1. Model selection practice --- My results are based on considering 4 possible X variables (X_1, X_2, X_3, and X_4). From your e-mails, some of you created all possible interactions and all possible quadratics. That is also a reasonable thing to do. I know the results are not the same for C_p because you're starting with a different 'full' model. I don't know whether AIC and BIC results differ. If you started with more X variables, your answers (if appropriate) should be marked correct, even if they differ from my answers.

a) $\hat{Y} = -1.36234 + 2.29157X_1 + 0.82934X_2 + 1.33143X_4$

Note: You could also indicate just the variables going into the model for full credit.

b) Yes, in this case we do.

c) The models with smallest Cp are the ones that include: (1) x1, x2, x4, (2) x1,x2,x3, and (3) all four variables.

No in Model	C(p)	R-Square	AIC	SBC	Variables in Model
3	3.0184	0.8069	140.2875	147.93557	x1 x2 x4
3	3.0736	0.8067	140.3488	147.99686	x1 x2 x3
4	5.0000	0.8070	142.2670	151.82715	x1 x2 x3 x4
3	6.5302	0.7919	144.0432	151.69133	x2 x3 x4

d) Using AIC we choose the same three models as in (c). However, using SBC instead of selecting the model with all four variables we select the one with x2, x3 and x4.

2. Residential house price models --- There are many possible answers here. What you learn by doing this problem is way more important than the points you get for you. Full credit goes to anyone who identified a model by any reasonable method (C_p, AIC or BIC), then evaluated diagnostics and reselected the model if major changes were made.

If you look at diagnostics, you will probably notice house 108. You may notice other curiosities, such as a house with 4 bedrooms and 7 bathrooms (lots of teenagers?). To repeat the moral one last time: if I was being paid to analyze these data, I would spend a lot of time looking for and checking curiosities before spending a lot of time on modeling. I would definitely fit a preliminary model or two to help me identify curiosities.

3. Drug side effects --- ANCOVA

My SAS code for all parts:

```
data heart;
  infile 'heart.txt' firstobs=2;
  input drug $ pre post;
run;
proc glm;
  class drug;
  model post = drug;
  lsmeans drug / stderr;
  estimate 'Drug A - placebo (C)' drug 1 0 -1;
  title 'ANOVA on post-trt values';
 run;
proc glm;
  class drug;
  model post = pre drug;
```

```
  lsmeans drug / stderr;
  estimate 'Drug A - placebo (C)' drug 1 0 -1;
  title 'ANCOVA using pre-trt values';
 run;
proc glm;
  class drug;
  model post = pre drug pre*drug;
  title 'Check interaction';
 run;
proc means;
  class drug;
  var pre;
  title 'Pre treatment means for each drug';
run;
```

a) No evidence. p-value for 2 d.f. F test of drugs = 0.49.

b) The means are A: 75.5, B: 77.9, C: 77.7. The difference between A and C is estimated to be -2.2, with s.e. = 2.3.

c) Yes, there is evidence of at least one difference among drugs. F for drugs = 3.62 with p=0.041

d) This question was poorly worded, because the means depend on which pre-treatment value you adjust to. SAS lsmeans uses the overall mean X as the default X value. That gives A: 74.6, B: 77.4, C: 79.3.

e) The difference between A and C is estimated to be -4.8 with s.e. = 1.8.

f) There is no evidence of different slopes. The p-value for the interaction pre*drug = 0.58

g) The estimated differences are not the same because drugs A and C have different mean pre-treatment values and the slope for pre is not zero. For A, the mean pre = 76, but for C, the mean pre = 72.9. The standard errors are different because the covariate accounts for a moderate amount of the variability in post-treatment responses. The sd. of the errors drops from 4.97 without the covariate to 3.72 with the covariate.

