

Results

Mediation

Introduction

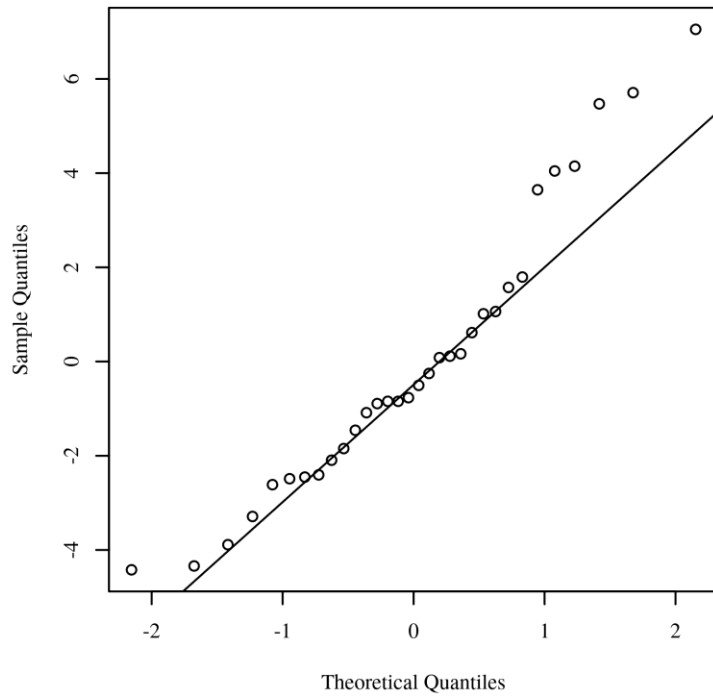
A causal mediation analysis was conducted to assess if cylinders mediated the relationship between displacement and miles_per_gallon.

Assumptions

Normality. The assumption of normality was assessed by plotting the quantiles of the model residuals against the quantiles of a Chi-square distribution, also called a Q-Q scatterplot (DeCarlo, 1997). For the assumption of normality to be met, the quantiles of the residuals must not strongly deviate from the theoretical quantiles. Strong deviations could indicate that the parameter estimates are unreliable. Figure 1 presents a Q-Q scatterplot of model residuals.

Figure 1

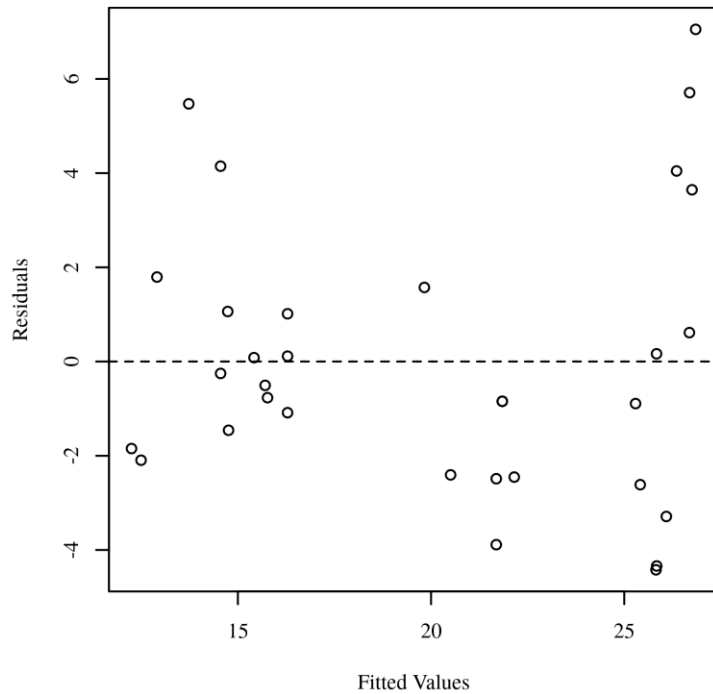
Q-Q scatterplot for normality of the residuals for the regression model.



Homoscedasticity. Homoscedasticity was evaluated by plotting the residuals against the predicted values (Bates et al., 2014; Field, 2017; Osborne & Walters, 2002). The assumption of homoscedasticity is met if the points appear randomly distributed with a mean of zero and no apparent curvature. Figure 2 presents a scatterplot of predicted values and model residuals.

Figure 2

Residuals scatterplot testing homoscedasticity



Multicollinearity. Variance Inflation Factors (VIFs) were calculated to detect the presence of multicollinearity between predictors. High VIFs indicate increased effects of multicollinearity in the model. VIFs greater than 5 are cause for concern, whereas VIFs of 10 should be considered the maximum upper limit (Menard, 2009). All predictors in the regression model have VIFs less than 10. Table 1 presents the VIF for each predictor in the model.

