Problem 1

a) Let Y be the vector of y_{ij}, and our eventual predictor to be $c'Y$ for $c \in \mathbb{R}^{ab}$. For our predictor to be unbiased, it must be the case that

$$\begin{aligned} 0 &= \mathrm{E}(c'Y - \mu - \alpha_1) \\ 0 &= c'\mathrm{E}(Y) - \mu \\ 0 &= \mu c'1 - \mu \\ 1 &= c'1 \end{aligned}$$

Furthemore, to make our LUP a BLUP, we need to minimize the variance around our estimate. Thus, making use of our constraint:

$$\begin{aligned} \mathrm{Var}(c'Y - \mu - \alpha_1) &= \mathrm{Var}\left(\sum_{i,j} c_{ij}(\mu + \alpha_i + \beta_j + e_{ij}) - \mu - \alpha_1\right) \\ &= \mathrm{Var}\left(\sum_{i,j} c_{ij}(\alpha_i + \beta_j + e_{ij}) - \alpha_1\right) \\ &= \mathrm{Var}\left(\sum_{i,j} c_{ij}(\alpha_i + \beta_j + e_{ij})\right) - 2\mathrm{Cov}\left(\alpha_1, \sum_{i,j} c_{ij}(\alpha_i + \beta_j + e_{ij})\right) + \mathrm{Var}(\alpha_1) \\ &= \mathrm{Var}\left(\sum_{i,j} c_{ij}(\alpha_i + \beta_j + e_{ij})\right) - 2\sigma_a^2 \sum_j c_{1j} + \sigma_a^2 \end{aligned}$$

If we let $\Sigma = \mathrm{Cov}(Y) = \sigma_e^2 I + \sigma_a^2 V_a + \sigma_b^2 V_b$ and $I_{\{i=1\}}$ be a vector who has entry 0 for ij where $i \neq 1$ and 1 when $i = 1$, we get

$$\begin{aligned} \mathrm{Var}(c'Y - \mu - \alpha_1) &= c'\Sigma c - 2\sigma_a^2 c' I_{\{i=1\}} + \sigma_a^2 \\ \nabla \mathrm{Var}(c'Y - \mu - \alpha_1) &= 2\Sigma c - 2\sigma_a^2 I_{\{i=1\}} \end{aligned}$$

Which is 0 when (we know σ is invertible due to its block structure):

$$c = \sigma_a^2 \Sigma^{-1} I_{\{i=1\}}$$

Making this $c'Y$ the BLUP.

a) Repeating a similar process, we have a constraint of

$$\begin{aligned} 0 &= \mathrm{E}(c'Y - \mu - \alpha_1 - \beta_1) \\ 0 &= c'\mathrm{E}(Y) - \mu \\ 0 &= \mu c'1 - \mu \\ 1 &= c'1 \end{aligned}$$

And find that our estimates variance is

$$\begin{aligned}
\mathrm{Var}(c'Y - \mu - \alpha_1) &= \mathrm{Var}\left(\sum_{i,j} c_{ij}(\mu + \alpha_i + \beta_j + e_{ij}) - \mu - \alpha_1 - \beta_1\right) \\
&= \mathrm{Var}\left(\sum_{i,j} c_{ij}(\alpha_i + \beta_j + e_{ij}) - \alpha_1 - \beta_1\right) \\
&= \mathrm{Var}\left(\sum_{i,j} c_{ij}(\alpha_i + \beta_j + e_{ij})\right) - 2\mathrm{Cov}\left(\alpha_1 + \beta_1, \sum_{i,j} c_{ij}(\alpha_i + \beta_j + e_{ij})\right) + \mathrm{Var}(\alpha_1 + \beta_1) \\
&= \mathrm{Var}\left(\sum_{i,j} c_{ij}(\alpha_i + \beta_j + e_{ij})\right) - 2\sigma_a^2 \sum_j c_{1j} - 2\sigma_b^2 \sum_i c_{1j} + \sigma_a^2 + \sigma_b^2 \\
&= c'\Sigma c - 2\sigma_a^2 c' I_{\{i=1\}} - 2\sigma_b^2 c' I_{\{j=1\}} + \sigma_a^2 + \sigma_b^2 \\
\nabla\mathrm{Var}(c'Y - \mu - \alpha_1 - \beta_1) &= 2\Sigma c - 2\sigma_a^2 I_{\{i=1\}} - 2\sigma_b^2 I_{\{j=1\}}
\end{aligned}$$

