SciMedCentral

JSM Mathematics and Statistics

Research Article

Comparing Tests of Homoscedasticity in Simple Linear Regression

Haiyan Su[1]* and Mark L Berenson[2]

[1]*Department of Mathematical Sciences, Montclair State University, USA*
[2]*Department of Information Management and Business Analytics, Montclair State University, USA*

***Corresponding author**
Haiyan Su, Department of Mathematical Sciences, Montclair State University, 1 Normal Ave, Montclair, NJ 07043, USA, Tel: 973-655-3279; Email: suh@montclair.edu

Submitted: 03 October 2017
Accepted: 10 November 2017
Published: 13 November 2017

Keywords
- Simple linear regression
- Homoscedasticity assumption
- Residual analysis
- Empirical and practical power

Abstract

Ongoing research puts into question the graphic literacy among introductory-level students when studying simple linear regression and the recommendation is to back up exploratory, graphical residual analyses with confirmatory tests. This paper provides an empirical power comparison of six procedures that can be used for testing the homoscedasticity assumption. Based on simulation studies, the GMS test is recommended when teaching an introductory-level undergraduate course while either the NWGQ test or the BPCW test can be recommended for teaching an introductory-level graduate course. In addition, to assist the instructor and textbook author in the selection of a particular test, areal data example is given to demonstrate their levels of simplicity and practicality.

INTRODUCTION

To demonstrate how to assess the aptness of a fitted simple linear regression model, all introductory textbooks use exploratory graphical residual analysis approaches to assess the assumptions of linearity, independence, normality, and homoscedasticity. However, our ongoing internal-review-board sponsored research (albeit at one university) indicates a deficiency in graphic literacy among introductory-level statistics students. Textbook authors and instructors should not assume that such students have sufficient experience/ability to appropriately assess the very important homoscedasticity assumption simply through graphic residual analyses.

This paper provides an empirical power comparison of six procedures that can be used for testing the homoscedasticity assumption in simple linear regression. In addition, an example is given that demonstrates the levels of simplicity of these six procedures. Based on Tukey's [1] concept of 'practical power,' the intent here is to assist introductory-level statistics course instructors and textbook authors in the selection of a particular test for homoscedasticity. More generally, however, it is recommended that a graphic residual analysis approach be coupled with a confirmatory approach for assessing all four simple linear regression modeling assumptions.

This paper is organized as follows. Section 2 provides a brief summary of approaches developed for assessing the assumption of homoscedasticity in regression analysis. Section 3 describes the six procedures selected for potential use in introductory-level statistics courses. Section 4 develops the Monte Carlo power study comparing these six procedures and Section 5 discusses the results from the Monte Carlo simulations. In Section 6 a small example illustrates how these six procedures are used and Section 7 provides the conclusions and recommendations.

ASSESSING THE ASSUMPTION OF HOMOSCEDASTICITY IN LINEAR REGRESSION

Introductory-level students learning simple linear regression analysis must be made aware of the consequences of a fitted model not meeting its assumptions. In particular, if the assumption of homoscedasticity is violated the ordinary least squares estimator of the slope is no longer efficient, estimates of the variances and standard errors are biased, and inferential methods are no longer valid.

In a seminal paper Anscombe [2] demonstrated the importance of graphical presentations to enhance understanding of what a data set is conveying and to assist in the model-building process for a regression analysis. In particular, he showed why a residual plot is a fundamental tool for uncovering violations in the assumptions of linearity and homoscedasticity.

Unfortunately, although inexperienced students may find the graphical demonstrations provided by Anscombe [2] to be clear, this does not imply they won't have difficulty in deciphering the underlying message residual plots are intended to be conveying in far more typical situations. On the contrary, subsequent research by Cleveland et al. [3], and Casey [4] have suggested a deficiency in graphic literacy among many inexperienced, introductory-level statistics students who have difficulty with reading and interpreting scatter plots. And our ongoing research on diagnostic plots corroborates the findings of Cleveland et al. [3], and Cook and Weisberg [5] that showed why the selected aspect ratio of the plot is essential to its understanding.

Given that residual plots are important, supplementing a graphical residual analysis of a model's assumptions with appropriate confirmatory approaches would surely enhance model development. The question that will be addressed in this paper is which confirmatory procedure is best for use in the introductory classroom.

Interestingly, in his earlier research Anscombe [6] dealt with more formal testing of the assumptions and his endeavors set in motion the development of several tests for the assumption of homoscedasticity.

In the 1960s Goldfeld and Quandt [7], Park [8], Glejser [9], and Ramsey [10] developed tests of homogeneity of variance still in use. Aside from the Park test [8], adaptations of the others have appeared in a few intermediate-level statistics texts. These tests are discussed further in Sections 3 and used in the Monte Carlo simulation in Section 5. Ramsey [10] had actually suggested four procedures, one of which (i.e., the RASET) seemed appropriate for introductory-level student audiences. The Park test [8], however, was not included in the Monte Carlo study because it would lack Tukey's "practical power" [1]. Too many introductory students seem to be insufficiently prepared for using logarithms and the procedure involves a secondary regression analysis of the log of the initial squared residuals on the log of the predictor variable.

In the 1970s Brown and Forsythe [11], Bickel [12], Harrison and McCabe [13], and Breusch-Pagan [14] derived other tests of homoscedasticity – the former and the latter have appeared in intermediate-level statistics texts and are discussed further in Sections 3 and 4 and also used in the Monte Carlo simulation in Section 6.

Bickel [12] investigated the power of Anscombe's procedures [6] and developed robust tests for homoscedasticity that are not intended for an introductory-level statistics course. On the other hand, Harrison and McCabe [13] proposed two tests, a bounds test and an "exact" test, and opined that the former had sufficient computational simplicity to merit use by practitioners. However, their Monte Carlo study indicated that their bounds test had "generally somewhat lower" power than the test by Goldfeld and Quandt [15] and was not considered further here.

