A randomisation model is where the random allocation of treatments to units is incorporated into the statistical model. A simple such model is as follows:

We have $n = 2r$ experimental units. We are going to pick $r$ of them at random, and assign these to the “treatment group”; the remaining $r$ units will be the “control group”.

We assume that each unit has two “potential responses”:

- if unit $i$ is in the control group, then its response will be $u_i$;
- if unit $i$ is in the control group, then its response will be $u_i + \Delta$.

Thus the unknown parameters in the model are the parameter of interest $\Delta$ (the “treatment effect”) and the nuisance parameters $u_1, \ldots, u_n$ (the “unit effects”).

We have yet to introduce random variables into the model. Let the vector $(I_1, \ldots, I_n)$ be randomly picked from the set of all vectors of length $n = 2r$ that have exactly $r$ 1’s and $r$ 0’s; there are $\binom{2r}{r}$ of these. This is the allocation vector; each $I_i \sim B(1, \frac{1}{2})$ but they are not independent.

The following table summarises things:

<table>
<thead>
<tr>
<th>Unit No.</th>
<th>Response if receiving</th>
<th>Allocation Vector</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control</td>
<td>Treatment</td>
</tr>
<tr>
<td>1</td>
<td>$u_1$</td>
<td>$u_1 + \Delta$</td>
</tr>
<tr>
<td>2</td>
<td>$u_2$</td>
<td>$u_2 + \Delta$</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$n$</td>
<td>$u_n$</td>
<td>$u_n + \Delta$</td>
</tr>
</tbody>
</table>

We are going to show via simulation that in a simple case ($n = 8$)

- the mean difference is approximately normally distributed about $\Delta$ and
- its variance can be estimated by $\frac{2\hat{\sigma}^2}{r}$ where $\hat{\sigma}^2$ is the residual mean square (pooled sample variance).

1. First simulate the randomisation model itself:
(a) Define a vector $u$ of length 8 using the last 8 digits of your student number.

(b) Randomly pick 4 numbers from 1, ..., 8 using `ind=sample(8,4)`.

(c) Define `control=u[ind]` and `treat=u[-ind]+5` (note the 5 is $\Delta$ from the general model).

(d) Compute the mean difference $d$ and pooled sample variance $v$.

2. Next we enumerate exhaustively the sampling distribution of the random pair $(d,v)$ (note there are only $\binom{8}{4} = 70$ different choices for the `ind` vector):

   (a) The R function `combinations(n,r)` computes a matrix whose rows contain all subsets of 1, ..., $n$ of size $r$ (seek assistance if R doesn’t seem to know about this function: it’s in the `gtools` library). Thus obtain a matrix `indmat` consisting of all subsets of size 4 from 1, ..., 8.

   (b) Define vectors `dvec=0` and `vvec=0`.

   (c) Using a `for` loop, and your commands for part 1., assign to `dvec[i]` and `vvec[i]` the mean difference and pooled sample variance you would obtain if `ind` in 1(c) is replaced by `indmat[i,]`.

3. The 70 pairs in `cbind(dvec,vvec)` constitute the “population” of equally likely possible outcomes (values of mean difference and residual mean square) under this randomisation model. We wish to show that the mean square is approximately normally distributed around $\Delta$ and that the residual mean square (times $\frac{2}{r}$) provides an unbiased estimate of the variance of the mean difference.

   (a) Find the expected value of the residual mean square under this randomisation model.

   (b) Find the expected value and variance of the mean difference under this randomisation model. Compare to part 3(a) and comment.

   (c) Show using a qq-plot that the distribution of the mean difference under this randomisation model is approximately normal (do not use `qqline()`). Add a line to this plot whose intercept and slope are respectively the mean and sd of the mean difference. Comment on what this plot tells us.