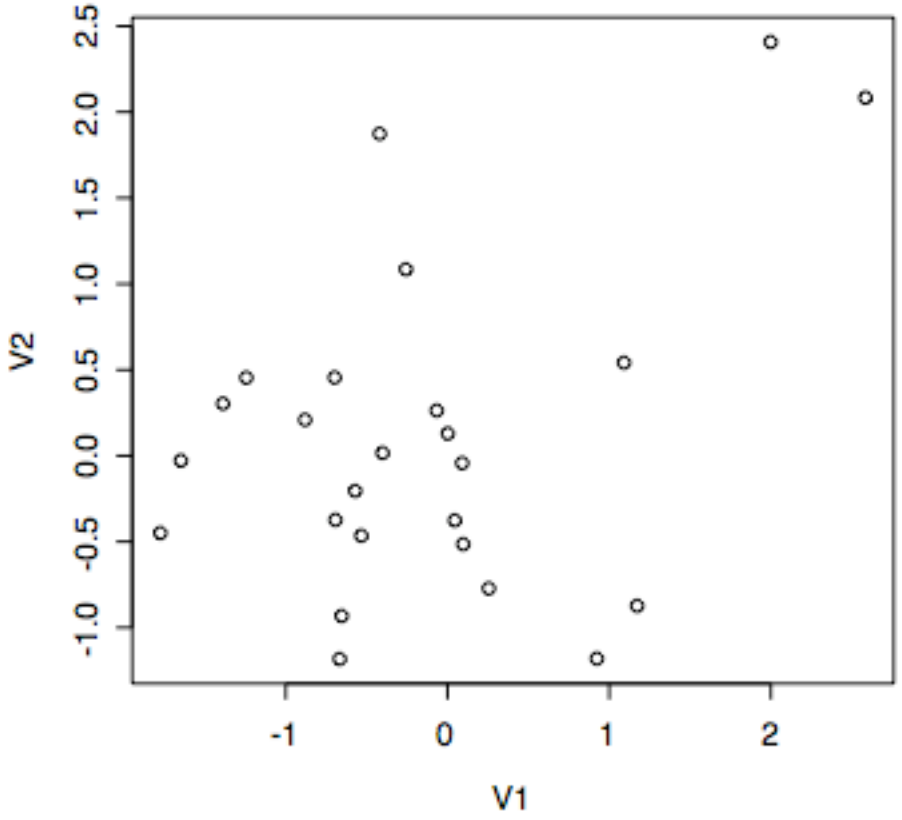
STAT 802 Applied Multivariate Analysis Sept. 14, 2005

Today: (Check notes from Sept 12 for some details re matrix algebra).

Some R demos for reviewing matrix arithmetic, statistical distance and projections.

An R program:

```
mat.ex
function (n=25,p=2,print=F,corr=.5)
{
        cov=matrix(ncol=p,nrow=p)
        for (i in 1:p){
                for (j in 1:p){
                        cov[i,j]=corr}
                        cov[i,i]=1
                        }
        data=as.data.frame(mvrnorm(n,c(rep(0,p)),cov))
        plot(data)
        cor=cor(data)
        eigen=eigen(cor)
        if (print==T){print(cov);print(data);print(cor);print(eigen)}
        invisible(list(data,eigen))
}
mat.ex(print=T)
```




```
[,1] [,2]
[1,]  1.0  0.5
[2,]  0.5  1.0
           V1          V2
1  -0.401134182  0.01673537
2  -0.066281715  0.26336694
3  -0.666406228 -1.18213983
4   1.172933559 -0.87296608
5   0.924188869 -1.18011096
6  -0.654135508 -0.93206532
7   1.998504588  2.40762006
8  -0.692424721 -0.37311506
9   1.091720134  0.54177933
10 -1.648034055 -0.02799312
11 -0.880500821  0.20944796
12 -1.775293274 -0.44941012
13  0.091212397 -0.04258021
14  2.585282192  2.08358000
15 -1.388156343  0.30383284
16 -1.243708536  0.45448712
```

