Question 1 (suggested for peer/personal tutorial)

Consider the roll of a fair six-sided die, and let X be the random variable that takes value 1 if the roll is an even number (i.e. $\{2,4,6\}$), and $X=0$ otherwise. Let Y be the random variable that takes the value 1 if the roll is a number greater than 3 (i.e. $\{4,5,6\}$), and $Y=0$ otherwise. Compute $\mathrm{E}\left(Y|X=1\right)$.

Solution to Question 1 (suggested for peer/personal tutorial)

There are two approaches to this problem, the first of which follows the definitions and is detailed, and the second of which takes a shortcut. Both approaches have their advantages, and while the second approach is more direct, it is shorter only because this problem is a special case. It is worthwhile to go through the calculation in the first approach because it is more general. **For tutors, the second method might be better for personal tutorials**.

Solution To Question 1: First method The definition of conditional expectation for discrete random variables means a direct calculation of $\mathrm{E}\left(Y|X=1\right)$ would be

$$\mathrm{E}\left(Y|X=1\right)=\sum_{y\in\{0,1\}} y p_{Y|X}\left(y|1\right)$$

where $p_{Y|X}\left(y|1\right)$ is the conditional probablity mass function. Recall that

$$p_{Y|X}\left(y|x\right)=\mathrm{P}\left(Y=y|X=x\right)=\frac{p_{X,Y}\left(x,y\right)}{p_X\left(x\right)},$$

so when $x=1$:

$$p_{Y|X}\left(y|1\right)=\mathrm{P}\left(Y=y|X=1\right)=\frac{p_{X,Y}\left(1,y\right)}{p_X\left(1\right)}.$$

We need to compute the marginal p.m.f. of X, and the joint p.m.f. of X and Y. First, let's take a step back, and define a few sets.

We start by defining the sample space of rolls as $\Omega=\{1,2,3,4,5,6\}$. Then, we define the sets

$$A=\{\omega\in\Omega : \omega \text{ is even}\}=\{2,4,6\},$$
$$B=\{\omega\in\Omega : \omega>3\}=\{4,5,6\}.$$

We can then define the random variables X and Y as:

$$X\left(\omega\right)=\begin{cases}1, & \text{if } \omega\in A,\\ 0, & \text{if } \omega\notin A.\end{cases}$$
$$Y\left(\omega\right)=\begin{cases}1, & \text{if } \omega\in B,\\ 0, & \text{if } \omega\notin B.\end{cases}$$

So, X and Y are actually indicator variables, i.e. $X=\mathbb{I}_A$ and $Y=\mathbb{I}_B$. Therefore (See Problem Sheet 9, Question 1), the p.m.f. of X can be calculated as

$$p_X\left(x\right)=\begin{cases}\mathrm{P}\left(A\right), & \text{if } x=1,\\ 1-\mathrm{P}\left(A\right), & \text{if } x=0.\end{cases}$$

We can compute $\mathrm{P}\left(A\right)$ naively by looking at the cardinalities of the sets:

$$\mathrm{P}\left(A\right)=\frac{\mathrm{card}(A)}{\mathrm{card}(\Omega)}=\frac{3}{6}=\frac{1}{2}.$$

Therefore,

$$p_X\left(x\right)=\begin{cases}\frac{1}{2}, & \text{if } x=1,\\ \frac{1}{2}, & \text{if } x=0.\end{cases}$$

For the joint p.m.f. of X and Y we need ot look at the intersection of the events A and B, and their complements. For example, when $X = 1$ and $Y = 1$, this corresponds to the event $A \cap B$. Therefore,

$$p_{X,Y}(x,y) = \begin{cases} \mathrm{P}(A^c \cap B^c), & \text{if } x = 0, y = 0, \\ \mathrm{P}(A^c \cap B), & \text{if } x = 0, y = 1, \\ \mathrm{P}(A \cap B^c), & \text{if } x = 1, y = 0, \\ \mathrm{P}(A \cap B), & \text{if } x = 1, y = 1. \end{cases}$$

where all the probabilities are computed by looking at the cardinalities of the sets, e.g.

