

## Binary Logistic Regression with PASW/SPSS®

---

Logistic regression is used to predict a categorical (usually dichotomous) variable from a set of predictor variables. With a categorical dependent variable, discriminant function analysis is usually employed if all of the predictors are continuous and nicely distributed; logit analysis is usually employed if all of the predictors are categorical; and logistic regression is often chosen if the predictor variables are a mix of continuous and categorical variables and/or if they are not nicely distributed (logistic regression makes no assumptions about the distributions of the predictor variables). Logistic regression has been especially popular with medical research in which the dependent variable is whether or not a patient has a disease.

For a logistic regression, the predicted dependent variable is a function of the probability that a particular subject will be in one of the categories (for example, the probability that Suzie Cue has the disease, given her set of scores on the predictor variables).

### Description of the Research Used to Generate Our Data

As an example of the use of logistic regression in psychological research, consider the research done by Wuensch and Poteat and published in the [\*Journal of Social Behavior and Personality\*](#), 1998, 13, 139-150. College students ( $N = 315$ ) were asked to pretend that they were serving on a university research committee hearing a complaint against animal research being conducted by a member of the university faculty. The complaint included a description of the research in simple but emotional language. Cats were being subjected to stereotaxic surgery in which a cannula was implanted into their brains. Chemicals were then introduced into the cats' brains via the cannula and the cats given various psychological tests. Following completion of testing, the cats' brains were subjected to histological analysis. The complaint asked that the researcher's authorization to conduct this research be withdrawn and the cats turned over to the animal rights group that was filing the complaint. It was suggested that the research could just as well be done with computer simulations.

In defense of his research, the researcher provided an explanation of how steps had been taken to assure that no animal felt much pain at any time, an explanation that computer simulation was not an adequate substitute for animal research, and an explanation of what the benefits of the research were. Each participant read one of five different scenarios which described the goals and benefits of the research. They were:

- COSMETIC -- testing the toxicity of chemicals to be used in new lines of hair care products.
- THEORY -- evaluating two competing theories about the function of a particular nucleus in the brain.
- MEAT -- testing a synthetic growth hormone said to have the potential of increasing meat production.

- VETERINARY -- attempting to find a cure for a brain disease that is killing both domestic cats and endangered species of wild cats.
- MEDICAL -- evaluating a potential cure for a debilitating disease that afflicts many young adult humans.

After reading the case materials, each participant was asked to decide whether or not to withdraw Dr. Wissen's authorization to conduct the research and, among other things, to fill out D. R. Forysth's Ethics Position Questionnaire (*Journal of Personality and Social Psychology*, 1980, 39, 175-184), which consists of 20 Likert-type items, each with a 9-point response scale from "completely disagree" to "completely agree." Persons who score high on the relativism dimension of this instrument reject the notion of universal moral principles, preferring personal and situational analysis of behavior. Persons who score high on the idealism dimension believe that ethical behavior will always lead only to good consequences, never to bad consequences, and never to a mixture of good and bad consequences.

Having committed the common error of projecting myself onto others, I once assumed that all persons make ethical decisions by weighing good consequences against bad consequences -- but for the idealist the presence of any bad consequences may make a behavior unethical, regardless of good consequences. Research by Hal Herzog and his students at Western Carolina has shown that animal rights activists tend to be high in idealism and low in relativism (see me for references if interested). Are idealism and relativism (and gender and purpose of the research) related to attitudes towards animal research in college students? Let's run the logistic regression and see.

## Using a Single Dichotomous Predictor, Gender of Subject

Let us first consider a simple (bivariate) logistic regression, using subjects' **decisions** as the dichotomous criterion variable and their **gender** as a dichotomous predictor variable. I have coded gender with 0 = Female, 1 = Male, and decision with 0 = "Stop the Research" and 1 = "Continue the Research".

Our regression model will be predicting the logit, that is, the natural log of the odds of having made one or the other decision. That is,

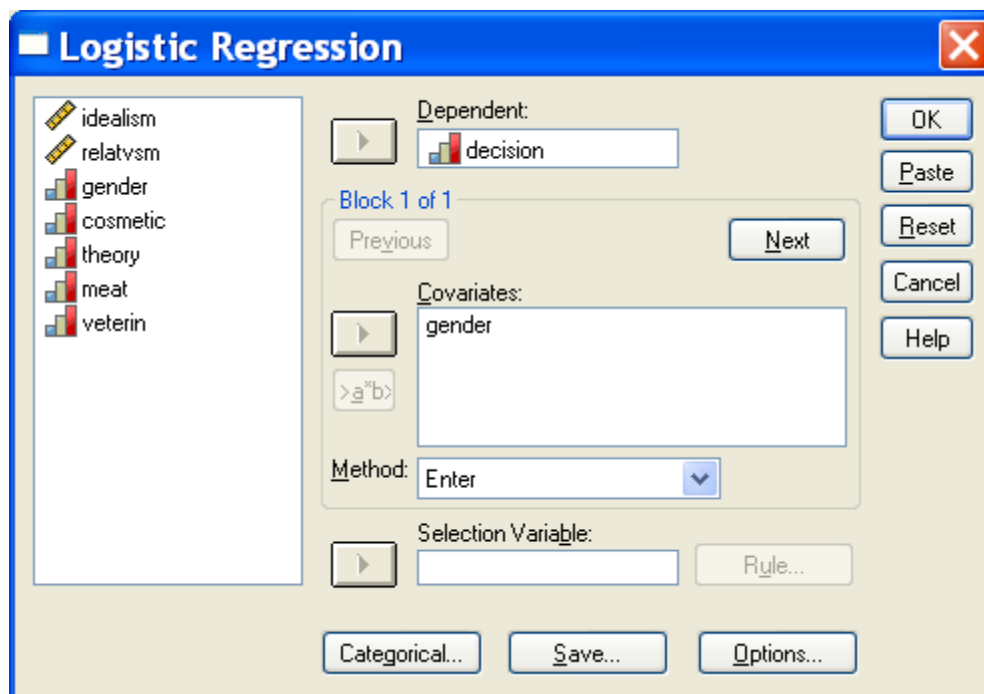
$$\ln(ODDS) = \ln\left(\frac{\hat{Y}}{1 - \hat{Y}}\right) = a + bX, \text{ where } \hat{Y} \text{ is the predicted probability of the event}$$

**which is coded with 1 (continue the research)** rather than with 0 (stop the research),  $1 - \hat{Y}$  is the predicted probability of the other decision, and  $X$  is our predictor variable, gender. Some statistical programs (such as SAS) predict the event which is coded with the smaller of the two numeric codes. By the way, if you have ever wondered what is "natural" about the natural log, you can find an answer of sorts at [http://www.math.toronto.edu/mathnet/answers/answers\\_13.html](http://www.math.toronto.edu/mathnet/answers/answers_13.html).

