



Assessing Student Ability to Interpret Regression Diagnostic Plots

Handanhal Ravinder
Montclair State University

Haiyan Su
Montclair State University

Mark Berenson
Montclair State University

Author Note

Handanhal Ravinder and Mark Berenson (Emeritus), Department of Information & Business Analytics; Haiyan Su, College of Science & Mathematics

Correspondence regarding this article should be addressed to Handanhal Ravinder, Information Management & Business Analytics Department, Feliciano School of Business, Montclair State University, 1 Normal Avenue, Montclair, NJ 07043; E-mail: ravinderh@mail.montclair.edu.

Abstract

Simple linear regression analysis is an important component of introductory statistics courses, and testing the validity of its assumptions is an important part of regression modeling and analysis. The traditional approach in introductory statistics courses is to use residual plots to assess possible departures from the key assumptions. Confirmatory statistics are typically not used. But do undergraduate students in these introductory courses have the ability to assess residual plots correctly? This paper presents the results of an Institutional Review Board (IRB) approved study to address this question. The results are interesting and have implications for the teaching of regression in introductory courses as well as for textbook authors who cater to these courses.

Keywords: regression assumptions, normality, homoscedasticity, residual analysis

Undergraduate students in introductory statistics classes are typically presented with residual plots when assessing the assumptions of regression modeling. Aside from the Durbin-Watson (1951) procedure for assessing the independence assumption, no introductory statistics text covers confirmatory approaches to residual analysis for the assumptions of linearity, normality and homoscedasticity. The basic question addressed in this paper is whether undergraduate students have enough experience and ability to accurately assess residual plots. If this is not the case it would be necessary to introduce a confirmatory, inferential approach to residual analysis that would supplement the graphical, exploratory approach that is currently in vogue.

The inability of humans to distinguish between random and non-random patterns in other contexts has been well-documented. When people are asked to produce a sequence of random coin tosses they produce sequences with far too many alternations in the erroneous belief that randomness means more alternations (Wagenaar, 1970). Other well-known examples where intuition regarding randomness generally fails, are the gambler's fallacy (the belief that because the run of bad luck has been so long it must turn in the gambler's favor) or the "hot hand" in basketball – the belief that players can have long streaks of scoring that cannot possibly be random (Gilovich, Vallone, & Tversky, 1985).

It is surprising then, that introductory statistics text treatments appear to assume that students are an exception to such findings and would not have trouble distinguishing random from non-random plots. Even journal articles on the teaching of regression assume, without question, that students can draw the appropriate conclusions about regression assumptions from an examination of residual plots (Richardson, Gabrosek, Reischman, & Curtis, 2004; Matson & Huguenard, 2007; Kuiperman, 2008).

When undertaking a study to determine how effective a graphical approach to residual analysis is, it becomes essential to assess contributing factors leading to correct student decisions regarding the viability of the regression model assumptions. Cleveland, McGill, and McGill (1988) and Cook and Weisberg (1994) demonstrated why the aspect ratio of a scatterplot is often so essential to its understanding. A second factor that would likely impact a student's graphic assessment of the assumptions is plot density, i.e., the number of plotted points observed in a residual plot. A third factor would be plot characteristic -- the type and severity of assumption violation encountered in the residual plot and the potential for confounding departures from linearity with departures from homoscedasticity.

The rest of this paper proceeds as follows. The next section is a brief review of the regression assumptions along with the approaches typically taken (and that could be taken) in introductory courses to verifying them. This is followed by a section which provides details of the experimental setup for the study's major and supplementary analyses. The results of the experiment are discussed in the fourth section and is divided in two parts – the first part dealing with the overall ability of student subjects to correctly identify the residual plots they were shown, and the second part discussing results in detail, focusing specifically on their ability to distinguish between linearity and homoscedasticity violations. The paper concludes by highlighting the findings and offering recommendations.

Simple Linear Regression Assumptions

The statistical model for simple linear regression is:

$$Y_i = \alpha + \beta X_i + \varepsilon_i \quad (1)$$

The assumptions underlying this model have to do with the errors ε_i and are well-known:

Linearity: $E(Y_i) = \alpha + \beta X_i$

Independence: $\varepsilon_i, \varepsilon_j$ are independent for any $i \neq j$.

Normality: $\varepsilon_i \sim N(0, \sigma^2)$

Homoscedasticity: $\text{Var}(\varepsilon_i) = \sigma^2$

The standard approach used in introductory statistics classes to assessing the validity of these assumptions is graphical:

- *Linearity:* Scatterplots of Y against the predictor variable X should be reasonably linear. Further, residuals are plotted against either fitted values or against the predictor variable X_i ; departures from random noise could indicate non-linearity.
- *Independence:* Residuals are plotted against time-ordered independent variables or in the sequence the data are presented. Non-random scatterplots would indicate departures from independence.
- *Normality:* A histogram, stem-and-leaf display, boxplot or normal probability plot of the residuals is used to visually assess this assumption.
- *Homoscedasticity:* Residuals are plotted against fitted values or against the predictor variable X_i . If these residual plots show any thickening or funnel-shaped behavior it could be taken as evidence of violation of this assumption.

