

**CSCI4390/6390 – Data Mining**  
**Fall 2006, Quiz II**  
**Total Points: 25**

Name: \_\_\_\_\_

1. (5 points) True or False: Changing the scale of measurement for the covariance matrix (e.g., from meters to kilometers) will change the value of the covariance.
2. (10 points) Recall that correlation between two variables is defined as:

$$\rho_{\mathbf{x}\mathbf{y}} = \frac{\sigma_{\mathbf{x}\mathbf{y}}}{\sigma_{\mathbf{x}}\sigma_{\mathbf{y}}}$$

where  $\sigma_{\mathbf{x}\mathbf{y}}$  is the covariance between  $\mathbf{x}$  and  $\mathbf{y}$ , and  $\sigma_{\mathbf{x}}$  and  $\sigma_{\mathbf{y}}$  give the standard deviation of  $\mathbf{x}$  and  $\mathbf{y}$ , respectively. Under what condition will the covariance matrix  $\Sigma$  be identical to the correlation matrix (whose  $(i, j)$  entry gives the correlation between attributes  $i$  and  $j$ )? What can you conclude about the two variables?

3. (10 points) Consider the count table shown below:

<b>x</b>	<b>y</b>	
	<b>A</b>	<b>B</b>
<b>C</b>	40	35
<b>D</b>	20	5

Compute the  $\chi^2$  statistic and determine whether  $\mathbf{x}$  and  $\mathbf{y}$  are independent or not using the chi-square value table for 1% area under the curve to the right of the given value:

Degrees of Freedom	$\chi^2$ Value
1	6.64
2	9.21
3	11.35
4	13.28