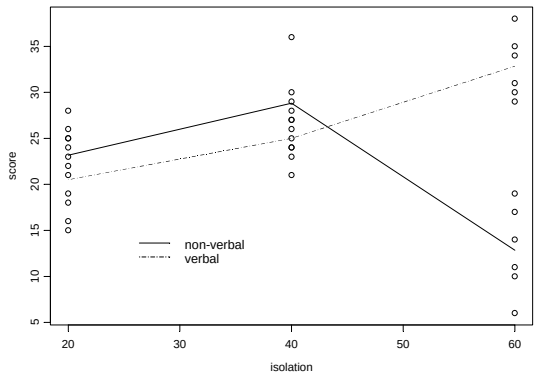
4. Children's memory – 2 way ANOVA.

My SAS code:

```
data memory;
  infile 'c:/philip/stat500/data/paragraph.txt';
  input reinf $ time $ score;
  both = trim(reinf) || '/' || time;
proc sort;
  by reinf time;
proc means noprint;
  by reinf time;
  var score;
  output out=means mean=mean;
proc print;
  id time;
  var reinf mean;
  title 'Mean scores for each group';
proc plot data=memory;
  plot score*time=reinf;
```

```
proc glm data=memory;
  class reinf time;
  model score = reinf time reinf*time;
  lsmeans reinf*time / slice=time;
  lsmeans reinf time / stderr;
  output out=resids r=resid p=yhat;
proc plot;
  plot resid*yhat;
proc glm;
  class both;
  model score = both;
  contrast 'reinforcement' both 1 1 1 -1 -1 -1;
  contrast 'isolation' both 1 0 -1 1 0 -1,
                       both 1 -1 0 1 -1 0;
  contrast 'interaction' both 1 0 -1 -1 0 1,
                         both 1 -1 0 -1 1 0;
run;
```

a) The plot of means shows that the lines connecting the verbal reinforcement means and the non-verbal reinforcement means cross. This suggests that there is an interaction. There might be a main effect of isolation because the values at 40 minutes are larger than those at 20; it is hard to see what is happening on average at 60 min without computing the marginal means. The main effect of reinforcement is hard to see (slightly + at two isolation times, strongly – at the last).



Note: SAS is not good for this sort of plot. We did it in another statistical package. Excel works; so does plotting the observations and adding the means and lines by hand.

b) ANOVA table shows reinforcement, isolation and their interaction are each significant at the .05 level.

Source	DF	Type III SS	Mean Square	F Value	Pr > F
reinf	1	196.000000	196.000000	12.42	0.0014
isol	2	156.222222	78.111111	4.95	0.0139
reinf*isol	2	1058.666667	529.333333	33.55	<.0001

c) The appropriate contrast coefficients depend on how you label the 6 groups. I numbered groups the reinforcement= 'none' groups as 1 through 3 and the reinforcement='verbal' as 4-6. The isolation = 20 groups were 1 and 4, =40 were 2 and 5 and =60 were 3 and 6. Hence, the appropriate contrast coefficientts (for SS and tests) were:

	None / 20	None / 40	None / 60	Verbal / 20	Verbal / 40	Verbal / 60
Reinforcement	1	1	1	-1	-1	-1
Isolation time	1	0	-1	1	0	-1
(2 df)	1	-1	0	1	-1	0
Interaction	1	0	-1	-1	0	1
(2 df)	1	-1	0	-1	1	0

The tests are exactly the same as before:

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	1410.888889	282.177778	17.88	<.0001
Error	30	473.333333	15.777778		
Corrected Total	35	1884.222222			

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
reinforce	1	196.000000	196.000000	12.42	0.0014
isolation	2	156.222222	78.111111	4.95	0.0139
interaction	2	1058.666667	529.333333	33.55	<.0001

Note: I used the SAS contrast statement, which will account for the covariance between two linear combinations. If you computed the SS for each contrast by hand and then added them, you would need to use orthogonal contrasts for each component of isolation time and the interaction.