Confirmatory statistical tests exist for the assumptions of linearity (e.g., F test for lack of fit – Berenson, Levine, & Goldstein, 1983), independence (e.g., Durbin & Watson, 1951), normality (e.g., Anderson & Darling, 1952), and homoscedasticity (e.g., Cook & Weisberg, 1983). However, in a first course in statistics at the undergraduate level, it is rare for the discussion to cover these tests. Although software like SPSS, Minitab, or JMP do provide students easy access to these tests, most statistics classes tend to rely on Excel and, unfortunately, its Data Analysis add-in does not provide these tests.

Evaluation of statistical plots, graphs, and displays has been an especially active area of research since the advent of personal computers. Pioneers like Tufte (1983), and Cleveland (1993) have forcefully made the case for careful design of visual displays if quantitative information is to be conveyed and comprehended accurately. The scatterplot has been one of the earliest tools for the presentation and analysis of bivariate information (Friendly & Denis, 2005). Much work has been done on how well scatterplots convey correlation (Pollock, 1960; Bobko & Karren, 1979; Cleveland, Diaconis, & McGill, 1982; Lauer & Post 1989; Meyer, Taieb, & Flascher, 2005; Doherty, Anderson, & Angott, 2007). The general finding from these studies is that correlations are hard to perceive when $|r| < 0.2$ and are typically underestimated when $|r|$ is between 0.2 and 0.6.

STUDENT ABILITY TO INTERPRET REGRESSION DIAGNOSTIC PLOTS

Regression diagnostic residual plots, however, do not involve estimation. The judgmental task is to assess whether a plot is random or shows non-random patterns. Departures from randomness might involve non-linearity or variability of spread, and research is needed to investigate the ability of students to distinguish between random and non-random plots in this context. Delmas, Garfield, Ooms, & Chance (2007), in their assessment of students' conceptual understanding after a first course in statistics, found that students did an adequate job of matching a scatterplot to a verbal description. But this task focused on bivariate linear relationships and not on regression diagnostic plots. Thus, there is no evidence on how effectively undergraduate students are able to use residual plots as regression diagnostic tools. In this sense, the present study represents a first attempt at answering this question.

The Experimental Setup

Overview

To keep the assessment task manageable for the student subjects, the study focused on only two of the regression assumptions – linearity and homoscedasticity. The other two assumptions were ignored for the following reasons: Independence is an important issue with time-series data; observations reported through survey data and often through experimental data are typically not time dependent. Normality is an important concern with small-sized samples; however, for studies containing data sets with large samples, the Central Limit Theorem would be expected to ensure at least approximate normality in the sampling distribution of the slope, making inferences viable.

Key Factors

Previous research on the assessment of correlations through scatterplots has revealed the importance of display factors such as the size of the plotting character, the overall size of the display, the orientation of the point cloud within the frame, and the size of the point cloud within the frame (Cleveland, Diaconis, & McGill, 1982; Cook & Weisburg, 1983). Lauer & Post (1989) demonstrate that density of scatterplots (i.e., the number of points in the plot) influences the estimation of correlation.

For the purposes of this study, three key factors were considered – plot characteristic, display, and plot density. The plot characteristic factor had to do with whether or not the residual plot incorporated a departure from the assumption of linearity and/or a departure from the assumption of homoscedasticity. Four different plot characteristics were thus possible: (i) linearity and homoscedasticity, (ii) linearity and heteroscedasticity, (iii) non-linearity and homoscedasticity, and (iv) non-linearity and heteroscedasticity. The display factor was the residual plot's aspect ratio – the ratio of its height to its width. Two ratios were used – the first had a default aspect ratio in R of 0.90:1, the other aspect ratio was half of this at 0.45:1. The plot density factor was the sample size or number of points on a residual plot. Two levels of plot density – low (i.e., 20 points) and high (i.e., 60 points) were used. A higher plot density would be expected to reveal plot characteristics more clearly in much the same way that a larger number of pixels improves the resolution of a picture.

The student subjects were presented with a series of plots that varied on the three factors discussed above and, for each, were asked to answer simple classification questions. How accurately they performed this classification is the central issue in this study.

STUDENT ABILITY TO INTERPRET REGRESSION DIAGNOSTIC PLOTS

Subjects

Subjects were undergraduate students enrolled in introductory statistics classes in a recent academic year. Participation in this IRB-approved study was voluntary but students were given extra credit by their instructors if they did so. Participating students were sent a link that took them to an online questionnaire. The responses were anonymous and not shared with the instructors of the participating classes; instructors were only given a list of participants so that the extra credit could be assigned.

