University of California, Los Angeles
Department of Statistics

Statistics 100A **Instructor: Nicolas Christou**

Covariance and correlation

Let random variables X, Y with means μ_X, μ_Y respectively. The covariance, denoted with $cov(X, Y)$, is a measure of the association between X and Y.

Definition:

$$\sigma_{XY} = cov(X, Y) = E(X - \mu_X)(Y - \mu_Y)$$

Note: If X, Y are independent then $E(XY) = (EX)E(Y)$ Therefore $cov(X, Y) = 0$.

Let W, X, Y, Z be random variables, and a, b, c, d be constants,

- Find $cov(a + X, Y)$

- Find $cov(aX, bY)$

- Find $cov(X, Y + Z)$

- Find $cov(aW + bX, cY + dZ)$

- Important:

  $$var(X + Y) = var(X) + var(Y) + 2cov(X, Y)$$

  Proof:

- Find $var(aX + bY)$

- In general: Let $X_1, X_2, \cdots, X_n$ be random variables, and $a_1, a_2, \cdots, a_n$ be constants. Find the variance of the linear combination $Y = a_1X_1 + a_2X_2 + \cdots + a_nX_n$.

- Example: Let X_1, X_2, X_3 be random variables with $EX_1 = 1, EX_2 = 2, EX_3 = -1, var(X_1) = 1, var(X_2) = 3, var(X_3) = 5, cov(X_1, X_2) = -0.4, cov(X_1, X_3) = 0.5, cov(X_2, X_3) = 2$. Let $U = X_1 - 2X_2 + X_3$. Find (a) $E(U)$, and (b) $var(U)$.

However, the covariance depends on the scale of measurement and so it is not easy to say whether a particular covariance is small or large. The problem is solved by standardize the value of covariance (divide it by $\sigma_X \sigma_Y$), to get the so called coefficient of correlation ρ_{XY}.

$$\rho = \frac{cov(X,Y)}{\sigma_X \sigma_Y}, \quad \text{Always, } -1 \leq \rho \leq 1, \text{ (see proof below).}$$

$$cov(X,Y) = \rho \sigma_X \sigma_Y$$

If X, Y are independent then $\cdots$

Show that $-1 \leq \rho \leq 1$:
Let X, Y be random variables with variances σ_X^2, σ_Y^2 respectively. Examine the following random expressions:

$\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}$	$\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}$

Example:
X and Y are random variables with joint probability density function
$f_{XY}(x, y) = x + y, 0 \leq x \leq 1, 0 \leq y \leq 1$. Find $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, cov(X, Y), \rho_{XY}$.

Example:
Let $X_1, X_2, \cdots, X_n$ be independent and identically distributed random variables having variance σ^2. Show that $cov(X_i - \bar{X}, \bar{X}) = 0$.

Portfolio risk and return

An investor has a certain amount of dollars to invest into two stocks (IBM and $TEXACO$). A portion of the available funds will be invested into IBM (denote this portion of the funds with a) and the remaining funds into TEXACO (denote it with b) - so $a+b=1$. The resulting portfolio will be $aX+bY$ where X is the monthly return of IBM and Y is the monthly return of $TEXACO$. The goal here is to find the most efficient portfolios given a certain amount of risk. Using market data from January 1980 to February 2001 we compute that $E(X)=0.010$, $E(Y)=0.013$, $Var(X)=0.0061$, $Var(Y)=0.0046$, and $Cov(X,Y)=0.00062$.
We first want to minimize the variance of the portfolio. This will be:

$$\begin{aligned} &\text{Minimize } \ \mathrm{Var}(aX+bY) \\ &\quad \text{subject to } \ a+b=1 \end{aligned}$$

Or

$$\begin{aligned} &\text{Minimize } \ a^2Var(X)+b^2Var(Y)+2abCov(X,Y) \\ &\qquad\qquad\qquad\qquad\qquad \text{subject to } \ a+b=1 \end{aligned}$$

Therefore our goal is to find a and b, the percentage of the available funds that will be invested in each stock. Substituting $b=1-a$ into the equation of the variance we get

$$a^2Var(X)+(1-a)^2Var(Y)+2a(1-a)Cov(X,Y)$$

To minimize the above exression we take the derivative with respect to a, set it equal to zero and solve for a. The result is:

$$a=\frac{Var(Y)-Cov(X,Y)}{Var(X)+Var(Y)-2Cov(X,Y)}$$

and therefore

$$b=\frac{Var(X)-Cov(X,Y)}{Var(X)+Var(Y)-2Cov(X,Y)}$$

The values of a and b are:

$$a=\frac{0.0046-0.0062}{0.0061+0.0046-2(0.00062)} \Rightarrow a=0.42.$$

and $b=1-a=1-0.42 \Rightarrow b=0.58$. Therefore if the investor invests 42% of the available funds into IBM and the remaining 58% into $TEXACO$ the variance of the portfolio will be minimum and equal to:

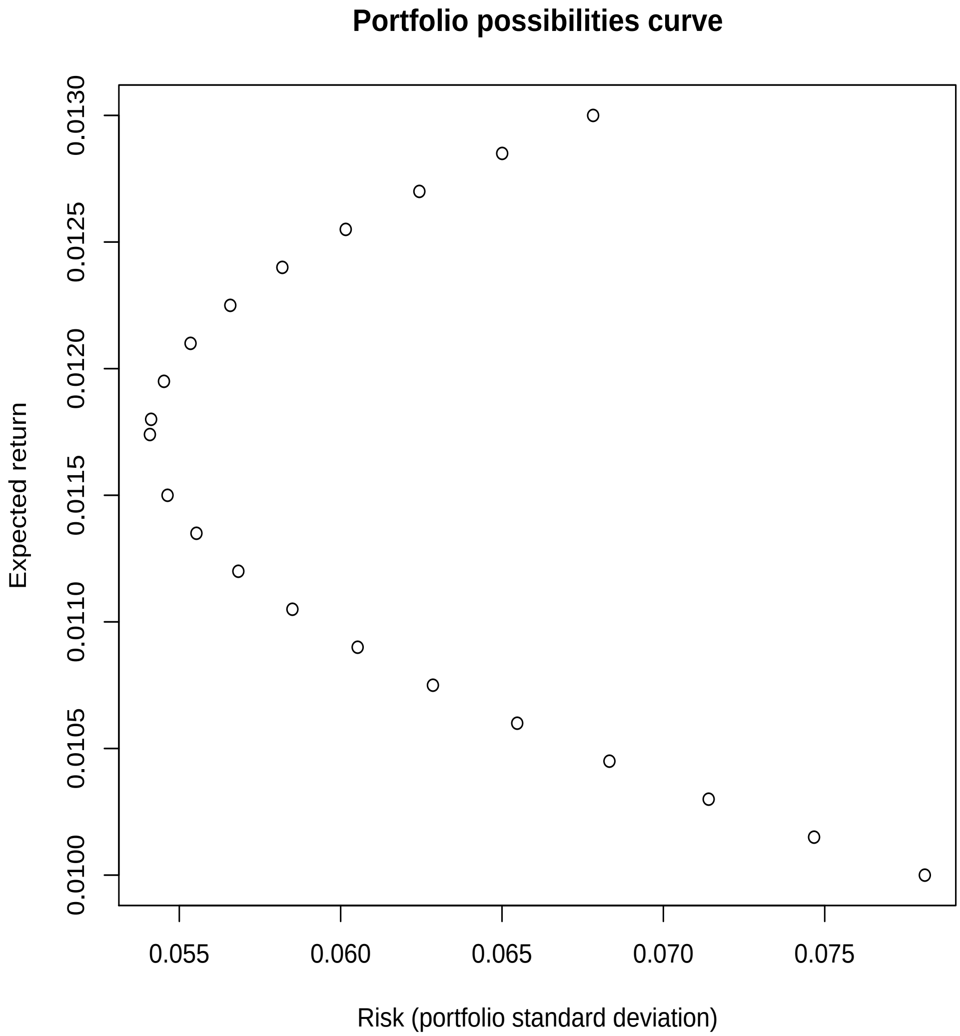
$$Var(0.42X+0.58Y)=0.42^2(0.0061)+0.58^2(0.0046)+2(0.42)(0.58)(0.00062)=0.002926.$$

The corresponding expected return of this porfolio will be:

$$E(0.42X+0.58Y)=0.42(0.010)+0.58(0.013)=0.01174.$$

We can try many other combinations of a and b (but always $a+b=1$) and compute the risk and return for each resulting portfolio. This is shown in the table below and the graph of return against risk on the next page.

a	b	Risk	Return
1.00	0.00	0.006100	0.01000
0.95	0.05	0.005576	0.01015
0.90	0.10	0.005099	0.01030
0.85	0.15	0.004669	0.01045
0.80	0.20	0.004286	0.01060
0.75	0.25	0.003951	0.01075
0.70	0.30	0.003663	0.01090
0.65	0.35	0.003423	0.01105
0.60	0.40	0.003230	0.01120
0.55	0.45	0.003084	0.01135
0.50	0.50	0.002985	0.01150
0.42	0.58	0.002926	0.01174
0.40	0.60	0.002930	0.01180
0.35	0.65	0.002973	0.01195
0.30	0.70	0.003063	0.01210
0.25	0.75	0.003201	0.01225
0.20	0.80	0.003386	0.01240
0.15	0.85	0.003619	0.01255
0.10	0.90	0.003899	0.01270
0.05	0.95	0.004226	0.01285
0.00	1.00	0.004600	0.01300


Portfolio possibilities curve
Expected return
0.0100
0.0105
0.0110
0.0115
0.0120
0.0125
0.0130
0.055
0.060
0.065
0.070
0.075
Risk (portfolio standard deviation)

Efficient frontier with three stocks

```
> summary(returns)
      ribm                 rxom                 rboeing
 Min.   :-0.2264526   Min.   :-0.5219233   Min.   :-0.34570
 1st Qu.:-0.0515524   1st Qu.:-0.0172273   1st Qu.:-0.04308
 Median :-0.0089916   Median : 0.0007013   Median : 0.01843
 Mean   : 0.0003073   Mean   :-0.0011666   Mean   : 0.01079
 3rd Qu.: 0.0462550   3rd Qu.: 0.0337488   3rd Qu.: 0.07357
 Max.   : 0.3537987   Max.   : 0.2269380   Max.   : 0.17483

> cov(returns)
                ribm        rxom      rboeing
ribm    9.930174e-03 0.001798962 3.020685e-05
rxom    1.798962e-03 0.006743820 1.781462e-03
rboeing 3.020685e-05 0.001781462 8.282167e-03
```