Intellectus Statistics™ Software

Smarter Tools for Enhanced Learning

Intellectus Statistics software combines a comprehensive breadth of statistical techniques, easy user interface, numerous learner-centered features, and the intelligence of teachers to create a statistical tool that provides traditional raw output AND an interpreted quantitative results document with APA formatted tables and figures in seconds.

Intellectus Statistics Team

Included Analyses: 1

- Linear Regression with Science_post predicted by Math_pre, Science_pre, and Reading_pre

Glossary of Terms and Symbols

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.

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

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

Homoscedasticity: refers to the relationship between the residuals and the independent variables; the assumption is met when we have no relationship; the assumption is met when the residuals plot has the points randomly distributed (with no pattern), and the distribution line is approximately straight

Dummy-code: done 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 (F ratio): used with the two df values to determine the p value of the overall model

df (degrees of freedom): used with the F to determine the p value

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

R^2 (R-squared statistic): tells how much variance in the dependent variable is explained by only the predictor variables

B (unstandardized beta): the slope of the predictor with the dependent variable

SE (standard error): how much the B is expected to vary

β (standardized beta): ranges from -1 to 1; gives the strength of the relationship between the predictor and dependent variable t (t-test statistic): used with the df to determine the p value; also can show the direction of the relationship between the predictor and dependent variable.

Descriptive Statistics

```
    Math_pre   Science_pre  Reading_pre  Science_post
 Min.   :55   Min.   :40   Min.   :40   Min.   :41  
 1st Qu.:74   1st Qu.:77   1st Qu.:65   1st Qu.:78  
 Median :78   Median :77   Median :65   Median :80  
 Mean   :76   Mean   :73   Mean   :67   Mean   :76  
 3rd Qu.:81   3rd Qu.:79   3rd Qu.:72   3rd Qu.:85  
 Max.   :90   Max.   :95   Max.   :88   Max.   :97  
```

Linear Regression Output

```
Call:
lm(formula = Science_post ~ Math_pre + Science_pre + Reading_pre, data = 
data)

Residuals:
    Min      1Q  Median      3Q     Max 
-6.3752 -2.0685 -0.8814  0.7108 12.7108 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.16652    5.41435   0.215   0.8306    
Math_pre     0.15508    0.09289   1.670   0.1037    
Science_pre  1.05229    0.05658  18.600   <2e-16 ***
Reading_pre -0.20716    0.10102  -2.051   0.0476 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.847 on 36 degrees of freedom
Multiple R-squared:  0.9311,	Adjusted R-squared:  0.9253 
F-statistic: 162.1 on 3 and 36 DF,  p-value: < 2.2e-16
```

Results

Intellectus Statistics version 1.01 was used for data analysis and narrative interpretation.

Multiple Linear Regression Analysis

A multiple linear regression analysis was conducted to assess whether a significant relationship existed between Math_pre, Science_pre, and Reading_pre and Science_post. The 'Enter' variable selection method was chosen for the linear regression model. This method forces all selected variables into the model. 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. 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. The assumption of homoscedasticity was assessed by plotting the model residuals against the predicted model values. The assumption is met if the points appear randomly distributed with a mean of zero and no apparent curvature. Figure 1 presents a Q-Q scatterplot of the model residuals. Figure 2 presents a scatterplot of predicted values and model residuals. The assumption is met if the points are not unevenly distributed and no curvature is apparent. Variance Inflation Factors (VIFs) were calculated to detect the presence of multicollinearity between predictors. High VIFs indicate increased effects of multicollinearity in the model. Variance Inflation Factors greater than 6 are cause for concern, whereas a VIFs of 10 should be considered the maximum upper limit. All predictors in the regression model have variance inflation factors (VIF) less than 10. Table 1 presents the VIF for each predictor in the model.

The results of the linear regression model were significant, $F(3,36) = 162.05$, $p < .001$, $R^2 = 0.93$, indicating that approximately 93% of the variance in Science_post is explainable by Math_pre, Science_pre, and Reading_pre. Math_pre was not a significant predictor of Science_post, $B = 0.16$, $t(36) = 1.67$, $p = 0.104$. Based on this sample, a one unit increase of Math_pre did not have a significant effect on Science_post. Science_pre significantly predicted Science_post, $B = 1.05$, $t(36) = 18.60$, $p < .001$. This indicates that on average, every one unit increase of Science_pre will result in a 1.05 unit change in Science_post. Reading_pre significantly predicted Science_post, $B = -0.21$, $t(36) = -2.05$, $p = 0.048$. This indicates that on average, every one unit increase of Reading_pre will result in a -0.21 unit change in Science_post. Table 2 summarizes the results of the regression model.

