Displays for Statistics 5401/8401

Lecture 5

September 16, 2005

Christopher Bingham, Instructor

612-625-1024, kb@umn.edu
372 Ford Hall

Class Web Page

http://www.stat.umn.edu/~kb/classes/5401

More on sample variance matrix

$$\mathbf{S} = \mathbf{S}_x = [s_{jk}] = (n-1)^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{X}}$$
$$= (n-1)^{-1}\sum_{1\le i\le n}\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i',$$

where $\tilde{\mathbf{X}} = [\tilde{\mathbf{x}}_1,\tilde{\mathbf{x}}_2,\ldots,\tilde{\mathbf{x}}_n]'$, $\tilde{\mathbf{x}}_i = \mathbf{x}_i - \overline{\mathbf{x}}$ is the matrix of deviations from the sample meam $\overline{\mathbf{x}}$.

- <u>Diagonal</u> elements:
  $s_{jj} = (n-1)^{-1}\sum_{1\le i\le n}(x_{ij} - \overline{x}_j)^2 = s_j^{\,2}$,
  the usual **sample variance**.
  $\sqrt{s_{jj}}$ = **sample standard deviation**
- <u>Off-diagonal</u> elements:
  $s_{jk} = (n-1)^{-1}\sum_{1\le i\le n}(x_{ij} - \overline{x}_j)(x_{ik} - \overline{x}_k)$, $j \ne k$
  = **sample covariance** of x_j and x_k
- $\mathbf{S}_x$ is *symmetric* ($\mathbf{S}_x' = \mathbf{S}_x$) since

$$s_{jk} = \sum_{1\le i\le n}(x_{ij} - \overline{x}_j)(x_{ik} - \overline{x}_k)$$
$$= \sum_{1\le i\le n}(x_{ik} - \overline{x}_k)(x_{ij} - \overline{x}_j) = s_{jk}$$

MacAnova:

```
Cmd> x
(1,1)         2.4        22.9          44
(2,1)        12.3        15.7        32.7
(3,1)        10.6        15.7        35.2
(4,1)        15.1        17.2        33.5
(5,1)        -1.3        22.5        26.7
Cmd> n <- nrows(x)
Cmd> xbar <- sum(x)/n # row vector, 1 by 3
Cmd> xtilde <- x - xbar # deviations from mean, 5 by 3
Cmd> # or xtilde <- x - rep(1/n,n)' %*% x
Cmd> df <- n - 1 # "degrees of freedom"
Cmd> s <- (xtilde' %*% xtilde)/df; s # sample variance matrix
(1,1)      48.337      -22.53       1.562
(2,1)      -22.53       13.07       3.775
(3,1)       1.562       3.775      38.947
```

Black box computation of **S** and $\overline{\mathbf{x}}$ using **tabs()**:

```
Cmd> tabs(x,covar:T)
(1,1)      48.337      -22.53       1.562
(2,1)      -22.53       13.07       3.775          S
(3,1)       1.562       3.775      38.947
Cmd> tabs(x,mean:T) # returns plain vector
(1)          7.82        18.8       34.42          Not row vector
Cmd> stats <- covar(x); stats # use pre-defined macro
component: n                                       sample size
(1)             5
component: mean
(1,1)        7.82        18.8       34.42          xbar, row vec.
component: covariance
(1,1)      48.337      -22.53       1.562
(2,1)      -22.53       13.07       3.775          S
(3,1)       1.562       3.775      38.947
Cmd> stats$covariance # extract S from structure stats
(1,1)      48.337      -22.53       1.562
(2,1)      -22.53       13.07       3.775
(3,1)       1.562       3.775      38.947
```


A matrix of Linear Combinations

Suppose $\mathbf{X}$ is a n by p data matrix and $\mathbf{A} = [\mathbf{a}_1,\ldots,\mathbf{a}_q]$ is a p by q matrix of constants, $\mathbf{a}_j = [a_{1j}, a_{2j}, \ldots, a_{pj}]'$. For example, you might have

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 0 & -2 \end{bmatrix}$$

Q. What is the $n \times q$ matrix
$\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_q] = \mathbf{XA} = [\mathbf{Xa}_1, \mathbf{Xa}_2 \ldots \mathbf{Xa}_q]$?

A. Each <u>column</u> $\mathbf{Y}_j = \mathbf{Xa}_j$ is a *linear combination* $\sum_{1\le\ell\le p}a_{\ell j}\mathbf{X}_\ell$ of the columns of $\mathbf{X}$ (e.g. $\mathbf{Y}_1 = \mathbf{X}_1+\mathbf{X}_2+\mathbf{X}_3$, $\mathbf{Y}_2=\mathbf{X}_1-\mathbf{X}_2$, ...)
Each <u>element</u> $y_{ij} = \mathbf{a}_j'\mathbf{x}_i = \sum_{1\le\ell\le p}a_{\ell j}x_{i\ell}$ is a linear combinations of the x-values for case i.

$\mathbf{Y}$ is a new data matrix derived from the original data matrix $\mathbf{X}$.

Sample mean and variance matrix of Y

- The sample mean $\overline{\mathbf{y}} = \sum_i \mathbf{y}_i/n$ of $\mathbf{Y}$ is

$$\overline{\mathbf{y}} = \mathbf{A}'\overline{\mathbf{x}} \qquad (q \times 1) \text{ column vector}$$

or, as a row vector,

$$\overline{\mathbf{y}}' = \overline{\mathbf{x}}'\mathbf{A} \qquad (1 \times q)$$

$$\overline{y}_j = \mathbf{a}_j'\overline{\mathbf{x}} \qquad \text{(univariate mean)}$$

- The *variance matrix* $\mathbf{S}_Y$ of $\mathbf{Y}$ is

$$\mathbf{S}_Y = \mathbf{A}'\mathbf{S}_X\mathbf{A} = [\mathbf{a}_j'\mathbf{S}_X\mathbf{a}_k]_{1\le j\le q, 1\le k\le q} \ (q \times q).$$

Diagonal elements are

$$s_{Y11} = s_{y_1}^{\ 2} = \mathbf{a}_1'\mathbf{S}_X\mathbf{a}_1, \ \ldots\ , s_{Yqq} = s_{y_q}^{\ 2} = \mathbf{a}_q'\mathbf{S}_X\mathbf{a}_q$$

This applies when the columns of $\mathbf{A}$ define the linear combinations.

Comment: You will find it useful to be able to recognize an expression like $\mathbf{a}'\mathbf{S}_X\mathbf{a}$ as representing the sample variance of a linear combination $\mathbf{a}'\mathbf{x}$.

MacAnova Example

```
Cmd> x # previously entered data matrix
(1,1)         2.4        22.9          44
(2,1)        12.3        15.7        32.7
(3,1)        10.6        15.7        35.2
(4,1)        15.1        17.2        33.5
(5,1)        -1.3        22.5        26.7

Cmd> a <- matrix(vector(1,1,1,1,-1,0,1,1,-2),3)

Cmd> a # matrix of linear combination coefficients
(1,1)           1           1           1
(2,1)           1          -1           1
(3,1)           1           0          -2

Cmd> y <- x %*% a; y # data matrix of linear combinations
(1,1)        69.3       -20.5       -62.7
(2,1)        60.7        -3.4       -37.4
(3,1)        61.5        -5.1       -44.1
(4,1)        65.8        -2.1       -34.7
(5,1)        47.9       -23.8       -32.2

Cmd> s_x <- tabs(x1, covar:T) # S_x

Cmd> s_y <- tabs(y, covar:T); s_y # S_y
(1,1)      65.968      33.054     -66.884
(2,1)      33.054      106.47      39.693
(3,1)     -66.884      39.693      150.79

Cmd> a' %*% s_x %*% a
(1,1)      65.968      33.054     -66.884
(2,1)      33.054      106.47      39.693
(3,1)     -66.884      39.693      150.79
```


Eigenvalues and Eigenvectors

Let $\mathbf{A} = [a_{ij}]$ be a p by p *square* matrix.

Vocabulary Suppose $\mathbf{u} \ne \mathbf{0}$ is a p by 1 vector that satisfies

$$\mathbf{A}\mathbf{u} = \lambda\mathbf{u} \text{ for some constant } \lambda.$$

Then $\mathbf{u}$ is an ***eigenvector*** of $\mathbf{A}$ with corresponding ***eigenvalue*** λ (also called *proper* or *characteristic* value and vector).

```
Cmd> a # note a is symmetric 2 by 2
(1,1)      2.9412     0.23529
(2,1)     0.23529      2.0588
```

Enter vectors $\mathbf{u}_1$ and $\mathbf{u}_2$

```
Cmd> u1 <- vector(4, 1); u2 <- vector(-1, 4)

Cmd> a %*% u1 # a %*% u1 = 3*u1
(1,1)          12              Eigenvalue = 3
(2,1)           3

Cmd> a %*% u2 # a %*% u2 = 2*u2
(1,1)          -2              Eigenvalue = 2
(2,1)           8
```

$$\mathbf{A}\mathbf{u}_1 = 3\times\mathbf{u}_1 \text{ and } \mathbf{A}\mathbf{u}_2 = 2\times\mathbf{u}_2$$

so

- $\mathbf{u}_1$ is an eigenvector with eigenvalue 3
- $\mathbf{u}_2$ is an eigenvector with eigenvalue 2

When $\mathbf{u}$ is an eigenvector with eigenvalue λ, so is $\mathbf{u}/c$ where c is a constant

Proof: $\mathbf{A}(\mathbf{u}/c) = \mathbf{A}\mathbf{u}/c = \lambda\mathbf{u}/c = \lambda(\mathbf{u}/c)$.

In particular $\tilde{\mathbf{u}} = \mathbf{u}/\|\mathbf{u}\|$ is an eigenvector such that $\|\tilde{\mathbf{u}}\| = 1$.

```
Cmd> u <- hconcat(u1,u2); sum(u^2) # u is 2 by 2
(1,1)          17          17  squared norms of u1 and u2
```

This shows $\|\mathbf{u}_1\| = \|\mathbf{u}_2\| = \sqrt{17}$

```
Cmd> u / sqrt(17) # columns are eigenvectors
(1,1)     0.97014    -0.24254
(2,1)     0.24254     0.97014
```

MacAnova:

- `eigenvals()` finds eigenvalues
- `eigen()` finds *both* eigenvalues and eigenvectors

```
Cmd> eigenvals(a) # just eigenvalues
(1)             3           2    In decreasing order

Cmd> eigs <- eigen(a); eigs
component: values
(1)             3           2
component: vectors
(1,1)    -0.97014     0.24254  Normalized columns
(2,1)    -0.24254    -0.97014  norms of columns are 1

Cmd> eigs$vectors # or eigs[2] #extract just vectors
(1,1)    -0.97014     0.24254
(2,1)    -0.24254    -0.97014
```

Note that signs are reversed from `u / sqrt(17)`. This doesn't matter.

Facts concerning eigenthings

When **A** is p×p *symmetric*, there are

- Exactly p *linearly independent* eigenvectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p$ with *real* elements
- with corresponding *real* eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots \geq \lambda_p$. The decreasing ordering is conventional.

When **A** is non-symmetric, eigenvectors and/or eigenvalues may be *complex*, requiring *imaginary* numbers

For example, you can check

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 1+i \\ 1-i \end{bmatrix} = -i \times \begin{bmatrix} 1+i \\ 1-i \end{bmatrix}, \; i = \sqrt{-1}$$

so $\mathbf{u} = \begin{bmatrix} 1+i \\ 1-i \end{bmatrix}$ is an eigenvector of

$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$, eigenvalue $\lambda = -i = -\sqrt{(-1)}$

Fact: If $\lambda_i \neq 0$, for all eigenvalues λ_i of **A**, then **A** is non-singular, i.e., $\mathbf{A}^{-1}$ exists.

Vocabulary

> A symmetric matrix **A** is ***positive definite*** if
> $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$ for *every* $\mathbf{x} \neq \mathbf{0}$

> A symmetric matrix **A** is ***positive semi-definite*** if
> $\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0$ for *every* $\mathbf{x} \neq \mathbf{0}$

- **Fact**: A symmetric matrix is positive definite if and only if $\lambda_i > 0$, i = 1,...,p
- **Fact**: A positive definite symmetric matrix is **always invertible** since all eigenvalues ≠ 0.
- A symmetric matrix is positive semi-definite if and only if $\lambda_i \geq 0$, i = 1,...,p

Let **A** = diag[$a_{11}, a_{22}, \ldots, a_{pp}$] be diagonal. Then

- λ_i's are the a_i's in decreasing order

Vocabulary: $\mathbf{x}'\mathbf{A}\mathbf{x}$ is an example of a ***quadratic form***.

Expanded in full a quadratic form is

$$\begin{aligned}
\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}'(\mathbf{A}\mathbf{x}) &= \textstyle\sum_{1 \leq i \leq p} x_i \left(\sum_{1 \leq j \leq p} a_{ij} x_j\right) \\
&= \textstyle\sum_{1 \leq i \leq p} \sum_{1 \leq j \leq p} a_{ij} x_i x_j \\
&= \textstyle\sum_i a_{ii} x_i^2 + \sum\sum_{i \neq j} a_{ij} x_i x_j \\
&= \textstyle\sum_i a_{ii} x_i^2 + 2\sum\sum_{i<j} a_{ij} x_i x_j
\end{aligned}$$

The last step is OK because **A** is symmetric so $a_{ji} = a_{ij}$.

When p = 2,

$$\begin{aligned}
\mathbf{x}'\mathbf{A}\mathbf{x} &= a_{11}x_1^2 + a_{22}x_2^2 + a_{12}x_1x_2 + a_{21}x_2x_1 \\
&= a_{11}x_1^2 + a_{22}x_2^2 + a_{12}x_1x_2 + a_{12}x_2x_1 \\
&= a_{11}x_1^2 + a_{22}x_2^2 + 2a_{12}x_1x_2
\end{aligned}$$

When p = 3,

$$\begin{aligned}
\mathbf{x}'\mathbf{A}\mathbf{x} &= a_{11}x_1^2 + a_{22}x_2^2 + a_{33}x_3^2 + \\
&\quad 2a_{12}x_1x_2 + 2a_{13}x_1x_3 + 2a_{23}x_2x_3
\end{aligned}$$

Suppose $\mathbf{S}_x$ is a p×p sample variance matrix and suppose $y = \mathbf{a}'\mathbf{x}$ is an arbitrary linear combination with $\mathbf{a} \neq \mathbf{0}$.

Recall that $\mathbf{a}'\mathbf{S}_x\mathbf{a} = s_y^2 \geq 0$. This shows that

> $\mathbf{S}_x$ is positive semi-definite

Now suppose $\mathbf{S}_x$ is not *positive* definite.

Then there is at least one $\mathbf{a} \neq \mathbf{0}$, with $\mathbf{a}'\mathbf{S}_x\mathbf{a} = 0$. In other words,

> $s_y^2 = 0$, where $y = \mathbf{a}'\mathbf{x} = \sum_j a_j x_j$

Also, when $\mathbf{S}_x$ is not positive definite there is at least one 0 eigenvalue and any **a** with $\mathbf{a}'\mathbf{S}_x\mathbf{a} = 0$ is an eigenvector with 0 eigenvalue.

Q. When can $s_y^2 = 0$ can happen?

A. Only when

$$y_1 = y_2 = \dots = y_n = \text{constant } c$$

Now $y_i = \sum_{1\le i\le p} a_j x_{ij}$.

This means that for any j with $a_j \ne 0$, the value for the j^{th} x variable is determined by the other variables:

$$x_{ij} = c - (a_1/a_j)x_{i1} - (a_2/a_j)x_{i2} - \dots - (a_{j-1}/a_j)x_{i,j-1} - (a_{j+1}/a_j)x_{i,j+1} - \dots - (a_p/a_j)x_{i,p}$$

so x_j is not really needed.

Vocabulary: In such a case, $x_1, x_2, \dots, x_p$ are **collinear**, there is an exact linear relationship between them.

When the the smallest eigenvalue λ_p of **S** is small relative to the largest λ_1, that is, $\lambda_p/\lambda_1 \approx 0$, $x_1, x_2, \dots, x_p$ are *nearly* **collinear**.

Properties of eigenvectors and eigenvalues

Let **A** be p×p symmetric.

Suppose we want to maximize (find the largest possible value of) the *quadratic form*

$$\mathbf{v}'\mathbf{A}\mathbf{v} = \sum_i a_{ii}v_i^2 + 2\sum\sum_{i<j} a_{ij}v_iv_j$$

over all choices for p×1 vector **v** such that

$$\|\mathbf{v}\|^2 = \sum_i v_i^2 = 1 \text{ ("unit" vector)}$$

Solution: $\mathbf{v} = \mathbf{u}_1$, the normalized *first* eigenvector.

The maximized value is $\mathbf{u}_1'\mathbf{A}\mathbf{u}_1 = \mathbf{u}_1'(\lambda_1\mathbf{u}_1) = \lambda_1\mathbf{u}_1'\mathbf{u}_1 = \lambda_1\|\mathbf{u}_1\|^2 = \lambda_1$, the largest eigenvalue.

Thus

$$\boxed{\max_{\|\mathbf{v}\|=1}\mathbf{v}'\mathbf{A}\mathbf{v} = \mathbf{u}_1'\mathbf{A}\mathbf{u}_1 = \lambda_1}$$

Similarly, when you want to *minimize* $\mathbf{v}'\mathbf{A}\mathbf{v}$ over all choices for **v** such that $\|\mathbf{v}\|^2 = 1$, the **solution** is $\mathbf{v} = \mathbf{u}_p$, the *last* eigenvector with the smallest eigenvalue.

The minimized value is

$$\lambda_p = \mathbf{u}_p'\mathbf{A}\mathbf{u}_p = \textit{minimum} \text{ eigenvalue.}$$

Thus

$$\boxed{\min_{\|\mathbf{v}\|=1}\mathbf{v}'\mathbf{A}\mathbf{v} = \mathbf{u}_p'\mathbf{A}\mathbf{u}_p = \lambda_p}$$

Now, for any **w**, $\mathbf{w}'\mathbf{A}\mathbf{w}/\|\mathbf{w}\|^2 = \mathbf{v}'\mathbf{A}\mathbf{v}$, where $\mathbf{v} \equiv \mathbf{w}/\|\mathbf{w}\|$ has $\|\mathbf{v}\| = 1$.

These two results imply that

$$\lambda_1 \ge \mathbf{w}'\mathbf{A}\mathbf{w}/\|\mathbf{w}\|^2 \ge \lambda_p, \text{ for all } \mathbf{w} \ne \mathbf{0}.$$

Or, multiplying by $\|\mathbf{w}\|^2$,

$$\lambda_1\|\mathbf{w}\|^2 \ge \mathbf{w}'\mathbf{A}\mathbf{w} \ge \lambda_p\|\mathbf{w}\|^2, \text{ for all } \mathbf{w}.$$

This are bounds on the values of the quadratic form (recall $\lambda_1 = \lambda_{max}, \lambda_p = \lambda_{min}$).

For p = 2, every unit vectors has the form

$$\mathbf{u} = \begin{bmatrix} \cos\theta \\ \\ \pm\sin\theta \end{bmatrix} \text{ for some } 0 \le \theta \le 2\pi$$

Here I computed and graphed $\mathbf{u}'\mathbf{A}\mathbf{u}$ for equally spaced values of θ

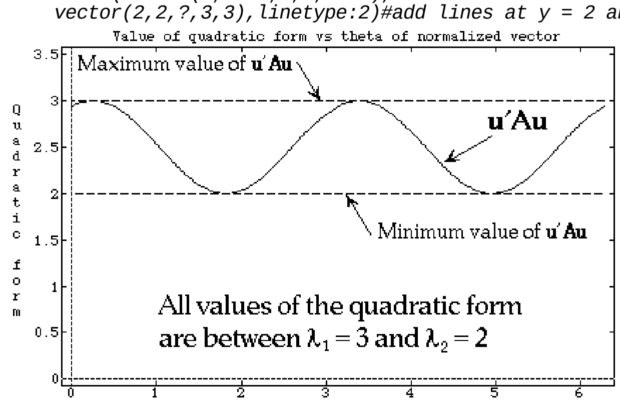
```
Cmd> theta <- 2*PI*run(0,250)/250 #equally spaced radian angles

Cmd> qformvals <- 0 * theta # create empty vector

Cmd> for(i,1,length(theta)){
        @u <- vector(cos(theta[i]), sin(theta[i]))
        qformvals[i] <- @u' %*% a %*% @u
    ;;} # compute quadratic form for each angle

Cmd> lineplot(theta,qformvals,xlab:"Theta",\
title:"Value of quadratic form vs theta of normalized vector",\
 ylab:"Quadratic form",ymin:0,ymax:3.5,show:F)

Cmd> addlines(vector(0,2*PI,?,0,2*PI),\
     vector(2,2,?,3,3),linetype:2)#add lines at y = 2 and 3
```



I gave an example of a *nonsymmetric* matrix whose eigenvectors and eigenvalues were composed of complex numbers rather than real.

There is one important *nonsymmetric* case where eigenvalues and eigenvectors *are* real:

Fact: Suppose **B** and **C** are *symmetric* p×p matrices. Then **A** = **BC** is (usually) nonsymmetric but its eigenvectors and eigenvalues are *real* .

You can always assume that the eigenvectors of a symmetric matrix **A** are *normalized*, that is

$$\|\mathbf{u}\| = \sqrt{(\sum_{1\le i\le p} u_i^2)} = 1.$$

MacAnova always produces normalized eigenvectors.

More Facts About Matrices

When **A** is a p×p square matrix, these five statements are either all true or all false.

1. **A** is non-singular (has an inverse)
2. rank(**A**) = p (**A** has "full rank")
3. **Ab** ≠ 0 for *all* **b** ≠ **0**
4. det(**A**) ≠ 0
5. All eigenvalues $\lambda_i \neq 0$

Here are some ways you might use this equivalence:

- If you have a **b** ≠ **0** but **Ab** = **0**, then **A** is singular, det(**A**) = 0 and **A** has at least one eigenvalue = 0 with eigenvector **b**.
- When det(**A**) = 0 there is a vector **b** ≠ **0** such that **Ab** = **0** and **A** has at least one zero eigenvalue

Suppose **u** is an eigenvector of **A** = **A**′ with eigenvalue λ. Then, *by definition*, **Au** = λ**u**.

This is the same as

$$\mathbf{Au} - \lambda\mathbf{u} = (\mathbf{A} - \lambda\mathbf{I}_p)\mathbf{u} = \mathbf{0}.$$

Because **u** ≠ **0**, property 4 tells us

$$P(\lambda) \equiv \det(\mathbf{A} - \lambda\mathbf{I}_p) = 0.$$

```
Cmd> write(a) # has eigenvalue 3 and 2
a:
(1,1)       2.94117647      0.235294118
(2,1)      0.235294118       2.05882353

Cmd> det(a - 3*dmat(2,1))# 3 is eigenvalue of a
(1)  2.5969e-16
```

Note: `write(a)` is the simplest way to print a vector or matrix with more significant digits (9) than the default (5).

When **A** = diag[a_{11}, a_{22}, ..., a_{pp}] is diagonal, its eigenvalues are a_{11}, a_{22}, ..., a_{pp} and the eigen vectors are $\mathbf{e}_j^p = [\underset{1}{0} \ldots \underset{j-1}{0}\ \underset{j}{1}\ \underset{j+1}{0} \ldots \underset{p}{0}]'$, the columns of $\mathbf{I}_p$.

$P(\lambda) \equiv \det(\mathbf{A} - \lambda\mathbf{I}_p)$ is actually a polynomial of degree p in λ:

$$P(\lambda) = (-1)^p\lambda^p + d_1\lambda^{p-1} + \ldots + d_{p-1}\lambda + d_p$$

Therefore the eigenvalues λ_i satisfy

$$P(\lambda_i) = 0,\ i = 1,\ldots,p.$$

If you can find the zeros of polynomial, you can compute eigenvalues.

If you like to do such things, for the matrix **a** in the example which has $\lambda_1 = 3$, $\lambda_2 = 2$, you can check that

$$P(\lambda) = \lambda^2 - 5\lambda + 6$$

For which

$$d_1 = -5 = -(\lambda_1 + \lambda_2),\ d_2 = 6 = \lambda_1\times\lambda_2$$

In general,

- $d_1 = (-1)^{p-1}\sum_{1\le j\le p}\lambda_j = (-1)^{p-1}\mathrm{trace}(\mathbf{A})$
- $d_p = \lambda_1\lambda_2\ldots\lambda_p = \det(\mathbf{A})$

```
Cmd> polyroot(-vector(-trace(a), det(a)))
(1,1)          2          0     Real and imaginary
(2,1)          3          0     parts of zeros
```