Business Mathematics (BK/IBA) – Quantitative Research Methods I (EBE)
Tutorial 1 – Full solutions

Instruction

In a tutorial session of 2 hours, we will obviously not be able to discuss all questions. Therefore, the following procedure applies:

- we expect students to prepare all exercises in advance;
- we will dicuss only a selection of exercises;
- exercises that were not discussed during class are nevertheless part of the course;
- students can indicate their wish list of exercises to be discussed during the session;
- teachers may invite students to answer questions, orally or on the blackboard.

!!! We further understand that your time is limited, and in particular that your time between lecture and tutorial may be limited. In case you have no time to prepare everything, we kindly advise you to give priority to the exercises that are indicated with the !!! sign.

Summation

Q1 (Sydsæter & Hammond, 4/E, 3.1.1.a)
Evaluate the following sum: $\sum_{i=1}^{10} i$

A1 55

Sol $\sum_{i=1}^{10} i = 1 + 2 + \cdots + 9 + 10 = 55$

Extra It may be interesting to write this as two rows: on top 1 2 3 4 5 and underneath 10 9 8 7 6, add a line under this, and add the sums per column: 11 11 11 11 11. This is a heuristic proof that $\sum_{i=1}^{n} i = \frac{n}{2}(n+1)$, at least when n is even.

!!! Q2 (Sydsæter & Hammond, 4/E, 3.1.1.c)
Evaluate the following sum: $\sum_{m=0}^{5}(2m+1)$

A2 36

Sol $\sum_{m=0}^{5}(2m+1) = (2\times 0+1) + (2\times 1+1) + (2\times 2+1) + (2\times 3+1) + (2\times 4+1) + (2\times 5+1) = 1+3+5+7+9+11 = 36$

Alternative solution:
$\sum_{m=0}^{5}(2m+1) = \sum_{m=0}^{5} 2m + \sum_{m=0}^{5} 1 = 2\sum_{m=0}^{5} m + \sum_{m=0}^{5} 1 = 2\times(0+1+2+3+4+5) + 6\times 1 = 2\times 15 + 6 = 36$

Extra Point out that sums need not start with $n = 1$.

Q3 (Sydsæter & Hammond, 4/E, 3.1.1.e)
Evaluate the following sum: $\sum_{i=1}^{10} 2$

A3 20

Sol $\sum_{i=1}^{10} 2 = 2 + 2 + \cdots + 2 + 2 = 10 \times 2 = 20$

Extra Point out that care should be taken what the result is, even when there is no index i inside the summation.

Q4 (Sydsæter & Hammond, 4/E, 3.1.3.a)
Express this sum in summation notation: $4 + 8 + 12 + 16 + \cdots + 4n$

A4 $4\sum_{i=1}^{n} i$ or equivalently $\sum_{i=1}^{n} 4i$.

Sol This is $4 \times (1 + 2 + \cdots + n)$, which is $4\sum_{i=1}^{n} i$ or equivalently $\sum_{i=1}^{n} 4i$.

Extra Stress that the index may be j, m, r, α, or whatever, but not n.

Q5 (Sydsæter & Hammond, 4/E, 3.1.3.d)
Express this sum in summation notation: $a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}$

A5 $\sum_{k=1}^{n} a_{ik}b_{kj}$

Sol Observe that i and j do not "run". But between i and j, there is an index that runs from 1 to n. So $\sum_{k=1}^{n} a_{ik}b_{kj}$

Extra This is in fact the formula for matrix multiplication: $(\mathbf{AB})_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}$. Matrix multiplication, however, is later in the programme.
Again, the index k can be anything (q, β, but not i, j or n), but the indices i and j and the constant n are given.

Q6 (Sydsæter & Hammond, 4/E, 3.1.3.e)
Express this sum in summation notation: $3x + 9x^2 + 27x^3 + 81x^4 + 243x^5$

A6 $\sum_{i=1}^{5} 3^i x^i$

Sol Observe the sequence 3, 9, 27: it looks like 3^1, 3^2, 3^3. So, $\sum_{i=1}^{5} 3^i x^i$ is the answer.

