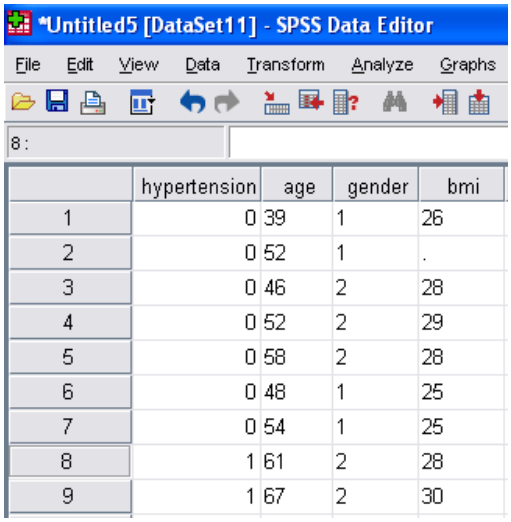
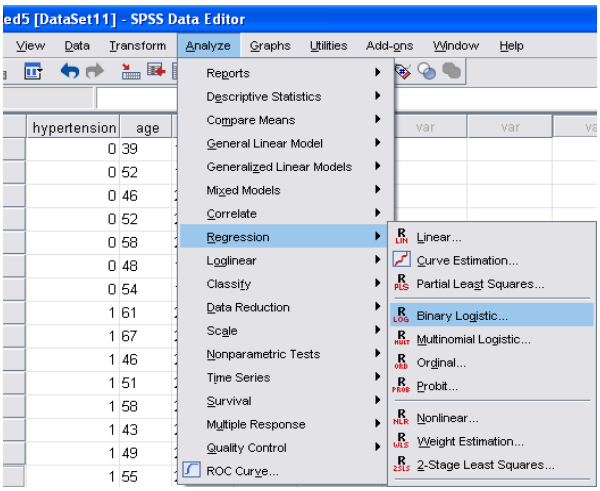
Suppose we are interested in investigating predictors of incident hypertension. The candidate predictor variables are age, gender, and body mass index. The dataset is available at U:_MT Student File Area\hjkim\STAT380\SPSS tutorial\hypertension.sav.

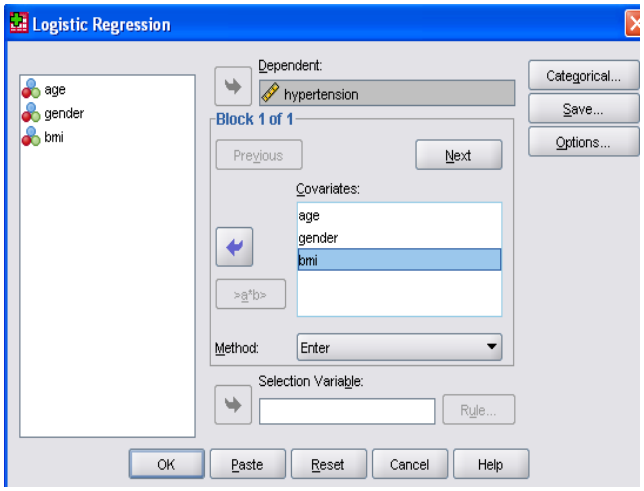


	hypertension	age	gender	bmi
1	0	39	1	26
2	0	52	1	.
3	0	46	2	28
4	0	52	2	29
5	0	58	2	28
6	0	48	1	25
7	0	54	1	25
8	1	61	2	28
9	1	67	2	30

Note that the hypertension variable binary variable. 0 means no hypertension and 1 means hypertension. Predictor variables are age, gender and body mass index. Age and bmi is quantitative and gender is categorical variable.

To perform a logistic regression analysis, select **Analyze-Regression-Binary Logistic** from the pull-down menu. Then place the hypertension in the dependent variable and age, gender, and bmi in the independent variable, we hit OK.





This generates the following SPSS output.

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	1072.808	3	.000
	Block	1072.808	3	.000
	Model	1072.808	3	.000

Overall Chi-square test

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

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

is rejected since p-value = .000.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1	age	.049	.002	398.729	1	.000	1.050
	gender	.218	.046	22.825	1	.000	1.244
	bmi	.150	.007	511.935	1	.000	1.161
	Constant	-5.602	.230	592.966	1	.000	.004

The Variables in the Equation table contains the coefficients for the (fitted) line and other relative information about the coefficients. The equation of the line found from the output is

$$\ln\left(\frac{\hat{p}(x)}{1-\hat{p}(x)}\right) = -5.602 + 0.049x_1 + 0.218x_2 + 0.150x_3$$

