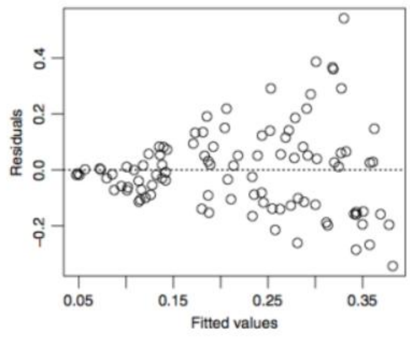
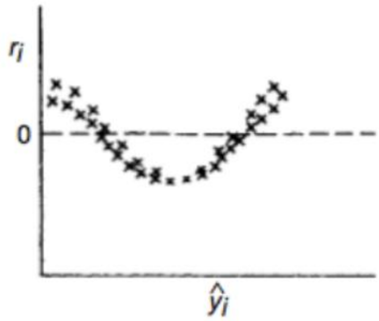
MID PRACTICES

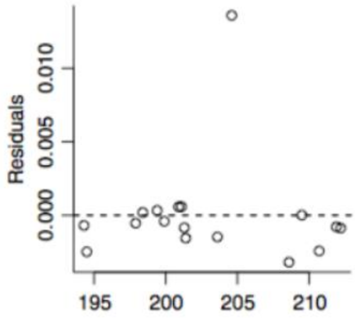
1. [15 pts] The following gives residual plots of three simple regression models. Are there any problems with these regressions? If so, specify the problems and describe in details how you would fix them. If not, give your reasons.



(a)



(b)



(c)

(a) Problem: variance not constant Fix: WLS

(b) Problem: non-linear Fix: Ass a quadratic term in the mean function

(c) Problem: outlier Fix: Perform outlier test. Consider remove outliers.

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.

Answers:

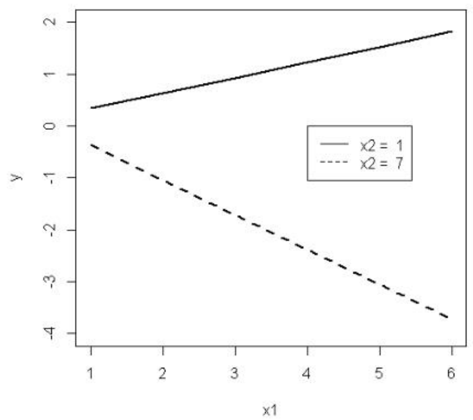
(a) The global F-test is used to test the null hypothesis $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$. The test statistic is F= 226.35 and p-value < α. There is a sufficient evidence to conclude that the model fit is a statistically useful predictor of the consumer's willingness to shop at a retailer's store in the future.

(b) H_0: $\beta_3 = 0$ vs Ha : $\beta_3 \neq 0$. Since we are testing an individual β parameter, a t-test is required. From the printout, the test statistic is t = -3.09. Since p-value < α, the null hypothesis is rejected and thus the interaction should be included in the model.

(c) $\hat{y} = 0.426x_1 + 0.044(1) - 0.157(1)x_1 \Rightarrow \hat{y} = 0.296x_1 + 0.044$

(d) $\hat{y} = 0.426x_1 + 0.044(7) - 0.157(7)x_1 \Rightarrow \hat{y} = -0.673x_1 + 0.308$

(e)



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.

Answer:

(a) To test whether the quadratic model is statistically useful. We conduct the global F test: $H_0 : \beta_1 = \beta_2 = 0$. $F = \frac{R^2}{(1-R^2)} \frac{n-k-1}{k} = \frac{.12}{(1-.12)} \frac{388-3}{2} = 26.25$

Critical value $F_{0.05}(2,385) \approx 3.00$ is less than the test statistic. So we reject H_0 and conclude that the model is significant at predicting the leadership ability score.

(b) H_0: $\beta_2 = 0$ vs H_0: $\beta_2 < 0$

(c) $t = -3.97$, and p-value $< 0.01/2 = 0.005$. We reject H_0 and conclude that this variable is significant at predicting leadership activity score and that leadership ability increases at a decreasing rate with assertiveness.

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.

Answer:

(a) The model is $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$

Where $x_1 = \begin{cases} 1, \text{ for low noise level} \\ 0, \text{ otherwise} \end{cases}$

$x_2 = \begin{cases} 1, \text{ for medium noise level} \\ 0, \text{ otherwise} \end{cases}$ $x_3 = \begin{cases} 1, \text{ for high noise level} \\ 0, \text{ otherwise} \end{cases}$

(b) β_0 = mean proportion of times (y) for noise level none.

β_1 = difference in mean proportion of times (y) between noise level low and noise level none.

β_2 = difference in mean proportion of times (y) between noise level medium and noise level none.