Extra i can be anything.

!!! Q7 (Sydsæter & Hammond, 4/E, 3.3.1.a)
Expand and compute the following double sum: $\sum_{i=1}^{3}\sum_{j=1}^{4} i \cdot 3^j$

A7 720

Sol Write out one of the sums first: $\sum_{i=1}^{3}(i \cdot 3^1 + i \cdot 3^2 + i \cdot 3^3 + i \cdot 3^4) = (3 + 9 + 27 + 81)\sum_{i=1}^{3} i = 120\sum_{i=1}^{3} i$. Next expand the remaining sum: $120(1 + 2 + 3) = 720$.

Extra The order of expanding the sums may be interchanged.

Q8 (Sydsæter & Hammond, 4/E, 3.3.1.b)
Expand and compute the following double sum: $\sum_{s=0}^{2}\sum_{r=2}^{4}\left(\frac{rs}{r+s}\right)^2$

A8 $5\frac{3113}{3600}$

Sol Write out one of the summations first: $\sum_{s=0}^{2}\left[\left(\frac{2s}{2+s}\right)^2 + \left(\frac{3s}{3+s}\right)^2 + \left(\frac{4s}{4+s}\right)^2\right]$. Next expand the remaining sum: $\left[\left(\frac{0}{2}\right)^2 + \left(\frac{0}{3}\right)^2 + \left(\frac{0}{4}\right)^2\right] + \left[\left(\frac{2}{3}\right)^2 + \left(\frac{3}{4}\right)^2 + \left(\frac{4}{5}\right)^2\right] + \left[\left(\frac{4}{4}\right)^2 + \left(\frac{6}{5}\right)^2 + \left(\frac{8}{6}\right)^2\right]$. This yields $[0 + 0 + 0] + \left[\frac{4}{9} + \frac{9}{16} + \frac{16}{25}\right] + \left[1 + \frac{36}{25} + \frac{64}{36}\right]$. Elaborate: $\frac{4}{9} + \frac{9}{16} + \frac{52}{25} + \frac{64}{36} + 1 = \frac{1600}{3600} + \frac{2025}{3600} + \frac{7488}{3600} + \frac{6400}{3600} + 1 = \frac{17513}{3600} + 1 = 5\frac{3113}{3600}$.

!!! Q9 Expand and compute the following product: $\prod_{i=1}^{3}(2i + 1)$

A9 315

Sol $\prod_{i=1}^{3}(2i + 1) = (2 + 1) \times (4 + 1) \times (6 + 1) = 5 \times 7 \times 9 = 315$

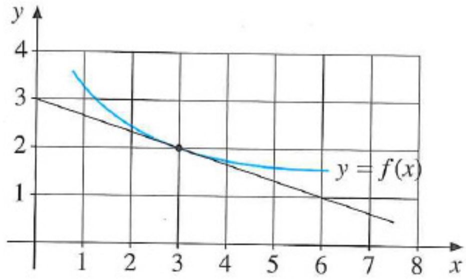
Q10 Expand and compute the following product: $\sum_{j=1}^{3}\prod_{i=1}^{j} ij$

A10 9

Sol $\sum_{j=1}^{3}\prod_{i=1}^{j} i = 1 + 1 \times 2 + 1 \times 2 \times 3 = 1 + 2 + 6 = 9$

Derivatives

Q1 (Sydsæter & Hammond, 4/E, 6.1.1)

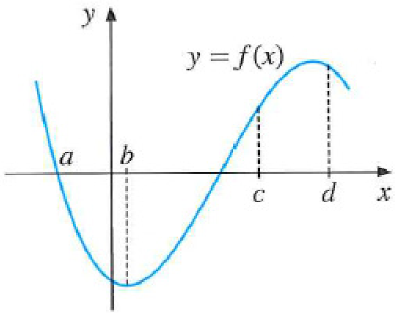


The figure shows the graph of a function f. Find the values of $f(3)$ and $f'(3)$.