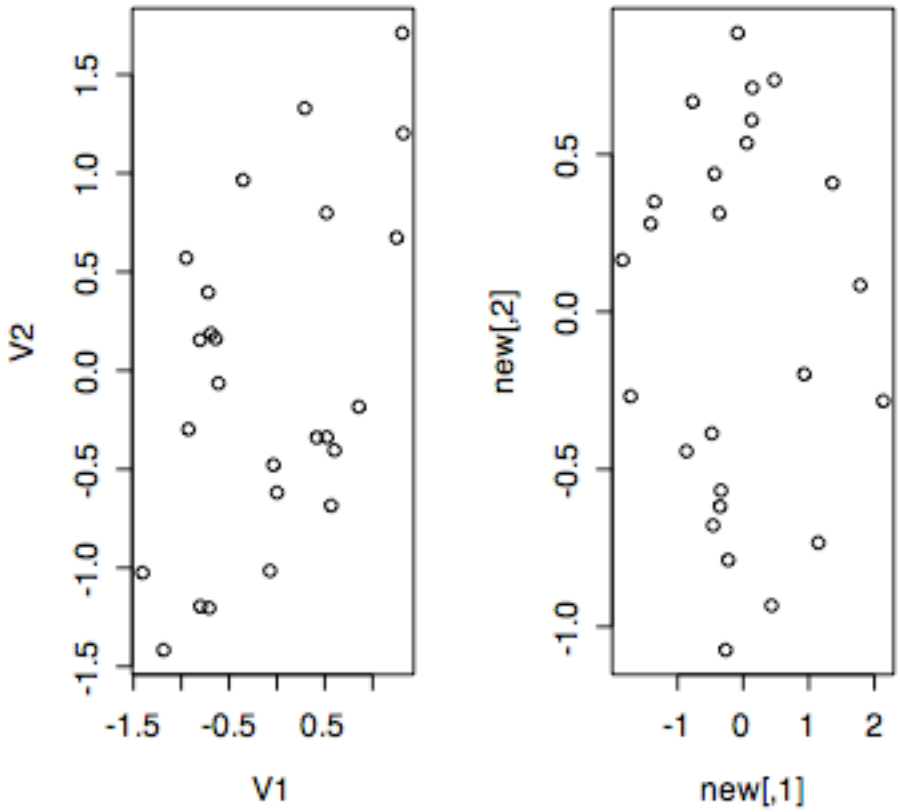
```
17  0.097436122 -0.51404878
18 -0.570465393 -0.20408117
19 -0.419638096  1.87387364
20  0.045792713 -0.37569927
21  0.255713267 -0.77215577
22 -0.532640519 -0.46494420
23  0.002334836  0.12964905
24 -0.256576938  1.08491527
25 -0.697396494  0.45577591
         V1        V2
V1 1.0000000 0.3776583
V2 0.3776583 1.0000000
$values
[1] 1.3776583 0.6223417

$vectors
         [,1]       [,2]
[1,] 0.7071068  0.7071068
[2,] 0.7071068 -0.7071068
```