In the 1980s White [16], Koenker [17], Ali and Giaccottto [6], and Engle [18] proposed tests of homogeneity of variance. A simplification provided by Berenson [19] enables White's test [16] to be considered for introductory-level statistics courses and is therefore presented in Sections 3 and used in the Monte Carlo simulation in Section 5. But none of the other three procedures were deemed appropriate for the introductory-level classroom. Koenker [17] modified the Breusch and Pagan [14] test by 'studentizing' the test statistic. Ali and Giaccotto [6] interpreted the assessment of heteroscedasticity as a shift in location of the distribution of the residuals or a shift in scale and proposed several nonparametric tests. Engle [18], working specifically in econometric modeling, developed a procedure for testing against conditional heteroscedasticity.

In the1990s, Godfrey [20] and Godfrey and Orme [15] provided modifications of the Glejser [9] and Koenker [17] procedures. At the turn of the new millennium, Im [21] and Marchado and Santos [22] independently developed additional modifications to the still popular Glejser test [9] so that it is more robust to skewness issues. Carapeto and Holt [23] developed a new test based on the Goldfeld–Quandt statistic [7] showed that it was very powerful at detecting heteroscedasticity, and then proposed a method for detecting particular forms of heteroscedasticity based on a neural network regression. Furno [24] developed a quantile regression approach using residuals estimated with least absolute deviations (LAD). Cai and Hayes [25] proposed a new hypothesis test for the overall significance of a multiple regression model by employing a heteroscedasticity-consistent covariance matrix (HCCM) estimator.

SIX TESTS FOR HOMOGENEITY OF VARIANCE

To assess the homoscedasticity assumption of fitted simple linear regression model, introductory level students have showed some difficulty and confusion in making the right conclusion using graphical approach based on our recent survey. In this section, we choose six confirmatory tests to supplement a traditional graphic residual analysis when assessing the homoscedasticity assumption in simple linear regression analysis. The selected six tests are relatively simple and methodological accessible to introductory level students. We compare the power of making the right decisions for testing homoscedasticity assumptions in simple regression models and then recommend the best ones for introductory level students. The six confirmatory are presented below.

NWRSR – The Neter-Wasserman / Ramsey / Spearman Rho T Test

Neter and Wasserman [26] suggested a procedure for assessing homoscedasticity using a *t*-test of Spearman's rank coefficient of correlation ρ_S based on a secondary analysis between the absolute value of the residuals from an initial linear regression analysis and the initial predictor variable. Their procedure yields identical results to Ramsey's RASET or Rank Specification Error Test [10], which employs the squared residuals in lieu of the absolute residuals proposed by Neter and Wasserman [26]. The NWRSR test is developed in four steps:

(**Step 1**) Perform an initial linear regression analysis and use the sample regression equation

$$\hat{Y}_i = b_0 + b_1 X_{1i}$$

to obtain the residual

$$e_i = Y_i - \hat{Y}_i$$

for each of the *n* observations.

(**Step 2**) Perform a secondary analysis to obtain the statistic r_s, the Spearman rank coefficient of correlation between $|e_i|$, the absolute value of the residuals, and the predictor variable X_{1i}, given by

$$r_S = \frac{\sum_{i=1}^{n}(R_{e_i} - \bar{R}_e)(R_{X_{1i}} - \bar{R}_{X_1})}{\sqrt{\sum_{i=1}^{n}(R_{e_i} - \bar{R}_e)^2 \sum_{i=1}^{n}(R_{X_{1i}} - \bar{R}_{X_1})^2}}$$

where R_{e_i} are the ranks from 1 to *n* of the $|e_i|$ and $R_{X_{1i}}$ are the ranks from 1 to *n* of the X_{1i_i} (such that any tied values are given the average of the ranks that would have been assigned had tied values not occurred) and where $\bar{R}_e$ and $\bar{R}_{X_1} = \frac{(n+1)}{2}$.

(**Step 3**) Under the null hypothesis that $\rho_S = 0$, the test statistic

$$t_{\rho_S} = \frac{r_s\sqrt{n-2}}{\sqrt{1-r_S^2}}$$

follows a t distribution with $n-2$ degrees of freedom.

(**Step 4**) Based on the *t*distribution with $n-2$ degrees of freedom of the NWRSE test statistic, a conclusion can be made about violation or no violation in the homoscedasticity assumption.

NWGQ – The Neter-Wasserman / Goldfeld-Quandt Test

Neter and Wasserman [26] proposed an extension of a procedure developed by Goldfeld and Quandt [7]. To develop this test the bivariate data set of *n* observations is first divided in half (or as close to half as possible, considering the value of *n* and possibility of ties) based on the ordered values of the independent variable X_{1i}. The NWGQ test is developed in four steps:

(**Step 1**) Perform a simple linear regression analysis based on the set of n_I observations with the *smaller* values of X_{1i}.

(**Step 2**) Perform a simple linear regression analysis based on the set of n_{II} observations with the *larger* values of X_{1i}.

(**Step 3**) Obtain the test statistic F_{NWGQ}, the ratio of the mean square error for the regression model based on the set of n_{II} observations with the *larger* values of X_{1i} to the mean square error for the regression model based on the set of n_I observations with the *smaller* values of X_{1i}. That is,

$$F_{NWGQ} = \frac{MSE_{II}}{MSE_I}.$$

Under the null hypothesis the test statistic F_{NWGQ} follows a F distribution with $n_{II} - 2$ and $n_I - 2$ degrees of freedom.

(**Step 4**) For this two-tailed test with level of significance *α*, if the test statistic F_{NWGQ} is greater than the F critical value with $n_{II} - 2$ and $n_I - 2$ degrees of freedom then the null hypothesis is rejected and the key homoscedasticity assumption in the regression analysis is deemed violated. Moreover, if the test statistic F_{NWGQ} is less than the reciprocal of the F critical value with $n_I - 2$ and $n_{II} - 2$ degrees of freedom then the null hypothesis is also rejected and the key homoscedasticity assumption in the regression analysis is deemed violated.