A1 $f(3) = 2$; $f'(3) = -\frac{1}{3}$

Sol $f(3) = 2$ (easy to read of); the tangent has slope $f'(3) = -\frac{1}{3}$ because the straight line has $\Delta y = -1$ when $\Delta x = 3$. E.g., the line runs from $(0,3)$ to $(6,1)$, so $\frac{\Delta y}{\Delta x} = \frac{1-3}{6-0} = \frac{-2}{6} = -\frac{1}{3}$

!!! Q2 (Sydsæter & Hammond, 4/E, 6.2.7)



The figure shows the graph of a function f. Determine the sign of the derivative $f'(x)$ at each of the four points a, b, c, and d.

A2 $f'(a) < 0$, $f'(b) = 0$, $f'(c) > 0$, $f'(d) < 0$

Q3 (Sydsæter & Hammond, 4/E, 6.4.1)
Let $C(x) = x^2 + 3x + 100$ be the cost function of a firm. Show that the average per unit rate of change when x is changed from 100 to $100 + h$ is

$$\frac{C(100+h) - C(100)}{h} = 203 + h \qquad (h \neq 0)$$

Sol $\frac{C(100+h)-C(100)}{h} = \frac{(100+h)^2+3(100+h)+100-(100^2+3\times100+100)}{h} = \frac{203h+h^2}{h} = \frac{(203+h)h}{h} = 203 + h$

$\Rightarrow C'(100) = \lim_{h\to0} \frac{C(100+h)-C(100)}{h} = \lim_{h\to0}(203 + h) = 203$

Using the basic derivatives, we find $C'(x) = 2x + 3 \Rightarrow C'(100) = 203$

Extra Mind that many students have never seen the limit concept and notation. However, it can be made clear intuitively.

!!! Q4 (Sydsæter & Hammond, 4/E, 6.6.3.g)
Find the derivative of the following: $\frac{3}{\sqrt[3]{x}}$

A4 $-\frac{1}{x\sqrt[3]{x}}$

Sol $f(x) = \frac{3}{\sqrt[3]{x}} = \frac{3}{x^{1/3}} = 3x^{-1/3}$

$\Rightarrow f'(x) = 3\left(-\frac{1}{3}\right)x^{-1/3-1} = -x^{-4/3} = -\frac{1}{x\sqrt[3]{x}}$

Extra	Note that changing $\sqrt{x}$ and $\frac{1}{x}$ etc. into $x^{1/2}$ and x^{-1} decreases the risk of errors.

Q5	(Sydsæter & Hammond, 4/E, 6.6.4.c)
	Compute the following: $\frac{d}{dA}\left(\frac{1}{A^2\sqrt{A}}\right)$
A5	$\frac{-5}{2A^3\sqrt{A}}$
Sol	$\frac{d}{dA}\left(\frac{1}{A^2\sqrt{A}}\right)=\frac{d}{dA}\left(\frac{1}{A^2A^{1/2}}\right)=\frac{d}{dA}\left(\frac{1}{A^{5/2}}\right)=\frac{d}{dA}\left(A^{-5/2}\right)=-\frac{5}{2}A^{-7/2}=-\frac{5}{2}\frac{1}{A^{7/2}}=-\frac{5}{2}\frac{1}{A^3A^{1/2}}=\frac{-5}{2A^3\sqrt{A}}$
Extra	Some students find it more difficult to differentiate with respect to A than wit respect to x. Please do not substitute A by x, then do the exercise with $\frac{d}{dx}$, and then move back to A. You should really get used to using other symbols than x and y.