$$\mathrm{P}(A^c \cap B) = \frac{\mathrm{card}(A^c \cap B)}{\mathrm{card}(\Omega)} = \frac{\mathrm{card}(\{1,3,5\} \cap \{4,5,6\})}{\mathrm{card}(\Omega)} = \frac{\mathrm{card}(\{5\})}{\mathrm{card}(\Omega)} = \frac{1}{6}$$

Therefore,

$$p_{X,Y}(x,y) = \begin{cases} \frac{1}{3}, & \text{if } x = 0, y = 0, \\ \frac{1}{6}, & \text{if } x = 0, y = 1, \\ \frac{1}{6}, & \text{if } x = 1, y = 0, \\ \frac{1}{3}, & \text{if } x = 1, y = 1. \end{cases}$$

We can use the joint p.m.f. to compute the conditional p.m.f.

$$p_{Y|X}(y|1) = \begin{cases} \dfrac{p_{X,Y}(1,0)}{p_X(1)}, & \text{if } y = 0, \\ \dfrac{p_{X,Y}(1,1)}{p_X(1)}, & \text{if } y = 1. \end{cases}$$

$$= \begin{cases} \dfrac{(1/6)}{(1/2)}, & \text{if } y = 0, \\ \dfrac{(1/3)}{(1/2)}, & \text{if } y = 1. \end{cases}$$

and therefore,

$$p_{Y|X}(y|1) = \begin{cases} \dfrac{1}{3}, & \text{if } y = 0, \\ \dfrac{2}{3}, & \text{if } y = 1. \end{cases}$$

It was a lot of effort to obtain this conditional probability mass function, but we can now compute the conditional expectation.

$$\begin{aligned} \mathrm{E}(Y|X=1) &= \sum_{y \in \{0,1\}} y p_{Y|X}(y|1) \\ &= 0 \cdot p_{Y|X}(0|1) + 1 \cdot p_{Y|X}(1|1) \\ &= 0 \cdot \frac{1}{3} + 1 \cdot \frac{2}{3} \\ \Rightarrow \mathrm{E}(Y|X=1) &= \frac{2}{3} \end{aligned}$$

Solution To Question 1: Second method The second solution takes a shortcut, by noticing that the summation for the conditional expectation is over $y \in \{0, 1\}$. Therefore, we first calculate

$$\begin{aligned}
\mathrm{E}\left(Y|X=1\right) &= \sum_{y\in\{0,1\}} y p_{Y|X}\left(y|1\right) \\
&= 0 \cdot p_{Y|X}\left(0|1\right) + 1 \cdot p_{Y|X}\left(1|1\right) \\
&= p_{Y|X}\left(1|1\right) \\
&= \mathrm{P}\left(Y=1|X=1\right).
\end{aligned}$$

So, instead of computing the whole p.m.f. $p_{Y|X}\left(x|y\right)$, we only need to compute the conditional probability $\mathrm{P}\left(Y=1|X=1\right)$. We recall Remark 5.2.1 in Prof. Veraart's notes, that when Ω is finite and the naive (classical) interpretation of probability can be used, we have

$$\mathrm{P}\left(B|A\right) = \frac{\mathrm{P}\left(B\cap A\right)}{\mathrm{P}\left(A\right)} = \frac{\frac{\mathrm{card}(B\cap A)}{\mathrm{card}(\Omega)}}{\frac{\mathrm{card}(A)}{\mathrm{card}(\Omega)}} = \frac{\mathrm{card}(B\cap A)}{\mathrm{card}(A)},$$

since the $\mathrm{card}(\Omega)$ terms cancel out. If we define events A and B as in the first solution above,

