Question #1

Statistical interaction includes a model which we allow a predictor's effect to differ according to levels of another predictor in the model. In other words, an interaction tells that the effect of one predictor varies by levels of another. In statistical interaction model, sometimes the magnitude of an effect varies (ordinal interaction) whereas other times the direction of an effect varies (disordinal interaction). Moreover, statistical interaction model (ordinal or disordinal) includes lines which are not parallel as opposed to main effects model where each predictor's effect is the same regardless of the values of all other predictors in the model (that is why the test detecting a statistical interaction is often called a test of parallelism). When we test the presence of a statistical interaction, we create a cross-product term, which is the product of two predictors. The hypothesis test looks whether H_0: $\beta_{cross-product} = 0$ or not. If the test rejects, it illustrates that the two predictors interact; if not, it shows that they each only have a main effect. Here, if the interaction term is statistically significant, it reveals that there is a statistically significant interaction between Over60 and Retire, and the effect of Over60 differs by retirement (Retire=1) and non-retirement communities (Retire=0).

Question #2

Our interaction term is ROver60, which is simply the product of Retire and Over60. The interaction effect model is as follows:

$\hat{PPE}$ = 6425.39 - 8.51*Over60 + 178.93*Retire - 21.34*ROver60

➔ $\hat{PPE}$ = 6425.39 - 8.51*Over60 + 178.93*Retire - 21.34*Retire*Over60

When Retire=0 (no retirement communities), $\hat{PPE}$ = 6425.39 - 8.51*Over60

When Retire=1 (there is a retirement community), $\hat{PPE}$ = 6604.32 - 29.85*Over60

The fact the cross-product term is not statistically significant points out that we do not have enough evidence to reach a conclusion that there is an interaction between Over60 and Retire in this uncontrolled model; thus, we go back to main effects model and continue to assume that the lines are parallel. The regression results reveal that, on average, in the districts without a retirement community,

one percentage point change in Over60 is associated with $8.51 lower PPE; similarly, in the districts with a retirement community, one percentage point change in Over60, on average, is associated with $29.85 lower PPE.

Question #3

Aside from Over60, L2Home, Poverty, and Retire, with the addition of ROver60 to the model, our most explanatory model became Model C (table 1 in the appendix) which best explains the variation in per pupil expenditure. In this controlled model, all of the predictors but Retire are statistically significant. Whereas looking at R^2 is not our first priority, of the three models, Model C has the highest R^2 with 27.4 percent and it brings a more accurate prediction by using all predictors. The fitted equation is as follows:

$\hat{PPE}$ = -3885.83 + 31.40 Over60 + 1540.03 L2Home + 77.05 Poverty + 245.04 Retire - 37.29 ROver60

Over60: Similar with previous model, controlling for other predictors, Over60 is still positively correlated with PPE and continues to be statistically significant. Every percentage point change in Over60 is associated with $31.40 higher PPE (previously was $17.68).

L2Home: Controlling for other predictors, every doubling in home prices now is associated with $1540 more PPE (as opposed to $1516).

Poverty: Controlling for other predictors, every percentage point change in this variable is associated with $77 higher PPE (previously was $73).

Retire: There is a noteworthy change in retirement communities. Previous model predicted that, controlling for other predictors, districts with retirement communities, on average, would have $388 less PPE whereas the new model shows that districts with retirement communities, on average, would have $245 more PPE.

Question #4

When creating the display given in the appendix, I preferred not to show L2Home and Poverty on the graph as my common sense would let me interpret that PPE would be positively associated with home prices, which are good indicator of wealth and socio-economic status. Besides, one of the leading monetary resources of public schools is property taxes. We also examined the effects of poverty in

assignment 4. After all, both home prices and poverty are economy-related variables, and this assignment is about the impact of age vis-a-vis PPE. Hence, I wanted to see the impacts of Over60 and Retire on PPE (Over60, question variable, is probably best displayed on the X-axis). The interaction in our controlled model tells that the effect of one predictor (Retire) varies by levels of another (Over60). Here, the direction of the effect varies and we conclude it as a disordinal interaction. The findings are mixed in addressing the gray peril hypothesis: Now that the cross-product term is statistically significant, our interpretations would be such that, first, the effect of Over60 differs by the existence of retirement communities (there is a positive association between Over60 and non-retirement communities whereas we observe a negative association between these two variables where there are retirement communities), and second, the effect of retirement communities differs by the percentage of citizens who are 60 and older. In short, longstanding and newly-arrived elderly differ in their financial support for public schools.