Table 1

Variance Inflation Factors for displacement and cylinders

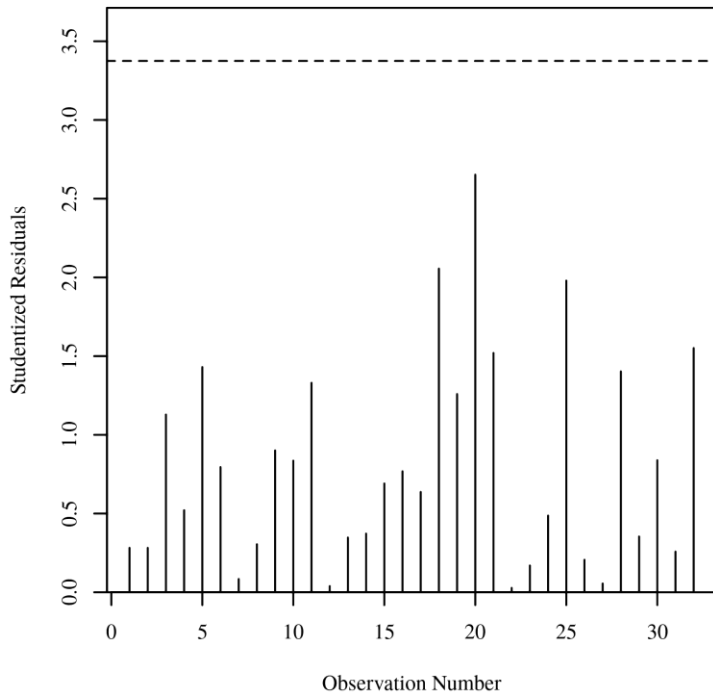
Variable	VIF
displacement	5.37
cylinders	5.37

Outliers. To identify influential points, Studentized residuals were calculated and the absolute values were plotted against the observation numbers (Field, 2017; Pituch & Stevens, 2015). Studentized residuals are calculated by dividing the model residuals by the estimated residual standard deviation. An observation with a Studentized residual greater than 3.37 in absolute value, the 0.999 quantile of a t distribution with 31 degrees of freedom, was considered

to have significant influence on the results of the model. Figure 3 presents the Studentized residuals plot of the observations. Observation numbers are specified next to each point with a Studentized residual greater than 3.37.

Figure 3

Studentized residuals plot for outlier detection



Results

Mediation was examined based on the indirect and direct effects using bootstrapping (N = 1,000) with percentile-based confidence intervals. The results are based on an alpha of .05. The regression model results are presented in Table 2 and Table 3. The mediation model diagram can be seen in Figure 4.

Table 2

Results for the Regression on miles_per_gallon

Variable	<i>B</i>	<i>SE</i>	95.00% CI	<i>t</i>	<i>p</i>
----------	----------	-----------	-----------	----------	----------

(Intercept)	34.66	2.55	[29.45, 39.87]	13.61	< .001
displacement	-0.02	0.01	[-0.04, 0.0004]	-2.01	.054
cylinders	-1.59	0.71	[-3.04, -0.13]	-2.23	.034

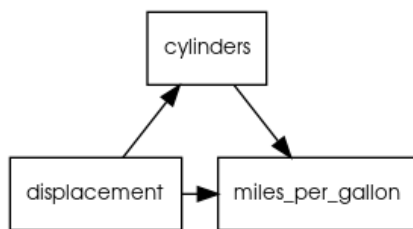
Table 3

Results for the Regression on cylinders

Variable	<i>B</i>	<i>SE</i>	95.00% CI	<i>t</i>	<i>p</i>
(Intercept)	3.19	0.30	[2.58, 3.79]	10.76	< .001
displacement	0.01	0.001	[0.01, 0.02]	11.45	< .001

Figure 4

Node diagram for the mediation analysis



Direct Effect. The average direct effect was not significant, $B = -0.02$, 95.00% CI [-0.04, 0.0004], $p = .054$. This indicates that displacement did not significantly predict miles_per_gallon.

Indirect Effect. The average indirect effect for displacement on miles_per_gallon through cylinders was significant, $B = -0.02$, 95.00% CI [-0.04, -0.009].

