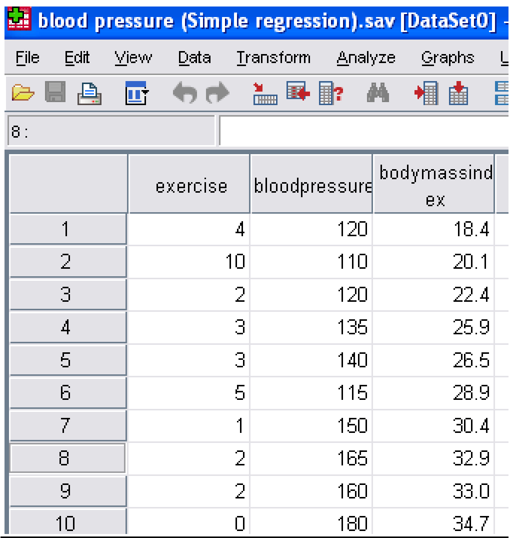
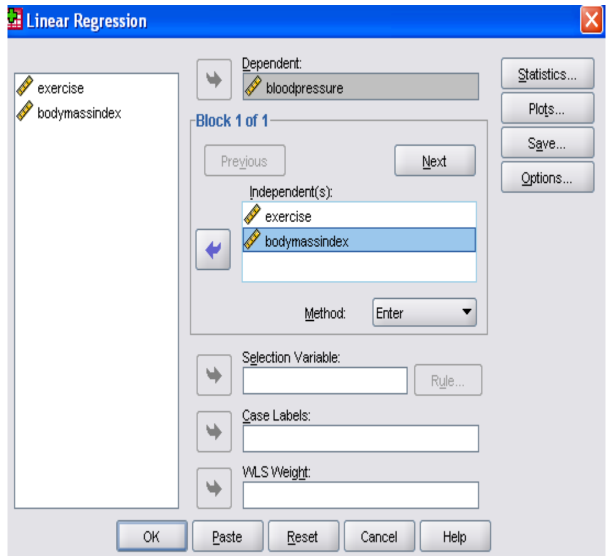
Suppose we are interested in how the exercise and body mass index affect on the blood pressure. A random sample of 10 males 50 years of age is selected and their height, weight, number of hours of exercise and the blood pressure are measured. Body mass index is calculated by the following fomula.

$$\text{BMI} = \frac{\text{(weight in pounds * 703 )}}{\text{height in inches}^2} \quad (\text{kg/m}^2)$$

The dataset is available at U:_MT Student File Area\hjkim\STAT380\SPSS tutorial\blood pressure.sav.




| | exercise | bloodpressure | bodymassindex |
|---|---|---|---|
| 1 | 4 | 120 | 18.4 |
| 2 | 10 | 110 | 20.1 |
| 3 | 2 | 120 | 22.4 |
| 4 | 3 | 135 | 25.9 |
| 5 | 3 | 140 | 26.5 |
| 6 | 5 | 115 | 28.9 |
| 7 | 1 | 150 | 30.4 |
| 8 | 2 | 165 | 32.9 |
| 9 | 2 | 160 | 33.0 |
| 10 | 0 | 180 | 34.7 |




Select **Analyze-Regression-Linear** from the pull-down menu.
Placing the variable we would like to predict, blood pressure, in the dependent variable and the variable we will use for prediction, exercise and body mass index in the independent variable, we hit OK. This generates the following SPSS output.

**Model Summary**[b]

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | .897[a] | .804 | .748 | 11.910 |

a. Predictors: (Constant), bodymassindex, exercise

b. Dependent Variable: bloodpressure

**ANOVA**[b]

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 4079.550 | 2 | 2039.775 | 14.380 | .003[a] |
| | Residual | 992.950 | 7 | 141.850 | | |
| | Total | 5072.500 | 9 | | | |

a. Predictors: (Constant), bodymassindex, exercise

b. Dependent Variable: bloodpressure

**Coefficients[a]**

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | 74.491 | 29.411 | | 2.533 | .039 |
| | exercise | -2.835 | 1.862 | -.332 | -1.523 | .172 |
| | bodymassindex | 2.712 | .914 | .647 | 2.965 | .021 |

a. Dependent Variable: bloodpressure

The interpretation of $R^2$ is same as before. We can see that 80.4% of the variation in Y is explained by the regression line. The fitted regression model found from the output is

$$\hat{Y} = 74.491 - 2.835X_1 + 2.712X_2$$

The next part of the output is the statistical analysis (ANOVA-analysis of variance) for the regression model. The ANOVA represents a hypothesis test with where the null hypothesis is

$H_o : \beta_i = 0$ for all i (In simple regression, i = 1)

$H_A : \beta_i \neq 0$ for at least 1 coefficient

In this example, p-value for this overall test is .003 concluding at least one of independent variables is significantly meaningful to explain the blood pressure.