β_3 = difference in mean proportion of times (y) between noise level high and noise level none.

(c) H_0: $\beta_1 = \beta_2 = \beta_3 = 0$, Ha: at least one $\beta_j \neq 0$, j=1,2,3

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., $\mathrm{lm}(\hat{\mathbf{e}} \sim \hat{\mathbf{Y}})$)?

(a) $RSS(\boldsymbol{\beta}) = (\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^\top(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})$. Setting $\partial_{\boldsymbol{\beta}} RSS(\boldsymbol{\beta}) = \mathbf{X}^\top(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}) = 0$ implies $\mathbf{X}^\top\mathbf{Y} - \mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} = 0$, i.e. $\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}$ if $(\mathbf{X}^\top\mathbf{X})^{-1}$ exists.

(b) $\mathbf{E}[\widehat{\boldsymbol{\beta}} \mid \mathbf{X}] = E[(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y} \mid \mathbf{X}] = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{E}[Y \mid \mathbf{X}] = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{E}[\mathbf{X}\boldsymbol{\beta} + \mathbf{e} \mid \mathbf{X}] = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top(\mathbf{X}\boldsymbol{\beta} + \mathbf{E}[\mathbf{e} \mid \mathbf{X}]) = \boldsymbol{\beta}$ where we used that $\mathbf{E}[\mathbf{e} \mid \mathbf{X}] = 0$ by model assumption.

$\mathrm{Var}(\widehat{\boldsymbol{\beta}}|\mathbf{X}) = \mathrm{Var}((\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y} \mid \mathbf{X}) = \mathrm{Var}((\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top(\mathbf{X}\boldsymbol{\beta}+\mathbf{e}) \mid \mathbf{X}) = \mathrm{Var}(\boldsymbol{\beta}+(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{e} \mid \mathbf{X}) = \mathrm{Var}((\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{e} \mid \mathbf{X}) = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathrm{Var}(\mathbf{e} \mid \mathbf{X})[(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top]^\top = \sigma^2 I_n(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1} = \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}$ where we used that $\mathrm{Var}(\mathbf{e} \mid \mathbf{X}) = \sigma^2 I_n$ by model assumption.

(c) By the normal equations we know $\mathbf{X}^\top(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) = \mathbf{X}^\top(\mathbf{Y} - \widehat{\mathbf{Y}}) = \mathbf{X}^\top\widehat{\mathbf{e}} = \mathbf{0}$. This is a set of p linear equations in matrix form. [Note: for full credit, we expect that you derive the normal equations by minimizing RSS, either here or in part (a)]

Consider the first equation, i.e. the inner product of the first row of $\mathbf{X}^\top$ times the first (only) column of $\widehat{\mathbf{e}}$. Since the first row of $\mathbf{X}^\top$ is a vector of $1'$s, this equation can be written as $\sum_{i=1}^n \widehat{e}_i = 0$, implying that $\frac{1}{n}\sum_{i=1}^n \widehat{e}_i = \bar{\widehat{\mathbf{e}}} = 0$.

(d) Since the sample mean of $\widehat{\mathbf{e}}$ is zero, then $\widehat{cov}(\widehat{\mathbf{e}}, \widehat{\boldsymbol{Y}}) = \frac{1}{n-1}(\widehat{\mathbf{Y}} - \bar{\hat{y}}\mathbf{1})^\top\widehat{\mathbf{e}} = \frac{1}{n-1}(\mathbf{X}\widehat{\boldsymbol{\beta}})^\top\widehat{\mathbf{e}} - \frac{1}{n-1}\bar{\hat{y}}\mathbf{1}^\top\widehat{\mathbf{e}} = \frac{1}{n-1}\widehat{\boldsymbol{\beta}}^\top\mathbf{X}\widehat{\mathbf{e}} - \frac{1}{n-1}\bar{\hat{y}}\bar{\widehat{\mathbf{e}}} = \frac{1}{n-1}\widehat{\boldsymbol{\beta}}^\top\mathbf{0} + 0 = 0$ again by normal equations and step (c).

(e) Let β_0 and β_1 the coefficients of the regression. We know that $\widehat{\beta}_1 = \widehat{cov}(\widehat{\mathbf{e}}, \widehat{\boldsymbol{Y}})/\widehat{var}(\widehat{\boldsymbol{Y}})$ but $\widehat{cov}(\widehat{\mathbf{e}}, \widehat{\boldsymbol{Y}}) = 0$ by point (d) then $\widehat{\beta}_1 = 0$. We also know that $\widehat{\beta}_0 = \bar{\widehat{\mathbf{e}}} - \widehat{\beta}_1\bar{\hat{y}}$ which is zero because $\bar{\widehat{\mathbf{e}}} = 0$ by point (d) and $\widehat{\beta}_1 = 0$ (just proved).