Questionnaire

Subjects had to answer the following two questions for each of the residual plots presented to them:

- a. *Do the points on this graph show any evidence of a curved pattern?*
☐ Yes
☐ No, the points appear to be scattered randomly throughout
- b. *Do the points on this graph show any evidence of increasing vertical spread as you look from left to right?*
☐ Yes, the points appear to show increasing vertical spread from left to right
☐ No

To prepare them for the assessment task, subjects were shown examples of the kinds of residual plots they would be evaluating along with the questions they would be answering. They then proceeded to the main assessment task. With four plot characteristics, two aspect ratios, and two plot densities, a total of 16 unique plots are possible. In addition, each subject was asked to assess two “check” plots – one clearly violating both assumptions and the other clearly compliant with both assumptions. This was done as a check for internal validity, to see if the subject understood the assessment task and was paying enough attention to it. Subjects who failed to identify these two plots correctly were removed from the analysis regardless of their responses to all the other plots.

Once subjects finished answering the above questions on these 18 plots, they were asked four simple demographic questions before exiting the online survey. These four demographic questions pertained to age, gender, class designation (i.e., freshmen, sophomore, junior, senior), and college or school in which they were enrolled. Subjects were instructed to allow about 15-20 minutes to complete the questionnaire. To get the extra credit for their efforts, subjects had to have answered both questions on all 18 residual plots. Responses to the demographic questions were optional.

Development of the Residual Plots

The residual plots generated had four possible plot characteristics – (i) linearity and homoscedasticity, (ii) linearity and heteroscedasticity, (iii) non-linearity and homoscedasticity, and (iv) non-linearity and heteroscedasticity. Details of the underlying mathematical structure for each of these plots are shown in Appendix A. Plots were generated in R (R Development Core Team 2008) using R’s default aspect ratio of 0.9:1 (height to width) denoted AR=D (for *default*) and another aspect ratio of 0.45:1, denoted AR=H (for *half* the default ratio). Plot density was either 20 points or 60 points. Appendix B shows some examples of the actual residual plots that

STUDENT ABILITY TO INTERPRET REGRESSION DIAGNOSTIC PLOTS

subjects were required to evaluate. The full questionnaire may be obtained by writing to the lead author.

Randomization of Treatments to Student Subjects

A student subject was presented with a randomly generated sequence of the 16 plots keeping in mind the need for balance of treatments. Included in each sequence were the two additional “check” plots as described earlier, making for a total of 18 plots that the subject had to address.

Mode of Response

For each of the residual plots presented, subjects were asked to answer two questions – one involving perception of violation of the linearity assumption and the other involving perception of the violation of the homoscedasticity assumption. Therefore, excluding the responses to the two “check” plots, this study resulted in 32 binary responses per student subject, each simply scored as correct (1) or incorrect (0). The total number of correct responses was then transformed to percent correct response and reported as a score on a 0 to 100 scale.

Sample Size Determination Based on Intended Effect Size and Power of Test

Based on an initial pilot assessment, the mean (i.e., percent) correct response and variability were estimated. Thus, for a level of significance α of 0.05 combined with a desired level of power $(1 - \beta)$ of 0.90 to detect a specific effect size difference in mean correct response between the two treatment conditions of $\delta=5$, the minimum sample size was 168 (Fleiss, 1986; Cohen, 1988).

Responses

515 students participated in the study. Of these, 153 students were eliminated for having responded incorrectly to at least one of the four internal check questions. Thus, the final sample size realized in the study was 362, well over twice the minimum needed.

An analysis of the group of 153 students excluded from the study showed that they did not deviate from the study group with respect to the age, gender, or class designation covariates, but did deviate significantly with respect to college or school enrollment ($\chi^2 = 9.44$ with 2 degrees of freedom; p -value = 0.0089). Table 1 below displays the breakdown of student subjects retained or excluded by college or school of enrollment. Although we could not ascertain why only 67.4% of the business school students were retained for the study versus 83.9% of science and mathematics students, we were able to compare the performance of the retained students from these two academic units. We found no significant difference between the two groups in the percentage correct responses to our experimental questions ($t=0.607$ with 352 degrees of freedom; p -value = 0.5444). Thus, we concluded that the enrollment differences were not a concern. So, for the purposes of the analysis described in detail in the next section, the entire study group was treated as one.

