Instructions: This is a closed-notes, closed-book exam. No calculator or electronic device of any kind may be used. Use nothing but a pen or pencil. Please write your name and answers on blank paper. Please do NOT write your answers on the pages with the questions. For questions that require extensive numerical calculations that you should not be expected to do without a calculator, simply set up the calculation and leave it at that. For example, $(3.45 - 1.67)/\sqrt{2.34}$ would be an acceptable answer. On the other hand, some quantities that are very difficult to compute one way may be relatively easy to compute another way. Part of this exam tests your ability to figure out the easiest way to compute things, based on the information provided and the relationships between various quantities. If you find yourself trying to do exceedingly complex or tedious calculations, there is probably a better way to solve the problem.

1. Consider an experiment with two factors: A and B. Suppose that the levels of factor A are indexed by $i = 1, 2$. Suppose the levels of factor B are indexed by $j = 1, 2$. For $i = 1, 2$ and $j = 1, 2$, let n_{ij} be the number of observations for the treatment combination of level i of factor A and level j of factor B. For $i = 1, 2$ and $j = 1, 2$ and $k = 1, \ldots, n_{ij}$, suppose

$$y_{ijk} = \mu_{ij} + e_{ijk},$$

where the μ_{ij} terms are unknown parameters and the e_{ijk} terms are independent and identically distributed as $N(0, \sigma^2)$. The following table contains response averages and the number of observations for each treatment group.

Level of Factor A	Level of Factor B	Average Response ($\bar{y}_{ij \cdot}$)	Number of Observations (n_{ij})
1	1	3.0	2
1	2	5.0	8
2	1	7.0	6
2	2	3.0	4

Find the Type II sum of squares for factor B.

2. Recall the example field experiment involving genotypes and fertilizer amounts that we discussed over the course of several lectures. The experiment was arranged as a split-plot design. At the whole-plot level, 3 genotypes were randomly assigned to 3 whole-plot experimental units within each of 4 blocks. At the split-plot level, 4 fertilizer amounts were randomly assigned to 4 split-plot experimental units within each whole-plot experimental unit. As was done in our course notes, assume that the response vector y is ordered first by block, then by genotype within each block, and then by fertilizer amount within each combination of block and genotype.

   Our notes included an ANOVA table for the data from such an experiment. The lines of this ANOVA table were ordered in a logical way to correspond to the split-plot experimental design. Provide matrices $X_1, \ldots, X_m$ with nested column spaces (as discussed in our notes on ANOVA) so that the sums of squares in the ANOVA table for our example split-plot experiment can be written in order as $y'(P_2 - P_1)y, \ldots, y'(P_m - P_{m-1})y, y'(I - P_m)y$. As in our course notes, these sequential sums of squares sum to the corrected total sum of squares, and $P_j = X_j(X_j'X_j)^- X_j'$ for all $j = 1, \ldots, m$. Because the matrices are large, it will be best to use Kronecker product notation to write down the matrices $X_1, \ldots, X_m$. You may also use notation like 0_k and 1_k to represent a k-dimensional vector of zeros or ones, respectively.

3. A total of 20 patients suffering from foot pain associated with the disease diabetes participated in an experiment to evaluate two oral medications (OM_1 and OM_2) and two foot treatments (FT_1 and FT_2) intended to relieve foot pain. The two oral medications were randomly assigned to patients using a balanced, completely randomized design, with 10 patients per oral medication. Within each oral medication treatment group, 5 patients were randomly selected to receive foot treatment FT_1 on their left foot and foot treatment FT_2 on their right foot. The other 5 patients in each oral medication treatment group received FT_2 on their left foot and foot treatment FT_1 on their right foot. After six weeks on their assigned oral medication and foot treatment regime, the pain experienced by each patient was measured separately for left and right feet to obtain a dataset with a total of 40 pain measurements. Consider the following R code and output.

```
> #y = vector of pain measurements
> #OM = factor specifying oral medication (levels 1 and 2)
> #FT = factor specifying foot treatment  (levels 1 and 2)
> #foot = factor specifying foot (levels L and R for left foot
> #                               and right foot, respectively)
> #subject = factor specifying the subject (levels 1 to 20)
> #d = data.frame containing the dataset
>
> head(d)
  OM subject foot FT   y
1  1       1    L   1 7.8
2  1       1    R   2 4.2
3  1       2    L   1 5.7
4  1       2    R   2 3.1
5  1       3    L   1 6.1
6  1       3    R   2 2.2
> library(nlme)
> o = lme(y ~ foot * OM * FT, random = ~ 1 | subject, data = d)
> o
Linear mixed-effects model fit by REML
  Data: d
  Log-restricted-likelihood: -44.95243
  Fixed: y ~ foot * OM * FT
  (Intercept)            footR               OM2              FT2
         6.44            -3.08             -1.94             1.22

    footR:OM2       footR:FT2         OM2:FT2
         0.52           -1.32            0.66

    footR:OM2:FT2
         -1.26


Random effects:
 Formula: ~1 | subject
        (Intercept)  Residual
StdDev:     1.13231 0.3941764
```

```
Number of Observations: 40
Number of Groups: 20
```

For the model fit to the data in the `R` code, the best linear unbiased estimate of any cell or marginal mean is equal to the corresponding response data average.

(a) Give the LSMEAN for level OM_1 (level 1 of the oral medication factor).

(b) Give the standard error for the estimate -3.08 that is labeled `footR` in the `R` output.

4. Consider an experiment with a completely randomized design involving three treatment groups. Suppose the sample sizes in the three treatment groups are n_1, n_2, and n_3, respectively. Suppose the n_1 responses for treatment group 1 are

$$y_{11}, \ldots, y_{1n_1} \overset{i.i.d.}{\sim} N(\mu_1, \sigma^2).$$

Suppose the n_2 responses for treatment group 2 are

$$y_{21}, \ldots, y_{2n_2} \overset{i.i.d.}{\sim} N(\mu_2, 4\sigma^2).$$

Suppose the n_3 responses from treatment group 3 are

$$y_{31}, \ldots, y_{3n_3} \overset{i.i.d.}{\sim} N(\mu_3, 9\sigma^2).$$

Furthermore, suppose all responses from all treatment groups are mutually independent. Note that the response variance differs across the three treatment groups.

(a) Provide a fully simplified expression for an unbiased estimator of σ^2 using summation notation. (Partial credit will be given for any unbiased estimator. Full credit will be given for the unbiased estimator that makes the best use of the response data.)

(b) Now suppose $\mu_1 = \mu_2 = \mu_3$, and let $\mu = \mu_1 = \mu_2 = \mu_3$. Provide the best linear unbiased estimator of μ.