$$\begin{aligned}
A &= \{\omega \in \Omega : \omega \text{ is even}\} = \{2,4,6\}, \\
B &= \{\omega \in \Omega : \omega > 3\} = \{4,5,6\},
\end{aligned}$$

then we immediately have

$$\mathrm{E}\left(Y|X=1\right) = \mathrm{P}\left(Y=1|X=1\right) = \mathrm{P}\left(B|A\right) = \frac{\mathrm{card}(B\cap A)}{\mathrm{card}(A)} = \frac{\mathrm{card}(\{4,6\})}{\mathrm{card}(\{2,4,6\})} = \frac{2}{3}.$$

Question 2

Suppose that $X_1, X_2, \ldots, X_n$ are independent random variables that follow a $\mathrm{N}\left(\mu, \sigma^2\right)$ distribution, and define $\overline{X} = \frac{1}{n}\sum_{i=1} X_i$ and $S^2 = \frac{1}{n-1}\left(\overline{X} - X_i\right)^2$, as usual. Show that the random variable T, where

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}},$$

can be written in the form

$$T = \frac{U}{\sqrt{V/p}},$$

where

- $U \sim \mathrm{N}\left(0, 1\right)$,
- p is some function of n,
- $V \sim \chi^2_p$, the chi-squared distribution with p degrees of freedom,
- U and V are independent random variables.

Solution to Question 2

We recall from Corollary 1.6.2 in the notes that, given the assumptions above, $X \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$. We also recall from Theorem 1.7.2 in the notes that $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$. These results will inform our approach to rewrite T in the desired form.

We perform several manipulations on the quantity T to obtain

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} = \frac{\overline{X} - \mu}{S/\sqrt{n}} \times \frac{\sqrt{n}}{\sqrt{n}} = \frac{\sqrt{n}\left(\overline{X} - \mu\right)}{S}$$

$$= \frac{\sqrt{n}\left(\overline{X} - \mu\right)}{S} \times \frac{\left(\frac{1}{\sigma}\right)}{\left(\frac{1}{\sigma}\right)} = \frac{\dfrac{\sqrt{n}(\overline{X} - \mu)}{\sigma}}{\dfrac{S}{\sigma}} = \frac{\dfrac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}}{\dfrac{S}{\sigma}} = \frac{\dfrac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\dfrac{S^2}{\sigma^2}}}$$

$$= \frac{\dfrac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\dfrac{S^2}{\sigma^2}}} \times \frac{1}{\left(\sqrt{\dfrac{n-1}{n-1}}\right)} = \frac{\dfrac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\left[\dfrac{(n-1)S^2}{\sigma^2}\right]/(n-1)}}$$

Therefore, by setting

$$U = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}, \qquad V = \frac{(n-1)S^2}{\sigma^2}$$

we have shown that we can write

$$\frac{\overline{X} - \mu}{S/\sqrt{n}} = \frac{U}{\sqrt{V/(n-1)}}.$$

and we can take $p = n - 1$. We still need to show that U and V follow the appropriate distributions.

Since $\overline{X}$ is a linear transformation of the independent normal $X_1, X_2, \ldots, X_n$, the sample mean $\overline{X}$ also follows a normal distribution (Corollary 1.6.2 in the notes). Furthermore, since U is a linear transformation of $\overline{X}$, U also follows a normal distribution. Now, from Proposition 1.2.6, since the $X_1, X_2, \ldots, X_n$ are independent with mean μ and variance σ^2 (since they are i.i.d. $\mathrm{N}\left(\mu, \sigma^2\right)$), we have that $\mathrm{E}\left(\overline{X}\right) = \mu$ and $\mathrm{Var}\left(\overline{X}\right) = \frac{\sigma^2}{n}$. Therefore, we can compute the mean and variance of U:

$$\mathrm{E}\left(U\right) = \mathrm{E}\left(\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right) = \frac{1}{\frac{\sigma}{\sqrt{n}}}\mathrm{E}\left(\overline{X}-\mu\right) = \frac{1}{\frac{\sigma}{\sqrt{n}}}\left[\mathrm{E}\left(\overline{X}\right)-\mu\right] = 0$$

$$\mathrm{Var}\left(U\right) = \mathrm{Var}\left(\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right) = \frac{1}{\left(\frac{\sigma}{\sqrt{n}}\right)^2}\mathrm{Var}\left(\overline{X}-\mu\right) = \frac{n}{\sigma^2}\mathrm{Var}\left(\overline{X}\right) = \frac{n}{\sigma^2}\left(\frac{\sigma^2}{n}\right) = 1.$$

Therefore, U is random variable following a $\mathrm{N}\left(0,1\right)$ distribution.

For V, Theorem 1.7.2 in the notes gives us that $V = \frac{(n-1)S^2}{\sigma^2}$ follows a χ^2_{n-1} distribution. %if $S^2 = \frac{1}{n-1}\left(\overline{X}-X_i\right)^2$, that

Finally, regarding independence, Theorem 1.7.2 also gives us that $V = \frac{(n-1)S^2}{\sigma^2}$ and $\overline{X}$ are independent. Since U is simply a linear transformation of $\overline{X}$, we therefore have that U and V are independent.

Therefore, we have shown that $T = \frac{\overline{X}-\mu}{S/\sqrt{n}}$ can be written as $\frac{U}{\sqrt{V/(n-1)}}$, where $U \sim \mathrm{N}\left(0,1\right)$ and $V \sim \chi^2_{n-1}$ and U and V are independent.

Question 3

Download the dataset `data_week18.csv` (link on Blackboard below problem sheet). This dataset contains 200 observations for each of the random variables X, Y and Z, where the ith row shows the simultaneous measurement of the three variables at time i. Using exploratory data analysis techniques (visualisations),

(a) Investigate whether or not there is any relationship between any of the variables.

(b) Guess the distributions of X, Y and Z.

Solution to Question 3

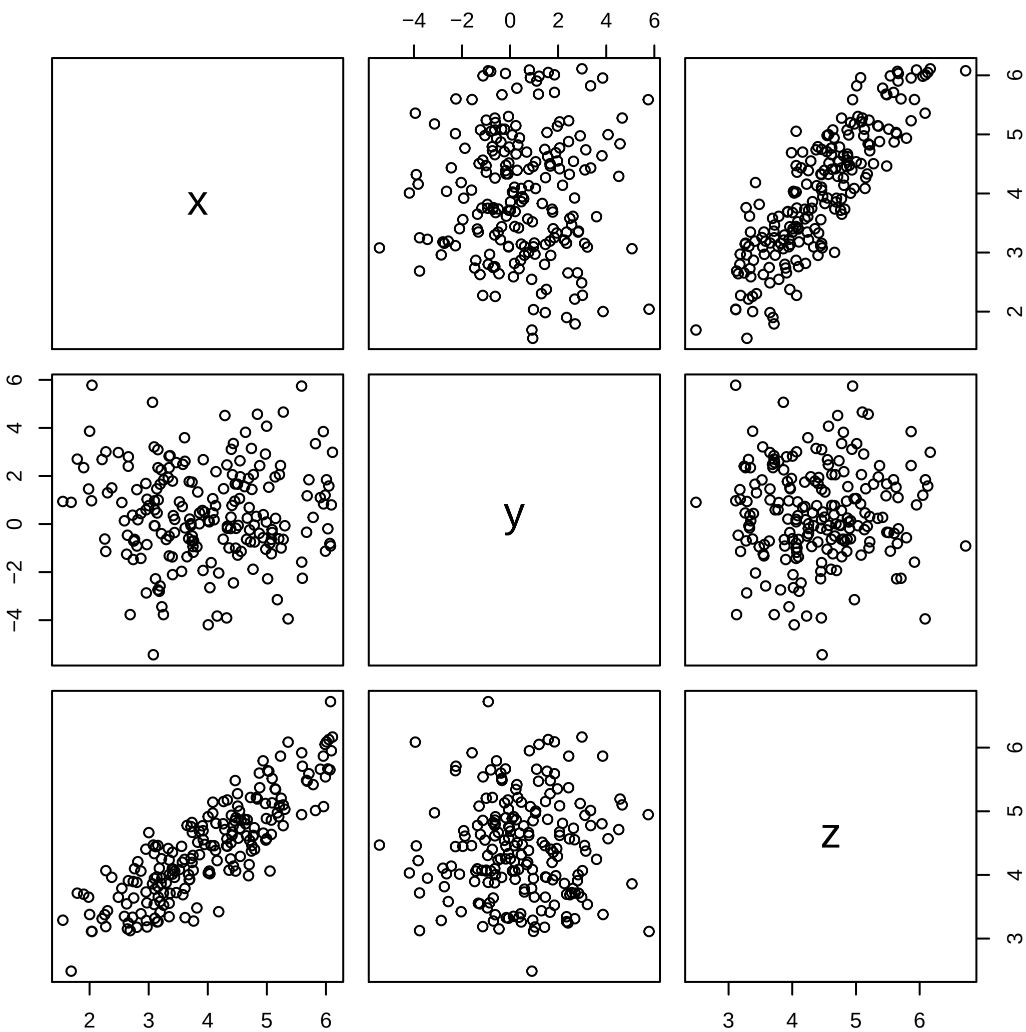
Part (a):