STUDENT ABILITY TO INTERPRET REGRESSION DIAGNOSTIC PLOTS

College	Kept	Removed	Total
SBUS	281	136	417
	67.4%	32.6%	100.0%
CSAM	73	14	87
	83.9%	16.1%	100.0%
Other	8	3	11
	72.7%	27.3%	100.0%
Total	362	153	515
	70.3%	29.7%	100.0%

Table 1. Subjects Included in Study by College

* (SBUS: School of Business; CSAM: College of Science and Mathematics)

Analysis of Results

We will analyze the results in two parts. In Part A, we will focus on the overall ability of our student subjects to identify violations of linearity and homoscedasticity in residual plots. This will be followed by Part B, where we focus on the two questions subjects answered – on curve and spread. In both these parts we discuss the impact of the type of violations, aspect ratio, and plot density on the ability of the subjects to correctly identify the plots.

Analysis-Part A

Table 2 below summarizes the main descriptive results obtained from the experiment. For each residual plot that was shown to them the subjects had to answer two questions, one pertaining to linearity (Qa) and the other to homoscedasticity (Qb). The table shows, for each of the 16 residual plots, the percent of respondents that answered both, one, or none of the questions correctly. Note that the internal validity “check” plot questions were answered correctly by all 362 subjects and are not shown in the table.

Overall, 62.5% of the respondents were able to answer both questions correctly and 5.1% got both questions wrong. 87.1% got Qa right and 70.3% got Qb right.

Research Question 1: Can student subjects identify violations of linearity and homoscedasticity by examining residual plots? A plot is identified correctly if *both* questions are answered correctly. Figure 1, derived from the data in Table 2, shows this percentage for the different kinds of violations.

Although the results shown in Figure 1 indicate a “visual literacy” significantly superior to what would be expected through mere guessing (i.e., 25.0 %), they nevertheless highlight a serious problem. Roughly three of every eight students do *not* have the ability to properly assess both the linearity and homoscedasticity assumptions in a simple linear regression analysis by looking at residual plots and it would be prudent for instructors of these introductory courses and textbook authors to supplement these with confirmatory approaches.

STUDENT ABILITY TO INTERPRET REGRESSION DIAGNOSTIC PLOTS

Plot Characteristic (Assumption Violated)	Aspect Ratio	Plot Density	Both Correct	Qa Corr, Qb Wrong	Qa Wrong, Qb Corr	Both Wrong
None	D	20	85.4%	11.6%	2.2%	0.8%
None	D	60	94.2%	5.2%	0.6%	0.0%
None	H	20	83.1%	9.9%	5.0%	1.9%
None	H	60	77.6%	7.2%	10.8%	4.4%
Linearity	D	20	63.0%	29.3%	4.7%	3.0%
Linearity	D	60	51.7%	46.7%	0.6%	1.1%
Linearity	H	20	64.4%	33.1%	1.4%	1.1%
Linearity	H	60	59.1%	39.5%	1.1%	0.3%
Homosced	D	20	47.5%	48.1%	2.5%	1.9%
Homosced	D	60	56.9%	26.5%	13.8%	2.8%
Homosced	H	20	50.0%	43.9%	2.8%	3.3%
Homosced	H	60	52.8%	24.9%	15.2%	7.2%
Both	D	20	52.2%	15.5%	17.4%	14.9%
Both	D	60	58.8%	20.4%	9.9%	10.8%
Both	H	20	31.8%	13.5%	30.9%	23.8%
Both	H	60	71.3%	18.0%	6.6%	4.1%
Average			62.5%	24.6%	7.8%	5.1%

Table 2. Summary of Results: Percent Response to Paired Residual Plot Questions.

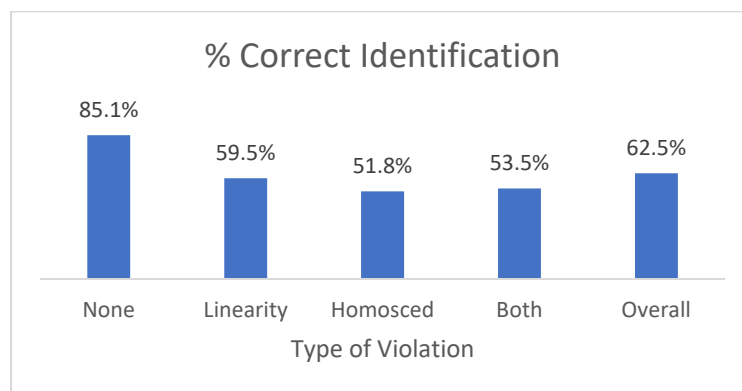


Figure 1. Correct Identification and Plot Characteristics.

From Table 2 it is clear that there are systematic differences in the ability to assess a plot correctly and that the type of violation, the aspect ratio, and the plot density all have impact on the extent of correct identification.

Research Question 2: Do subjects make more correct identifications for some violations than for others? Using a randomized complete block design, each of the 362 subjects answered 8 questions pertaining to each of the 4 plot types (*None, Linearity, Homoscedasticity, Both*). A 2-way ANOVA based on 3 and 1083 degrees of freedom found significant differences in students' abilities to correctly identify the different kinds of violations ($F=108.16$; p -value=0.0000). Further, each pairwise comparison based on Tukey's *a posteriori* HSD procedure indicated significant differences, confirming the message of Figure 1.