Q6	(Sydsæter & Hammond, 4/E, 6.7.3.e)
	Differentiate the function defined by the formula: $\frac{x+1}{x^5}$
A6	$-\frac{4x+5}{x^6}$
Sol	$f(x)=\frac{x+1}{x^5}$. $f'(x)=\frac{1\cdot x^5-5x^4(x+1)}{(x^5)^2}=\frac{-4x^5-5x^4}{x^{10}}=\frac{x^4(-4x-5)}{x^{10}}=\frac{-4x-5}{x^6}=-\frac{4x+5}{x^6}$.
	Alternative way, without introducing f: $\frac{d}{dx}\left(\frac{x+1}{x^5}\right)=\cdots=-\frac{4x+5}{x^6}$.
Extra	The result can also be written as $-4x^{-5}-5x^{-6}$

!!!	Q7	(Sydsæter & Hammond, 4/E, 6.7.7.b)
		Find the equations for the tangents to the graph of the following function at the specified point: $y=\frac{x^2-1}{x^2+1}$ at $x=1$
	A7	$y=x-1$
	Sol	$y=f(x)=\frac{x^2-1}{x^2+1}$
		$\Rightarrow \frac{dy}{dx}=f'(x)=\frac{2x(x^2+1)-2x(x^2-1)}{(x^2+1)^2}=\frac{4x}{(x^2+1)^2}$.
		At $x=1$, $f'(1)=\frac{4}{4}=1$.
		At $x=1$, we have $y=f(1)=\frac{0}{2}=0$. Tangent in point (1,0).
		Equation for tangent: $y=ax+b$, with $a=1$. Take the point (1,0): $0=1+b \Rightarrow b=-1$.
		The equation of the tangent is $y=x-1$.

!!!	Q8	If $D(P)$ denotes the demand for a product when the price per unit is P, then the revenue function $R(P)$ is given by $R(P)=P\cdot D(P)$. Find an expression for $R'(P)$.
	A8	$R'(P)=D(P)+PD'(P)$
	Sol	$R(P)=P\cdot D(P)\Rightarrow R'(P)=1\cdot D(P)+P\cdot D'(P)=D(P)+PD'(P)$
	Extra	Mind the use of the product rule, and the fact that we leave an unspecified $D'(P)$ in the answer.

Q9	(Sydsæter & Hammond, 4/E, 6.8.3.a)
	Find the derivative of the following function: $y=\frac{1}{(x^2+x+1)^5}$
A9	$f'(x)=-\frac{5(2x+1)}{(x^2+x+1)^6}$
Sol	$y=f(x)=\frac{1}{(x^2+x+1)^5}=(x^2+x+1)^{-5}$
	$\Rightarrow \frac{dy}{dx}=f'(x)=-5(x^2+x+1)^{-6}(2x+1)=-\frac{5(2x+1)}{(x^2+x+1)^6}$

Sol Again, you may omit the introduction of the symbol f by using $\frac{d}{dx}\left(\frac{1}{(x^2+x+1)^5}\right) = \cdots$.

Q10 (Sydsæter & Hammond, 4/E, 6.8.6)
Compute dx/dp for the demand function $x = b - \sqrt{ap - c}$, where a, b, and c are positive constants, while x is the number of units demanded, and p is the price per unit, with $p > c/a$.

A10 $\frac{dx}{dp} = -\frac{a}{2\sqrt{ap-c}}$.

Sol $x = f(p) = b - \sqrt{ap - c} = b - (ap - c)^{\frac{1}{2}}$
$\Rightarrow \frac{dx}{dp} = f'(p) = 0 - \frac{1}{2}(ap - c)^{-\frac{1}{2}}a = -\frac{a}{2(ap-c)^{\frac{1}{2}}} = -\frac{a}{2\sqrt{ap-c}}$.
Note that $ap - c > 0 \Rightarrow ap > c \Rightarrow p > \frac{c}{a}$, because a, p, and c are positive.