First, we need to read in the data:

```
# read the data into a data frame (see Problem sheet 9)
df <- read.table("data_week18.csv", sep=",", header=T)
```

Now, to investigate any relationships between the random variables `x`, `y` and `z`, since there are exactly 200 observations each, we could plot the random variables against each other, e.g. `x` vs `y`, etc. R has a function called `pairs` which does this in one line and produces the plots in a nice layout:

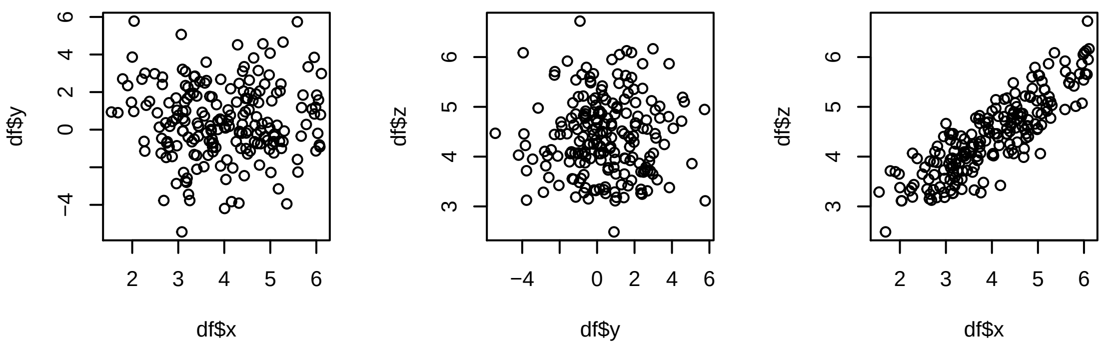
```
# use the 'pairs' function to plot variables against each other
pairs(df)
```



One reads the figures as row vs column; i.e. in the 3×3 matrix of figures above, the (sub)plot in the second row and first column is `y` vs `x`, i.e. `y` on the y-axis and `x` on the x-axis.

