ECNS561, Fall 2011

Take-home Graded Practice Opportunity

Due date: September 23, 2010, 5:00 p.m. (to Jane or Rebecca)

1. Evaluate the following sums:

$$a) \qquad \sum_{i=1}^{3}\sum_{j=1}^{3} 2$$

$$b) \qquad \sum_{i=1}^{5} i \cdot j$$

$$c) \qquad \sum_{i=1}^{3} \frac{Y-i}{i}$$

$$d) \qquad \sum_{i=1}^{3}\sum_{j=1}^{3} -1^{i} \cdot Y_i \cdot Y_j$$

   (a) $\sum_{i=1}^{3}\sum_{j=1}^{3} 2 = 3 \times 3 \times 2 = 18$

   (b) $\sum_{i=1}^{5} i \cdot j = 15j$

   (c) $\sum_{i=1}^{3} \frac{Y-i}{i} = \frac{Y-1}{1} + \frac{Y-2}{2} + \frac{Y-3}{3} = \frac{11}{6}Y - 3$

   (d) $\sum_{i=1}^{3}\sum_{j=1}^{3} -1^{i} \cdot Y_i \cdot Y_j = Y_2^2 - Y_1^2 - Y_3^2 - 2Y_1Y_3$

2. Suppose that you trying to determine the number of different combinations of name initials.

   (a) If all people have two given names (a first and a middle) and one surname, how many total unique combinations are possible?

   There are 26 letters in the English alphabet. If everyone had three initials, then there are 26^3 *possible unique combinations.*

   (b) If some people have only one given name and surname and other have two given names, how many total unique combinations of initials are possible?

We already know that there are 26^3 initial combinations if people had three initials. However, if people had only two initials, then there would be 26^2 unique combinations. So, the total number of combinations if people could have two or three initials is $26^2 + 26^3$.

3. Suppose that the following table describes the population of strikeouts, walks, and earned run averages for the starting pitcher rotation of a baseball team.

Strikeouts	**Walks**	**Earned Run Average (ERA)**
80	130	2.5
46	88	3.25
112	90	4.75
75	52	5.5
92	75	3.5

Calculate the following:

(a) Expected value of each variable.

We don't know the probability of each occurrence, so we need to assume that each is given an equal likelihood – 1/5. Therefore, we can retrieve each of the expected values by calculating: $\frac{1}{5}\sum_i^5 y_i$, where y is each of the three variables.

$E[strikeouts] = 81$
$E[walks] = 87$
$E[ERA] = 3.9$

(b) Variance of each variable.
We can use the formula $\sigma^2 = E[Y^2] - E[Y]^2$ to determine each of the variances.

$\sigma^2_{so} = \{(80^2)\times0.2+(46^2)\times0.2+(112^2)\times0.2+(75^2)\times0.2+(92^2)\times0.2\}-81^2 = 468.8$

$\sigma^2_{walk} = 645.5$

$\sigma^2_{era} = 1.17$

(c) Standard deviation of each variable.

The standard deviation is just the square root of the variance.

$\sigma_{so} = \sqrt{468.8} = 21.65$

$\sigma_{walk} = 25.41$

$\sigma_{era} = 1.08$

(d) Interpret the standard deviations.

Each standard deviation gives information about the dispersion of each random variable in the units of the random variable. So, within a one-standard deviation, a pitcher may throw between approximately 60 and 102 strikeouts, walk between 62 and 112 batters, and have an ERA that is between 2.82 and 4.98.

4. You are studying youth school violence and its relationship to the number of police officers in a school. You know that the joint probability function is as follows:

		Number of Police Officers						
		0	1	2	3	4	5	6
Youth violence cases	0	0.03324	0.04378	0.02869	0.02324	0.02119	0.00198	0.00206
	1	0.11545	0.07189	0.0499	0.02901	0.01155	0.00144	0.00033
	2	0.11158	0.09628	0.04982	0.01762	0.00659	0.00458	0.00038
	3	0.09386	0.04166	0.01632	0.005	0.00096	0.00061	0.00005
	4	0.04147	0.01699	0.00311	0.0025	0.00022	0	0
	5	0.02732	0.00428	0.00429	0.00066	0.00003	0	0
	6	0.00927	0.00091	0.00056	0.00005	0	0	0

Calculate the following:

(a) Marginal probability function of each individual variable. What do these marginal pdfs imply about the likelihood of observation youth violence and a particular number of police officers?