References

- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4: arXiv preprint arXiv, *Journal of Statistical Software*.
<https://doi.org/10.18637/jss.v067.i01>
- DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods*, 2(3), 292-307. <https://doi.org/10.1037/1082-989X.2.3.292>
- Field, A. (2017). *Discovering statistics using IBM SPSS statistics: North American edition*. Sage Publications
- Intellectus Statistics [Online computer software]. (2023). Intellectus Statistics.
<https://statistics.intellectus360.com>
- Menard, S. (2009). *Logistic regression: From introductory to advanced concepts and applications*. Sage Publications. <https://doi.org/10.4135/9781483348964>
- Osborne, J., & Waters, E. (2002). Four assumptions of multiple regression that researchers should always test. *Practical Assessment, Research & Evaluation*, 8(2), 1-9.
- Pituch, K. A., & Stevens, J. P. (2015). *Applied multivariate statistics for the social sciences* (6th ed.). Routledge Academic. <https://doi.org/10.4324/9781315814919>

Glossaries

Mediation

Mediation is conducted in order to establish if a mediator variable explains the relationship between two continuous variables (independent and dependent variable). In other words, the mediator variable is the intervening or the process variable. The Hayes' process analysis is used, which examines mediation through a series of regressions. Moderation can also be included in the model, which are examined through conditional effects or simple slopes.

Confidence Interval (CI): An interval that is expected to contain the true value of a statistic in $n\%$ of repeated samples from the same probability distribution. n is based on the confidence level of the confidence interval.

D : The test statistic used for determining significance in the Kolmogorov-Smirnov test.

Degrees of Freedom (df): Used with the F ratio to determine the p -value.

Direct Effect: The effect that an independent variable has on the dependent variable without including any indirect effects.

Dummy-Code: Performed in order to add a nominal or ordinal independent variable into the regression model; turns the one variable into a series of dichotomous "yes/no" variables, one for each category; one of the categories are left out of the regression as the reference group that all other categories are compared to.

F Ratio (F): Used with the two df values to determine the p -value of the overall model.

Homoscedasticity: Refers to the relationship between the residuals and the fitted values; the assumption is met when the residuals plot has the points randomly distributed (with no pattern), and the distribution line is approximately straight.

Indirect Effect: The effect that an independent variable has on the dependent variable through a mediating variable. The indirect shows the degree to which mediation occurs.

Multicollinearity: A state of very high intercorrelations or inter-associations among a set of variables.

Normality: Refers to the distribution of the residuals; the assumption is that the residuals follow a bell-shaped curve; the assumption is met when the q-q plot has the points distributed approximately on the normality line.

Outlier: A data point that is abnormally distant from a set of observations.

p -value: The probability that the null hypothesis (no relationship in the dependent variable by the independent variable) is true.

Residuals: Refers to the difference between the predicted value for the dependent variable and the actual value of the dependent variable.

R-Squared Statistic (R^2): Tells how much variance in the dependent variable is explained by only the predictor variables.

Standard Error (SE): How much we expect B to vary.

Studentized Residuals: Residuals that are scaled by dividing each residual by the estimated standard deviation of the residuals.

***t*-Test Statistic (*t*):** Used with the *df* to determine the *p*-value; also can show the direction of the relationship between the predictor and dependent variable.

Unstandardized Beta (*B*): The slope of the predictor with the dependent variable.

W The test statistic used for determining significance in the Shapiro-Wilk test.

Variance Inflation Factors: A measurement to assess the amount of multicollinearity present in regression analysis.

Raw Output

Mediation Regression Analysis

Included Variables:

miles_per_gallon, cylinders, and displacement

Sample Size (Complete Cases):

N = 32

Variance Inflation Factors:

Variable	VIF
displacement	5.367
cylinders	5.367

Results for the Regression on miles_per_gallon:

Variable	B	SE	95.00% CI	t	p
(Intercept)	34.661	2.547	[29.452, 39.870]	13.609	4.023×10^{-14}
displacement	-0.0206	0.0103	[-0.0416, 3.953×10^{-04}]	-2.007	0.0542
cylinders	-1.587	0.712	[-3.043, -0.131]	-2.230	0.0337

Results for the Regression on cylinders:

Variable	B	SE	95.00% CI	t	p
(Intercept)	3.189	0.296	[2.583, 3.794]	10.758	8.122×10^{-12}
displacement	0.0130	0.00114	[0.0107, 0.0153]	11.445	1.803×10^{-12}

Direct Effect for displacement on miles_per_gallon:

Effect = -0.0206, SE = 0.0103, 95.000% CI [-0.0416, 3.953×10^{-04}], p = 0.0542

Indirect Effect for displacement on miles_per_gallon through cylinders:

Effect = -0.0206, SE = 0.00733, 95.000% CI [-0.0379, -0.00944]

Bootstrap Results for Regression on miles_per_gallon:

Variable	B	M	SE	95.00% CI
(Intercept)	34.661	34.687	2.517	[30.159, 40.071]
displacement	-0.0206	-0.0200	0.00838	[-0.0338, -0.00195]
cylinders	-1.587	-1.615	0.531	[-2.738, -0.733]

Bootstrap Results for Regression on cylinders:

Variable	B	M	SE	95.00% CI
(Intercept)	3.189	3.160	0.282	[2.621, 3.751]

displacement	0.0130	0.0132	0.00128	[0.0110, 0.0160]
--------------	--------	--------	---------	------------------