BF – The Brown-Forsythe test

Brown and Forsythe [11] developed a procedure for assessing homoscedasticity using a pooled-variance *t*-test based on a secondary analysis of the residuals obtained from an initial linear regression analysis. The BF test is developed in four steps:

(**Step 1**) Perform an initial linear regression analysis and use the sample regression equation

$$\hat{Y}_i = b_0 + b_1 X_{1i}$$

to obtain the residual

$$e_i = Y_i - \hat{Y}_i$$

for each of the *n* observations.

(**Step 2**) Divide the data set in half (or as close to half as possible considering possible ties in the measurements) based on the ordered values of the initial predictor variable X_{1i}. Group *I* contains the set of n_I residual observations based on the *smaller* values of X_{1i} and Group *II* contains the set of n_{II} residual observations based on the *larger* values of X_{1i} so that $n = n_I + n_{II}$. Determine M_{e_I} and $M_{e_{II}}$, the median residual for Groups *I* and *II*, and then obtain the sets of absolute residual differences $a_{I_i} = |e_{I_i} - M_{e_I}|$ and $a_{II_i} = |e_{II_i} - M_{e_{II}}|$ for the two groups. Perform a secondary analysis of the 'difference in the absolute residual differences' to obtain the Brown-Forsythe test statistic t_{BF}.

(**Step 3**) Under the null hypothesis of homoscedasticity the test statistic

$$t_{BF} = \frac{\bar{a}_{II} - \bar{a}_I}{S_{\bar{a}_{II} - \bar{a}_I}}$$

where

$$S_{\bar{a}_{II} - \bar{a}_I} = \sqrt{\left(\frac{\sum_{i=1}^{n_I}(a_{I_i} - \bar{a}_I)^2 + \sum_{i=1}^{n_{II}}(a_{II_i} - \bar{a}_{II})^2}{n-2}\right)\left(\frac{1}{n_I} + \frac{1}{n_{II}}\right)}$$

Follows a t distribution with $n-2$ degrees of freedom.

(**Step 4**) Based on the *t* distribution with $n-2$ degrees of freedom of the BF test statistic, a conclusion can be made about violation or no violation in the homoscedasticity assumption.

W – The White Test

White's test ([27],[19]) is based on a secondary regression analysis with the squared residuals from the initial linear regression analysis used as the dependent variable. The independent variables in the secondary regression analysis consist of the initial predictor variable and its square. The W test is developed in four steps:

(**Step 1**) Perform an initial linear regression analysis and use the sample regression equation

$$\hat{Y}_i = b_0 + b_1 X_{1i}$$

to obtain the residual

$$e_i = Y_i - \hat{Y}_i$$

for each of the *n* observations.

(**Step 2**) Perform a secondary regression analysis with the squared residuals e_i^2 as the dependent variable and the initial predictor variable and its square as the two independent variables. The secondary regression equation is

$$\hat{e}_i^2 = b_0^{'} + b_1^{'} X_{1i} + b_2^{'} X_{1i}^2$$

(**Step 3**) Use the secondary regression analysis to obtain the White test statistic $n\ {}^2$, the product of the sample size *n* and r^2, the "unadjusted" coefficient of multiple determination. Under the null hypothesis the White test statistic $n\ {}^2$ follows a χ^2 distribution with 2 degrees of freedom (i.e., the number of independent variables in the secondary regression equation).

(**Step 4**) Based on the χ^2 distribution with 2 degrees of freedom of the white test statistic, a conclusion can be made about violation or no violation in the homoscedasticity assumption.

BPCW – The Breusch-Pagan / Cook-Weisberg Scores Test

The Cook and Weisberg scores test ([28],[5]) is a generalized procedure that reduces to the Breusch and Pagan test ([14],[29]). This test is based on a secondary regression analysis with the squared residuals from the initial linear regression analysis used as the dependent variable. The independent variable in the secondary regression analysis consists of the initial predictor variable. The BPCW test is developed in four steps:

(**Step 1**) Perform an initial linear regression analysis and use the sample regression equation

$$\hat{Y}_i = b_0 + b_1 X_{1i}$$

to obtain the residual

$$e_i = Y_i - \hat{Y}_i$$

For each of the *n* observations

(**Step 2**) Perform a secondary regression analysis with the squared residuals e_i^2 as the dependent variable and the initial predictor variable as the independent variable. The secondary regression equation is

$$\hat{e}_i^2 = b_0^{+} + b_1^{+} X_{1i}$$

(**Step 3**) Use the secondary regression analysis to obtain the Breusch-Pagan / Cook-Weisberg test statistic $\chi_{BPCW}^{\ \ 2}$, computed as the ratio of the sum of squares due to regression in the secondary analysis

$$SSR^{+} = \sum_{i=1}^{n}\left(\hat{e}_i^2 - \frac{\sum_{i=1}^{n}\hat{e}_i^2}{n}\right)^2$$

to twice the square of the biased estimate of the mean square error in the initial linear regression analysis

$$2\left[\left(\frac{SSE}{n}\right)^2\right] = 2\left[\left(\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n}\right)^2\right].$$

Under the null hypothesis the test statistic $\chi_{BPCW}^{\ \ 2}$ follows a χ^2 distribution with 1 degree of freedom (i.e., the number of independent variables in the secondary regression equation).

(**Step 4**) Based on the χ^2 distribution with 1 degree of freedom of the BPCW test statistic, a conclusion can be made about violation or no violation in the homoscedasticity assumption.