d) Residual plots appear consistent with constant variance. The normal probability plots exhibit fairly straight lines indicating the normal error assumption is met, and no outliers are present. The assumption of independence is reasonable since the students were randomly assigned to the treatments.

e) Using the pooled error and slicing the interaction by isolation time gave me:

Time	SS	F	P
20 min	21.3	1.35	0.25
40 min	33.3	2.11	0.16
60 min	1200	76.1	< 0.001

Note: For this specific question, you could also use unadjusted p-values from pairwise comparisons, since the comparison at each isolation time are between two specific treatments. You could also use specific contrasts. Such contrasts would include both a reinf term and an interaction term. To see this, you should write out the desired contrasts as comparisons of cell means, then substitute the factor effects equations and do the subtraction.

Note: Splitting the data into thirds and doing separate t-tests (or 1-way ANOVA with 2 groups) for each time is less appropriate. Splitting the data gives you the same hypothesis SS but different F's and p's (you should be able to explain why). Since the sample size is small and the data suggest equal variances, I prefer to use all the data to estimate a single pooled error variance.

Note: it is hard to decide whether to use a multiple comparisons adjustment here. The three comparisons are orthogonal and we can't do too many, so most folks would not adjust. However, you are doing three (here), or possibly quite a few (in general), tests. I accepted either answer.

f) There are lots of possible answers here. Here is my summary. Some features to notice: I emphasize the magnitudes of effects, not the p-values. My choice of features to include in the summary depended on which effects were significant in the ANOVA table. I did not describe all pairs of means, but instead used the treatment structure to focus my summary. Other folks might emphasize different points; traditions vary across scientific disciplines.

On average, verbal reinforcement increases memory scores by 4.7 units, but this difference is not consistent across the isolation times. Considered separately at each isolation time, there is no evidence of an effect of reinforcement at 20 or 40 minutes, but verbal reinforcement increases retention by 20 units ($p < 0.0001$) at 60 minutes. With verbal reinforcement, longer isolation time tends to enhance children's memory. Without verbal reinforcement, moderate isolation time (40 min.) leads to improved performance, but too much isolation time (60 min.) reduces children's performance.

The population is fourth graders from this specific city, but I imagine the folks who collected these data will use them to make general conclusions.

5. Doughnut fats and ANOVA parameterization (optional)

a) SS treatments: 1636.5
SS error: 2018
Fat 1 mean: 172.0, se = 4.1
mean diff between fat1 and 4: 10.0, se = 5.8

b) For the fat means: $\mu_1 = B_0+\beta_1$, $\mu_2 = B_0+\beta_2$, $\mu_3 = B_0+\beta_3$, $\mu_4 = B_0$.
For the parameters in the regression model: $\beta_0 = \mu_4$, $\beta_1 = \mu_1 - \mu_4$, $\beta_2 = \mu_2 - \mu_4$, $\beta_3 = \mu_3 - \mu_4$,

c) SS treatments: 1636.5
SS error: 2018
β_1 estimate: 10.0, s.e. = 5.8
Fat 1 mean: 162.0 + 10.0 = 172.0, s.e. = 4.1 (obtained from estimate ' ' intercept 1 a1 1;)
mean diff between fat1 and 4: This is the estimate of $\beta_1 = 10.0$, s.e. = 5.8

The values for SS, means, and the diff. in means are the same as those in (a). The slope for fat 1 is not.

d) SS treatments: 1636.5
SS error: 2018
β_1 estimate: -1.75, s.e. = 3.55
Fat 1 mean: 173.5 + (-1.75) = 172.0, s.e. = 4.1
mean diff between fat1 and 4: again 10 with s.e. = 5.8.
Note, this is related to the estimated slopes because the mean for fat 4 = $B_0 - \beta_1 - \beta_2 - \beta_3$, so fat 1 – fat 4 = $B_0 + \beta_1 - (B_0 - \beta_1 - \beta_2 - \beta_3) = 2\beta_1 + \beta_2 + \beta_3$
Everything is the same except for the slope estimate.