1. A large group of ISU chemists and biologists is studying a chemical produced by *Echinacea purpurea* plants. This chemical is one of the main "active" chemicals in Echinacea herbal medicines. Measuring the concentraion of the active chemical requires 3 separate steps: extraction, clarification, and measurment. The data analyzed here was collected by growing 25 plants then separating each plant into three tissues: leaves, flowers, and roots. Each of the 75 biological samples was divided into two parts; each biological part was extracted separately. Each extract was divided into two parts; each extract part was clarified separately. Each clarified sample was measured 2 times. In summary: 25 plants, 75 biological samples, 150 extracts, 300 clarified samples, and 600 measurements. A potential model is:

$$\begin{aligned}
y_{ijklm} &= \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \gamma_{ijk} + \nu_{ijkl} + \epsilon_{ijklm}, \\
i &\in \{1, \ldots, 25\} \text{ identifies plants} \\
j &\in \{L, F, R\} \text{ identifies tissue} \\
k &\in \{1, 2\} \text{ identifies extract within biological sample} \\
l &\in \{1, 2\} \text{ identifies clarified sample within extract} \\
m &\in \{1, 2\} \text{ identifies measurement within clarified sample}
\end{aligned}$$

$$\begin{aligned}
\alpha_i &\sim N(0, \sigma^2_{plants}) \\
\alpha\beta_{ij} &\sim N(0, \sigma^2_{biol.samp}) \\
\gamma_{ijk} &\sim N(0, \sigma^2_{extract}) \\
\nu_{ijkl} &\sim N(0, \sigma^2_{clar.samp}) \\
\epsilon_{ijklm} &\sim N(0, \sigma^2_{measurement})
\end{aligned}$$

(a) 10 pts. Write out the sources of variation and corresponding degrees of freedom for the ANOVA table corresponding to this model

The investigators are not sure about the appropriate model for the random effects. For example, is $\sigma^2_{biol.samp}$ 0 or > 0? One way to evaluate this is to compare a model with the $\alpha\beta_{ij}$ term to a model without that term. Similarly, $\sigma^2_{extract}$ may be 0 or > 0 and $\sigma^2_{clar.samp}$ may be 0 or > 0. There are 8 possible models representing all possible combinations of these three variance components. The investigators also consider two models where plant effects (α_i) are considered fixed effects. Each model was fit using REML. AIC was calculated using k = number of random effect parameters. AIC statistics for each of the 10 models are:

Model	Plants	$\sigma^2_{biol.samp}$	$\sigma^2_{extract}$	$\sigma^2_{clar.samp}$	AIC
1	Random	>0	>0	>0	105.1
2	Random	>0	>0	0	103.2
3	Random	>0	0	>0	110.7
4	Random	>0	0	0	108.9
5	Random	0	>0	>0	115.4
6	Random	0	>0	0	113.5
7	Random	0	0	>0	118.1
8	Random	0	0	0	116.2
9	Fixed	>0	>0	0	98.7
10	Fixed	0	0	0	113.0

(b) 5 pts. Calculate the REML log likelihood for model 1.

(c) 10 pts. Based on the data, which of the 10 models is the most reasonable? Explain your choice.

(d) 5 pts. Are there any other models that might also be reasonable to consider? If so, which model(s)? Explain your choice(s).

2. The following data are based on a study of the effect of video games on reaction time, defined as the time required for an individual to respond to a stimulus. The two treatments are to play a video game (treated group) or to listen to music from that video game (control group). A subject does one of those two treatments for 10 minutes, then their reaction time is measured. That subject then does the other treatment for 10 minutes and their reaction time is measured. The order of the treatments is randomly assiged to each subject. There are a total of 92 subjects in the study. The data are analyzed using the model:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij},$$

where: Y_{ij} is the reaction time for subject i, $i \in \{1, \ldots, 92\}$ and $j \in \{T, C\}$. The response of interest is the ratio of mean reaction times,

$$\theta = \frac{\mu + \bar{\alpha}_. + \beta_T}{\mu + \bar{\alpha}_. + \beta_C}$$

The investigators can estimate this ratio and the variance of this ratio from the data (the details are not important; the estimators are reasonable).