Research Question 3: Is there evidence of a difference in subjects' ability to assess the linearity assumption versus the homoscedasticity assumption in a residual plot? Each subject examined 16 plots and for each one answered a question about linearity and another about homoscedasticity. A paired t -test compared the percent correct responses to the linearity and the homoscedasticity questions for the 362 subjects, and found that subjects had highly significantly better perception of the assumption of linearity than of homoscedasticity. ($t = 16.05$ with 361 degrees of freedom; p -value = 0.0000).

Research Question 4: Is there evidence of a difference in subjects' ability to correctly assess the residual plots based on plot density? A paired t -test was used to compare the percent correct responses by the 362 subjects to the 16 questions with low plot density versus the corresponding 16 questions with high plot density. The test indicated a significant increase in the percentage of correct responses when the plot density was high. ($t = 6.31$ with 361 degrees of freedom; p -value = 0.0000).

Research Question 5: Is there evidence of a difference in subjects' ability to correctly assess the residual plots based on aspect ratio? A paired t -test was used to compare the percent correct responses by the 362 subjects to the 16 questions with the AR=D aspect ratio versus the corresponding 16 questions with the AR=H aspect ratio. The test indicated a significantly larger percentage of correct responses with the AR=D aspect ratio ($t=3.19$ with 361 degrees of freedom; p -value=0.0015).

Analysis-Part B

As the results of the previous section show, residual plots are helpful in identifying the violations of the linearity and homoscedasticity assumptions of linear regression. But some combinations of violation types seem to be easier to detect than others, and violations of linearity are easier to detect than violations of homoscedasticity. Further, both aspect ratio and plot density seem to influence the correct identification of plots. The AR=D aspect ratio and higher plot density, in general, are associated with higher percentages of correct identification. However, as can be seen from Table 3a, there are inconsistencies. For example, when only the assumption of linearity is violated, it is observed that students correctly identify this more frequently with low aspect ratio plots and low plot density than with other combinations. In this section, we look at these issues in more detail.

In the rest of this section we will focus on the two questions we asked the student subjects and analyze their responses.

The Curve Question: Do the points on this graph show any evidence of a curved pattern? Table-3a below shows the percentage of *incorrect* answers of the curve question

STUDENT ABILITY TO INTERPRET REGRESSION DIAGNOSTIC PLOTS

(regardless of how the accompanying spread question was answered), based on plot type and aspect ratio.

Type of Violation	AR=D	AR=H	Overall
Both	26.3%	32.1%	29.2%
Homoscedasticity	10.4%	14.2%	12.3%
Linearity	4.2%	1.7%	3.0%
None	2.0%	10.8%	6.4%
Overall	10.7%	14.7%	12.7%

Table 3a. The Curve Question - % Incorrect Identification and Aspect Ratio

Students seem to have most trouble with this question when homoscedasticity is violated – by itself or in combination with non-linearity. This suggests that in the presence of heteroscedasticity, students are, to a large degree, confused. They don't see curvature when it is present, and they see curvature when it is not present. As Table 3a shows, these findings hold good for either aspect ratio. In general, the AR=D aspect ratio is associated with lower percentage of incorrect identifications (10.7% vs 14.7%). This holds good for each plot type except in the case where only linearity is violated.

Table 3b below shows the impact of plot density on percent incorrect identifications.

Type of Violation	PD=20	PD=60	Overall
Both	43.0%	15.5%	29.2%
Homoscedasticity	5.1%	19.6%	12.3%
Linearity	4.5%	1.4%	3.0%
None	4.9%	7.9%	6.4%
Overall	14.4%	11.1%	12.7%

Table 3b. The Curve Question - % Incorrect identification and Plot Density

As we would expect, overall, a higher plot density seems to be associated with a lower percentage of misidentifications. But looking at the individual cases, the effect is mixed. When both types of violations are present or only linearity is violated, a higher PD is associated with lower misidentification percentages. When there are no violations or only homoscedasticity is violated, a higher plot density is associated with *higher* misidentification percentages. This suggests that when there is curvature, it is revealed more clearly with higher PD than with lower PD. But a larger PD seems to encourage students to see curvature where there isn't any (when only homoscedasticity is violated or there are no violations). These explanations are very tentative and have to be validated through further investigation.

The Spread Question: Do the points on this graph show any evidence of increasing vertical spread as you look from left-to-right? Table 4a below shows percentage incorrect answers to this question broken down by plot type and aspect ratio. Again, students seem to do best when neither linearity nor homoscedasticity is violated. With any violation, the increase in incorrect answers to this question is dramatic. But the results are puzzling. One possible

STUDENT ABILITY TO INTERPRET REGRESSION DIAGNOSTIC PLOTS

explanation is that changes in spread of points in a residual plot are just hard for students to identify – hence the high percentages of misidentifications (30.1% and 41.1%) in the two cases where homoscedasticity is violated. In the case where only linearity is violated, the percentage of misidentifications was 38.2%. A tentative explanation that has to be verified through further investigation is that students mistake linearity violations for spread.