Our model will be constructed by **an iterative maximum likelihood procedure**. The program will start with arbitrary values of the regression coefficients and will construct an initial model for predicting the observed data. It will then evaluate errors in such prediction and change the regression coefficients so as make the likelihood of the observed data greater under the new model. This procedure is repeated until the model

converges -- that is, until the differences between the newest model and the previous model are trivial.

Open the data file at <http://core.ecu.edu/psyc/wuenschk/SPSS/Logistic.sav>. Click **Analyze, Regression, Binary Logistic**. Scoot the decision variable into the Dependent box and the gender variable into the Covariates box. The dialog box should now look like this:



Click **OK**.

Look at the statistical output. We see that there are 315 cases used in the analysis.

**Case Processing Summary**

Unweighted Cases <sup>a</sup>		N	Percent
Selected Cases	Included in Analysis	315	100.0
	Missing Cases	0	.0
	Total	315	100.0
Unselected Cases		0	.0
Total		315	100.0

a. If weight is in effect, see classification table for the total number of cases.

The **Block 0** output is for a model that includes only the intercept (which SPSS calls the constant). Given the base rates of the two decision options ( $187/315 = 59\%$  decided to stop the research,  $41\%$  decided to allow it to continue), and no other information, the best strategy is to predict, for every case, that the subject will decide to stop the research. Using that strategy, you would be correct  $59\%$  of the time.

Classification Table<sup>a,b</sup>

Observed			Predicted		
			decision		Percentage Correct
			stop	continue	
Step 0	decision	stop	187	0	100.0
		continue	128	0	.0
Overall Percentage					59.4

a. Constant is included in the model.

b. The cut value is .500

Under **Variables in the Equation** you see that the intercept-only model is  **$\ln(\text{odds}) = -.379$** . If we exponentiate both sides of this expression we find that our predicted odds  $[\text{Exp}(B)] = .684$ . That is, the predicted odds of deciding to continue the research is .684. Since 128 of our subjects decided to continue the research and 187 decided to stop the research, our **observed odds are  $128/187 = .684$** .

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 0 Constant	-.379	.115	10.919	1	.001	.684

Now look at the **Block 1** output. Here SPSS has added the gender variable as a predictor. **Omnibus Tests of Model Coefficients** gives us a Chi-Square of 25.653 on 1 *df*, significant beyond .001. This is a test of the null hypothesis that adding the gender variable to the model has not significantly increased our ability to predict the decisions made by our subjects.

Omnibus Tests of Model Coefficients

	Chi-square	df	Sig.
Step 1 Step	25.653	1	.000
Block	25.653	1	.000
Model	25.653	1	.000

Under **Model Summary** we see that the **-2 Log Likelihood** statistic is 399.913. This statistic **measures how poorly the model predicts the decisions** -- the smaller the statistic the better the model. Although SPSS does not give us this statistic for the model that had only the intercept, I know it to be 425.666. Adding the gender variable reduced the -2 Log Likelihood statistic by  $425.666 - 399.913 = 25.653$ , the  $\chi^2$  statistic we just discussed in the previous paragraph. The **Cox & Snell  $R^2$**  can be interpreted like  $R^2$  in a multiple regression, but cannot reach a maximum value of 1. The **Nagelkerke  $R^2$**  can reach a maximum of 1.

**Model Summary**

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	399.913 <sup>a</sup>	.078	.106

a. Estimation terminated at iteration number 3 because parameter estimates changed by less than .001.

The **Variables in the Equation** output shows us that the regression equation is  $\ln(ODDS) = -.847 + 1.217\text{Gender}$ .

**Variables in the Equation**

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1	gender	1.217	.245	24.757	1	.000	3.376
	Constant	-.847	.154	30.152	1	.000	.429

a. Variable(s) entered on step 1: gender.

We can now use this model to **predict the odds** that a subject of a given gender will decide to continue the research. The odds prediction equation is  $ODDS = e^{a+bX}$ . If our subject is a woman (gender = 0), then the  $ODDS = e^{-.847 + 1.217(0)} = e^{-.847} = 0.429$ . That is, a woman is only .429 as likely to decide to continue the research as she is to decide to stop the research. If our subject is a man (gender = 1), then the  $ODDS = e^{-.847 + 1.217(1)} = e^{.37} = 1.448$ . That is, a man is 1.448 times more likely to decide to continue the research than to decide to stop the research.

We can easily **convert odds to probabilities**. For our women,

$$\hat{Y} = \frac{ODDS}{1 + ODDS} = \frac{0.429}{1.429} = 0.30$$

That is, our model predicts that 30% of women will

decide to continue the research. For our men,  $\hat{Y} = \frac{ODDS}{1 + ODDS} = \frac{1.448}{2.448} = 0.59$ . That is, our model predicts that 59% of men will decide to continue the research

The **Variables in the Equation** output also gives us the **Exp(B)**. This is better known as **the odds ratio** predicted by the model. This odds ratio can be computed by **raising the base of the natural log to the  $b^{\text{th}}$  power, where  $b$  is the slope from our logistic regression equation**. For our model,  $e^{1.217} = 3.376$ . That tells us that the model predicts that the odds of deciding to continue the research are 3.376 times higher for men than they are for women. For the men, the odds are 1.448, and for the women they are 0.429. The odds ratio is  $1.448 / 0.429 = 3.376$ .

The results of our logistic regression can be used to **classify subjects** with respect to what decision we think they will make. As noted earlier, our model leads to the prediction that the probability of deciding to continue the research is 30% for women and 59% for men. Before we can use this information to classify subjects, we need to

have a decision rule. Our **decision rule** will take the following form: If the probability of the event is greater than or equal to some threshold, we shall predict that the event will take place. By default, SPSS sets this threshold to .5. While that seems reasonable, in many cases we may want to set it higher or lower than .5. More on this later. Using the default threshold, SPSS will classify a subject into the “Continue the Research” category if the estimated probability is .5 or more, which it is for every male subject. SPSS will classify a subject into the “Stop the Research” category if the estimated probability is less than .5, which it is for every female subject.