The marginal probabilities are simply the sum of all of the joint probabilities in a particular row or column of the joint probability density function table. For example, to get the marginal probability of having zero youth violence cases, you will add up the joint probabilities in the row where "Youth violence cases = 0" and across all of the columns of the police officers random variable.

$P[YV = 0] = 0.03324 + 0.04378 + 0.02869 + 0.02324 + 0.02119 + 0.00198 + 0.00206 = 0.154$

Similarly:

$P[YV = 1] = 0.279$
$P[YV = 2] = 0.287$
$P[YV = 3] = 0.159$
$P[YV = 4] = 0.064$
$P[YV = 5] = 0.037$
$P[YV = 6] = 0.011$

For the marginal probabilities of the number of police officers, you would add across all rows in each column:
$P[PO = 0] = 0.03324 + 0.11545 + 0.11158 + 0.09386 + 0.04147 + 0.02732 + 0.00927 = 0.432$

Similarly:
$P[PO = 1] = 0.276$
$P[PO = 2] = 0.152$
$P[PO = 3] = 0.078$
$P[PO = 4] = 0.041$
$P[PO = 5] = 0.009$
$P[PO = 6] = 0.003$

These tell marginals tell us that it is most likely that we would find a school without a single police officer and where there are two youth violence cases.

(b) Expected value of youth violence cases.

The basic approach is shown in problem (3). $E[YV] = 1.833$ *cases.*

(c) Expected value of police officers.

$E[PO] = 1.037$ officers.

(d) Variances around each expected value.

The basic approach is shown in problem (3).

$\sigma^2_{yv} = 1.823$

$\sigma^2_{po} = 1.479$

(e) Correlation between the number of youth violence cases and the number of police officers.

There are three components necessary to calculate the correlation. We require the covariance between the two random variables, and each of the standard deviations.

Calculating the standard deviations is straightforward:

$\sigma_{yv} = \sqrt{\sigma^2_{yv}} = 1.35$ *officers.*
$\sigma_{po} = \sqrt{\sigma^2_{po}} = 1.22$ *cases.*

The covariance is calculated using the formula $E[YV \times PO] - E[YV]E[PO]$. *The second term is relatively straightforward using answers from parts (a) and (b). The first term must be calculated using the joint pdf table above. That is, for each combination of outcomes, we need to determine the value:* $\{yv_i \times po_j \times P[YV = y_i, PO = po_j]\}$. *Essentially, you can derive another* 7×7 *table of these values. After doing so, you sum those values up, providing you with the term* $E[YV \times PO]$.

$E[YV \times PO] = 1.418$
$\sigma_{YV,PO} = -0.483$
$\rho_{YV,PO} = -0.435$

The correlation implies that, as expected, schools in which there is a higher number of police officers are expected to have less youth violence cases.

5. Applied practice. Please turn in your code and a summary of your output. If you are going to do the assignment in anything other than SAS, please also email me your code. You don't have to print the entire output window; just provide the relevant statistics asked for in the problems.

   You are presented with two probability density functions, one that characterizes the likelihood of a person being of a particular political view and another characterizing the likelihood of a person owning a particular number of pets. Numbers in parentheses are used to represent each of the discrete outcomes.

Party	P[Party = party]
Very Liberal (1)	0.2
Moderate (2)	0.7
Very Conservative (3)	0.1

# of Pets	P[# Pets = # pets]
0	0.4
1	0.3
2	0.2
3	0.1

Perform the following in IML (or another matrix language):

(a) Create a dataset with 60 observations of each variable, such that the observations are generated using the pdfs above.

(b) Again in IML, create a joint probability density table using each of the marginals.

(c) Calculate the expected value and variance of each variable. Don't use any of the packaged procedures – they assume equally likely outcomes, which is not the case in this problem.

(d) Create a frequency table of the created data set to check whether the data follow the appropriate pdf. In SAS, you can do this using the `proc freq` procedure (see your SAS lab notes on the use of the `proc freq` procedure). To output your matrices into a SAS dataset, use the following SAS code:

```
/* Substitute the output SAS dataset name for ``SASdata'' and
   the IML matrix name for ``IMLdata'' */

varnames = {``Party'' ``Pets''};
create SASdata from IMLdata[c=varnames];
append from IMLdata;
quit;
```

Solution code is posted under the "Homework" link of the course website.