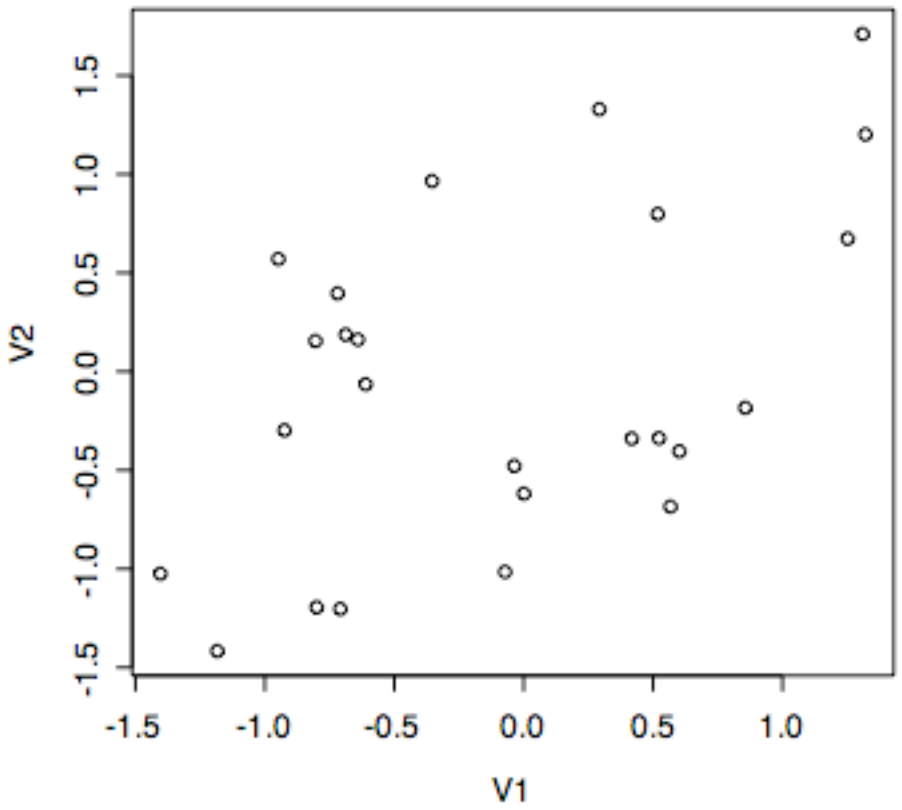
What are new coordinates (of rotated, uncorrelated, variables)?