The **Classification Table** shows us that this rule allows us to correctly classify  $68 / 128 = 53\%$  of the subjects where the predicted event (deciding to continue the research) was observed. This is known as the **sensitivity** of prediction, the  $P(\text{correct} \mid \text{event did occur})$ , that is, the percentage of occurrences correctly predicted. We also see that this rule allows us to correctly classify  $140 / 187 = 75\%$  of the subjects where the predicted event was not observed. This is known as the **specificity** of prediction, the  $P(\text{correct} \mid \text{event did not occur})$ , that is, the percentage of nonoccurrences correctly predicted. Overall our predictions were correct 208 out of 315 times, for an **overall success rate** of 66%. Recall that it was only 59% for the model with intercept only.

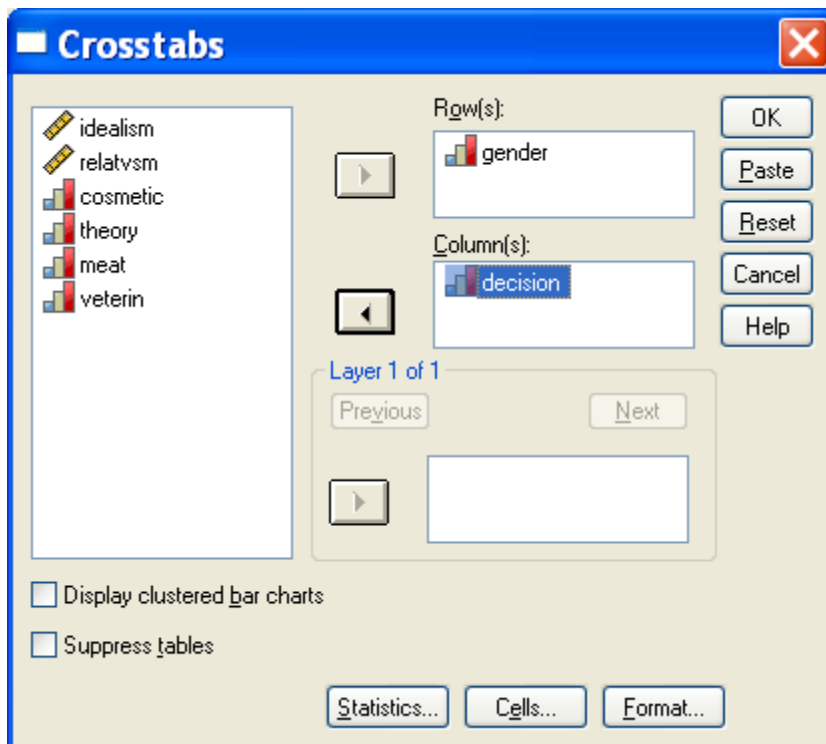
**Classification Table<sup>a</sup>**

Observed			Predicted		
			decision		Percentage Correct
			stop	continue	
Step 1	decision	stop	140	47	74.9
		continue	60	68	53.1
Overall Percentage					66.0

a. The cut value is .500

We could focus on error rates in classification. A **false positive** would be predicting that the event would occur when, in fact, it did not. Our decision rule predicted a decision to continue the research 115 times. That prediction was wrong 47 times, for a false positive rate of  $47 / 115 = 41\%$ . A **false negative** would be predicting that the event would not occur when, in fact, it did occur. Our decision rule predicted a decision not to continue the research 200 times. That prediction was wrong 60 times, for a false negative rate of  $60 / 200 = 30\%$ .

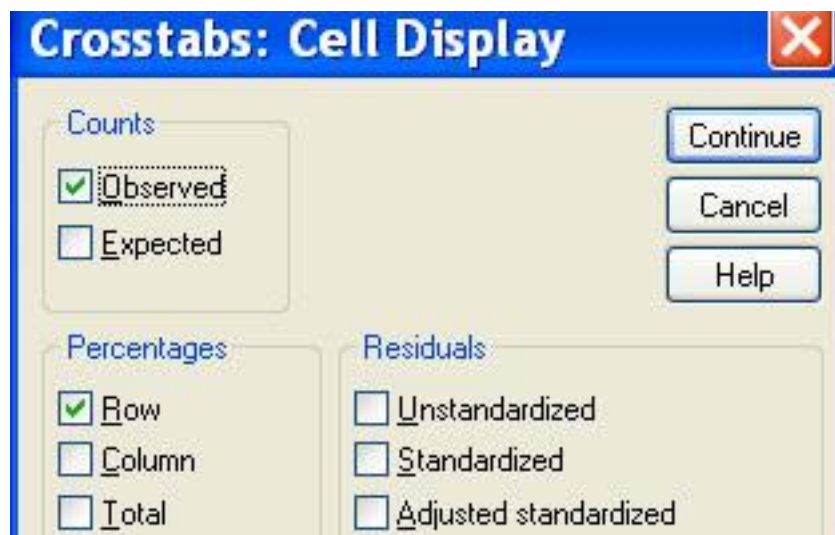
It has probably occurred to you that you could have used a simple **Pearson Chi-Square Contingency Table Analysis** to answer the question of whether or not there is a significant relationship between gender and decision about the animal research. Let us take a quick look at such an analysis. In SPSS click **Analyze, Descriptive Statistics, Crosstabs**. Scoot gender into the rows box and decision into the columns box. The dialog box should look like this:



Now click the **Statistics** box. Check **Chi-Square** and then click **Continue**.



Now click the **Cells** box. Check **Observed Counts** and **Row Percentages** and then click Continue.





Back on the initial page, click **OK**.

In the **Crosstabulation output** you will see that 59% of the men and 30% of the women decided to continue the research, just as predicted by our logistic regression.

**gender \* decision Crosstabulation**

			decision		Total
			stop	continue	
gender	Female	Count	140	60	200
		% within gender	70.0%	30.0%	100.0%
	Male	Count	47	68	115
		% within gender	40.9%	59.1%	100.0%
Total		Count	187	128	315
		% within gender	59.4%	40.6%	100.0%

You will also notice that the **Likelihood Ratio Chi-Square** is 25.653 on 1 *df*, the same test of significance we got from our logistic regression, and the Pearson Chi-Square is almost the same (25.685). If you are thinking, “Hey, this logistic regression is nearly equivalent to a simple Pearson Chi-Square,” you are correct, in this simple case. Remember, however, that we can add additional predictor variables, and those additional predictors can be either categorical or continuous -- you can’t do that with a simple Pearson Chi-Square.

