1. The dataset `data7.csv` contains 60 observations with the following variables:

   - `w`: Categorical baseline variable (levels 1,2,3,4)

   - `y`: Response (outcome)

   - `z`: Treatment

   The dataset is sorted by `w` with the following frequencies:

w:	1	2	3	4
	8	12	20	20

   Treatment (`z`) was assigned using permuted blocks of size 4 within each stratum defined by levels of `w`. For simplicity, there are no incomplete blocks (the size of each stratum is a multiple of 4). Let β be the difference between treatment groups.

   (a) Find the estimates of β unadjusted and adjusted (for `w`) and associated standard errors using ordinary least squares (`lm`). Comment on the differences between these two models.

   ```
   > summary(lm(y~z,data=data7))
   ...
   Coefficients:
               Estimate Std. Error t value Pr(>|t|)
   (Intercept)  119.600      6.043  19.790    <2e-16 ***
   z             16.307      8.547   1.908    0.0614 .
   ...
   Residual standard error: 33.1 on 58 degrees of freedom
   ...
   > summary(lm(y~z+w,data=data7))
   ...
   Coefficients:
               Estimate Std. Error t value Pr(>|t|)
   (Intercept)   84.784      9.811   8.642 7.94e-12 ***
   z             16.307      6.730   2.423  0.01871 *
   w2            14.096     11.897   1.185  0.24119
   w3            37.523     10.904   3.441  0.00111 **
   w4            58.468     10.904   5.362 1.68e-06 ***
   ...
   Residual standard error: 26.07 on 55 degrees of freedom
   ...
   ```

   The unadjusted and adjusted estimates of β are identical. This is a consequence of the permuted block randomization that guarantees that the distribution of `w` is identical for the two treatment groups.

   The standard errors for $\widehat{\beta}$ are different—the SE of the adjusted $\widehat{\beta}$ is smaller—because

the adjusted model accounts for the variability explained by `w`, whereas the unadjusted model does not. The additional variability accounted for by the adjusted model is reflected in the smaller residual standard error (26.07 versus 33.1).

(b) Find the size of the reference set for

    i. complete randomization (independent unbiased coin flips),

    ii. random allocation rule (30 subjects assigned `z=0`, the rest `z=1`, ignoring `w`) and

    iii. permuted block randomization (block size 4) stratified by `w`.

    i. $2^{60} = 1.153 \times 10^{18}$

    ii. $\binom{60}{30} = 1.18 \times 10^{17}$

    iii. $6^{15} = 4.70 \times 10^{11}$ (15 blocks, each with 6 possible allocations.)

Each of these is too large enumerate so the exact randomization distribution cannot be easily obtained.

(c) Since the reference sets are too large to enumerate, by randomly sampling from each of the randomization distributions in the part (b), (10,000 should only take a few seconds), calculate the

- variance and standard error of both the adjusted and unadjusted $\widehat{\beta}$.

- two-sided randomization p-values.

Compare these to the models in part (a). In particular, comment on the relative efficiency of adjustment in the analysis versus blocking the randomization.

Using 10,000 replications, I get the following (because this is based on random samples, your answers will vary a little from these.)

The randomization distribution for the unadjusted $\widehat{\beta}$ yields:

```
> betaC <- replicate(10000, {data7$z <- sample(0:1,60, repl=T)
+    lm(y~z,data=data7)$coef[2]})
> var(betaC)
[1] 79.36542
> sd(betaC)
[1] 8.908727
> betaR <- replicate(10000, {data7$z <- sample(rep(0:1,30), 60, repl=F)
+    lm(y~z,data=data7)$coef[2]})
> var(betaR)
[1] 75.98217
> sd(betaR)
[1] 8.716775
```

```
> betaB <- replicate(10000, {data7$z <- c(replicate(15,sample(rep(1:2,2),4)))
+    lm(y~z,data=data7)$coef[2]})
> var(betaB)
[1] 50.01049
> sd(betaB)
[1] 7.071809
```

The SEs for complete randomization and the random allocation rule are comparable to one another and to the variance from the unadjusted model in part (a).

The SE for permuted block randomization is smaller than the other two and comparable to the SE from the adjusted model in part (a).

The randomization distribution for the adjusted $\widehat{\beta}$ yields:

```
> betaC <- replicate(10000, {data7$z <- sample(0:1,60, repl=T)
+    lm(y~z+w,data=data7)$coef[2]})
> var(betaC)
[1] 53.32874
> sd(betaC)
[1] 7.302653
>  betaR <- replicate(10000, {data7$z <- sample(rep(0:1,30), 60, repl=F)
+    lm(y~z+w,data=data7)$coef[2]})
>  var(betaR)
[1] 51.99803
>  sd(betaR)
[1] 7.210966
```

The SEs for the adjusted $\widehat{\beta}$ are comparable to one another and the SE for permuted block randomization and the adjusted model in part (a).

In particular, if we use the adjusted model, it makes little difference which randomization scheme we had used.

2. Suppose that we conduct a randomized trial with a survival outcome and we observe the following times, y_i ("+" indicates censoring):

Group 1: 8, 11, 21+, 35, 38,
Group 2: 37, 41+, 76, 83+, 119

For each failure time, t_j, let R_j be the number of subjects at risk over both treatment groups. Assign each subject a score as follows:

$$
C_i = \begin{cases} \displaystyle\sum_{t_j \leq y_i} \frac{1}{R_j} & \text{if censored} \\ \displaystyle\sum_{t_j \leq y_i} \frac{1}{R_j} - 1 & \text{if not censored} \end{cases}
$$

Calculate a test statistic, $T = \sum_{\text{group 2}} C_i$.

"By hand," we begin by sorting by follow-up times, and we have the following table:

y_i	δ_i	z_i	R_j	C_i		
8	1	1	10	$1/10 - 1$	$=$	-0.900
11	1	1	9	$1/10 + 1/9 - 1$	$=$	-0.789
21	0	1	–	$1/10 + 1/9$	$=$	0.211
35	1	1	7	$1/10 + 1/9 + 1/7 - 1$	$=$	-0.646
37	1	2	6	$\mathbf{1/10 + 1/9 + 1/7 + 1/6 - 1}$	$=$	-0.479
38	1	1	5	$1/10 + 1/9 + 1/7 + 1/6 + 1/5 - 1$	$=$	-0.279
41	0	2	–	$\mathbf{1/10 + 1/9 + 1/7 + 1/6 + 1/5}$	$=$	0.721
76	1	2	3	$\mathbf{1/10 + 1/9 + 1/7 + 1/6 + 1/5 + 1/3 - 1}$	$=$	0.054
83	0	2	–	$\mathbf{1/10 + 1/9 + 1/7 + 1/6 + 1/5 + 1/3}$	$=$	1.054
119	1	2	1	$\mathbf{1/10 + 1/9 + 1/7 + 1/6 + 1/5 + 1/3 + 1/1 - 1}$	$=$	1.054

where δ_i is censoring indicator and z_i is treatment. Note that times 21, 41, 83 are censoring times, so they don't contribute a $1/R_j$. Among subjects in group 2, we have

$$T = 5 \times 1/10 + 5 \times 1/9 + 5 \times 1/7 + 5 \times 1/6 + 4 \times 1/5 + 3 \times 1/3 + 1 \times 1/1 - 3 = 2.4032$$

(Aside: note that overall, we have $\sum_i C_i = 10/10 + 9/9 + 7/7 + 6/6 + 5/5 + 3/3 + 1/1 - 7 = 0$ so as defined, the scores have mean zero.)

In R, suppose we create a `data.frame` called `data1` with variables y, s, z. Sort the dataset by y and create the scores:

```
> data1 <- data1[order(data1$y),]
> data1$R <- 10:1                    ## No ties, otherwise we need to be more clever
> data1$C <- cumsum(data1$s/data1$R) - data1$s
> data1
     y s z  R            C
1    8 1 1 10 -0.90000000
2   11 1 1  9 -0.78888889
4   21 0 1  8  0.21111111
5   35 1 1  7 -0.64603175
6   37 1 2  6 -0.47936508
3   38 1 1  5 -0.27936508
7   41 0 2  4  0.72063492
8   76 1 2  3  0.05396825
9   83 0 2  2  1.05396825
10 119 1 2  1  1.05396825
> Tstat <- sum(data1$C[data1$z==2])
> Tstat
[1] 2.403175
```

(a) Using the random allocation rule (randomly assigning 5 subjects to group 1, the remainder to group 2), find the two-sided randomization p-value for the observed T. Also calculate the variance of the randomization distribution of T, the corresponding chi-square test statistic, and the chi-square p-value.

Create the set of all allocations of treatment to subjects

```
> combs <- combn(10,5)
> dim(combs)
[1]    5 252
> combs[,1:10]    ## first 10 columns
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
[1,]    1    1    1    1    1    1    1    1    1     1
[2,]    2    2    2    2    2    2    2    2    2     2
[3,]    3    3    3    3    3    3    3    3    3     3
[4,]    4    4    4    4    4    4    5    5    5     5
[5,]    5    6    7    8    9   10    6    7    8     9
```

```
> Tstar <- apply(combs, 2, function(i) sum(lrData$C[i]))
> length(Tstar)
[1] 252
> mean(Tstar)  ## Tstar has mean zero (up to round off error)
[1] -1.149736e-16
> mean(abs(Tstar)>=abs(Tstat))  ## proportion at least as large as Tstat
[1] 0.03968254
> var(Tstar)
[1] 1.379371
> Tstat^2/var(Tstar)
[1] 4.18687
> pchisq(Tstat^2/var(Tstar), df=1, lower=F)
[1] 0.04073826
```

The randomization p-value is very close to the p-value from the chi-square approximation.