Extra Observe that x is here the dependent variable (a function of p), while in most cases the dependent variable depends on x. That's why we have dx in the numerator, not in the denominator. If you are too much obsessed by always using $y = f(x)$, you would in this case have to rework the equation into $y = b - \sqrt{ax - c}$, find $\frac{dy}{dx} = -\frac{a}{2\sqrt{ax-c}}$, and then rewrite the answer. This is a lot more work, with a high risk of making mistakes!

Q11 (Sydsæter & Hammond, 4/E, 6.8.9)
Suppose that $C = 20q - 4q\left(25 - \frac{1}{2}x\right)^{1/2}$, where q is a constant and $x < 50$. Find dC/dx.

A11 $\frac{dC}{dx} = \frac{q}{\sqrt{25-\frac{1}{2}x}}$.

Sol $C = f(x) = 20q - 4q\left(25 - \frac{1}{2}x\right)^{1/2}$
$\frac{dC}{dx} = f'(x) = 0 - 4q \cdot \frac{1}{2}\left(25 - \frac{1}{2}x\right)^{-1/2}\left(-\frac{1}{2}\right) = \frac{q}{\sqrt{25-\frac{1}{2}x}}$.

Extra Note that this requires $25 - \frac{1}{2}x > 0 \Rightarrow \frac{1}{2}x < 25 \Rightarrow x < 50$.

!!! Q12 (Sydsæter & Hammond, 4/E, 6.10.1.f)
Find the first-order derivative of: $y = x^3e^x$

A12 $\frac{dy}{dx} = (3 + x)x^2e^x$

Sol $y = f(x) = x^3e^x \Rightarrow \frac{dy}{dx} = f'(x) = 3x^2e^x + x^3e^x = (3x^2 + x^3)e^x = (3 + x)x^2e^x$

Q13 (Sydsæter & Hammond, 4/E, 6.10.3.d)
Find the first and second-order derivatives of: $y = 5e^{2x^2-3x+1}$

A13 $\frac{dy}{dx} = (20x - 15)e^{2x^2-3x+1}$
$\frac{d^2y}{dx^2} = 5[16x^2 - 24x + 13]e^{2x^2-3x+1}$

Sol $y = f(x) = 5e^{2x^2-3x+1}$
$\Rightarrow \frac{dy}{dx} = f'(x) = 5e^{2x^2-3x+1}(4x - 3) = (20x - 15)e^{2x^2-3x+1}$
$\Rightarrow \frac{d^2y}{dx^2} = f''(x) = 20e^{2x^2-3x+1} + 5(4x - 3)^2e^{2x^2-3x+1} = [20 + 5(4x - 3)^2]e^{2x^2-3x+1} =$
$[20 + 80x^2 - 120x + 45]e^{2x^2-3x+1} = [80x^2 - 120x + 65]e^{2x^2-3x+1} = 5[16x^2 - 24x +$
$13]e^{2x^2-3x+1}$

Extra Put emphasis on simplifying the result. In an exam set-up part of the credits would go to applying the rules of differentiation, and part to simplifying the result.

Q14 (Sydsæter & Hammond, 4/E, 6.10.7.b)
Differentiate: $y = x2^x$

A14 $\frac{dy}{dx} = (1 + x\ln 2)2^x$

Sol $y = f(x) = x2^x$
$\frac{dy}{dx} = f'(x) = 1 \cdot 2^x + x \cdot 2^x \ln 2 = (1 + x\ln 2)2^x$

!!! Q15 (Sydsæter & Hammond, 4/E, 6.11.1.b)
Compute the first and second-order derivatives of: $y = x^2 - 2\ln x$

A15 $\frac{dy}{dx} = 2x - \frac{2}{x}$
$\frac{d^2y}{dx^2} = 2 + \frac{2}{x^2}$

Sol $y = f(x) = x^2 - 2\ln x$
$\frac{dy}{dx} = f'(x) = 2x - \frac{2}{x}$ $(= 2x - 2x^{-1})$
$\frac{d^2y}{dx^2} = f''(x) = 2 - 2(-1)x^{-2} = 2 + \frac{2}{x^2}$

Indexing

!!! Q1 Suppose we develop a model for regional unemployment in The Netherlands, distinguishing 12 provinces (Groningen, Friesland, etc.). How would you indicate the unemployment numbers in these provinces in an efficient way?