**Chi-Square Tests**

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	25.685 <sup>b</sup>	1	.000
Likelihood Ratio	25.653	1	.000
N of Valid Cases	315		

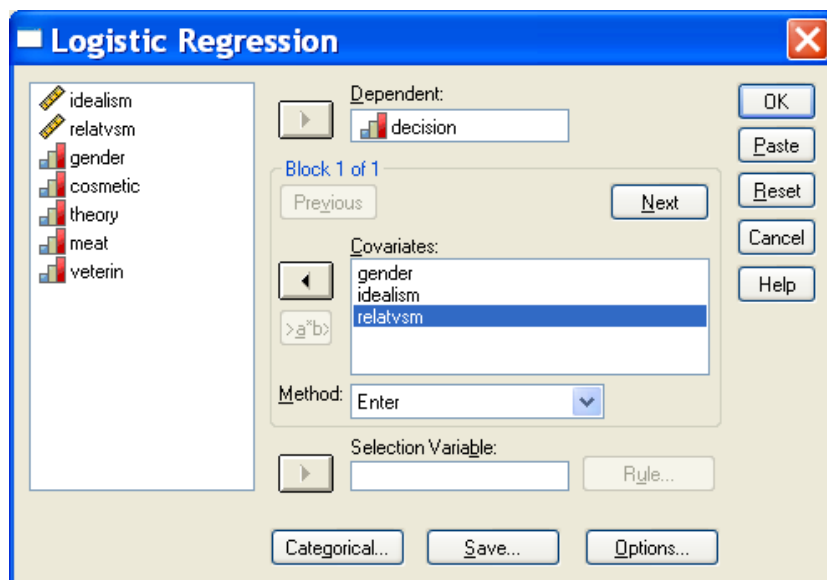
a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 46.73.

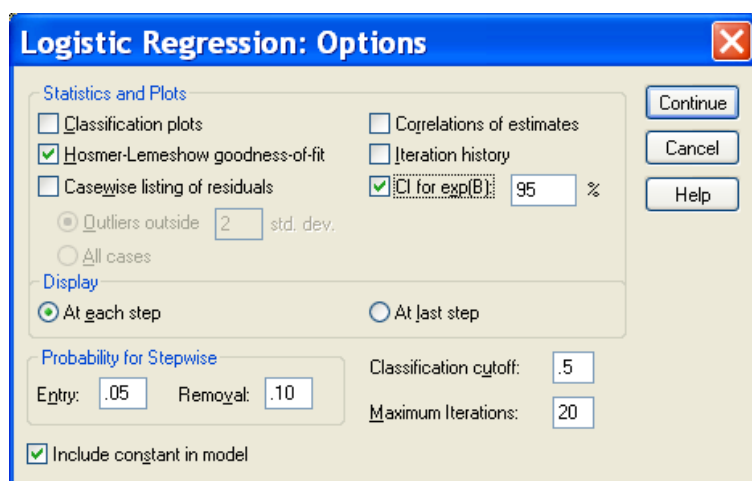
## Multiple Predictors, Both Categorical and Continuous

Now let us conduct an analysis that will better tap the strengths of logistic regression. Click **Analyze, Regression, Binary Logistic**. Scoot the decision variable in the **Dependent** box and gender, idealism, and relativism into the **Covariates** box.





Click **Options** and check “Hosmer-Lemeshow goodness of fit” and “CI for exp(B) 95%.”



Continue, OK. Look at the output.

In the **Block 1** output, notice that the -2 Log Likelihood statistic has dropped to 346.503, indicating that our expanded model is doing a better job at predicting decisions than was our one-predictor model. The  $R^2$  statistics have also increased.

**Model Summary**

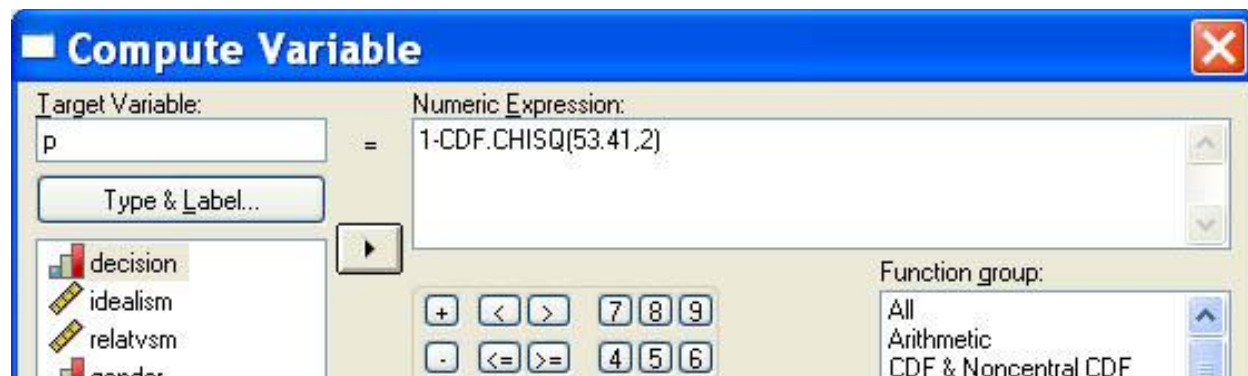
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	346.503 <sup>a</sup>	.222	.300

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

We can test the significance of the difference between any two models, as long as one model is nested within the other. Our one-predictor model had a

–2 **Log Likelihood** statistic of 399.913. Adding the ethical ideology variables (idealism and relativism) produced a **decrease of 53.41**. This difference is a  $\chi^2$  on 2 *df* (one *df* for each predictor variable).

To **determine the p value** associated with this  $\chi^2$ , just click **Transform, Compute**. Enter the letter p in the **Target Variable** box. In the **Numeric Expression** box, type **1-CDF.CHISQ(53.41,2)**. The dialog box should look like this:



Click **OK** and then go to the **SPSS Data Editor, Data View**. You will find a new column, p, with the value of .00 in every cell. If you go to the **Variable View** and set the number of **decimal points** to 5 for the p variable you will see that the value of p is .00000. **We conclude that adding the ethical ideology variables significantly improved the model,  $\chi^2(2, N = 315) = 53.41, p < .001$ .**

Note that our overall success rate in classification has improved from 66% to 71%.

**Classification Table<sup>a</sup>**

Observed			Predicted		
			decision		Percentage Correct
			stop	continue	
Step 1	decision	stop	151	36	80.7
		continue	55	73	57.0
Overall Percentage					71.1

a. The cut value is .500

The **Hosmer-Lemeshow** tests the null hypothesis that there is a linear relationship between the predictor variables and the log odds of the criterion variable. Cases are arranged in order by their predicted probability on the criterion variable. These ordered cases are then divided into ten groups (lowest decile [prob < .1] to highest decile [prob > .9]). Each of these groups is then divided into two groups on the basis of actual score on the criterion variable. This results in a 2 x 10 contingency table. Expected frequencies are computed based on the assumption that there is a linear relationship between the weighted combination of the predictor variables and the log odds of the criterion variable. For the outcome = no (decision = stop for our data) column, the expected frequencies will run from high (for the lowest decile) to low (for the highest decile). For the outcome = yes column the frequencies will run from low to high.

