Solutions week 5

5.1

Programming exercise...

5.2

Assume both have had the possibility of achieving 100 grade points on total, but got grades split into math and physics the following way:

	X		Y
Math	$\frac{50}{80}$	$>$	$\frac{10}{20}$
Physics	$\frac{20}{20}$	$>$	$\frac{70}{80}$
Total	$\frac{70}{100}$	$<$	$\frac{80}{100}$

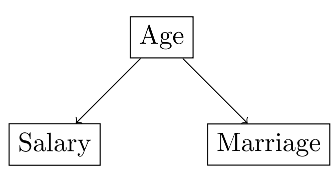
Here X has higher average grades both in math and physics, but X has lower total average grade. (It seems it was easier to get higher grades in physics. And Y has taken more such courses, explaining the result).

5.3

There might be many correct answers for these, here we just give an example of what could be an explanation.

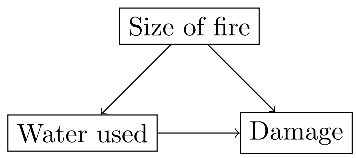
(a)

Age is probably a confounder of being married and salary.



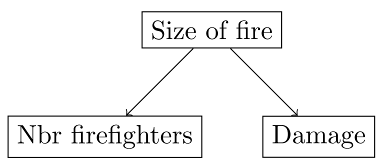
(b)

Size of fire is likely a confounder of water used and damage, but more water can also cause damage.



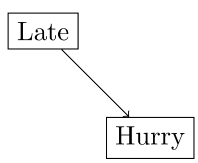
(c)

Size of fire is likely a confounder of number of firefighters and damage.



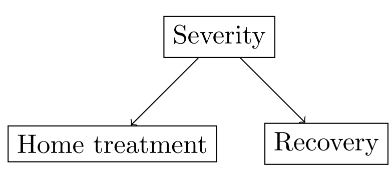
(d)

It is most likely that the causation should be that being late makes you hurry to work.



(e)

Severity of illness might be a confounder here.



(f)

Altitude is also a very strong direct cause of survival, and not taking this into account can lead you to misleading conclusions. Have a look at the picture at the end of the article, and you will see that everyone in the test jumped out from about one meter.

$$
\begin{tikzcd}
& \boxed{\text{Altitude}} \arrow[d] \\
\boxed{\text{Parachute}} \arrow[r] & \boxed{\text{Survival}}
\end{tikzcd}
$$

5.4

(a)

We start with the knowledge that

$$
\begin{aligned}
X \perp\!\!\!\perp Y &\Leftrightarrow p(x,y) = p(x)p(y) \\
p(x \mid y) &= \frac{p(x,y)}{p(y)}
\end{aligned}
$$

With this we can then get the implication in both directions

$$
X \perp\!\!\!\perp Y \Rightarrow p(x,y) = p(x)p(y) \Rightarrow p(x \mid y) = \frac{p(x)p(y)}{p(y)} = p(x)
$$

$$
p(x \mid y) = p(x) \Rightarrow p(x,y) = p(x)p(y) \Rightarrow X \perp\!\!\!\perp Y
$$

(b)

We start with the knowledge that

$$
\begin{aligned}
X \perp\!\!\!\perp Y &\Leftrightarrow p(x,y) = p(x)p(y) \\
\mathbb{E}[f(x,y)] &= \int f(x,y)p(x,y)dxdy
\end{aligned}
$$

With this we then get that

$$
\begin{aligned}
\mathbb{E}[f(x)g(y)] &= \int f(x)g(y)p(x,y)dxdy \\
&= \int f(x)g(y)p(x)(y)dxdy \\
&= \int f(x)p(x)dx \int g(y)p(y)dy \\
&= \mathbb{E}[f(x)]\mathbb{E}[g(y)]
\end{aligned}
$$

(c)

$$
\begin{aligned}
p(x \mid y, z) &= p(x \mid z) && (1)\\
\frac{p(x,y,z)}{p(y,z)} &= \frac{p(x,z)}{p(z)}\\
\frac{p(x,y,z)}{p(x,z)} &= \frac{p(y,z)}{p(z)}\\
p(y \mid x, z) &= p(y \mid z) && (2)\\
\frac{p(x,y,z)}{p(z)} &= \frac{p(x,z)p(y,z)}{p(z)p(z)}\\
p(x, y \mid z) &= p(x \mid z)p(y \mid z) && (3)\\
p(x,y,z)p(z) &= p(x,z)p(y,z) && (4)
\end{aligned}
$$

5.5

(a)

We know that

$$
\begin{aligned}
n_Z, n_X, n_Y &\sim \mathcal{N}(0,1)\\
Z &= n_Z\\
X &= Z + n_X\\
Y &= \theta_{true}X + Z + n_Y\\
\mathbb{E}[n_i n_j] &= 0\\
V(n_i) &= \mathbb{E}[n_i^2] - (\mathbb{E}[n_i])^2\\
\mathbb{E}[n_i^2] &= V(n_i) - (\mathbb{E}[n_i])^2 = \sigma_{n_i}^2 - 0 = 1
\end{aligned}
$$

and then we try to model it using $Y = \theta X + e$ where e is the noise from Z and n_Y. This becomes a problem since X and Z is correlated, and will give the wrong answer from least squares. Let's check how $\hat{\theta}$ relates to θ_{true} if we use normal least square, $X^T X\hat{\theta} = X^T Y$. We assume with many samples in X and Y we converge towards a constant $\hat{\theta}$ and thus can solve the expected values of $X^T X$ and $X^T Y$ separately and get $\hat{\theta}$ from them.

