

1 Model Diagnostics for Logistic Regression

2 Model Diagnostics for Logistic Regression

2.1 Omnibus Test for Model Coefficients

The overall significance is tested using what SPSS calls the ***Model Chi-square***, which is derived from the likelihood of observing the actual data under the assumption that the model that has been fitted is accurate. There are two hypotheses to test in relation to the overall fit of the model:

H_0 The model is a good fitting model.

H_1 The model is not a good fitting model (i.e. the predictors have a significant effect).

- For the example provided here, the model chi square has 2 degrees of freedom, a value of 24.096 and a probability of $p < 0.000$. (*see Next Page for SPSS Output*).
- Thus, the indication is that the model has a poor fit, with the model containing only the constant indicating that the predictors do have a significant effect and create essentially a different model.
- So we need to look closely at the predictors and from later tables determine if one or both are significant predictors.

Remark upon “Steps”

- This table has 1 step. This is because we are entering both variables and at the same time providing only one model to compare with the constant model.
- In stepwise logistic regression there are a number of steps listed in the table as each variable is added or removed, creating different models.
- The step is a measure of the improvement in the predictive power of the model since the previous step.
- I will revert to this next class.

2.2 Omnibus Test for Model Coefficients

The overall significance is tested using what SPSS calls the ***Model Chi-square***, which is derived from the likelihood of observing the actual data under the assumption that the model that has been fitted is accurate. There are two hypotheses to test in relation to the overall fit of the model:

H_0 The model is a good fitting model.

H_1 The model is not a good fitting model (i.e. the predictors have a significant effect).

- For the example provided here, the model chi square has 2 degrees of freedom, a value of 24.096 and a probability of $p < 0.000$. (*see Next Page for SPSS Output*).
- Thus, the indication is that the model has a poor fit, with the model containing only the constant indicating that the predictors do have a significant effect and create essentially a different model.
- So we need to look closely at the predictors and from later tables determine if one or both are significant predictors.

Remark upon “Steps”

- This table has 1 step. This is because we are entering both variables and at the same time providing only one model to compare with the constant model.
- In stepwise logistic regression there are a number of steps listed in the table as each variable is added or removed, creating different models.
- The step is a measure of the improvement in the predictive power of the model since the previous step.
- I will revert to this next class.

2.3 Model Summary Table

- The likelihood function can be thought of as a measure of how well a candidate model fits the data (although that is a very simplistic definition).
- Later on in the course, we will meet the AIC Function. The AIC criterion is based on the Likelihood function.
- The likelihood function of a fitted model is commonly re-expressed as $-2LL$ (i.e. The log of the likelihood times minus 2). The 2LL value from the Model Summary table below is 17.359.
- It is not immediately clear how to interpret $-2LL$ yet. However this measure is useful in comparing various candidate models, and is therefore used in **Variable Selection Procedure**, such as backward and forward selection.

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	24.096	2	.000
	Block	24.096	2	.000
	Model	24.096	2	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	17.359 ^a	.552	.737

a. Estimation terminated at iteration number 8 because parameter estimates changed by less than .001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	6.378	8	.605

2.4 Model Summary Table

- The likelihood function can be thought of as a measure of how well a candidate model fits the data (although that is a very simplistic definition).
- Later on in the course, we will meet the AIC Function. The AIC criterion is based on the Likelihood function.
- The likelihood function of a fitted model is commonly re-expressed as $-2LL$ (i.e. The log of the likelihood times minus 2). The 2LL value from the Model Summary table below is 17.359.
- It is not immediately clear how to interpret $-2LL$ yet. However this measure is useful in comparing various candidate models, and is therefore used in **Variable Selection Procedure**, such as backward and forward selection.

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	24.096	2	.000
	Block	24.096	2	.000
	Model	24.096	2	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	17.359 ^a	.552	.737

a. Estimation terminated at iteration number 8 because parameter estimates changed by less than .001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	6.378	8	.605

2.5 Hosmer-Lemeshow Goodness-of-Fit Test

- The Hosmer-Lemeshow Goodness-of-Fit test tells us whether we have constructed a valid overall model or not, and is an alternative to model chi square procedure.
- If the model is a good fit to the data then the Hosmer-Lemeshow Goodness-of-Fit test should have an associated p-value greater than 0.05.

If the H-L goodness-of-fit test statistic is greater than .05, as we want for well-fitting models, we fail to reject the null hypothesis that there is no difference between observed and model-predicted values, implying that the models estimates fit the data at an acceptable level. That is, well-fitting models show non-significance on the H-L goodness-of-fit test. This desirable outcome of non-significance indicates that the model prediction does not significantly differ from the observed.