A chi-square statistic is computed comparing the observed frequencies with those expected under the linear model. A nonsignificant chi-square indicates that the data fit the model well.

**Hosmer and Lemeshow Test**

Step	Chi-square	df	Sig.
1	8.810	8	.359

**Contingency Table for Hosmer and Lemeshow Test**

		decision = stop		decision = continue		Total
		Observed	Expected	Observed	Expected	
Step 1	1	29	29.331	3	2.669	32
	2	30	27.673	2	4.327	32
	3	28	25.669	4	6.331	32
	4	20	23.265	12	8.735	32
	5	22	20.693	10	11.307	32
	6	15	18.058	17	13.942	32
	7	15	15.830	17	16.170	32
	8	10	12.920	22	19.080	32
	9	12	9.319	20	22.681	32
	10	6	4.241	21	22.759	27

## Using a $K > 2$ Categorical Predictor

We can use a categorical predictor that has more than two levels. For our data, the stated purpose of the research is such a predictor. While SPSS can dummy code such a predictor for you, I prefer to set up my own dummy variables. You will need **K-1 dummy variables** to represent K groups. Since we have five levels of purpose of the research, we shall need 4 dummy variables. Each of the subjects will have a score of either 0 or 1 on each of the dummy variables. For each dummy variable a score of 0 will indicate that the subject does not belong to the group represented by that dummy variable and a score of 1 will indicate that the subject does belong to the group represented by that dummy variable. One of the groups will not be represented by a dummy variable. If it is reasonable to consider one of your groups as a **reference group to which each other group should be compared**, make that group the one which is not represented by a dummy variable.

I decided that I wanted to compare each of the cosmetic, theory, meat, and veterinary groups with the medical group, so I set up a dummy variable for each of the groups except the medical group. Take a look at our data in the **data editor**. Notice that the first subject has a score of 1 for the cosmetic dummy variable and 0 for the other three dummy variables. That subject was told that the purpose of the research was to test the safety of a new ingredient in hair care products. Now scoot to the bottom of the data file. The last subject has a score of 0 for each of the four dummy

variables. That subject was told that the purpose of the research was to evaluate a treatment for a debilitating disease that afflicts humans of college age.

Click **Analyze, Regression, Binary Logistic** and add to the list of covariates the four dummy variables. You should now have the decision variable in the **Dependent** box and all of the other variables (but not the p value column) in the **Covariates** box. Click **OK**.

The Block 0 “Variables not in the Equation” show how much the -2LL would drop if a single predictor were added to the model (which already has the intercept)

Variables not in the Equation					
			Score	df	Sig.
Step 0	Variables	gender	25.685	1	.000
		idealism	47.679	1	.000
		relativism	7.239	1	.007
		cosmetic	.003	1	.955
		theory	2.933	1	.087
		meat	.556	1	.456
		veterin	.013	1	.909
		Overall Statistics	77.665	7	.000

Look at the output, **Block 1**. Under **Omnibus Tests of Model Coefficients** we see that our latest model is significantly better than a model with only the intercept.

Omnibus Tests of Model Coefficients				
		Chi-square	df	Sig.
Step 1	Step	87.506	7	.000
	Block	87.506	7	.000
	Model	87.506	7	.000

Under **Model Summary** we see that our  $R^2$  statistics have increased again and the **-2 Log Likelihood** statistic has dropped from 346.503 to 338.060. Is this drop statistically significant? The  $\chi^2$ , is the difference between the two -2 log likelihood values, 8.443, on 4 *df* (one *df* for each dummy variable). Using **1-CDF.CHISQ(8.443,4)**, we obtain an upper-tailed *p* of .0766, short of the usual standard of statistical significance. I shall, however, retain these dummy variables, since I have an a priori interest in the comparison made by each dummy variable.

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	338.060 <sup>a</sup>	.243	.327

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

In the **Classification Table**, we see a small increase in our overall success rate, from 71% to 72%.

**Classification Table**

Observed			Predicted		
			decision		Percentage Correct
			stop	continue	
Step 1	decision	stop	152	35	81.3
		continue	54	74	57.8
Overall Percentage					71.7

a. The cut value is .500

I would like you to compute the values for Sensitivity, Specificity, False Positive Rate, and False Negative Rate for this model, using the default .5 cutoff.

Sensitivity	percentage of occurrences correctly predicted
Specificity	percentage of nonoccurrences correctly predicted
False Positive Rate	percentage of predicted occurrences which are incorrect
False Negative Rate	percentage of predicted nonoccurrences which are incorrect

Remember that the predicted event was a decision to continue the research.

Under **Variables in the Equation** we are given regression coefficients and odds ratios.

**Variables in the Equation**

		B	Wald	df	Sig.	Exp(B)	95.0% C.I. for EXP(B)	
							Lower	Upper
Step 1	gender	1.255	20.586	1	.000	3.508	2.040	6.033
	idealism	-.701	37.891	1	.000	.496	.397	.620
	relativism	.326	6.634	1	.010	1.386	1.081	1.777
	cosmetic	-.709	2.850	1	.091	.492	.216	1.121
	theory	-1.160	7.346	1	.007	.314	.136	.725
	meat	-.866	4.164	1	.041	.421	.183	.966
	veterin	-.542	1.751	1	.186	.581	.260	1.298
	Constant	2.279	4.867	1	.027	9.766		

a. Variable(s) entered on step 1: gender, idealism, relativism, cosmetic, theory, meat, veterin.

We are also given a statistic I have ignored so far, the **Wald Chi-Square** statistic, which tests the unique contribution of each predictor, in the context of the other predictors -- that is, holding constant the other predictors -- that is, eliminating any overlap between predictors. Notice that each predictor meets the conventional .05 standard for statistical significance, except for the dummy variable for cosmetic research and for veterinary research. I should note that the Wald  $\chi^2$  has been criticized

for being too **conservative**, that is, lacking adequate power. An alternative would be to test the significance of each predictor by eliminating it from the full model and testing the significance of the increase in the -2 log likelihood statistic for the reduced model. That would, of course, require that you construct  $p+1$  models, where  $p$  is the number of predictor variables.

Let us now interpret the odds ratios.