A review of the table also indicates several other statistical tests that SPSS is performing. You'll note that SPSS test both of the coefficients to see if they are equal to zero with Wald chi square tests. We can see that all of the coefficients are significantly different from zero. (p-values are 0.000)

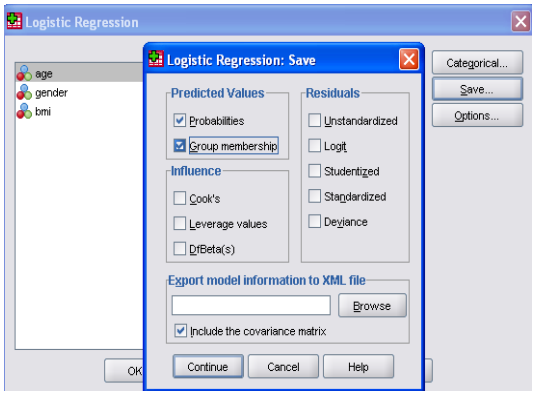
$$H_o : \beta_1 = 0 \quad H_A : \beta_1 \neq 0$$
$$H_o : \beta_o = 0 \quad H_A : \beta_o \neq 0$$

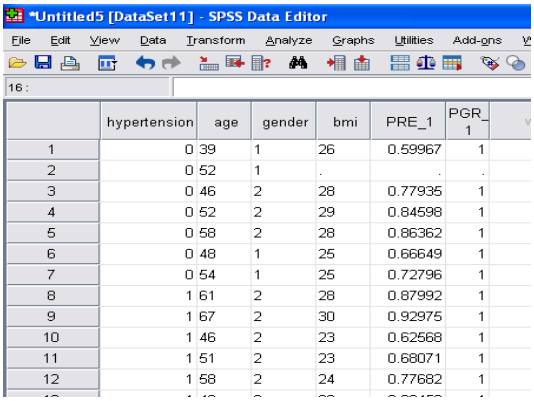
Odds ratio,

$\widehat{OR} = \exp(\widehat{\beta_1})$ = Exp(B), the last column of the Variables in the Equation table.

<u>Creating probability estimate and the group</u>

Conduct the logistic regression as before by selecting **Analyze-Regression-Binary Logistic** from the pull-down menu. In the window select the save button on the right hand side. This will bring up the **Logistic Regression: Save** window. Check the box for Probabilities and Group membership hit continue. This will create a new output in the data screen.





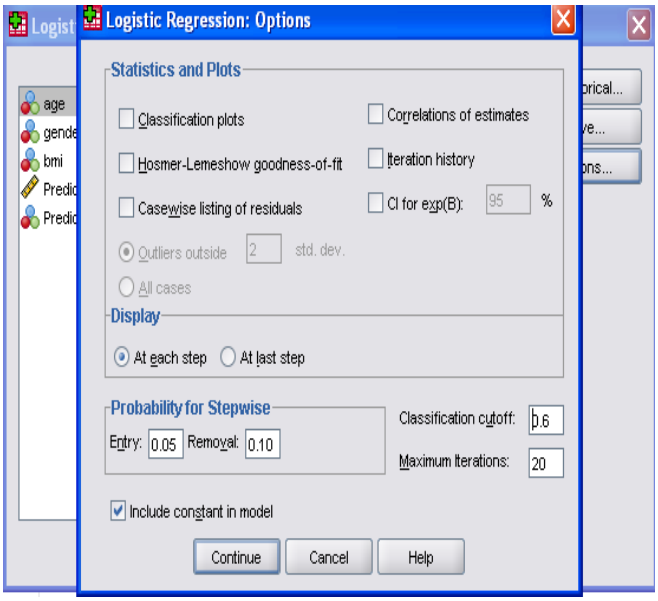
<u>Model Validation</u>

Classification table from output result summarizes the observed group and the predicted group classification. For example, the overall correctly specified group percentage is 74.6%.
Here, the cutoff point is 0.5 by default. This can be changed by going options under logistic regression window and change classification cutoff.

