1. Split plot experiment

   (a)
   $$Y_{ijkl} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \gamma_{ik} + \epsilon_{ijkl}$$

   where $i, j, k, l = 1, 2$. i indexes temperatures, j indexes catalysts, k indexes growth chamber, and l indexes tube.

   $$\begin{aligned} \gamma_{ik} &\sim (0, \sigma_G^2) \text{ or } N(0, \sigma_G^2) \\ \epsilon_{ijkl} &\sim (0, \sigma_e^2) \text{ or } N(0, \sigma_e^2) \end{aligned}$$

   All random effects are independent.

   (b) Write out the $\mathbf{Z}$ matrix for this study.
   the $\mathbf{Z}$ matrix is

   $$\mathbf{Z} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}_{16 \times 4}$$

   (c) Write out $\mathbf{\Sigma}$, the variance-covariance matrix of the observations, as a 16×16 matrix using the appropriate model parameters.

   $$\Sigma = \begin{pmatrix} B & 0 & 0 & 0 \\ 0 & B & 0 & 0 \\ 0 & 0 & B & 0 \\ 0 & 0 & 0 & B \end{pmatrix}_{16 \times 16}$$

   where 0 is 4×4 zero matrix and

   $$B = \begin{pmatrix} \sigma_e^2 + \sigma_G^2 & \sigma_G^2 & \sigma_G^2 & \sigma_G^2 \\ \sigma_G^2 & \sigma_e^2 + \sigma_G^2 & \sigma_G^2 & \sigma_G^2 \\ \sigma_G^2 & \sigma_G^2 & \sigma_e^2 + \sigma_G^2 & \sigma_G^2 \\ \sigma_G^2 & \sigma_G^2 & \sigma_G^2 & \sigma_e^2 + \sigma_G^2 \end{pmatrix}_{4 \times 4}$$

1

    (d) Write $\boldsymbol{\Sigma}$ using Kronecker product notation Denote $\mathbf{1}_{4\times 1} = (1,1,1,1)^T$, and $I_{p\times p}$ a $p \times p$ identity matrix,
$$\Sigma = \sigma_G^2 I_{4\times 4} \otimes (\mathbf{1}_{4\times 1}\mathbf{1}'_{4\times 1}) + \sigma_e^2 I_{16\times 16}$$

2. Gauge R&R study

    (a)
$$Y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk}$$
where $i, j, k = 1, 2$. i indexes parts, j indexes operators, and k indexes measurements.

$$
\begin{aligned}
\alpha_i &\sim (0, \sigma_p^2) \text{ or } N(0, \sigma_p^2) \\
\beta_j &\sim (0, \sigma_R^2) \text{ or } N(0, \sigma_R^2) \\
\alpha\beta_{ij} &\sim (0, \sigma_I^2) \text{ or } N(0, \sigma_I^2) \\
\epsilon_{ijk} &\sim (0, \sigma_e^2) \text{ or } N(0, \sigma_e^2)
\end{aligned}
$$

    All random variables are assumed to be independent.

    (b) Write out the $\mathbf{Z}$ matrix for this study. The $\mathbf{Z}$ matrix is:

$$\mathbf{Z} = \begin{bmatrix}
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 1
\end{bmatrix}_{8\times 8}$$

    (c) Write out $\boldsymbol{\Sigma}$, the variance-covariance matrix of the observations, as a 8×8 matrix using the appropriate model parameters.
The variance-covariance matrix $\Sigma = ZGZ' + R$ is a 8×8 compound symmetric matrix:

$$\Sigma = \begin{bmatrix}
I & P & R & 0 \\
P & I & 0 & R \\
R & 0 & I & P \\
0 & R & P & I
\end{bmatrix}_{8\times 8}$$

where 0 is 2×2 zero matrix,

$$I = \begin{pmatrix} \sigma_p^2 + \sigma_R^2 + \sigma_I^2 + \sigma_e^2 & \sigma_p^2 + \sigma_R^2 + \sigma_I^2 \\ \sigma_p^2 + \sigma_R^2 + \sigma_I^2 & \sigma_p^2 + \sigma_R^2 + \sigma_I^2 + \sigma_e^2 \end{pmatrix}_{2\times 2},$$

$$P = \begin{pmatrix} \sigma_p^2 & \sigma_p^2 \\ \sigma_p^2 & \sigma_p^2 \end{pmatrix}_{2\times 2},$$

and

$$R = \begin{pmatrix} \sigma_R^2 & \sigma_R^2 \\ \sigma_R^2 & \sigma_R^2 \end{pmatrix}_{2\times 2}.$$

(d) Write Σ using Kronecker product notation

As $\Sigma = ZGZ' + R$ where $Z = (I_{2\times 2} \otimes 1_{4\times 1}, 1_{2\times 1} \otimes (I_{2\times 2} \otimes 1_{2\times 1}), I_{4\times 4} \otimes 1_{2\times 1})$, hence,

$$\Sigma = \sigma_p^2 I_{2\times 2} \otimes (1_{4\times 1} 1'_{4\times 1}) + \sigma_R^2 (1_{2\times 1} 1'_{2\times 1} \otimes (I_{2\times 2} \otimes (1_{2\times 1} 1'_{2\times 1})))$$
$$+ \sigma_I^2 I_{4\times 4} \otimes (1_{2\times 1} 1'_{2\times 1}) + \sigma_e^2 I_{8\times 8}.$$

Note: the Kronecker product question for this model was definitely a hard problem! My hat off to those who got it!

3. Plankton tows

(a) The appropriate model includes spp and tow, both as class variables (i.e. defining groups). The desired test is is the F test for spp. F = 251.7, p < 0.0001.

(b) The estimate is -10.4 with a s.e. of 12.3 and a p-value of 0.40.

(c) The plots are not enclosed. The residual plot shows strong curvature and unequal variance. The obs vs block means plot clearly shows fanning lines. Neither assumption seems reasonable.

(d) The SAS/R code for parts d and e is identical to the above code, but logabdn is used as the response instead of abdn. You could choose either mean abundance or mean log abundance as the X axis in the plot. That's really a minor issue that doesn't affect the conclusions.

The residual plot looks very close to a flat band; the log abundance vs block mean plot looks very close to parallel lines. Both assumptions seem to be reasonable on the log scale.

Note: The above are my interpretations. You get full credit if you have a different interpretation so long as it is reasonable and you provide a good explanation.

(e) On the log scale, the difference is -0.982 with a s.e. of 0.078. A very different conclusion after log transformation.

Note: One way of interpreting differences of log transformed responses is to exponentiate the difference. $\exp(-0.964) = 0.38$. One appropriate interpretation is that the median abundance of species 1 is 38% the median abundance of species 2.

(f) The estimated mean for species 1 is 1.83 with an se of 0.055

(g) With random blocks, the estimated difference on the log scale is -0.982 with a s.e. of 0.078.

Note: Same as the fixed block results

(h) With random blocks, the estimated mean for species 1 is 1.83 with an se of 0.071.

(i) The results (estimate, s.e.) for the the difference are identical. This is because block effects cancel out of the difference. The estimated species mean is identical. All species are present in all blocks, so there is no 'between block' information. The s.e. of the mean is larger in the random block model. That is because the fixed block model makes inferences to this specific set of 12 tows. The random blocks model makes inferences to a set of 12 new tows, which will have different block effects than the current set.

4. sensory evaluation study

(a) est(a)=0.4282 se(a)= 0.3191
Note: if you got 0.3228 with se=0.6382, you used treatment contrasts in R and forgot that the difference in marginal means includes various interaction effects when you use that parameterization. That's wrong.

(b) est(a)= 0.3449 se(a)= 0.2920
Note: if you got 0.3377 with se = 0.2933, you used the ANOVA estimates of variance components in SAS (mixed method=type3). This gives a slightly different estimate of $\sigma_b^2 lock$ and hence a different estimate of the main effect. (This is a situation where the GLS estimator depends on the variances and is not equal to the OLS estimator). Accepted for full credit.

(c) Two possible explanations:
1) This an incomplete block design where each participant(block) only evaluates 4 of the 8 treatments. Block effects don't cancel out of the difference. The estimated A effect isn't identical.
2) The fixed block estimator uses only within block information (difference between treatments in the same block). The random block estimator recovers inter-block information. That estimate is a linear combination of the estimate from the fixed block model and the estimate from between-block information.

(d) Recovery of the 'between block' variation provides a more precise estimate.

(e) efficiency for the difference=$\frac{0.3191^2}{0.2920^2}$=1.194. Slightly different if you used ANOVA estimates of the variance components. Accepted for full credit.
Note: I was expecting you to work with the differences. You could also calculate the efficiency for the estimate of the marginal mean. This is $(0.2174/0.2065)^2 = 1.108$.

(f) n per treatment = 7*1.194 = 8.36
Or 7*1.108 = 7.76.

My R code for problem 3, parts d-h (a-c similar, but with abdn not logabdn):

```
plankton <- read.table('plankton.txt',as.is=T, header=T)

plankton$tow.f <- factor(plankton$tow)
plankton$species.f <- factor(plankton$species)
```

```
plankton$logabdn <- log(plankton$abundance)

p.b <- predict(lm(logabdn~tow.f,data=plankton))
p.m1 <- lm(logabdn ~ tow.f + species.f, data=plankton)

plot(predict(p.m1),resid(p.m1))
plot(p.b, plankton$logabdn, pch=18+as.integer(plankton$species.f))

library(lme4)

p.m2 <- lmer(logabdn ~ species.f + (1|tow.f), data=plankton)

options(contrasts=c('contr.sum', 'contr.poly'))

p.m3 <- lm(logabdn ~ tow.f + species.f, data=plankton)

p.m4 <- lmer(logabdn ~ species.f + (1|tow.f), data=plankton)

# just for fun,
#  look at the relevant bits of the variance-covariance matrix,
#  that is the intercept and the three species effects

vcov(p.m3)[c(1,13:15),c(1,13:15)]

vcov(p.m4)

# the differences among species have the same variance in both models
#   but var(intercept) is 3.7 times larger in the random block model

c.m3 <- c(1, rep(0,11),1,0,0)
  # mu + alpha_1 = species 1 mean in fixed block model.
  # block effects are 0 because sum-to-zero constraint
  # if used treatment contrasts (set first to 0), would need 1/12 for all
  # block effects.  Otherwise, are calculating species 1 mean in block 1

c.m4 <- c(1,1,0,0)  # mu + 2*alpha_1 in random block model

# calculate estimated means:
t(c.m3) %*% coef(p.m3)
t(c.m4) %*% fixef(p.m4)
```

```
# and the se's of the means:
sqrt(t(c.m3) %*% vcov(p.m3) %*% c.m3)
sqrt(t(c.m4) %*% vcov(p.m4) %*% c.m4)
```

My SAS for problem 3, parts a-c:

```
data plankton;
  infile 'plankton.txt'; input tow spp $ abdn; logabdn = log(abdn);
proc glm;
  class tow spp; model abdn = tow spp;
  lsmeans spp / stderr;
  estimate 'diff spp 1 - 2' spp 1 -1 0 0;
  output out=resids r=resid p=yhat;
proc sgplot;
  scatter x=yhat y=resid;
proc sort data=plankton;
  by tow;
proc means noprint data=plankton;
  by tow;  var abdn; output out=means mean=meanabdn;
data new;
  merge plankton means;
  by tow;
proc sgplot;
  scatter x=meanabdn y=abdn;
  run;
proc mixed data=plankton;
  class spp tow;
  model logabdn = spp;
  random tow;
  lsmeans spp;
  estimate 'difference between spp 1 and 2' spp 1 -1;
  run;
```

You could also get the block means as predicted values from a model with only blocks as a
fixed effect. That is, do something like this:

```
data plankton; ...
proc glm;
  class tow; model abdn = tow;
  output out=blockmeans p=bmean;
  run;
proc glm;
```

```
  class tow spp; model abdn = tow spp;
  estimate 'diff spp 1 - 2' spp 1 -1 0 0;
  output out=resids r=resid p=yhat;
proc sgplot;
  scatter x=bmean y=resid;
  run;
```

This is especially useful in R unless you know how to merge data.

Senniang's R code for problem 4.

```
bibd=read.csv("bibd.csv",head=TRUE)
# block and treatment labels are numbers, so not converted to factors by default
options(contrasts=c('contr.sum', 'contr.poly'))
#  if you use sum to zero constraints, the coefficient for a is 1/2 the difference
#     between the two marginal means
#  if you use treatment contrasts, need to include interaction coefficients in the
#     difference.
bibd$blockf=as.factor(bibd$block)
bibd$af=as.factor(bibd$a)
bibd$bf=as.factor(bibd$b)
bibd$cf=as.factor(bibd$c)
bibd.lm=lm(y~blockf+af*bf*cf,data=bibd)
summary(bibd.lm)
2*c(0.21408, 0.15956)

library(lme4)
bibd.mix=lmer(y~af*bf*cf+(1|blockf),data=bibd)
summary(bibd.mix)
eff=0.638225^2/ 0.58399^2
n=eff*length(bibd$a[bibd$a==1])
```