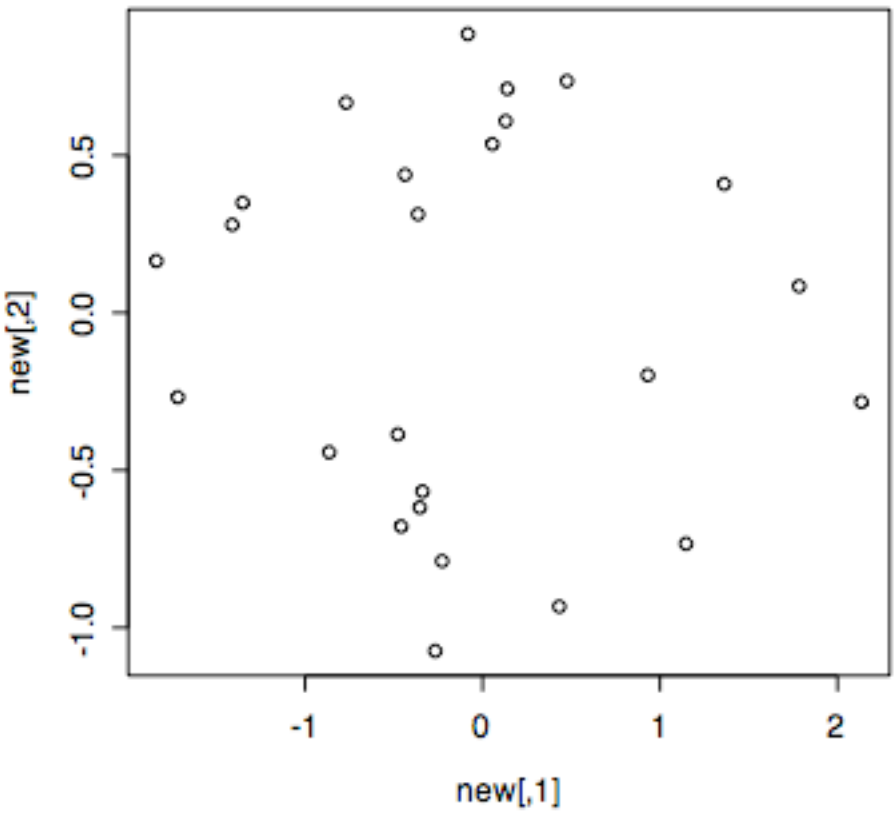
```
> a=mat.ex(print=T,corr=.7)
     [,1] [,2]
[1,]  1.0  0.7
[2,]  0.7  1.0
            V1          V2
1   0.856777697 -0.18353319
2   0.518490028  0.79870863
3   0.001430568 -0.61841160
4  -0.923719969 -0.29745391
5  -0.035843156 -0.47826208
6  -0.687189884  0.18704752
7   1.252204724  0.67356917
8   0.522859971 -0.33771509
9   0.417979143 -0.33961463
10  0.601965991 -0.40375589
11 -0.718405818  0.39638399
12 -0.071754873 -1.01497219
13 -0.610438192 -0.06429601
14 -1.183292009 -1.41601671
15 -0.640266435  0.16182161
16  0.567492931 -0.68416784
17 -0.947626360  0.57086653
18 -0.708182330 -1.20280735
19  0.292521315  1.32956262
```

```
20 -0.803829537  0.15512070
21 -0.353582226  0.96562814
22  1.321002496  1.20192461
23 -0.799358503 -1.19484412
24 -1.403571304 -1.02436650
25  1.309919645  1.71067712
          V1        V2
V1 1.0000000 0.4879367
V2 0.4879367 1.0000000
$values
[1] 1.4879367 0.5120633

$vectors
          [,1]       [,2]
[1,] 0.7071068  0.7071068
[2,] 0.7071068 -0.7071068

> data=as.matrix(a[[1]])
> ev=as.matrix(a[[2]][[2]])
> new=data%*%ev
> new
          [,1]       [,2]
1   0.47605576  0.73561088
2   0.93140011 -0.19814448
3  -0.43627147  0.43829460
4  -0.86350033 -0.44283698
5  -0.36352730  0.31283742
6  -0.35365406 -0.61817920
7   1.36172778  0.40915712
8   0.13091720  0.60851846
9   0.05541208  0.53569969
10  0.14015570  0.71115276
11 -0.22770382 -0.78827543
12 -0.76843207  0.66695536
13 -0.47710913 -0.38618084
14 -1.83798882  0.16456121
15 -0.33831158 -0.56716189
16 -0.08250162  0.88505782
17 -0.26640943 -1.07373662
18 -1.35127376  0.34975270
19  1.14698655 -0.73329894
20 -0.45870642 -0.67808021
21  0.43278181 -0.93282259
22  1.78397887  0.08420078
23 -1.41011419  0.27965056
24 -1.71681128 -0.26813829
```

25 2.13588445 -0.28337833
> plot(new)

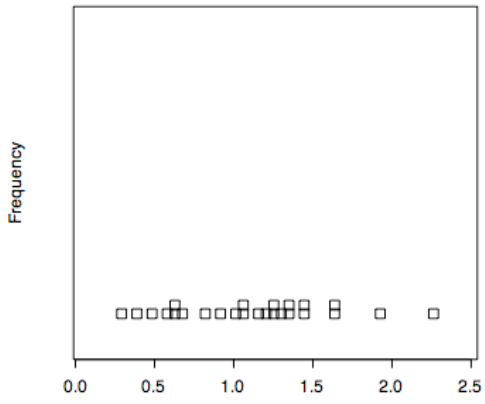



V2
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-1.5
-1.0
-0.5
0.0
0.5
1.0
V1



Now compute statistical distances: (I have used newly generated data for this ..)

