1. New Zealand pasture fertilization study.

   (a) Degrees of freedom:

Source	df
variety	3
block	3
v*b	9
fertilizer	1
v*f	3
residual	12
c. total	31

   (b) E MS for variety: E SS $= E\left[\sum_{i=1}^{4}\sum_{j=1}^{4}\sum_{k=1}^{e}\left(\overline{y}_{.i.}-\overline{y}_{...}\right)^2\right]$

$$= 8\sum_{j=1}^{4} E\ \left(\overline{y}_{.i.}-\overline{y}_{...}\right)^2 \tag{1}$$

$$= 8\sum_{j=1}^{4} E\ \left[(\alpha_j-\overline{\alpha}_{.})+(\alpha\beta_{.j}-\overline{\alpha\beta}_{..})+(\alpha\tau_{j.}-\overline{\alpha\tau}_{..})+(\epsilon_{.j.}-\overline{\epsilon}_{...})\right]^2 \tag{2}$$

$$= 8\sum_{j=1}^{4} E\ \left(\alpha_{j.}+\alpha\tau_{j.}-\overline{\alpha}_{.}-\overline{\alpha\tau}_{..}\right)^2 + 8\,E\sum_{j=1}^{4}(\alpha\beta_{.j}-\overline{\alpha\beta}_{..})^2 + 8\,E\sum_{j=1}^{4}(\epsilon_{.j.}-\overline{\epsilon}_{...})^2 \tag{3}$$

   Hence, the expected MS is E SS / df = E SS / 3:

$$= \frac{8}{3}\sum_{j=1}^{4}\left(\alpha_j+\overline{\alpha\tau_{j.}}-\overline{\alpha}_{.}-\overline{\alpha\tau}_{..}\right)^2 + 8\ \mathrm{Var}(\overline{\alpha\beta}_{.j}) + 8\ \mathrm{Var}(\overline{\epsilon}_{.j.}) \tag{4}$$

$$= \frac{8}{3}\sum_{j=1}^{4}\left(\alpha_j+\overline{\alpha\tau_{j.}}-\overline{\alpha}_{.}-\overline{\alpha\tau}_{..}\right)^2 + 2\sigma_m^2+\sigma_s^2. \tag{5}$$

   The evaluation of the expected SS is mostly algebra. The step from (1) to (2) is expanding the model equation, multiplying terms and realizing that the expected value of each cross product terms is zero. The cross product terms take two forms. A sum of fixed effects times a random variable, which has expectation 0 because the random variable has expectation 0. Or, a sum of products of random variables, which expectation 0 because of independence of random variables.

   The two quantities in (3) involving random variables have expectations 3 σ^2, which is why the 3's cancel out in (4). The two variances in (4) are averages of 4 observations and 8 observations, so the coefficients on the variance components are 2 and 1.

   (c) No. The block effects subtract out of the equation defining the expected SS, so their role in the model is irrelevant.

(d) The ANOVA table is:

Source	df	Sum Sq	Mean Sq	F value	p-value
variety	3	11221	3740	4.1532	0.042
block	3	13879	4626		
variety*block	9	8105	901		
fertilizer	1	138864	138864	185.56	< 0.0001
variety*fertilizer	3	13758	4586	6.13	0.009
residual	12	8980	748		
cor.total	31				

Notes: If you use R, it really helps that these data are balanced. When the design is balanced, the SS can be obtained by fitting a sequence of fixed effects models (or using anova() after lm() to get sequential SS on the fixed effects model block + variety + block:variety + fertilizer + variety:fertilizer). Then construct the F tests yourself.

If you use SAS, both the sequential and partial ANOVA tables are part of the proc mixed output. Variety*block is in the random statement. Block is either in the random or model statement. The other terms are in the model statement.

(e) The simple effect of fertilization for each variety is given by $\bar{Y}_{.j1} - \bar{Y}_{.j2}$ with variance $\sigma_s^2/2$. Thus, the common s.e is $\sqrt{748/2} = 19.34$.

variety	est	t value	p value
Kent	151.5	7.83	< 0.0001
NZ	91	4.71	0.0005
S23	94.25	4.872	0.0004
X	190.25	9.84	< 0.0001

Notes: In R, there are a variety of ways to get the estimates. The most reliable is to set up a $\boldsymbol{C}$ vector for each estimate. The coefficients depend on the choice of full-rank parameterization. Most parameterizations require that you include differences of interaction coefficients as well differences of main effect coefficients. Setting up $\boldsymbol{C}$ is the hard part. Once specified, you get $\boldsymbol{C}\hat{\boldsymbol{\beta}}$ and Var $\boldsymbol{C}\hat{\boldsymbol{\beta}}$ by computing the obvious matrix equations.

Because the data are balanced, the simple averages for each cell (combination of variety and fertilization) are also the lsmeans. tapply() will give you those means. Then you know the form of the variance for a difference within variety, so you can hand calculate the variance of the effects and do the t-test.

In SAS, you need to add something like:

estimate 'Kent' fert 1 -1 variety*fert 1 -1 0 0 0 0 0 0;

to the proc mixed ocde.