Performing the Hosmer-Lemeshow Goodness-of-Fit Test

- The Hosmer-Lemeshow test of goodness of fit is not automatically a part of the SPSS logistic regression output. To get this output, we need to go into **options** and tick the box marked Hosmer-Lemeshow test of goodness of fit.

In our example, this gives us the following output:

Step	Chi-square	df	Sig.
1	142.032	6	.000

Therefore, our model is significant, suggesting it does not fit the data. However, as we have a sample size of over 13,000, even very small divergencies of the model from the data would be flagged up and cause significance. Therefore, with samples of this size it is hard to find models that are parsimonious (i.e. that use the minimum amount of independent variables to explain the dependent variable) and fit the data. Therefore, other fit indices might be more appropriate.

Computing the Hosmer and Lemeshow Test Statistic

- The Hosmer and Lemeshow test divides subjects into 10 ordered groups of subjects and then compares the number actually in the each group (observed) to the number predicted by the logistic regression model (predicted).
- The 10 ordered groups are created based on their estimated probability; those with estimated probability below .1 form one group, and so on, up to those with probability .9 to 1.0.
- Each of these categories is further divided into two groups based on the actual observed outcome variable (success, failure). The expected frequencies for each of the cells are obtained from the model. A probability (p) value is computed from the chi-square distribution with 8 degrees of freedom to test the fit of the logistic model.

		take solar panel offer = decline offer		take solar panel offer = take offer		Total
		Observed	Expected	Observed	Expected	
Step 1	1	3	2.996	0	.004	3
	2	3	2.957	0	.043	3
	3	3	2.629	0	.371	3
	4	2	2.136	1	.864	3
	5	2	1.881	1	1.119	3
	6	0	.833	3	2.167	3
	7	0	.376	3	2.624	3
	8	1	.171	2	2.829	3
	9	0	.019	3	2.981	3
	10	0	.000	3	3.000	3

Figure 1: Hosmer and Lemeshow Table

Sample Adequacy

The H-L statistic assumes sampling adequacy, with a rule of thumb being enough cases so that 95% of cells (typically, 10 decile groups times 2 outcome categories = 20 cells) have an expected frequency > 5 . Our H-L statistic has a significance of .605 which means that it is not statistically significant and therefore our model is quite a good fit.

Hosmer and Lemeshow Test			
Step	Chi-square	df	Sig.
1	6.378	8	.605

Figure 2: Hosmer and Lemeshow Statistic

2.6 Hosmer-Lemeshow Goodness-of-Fit Test

- The Hosmer-Lemeshow Goodness-of-Fit test tells us whether we have constructed a valid overall model or not, and is an alternative to model chi square procedure.
- If the model is a good fit to the data then the Hosmer-Lemeshow Goodness-of-Fit test should have an associated p-value greater than 0.05.

If the H-L goodness-of-fit test statistic is greater than .05, as we want for well-fitting models, we fail to reject the null hypothesis that there is no difference between observed and model-predicted values, implying that the models estimates fit the data at an acceptable level. That is, well-fitting models show non-significance on the H-L goodness-of-fit test. This desirable outcome of non-significance indicates that the model prediction does not significantly differ from the observed.

Performing the Hosmer-Lemeshow Goodness-of-Fit Test

- The Hosmer-Lemeshow test of goodness of fit is not automatically a part of the SPSS logistic regression output. To get this output, we need to go into **options** and tick the box marked Hosmer-Lemeshow test of goodness of fit.

In our example, this gives us the following output:

Step	Chi-square	df	Sig.
1	142.032	6	.000

Therefore, our model is significant, suggesting it does not fit the data. However, as we have a sample size of over 13,000, even very small divergencies of the model from the data would be flagged up and cause significance. Therefore, with samples of this size it is hard to find models that are parsimonious (i.e. that use the minimum amount of independent variables to explain the dependent variable) and fit the data. Therefore, other fit indices might be more appropriate.

Computing the Hosmer and Lemeshow Test Statistic

- The Hosmer and Lemeshow test divides subjects into 10 ordered groups of subjects and then compares the number actually in the each group (observed) to the number predicted by the logistic regression model (predicted).
- The 10 ordered groups are created based on their estimated probability; those with estimated probability below .1 form one group, and so on, up to those with probability .9 to 1.0.
- Each of these categories is further divided into two groups based on the actual observed outcome variable (success, failure). The expected frequencies for each of the cells are obtained from the model. A probability (p) value is computed from the chi-square distribution with 8 degrees of freedom to test the fit of the logistic model.