(b) Compare treatment groups using the log-rank test.

Using `survdiff`,

```
> survdiff(Surv(y,s)~z,data=data1)
...
     N Observed Expected (O-E)^2/E (O-E)^2/V
z=1 5        4      1.6      3.62      5.78
z=2 5        3      5.4      1.07      5.78

 Chisq= 5.8  on 1 degrees of freedom, p= 0.0162
```

"By hand," we have 7 distinct failure times and the following table:

t_j	d_{j1}	n_{j1}	d_{j2}	n_{j2}	R_j	$E[d_{j2}]$	$\text{Var}(d_{j2})$	
8	1	5	0	5	10	$1 \times 5/10$	$1 \times 9 \times 5 \times 5/10^2 \times 9$	$= 0.25000$
11	1	4	0	5	9	$1 \times 5/9$	$1 \times 8 \times 4 \times 5/9^2 \times 8$	$= 0.24691$
35	1	2	0	5	7	$1 \times 5/7$	$1 \times 6 \times 2 \times 5/7^2 \times 6$	$= 0.20408$
37	0	1	1	5	6	$1 \times 5/6$	$1 \times 5 \times 1 \times 5/6^2 \times 5$	$= 0.13889$
38	1	1	0	4	5	$1 \times 4/5$	$1 \times 4 \times 1 \times 4/5^2 \times 4$	$= 0.16000$
76	0	0	1	3	3	$1 \times 3/3$	$1 \times 3 \times 0 \times 3/3^2 \times 2$	$= 0.0$
119	0	0	1	1	1	$1 \times 1/1$	$1 \times 1 \times 0 \times 1/1^2 \times 0$	$= 0.0$
Totals:			3			5.403		0.9998

The log-rank statistic is

$$\frac{(3 - 5.403)^2}{0.9998} = 5.776$$

The variance is smaller than the randomization variance, and the (2-sided) p-value is 0.0162, smaller than the randomization p-value.

(c) Comment on the similarities and differences among these tests.

Except for the sign, the sum of the scores in group 2 is exactly equal to the observed minus expected $(3 - 5.403 = -2.403)$ from the log-rank test. The sum of the $E[d_{k2}]$ column in the log-rank table is exactly the first 7 terms of T in part (a): $5/10 + 5/9 + 5/7 + 5/6 + 4/5 + 3/3 + 1/1$. Therefore, (at least when there are no ties) the log-rank test can be constructed as a test based on scores assigned to each subject in a way similar to the rank scores in the Wilcoxon rank-sum test. (For observations $x_1, x_2, \ldots$, with no ties, the rank for observation x_i can be written as $\sum_{xj \leq x_j} 1$, so the log-rank scores are computed in a similar, although more complex way, than the ranks used in the Wilcoxon test.)

The one-sided randomization p-value is also larger (0.0198) than the one based on the chi-square approximation (0.0081). This is because the randomization p-value is based on a restricted sample space, conditional on the observed data, and in particular exactly 7 failures. The chi-square test is based on the population model which is not conditional on the observed data, and in particular, does not condition on the total number of failures.

For the log-rank, we required the assumption that the contributions from each table were uncorrelated in order to add the variances. By assigning scores to subjects rather than considering each failure time separately, we can conduct valid inference without requiring this assumption, essentially validating the use of the log-rank test.

3. Suppose that we are using minimization as a covariate adaptive allocation scheme and we wish to balance with respect to smoking status and sex. Using the notation from class let $G_t = |x_{11}^t - x_{12}^t| + |x_{21}^t - x_{22}^t|$.

Suppose that the next subject is a non-smoking female and we have:

| | Smoker | | Sex | | |
Group	Y	N	M	F	Total
1	15	26	19	22	41
2	16	28	21	23	44

To which treatment group should the next subject be allocated?

If the next subject is allocated to group 1, the table will be:

| | Smoker | | Sex | | |
Group	Y	N	M	F	Total
1	15	27	19	23	42
2	16	28	21	23	44

so $G_1 = |27 - 28| + |23 - 23| = 1$. If the next subject is allocated to group 2, the table will be:

| | Smoker | | Sex | | |
Group	Y	N	M	F	Total
1	15	26	19	22	42
2	16	29	21	24	44

so $G_1 = |26 - 29| + |22 - 24| = 5$.

Therefore, the next subject should be allocated to group 1.