Table 1

Variance Inflation Factors for Math_pre, Science_pre, and Reading_pre

Variable	*VIF*
Math_pre	1.90
Science_pre	1.52
Reading_pre	2.42

Table 2

Results for Multiple Linear Regression with Math_pre, Science_pre, and Reading_pre predicting Science_post

Variable	*B*	*SE*	β	*t*	*p*
(Intercept)	1.17	5.41	0.00	0.22	.831
Math_pre	0.16	0.09	0.10	1.67	.104
Science_pre	1.05	0.06	1.00	18.60	< .001
Reading_pre	-0.21	0.10	-0.14	-2.05	.048

Note. $F(3,36) = 162.05$, $p < .001$, $R^2 = 0.93$

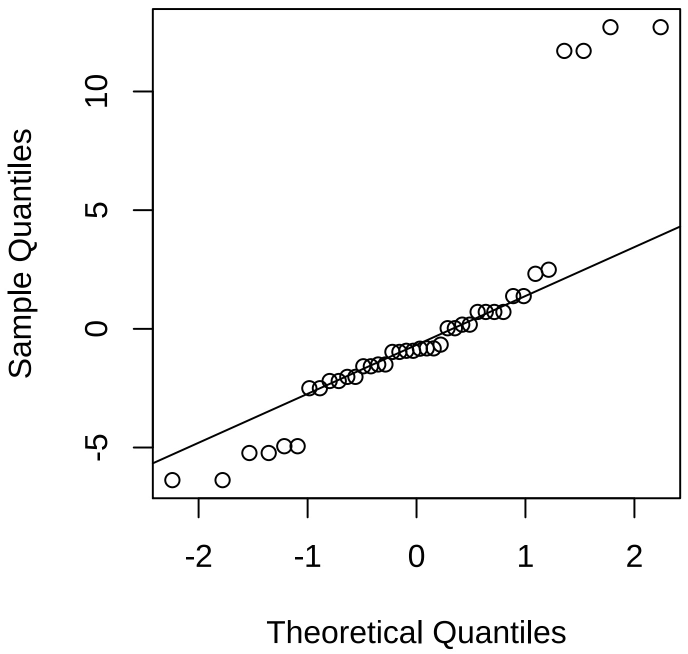

Figure 1. Q-Q scatterplot for normality for Math_pre, Science_pre, and Reading_pre predicting Science_post

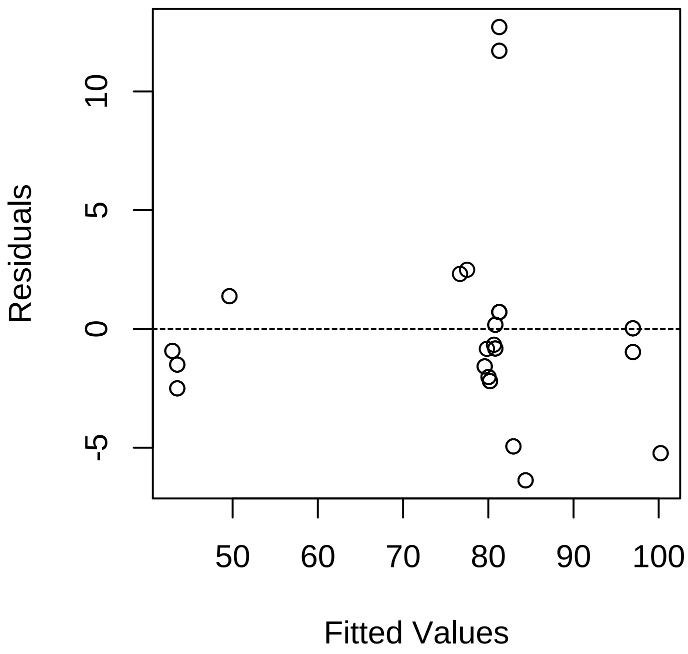

Figure 2. Residuals scatterplot for homoscedasticity for Math_pre, Science_pre, and Reading_pre predicting Science_post

References

Statistics Solutions. (2016). Intellectus Statistics (Version 1.01) [Online computer software]. Retrieved from http://ssp.statisticssolutions.com/