A1 $u_1, u_2, \dots, u_{12}$, where 1 codes for Groningen, 2 for Friesland, etc.

Extra You may further group this in an unemployment vector $\mathbf{u} = (u_1, u_2, \dots, u_{12})$ in a 12-dimensional space.

Q2 A consultant proposes to use the symbols g for unemployment in Groningen, f for unemployment in Friesland, etc. Why is this not a good idea?

A2 Lack of ease, clearity, and flexibility.

Sol There are different reasons.
Lack of ease: coding the total employment would need a cumbersome formula $(f + g + \cdots + l)$ while with indices it would be $\sum_{i=1}^{12} u_i$.
Lack of clarity: it is much clearer to always use the same symbol (e.g., u) for quantities of the same type (e.g., unemployment), and use indices to refer to the regional specification.
Lack of flexibility: if the subdivision changes (e.g., if provinces merge or split), the changes in the model are much smaller.

Q3 Because we are interested in the phenomenon of "brain drain", we wish to subdivide the model into 3 different labour skills: high, medium, and low. What symbols and indices do you propose?

A3 $u_{1,1}, u_{1,2}, u_{13}, u_{2,1}, u_{2,2}, u_{2,3}, \dots u_{12,3}$, where the first index codes for the 12 provinces as above, and the second index codes for labour skill according to 1=high, 2=medium, 3=low.

Extra The order of the indices may be changed, as long as it is done consistently. Also the notation $u_{i,j}$ with $i = 1, \dots, 12$ and $j = 1, \dots, 2$ is a good answer.

Descriptive statistics

Q1 Compute the mean of the data set $(5,6,4,3,0,-4)$.

A1 $2\frac{1}{3}$

	Sol	Use $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{6}(5+6+4+3+0+-4) = \frac{1}{6} \times 14 = \frac{14}{6} = 2\frac{1}{3}$
	Extra	Observe that the data contains a 4 and a -4 which cancel, and need not be included in the calculation. However, this has the danger of dividing by 4 in the end, instead of by 6.
!!!	Q2	Compute the variance of the above data set.
	A2	13.87
	Sol	Use $s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{6-1}\left[\left(5-\frac{7}{3}\right)^2 + \left(6-\frac{7}{3}\right)^2 + \left(4-\frac{7}{3}\right)^2 + \left(3-\frac{7}{3}\right)^2 + \left(0-\frac{7}{3}\right)^2 + \left(-4-\frac{7}{3}\right)^2\right] = \frac{1}{5}[69.33] = 13.87$
	Extra	Observe that the 4 and -4 do not cancel here. That is, both values deviate from the mean, and they both contribute to the variance.
!!!	Q3	Given a data set of size $n = 12$ with mean $\bar{x} = 8$. What is the sum of the data?
	A3	96
	Sol	Use $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \Rightarrow \sum_{i=1}^{n} x_i = n\bar{x} = 8 \times 12 = 96$
	Q4	A data set of a variable measured in euro has a mean of 23 and a variance of 4. What is the standard deviation if we measure the data in cent?
	A4	200 cent
	Sol	The standard deviation in original units is $\sqrt{4} = 2$ euro. This scales linearly to 200 cent. Alternatively, the variance scales in a quadratic way with the data, so it will become 40000 cent^2. The standard deviation is then $\sqrt{40000} = 200$ cent. Observe that more data is given than needed (here: the mean is not used). This will happen more often in our exercises. After all, you should be able to do the correct math in real life also when they have too much information.
	Q5	In which cases does it make sense to compute a correlation coefficient of two data sets x and y? (a) if x is the income of a random sample of married men and y the income of a random sample of married women. (b) if in a random sample of married couples, x is the income of the man and y the income of the woman. (c) if x denotes the stock values of 300 companies in 2012 and y the stock value of the same 300 companies in 2013. (d) if x denotes the stock values of 300 Dutch companies and y the stock value of 300 German companies. There may be 1 correct answer, or more, or less.
	A5	(b) and (c)
	Sol	Correlation only makes sense for paired observations. (b) and (c) are paired, (a) and (d) are not. Therefore, the correlation coefficient is meaningful for (b) and (c).
	Q6	Prove that the mean deviation, defined by $\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})$ always is equal to 0, and therefore does not convey information on the data.
	A6	$\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}) = \frac{1}{n-1}\sum_{i=1}^{n} x_i - \frac{1}{n-1}\sum_{i=1}^{n}\bar{x} = \frac{1}{n-1}\sum_{i=1}^{n} x_i - \frac{1}{n-1}n\bar{x} = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i - n\bar{x}\right) = \frac{1}{n-1}0 = 0$ QED
	Extra	Although we don't expect you to prove most things, this is a good exercise in applying summation and indexed data.

