Problem 1

a. The distribution of Z is identical to the distribution of the described variance estimator when the error is iid $N(0,1)$. We can see this by the correspondence between its creation and the prescribed procedure:

   i. S_Y^2 has the same distribution as the residual sum of squares from the model where 2-factor interactions are assumed (giving residuals with five degrees of freedom), and S_X^2 has the same distribution as the marginal sum of squares attributable to the two factor interactions. This makes $\frac{S_X^2}{S_Y^2}$ identically distributed to the F statistic for the comparison of the model without two factor interactions to the model with two factor interactions.

   ii. Thus, $\frac{S_X^2}{S_Y^2} < 4.95 = F_{.95}(6,5)$ is equivalent to not rejecting the null that there are no two factor interactions. In the procedure, when this occurs the two factor interactions are pooled into the error and the error variance is estimated on 11 degrees of freedom. This is equivalent to $Z = \frac{5S_Y^2+6S_X^2}{11}$.

   iii. $\frac{S_X^2}{S_Y^2} \geq 4.95$ is equivalent to rejecting the null, so the error is estimated based on the original 5 degrees of freedom. In this case, that is distributed the same as $Z = S_Y^2$.

b. Over 1000 simulations, the mean of Z was .951, and the mean squared error was .180. I simulated it by the same method as described in part a. These values are meaningfully lower than both the true value $\sigma^2 = 1$ and the mean squared error of $S_Y^2 = .4$. We can derive the latter by noting $S_Y^2 \sim \frac{\chi_5^2}{5}$, so

$$\mathrm{E}((S_Y^2 - 1)^2) = \mathrm{E}(S_Y^2 - 1)^2 + \mathrm{Var}(S_Y^2) = \frac{\mathrm{Var}(\chi_5^2)}{25} = \frac{10}{25}$$

   The reason we see the underestimate of σ^2 is that with each β set to 0, we are essentially drawing 11 χ^2 variables, and throwing out the last 6 if they are too much larger than the first 5, then averaging over however many remain. The end result is that the average will be biased downwards. The mean square error is smaller because 95% of the time it is average of 11 instead of 5 χ^2 variables, reducing the variance more than enough to compensate for the small additional bias.

c. When β_1 is simulated 1000 times at different values, this is the mean and MSE I got:

	0.5	1	2	4	8
Z Mean	1.016	1.220	1.651	1.191	0.996
MSE	0.235	0.430	1.576	1.858	0.395

   As you would expect, initially both the Mean and MSE of Z increase with β_1. Since the F test has relatively low power for small, positive β, the null is rarely rejected, and the mean and MSE both go up with S_x^2. However once β_1 is large enough and the power is high, the test rejects more often, so both the mean and MSE drop back down again. We see this clearest at $\beta_1 = 8$, where the test should reject the vast majority of the time, making it so that $Z \approx S_Y^2$.

d. The obvious disadvantage of pooling the interactions into the error when a test determines them insignificant is the potential for bias. When there are no interactions, the estimate of the variance is biased downwards, and when there are nonnegative interactions and low power, the variance is significantly biased upwards. The advantage is that for small or non-existent interactions, you still have a better MSE due to the decreased variance of your estimator. My advice is that this is probably safe if your significance test has high power (though this probably makes the pooling unnecessary), but otherwise I would avoid employing this technique; as our tests have shown, it seems that you generally end up with a worse estimator.