$$
\begin{aligned}
\mathbb{E}[X^T X\hat{\theta}] &= \mathbb{E}\left[\sum_i^N x_i^2\hat{\theta}\right] = \hat{\theta}\mathbb{E}\left[\sum_i^N x_i^2\right] = \hat{\theta}\sum_i^N \mathbb{E}[x_i^2] \\
&= \hat{\theta}N\mathbb{E}[(n_X + n_Z)^2] \\
&= \hat{\theta}N\mathbb{E}[n_X^2 + n_Z^2 + 2n_X n_Z] \\
&= \hat{\theta}N\left(\mathbb{E}[n_X^2] + \mathbb{E}[n_Z^2] + 2\mathbb{E}[n_X n_Z]\right) \\
&= \hat{\theta}N(\sigma_{n_X}^2 + \sigma_{n_Z}^2 + 0) = 2N\hat{\theta}
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}[X^T Y] &= \mathbb{E}\left[\sum_i^N x_i y_i\right] = \sum_i^N \mathbb{E}[x_i y_i] \\
&= N\mathbb{E}[(n_X + n_Z)(\theta_{true}(n_X + n_Z) + n_Z + n_Y)] \\
&= N\mathbb{E}[\theta_{true}(n_X^2 + n_Z^2 + 2n_X n_Z) + n_X n_Z + n_X n_Y + n_Z^2 + n_Z n_Y] \\
&= N\left(\theta_{true}\mathbb{E}[n_X^2] + (\theta_{true} + 1)\mathbb{E}[n_Z^2]\right) \\
&= N(\theta_{true}\sigma_{n_X}^2 + (\theta_{true} + 1)\sigma_{n_Z}^2) = N(2\theta_{true} + 1)
\end{aligned}
$$

If we now want these to be equal we get $2\hat{\theta} = 2\theta_{true} + 1$ so for $\theta_{true} = 3$ we get $\hat{\theta} = 3.5$ instead, and we see that we will not manage to estimate it correctly since the least squares assumes independent noise which is not true here.

(b)

For this problem we have

$$
Y = \theta_1 X + \theta_2 Z + e = [X\ Z]\begin{bmatrix}\theta_1 \\ \theta_2\end{bmatrix} + e
$$

which then has the least square solution

$$
\begin{bmatrix}X^T \\ Z^T\end{bmatrix}[X\ Z]\begin{bmatrix}\hat{\theta}_1 \\ \hat{\theta}_2\end{bmatrix} = \begin{bmatrix}X^T \\ Z^T\end{bmatrix}Y
$$

and the expected value of θ_1 is then found by finding the expected value of each side. We start with the left equation:

$$\begin{aligned}
\mathbb{E}\left[\begin{bmatrix} X^TX & X^TZ \\ Z^TX & Z^tZ \end{bmatrix}\begin{bmatrix}\hat\theta_1 \\ \hat\theta_2\end{bmatrix}\right] &= \mathbb{E}\left[\begin{bmatrix} X^TX & X^TZ \\ Z^TX & Z^tZ \end{bmatrix}\right]\begin{bmatrix}\hat\theta_1 \\ \hat\theta_2\end{bmatrix} \\
&= \mathbb{E}\left[\begin{bmatrix} \sum_i^N x_i^2 & \sum_i^N x_iz_i \\ \sum_i^N x_iz_i & \sum_i^N z_i^2 \end{bmatrix}\right]\begin{bmatrix}\hat\theta_1 \\ \hat\theta_2\end{bmatrix} \\
&= \begin{bmatrix} N\mathbb{E}[x_i^2] & N\mathbb{E}[x_iz_i] \\ N\mathbb{E}[x_iz_i] & N\mathbb{E}[z_i^2] \end{bmatrix}\begin{bmatrix}\hat\theta_1 \\ \hat\theta_2\end{bmatrix} \\
&= \begin{bmatrix} N(\sigma_{n_Z}^2 + \sigma_{n_X}^2) & N\sigma_{n_Z}^2 \\ N\sigma_{n_Z}^2 & N\sigma_{n_Z}^2 \end{bmatrix}\begin{bmatrix}\hat\theta_1 \\ \hat\theta_2\end{bmatrix} \\
&= N\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix}\hat\theta_1 \\ \hat\theta_2\end{bmatrix}
\end{aligned}$$