Classification Table[a]

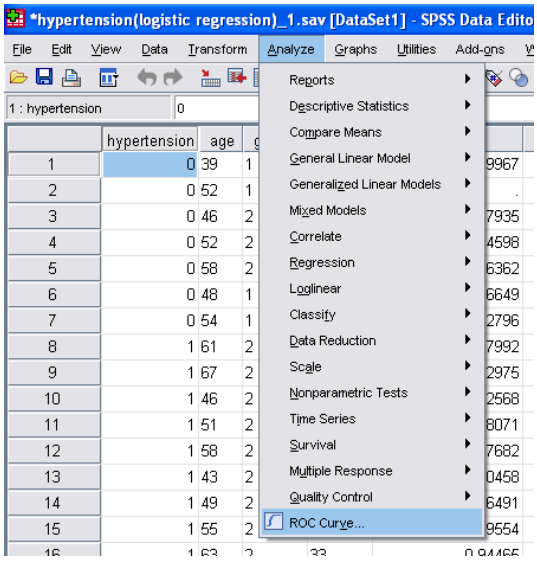
Observed			Predicted		
			hypertension		
			No	Yes	Percentage Correct
Step 1	hypertension	No	293	2682	9.8
		Yes	261	8339	97.0
		Overall Percentage			74.6

a. The cut value is .500

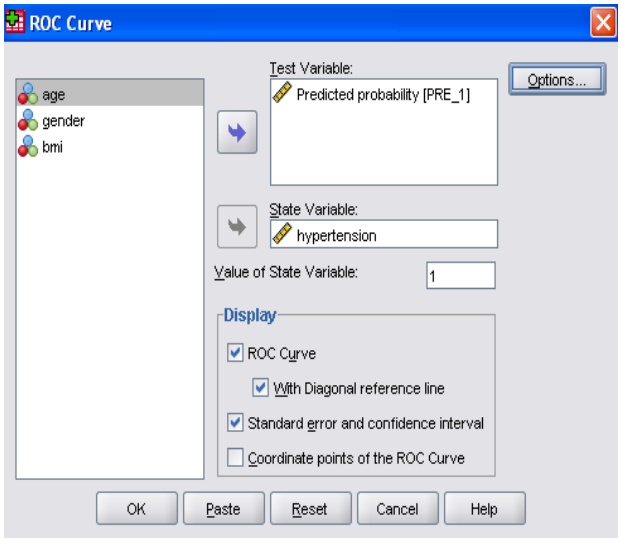


ROC curve

A measure of goodness-of-fit often used to evaluate the fit of a logistic regression model is based on the simultaneous measure of sensitivity (True positive) and specificity (True negative) for all possible cutoff points. First, we calculate sensitivity and specificity pairs for each possible cutoff point and plot sensitivity on the y axis by (1-specificity) on the x axis. This curve is called the receiver operating characteristic (ROC) curve. The area under the ROC curve ranges from 0.5 and 1.0 with larger values indicative of better fit.



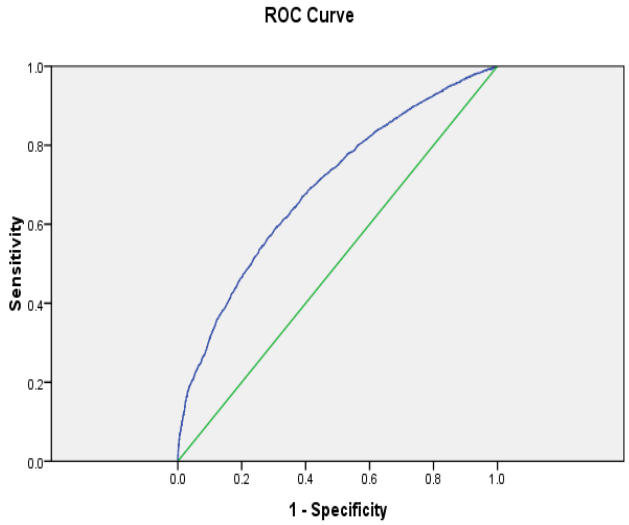
To obtain ROC curve, first the predicted probabilities should be saved. Conduct the logistic regression as before by selecting **Analyze-Regression-Binary Logistic** from the pull-down menu. In the window select the save button on the right hand side. This will bring up the **Logistic Regression: Save** window. Check the box for Probabilities. This will save the predicted probabilities. Then select **Analyze-ROC curve.**



Test variables are often composed of probabilities from logistic regression. The state variable can be the true category to which a subject belongs. The value of the state variable indicates which category should be considered positive.

Move Predicted probability to Test Variable, hypertension to State Variable, define Value of State Variable as 1. Click Display options. With Diagonal reference line will give the ROC curve with the diagonal line. Standard error and confidence interval option will provide the area under the ROC curve with inference statistics about the curve.

SPSS output shows ROC curve. The area under the curve is .694 with 95% confidence interval (.683, 704). Also, the area under the curve is significantly different from 0.5 since p-value is .000 meaning that the logistic regression classifies the group significantly better than by chance.



Diagonal segments are produced by ties.

Area Under the Curve

Test Result Variable(s):Predicted probability

Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.694	.005	.000	.683	.704

The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5