University of California, Los Angeles
Department of Statistics

Statistics 13 **Instructor: Nicolas Christou**

Homework 2

EXERCISE 1
Consider the following data on `cadmium`, `copper`, `lead`, and `zinc` at 6 locations define by the coordinates `x` and `y`.

```
       x      y cadmium copper lead zinc
1 181072 333611    11.7     85  299 1022
2 181025 333558     8.6     81  277 1141
3 181165 333537     6.5     68  199  640
4 181298 333484     2.6     81  116  257
5 181307 333330     2.8     48  117  269
6 181390 333260     3.0     61  137  281
```

Compute the following by hand:

a. The standard deviation of `cadmium`.

b. The standard deviation of `lead`.

c. The estimates of β_0 and β_1 of the model

$$\texttt{cadmium}_i = \beta_0 + \beta_1 \texttt{lead}_i + \epsilon_i$$

d. The covariance between `cadmium` and `lead`.

e. The correlation coefficient between `cadmium` and `lead`.

EXERCISE 2
Consider the simple regression model:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

a. Show that the sum of the residuals is always equal to zero:

$$\sum_{i=1}^{n} e_i = 0, \quad \text{where} \quad e_i = y_i - \hat{y}_i$$

b. Show that the estimate of β_1 can be computed also using:

$$\hat{\beta}_1 = r \frac{\texttt{sd(y)}}{\texttt{sd(x)}}$$

c. Use the result of part (b) to compute again $\hat{\beta}_1$ of exercise 1.

EXERCISE 3
Access the following data in `R` (see other side for variable description):

```
a <- read.table("http://www.stat.ucla.edu/~nchristo/statistics13/house_data.txt", header=TRUE)
```

a. Use `R` to run the regression of `PRICE` on `FLR` and `RMS`. Write the fitted regression equation.

b. Use `R` to predict the house prices for the following data on `FLR` and `RMS`:

FLR	RMS
841	4
890	4
1050	5
1560	6
2180	8

Data for exercise 3:

$PRICE$	BDR	FLR	FP	RMS	ST	LOT	TAX	BTH	CON	GAR	CDN	$L1$	$L2$
53	2	967	0	5	0	39	652	1.5	1	0	0	1	0
55	2	815	1	5	0	33	1000	1	1	2	1	1	0
56	3	900	0	5	1	35	897	1.5	1	1	0	1	0
58	3	1007	0	6	1	24	964	1.5	0	2	0	1	0
64	3	1100	1	7	0	50	1099	1.5	1	1.5	0	1	0
44	4	897	0	7	0	25	960	2	0	1	0	1	0
49	5	1400	0	8	0	30	678	1	0	1	1	1	0
70	3	2261	0	6	0	29	2700	1	0	2	0	1	0
72	4	1290	0	8	1	33	800	1.5	1	1.5	0	1	0
82	4	2104	0	9	0	40	1038	2.5	1	1	1	1	0
85	8	2240	1	12	1	50	1200	3	0	2	0	1	0
45	2	641	0	5	0	25	860	1	0	0	0	0	1
47	3	862	0	6	0	25	600	1	1	0	0	0	1
49	4	1043	0	7	0	30	676	1.5	0	0	0	0	1
56	4	1325	0	8	0	50	1287	1.5	0	0	0	0	1
60	2	782	0	5	1	25	834	1	0	0	0	0	1
62	3	1126	0	7	1	30	734	2	1	0	1	0	1
64	4	1226	0	8	0	37	551	2	0	2	0	0	1
66	2	929	1	5	0	30	1355	1	1	1	0	0	1
35	4	1137	0	7	0	25	561	1.5	0	0	0	0	0
38	3	743	0	6	0	25	489	1	1	0	0	0	0
43	3	596	0	5	0	50	752	1	0	0	0	0	0
46	2	803	0	5	0	27	774	1	1	0	1	0	0
46	2	696	0	4	0	30	440	2	1	1	0	0	0
50	2	691	0	6	0	30	549	1	0	2	1	0	0
65	3	1023	0	7	1	30	900	2	1	1	0	1	0

The variables above represent:

$PRICE$	= Selling price of house in thousands of dollars
BDR	= Number of bedrooms
FLR	= Floor space in sq.ft.
FP	= Number of firplaces
RMS	= Number of rooms
ST	= Storm windows (1 if present. 0 if absent)
LOT	= Front footage of lot in feet
TAX	= Annual taxes
BTH	= Number of bathrooms
CON	= Construction (0 if frame, 1 if brick)
GAR	= Garage size (0=no garage, 1=one-car garage, etc.)
CDN	= Condition (1=needs workk, 0 otherwise)
$L1$	= Location (L1=1 if property is in zone A, L1=0 otherwise)
$L2$	= Location (L2=1 if property is in zone B, L2=0 otherwise)

EXERCISE 4

Three stocks A, B, C have the following expected (mean) returns and standard deviations:

	μ	σ
A	0.20	0.08
B	0.10	0.04
C	0.15	0.06

Also, the correlation coefficients are: $\rho_{AB} = 0.5, \rho_{AC} = 0.2$, and $\rho_{BC} = 0.1$.

a. What is the mean return and risk (variance) on a portfolio of $\frac{3}{4}$ `A` and $\frac{1}{4}$ `B`?

b. What is the mean return and risk (variance) on a portfolio of 20% stock `A`, 50% stock `B`, and 30% stock `C`?

c. Consider only stocks `A` and `C`. Find the composition of the minimum risk portfolio. Using many combinations of stocks `A` and `C` construct the portfolio possibilities curve and identify the efficient frontier.