- The .496 odds ratio for idealism indicates that the odds of approval are more than cut in half for each one point increase in respondent's idealism score. Inverting this odds ratio for easier interpretation, for each one point increase on the idealism scale there was a doubling of the odds that the respondent would not approve the research.
- Relativism's effect is smaller, and in the opposite direction, with a one point increase on the nine-point relativism scale being associated with the odds of approving the research increasing by a multiplicative factor of 1.39.
- The odds ratios of the scenario dummy variables compare each scenario except medical to the medical scenario. For the theory dummy variable, the .314 odds ratio means that the odds of approval of theory-testing research are only .314 times those of medical research.
- Inverted odds ratios for the dummy variables coding the effect of the scenario variable indicated that the odds of approval for the medical scenario were 2.38 times higher than for the meat scenario and 3.22 times higher than for the theory scenario.

Let us now revisit the issue of the **decision rule** used to determine into which group to classify a subject given that subject's estimated probability of group membership. While the most obvious decision rule would be to classify the subject into the target group if  $p > .5$  and into the other group if  $p < .5$ , you may well want to choose a different decision rule given the relative seriousness of making one sort of error (for example, declaring a patient to have breast cancer when she does not) or the other sort of error (declaring the patient not to have breast cancer when she does).

Repeat our analysis with classification done with a different decision rule. Click **Analyze, Regression, Binary Logistic, Options**. In the resulting dialog window, change the **Classification Cutoff** from .5 to .4. The window should look like this:

**Logistic Regression: Options**

**Statistics and Plots**

☐ Classification plots

☒ Hosmer-Lemeshow goodness-of-fit

☐ Casewise listing of residuals

☒ Correlations of estimates

☐ Iteration history

☒ CI for exp(B): 95 %

☒ Outliers outside 2 std. dev.

☐ All cases

**Display**

☒ At each step

☐ At last step

**Probability for Stepwise**

Entry: .05 Removal: .10

**Classification cutoff:** .4

**Maximum Iterations:** 20

☒ Include constant in model

Continue Cancel Help

Click Continue, OK.

Now SPSS will classify a subject into the "Continue the Research" group if the estimated probability of membership in that group is .4 or higher, and into the "Stop the Research" group otherwise. Take a look at the classification output and see how the change in cutoff has changed the classification results. Fill in the table below to compare the two models with respect to classification statistics.

Value When Cutoff =	.5	.4
Sensitivity		
Specificity		
False Positive Rate		
False Negative Rate		
Overall % Correct		

SAS makes it much easier to see the effects of the decision rule on sensitivity etc. Using the ctable option, one gets output like this:



---

The LOGISTIC Procedure

## Classification Table

Prob Level	Correct		Incorrect		Correct	Percentages		False POS	False NEG
	Event	Non- Event	Event	Non- Event		Sensi- tivity	Speci- ficity		
0.160	123	56	131	5	56.8	96.1	29.9	51.6	8.2
0.180	122	65	122	6	59.4	95.3	34.8	50.0	8.5
0.200	120	72	115	8	61.0	93.8	38.5	48.9	10.0
0.220	116	84	103	12	63.5	90.6	44.9	47.0	12.5
0.240	113	93	94	15	65.4	88.3	49.7	45.4	13.9
0.260	110	100	87	18	66.7	85.9	53.5	44.2	15.3
0.280	108	106	81	20	67.9	84.4	56.7	42.9	15.9
0.300	105	108	79	23	67.6	82.0	57.8	42.9	17.6
0.320	103	115	72	25	69.2	80.5	61.5	41.1	17.9
0.340	100	118	69	28	69.2	78.1	63.1	40.8	19.2
0.360	97	120	67	31	68.9	75.8	64.2	40.9	20.5
0.380	96	124	63	32	69.8	75.0	66.3	39.6	20.5
0.400	94	130	57	34	71.1	73.4	69.5	37.7	20.7
0.420	88	134	53	40	70.5	68.8	71.7	37.6	23.0
0.440	86	140	47	42	71.7	67.2	74.9	35.3	23.1
0.460	79	141	46	49	69.8	61.7	75.4	36.8	25.8
0.480	75	144	43	53	69.5	58.6	77.0	36.4	26.9
0.500	71	147	40	57	69.2	55.5	78.6	36.0	27.9
0.520	69	152	35	59	70.2	53.9	81.3	33.7	28.0
0.540	67	157	30	61	71.1	52.3	84.0	30.9	28.0
0.560	65	159	28	63	71.1	50.8	85.0	30.1	28.4
0.580	61	159	28	67	69.8	47.7	85.0	31.5	29.6
0.600	56	162	25	72	69.2	43.8	86.6	30.9	30.8
0.620	50	165	22	78	68.3	39.1	88.2	30.6	32.1
0.640	48	166	21	80	67.9	37.5	88.8	30.4	32.5
0.660	43	170	17	85	67.6	33.6	90.9	28.3	33.3
0.680	40	170	17	88	66.7	31.3	90.9	29.8	34.1
0.700	36	173	14	92	66.3	28.1	92.5	28.0	34.7
0.720	30	177	10	98	65.7	23.4	94.7	25.0	35.6
0.740	28	178	9	100	65.4	21.9	95.2	24.3	36.0
0.760	23	180	7	105	64.4	18.0	96.3	23.3	36.8
0.780	22	180	7	106	64.1	17.2	96.3	24.1	37.1
0.800	18	181	6	110	63.2	14.1	96.8	25.0	37.8
0.820	17	182	5	111	63.2	13.3	97.3	22.7	37.9
0.840	13	184	3	115	62.5	10.2	98.4	18.8	38.5
0.860	12	185	2	116	62.5	9.4	98.9	14.3	38.5
0.880	8	185	2	120	61.3	6.3	98.9	20.0	39.3
0.900	7	185	2	121	61.0	5.5	98.9	22.2	39.5
0.920	5	187	0	123	61.0	3.9	100.0	0.0	39.7
0.940	1	187	0	127	59.7	0.8	100.0	0.0	40.4
0.960	1	187	0	127	59.7	0.8	100.0	0.0	40.4
0.980	0	187	0	128	59.4	0.0	100.0	.	40.6

---

The classification results given by SAS are a little less impressive because SAS uses a jackknifed classification procedure. Classification results are biased when the coefficients used to classify a subject were developed, in part, with data provided by that same subject. SPSS' classification results do not remove such bias. With jackknifed classification, SAS eliminates the subject currently being classified when computing the coefficients used to classify that subject. Of course, this procedure is

computationally more intense than that used by SPSS. If you would like to learn more about conducting logistic regression with SAS, see my document at <http://core.ecu.edu/psyc/wuenschk/MV/MultReg/Logistic-SAS.doc>.

