Results

Linear Regression Analysis

Introduction

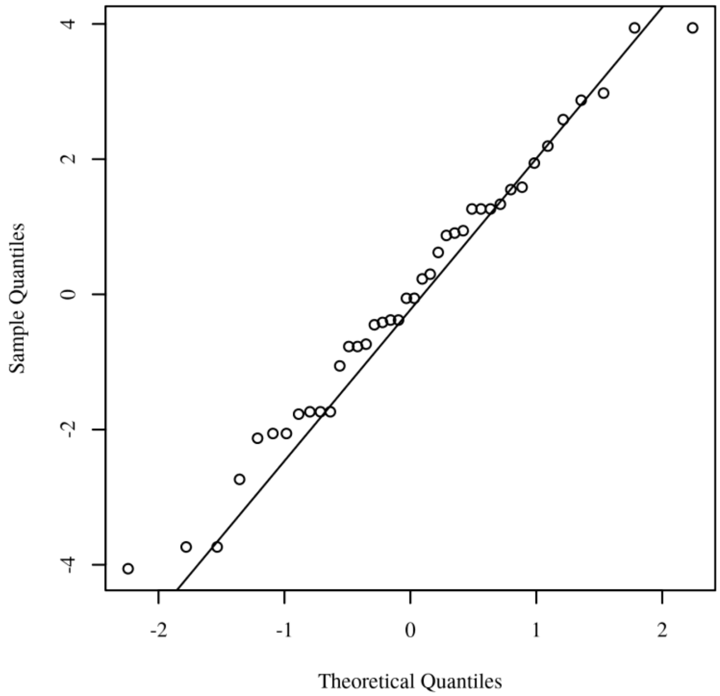
A linear regression analysis was conducted to assess whether DEVELOP significantly predicted PERSON.

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 the model residuals.

Figure 1

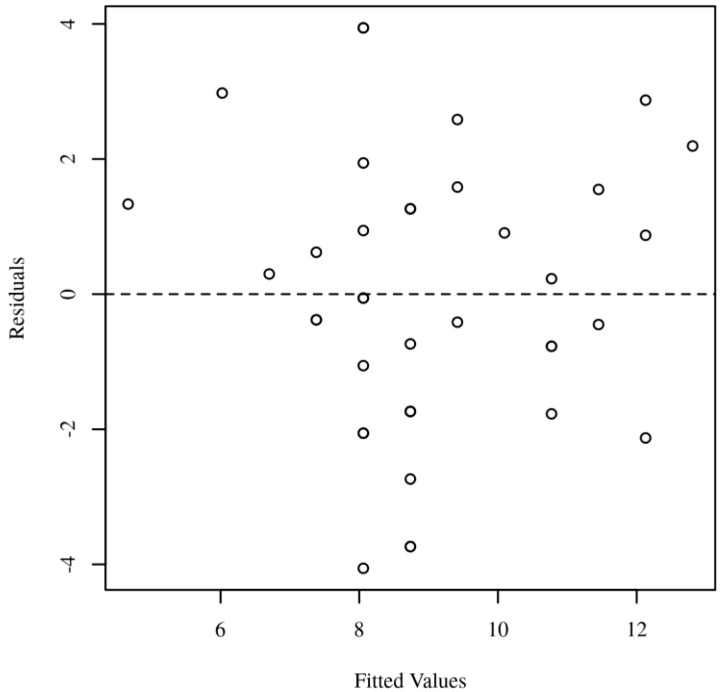
Q-Q scatterplot for normality of the residuals for the regression model.