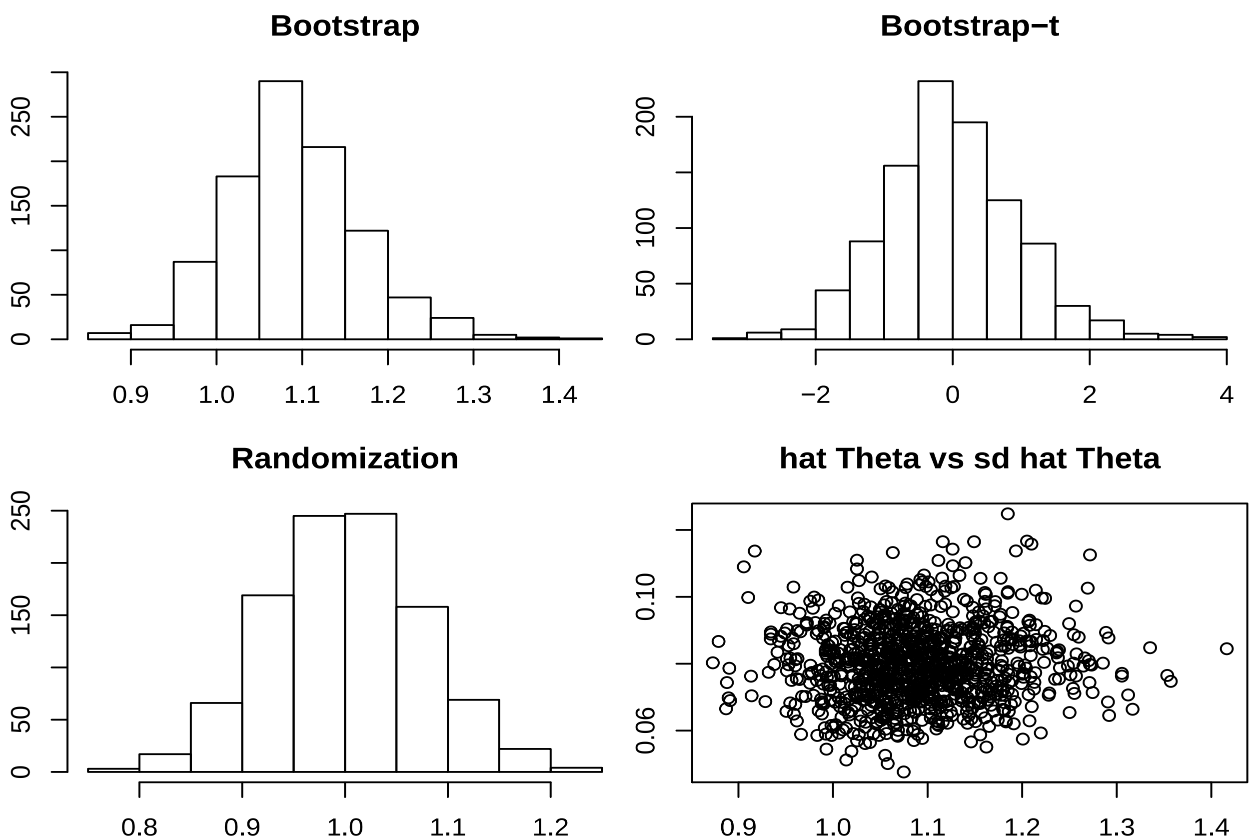
Some summary statistics, calculated from the data are:

$\hat{\theta}$:	1.091
$\sqrt{\widehat{\text{Var}}\,\hat{\theta}}$:	0.079

The investigators use computer-intensive methods for inference on θ. They consider three methods. 1000 values are simulated for each method. The 10 smallest and 10 largest values from the simulated distribution are shown for each method. The three methods are:

- A percentile bootstrap. Subjects are randomly sampled with replacement.
  0.873 0.879 0.887 0.888 0.890 0.890 0.891 0.906 0.910 0.913
  1.291 1.292 1.305 1.305 1.312 1.317 1.335 1.353 1.357 1.416
- A percentile-t bootstrap. Subjects are randomly sampled with replacement.
  -3.075 -2.903 -2.894 -2.738 -2.723 -2.558 -2.521 -2.452 -2.374 -2.336
  2.775 2.804 2.875 2.905 3.106 3.121 3.393 3.423 3.555 3.845
- Randomization. The two treatment labels (T, C) are randomly reassigned to the two response times within each subject.
  0.765 0.765 0.783 0.801 0.808 0.810 0.815 0.820 0.822 0.823
  1.176 1.188 1.196 1.196 1.196 1.198 1.201 1.203 1.214 1.234

Histograms of each distribution and a plot of the sd and mean of the bootstrap samples:



(a) 10 pts. Calculate the endpoints of the 99% confidence interval, using the most appropriate of these three methods.
If this is not possible with the information provided, say what additional information you need.

(b) 10 pts. The investigators also want a p-value for the test of Ho: $\theta = 1$. Calculate the p-value, either exactly or within a region (e.g. < 0.0001), using the most appropriate of these three methods.
If this is not possible with the information provided, say what additional information you need.

(c) 10 pts. A colleague suggests using a different bootstrap method. Their suggested method is sample from the 92 control observatations independently of the 92 treatment observations. Their method gives a 99% confidence interval for θ of (1.001, 1.197). Should you report this confidence interval instead of the one you computed in part 2a? Explain why or why not.

3. In an extenstion of the video game study, the investigators considered six different types of subjects. The quantities of interest are now the mean reaction times for the two treatments in different groups of subjects. They recruited male and female subjects in three different age groups (18-21, 25-35, and 60-70 years old). The study was intended to have 10 subjects in each of the 6 groups and two measurements of reaction time for each subject (one for the "play game" treatment; the second for the "listen" control). Unfortunately, the machinery to measure reaction time sometimes broke down, so some measurements are missing. One possible model for these data is

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \tau_k + \alpha\tau_{ik} + \beta\tau_{jk} + \alpha\beta\tau_{ijk} + \nu_{ijl} + \epsilon_{ijkl},$$

where $i \in \{M, F\}$, $j \in \{18, 25, 60\}$, $k \in \{L, P\}$, and $l \in \{1, \ldots, 10\}$

$$\begin{aligned} \nu_{ijl} &\sim \sigma^2_m \\ \epsilon_{ijkl} &\sim \sigma^2_s \end{aligned}$$

This model was fit to the non-missing observations. Sources of variation, degrees of freedom, Type III Sums of squares, and Expected Mean Squares are:

Source	d.f.	Type III SS	E MS
Sex	1	4.157	$\sigma^2_s + 1.529\sigma^2_m + Q()$
Age	2	46.884	$\sigma^2_s + 1.530\sigma^2_m + Q()$
Sex*Age	2	18.557	$\sigma^2_s + 1.530\sigma^2_m + Q()$
Treatment	1	7.094	$\sigma^2_s + Q()$
Trt*Sex	1	0.174	$\sigma^2_s + Q()$
Trt*Age	2	7.016	$\sigma^2_s + Q()$
Trt*Sex*Age	2	5.508	$\sigma^2_s + Q()$
Subject(Sex*Age)	41	185.747	$\sigma^2_s + 1.683\sigma^2_m$
Residual	28	36.327	σ^2_s

Q() in the expressions for the E MS represents a quadratic function of some or all of the fixed effects.

(a) 5 pts. The experiment was intended to include 120 measurements. How many times did the machinery break down producing a missing measurement of reaction time?

The experiment was intended to include data from 60 subjects. How many times did the machinery break down for both measurements from an individual subject, so that the subject is completely omitted from the analysis?

(b) 5 pts. Given below are the first five non-missing observations in the data set. Write out the rows of the Z matrix corresponding to these five observations. **OMIT any columns of Z for which the first five values in the column are all 0.**

Subject (l)	Sex	Age	Trt	time
1	M	60	Play	6.05
2	M	60	Play	5.55
2	M	60	Listen	5.62
1	F	18	Play	3.03
1	F	18	Listen	3.60

(c) 5 pts. Write out the 5 x 5 block of the $\mathbf{\Sigma}$ matrix corresponding to the five observations given in part 3b.

(d) 5 pts. What is the method-of-moments estimate of σ^2_m?

(e) 5 pts. Calculate the F statistic to test Ho: no interaction between treatment and age.

(f) 5 pts. Using the Cochran-Satterthwaite approximation where appropriate, what are the numerator and denominator degrees of freedom for the test in part 3e

(g) 5 pts. Calculate the F statistic to test Ho: no difference between the two sexes, averaged over treatment and age (i.e. no main effect of sex).

(h) 5 pts. Using the Cochran-Satterthwaite approximation where appropriate, what are the numerator and denominator degrees of freedom for the test in part 3g

4. A student in aerospace engineering is studying the relationship between the curvature of the leading edge of an airplane wing and the lift it produces. They build 6 plastic wings. Two are randomly assigned to have no curvature, two are randomly assigned to have slight curvature and the last two have moderate curvature. The lift generated by each wing is measured 10 times in a wind tunnel, so there are a total of 60 observations.

(a) 10 pts. Write out the skeleton ANOVA table corresponding to a reasonable model for this study. (The skeleton ANOVA table includes sources of variability and the d.f. for each source).

(b) 5 pts. If the investigators want to test the 2 d.f. hypothesis of no effect of curvature, what is the appropriate MS to use as the denominator of the F statistic, i.e. what is the appropriate error term to test this hypothesis?

5. An ISU research center is studying how to efficiently produce fuels from biological material (e.g. corn stalks). The process involves many steps and some complicated equipment. One experiment compared the amount of fuel produced at 3 different temperatures (150, 180, and 200) for one of the key steps in the process. Because there is only one piece of equipment, the investigators can run only one temperature at a time. The investigators know that the equipment gets clogged and dirty over time, so later runs are less efficient than earlier runs. So, they arrange runs into blocks of three runs each. The three temperatures are randomly assigned to runs within a block. They run the experiment 15 times (5 blocks of 3 runs each). Each of these 15 runs starts with material randomly chosen from a batch of corn stalks.

   They then tear apart the equipment, clean it thoroughly, and rebuild it. They then repeat the experiment using the same design (5 blocks of 3 runs each) with a completely new batch of corn stalks. Corn stalks differ in their composition, so this new batch may give different results than the first batch.

   They then tear apart, clean, and rebuild the equipment a third time and repeat the experiment using another 5 blocks of 3 runs each. This repetition of the experiment uses a third batch of corn stalks.

   The investigators are interested in the mean efficiency for each temperature and differences in efficiency between temperatures.

   In a picture (where the numbers are the temperature used in that run), the study is:

Expt	Block 1			Block 2			···	Block 5		
1	150	200	180	180	150	200	...	200	150	180
2	200	180	150	180	200	150	...	150	180	200
3	150	180	200	200	180	150	...	150	200	180

   There are 15 observations in each experiment, for a total of 45 observations.

   (a) 10 pts. Write out the skeleton ANOVA table corresponding to a reasonable model for this study. (The skeleton ANOVA table includes sources of variability and the d.f. for each source).

   (b) 5 pts. If the investigators are interested in broad sense inference about treatment effects, what is the appropriate MS to use as the denominator of the F statistic i.e. what is the appropriate error term for testing Ho: no differences among temperatures)?