## Beyond An Introduction to Logistic Regression

I have left out of this handout much about logistic regression. We could consider logistic regression with a criterion variable with more than two levels, with that variable being either qualitative or ordinal. We could consider testing of interaction terms. We could consider sequential and stepwise construction of logistic models. We could talk about detecting outliers among the cases, dealing with multicollinearity and nonlinear relationships between predictors and the logit, and so on. If you wish to learn more about logistic regression, I recommend, as a starting point, Chapter 10 in *Using Multivariate Statistics*, 5<sup>th</sup> edition, by Tabachnick and Fidell (Pearson, 2007).

## Presenting the Results

Let me close with an example of how to present the results of a logistic regression. In the example below you will see that I have included both the multivariate analysis (logistic regression) and univariate analysis. I assume that you all already know how to conduct the univariate analyses I present below.

Table 1  
*Effect of Scenario on Percentage of Participants Voting to Allow the Research to Continue and Participants' Mean Justification Score*

Scenario	Percentage Support
Theory	31
Meat	37
Cosmetic	40
Veterinary	41
Medical	54

As shown in Table 1, only the medical research received support from a majority of the respondents. Overall a majority of respondents (59%) voted to stop the research. Logistic regression analysis was employed to predict the probability that a participant would approve continuation of the research. The predictor variables were participant's gender, idealism, relativism, and four dummy variables coding the scenario. A test of the full model versus a model with intercept only was statistically significant,  $\chi^2(7, N = 315) = 87.51, p < .001$ . The model was able correctly to classify 73% of those who approved the research and 70% of those who did not, for an overall success rate of 71%.

Table 2 shows the logistic regression coefficient, Wald test, and odds ratio for each of the predictors. Employing a .05 criterion of statistical significance, gender, idealism, relativism, and two of the scenario dummy variables had significant partial effects. The odds ratio for gender indicates that when holding all other variables constant, a man is 3.5 times more likely to approve the research than is a woman. Inverting the odds ratio for idealism reveals that for each one point increase on the nine-

point idealism scale there is a doubling of the odds that the participant will not approve the research. Although significant, the effect of relativism was much smaller than that of idealism, with a one point increase on the nine-point idealism scale being associated with the odds of approving the research increasing by a multiplicative factor of 1.39. The scenario variable was dummy coded using the medical scenario as the reference group. Only the theory and the meat scenarios were approved significantly less than the medical scenario. Inverted odds ratios for these dummy variables indicate that the odds of approval for the medical scenario were 2.38 times higher than for the meat scenario and 3.22 times higher than for the theory scenario.

Table 2

*Logistic Regression Predicting Decision From Gender, Ideology, and Scenario*

Predictor	<i>B</i>	Wald $\chi^2$	<i>p</i>	Odds Ratio
Gender	1.25	20.59	< .001	3.51
Idealism	-0.70	37.89	< .001	0.50
Relativism	0.33	6.63	.01	1.39
Scenario				
Cosmetic	-0.71	2.85	.091	0.49
Theory	-1.16	7.35	.007	0.31
Meat	-0.87	4.16	.041	0.42
Veterinary	-0.54	1.75	.186	0.58

Univariate analysis indicated that men were significantly more likely to approve the research (59%) than were women (30%),  $\chi^2(1, N = 315) = 25.68, p < .001$ , that those who approved the research were significantly less idealistic ( $M = 5.87, SD = 1.23$ ) than those who didn't ( $M = 6.92, SD = 1.22$ ),  $t(313) = 7.47, p < .001$ , that those who approved the research were significantly more relativistic ( $M = 6.26, SD = 0.99$ ) than those who didn't ( $M = 5.91, SD = 1.19$ ),  $t(313) = 2.71, p = .007$ , and that the omnibus effect of scenario fell short of significance,  $\chi^2(4, N = 315) = 7.44, p = .11$ .

## Interaction Terms

Interaction terms can be included in a logistic model. When the variables in an interaction are continuous they probably should be centered. Consider the following research: Mock jurors are presented with a criminal case in which there is some doubt about the guilt of the defendant. For half of the jurors the defendant is physically attractive, for the other half she is "plain." Half of the jurors are asked to recommend a verdict without having deliberated, the other half are asked about their recommendation only after a short deliberation with others. The deliberating mock jurors were primed with instructions predisposing them to change their opinion if convinced by the arguments of others. We could use a logit analysis here, but elect to use a logistic regression instead. The article in which these results were published is: Patry, M. W.

(2008). Attractive but guilty: Deliberation and the physical attractiveness bias. *Psychological Reports*, 102, 727-733.

The data are in Logistic2x2x2.sav at my [SPSS Data Page](#). Download the data and bring them into SPSS. Each row in the data file represents one cell in the three-way contingency table. Freq is the number of scores in the cell.

	Delib	Plain	Guilty	Freq
1	0	0	0	13
2	0	0	1	14
3	0	1	0	8
4	0	1	1	27
5	1	0	0	22
6	1	0	1	8
7	1	1	0	29
8	1	1	1	1

Tell SPSS to weight cases by Freq. Data, Weight Cases:

**Weight Cases**

Delib  
Plain  
Guilty

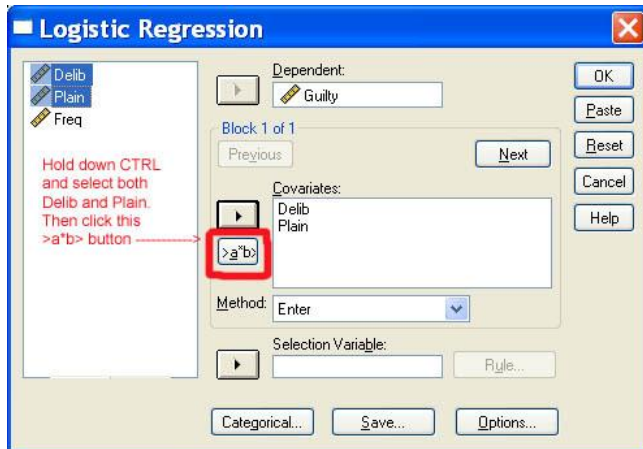
☐ Do not weight cases  
☒ Weight cases by