Contingency Table for Hosmer and Lemeshow Test						
		take solar panel offer = decline offer		take solar panel offer = take offer		Total
		Observed	Expected	Observed	Expected	
Step 1	1	3	2.996	0	.004	3
	2	3	2.957	0	.043	3
	3	3	2.629	0	.371	3
	4	2	2.136	1	.864	3
	5	2	1.881	1	1.119	3
	6	0	.833	3	2.167	3
	7	0	.376	3	2.624	3
	8	1	.171	2	2.829	3
	9	0	.019	3	2.981	3
	10	0	.000	3	3.000	3

Figure 3: Hosmer and Lemeshow Table

Sample Adequacy

The H-L statistic assumes sampling adequacy, with a rule of thumb being enough cases so that 95% of cells (typically, 10 decile groups times 2 outcome categories = 20 cells) have an expected frequency > 5 . Our H-L statistic has a significance of .605 which means that it is not statistically significant and therefore our model is quite a good fit.

Hosmer and Lemeshow Test			
Step	Chi-square	df	Sig.
1	6.378	8	.605

Figure 4: Hosmer and Lemeshow Statistic

3 R Squared Diagnostics

- In order to understand how much variation in the dependent variable can be explained by the model (the equivalent of R^2 in multiple regression), you should consult **Model Summary** statistics.
- Logistic regression does not have an equivalent to the R-squared that is found in OLS regression; however, many researchers have tried to come up with one.
- The SPSS output table below contains the *Cox & Snell R Square* and *Nagelkerke R Square* values, which are both methods of calculating the explained variation. These values are sometimes referred to as pseudo R^2 values (and will have lower values than in multiple regression).

- However, they are interpreted in the same manner, but with more caution. Therefore, the explained variation in the dependent variable based on our model ranges from 24.0% to 33.0%, depending on whether you reference the Cox & Snell R^2 or Nagelkerke R^2 methods, respectively.
- Nagelkerke R^2 is a modification of Cox & Snell R^2 , the latter of which cannot achieve a value of 1. For this reason, it is preferable to report the Nagelkerke R^2 value.

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	102.088 ^a	.240	.330

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Figure 5: SPSS output

- Although there is no close analogous statistic in logistic regression to the coefficient of determination R^2 the Model Summary Table provides some approximations. Cox and Snell's R-Square attempts to imitate multiple R-Square based on likelihood, but its maximum can be (and usually is) less than 1.0, making it difficult to interpret.
- Here it is indicating that 55.2% of the variation in the DV is explained by the logistic model.
- Logistic regression does not have an equivalent to the R-squared that is found in OLS regression; however, many people have tried to come up with one. Cox and Snell R Square and Nagelkerke R Square - These are pseudo R-squares.
- Nagelkerke's R-Square is a further modification of the Cox and Snell coefficient to assure that it can vary from 0 to 1. Nagelkerke's R-Square will normally be higher than the Cox and Snell measure. It is part of SPSS output and is the most-reported of the R-squared estimates.
- The Nagelkerke modification that does range from 0 to 1 is a more reliable measure of the relationship. Nagelkerke's R^2 will normally be higher than the Cox and Snell measure.
- **Cox and Snell's R-Square** is an attempt to imitate the interpretation of multiple R-Square based on the likelihood, but its maximum can be (and usually is) less than 1.0, making it difficult to interpret. It is part of SPSS output.

3.0.1 Nagelkerke's R-Square

- Nagelkerke's R^2 is part of SPSS output in the Model Summary table and is the most-reported of the R-squared estimates.
- In our case it is 0.737, indicating a moderately strong relationship of 73.7% between the predictors and the prediction.

3.1 Pseudo R-squares

Cox & Snell R Square and Nagelkerke R Square are two measures from the **pseudo R-squares** family of measures.

There are a wide variety of pseudo-R-square statistics (these are only two of them). Because this statistic does not mean what R-squared means in OLS regression (the proportion of variance explained by the predictors), we suggest interpreting this statistic with great caution.

4 Likelihood Ratio Test

- The likelihood ratio test is a test of the difference between -2LL for the full model with predictors and -2LL for initial chi-square in the null model.
- When probability fails to reach the 5% significance level, we retain the null hypothesis that knowing the independent variables (predictors) has no increased effects (i.e. make no difference) in predicting the dependent.