GMS – The Glejser / Mendenhall-Sincich test

Mendenhall and Sincich [30] adopt a procedure proposed by Glejser [9] that is based on a secondary regression analysis with the absolute value of the residuals from the initial linear regression analysis used as the dependent variable. The independent variable in the secondary regression analysis is the initial predictor variable. The GMS test is developed in four steps:

(**Step 1**) Perform an initial linear regression analysis and use the sample regression equation

$$\hat{Y}_i = b_0 + b_1 X_{1i}$$

to obtain the residual

$$e_i = Y_i - \hat{Y}_i$$

for each of the *n* observations.

(**Step 2**) Perform a secondary regression analysis with the absolute value of the residuals $|e_i|$ as the dependent variable and the initial predictor variable as the independent variable. The secondary regression equation is

$$|\hat{e}_i| = b_0^{*} + b_1^{*} X_{1i}$$

(**Step 3**) Use the secondary regression analysis to obtain the test statistic $t_{GMS} = b_1^{*} / S_{b_1^{*}}$, the ratio of the slope to its standard error, which enables a *t*-test of the population slope β_1^{*} from the secondary analysis. Under the null hypothesis, the Glejser / Mendenhall-Sincich test statistic t_{GMS} follows a t distribution with $n-2$ degrees of freedom.

(**Step 4**) Based on the *t* distribution with $n-2$ degrees of freedom of the GMF test statistic, a conclusion can be made about violation or no violation in the homoscedasticity assumption.

THE MONTE CARLO SIMULATION STUDY

Table 1 presents the set of 24 condition-combinations used for a Monte Carlo power study. The imposition of particular condition-combinations was intended to determine:

1. Whether the six tests are "valid" and hold their specified α = 0.05 level of significance under the null hypothesis of homoscedasticity or whether any of the tests can be classified as either liberal or conservative.

2. Whether any test can be declared empirically most powerful in detecting heteroscedasticity when the standard deviation of the random error around the regression line is moderate to large, regardless of sample size, or whether one or more tests display superior power when the sample size is small while others provide superior power when the sample size is larger.

A total of 5,000 sets of bivariate data were generated for each of the above 24 condition-combinations. Without losing of generality, predictor variable *X* was generated from the uniform distribution from 5 to 10. Random error *e* was generated from the standardized normal distribution N (0,1). The constant *c* (*c* = 0.5 or *c* = 1) was used to control the standard deviation of the random error term in the model. In the homoscedasticity case, $Y = 1 + 2X + ce$ (where the random error *ce* has constant variance); in the heteroscedasticity Case I, $Y = 1 + 2X + cXe$ (where the error term *cXe* is correlated with the predictor *X*), and in the heteroscedasticity Case II, $Y= 1 + 2X + (X^c)\ e$ (where the error term $(X^c)e$ is correlated with the predictor *X*). The sample sizes considered were 20, 50, 100, and 200, which covered a range from "small" sample to "medium" sample to "relatively large" sample. Level of significance is set to be 5%. Statistical software R [31] was used for all the simulation studies.

For each of the data sets developed under the 24 condition-combinations the six tests of homoscedasticity were performed and a record was recorded of both the *p* value and a tally (1 or 0) pertaining to whether the actual violation in homoscedasticity was found. The empirical α levels and the empirical powers from the 5,000 repetitions were then obtained and the respective results are displayed in Tables 2-4.

RESULTS

Under the null hypothesis of homoscedasticity the empirical α levels were obtained for each of the six test procedures using the condition combinations *c* = 0.5 or 1.0 and *n* = 20, 50, 100, or 200. Since the standard error associated with a stated 0.05 α level of significance is 0.00308 (based on 5,000 repetitions), it is noted from Table 2 that the NWGQ, W, and GMS tests are "valid" for the conditions considered [32] since their empirical α levels or proportion of incorrect rejections never exceeded three standard errors of the nominal α level. On the other hand, the NWRSR test was liberal when *c* = 0.5 and *n* = 20 and valid otherwise while the BPCW test was conservative for very small sample size (*n* = 20) and valid otherwise. The BF test, however, was conservative for both smaller sample sizes (*n* = 20 and *n* = 50) when *c* = 1.0 and for the very small sample size (*n* = 20) when *c* = 0.5 and, therefore, cannot be recommended for use under such conditions.

With respect to empirical power, the Cochran Q test [33] showed that there were real differences among the six tests at the = 0.05 level of significance for each of the sixteen condition-combinations *c* = 0.5 or 1.0 (in Case I), c=0.75 or 1.25 (in Case II) and *n* = 20, 50, 100, or 200. Using the Benjamini-Hochberg [34] multiple comparisons approach for each of these condition-combinations, Table 5 and Table 6 display the rankings of these six test procedures based on their empirical power under heteroscedasticity Case I and Case II. For pairwise comparisons that are not significantly different from each other, the procedure with the higher empirical power is listed first but both members of the pair are displayed with underscore or with italics.

From Table 5, there is consistency in the rankings of the six tests over the two levels of *c*. The NWGQ test had the highest empirical power for smaller sized samples (when *n* = 20 or 50) and ranked first, second, or third when *n* = 100 or 200. The BPCW test displayed high empirical power for larger sized samples (when *n* = 100 or 200), ranked second when *n* = 50, and even ranked fourth when *n* = 20 where the procedure was deemed conservative. The GMS test performed well for all eight conditions, ranking second or third. The W test performed poorly with respect to competitors, ranking fourth for larger sized samples and lower when *n* = 20 or 50. The NWRSR test fared better with smaller *n* (ranking third when *n* = 20 and the procedure was deemed liberal) but continued to fall in the rankings as *n* increased, performing even worse (albeit not significantly) than the BF test when *n* = 200. The BF test, deemed conservative when *n* = 20 or 50, displayed weaker empirical power than its competitors, ranking either fifth or sixth for all eight condition combinations. Table 6 gives similar results of the ranking for the six tests under heterogeneity Case II.