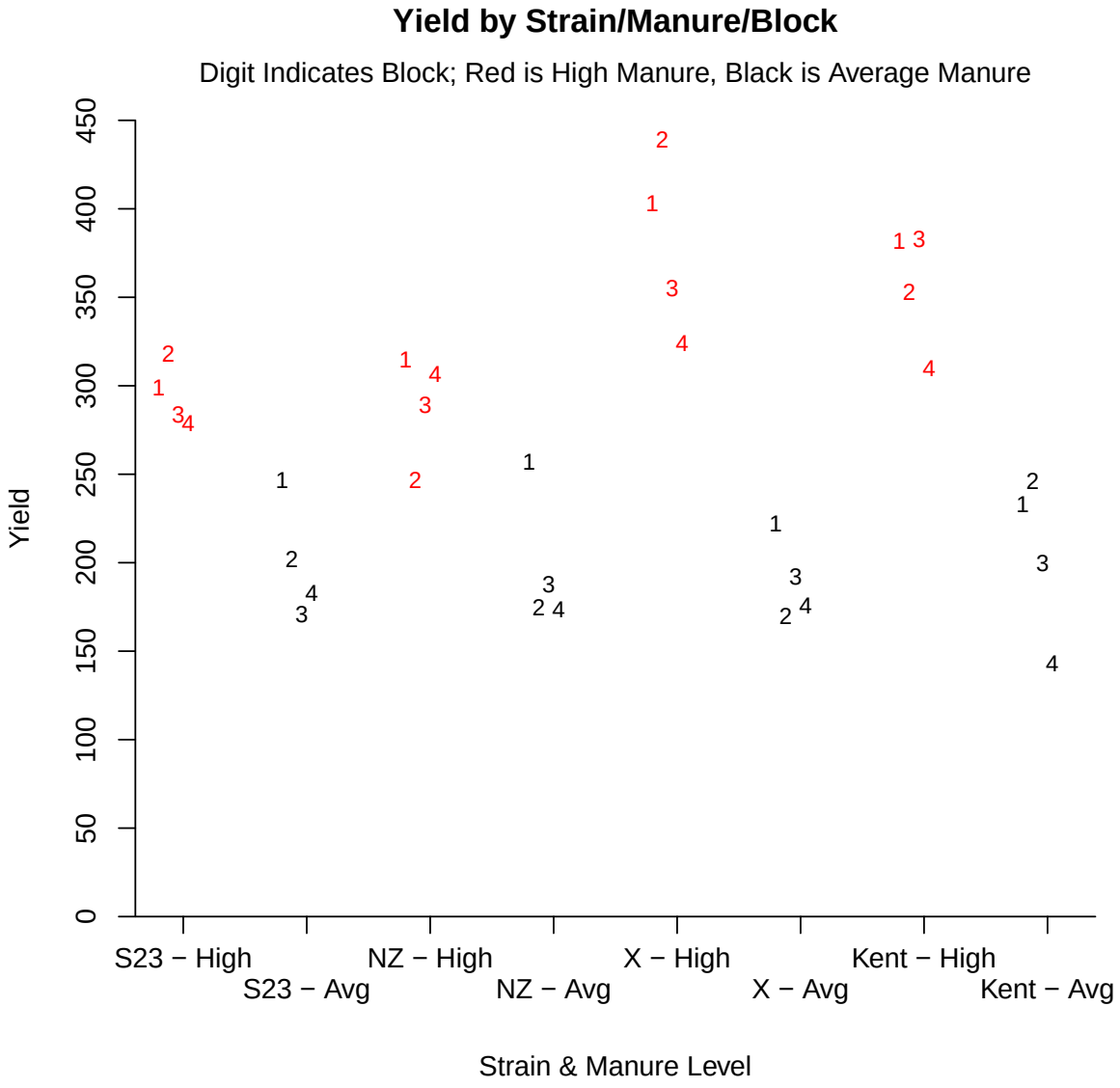
Which is 0 when

$$c = \Sigma^{-1}(\sigma_a^2 I_{\{i=1\}} + 2\sigma_b^2 I_{\{j=1\}})$$

Making this $c'Y$ the BLUP

Problem 2

a) As we can see from this plot, it appears that the yield of all strains is increased by heavy manure. In addition, yield seems to vary with plot, and it seems that there is an interaction between high yield and strain.