Using the commands below, these are the "statistical distances" …




```
> out=mat.ex(corr=.7,print=T,p=2)
     [,1] [,2]
```

```
[1,]  1.0  0.7
[2,]  0.7  1.0
           V1          V2
1  -0.39941879  0.47828780
2   0.09335106  0.65288681
3  -0.82252322 -1.41982213
4   0.62799823  0.74132598
5  -0.50599435 -0.28386755
6  -1.76882885 -0.62860855
7  -0.25088953  1.55589986
8   1.07080601  0.49449178
9  -0.14348309  0.72133230
10  1.04758650  1.74353786
11 -0.40127929 -0.91463645
12  0.54301087  0.88971003
13 -0.82468044 -0.10912409
14 -0.26764256  1.03378413
15  0.92890018 -0.55209461
16 -1.56681511 -0.02930754
17  0.65922713 -0.18960179
18  0.27437677  0.37263011
19  0.73328653  0.99427499
20 -1.09677842  0.50885170
21  0.05113928  0.13619687
22 -0.72827338 -1.23634421
23 -0.82217403  0.13444789
24 -0.93184240 -0.80825216
25 -1.17995845 -1.57653427
          V1        V2
V1 1.0000000 0.5204037
V2 0.5204037 1.0000000
$values
[1] 1.5204037 0.4795963

$vectors
          [,1]       [,2]
[1,] 0.7071068  0.7071068
[2,] 0.7071068 -0.7071068

> data=as.matrix(out$data)
> ev=as.matrix(out$eigen$vectors)
> new=data%*%ev
> new
          [,1]       [,2]
1   0.05576881 -0.62063229
2   0.52766986 -0.39565153
```

```
3  -1.58557760  0.42235412
4   0.96825843 -0.08013482
5  -0.55851670 -0.15706737
6  -1.69524424 -0.80625751
7   0.92278165 -1.27759303
8   1.10683268  0.40751570
9   0.40860110 -0.61151683
10  1.97362297 -0.49211192
11 -0.93049294  0.36299833
12  1.01308667 -0.24515333
13 -0.66029951 -0.50597475
14  0.54174390 -0.92024763
15  0.26644178  1.04722146
16 -1.12862915 -1.08718203
17  0.33207526  0.60021269
18  0.45750295 -0.06947561
19  1.22157047 -0.18454671
20 -0.41572697 -1.13535195
21  0.13246666 -0.06014480
22 -1.38919442  0.35926033
23 -0.48629581 -0.67643385
24 -1.23043266 -0.08739149
25 -1.94913470  0.28042145
> new=as.data.frame(new)
> mean(new$V1)
[1] -0.08404486
> mean(new$V2)
[1] -0.2373153
> centroid=c(-0.084,-0.237)
> diff=new-centroid
> diff
          V1           V2
1   0.1397688 -0.383632285
2   0.7646699 -0.311651528
3  -1.5015776  0.659354116
4   1.2052584  0.003865181
5  -0.4745167  0.079932634
6  -1.4582442 -0.722257507
7   1.0067817 -1.040593031
8   1.3438327  0.491515705
9   0.4926011 -0.374516825
10  2.2106230 -0.408111924
11 -0.8464929  0.599998326
12  1.2500867 -0.161153327
13 -0.5762995 -0.268974745
14  0.7787439 -0.836247632
```

```
15  0.3504418  1.284221459
16 -0.8916292 -1.003182031
17  0.4160753  0.837212688
18  0.6945029  0.014524394
19  1.3055705  0.052453287
20 -0.1787270 -1.051351945
21  0.2164667  0.176855199
22 -1.1521944  0.443260330
23 -0.4022958 -0.439433845
24 -0.9934327 -0.003391491
25 -1.8651347  0.517421447
> distance=(diff$V1^2+diff$V2^2)^.5
> distance
 [1] 0.4083002 0.8257401 1.6399644 1.2052646 0.4812020 1.6273083
 [7] 1.4479099 1.4308997 0.6188042 2.2479789 1.0375685 1.2604313
[13] 0.6359784 1.1426951 1.3311777 1.3421538 0.9349030 0.6946548
[19] 1.3066237 1.0664353 0.2795274 1.2345168 0.5957718 0.9934385
[25] 1.9355755
> dotplot(distance)
```