Problem 2

a. From those two defining relationships, we can figure out what is aliased with the identity via :

$$E = ABC \,\&\, F = BCD \Longrightarrow I = ABCE \,\&\, I = BCDF$$

Multiplying ABCE by BCDF gives us the last element aliased with I:

$$I = ABCE = BCDF = ADEF$$

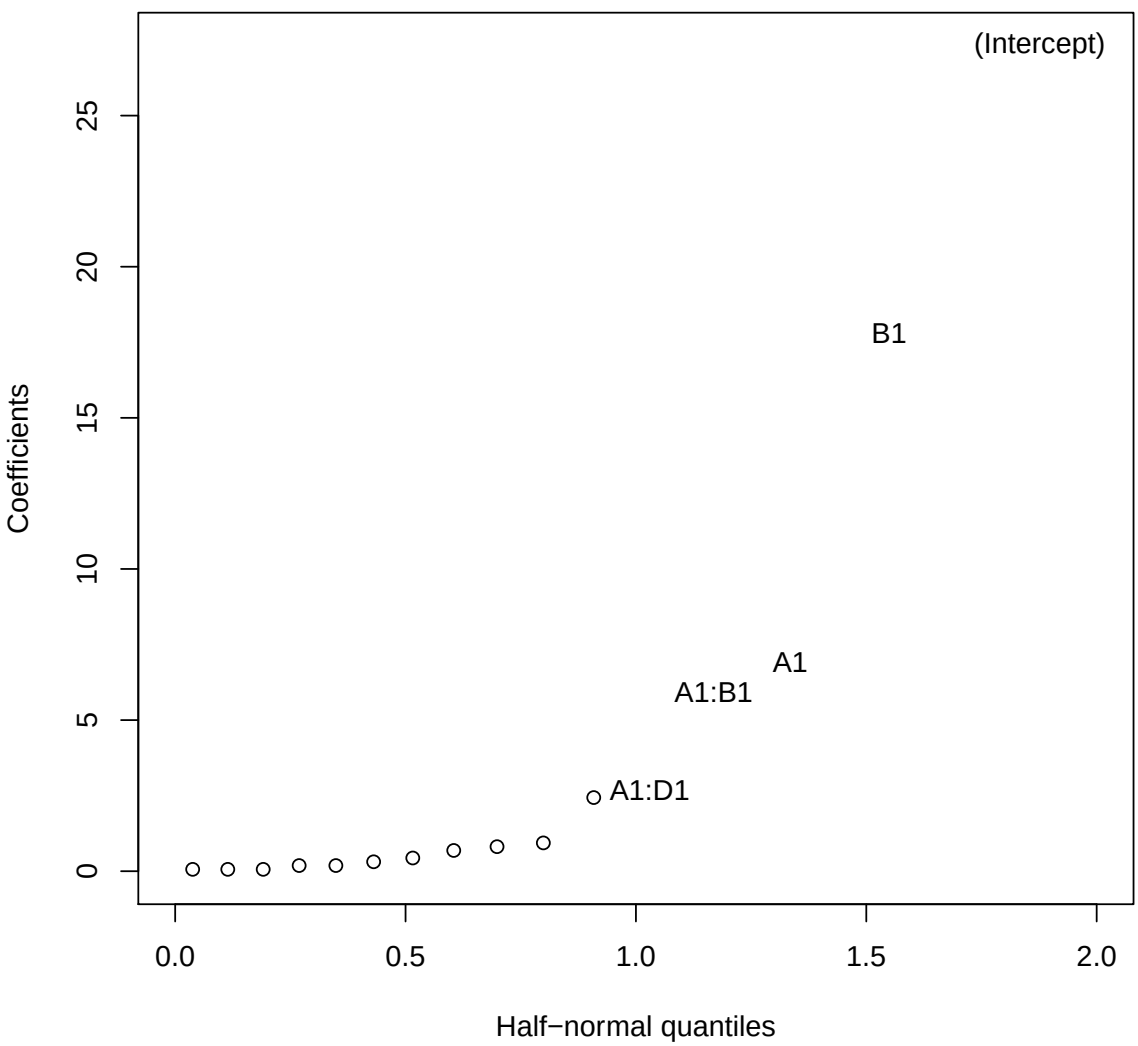
Now, multiplying through by remaining terms in the data set gives us the rest of the relationships:

$$\begin{aligned}
A &= BCE = ABCDF = DEF \\
B &= ACE = CDF = ABDEF \\
AB &= CE = ACDF = BDEF \\
C &= ABE = ADF = ADCEF \\
AC &= BE = ABDF = CDEF \\
BC &= AE = ABCDEF = DF \\
ABC &= E = BCDEF = ADF \\
D &= ABCDE = BCF = AEF \\
AD &= BCDE = ABCF = EF \\
BD &= ACDE = CF = ABEF \\
ABD &= CDE = ACF = BEF \\
CD &= ABDE = BF = ACEF \\
ACD &= BDE = ABF = CEF \\
BCD &= ABCEF = F = ADE \\
ABCD &= BCEF = AF = DE
\end{aligned}$$

b. As I'll show from the regression model, the best combination seems to be A, B, D at the "-" level, and C, E, and F at the "+" level, though besides A and B, the rest don't seem to have a significant impact. Running a fully interacted model, we get these estimations (effects are under the sum constraint and for the "-" level):

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	27.312			
A1	-6.938			
B1	-17.812			
C1	0.437			
D1	-0.687			
A1:B1	5.938			
A1:C1	-0.813			
B1:C1	-0.937			
A1:D1	-2.687			
B1:D1	-0.062			
C1:D1	-0.062			
A1:B1:C1	-0.188			
A1:B1:D1	-0.063			
A1:C1:D1	2.438			
B1:C1:D1	-0.187			
A1:B1:C1:D1	0.312			

Based off this, it seems like A- and B- is unquestionably a must, since their individual effects are so large negative, outweighing their positive interaction. Then, with D- having a negative coefficient and A-D- having a relatively large negative coefficient, D- seems a good choice too. A-B-C- has a relatively large positive coefficients, so C+ is the better choice. F and E, as determined by the aliasing relationship in the data, should be thus set to "+".
However, looking at the half normal plot of the absolute value of the coefficients, besides A and B its unclear if any of the estimates are significant enough to worry about:



Clearly B-, A-, and A-B- have meaningful effects. A-D- and its neighbor, A-C-D- are borderline, only maybe lying above the expected line for normal noise which the remaining coefficients fall on or below.
Finally, I tried to look into whether there was predictable variation in shrinkage. Obviously, with a fully interacted model, we can't estimate the variance at all, but since the interactions generally seemed small, I estimated a model with no interactions (but including E and F, as determined by the aliasing). I then fit a model predicting the log absolute value of the residual:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.453	0.249	5.845	0.000
A	-0.106	0.249	-0.427	0.679
B	0.066	0.249	0.266	0.796
C	-0.304	0.249	-1.224	0.252
D	0.032	0.249	0.130	0.900
E	0.086	0.249	0.346	0.737
F	-0.603	0.249	-2.427	0.038

The one 'significant' p-value is borderline, and based on the F-statistic comparing this to a mean model, we get a p-value of .348. Thus, we don't have evidence to say the shrinkage variation changes with any of our variables.