In summation, with respect to empirical power the BPCW test is best with larger data sets (i.e., $n \geq 100$) while the NWGQ test is clearly best for smaller sized samples (*n*< 100) and second best with larger data sets. The GMS test is always a good test and superior to the W test, the NWRSR test, and the BF test, three procedures that cannot be recommended based on empirical power considerations.

APPLYINGTHESIXTESTSOFHOMOSCEDASTICITY: AN ILLUSTRATIVE EXAMPLE

The data set is from a classroom example pertaining to a sample of 50 restaurants surveyed in a Zagat Restaurant Guide. The objective is to develop a simple linear regression model that could be used to predict the average cost per meal (in dollars) based on a predictor variable that represents the summated ratings given for food taste, service received, and restaurant ambiance or décor – each measured on scales of 1 (low) to 30 (high).

Although the fitted model is statistically significant (*p*-value = 0.000), both the scatter plot and a set of residual plots obtained from g (Figures 1 and 2) indicate potential violations to homoscedasticity and perhaps linearity. Note, however, that there is no evidence of any gross violation to the assumption of normality and note further that there is no violation to the independence assumption; the listing of the initial data set is organized by type of restaurant food (either Asian or North American / European) which may be considered a lurking dummy variable, significantly correlated with both the dependent variable and the predictor variable.

If a violation in the important assumption of homoscedasticity is found, an appropriate remedy should be employed prior to using a model for purposes of prediction. In introductory-level courses an appropriate transformation of either the dependent or predictor variable should be made and the regression model refitted. In higher-level courses, students can also use weighted least squares or robust estimation methods to improve the model.

Table 7 displays the *p* values for each of the six test procedures used on the meal-cost data and indicates whether (1) or not (0) the tests call for the rejection of the null hypothesis of homoscedasticity at a traditional 0.05 level of significance. Based on the Monte Carlo power study described in Sections 4 and 5, for data set samples of size 50 the three recommended procedures are, respectively, NWGQ, BPCW and GMS. Interestingly, for the meal cost data, the NWGQ test did not "officially detect" [27] heteroscedasticity, but both BPCW and GMS tests had "formal

Table 1: Controlled conditions used in the Monte Carlo power study.		
Condition	**Levels**	**Description**
Distribution	1	Normal
Homoscedasticity	2	$Y = 1 + 2X + ce$, c=0.5, 1.0
Heteroscedasticity: Case I **Case II**	2 2	$Y = 1 + 2X + cXe$, c=0.5, 1.0 $Y = 1 + 2X + (X^{\wedge}c)e$, c=0.75,1.25
Sample Size	4	n = 20, 50, 100, 200
Level of Significance	1	$\alpha = 0.05$

Table 2: Type I error of the test (α = 0.05) with 5,000 simulation runs.								
Conditions								
c=0.5					c=1.0			
Test	n=20	n=50	n=100	n=200	n=20	n=50	n=100	n=200
NWRSR	0.0596	0.0546	0.0546	0.0476	0.0556	0.0522	0.0468	0.0524
NWGQ	0.0508	0.0478	0.0492	0.0508	0.0530	0.0486	0.0430	0.0550
BF	0.0374	0.0424	0.0480	0.0494	0.0378	0.0378	0.0444	0.0500
W	0.0454	0.0430	0.0474	0.0496	0.0408	0.0436	0.0464	0.0518
BPCW	0.0384	0.0436	0.0454	0.0526	0.0386	0.0504	0.0468	0.0512
GMS	0.0586	0.0540	0.0514	0.0534	0.0528	0.0546	0.0478	0.0554

Table 3: Power of the test (α = 0.05) with 5,000 simulation runs (Case I).								
Conditions								
c=0.5					c=1.0			
Test	n=20	n=50	n=100	n=200	n=20	n=50	n=100	n=200
NWRSR	0.1444	0.3126	0.5544	0.8446	0.1342	0.2966	0.5584	0.8438
NWGQ	0.2244	0.4642	0.7348	0.9380	0.2124	0.4432	0.7154	0.9418
BF	0.0900	0.2596	0.5380	0.8476	0.0902	0.2420	0.5306	0.8506
W	0.0980	0.2542	0.5628	0.9110	0.0916	0.2538	0.5586	0.9080
BPCW	0.1444	0.4008	0.7260	0.9604	0.1298	0.3994	0.7234	0.9566
GMS	0.1618	0.3814	0.6824	0.9430	0.1518	0.3710	0.6946	0.9386

Table 4: Power of the test (α = 0.05) with 5,000 simulation runs (Case II).								
Conditions								
c=0.75					c=1.25			
Test	n=20	n=50	n=100	n=200	n=20	n=50	n=100	n=200
NWRSR	0.1002	0.1978	0.3486	0.6070	0.1756	0.4230	0.7364	0.9574
NWGQ	0.1594	0.3130	0.5218	0.7874	0.2698	0.6002	0.8642	0.9906
BF	0.0616	0.1542	0.3392	0.6066	0.1094	0.3632	0.7106	0.9592
W	0.0716	0.1530	0.3482	0.6736	0.1180	0.3582	0.7594	0.9858
BPCW	0.0944	0.2540	0.4772	0.7984	0.1792	0.5540	0.8850	0.9950
GMS	0.1084	0.2494	0.4464	0.7550	0.2090	0.5236	0.8574	0.9898

Table 5: Empirical power rankings (1 = best) for six tests under eight condition-combinations (Case I).						
Condition Combination	NWRSR Test	NWGQ Test	BF Test	W Test	BPCW Test	GMS Test
c = 0.5 and n = 20	3.5	1	6	5	3.5	2
c = 0.5 and n = 50	4	1	5	6	2	3
c = 0.5 and n = 100	5	1	6	4	2	3
c = 0.5 and n = 200	6	3	5	4	1	2
c = 1.0 and n = 20	3	1	6	5	4	2
c = 1.0 and n = 50	4	1	6	5	2	3
c = 1.0 and n = 100	5	2	6	4	1	3
c = 1.0 and n = 200	6	2	5	4	1	3

Table 6: Empirical power rankings (1 = best) for six tests under eight condition-combinations (Case II).