And then the right equation:

$$\begin{aligned}
\mathbb{E}\left[\begin{bmatrix} X^TY \\ Z^TY \end{bmatrix}\right] &= \begin{bmatrix} N\mathbb{E}[x_iy_i] \\ N\mathbb{E}[z_iy_i] \end{bmatrix} \\
&= \begin{bmatrix} N\mathbb{E}[(\theta_{true}+1)n_Z^2 + \theta_{true}n_X^2] \\ N\mathbb{E}[(\theta_{true}+1)n_Z^2] \end{bmatrix} \\
&= N\begin{bmatrix} (\theta_{true}+1)\sigma_{n_Z}^2 + \theta_{true}\sigma_{n_X}^2 \\ (\theta_{true}+1)\sigma_{n_Z}^2 \end{bmatrix} \\
&= N\begin{bmatrix} 2\theta_{true}+1 \\ \theta_{true}+1 \end{bmatrix} \\
&= N\begin{bmatrix} 7 \\ 4 \end{bmatrix}
\end{aligned}$$

They can now be put together for the solution:

$$\begin{aligned}
N\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix}\hat\theta_1 \\ \hat\theta_2\end{bmatrix} &= N\begin{bmatrix} 7 \\ 4 \end{bmatrix} \\
\begin{bmatrix}\hat\theta_1 \\ \hat\theta_2\end{bmatrix} &= \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}^{-1}\begin{bmatrix} 7 \\ 4 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}
\end{aligned}$$

And we see that now we get the correct answer since the error was actually independent.

5.6

(a)

We look at $X^T X\hat{\theta} = X^T Y$ and find the expected value for both sides.

$$\begin{aligned}
\mathbb{E}\left[X^T X\hat{\theta}\right] &= N\mathbb{E}[n_X^2]\hat{\theta} = N\sigma_{n_X}^2\hat{\theta} \\
\mathbb{E}\left[X^T Y\right] &= N\mathbb{E}[n_X n_Y] = 0 \\
N\sigma_{n_X}^2\hat{\theta} &= 0 \Rightarrow \hat{\theta} = 0
\end{aligned}$$

So here we correctly find that $\hat{\theta} = 0$.

(b)

For this problem we have

$$Y = \theta_1 X + \theta_2 Z + e = [X\ Z]\begin{bmatrix}\theta_1 \\ \theta_2\end{bmatrix} + e$$

and the least square solution

$$\begin{bmatrix}X^T \\ Z^T\end{bmatrix}[X\ Z]\begin{bmatrix}\hat{\theta}_1 \\ \hat{\theta}_2\end{bmatrix} = \begin{bmatrix}X^T \\ Z^T\end{bmatrix} Y$$

The expected value of θ_1 is then found by finding the expected value of each side.

$$\begin{aligned}
\mathbb{E}\left[\begin{bmatrix}X^T X & X^T Z \\ Z^T X & Z^t Z\end{bmatrix}\begin{bmatrix}\hat{\theta}_1 \\ \hat{\theta}_2\end{bmatrix}\right] &= \mathbb{E}\left[\begin{bmatrix}X^T X & X^T Z \\ Z^T X & Z^t Z\end{bmatrix}\right]\begin{bmatrix}\hat{\theta}_1 \\ \hat{\theta}_2\end{bmatrix} \\
&= \begin{bmatrix}N\mathbb{E}[x_i^2] & N\mathbb{E}[x_i z_i] \\ N\mathbb{E}[x_i z_i] & N\mathbb{E}[z_i^2]\end{bmatrix}\begin{bmatrix}\hat{\theta}_1 \\ \hat{\theta}_2\end{bmatrix} \\
&= N\begin{bmatrix}\mathbb{E}[n_X^2] & \mathbb{E}[n_X(n_X + n_Z)] \\ \mathbb{E}[n_X(n_X + n_Z)] & \mathbb{E}[(n_X + n_Y + n_Z)^2]\end{bmatrix}\begin{bmatrix}\hat{\theta}_1 \\ \hat{\theta}_2\end{bmatrix} \\
&= N\begin{bmatrix}\sigma_{n_X}^2 & \sigma_{n_X}^2 \\ \sigma_{n_X}^2 & \sigma_{n_Z}^2 + \sigma_{n_X}^2 + \sigma_{n_Y}^2\end{bmatrix}\begin{bmatrix}\hat{\theta}_1 \\ \hat{\theta}_2\end{bmatrix} \\
&= N\begin{bmatrix}1 & 1 \\ 1 & 3\end{bmatrix}\begin{bmatrix}\hat{\theta}_1 \\ \hat{\theta}_2\end{bmatrix}
\end{aligned}$$

$$
\begin{aligned}
\mathbb{E}\left[\begin{bmatrix} X^T Y \\ Z^T Y \end{bmatrix}\right] &= \begin{bmatrix} N\mathbb{E}[x_i y_i] \\ N\mathbb{E}[z_i y_i] \end{bmatrix} \\
&= \begin{bmatrix} N\mathbb{E}[n_Z n_X] \\ N\mathbb{E}[n_Z(n_X + n_Y + n_Z)] \end{bmatrix} \\
&= N \begin{bmatrix} 0 \\ 1 \end{bmatrix}
\end{aligned}
$$

Which gives us

$$
\begin{aligned}
N \begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} &= N \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\
\begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} &= \begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -0.5 \\ 0.5 \end{bmatrix}
\end{aligned}
$$

5.7

This is for you to think about...