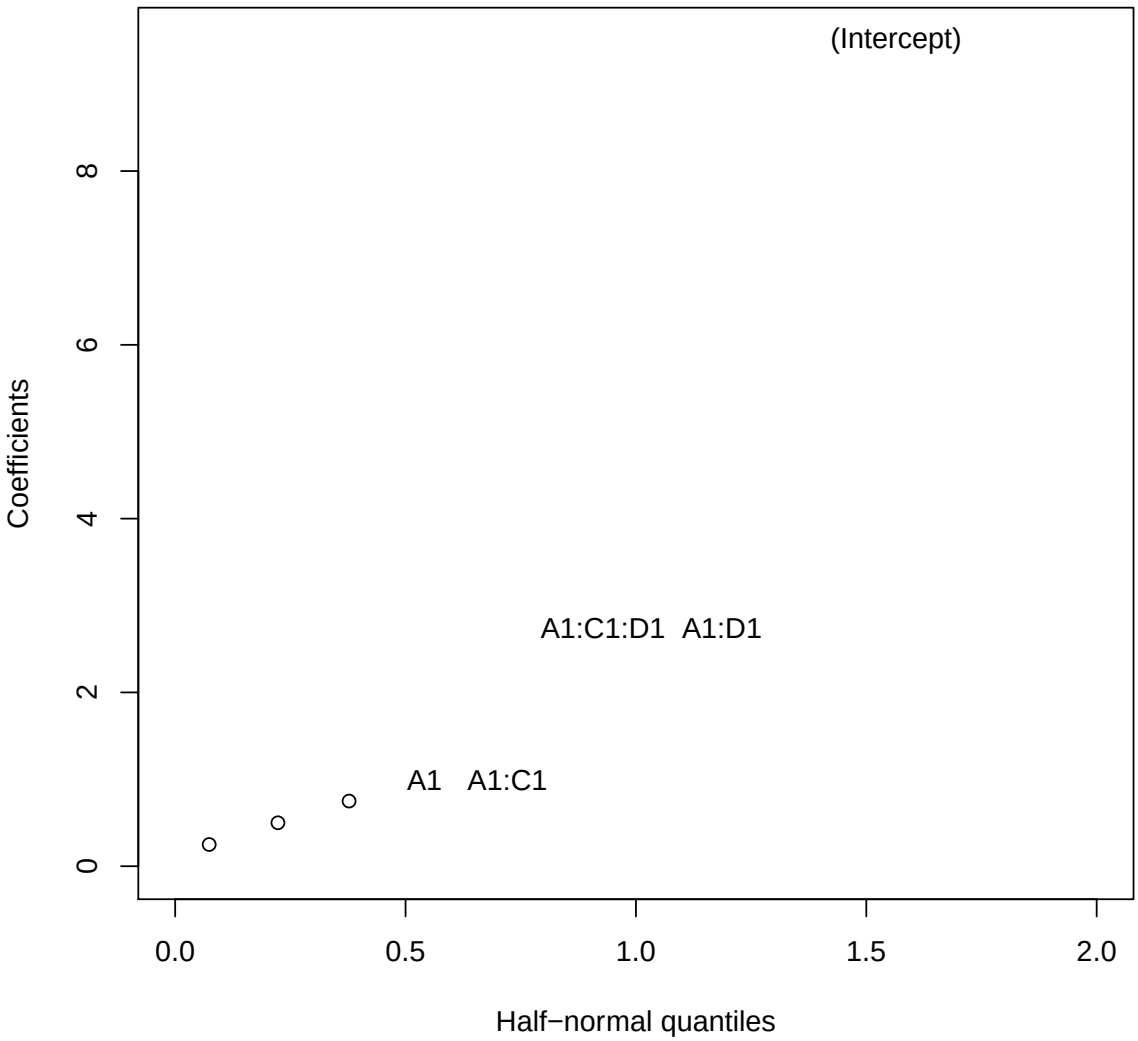
c. Having B set to "-" was clearly the most important choice, so I used that as the basis for this analysis. Now, the Aliasing is:

$$\begin{aligned}
I &= ACE = CDF = ADEF \\
A &= CE = ACDF = DEF \\
C &= AE = ADF = ADCEF \\
AC &= E = ADF = CDEF \\
D &= ACDE = CF = AEF \\
AD &= CDE = ACF = EF \\
CD &= ADE = F = ACEF \\
ACD &= DE = AF = CEF
\end{aligned}$$

Repeating our analysis from before, we get the following coefficients from the regression:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	9.500			
A1	-1.000			
C1	-0.500			
D1	-0.750			
A1:C1	-1.000			
A1:D1	-2.750			
C1:D1	-0.250			
A1:C1:D1	2.750			

Which I have plotted on a half norm plot here:



With so few data points, its hard to say if any of these are significant; besides the intercept, this could easily be a null plot. However, if we have to make a choice, A-D- still has the biggest negative coefficient, even if the coefficient on just A- has shrunk, while A-C-D- has an equally large positive coefficient. Thus, The same combination, A-, C+, D-, E+, F+ still seems like the best option.