Condition Combination	NWRSR Test	NWGQ Test	BF Test	W Test	BPCW Test	GMS Test
c = 0.75 and n = 20	3	1	6	5	4	2
c = 0.75 and n = 50	4	1	5	6	2	3
c = 0.75 and n = 100	4	1	6	5	2	3
c = 0.75 and n = 200	5	2	6	4	1	3
c = 1.25 and n = 20	4	1	*6*	*5*	3	2
c = 1.25 and n = 50	4	1	5	6	2	3
c = 1.25 and n = 100	*5*	2	6	*4*	1	3
c = 1.25 and n = 200	*6*	2	*5*	4	1	3

Table 7: A comparison of p-values for six tests on the meal cost data (α = 0.05).

Result\Test	NWRSR	NWGQ	BF	W	BPCW	GMS
***P*-Value**	0.0726	0.0825	0.1627	0.1278	0.0498	0.0396
Decision	0	0	0	0	**1**	**1**

Table 8: A comparison of p-values for six tests on the logs of the meal cost data (α = 0.05).

Result\Test	NWRSR	NWGQ	BF	W	BPCW	GMS
***P*-Value**	0.7832	0.5647	0.8520	0.4554	0.7464	0.7683
Decision	0	0	0	0	0	0

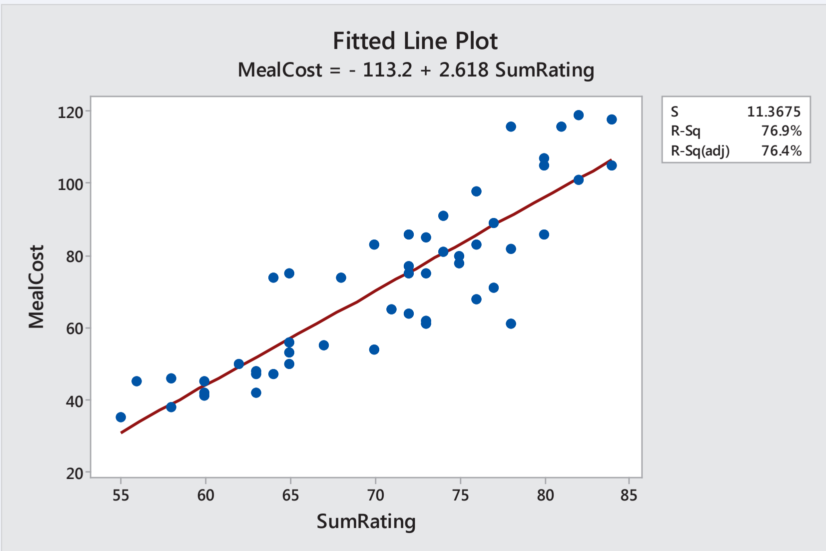

Figure 1 Minitab developed scatter plot of cost of a meal (in $) and summated rating of food, service and décor.

confirmation" of the visual inspection that indeed the assumption of homoscedasticity has been violated and the assumption of linearity (i.e., model specification) may also be questioned.

Given that the regression diagnostics via graphical residual analysis indicates possible issues with the assumptions of linearity and homoscedasticity, the "ladder of powers" suggested by Mosteller and Tukey [35] call for a lower power of Y or higher power of X to "straighten out curvature" and to reduce any observed heteroscedasticity. It was decided to use the natural logarithm of Y as a linear function of X and refit the model.

The coefficient of determination for the refitted model has now increased by 4.7 % over the initially fitted model (i.e., a rise from 76.9% to 80.5%).The refitted model is statistically significant (p-value = 0.000) and visual inspection of both the scatter plot and set of residual plots obtained from Minitab (Figures 3 and 4) indicate that the log transformation on the dependent variable successfully reduced the issues of linearity and homoscedasticity previously observed and the new model appears to fit adequately. As shown in Table 8, any of the formal tests of the homoscedasticity assumption confirm this exploratory, graphical observation.