!!! Q7 A data set has a variance of 5 m^2 and a coefficient of variation of 1.2. What is the mean?

A7 1.863 m

Sol Use $CV_x = \frac{s_x}{\bar{x}} = \frac{\sqrt{s_x^2}}{\bar{x}} \Rightarrow \bar{x} = \frac{\sqrt{s_x^2}}{CV_x} = \frac{\sqrt{5}}{1.2} \approx 1.863$ m

Extra Do not forget the unit.

Q8 Given are three data pairs (p, q): $(3, -7)$, $(5,2)$, and $(-12,0)$. Calculate the correlation coefficient.

A8 -0.201

Sol Use $r_{p,q} = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(p_i-\bar{p})(q_i-\bar{q})}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(p_i-\bar{p})^2}\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(q_i-\bar{q})^2}} = \frac{\sum_{i=1}^{n}(p_i-\bar{p})(q_i-\bar{q})}{\sqrt{\sum_{i=1}^{n}(p_i-\bar{p})^2}\sqrt{\sum_{i=1}^{n}(q_i-\bar{q})^2}}$

Determine $\bar{p} = -1.333$ and $\bar{q} = -1.667$. (That should be easy now).

$r_{p,q} = \frac{(3+1.333)(-7+1.667)+(5+1.333)(2+1.667)+(-12+1.333)(0+1.667)}{\sqrt{(3+1.333)^2+(5+1.333)^2+(-12+1.333)^2}\sqrt{(-7+1.667)^2+(2+1.667)^2+(0+1.667)^2}}$

This gives -0.201

Extra A useful check is always if the result is between -1 and 1. Further, observe the convenience of skipping three divisions by $n-1$.

Q9 Two paired data vectors q and h have $n = 82$ has $s_q^2 = 23$, $s_h^2 = 0.45$ and $r_{q,h} = 0.19$. Calculate the covariance.

A9 0.611

Sol Use $r_{q,h} = \frac{s_{q,h}^2}{s_q s_h} \Rightarrow s_{q,h}^2 = r_{q,h} s_q s_h = r_{q,h}\sqrt{s_q^2 s_h^2} = 0.19 \times \sqrt{23 \times 0.45} = 0.611$

Extra The covariance and the correlation coefficent always have the same sign. Further $-1 \leq r \leq 1$ provides a check.

!!! Q10 A table with prices and some other data is given below.

