4. One possibility is to block by I = ±ABC = ±ACD (= ±BD). Now there are 4 blocks per replicate, or 12 blocks in all, with sum of squares computed as $4(\bar{y}_{block\ 1} - \bar{y}_{all})^2 + ... + 4(\bar{y}_{block\ 12} - \bar{y}_{all})^2$. Effect estimates are the same, except that ABC, ACD, and BD are now confounded with blocks and so $(\alpha\beta\gamma)$, $(\alpha\gamma\delta)$ and $(\beta\delta)$ are no longer estimable. In the ANOVA decomposition, the treatment degrees of freedom are reduced by 3, and the sum of squares is reduced by what can be computed, as though there are no blocks, as $N \times ((\widehat{\alpha\beta\gamma})^2 + (\widehat{\alpha\gamma\delta})^2 + (\widehat{\beta\delta})^2)$:

source	d.f.	sum of squares	mean square
blocks	11	810.604	
treatments	12	3211.067	267.589
residual	24	471.091	19.629
corrected total	47	4492.762	

5. All 15 estimates take the same values as in exercises 2 and 3. The ANOVA decomposition *can* actually be constructed in (at least) two different ways, depending on interpretation:

   - Suppose all 12 blocks are randomly drawn from the same source. That is, we regard two blocks within a rep as having the same relationship as two blocks from different replicates. Then ABC, ACD, and BD are "whole-plot effects" with 3 degrees of freedom, and a sum of squares computed as $48(\widehat{\alpha\beta\gamma})^2 + 48(\widehat{\alpha\gamma\delta})^2 + 48(\widehat{\beta\delta})^2$. The whole-plot corrected total line has 11 degrees of freedom, and is the "block" component in exercise (4). The whole-plot residual = whole-plot corrected total - whole-plot treatments. The remaining 12 factorial effects are "split-plot effects," with 12 degrees of freedom and sum of squares as computed for "treatments" in exercise 4.2, tested against split-plot residual variation computed as "residual" in exercise 4.
   - Suppose there may be systematic (unknown, nonrandom) differences among replicates, but that within a replicate the 4 blocks represent random draws from a single source. Now, we regard two blocks within a rep as potentially more similar than two blocks from different reps. The ANOVA decomposition is similar to what is required under the first interpretation, but here there is a need to reintroduce "replicate" as a fixed effect line in the whole-plot section of the ANOVA table; this has 2 degrees of freedom and is orthogonal to all factorial effects. Whole-plot residual is now *within-replicate* variability among blocks, apart from that which is associated with the 3 whole-plot effects. Degrees of freedom and sums of squares for whole- and split-plot treatments and residuals are the same as in the first interpretation, so the split-plot analysis is identical.

6. One possibility is:

   - Rep 1: I = ABC = ACD (= BD)
   - Rep 2: I = BCD = ABD (= AC)
   - Rep 3: I = AB = CD (= ABCD).

   Then, for example, $(\widehat{\alpha\beta\delta}) = [-48.87 - 54.37 + 31.52 + 34.88... + 56.48]/32$, and the other 8 effects involved in the blocking scheme are similarly computed from subsets of 32 data values. The remaining

15 - 9 effects are estimated as if there were no blocking (i.e. all data are used, and $N = 48$). Overall treatment sum of squares is the sum of squared least-squares estimates, each multiplied by the number of data values used *in that estimate*. Residual degrees of freedom are 47 (c. total) - 11 (reps and blocks) - 15 (effects) = 21, and the associated sum of squares is also computed by subtraction.

3.

(a) Letting Glucose = 'A', Inoculum Size = 'B', Aeration = 'C', Temperature = 'D', and Sodium = 'E',

$$\mathrm{I} = +\mathrm{ACD} = +\mathrm{BCE}\ (= +\mathrm{ABDE}).$$

(b) The estimable strings of effects are:

A+CD+ABCE+BDE
B+ABCD+CE+ADE
C+AD+BE+ABCDE
D+AC+BCDE+ABE
E+ACDE+BC+ABD
AB+BCD+ACE+DE
AE+CDE+ABC+BD

If we make a tentative assumption that interactions are not present (or are small relative to main effects) we may justify using 2 degrees of freedom corresponding to the last two strings as an estimate of σ^2, i.e. for strain A data,

$$MSE = 8 \times ((-0.2600)^2 + (0.2375)^2)/2 = 0.4960.$$

Using this mean square as the basis for the denominator of t statistics and the analogous value for strain B data, tests for the first five strings could be based on:

	strain A		strain B	
string	estimate	t	estimate	t
A	0.1625	0.6526	0.2663	2.6122
B	0.1625	0.6526	0.2463	2.4160
C	0.0875	0.3514	-0.2287	-2.2443
D	-0.7900	-3.1726	-1.1188	-10.9760
E	-0.5600	-2.2490	-0.6638	-6.5120

(c) If strain A is of primary interest, then factor D may be of main concern in the follow-up experiment; in this case I = -ACD = +BCE (= -ABDE) might be considered, producing the half-fraction I = +BCE for which the interaction aliased with the factor D main effect is BCDE. If strains A and B are to be followed up, and/or factors D and E are both regarded as potentially interesting, it might be more reasonable to double with I = -ACD = -BCE (= +ABDE), for I = +ABDE in the combined experiment. Here the aliases of D and E are each three-factor interactions, and the previous aliasing relationships between D and AC, and E and BC are broken. Doubling to I = +BCE or I = +ACD would leave one of the two "interesting" main effects confounded with a two-factor interaction.

5.

(a) The model contains 1 (intercept) + 8 (main effects) + 28 (two-factor interactions) = 37 parameters. The smallest power of 2 greater than 37 is $2^6 = 64$, so the fraction must contain $2^{8-2} = 64$ treatments.

(b) I = +ABCDE = +ABFGH (= + CDEFGH).

8. I = +ABC = +DEF (= +ABCDEF).