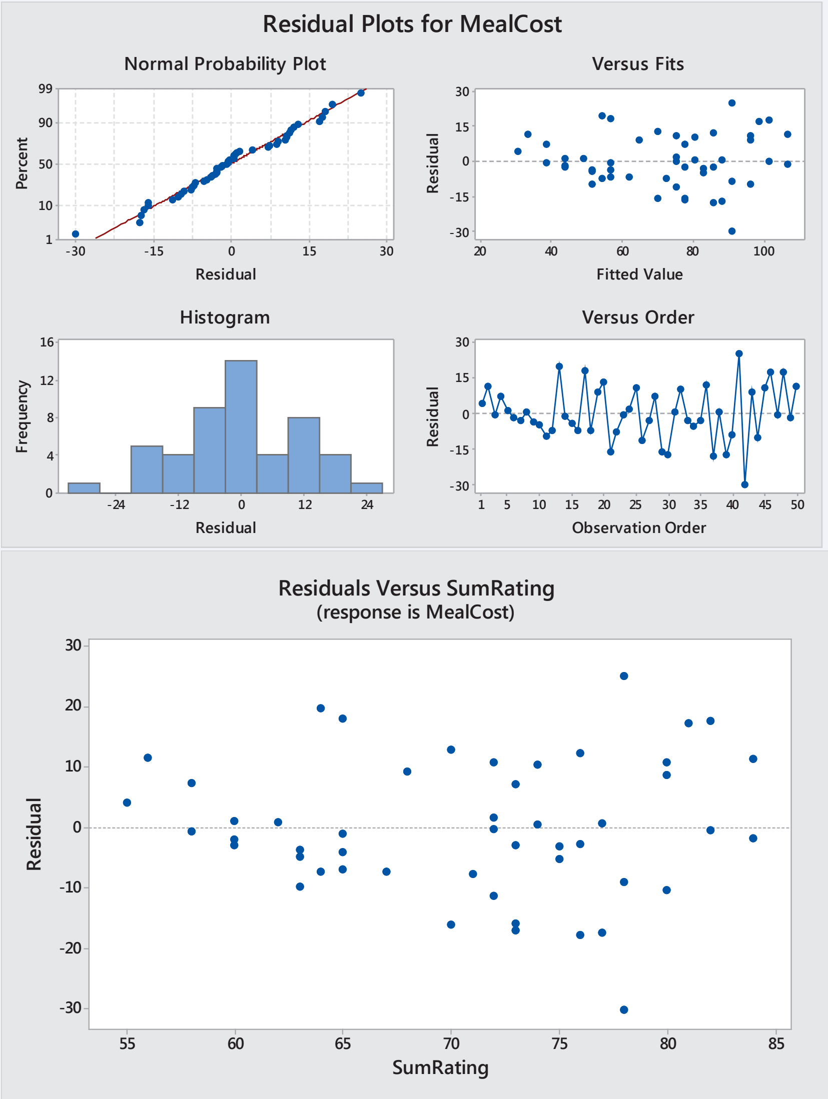

Figure 2 Minitab developed residual plots for a simple linear regression model.

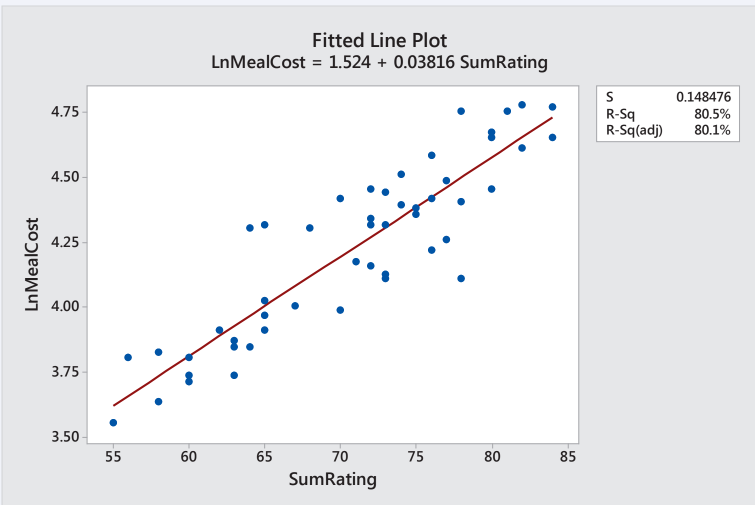

Figure 3 Minitab developed scatter plot of the log meal cost and summated rating of food, service and décor.

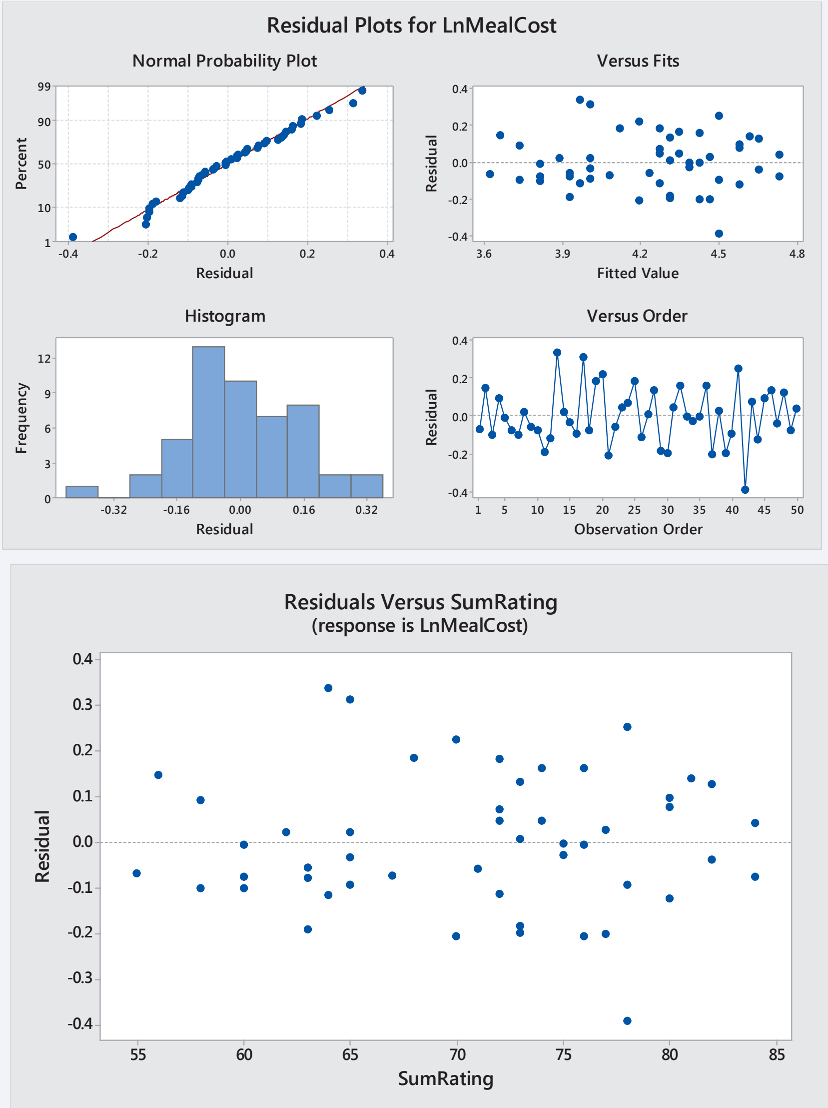

Figure 4 Minitab developed residual plots for a refitted simple linear regression model.

