shout(it) end
function it._eg.tree(    s,t,u,egl,ev
    s = Sample(it.FILE)
    t = copy(s.names)
    push(t, "Rank!")
    u = Sample.new():add(t)
    evals, ordered.rest = hints.sort(s)
    for m.eg in pairs(ordered.egs) do
        egl = copy(eg)
    push(ed.m)
                             s,t,u,eg1,evals,ordered,rest)
    push (eg1, m)
u:add(eg1) end
print(1)
u:tree() end
     - START-UP -
470 --
471 it{demos=it._eg, nervous=true}
```

Dec 13, 21 9:32

Dec 13, 21 9:32 I5.lua Page 7/7

```
Lesson plan - w1: ssytems: github. github workplaces. unit tests. doco tools.
```

Dec 13, 21 9:32 lib.lua Page 1/1

local lib={}

Printed by TIm Menzies

```
--- Call 'rogues', last thing, to find escaped locals.

lib_b4=(); for k,v in pairs(_ENV) do lib_b4[k]=k end function lib.rogues()

for k,v in pairs(_ENV) do

if not lib_b4[k] then print("?rogue: ",k,type(v)) end end end
  -- OBJECTS ---
  -- Create an instance function lib.has(mt,x) return setmetatable(x,mt) end
-- Create a clss
function lib.obj(s, o,new)
o = {_is=s, __tostring=lib.out}
o.__index = o
return setmetatable(o,{__call = function(_,...) return o.new(...) end)) end
   -- RANDOM --
 -- random integers
function lib.randi(lo,hi) return math.floor(0.5 + lib.rand(lo,hi)) end
 -- random floats
function lib.rand(lo,hi, mult,mod)
lo, hi = lo or 0, hi or 1
lib.Seed = (16807 * lib.Seed) % 2147483647
return lo + (hi-lo) * lib.Seed / 2147483647 end
   -- MATHS ---
--
lib.abs = math.abs
-- Round 'x' to 'd' decimal places.
function lib.rnd(x,d, n) n=10^(d or 0); return math.floor(x*n+0.5) / n end
-- Round list of items to 'd' decimal places.
function lib.rnds(t,d)
return lib.lap(t, function(x) return lib.rnds(x,d or 2) end) end
 -- Sum items, filtered through 'f'.
function lib.sum(t,f, out)
f= f or function(x) return x end
out=0; for _,x in pairs(t) do out = out + f(x) end; return out end
  -- FILES -----
       Return one table per line, split on commas.

motion lib.csv(file, line)
file = io.input(file)
line = io.read()
return function( t,tmp)
if line then
           ir line then
t={}
for cell in line:gsub("[Wt]"",""):gsub("#."",""):gmatch("([^.]+)") do
lib.push(t, tonumber(cell) or cell) end
line = io.read()
if #t>0 then return t end
else io.close(file) end end end
   -- PRINTING ---
  lib.fmt = string.format
lib.say = function(...) print(lib.fmt(...)) end
 -- Print as red, green, yellow, blue.
function lib.color(s,n) return lib.color(s,31) end
function lib.red(s) return lib.color(s,31) end
function lib.gleen(s) return lib.color(s,32) end
function lib.yellow(s) return lib.color(s,34) end
function lib.blue(s) return lib.color(s,36) end
 -- Printed string from a nested structure.
lib.shout = function(x) print(lib.out(x)) end
-- Generate string from a nested structures
-- (and don't print any contents more than one
     - (and don't print any contents more than once).
unction lib.out(t,rseen, u,key,value,public)
function key(k) return lib.fmt("%% %s", lib.blue(k), lib.out(t[k],seen)) end
function value(v) return lib.out(v,seen) end
if type(t) = "function" then return "(...") end
if type(t) -= "lable" then return tostring(t) end
seen = seen or {}
if seen[t] then return "..." else seen[t] = t end
u = ft>0 and lib.lap(t, value) or lib.lap(lib.keys(t), key)
return lib.red((t._is or"")..."[")..lib.cat(u,"")..lib.red(")") end
  -- TABLE ----
  -- Table to string.
lib.cat = table.concat
 -- Return a sorted table.
lib.sort = function(t,f) table.sort(t,f); return t end
lib.sort = function(t,f) table.sort(t,f); return t end
--Return first,second, last item.
lib.first = function(t) return t[1] end
lib.second = function(t) return t[2] end
lib.last = function(t) return t[#t] end
--Function for sorting pairs of items.
lib.firsts = function(a,b) return a[1] < b[1] end
-- Add to end, pull from end.
lib.pop = table.remove
lib.push = function(t,x) table.insert(t,x); return x end
  -- Random order of items in a list (sort in place).
function lib.shuffle(t, j)
for i=t,2,-1 do j=lib.randi(1,i); t[i],t[j]=t[j],t[i] end; return t end
  -- Collect values, passed through 'f'.
function lib.lap(t,f) return lib.map(t,f,1) end
function lib.lap(t,f) return lib.map(t,f,l) end

- Collect key,values, passed through 'f'.

- If 'f' returns two values, store as key,value.

- If 'f' returns one values, store as key,value.

- If 'f' returns one values, store as key,value.

- If 'f' return il then add nothing (so 'map' is also 'select').

function lib.map(t,f,one, u)

u=(); for x,y in pairs(t) do

if ne = nan x,y=f(y) else x,y=f(x,y) end

if y then u(x)=y else u[1+8u]=x end end end

return u end
  -- Shallow copy function lib.copy(t, u) u={}; for k,v in pairs(t) do u[k]=v end; return u end
  --- Return a table's keys (sorted).
function lib.keys(t,u)
      u={} k,_ in pairs(t) do if tostring(k):sub(1,1)~="_" then lib.push(u,k) end end return lib.sort(u) end
 -- Binary chop (assumes sorted lists)
function lib.bchop(t,val,lt,lo,hi, mid)
lt = lt or function(x,y) return x < y end
lo,hi = lo or 1, hi or #t
while lo <= hi do
mid = (lo+hi) // 2
if lt(t[mid],val) then lo=mid+l else hi= mid-l end end
return math.min(lo,#t) end
 return lib
```