It would not be appropriate to have fixed effects for each of the three variables, since only the effect of the specific manure level and strain is of interest. On the other hand, no other land manager would care about the particular effect of these specific plots; their inclusion in a model is only so their effect can be factored out of the estimates for manure level and strain and the variance from it can be incorporated.

b) I fit a model with strain and manure level, along with an interaction between the two, as fixed effects, and plot as a random effect. At eight parameters and four random effects, this is a relatively high number of variables to estimate for a data set of 32 observations, but the manure - strain interaction seems to meaninful to leave out. These are the estimates of the fixed effects:

	Fixed Effect
(Intercept)	205.50
StrainNZ	-7.00
StrainS23	-4.75
StrainX	-15.50
ManureH	151.50
StrainNZ:ManureH	-60.50
StrainS23:ManureH	-57.25
StrainX:ManureH	38.75

These are the estimated effects for the manure level, strain, and the interaction between the two. They can be thought of as constant for these strains; there was no randomness in their manifestation, and they don't provide information on any other strain or manure level. Then, these are the predictions of random effects for each plot:

	Random Effect
1	24.88
2	3.45
3	-5.61
4	-22.72

These are the particular manifestation of a random variable which dictates how an arbitrary plot effect yields. These estimates aren't so interesting in themselves, but their distribution is important to understand how crop yield will vary from place to place.

c) I've calculated the standard errors and p-values two different ways. The first is by using the built in nlme package assumption that the coefficients are t-distributed in the normal fashion with their rules for degrees of freedom. The difference between manure and not for each plant can then be calculated as the sum of the manure and the interaction coefficients (with standard errors based on the covariance matrix):

	Coefficient	SE	P-Value
Kent	151.50	20.17	0.00
NZ	91.00	20.17	0.00
S23	94.25	20.17	0.00
X	190.25	20.17	0.00

I checked, and the residuals appear to be approximately normally distributed, though with only four estimated random effect predictions, its impossible to say normality holds. Thus, I also estimated the standard errors and p-values using a non-parametric bootstrap. For each boot, I drew randomly drew, with replacement, residuals and random effects from the observed pool of each, and added these on to the model estimates of the mean. Then, for each boot, I estimated the parameters again, using the sample over all boots to get the statistics of interest. As you can see, these standard errors are somewhat smaller (though the p-values are extremely small either way):

	Coefficient	SE	P-Value
Kent	151.50	17.20	0.00
NZ	91.00	16.74	0.00
S23	94.25	16.36	0.00
X	190.25	16.18	0.00

This difference is largely due to the small pool of realized random effects being drawn from. With a parametric bootstrap, the estimates are very close (as you would expect). Its hard to say which is better; in both cases we are hampered by the small pool of realized random effects, and we lack enough data to reject the normal assumption. However, with such small p-values either way, we don't have to worry too much about a false positive.

d) Using the normal assumption again, we can calculate the difference between S23 and NZ when heavily manured as

$$\hat{d} = \mu + \beta_{S23} + \beta_H + \beta_{S23,H} - \mu - \beta_{NZ} - \beta_H - \beta_{NZ,H} = \beta_{S23} + \beta_{S23,H} - \beta_{NZ} - \beta_{NZ,H}$$

With the usual variance for a linear combination of a multivariate normal RV. Using the 21 degrees of freedom given by lme, this gives us a confidence interval of:

$$\hat{d} \pm \sqrt{\text{Var}(\hat{d})} * t_{.025,21} = (-4.783, 15.783)$$