Frequency Variable:  
Freq

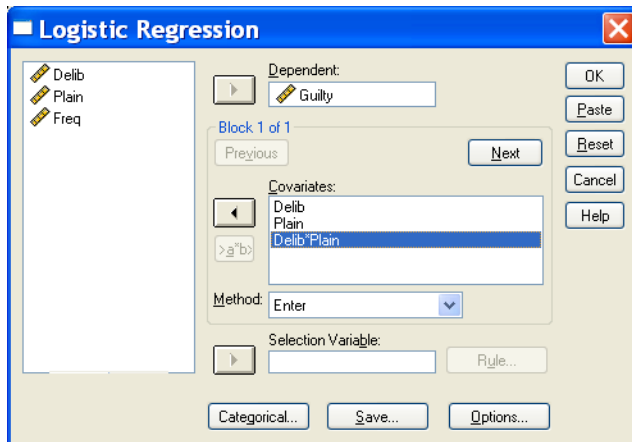
Current Status: Weight cases by Freq

OK  
Paste  
Reset  
Cancel  
Help

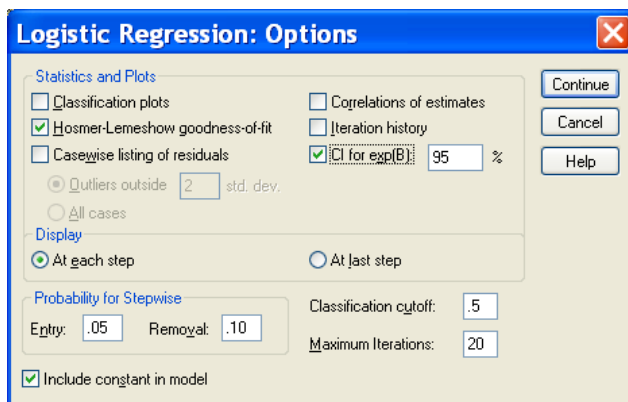
Analyze, Regression, Binary Logistic. Slide Guilty into the Dependent box and Delib and Plain into the Covariates box. Highlight both Delib and Plain in the pane on the left and then click the >a\*b> box.



This creates the interaction term. It could also be created by simply creating a new variable,  $\text{Interaction} = \text{Delib} * \text{Plain}$ .



Under Options, ask for the Hosmer-Lemeshow test and confidence intervals on the odds ratios.



You will find that the odds ratios are .338 for Delib, 3.134 for Plain, and 0.030 for the interaction.

Variables in the Equation

		Wald	df	Sig.	Exp(B)	95.0% C.I. for EXP(B)	
						Lower	Upper
Step 1 <sup>a</sup>	Delib	3.697	1	.054	.338	.112	1.021
	Plain	4.204	1	.040	3.134	1.052	9.339
	Delib by Plain	8.075	1	.004	.030	.003	.338
	Constant	.037	1	.847	1.077		

a. Variable(s) entered on step 1: Delib, Plain, Delib \* Plain .

Those who deliberated were less likely to suggest a guilty verdict (15%) than those who did not deliberate (66%), but this (partial) effect fell just short of statistical significance in the logistic regression (but a 2 x 2 chi-square would show it to be significant).

Plain defendants were significantly more likely (43%) than physically attractive defendants (39%) to be found guilty. This effect would fall well short of statistical significance with a 2 x 2 chi-square.

We should not pay much attention to the main effects, given that the interaction is powerful.

The interaction odds ratio can be simply computed, by hand, from the cell frequencies.

- For those who did deliberate, the odds of a guilty verdict are 1/29 when the defendant was plain and 8/22 when she was attractive, yielding a conditional odds ratio of 0.09483.

Plain \* Guilty Crosstabulation<sup>a</sup>

			Guilty		Total
			No	Yes	
Plain	Attractive	Count	22	8	30
		% within Plain	73.3%	26.7%	100.0%
Plain	Plain	Count	29	1	30
		% within Plain	96.7%	3.3%	100.0%
Total		Count	51	9	60
		% within Plain	85.0%	15.0%	100.0%

a. Delib = Yes

- For those who did not deliberate, the odds of a guilty verdict are 27/8 when the defendant was plain and 14/13 when she was attractive, yielding a conditional odds ratio of 3.1339.

Plain \* Guilty Crosstabulation

			Guilty		Total
			No	Yes	
Plain	Attractive	Count	13	14	27
		% within Plain	48.1%	51.9%	100.0%
Plain	Plain	Count	8	27	35
		% within Plain	22.9%	77.1%	100.0%
Total		Count	21	41	62
		% within Plain	33.9%	66.1%	100.0%

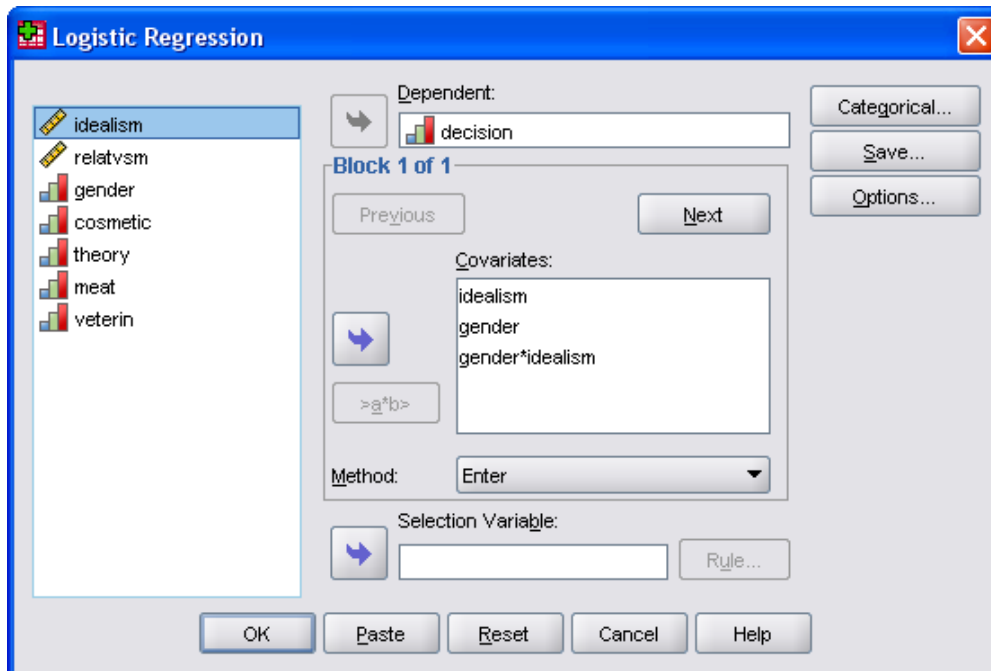
a. Delib = No

- The interaction odds ratio is simply the ratio of these conditional odds ratios – that is,  $.09483/3.1339 = 0.030$ .

Follow-up analysis shows that among those who did not deliberate the plain defendant was found guilty significantly more often than the attractive defendant,  $\chi^2(1, N = 62) = 4.353, p = .037$ , but among those who did deliberate the attractive defendant was found guilty significantly more often than the plain defendant,  $\chi^2(1, N = 60) = 6.405, p = .011$ .

### