

## Eigen Vector and Eigen Value Lab

---

We have seen this syntax to compute a correlation matrix among four selected variables from the TIMSS dataset. This includes a couple modifications to combine some of the steps.

MATRIX.	
Get X / Variables = club, day6, age, self_efy.	"Get" the variables you want to use
compute ones = make(nrow(x),1,1).	compute a "ones" vector
compute means = csum(x)/nrow(x).	compute "means" by taking the column sum divided by n
compute dx = x - ones*means.	compute a deviation score vector "dx" by subtracting a means matrix "ones*means" from our X matrix
compute sigma = sscp(dx)/nrow(x).	compute "sigma" the variance-covariance matrix by dividing the sums of squares and cross-products "sscp" matrix by n
print sigma.	
compute s = inv(mdiag(sqrt(diag(sigma)))).	create a diagonal matrix of standard deviations "s" by turning the square roots of the diagonal of sigma into a diagonal matrix
compute r = s*sigma*s.	compute the correlation matrix "r" by standardizing the variance-covariance matrix sigma
print r.	
END MATRIX.	

Accompanying this computer assignment is the data set of 150 students who recently completed one year of college. The data set contains the following information:

1. Gender (1=female, 0=male) and Minority status (1=minority, 0=non-minority student).
2. High School GPA.
3. High School Proficiency Test scores (science, math, writing, reading) taken during junior year in high school.
4. ACT test scores (English, math, reading, science, composite) taken during senior year. Notice that the composite is the average of the four subject tests.
5. SAT test scores (math, verbal) taken during senior year.
6. College GPA in math courses, English courses, and science courses.

## Eigen Value/Vector Lab

---

Create a correlation matrix with the four ACT subject tests and the composite score: actengl, actmath, actread, actscien, actcomp.

- a. Print the variance-covariance matrix ( $\sigma$ ) and the correlation matrix ( $r$ ).
- b. Compute the determinant, trace, and rank of both; print all.
- c. Compute the eigen values and eigen vectors of both; print all.
- d. Compute the inverse of  $\sigma$ ; print.

Here is some syntax that will help with b, c, and d.

```
compute determs = det(sigma).
compute traces = trace(sigma).
compute ranks = rank(sigma).
call eigen(sigma,a,b).
print determs.
print traces.
print ranks.
print a.
print b.
```

```
compute determr = det(r).
compute tracer = trace(r).
compute rankr = rank(r).
call eigen(r,c,d).
print determr.
print tracer.
print rankr.
print c.
print d.
```

```
compute invsigma = inv(sigma).
print invsigma.
```

***Answer the following questions:***

1. Based on the output, what do the determinant, trace, and rank tell you about the  $\sigma$  and  $r$  matrices? Interpret result.
2. You should expect an error message when you compute the inverse of  $\sigma$  – what is that error message that you received? What does this mean?
3. Based on your analysis of  $\sigma$ , what is the total variance that the eigen values could possibly account for (*hint*: trace of  $\sigma$ )?
4. How much variance does the first eigen vector account for – what is the proportion of variance accounted for in  $\sigma$ ?
5. Based on your analysis of  $r$ , what is the total variance that the eigen values could possibly account for in  $r$ ?
6. How much variance does the principal component (first eigen vector) account for in  $r$ ?
7. What is the number of eigen vectors needed to account for 100% of the variance among these five variables? What matrix summary index did you compute that also gives us this information?