Type of Violation	AR=D	AR=H	Overall
Both	30.8%	29.4%	30.1%
Homoscedasticity	39.7%	42.4%	41.1%
Linearity	39.9%	36.6%	38.2%
None	7.2%	10.8%	9.0%
Overall	29.4%	29.8%	29.6%

Table 4a. The Spread Question - % Incorrect Identification and Aspect Ratio

Aspect ratio seems to make statistically significant differences (given our large sample size), but its practical significance is questionable. Further, the direction of the impact is mixed, with one aspect ratio (AR=D) being superior in two cases and worse in the other two cases.

Table 4b below shows the break-down by plot density to highlight its effect.

Type of Violation	PD=20	PD=60	Overall
Both	34.1%	26.2%	30.1%
Homoscedasticity	51.0%	31.1%	41.1%
Linearity	32.7%	43.8%	38.2%
None	10.7%	7.3%	9.0%
Overall	32.1%	27.1%	29.6%

Table 4b. The Spread Question - % Incorrect Identification and Plot Density

The larger plot density is overall better than the smaller plot density. The differences are quite large (nearly 20% when only homoscedasticity is violated). This is true for each plot type except for the case where only linearity is violated.

Summary of Key Findings and Recommendations

This research effort was undertaken to examine the ability of student subjects to correctly identify residual plots that differed in the presence or absence of violations of the linearity and homoscedasticity assumptions of linear regression. The plots also varied in terms of their aspect ratios and plot densities.

In summary, our results show the following:

1. Students were able to correctly identify 62.5% of the residual plots in terms of their indicating departures from the linearity and homoscedasticity assumptions and this is

STUDENT ABILITY TO INTERPRET REGRESSION DIAGNOSTIC PLOTS

- significantly better than the 25% identification rate they would have achieved if they were simply guessing.
2. Nevertheless, the results flag a serious problem – 37.5% or three of every eight students were unable to correctly assess both the linearity and homogeneity assumptions in a simple linear regression analysis through residual plots.
 3. Some kinds of violations were more readily identified correctly than others. Plots with no violations in them had the highest correct identification percentage (85.1%). Once there were violations of some kind, correct identifications fell precipitously to 59.5% when only linearity was violated, and to 51.8% when only homoscedasticity was violated. These numbers have to be of concern to anyone who teaches regression or writes about it. The exclusive reliance on residual plots in introductory statistics textbooks to assess regression assumptions seems misplaced.
 4. The results show that questions of spread were harder to correctly answer than questions of linearity. The reasons for this are not obvious and must be the subject of further investigation. Plausible explanations have to do with unfamiliarity with the very concept of homoscedasticity. Students are exposed to the concept of linearity in other contexts like testing for normality (probability plots), correlation, lines of best fit, etc. But their first exposure to the concept of homoscedasticity comes when they discuss the regression assumptions. Therefore, it is possible that student subjects confused non-linear plots with heteroscedastic plots and vice versa.
 5. Aspect ratio and plot density made a difference. In general, larger aspect (height-to-width) ratio and larger plot densities aided the correct identification of plots. However, their impact was situation specific. In some instances, their effect was not what we would have expected.

The findings herein clearly demonstrate that although residual plots are important, too many inexperienced individuals have difficulty in deciphering the underlying message of the residual plots, particularly when heteroscedasticity is present. Supplementing a graphical residual analysis of a model's assumptions with appropriate confirmatory approaches would surely enhance the student's ability in this regard. For the assumption of homoscedasticity, a confirmatory procedure developed by Glejser (1969) could be employed in a one-semester undergraduate introductory statistics class. Glejser's procedure is a *t*-test for the significance of the slope in a secondary simple linear regression analysis that uses the absolute residual values as a function of the independent variable.

As mentioned in the section titled 'Simple Linear Regression Assumptions', confirmatory approaches are available in the popular statistical packages for all the regression assumptions. We realize that the details of some of these tests which are based on the χ^2 or the *F* statistic are beyond the scope of a one-semester undergraduate introductory statistics course. However, it should be possible to show students the results of the tests along with *p*-values which they can be taught to interpret. This will strengthen their intuition regarding residual plots and lead to more accurate diagnoses and conclusions about the regression assumptions. Writers and teachers should incorporate these approaches in their books and classes without delay.