Commodity		28-Dec-01	11-Jul-08	Price of Commodity on 11-Jul-08 if the USD/ EURO exchange rate remained at 0.8912 (28-Dec-01)	Exchange-rate Contribution to the Total Change in Commodity Price	Direction of Real Supply-Demand Fundamentals
Rough Rice	(cents/cwt.)	369.00	1790.00	1,000.91	55.53%	+
Soybeans	(cents/bushel)	421.00	1615.50	903.33	59.62%	+
Corn	(cents/bushel)	209.00	691.00	386.38	63.20%	+
Coffee	(cents/pound)	46.20	142.25	79.54	65.29%	+
Wheat	(cents/bushel)	289.00	830.75	464.53	67.60%	+
Oats	(cents/bushel)	195.75	449.50	251.35	78.09%	+
Cocoa	(USD/mt.)	1,310.00	2912.00	1,628.29	80.13%	+
Sugar #11	(cents/pound)	7.39	13.99	7.82	93.44%	+
Live Cattle	(cents/pound)	68.17	101.20	56.59	135.07%	–
Orange Juice	(cents/pound)	89.10	123.05	68.81	159.78%	–
Lean Hogs	(cents/pound)	57.05	74.65	41.74	186.98%	–
Gold	(USD/troy oz.)	279.00	960.40	537.02	62.13%	+
Crude Oil	(USD/barrel)	19.84	145.66	81.45	51.03%	+

Suppose we have converted these data into a data matrix **X**, ignoring qualitative data, so that, for example, the number in row 1 column 1 is 369.00.

Give a formula to calculate the average price at 11-Jul-08. You don't need to do the calculation.

A10 $\overline{x_2} = \frac{1}{13}\sum_{i=1}^{13} x_{i,2}$

Sol We need to calculate the average of the values of the cells in colum 2 of **X**, hence of the values $x_{1,2}, x_{2,2}, \dots, x_{13,2}$. The formula is then $\overline{x_2} = \frac{1}{13}\sum_{i=1}^{13} x_{i,2}$.

Extra Using n instead of 13 is also fine, of course. By the way, this average is in fact ill-defined, because the data are in incompatible units. Also observe that you must in this case use commas, as x_{132} is unclear.

Q11 Use the same data as above. Give a formula for the correlation coefficient between the prices at 28-Dec-01 and 11-Jul-08.

A11 $r_{x_1,x_2} = \frac{\sum_{i=1}^{13}(x_{i,1}-\overline{x_1})(x_{i,2}-\overline{x_2})}{\sqrt{\sum_{i=1}^{13}(x_{i,1}-\overline{x_1})^2}\sqrt{\sum_{i=1}^{13}(x_{i,2}-\overline{x_2})^2}}$

Extra Observe that we left out three fractions $\frac{1}{13-1}$. Further, the same concptual problem as in Q10 is present: the cells in one column have different units, so r is meaningless. However, the purpose is to demonstrate how to apply the statistical formulas with indexing.

Q12 We have seen in the lectures that the covariance of a data vector **x** and itself is equal to the variance of **x**. What about the correlation coefficient of a data vector **x** and itself?

A12 It is always 1 and hence conveys no information about the data set.

Sol $r_{x,y} = \frac{s_{x,y}^2}{s_x s_y} \Rightarrow r_{x,x} = \frac{s_{x,x}^2}{s_x s_x} = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}} = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2} = 1\ (n > 1)$

Extra It may be good to stress that this means that the result is meaningless because it is a data-independent constant.

!!! Q13 Which statement(s) is/are true?

(a) The mean of a data set is never 0.
(b) The variance of a data set is always positive.
(c) The variance of a data set is never negative.
(d) A large value of the covariance means that all data points fall almost on a straight line.
(e) A large value of the correlation coefficient means that all data points fall almost on a straight line.
(f) The coefficient of variation always falls between 0 and 1.
(g) When the mean of a data set is 0, the coefficient of variation is 0 as well.
(h) When you change measurment units (e.g., from euro to dollar), the coefficient of variation is unaffected.

A13 (c), (e), (h)

Extra The difference between positive (b) and non-negative (c) may be unknown to some. Also observe in (g) that CV is not defined.