CONCLUSIONS AND RECOMMENDATIONS

When evaluating the worth of a statistical procedure, Tukey [1] defined practical power as the product of statistical power and the utility of the statistical technique (i.e., the ease in which the test can be learned and likelihood it would be used). Given the potential limitations for each of the six tests of homoscedasticity that were described in Section 2, it appears that the GMS test would have most practical power when teaching an introductory-level undergraduate course while either NWGQ test or BPCW test can be recommended for teaching an introductory-level graduate course[36]. Since the former displays more statistical power with smaller sized samples and the latter performers better with larger data sets it is recommended that both these procedures should be considered.

ACKNOWLEDGMENTS

Dr. Berenson's work was supported in part through the Khubani/Telebrands Faculty Research Fellowship in the Feliciano School of Business at Montclair State University. Comparing Tests of Homoscedasticity in Simple Linear Regression

REFERENCES

1. Tukey JW. A quick, compact two-sample test to Duckworth's specifications. Technometrics. 1959; 1: 31-48.

2. Anscombe FJ. Examination of residuals. Proceedings of the Fourth Berkeley. 1961.

3. Cleveland WS, McGill ME, McGill R. The shape parameter of a two-variable graph. J Am Stat Assoc. 1998; 83: 289-300.

4. Casey SA. "Examining Student Conceptions of Covariation: A Focus on the Line of Best Fit". J Stat Edu. 2015; 23.

5. Cook RD, Weisberg S. An Introduction to Regression Graphics. New York: Wiley, 1994.

6. Ali M, Giaccotto C. "A Study of Several New and Existing Tests for Heteroscedasticity in the General Linear Model," J Econom. 1984; 26: 355-373.

7. Goldfeld S, Quandt R. Some tests for homoscedasticity. J Am Stat Assoc. 1965; 60: 539-547.

8. Park RE. Estimation with heteroscedastic error terms. Econometrica. 1966; 34: 888.

9. Glejser H. A new test for heteroskedasticity. J Am Stat Assoc. 1969; 64: 315-323.

10. Ramsey JB. Tests for specification error in the general linear model. J Royal Stat Soc. 1969; 31: 350-371.

11. Brown M, Forsythe A. Robust tests for equality of variances. J Amer Stat Assoc. 1974; 69: 364-367.

12. Bickel P. Tests for heteroscedasticity, nonlinearity. Ann Stat. 1978; 6: 266-291.

13. Harrison MJ, Mc Cabe BP. A test for heteroscedasticity based on ordinary least squares residuals. J Am Stat Assoc.1979; 74: 494-499.

14. Breusch T, Pagan A. A simple test for heteroscedasticity and random coefficient variation. Econometrica. 1979; 47: 1287-1294.

15. White H. A heteroscedasticity-consistent covariance matrix estimator and a direct test for Heteroscedasticity. Econometrica. 1980; 48: 817-838.

16. Godfrey LG, Orme CD. The robustness, reliability and power of heteroskedasticity tests. Econom Rev. 1999; 18: 169-194.

17. Koenker R. A note on studentizing a test for heteroscedasticity. J Econom. 1981; 17: 107-112.

18. Engle R. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica. 1988; 96: 893-920.

19. Berenson ML. Using Excel for White's test – an important technique for evaluating the equality of variance assumption and model specification in a regression analysis. Dec Sci J Innovative Ed. 2013; 11: 243-262.

20. Godfrey LG. Some results on the Glejser and Koenker tests for heteroskedasticity. J Econom. 1996; 72: 275.

21. Im KS. Robustifying Glejser test of heteroskedasticity. J Econom. 2000; 97: 179.

22. Machado JA, Santos JM. Glejser's test revisited. J Econom. 2000; 97: 189-202.

23. Carapeto M, Holt W. "Testing for Heteroscedasticity in Regression Models." J App Stat, 2003; 30: 13-20.

24. Furno M. The Glesjer test and the median regression. Sankhya. 2005; 67: 335-358.

25. Cai L, Hayes AF. "A New Test of Linear Hypotheses in OLS Regression under Heteroscedasticity of Unknown Form." J Edu Behav Stat. 2008; 33: 21-40.

26. Neter J, Wasserman W. Applied Linear Statistical Models. Homewood, IL: Richard D. Irwin. 1974.

27. Wasserstein RL and Lazar NA. The ASA's statement on p-values: context, process, and purpose, Am Statistical Assoc. 2016: 129-133.

28. Cook RD, Weisberg, S. Diagnostics for heteroscedasticity in regression. Biometrika. 1983; 70: 1-10.

29. Klibanoff P, Sandroni A, Moselle B, Saraniti B. Managerial Statistics: A Case-Based Approach. Marion, OH: Thomson/South-Western (CengageLearning). 2006.

30. Mendenhall W, Sincich TA. Second Course in Statistics: Regression Analysis, 6th ed. Upper Saddle River, NJ: Prentice Hall (Pearson HigherEducation). 2003.

31. R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, ISBN 3-900051-07-0. 2008.

32. Gibbons JD. Nonparametric Methods for Quantitative Analysis. New York: Holt, Rinehartand Winston. 1976.

33. Cochran WG. The comparison of percentages in matched samples. Biometrika. 1950; 37: 256- 266.

34. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy. Statist Soc Ser B. 1995; 57: 289-300.

35. Mosteller F, Tukey JW. Data Analysis and Regression: A Second Course in Statistics. Reading, MA: Addison-Wesley. 1977.

36. Anscombe FJ. Graphs in statistical analysis. Am Stat. 1973; 27: 17-21.

Cite this article

Su H, Berenson ML (2017) Comparing Tests of Homoscedasticity in Simple Linear Regression. JSM Math Stat 4(1): 1017.