As previously stated, further research in this area is clearly needed. We intend to build on the results of this study and systematically evaluate not just the type of violation but also the severity of the violation. Additional levels of aspect ratio and plot density will be included in the

STUDENT ABILITY TO INTERPRET REGRESSION DIAGNOSTIC PLOTS

design and the research will specifically look at the confounding of nonlinearity with heteroscedasticity in the residual plots.

References

- Anderson, T. W., & Darling, D. A. (1952). Asymptotic theory of certain “goodness of fit” test criteria based on stochastic processes. *Annals of Mathematical Statistics* 23, 193-212.
- Anscombe, F. J. (1973). Graphs in statistical analysis. *The American Statistician* 27, 17-21.
- Berenson, M. L., Levine, D. M., & Goldstein, M. (1983). *Intermediate statistical methods and applications*. Englewood Cliffs, NJ: Prentice-Hall.
- Bobko, P., & Karren, R. (1979). The perception of Pearson product moment correlations from bivariate scatterplots. *Personnel Psychology* 32, 313-325.
- Cleveland, W. S., Diaconis, P., & McGill, R. (1982). Variables on scatterplots look more highly correlated when scales are increased. *Science* 216, 1138-1141.
- Cleveland, W. S., McGill, M. E., & McGill, R. (1988). The shape parameter of a two-variable graph. *Journal of the American Statistical Association* 83, 289-300.
- Cleveland, W. S. (1993). *Visualizing data*. Summit, NJ: Hobart Press.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Earlbaum Associates.
- Cook, R. D., & Weisberg, S. (1983). Diagnostics for heteroscedasticity in regression. *Biometrika* 70, 1-10.
- Cook, R. D., & Weisberg, S. (1994). *An introduction to regression graphics*. NY: Wiley.
- Delmas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students’ conceptual understanding after a first course in Statistics. *Statistics Education Research Journal* 6, 28-58.
- Doherty, M. E., Anderson, R. B., Angott, A. M., & Klopfer, D. S. (2007). The perception of scatterplots. *Perception & Psychophysics* 69, 1261-1272.
- Durbin, J., & Watson, G.S., (1951). Testing for serial correlation in least squares regression (II). *Biometrika* 38, 159-178.
- Fleiss, J. L. (1986). *The design and analysis of clinical experiments*. NY: Wiley.
- Friendly, M., & Denis, D. (2005). The early origins and development of the scatterplot. *Journal of the History of the Behavioral Sciences* 41, 103-130.
- Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology* 17 (3), 295-314.

STUDENT ABILITY TO INTERPRET REGRESSION DIAGNOSTIC PLOTS

- Glejser, H. (1969). A new test for heteroscedasticity. *Journal of the American Statistical Association* 64 (3), 315-323.
- Kuiperman, S. (2008). Introduction to multiple regression: How much is your car worth? *Journal of Statistics Education* 16, 8.
- Lauer, T. W., & Post G. V. (1989). Density in scatterplots and the estimation of correlation. *Behaviour & Information Technology* 8, 235-244.
- Matson, J.E. & Huguenard, B.R. (2007). Evaluating aptness of a regression model. *Journal of Statistics Education* 15, 2.
- Meyer, J., Taieb, M., & Flascher, I. (2005). Correlation estimates as perceptual judgments. *Journal of Experimental Psychology: Applied* 3, 3-20.
- Pollock, I. (1960). Identification of visual correlational scatterplots. *Journal of Experimental Psychology* 59, 351-360.
- Richardson, M., Gabrosek, J., Reischman, D., & Curtiss, P. (2004). Morse code, Scrabble, and the alphabet. *Journal of Statistics Education* 12, 3.
- Tufte, E.R. (1983). *The visual display of quantitative information*. Cheshire, CT: Graphics Press.
- Wagenaar, W. A. (1972). "Generation of random sequences by human subjects: A critical survey of literature. *Psychological Bulletin* 77 (1), 65.

Appendix A-Developing the Residual Plots

The residual plots were generated by using the following models:

1. Linear model with constant variances: $Y=5+x+e$, where x is generated from $Uniform(0, 50)$, the random error term e is generated from $N(0,5)$
2. Nonlinear model with constant variances: $Y=5+x+0.05*x^2+e$, where x is generated from $Uniform(0, 50)$, e is generated from $N(0, 5)$.
3. Linear model with non-constant variances: $Y=5+x+0.5*x*e$, where $x \sim Uniform(0,50)$, $e \sim N(0, 5)$.
4. Nonlinear model with non-constant variances: $Y=5+x+0.2x^2+0.5*x*e$, where $x \sim Uniform(0, 50)$, $e \sim N(0,5)$

The four models were used to generate plots with two sample sizes: low plot density or small sample ($n=20$) and high plot density or large sample ($n=60$).