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

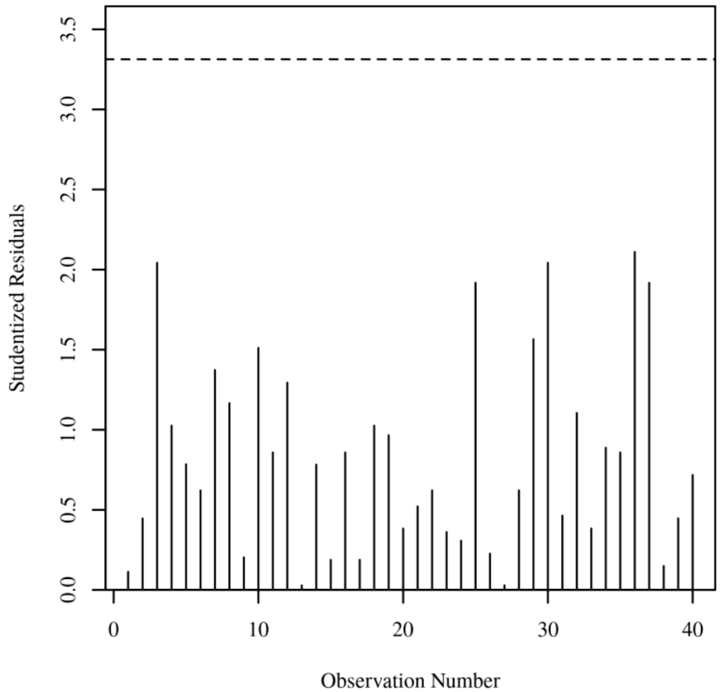


Multicollinearity. Since there was only one predictor variable, multicollinearity does not apply, and Variance Inflation Factors were not calculated.

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 3.31 in absolute value, the 0.999 quantile of a *t* distribution with 39 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 3.31.

Figure 3

Studentized residuals plot for outlier detection



Results

The results of the linear regression model were significant, $F(1,38) = 28.88$, $p < .001$, $R^2 = .43$, indicating that approximately 43.18% of the variance in PERSON is explainable by DEVELOP. DEVELOP significantly predicted PERSON, $B = 0.68$, $t(38) = 5.37$, $p < .001$. This indicates that on average, a one-unit increase of DEVELOP will increase the value of PERSON by 0.68 units. Table 1 summarizes the results of the regression model.

Table 1

Results for Linear Regression with DEVELOP predicting PERSON

Variable	*B*	*SE*	95.00% CI	β	*t*	*p*
(Intercept)	2.63	1.23	[0.14, 5.13]	0.00	2.13	.039
DEVELOP	0.68	0.13	[0.42, 0.93]	0.66	5.37	< .001

Note. Results: $F(1,38) = 28.88$, $p < .001$, $R^2 = .43$
Unstandardized Regression Equation: PERSON = 2.63 + 0.68*DEVELOP

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.io1

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

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

Multiple Linear Regression

The multiple linear regression is the most common form of linear regression analysis. As a predictive analysis, the multiple linear regression is used to explain the relationship between one continuous dependent variable from two or more independent variables. It does this by creating a linear combination of all the independent variables to predict the dependent variable. The independent variables can be continuous or categorical (dummy coded as appropriate). The R^2 statistic is used to assess how well the regression predicted the dependent variable. While the unstandardized beta (*B*) describes the increase or decrease of the independent variable(s) with the dependent variable.

Fun Fact! *Linear regression and multiple linear regression are commonly used in economics and finance as a way to predict risk of investment, consumption spending, inventory investment, labor demand, and more!*

95% Confidence Interval (95% CI): An interval that estimates the range one would expect *B* to lie in 95% of the time given the samples tested comes from the same distribution.

Akaike's Information Criterion (*AIC*): A measure of model quality or fit. It uses the maximized log likelihood value as a baseline for model fit, and adds a penalty for estimating additional parameters. Smaller *AIC* values represent better model fit.

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

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.

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.

Standardized Beta (β): Ranges from -1 to 1; gives the strength of the relationship between the predictor and dependent variable.

Studentized Residuals: Residuals that are scaled by diving the 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.

Standard Error (*SE*): How much the *B* is expected to vary.

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

Raw Output

Linear Regression with PERSON predicted by DEVELOP

Included Variables:
PERSON and DEVELOP

Sample Size (Complete Cases):
N = 40

Linear Regression Coefficients:

Variable	B	SE	95.00% CI	β	t	p
(Intercept)	2.632	1.233	[0.136, 5.127]	0.000	2.135	0.0393
DEVELOP	0.678	0.126	[0.423, 0.934]	0.657	5.374	4.111×10^{-06}

Model Fit Statistics:
$F(1,38) = 28.876$, $p = 4.111 \times 10^{-06}$, $R^2 = 0.432$, adj. $R^2 = 0.417$

Partial Correlations:

Variable	Estimate	t	p
DEVELOP	0.657	5.374	4.111×10^{-06}

Semi-Partial Correlations:

Variable	Estimate	t	p
DEVELOP	0.657	5.374	4.111×10^{-06}