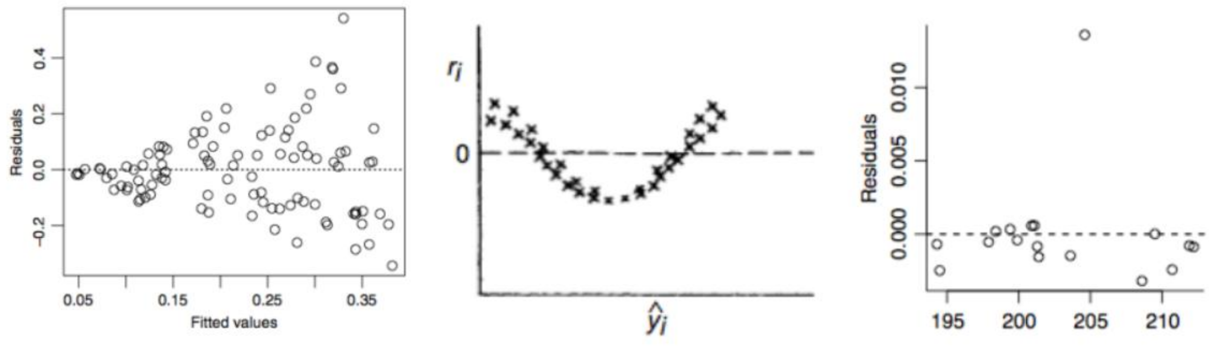
1. [15 pts] The following gives residual plots of three simple regression models. Are there any problems with these regressions? If so, specify the problems and describe in details how you would fix them. If not, give your reasons.



(a) (b) (c)

1. Retail interest is defined by markers as the level of interest a consumer has in a given retail store. Using survey data collected for n= 375 consumers, an interaction model is developed for y=willingness of the consumer to shop at a retailer's store in the future (called "repatronage intensions") as the function of x_1= consumer satisfaction and x_2=retailer interest. The regression results are shown below:

Variable	Estimated β	t-value	p-value
Satisfaction (x1)	.426	7.33	< .01
Retailer Interest (x2)	.044	0.85	> .10
Satisfaction x Retailer Interest (x1x2)	-.157	-3.09	< .01
R^2=.65, F = 226.35, p-value < .001			

(a) Is the overall model statistically useful for predicting y? Test using α=0.05.
(b) Conduct a test for interaction at α=0.05.
(c) Use the β-estimates to sketch the estimated relationship between y and satisfaction (x1) when retailer interest is x2=1 (a low value).
(d) Repeat part c when retailer interest is x2= 7 (a high value).
(e) Sketch the two lines, part c and d on the same graph to illustrate the nature of the interaction.

2. The sample was compared of 388 people enrolled in a full-time MBA program. Based on answers to a questionnaire, the researchers measured two variables for each subject: assertiveness score (x) and leadership ability score (y). A quadratic regression model was fit to the date with the following results:

Independent variable	β Estimate	t-value	p-value
x	0.57	2.55	0.01
x^2	-0.88	-3.97	<0.01
Model R^2 = 0.12			

(b) The researchers hypothesized that leadership ability will increase at a decreasing rate with assertiveness. Set up the null and alternative hypothesis to test this theory.
(c) Use the reported results to conduct the test part b. Give your conclusion (at α=0.05) in the words of the problem.

4.60 Stop, look, and listen.

Where do you look when you are listening to someone speak? Researchers have discovered that listeners tend to gaze at the eyes or mouth of the speaker. Subjects watched a videotape of a speaker gathering. The level of background noise (multilingual voices and music) was varied during the listening sessions. Each subject wore a pair of clear plastic goggles on which an infrared corneal detection system was mounted, enabling the researchers to monitor the subject's eye movements. One response variable of interest was the proportion y of times the subject's eyes fixated on the speaker's mouth.

(a) The researchers wanted to estimate E(y) for four different noise levels: none, low, medium, and high. Hypothesize a model that will allow the researchers to obtain these estimates.

(b) Interpret the β's in the model, part a.

(c) Explain how to test the hypothesis of no differences in the mean proportions of mouth fixations for the four background noise levels.

2. [20 pts] Consider a multiple linear regression with n observations and p predictors, $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, where $\mathbf{Y} = (Y_1, \ldots, Y_n)^\top$ is the $n \times 1$ vector of responses; $\mathbf{X}$ is the $n \times (p+1)$ matrix of predictors, including a column of 1's for the intercept; and $\mathbf{e} = (e_1, \ldots, e_n)^\top$ is the $n \times 1$ vector of statistical errors. Let $\hat{\mathbf{Y}} = (\hat{Y}_1, \ldots, \hat{Y}_n)^\top$ denote the fitted values and $\hat{\mathbf{e}} = (\hat{e}_1, \ldots, \hat{e}_n)^\top$ denote the residuals. (Note: show all the derivation steps for the following questions)

(a) [5 pts] Derive the formula of the ordinary least square (OLS) estimator $\widehat{\boldsymbol{\beta}}$.

(b) [5 pts] Derive the expectation and covariance matrix of $\widehat{\boldsymbol{\beta}}$.

(c) [3 pts] Show that the sample mean of $\hat{\mathbf{e}}$ is 0.

(d) [3 pts] Show that the sample covariance of $\hat{\mathbf{e}}$ and $\hat{\mathbf{Y}}$ is 0.

(e) [4 pts] What are the intercept and slope of the simple regression of $\hat{\mathbf{e}}$ on $\hat{\mathbf{Y}}$ (i.e., $\mathrm{lm}(\hat{\mathbf{e}} \sim \hat{\mathbf{Y}})$)?