EMTH211 Tutorial 11 Solutions

Problem 1
Suppose that we have 50 measurements of a variable $\mathbf{x}$ and 82 measurements of a variable $\mathbf{y}$. If the norm of $\|\tilde{\mathbf{x}}\|$ is 14 and the norm of $\|\tilde{\mathbf{y}}\|$ is 18, which variable has the greater variance?

Solution Both variables have the same variance.

$$var(\mathbf{x}) = \frac{1}{50-1}\|\tilde{\mathbf{x}}\|^2 = \frac{14^2}{49} = \frac{(7^2)(2^2)}{7^2} = 4$$
$$var(\mathbf{y}) = \frac{1}{82-1}\|\tilde{\mathbf{y}}\|^2 = \frac{18^2}{81} = \frac{(9^2)(2^2)}{9^2} = 4$$

Problem 2
Suppose that we have 170 measurements of a variable $\mathbf{x}$ and 145 measurements of a variable $\mathbf{y}$. If $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{y}}$ have the same norm of 26, which variable has greater standard deviation?

Solution

$$sd(\mathbf{x}) = \frac{1}{\sqrt{170-1}}\|\tilde{\mathbf{x}}\| = \frac{1}{\sqrt{170-1}}\|\tilde{\mathbf{y}}\| < \frac{1}{\sqrt{145-1}}\|\tilde{\mathbf{y}}\| = sd(\mathbf{y}).$$

Problem 3
The number of birds observed at a feeder is observed to be smaller on cold days. Is the correlation between number of birds and temperature (°C) positive, negative, or zero? Would the correlation change if we measured temperature in °F?

Solution When the temperature decreases, the number of birds decreases. Both variables tend to go into the same direction. Hence, the correlation is positive. If we change from °C to °F, one of the variables is shifted and multiplied by a positive constant. This does not change the correlation.

Problem 4
If more shipwrecks happen near the shore than further away, is the correlation between distance to the shore and number of shipwrecks positive, negative or zero?

Solution When the distance to the shore decreases, the number of wrecks increases and vice versa. The variables tend to go into opposite directions. Hence, the correlation is negative.

Problem 5
If $2.5cm = 1inch$, how will the standard deviation of some measurements in cm change if we rewrite them in inches? How will the variance of the measurements change?

Solution The standard deviation is in the same unit of measurement as the variable $sd(c\mathbf{x}) = |c|sd(\mathbf{x})$. When $\mathbf{x}$ is in cm, then $1/(2.5)\mathbf{x}$ is in inch. Hence, the standard deviation decreases by the factor $2/5$. The variance will decrease by $(2/5)^2 = 4/25$ since it is the squared standard deviation.

Problem 6
A person's body mass index (BMI) is defined by

$$\text{BMI} = \frac{\text{mass in } kg}{(\text{height in } m)^2}$$

If we measure the mass and BMI of a number of people who are all $1.8m$ tall, what will be the correlation between mass and BMI?

Solution In this case BMI $= 1/(1.8)^2$mass in kg. This is a perfect linear relation with positive slope. The correlation is 1.

Problem 7
If we take a group of people who all weigh $100kg$ and measure their BMI and height, will the correlation between height and BMI be positive, negative or zero?

Solution When the height decreases, the BMI increases and vice versa. The variables tend to go into opposite directions. Hence, the correlation is negative. The correlation will be larger than -1 since BMI and height are not related linearly.

Problem 8
If we measure the mass in pounds instead of kg, how will the correlation change?

Solution The correlation does not change if a variable is multiplied by a positive constant.

Problem 9
If $\mathbf{x}$ is a vector, what is the correlation between $\mathbf{x}$ and $-\mathbf{x}$?

Solution $cor(\mathbf{x}, -\mathbf{x}) = cor(\mathbf{x}, (-1)\mathbf{x}) = sign(-1)cor(\mathbf{x}, \mathbf{x}) = -cor(\mathbf{x}, \mathbf{x}) = -1.$

Problem 10 (Matlab)
Consider the pressure and temperature measurements for a boiler in Table 1. These data consist of n = 6 measurements each of two variables: $\mathbf{x}$ = temperature and $\mathbf{y}$ = pressure.

(i) Compute mean, variance, and standard deviation for $\mathbf{x}$ and $\mathbf{y}$.

(ii) Compute covariance and correlation of $\mathbf{x}$ and $\mathbf{y}$.

(iii) Produce a scatter plot of $\mathbf{y}$ against $\mathbf{x}$, and of $\tilde{\mathbf{y}}$ against $\tilde{\mathbf{x}}$.

Temp (°C)	Pressure (kPa)
0	91
10	95
20	100
30	101
40	107
50	112

Solution

(i)
```
x = [0,10,20,30,40,50]
y = [91,95,100,101,107,112]
mean(x)
mean(y)
var(x)
var(y)
std(x)
std(y)
```

(ii)
```
cov(x,y)
```

Note that this gives you not only $cov(x, y)$ but the covariance matrix

$$\begin{pmatrix} cov(\mathbf{x},\mathbf{x}) & cov(\mathbf{x},\mathbf{y}) \\ cov(\mathbf{y},\mathbf{x}) & cov(\mathbf{y},\mathbf{y}) \end{pmatrix} = \begin{pmatrix} var(\mathbf{x}) & cov(\mathbf{x},\mathbf{y}) \\ cov(\mathbf{x},\mathbf{y}) & var(\mathbf{y}) \end{pmatrix}.$$

```
corr(x',y')
```

The command `corr` makes a difference between row and column vectors. We have to use column vectors to get the right result.

(iii)
```
scatter(x,y)
scatter(x-mean(x),y-mean(y))
```