APPENDIX

```
options nodate nocenter nonumber;

title1 "S-030: Intermediate Statistics";
title2 "The case of Michigan on the basis of PPE and -newly arrived and longstanding- elderly";
title3 "Yansi Eraslan";
footnote1 "Program: Educational Expenditures vs. Age in Michigan -- e:\S030_Michigan2.txt";

data one;
infile "E:\S030_Michigan2.txt";
input ID 1-3 PPE 5-11 HOME 14-20 POVERTY 22-26 OVER60 28-32 RETIRE 34;
L2Home = log2(HOME);
ROVER60=RETIRE*OVER60;
RL2Home=Retire*L2Home;
RPoverty=Retire*Poverty;

proc sort data=one;
by Retire PPE;

proc univariate plot data=one;
title4 "Descriptive statistics";
var PPE Over60 Retire;
id id;

proc reg data=one;
title4 "Regression results - PPE vs. Retire";
model PPE=Retire;

proc sort data=one;
by id;

proc print data=one;
title4 "Listing data on observations 1-25 for inspection";
where 1 <= id <= 25;
var id Retire Over60 ROver60;

proc reg data=one;
title4 "Model A: Interaction effect model";
model PPE=Over60 Retire ROver60;

proc reg data=one;
title4 "Model B: PPE vs. Over60, L2Home, Poverty, and Retire";
model PPE=Over60 L2Home Poverty Retire;
```

```
proc reg data=one;
title4 "Model C: PPE vs. Over60, L2Home, Poverty, Retire, and ROver60";
model PPE=Over60 L2Home Poverty Retire ROver60;

proc means data=one;
title4 “PPE univariate summary information in tables for Retire=0 & Retire=1”;
by Retire;
var PPE;

proc reg data=one;
title4 "What happens if we fit separate models for PPE within retirement communities?";
by retire;
model PPE=Over60;

proc reg data=one;
title4 "Interactions for remaining predictors; uncontrolled";
model PPE=L2Home Retire Rl2Home;
model PPE=Poverty Retire RPoverty;

proc reg data=one;
title4 "Interactions for predictors, controlling for all main effects";
model PPE=Over60 L2Home Poverty Retire;
model PPE=Over60 L2Home Poverty Retire ROver60;
model PPE=Over60 L2Home Poverty Retire Rpoverty;
model PPE=Over60 L2Home Poverty Retire RL2Home;

run;
quit;
```

Table 1. Comparison of regression models predicting PPE in Michigan districts (n=471)			
Predictor	Model A	Model B	Model C
Intercept	6425.39	-3461.08	-3885.83
	(198.40)	(809.53)	(824.36)
	32.39***	-4.28***	-4.71***
Elder citizens	-8.51	17.68	31.40
	(10.83)	(8.14)	(9.89)
	-0.79	2.17*	3.17**
Log_2 (home prices)		1515.87	1540.03
		(120.29)	(120.09)
		12.60***	12.82***
Poverty		73.15	77.05
		(8.48)	(8.59)
		8.62***	8.97***
Retire	178.93	-388.16	245.04
	(315.73)	(85.48)	(275.73)
	0.57	-4.54***	0.89
ROver60	-21.34		-37.29
	(17.57)		(15.45)
	-1.21		-2.41*
R^2	1.77	26.53	27.44
F	2.80	42.06	35.17
(df)	(3, 467)	(4, 466)	(5, 465)
P	0.0394	<0.0001	<0.0001
Cell entries are estimated regression coefficients, (standard errors) and t-statistics. ~ p<0.10, * p<0.05, ** p<0.01, *** p<0.001			

Figure 1. Fitted values for the relationship between per pupil expenditure and elderly citizens for districts with and without retirement communities