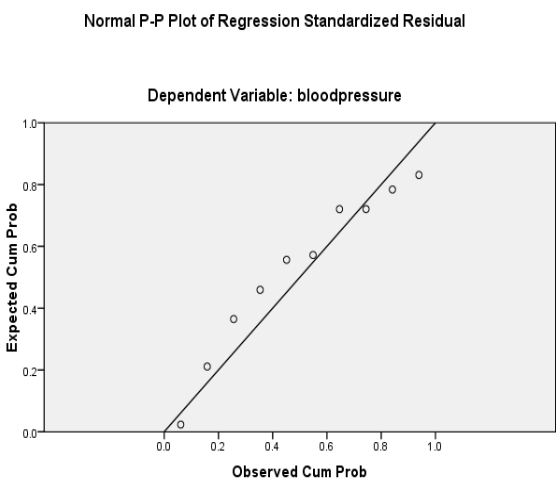
The individual t-test can also be performed.

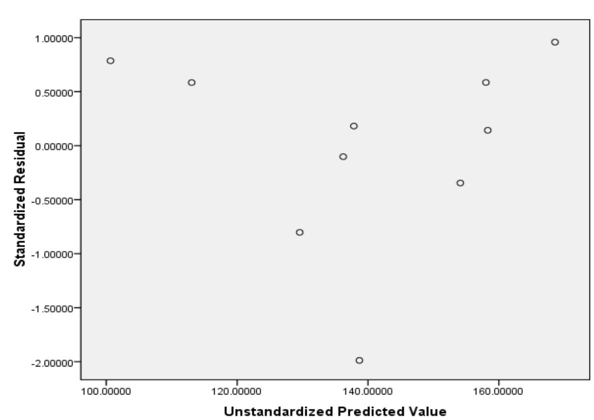
$H_o : \beta_1 = 0 \quad H_A : \beta_1 \neq 0$

$H_o : \beta_2 = 0 \quad H_A : \beta_2 \neq 0$

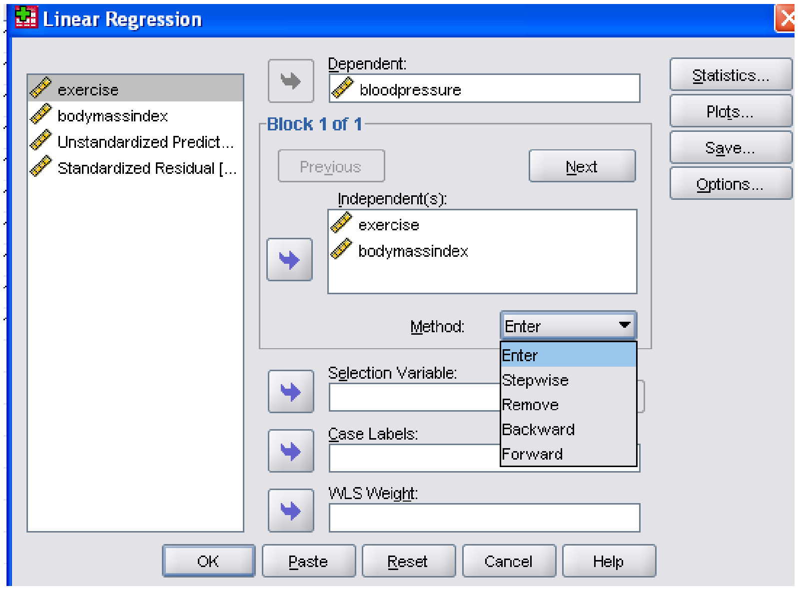
In this example p-value is .172 and .021. Thus, $\beta_1$ is not significantly different from zero when body mass index is in the model, and $\beta_2$ is significantly different from zero when body mass index is in the model.

Model assumption checking and prediction interval can be done in the similar manner as the simple regression analysis. Normal probability plot and residual plot can be obtained as follows.



Model selection (Variable selection)




For variable selection procedure we can choose stepwise, remove, backward, and forward. This procedure automatically selects variables that are significantly important in the model using different procedure. The detail procedure can be found in class handout.
Even though it is not provided in SPSS, model selection criteria such as AIC AICc, BIC, and Cp are also very commonly used to choose the best candidate model.