(f) The variance of the difference between the NZ and Kent varieties when heavily manured is $\sigma_m^2/2 + \sigma_s^2/2$. The first thing to do is figure out the linear combination of mean squares that estimates this. That is:

$$\frac{1}{4}\left(EMS_{res} + EMS_{var*block}\right)$$

Then, using Cochran-Satterthwaite, you get:

$$\hat{\nu} = \frac{(\sum a_i MS_i)^2}{(\sum a_i^2 MS_i^2)/df_i} = \frac{\left(\frac{1}{4}EMS_{res} + \frac{1}{4}EMS_{var*block}\right)^2}{\left(\frac{1}{4}EMS_{res}\right)^2/9 + \left(\frac{1}{4}EMS_{var*block}\right)^2/12} = 19.88$$

The estimated difference is 295-289.5 = 5.5. The standard error is 20.30. The 0.975 quantile of a $T_{19.88}$ distribution is 2.087, or you can use a T_{20} = 2.086. The 95% ci is (-36.87,47.87).

2. Porosity of soils:

(a) The df are:

Source	df	
Field	14	= 15 - 1
Section(field)	15	= 30 - 15, or (2-1)*15
Location(section,field)	30	= 60 - 30, or (2-1)*30
c.total	59	

Notes: The first calculation is the change in error df between two models (e.g. intercept only and field for the first line). The second calculation comes from counting the number of df contributed by the variation at one level, pooled over the higher level. E.g. the 2 sections in field 1 give 2-1 = 1 df. Pooled over 15 fields gives 15. The first calculation works for unequal replication; the second only for each replication.

(b) The E MS for fields is:

$$\begin{aligned}
&= \frac{1}{14}E\sum_{i=1}^{15}\frac{1}{14}\sum_{j=1}^{2}\frac{1}{14}\sum_{k=1}^{2}(\overline{y}_{i..} - \overline{y}_{...})^2 \\
&= \frac{4}{14}\frac{1}{14}\sum_{i=1}^{15}E\left[(\alpha_i - \overline{\alpha}) + (\overline{\beta}_{i.} - \overline{\beta}_{..}) + (\overline{\epsilon}_{i..} - \overline{\epsilon}_{...})\right]^2 \\
&= 4\left[E\frac{1}{14}\sum_{i=1}^{15}(\alpha_i - \overline{\alpha})^2\right] + 4\left[E\frac{1}{14}\sum_{i=1}^{15}(\overline{\beta}_{i.} - \overline{\beta}_{..})^2\right] + 4\left[E\frac{1}{14}\sum_{i=1}^{15}(\overline{\epsilon}_{i..} - \overline{\epsilon}_{...})^2\right] \\
&= 4\left(\text{Var }\alpha_i + \text{Var }\overline{\beta}_{i.} + \text{Var }\overline{\epsilon}_{i..}\right) \\
&= 4\sigma_f^2 + 2\sigma_s^2 + \sigma_l^2
\end{aligned}$$

(c) The df for the real study are:

Source	df	
Field	14	= 15 - 1
Section(field)	15	= 30 - 15, or (2-1)*15
Location(section,field)	6	= 36 - 30
c.total	35	

(d) The unbalanced design has more df associated with fields and sections, so it provides more precise estimates of those variance components. The balanced design provides many more df to estimate variability among locations.

(e) The sequential SS (Type I) associated with each source of variation and type I ANOVA estimates of the three variance components are:

Source	df	Sum Sq	Mean Sq
field	14	15.8094	1.1292
section(field)	15	10.7097	0.7140
location(section,field)	6	8.7989	1.4665

Thus, $\hat{\sigma}^2_{location} = 1.4665$, $\hat{\sigma}^2_{section} = -0.6271$, $\hat{\sigma}^2_{field} = 0.1719$

(f) Using the corrected data, the SS and estimated variance components are:

Source	df	Sum Sq	Mean Sq
field	14	14.7239	1.0517
section(field)	15	13.0977	0.8732
location(section,field)	6	2.4749	0.4125

Thus, $\hat{\sigma}^2_{location} = 0.4125$, $\hat{\sigma}^2_{section} = 0.3839$, $\hat{\sigma}^2_{field} = 0.0765$

Note: one outlier can really influence the estimated variance components. I would not just force $\sigma^2_{section}$ to be non-negative.

(g) F=0.8732/0.4125=2.117 with p-value 0.181 under distribution $F_{15,6}$

(h) The needed linear combination of Mean Squares is:

$$\frac{1.1905}{1.2} MS_{section} + \frac{1.2 - 1.1905}{1.2} MS_{location} = 0.9921 MS_{section} + 0.0079 MS_{location}$$

(i) The CS approximate d.f. is:

$$\hat{\nu} = \frac{(\sum a_i MS_i)^2}{(\sum a_i^2 MS_i^2)/df_i} = 15.11$$

(j) F=$\dfrac{1.0517}{0.9921 * 0.8732 + 0.0079 * 0.4125} = 1.21$ with p-value 0.36 under distribution $F_{14,15.11}$

(k) Under the null hypothesis,

$$\frac{MS_{section}}{MS_{location}} \sim F_{15,6}$$

and the corresponding upper 95% quantile is 3.94.

Under the alternative hypothesis,

$$\frac{MS_{section}}{MS_{location}} \sim \frac{\sigma^2_{location} + 1.2\sigma^2_{section}}{\sigma^2_{location}} F_{15,6}$$

To have 80% power requires that

$$P\left[\frac{\sigma^2_{location} + 1.2\sigma^2_{section}}{\sigma^2_{location}} F_{15,6} > 3.94 = 0.8\right]$$

$$\Rightarrow \frac{3.94\sigma^2_{location}}{\sigma^2_{location} + 1.2\sigma^2_{section}} = F_{(15,6),20\%} \approx 0.6 \Rightarrow \sigma^2_{section} \approx 1.76$$