Alternatively, one could use the following code to do this 'manually':

```
# create a plot with 1 row and 3 columns
par(mfrow=c(1,3))
plot(x=df$x, y=df$y)
plot(x=df$y, y=df$z)
plot(x=df$x, y=df$z)
```

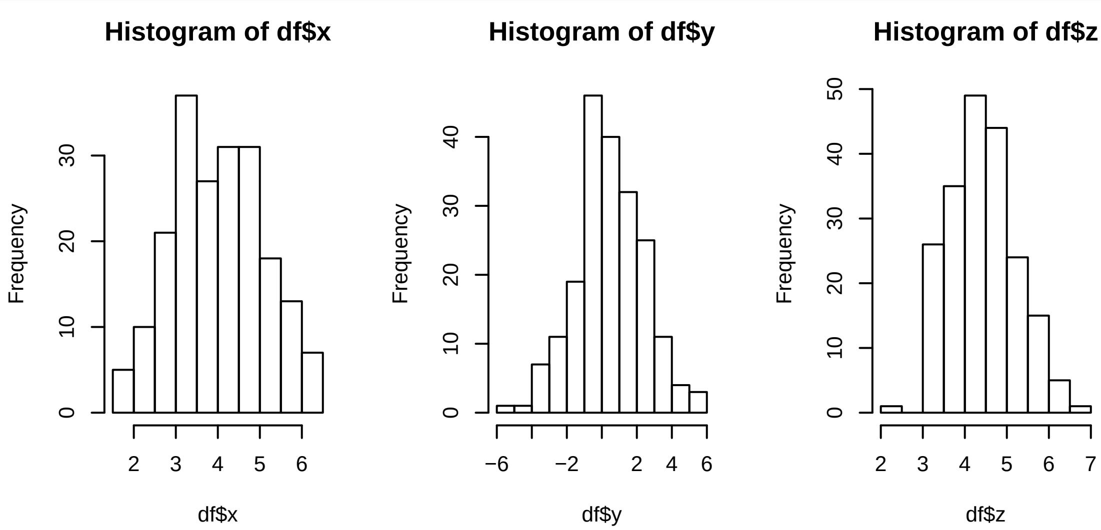


From the figures, there does not seem to be any obvious relationship between `x` and `y`, or between `y` and `z`, since the scatterplots show a somewhat random cloud of points. However, looking at `z` vs `x`, it seems that when a small `x` value was measured at time i, a small `z` value was measured at time i. This suggests there is some relationship between `x` and `z`.

Part (b):

To get an overview of the distribution of a random variable, a histogram is a useful plot. The following code plots histograms of the `x`, `y` and `z`.

```
# create a plot with 1 row and 3 columns
par(mfrow=c(1,3))
hist(df$x)
hist(df$y)
hist(df$z)
```



It is difficult to draw conclusions from these histograms, but they do suggest that the random variables follow normal distributions; perhaps the histogram for `y` most clearly suggests this.

We can compute the mean and standard deviation of these datasets to obtain:

```
x_param <- c(mean(df$x), sd(df$x))
y_param <- c(mean(df$y), sd(df$y))
z_param <- c(mean(df$z), sd(df$z))
df_param <- data.frame(x=x_param, y=y_param, z=z_param)
row.names(df_param) <- c("mean", "sd")
print(df_param)
#>          x     y      z
#> mean 3.999 0.457 4.4206
#> sd   1.074 1.972 0.7801
```

For example, from these values, one might say that Y follows a normal distribution with mean approximately 0.46 and standard deviation approximately 2. However, we cannot be certain.

It might be interesting to see how this data was generated. Here is the code used to generate this data:

```
set.seed(2)
filename <- "data_week18.csv"
n <- 200
rho <- 0.6
x <- rnorm(n, mean=4, sd=1)
y <- rnorm(n, mean=0, sd=2)
z <- rnorm(n, mean=5, sd=1)
z <- (1-rho) * z + rho * x
df <- data.frame(x, y, z)
write.csv(df, file=filename, quote=F, row.names=F)
```

So, in fact X and Y are normal, but Z is a mixture between X and another normal distribution; this explains the relationship between X and Z that we observed in Part (a).