In addition, two control cases were generated to screen participants:

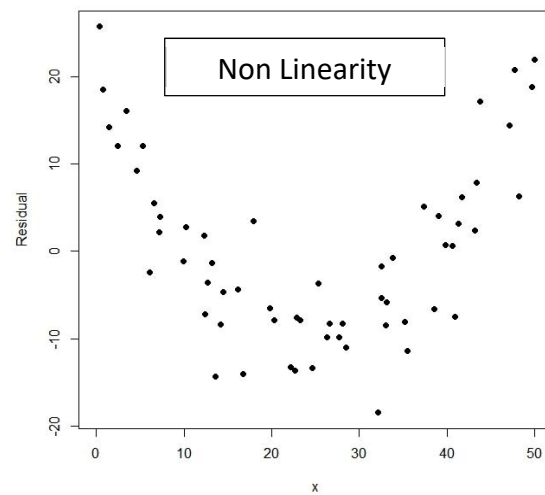
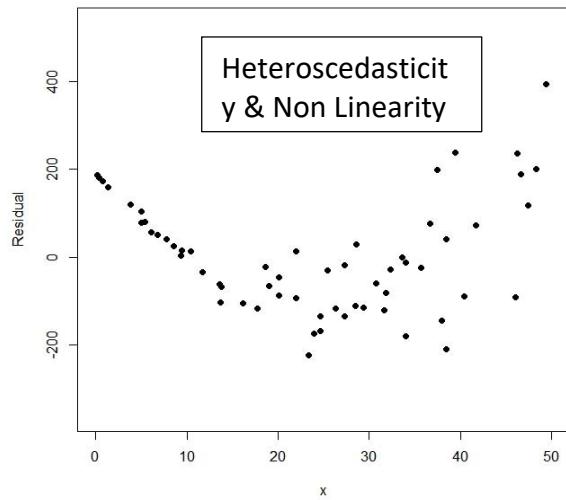
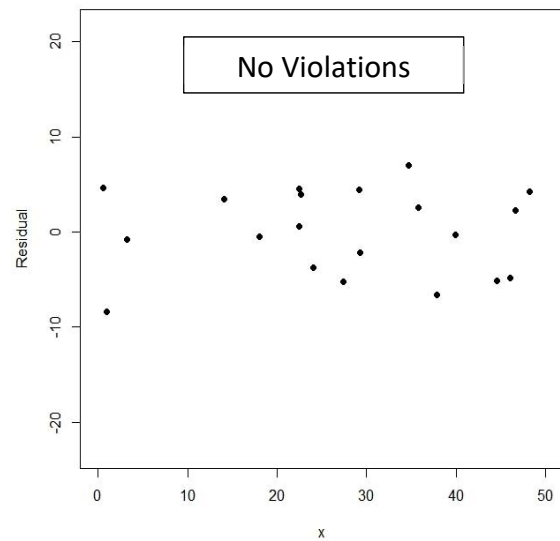
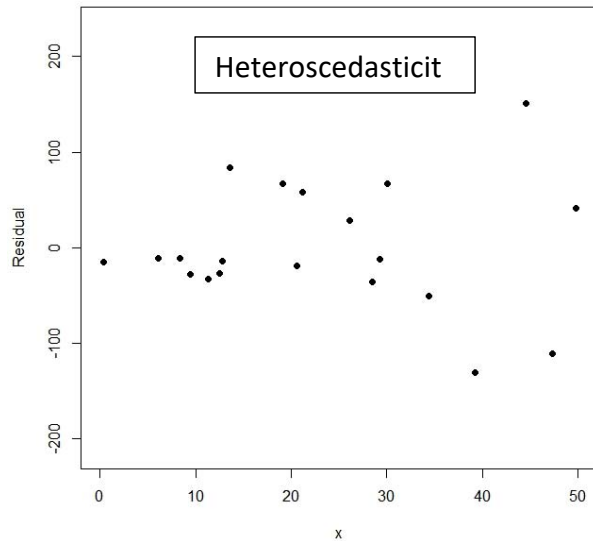
Control Case 1: Obvious ‘no violation’ case for small sample ($n=20$):
 $Y=5+x+e$, $x \sim Uniform(0, 50)$, $e \sim N(0, 1)$.

Control Case 2: Obvious violation of both linearity and homoscedasticity for large sample ($n=60$).
 $Y=5+x+0.5x^2+x*e$, $x \sim Uniform(0, 50)$, $e \sim N(0, 5)$.

All the figures were plotted in Statistical software R (R Development Core Team (2008)).

STUDENT ABILITY TO INTERPRET REGRESSION DIAGNOSTIC PLOTS

Appendix B-Example Residual Plots



Subjects were shown 18 such plots. For each plot they had to answer the following two questions:

Do the points on this graph show any evidence of a curved pattern?

- ☐ Yes
- ☐ No, the points appear to be scattered randomly throughout

Do the points on this graph show any evidence of increasing vertical spread as you look from left to right?

- ☐ Yes, the points appear to show increasing vertical spread from left to right
- ☐ No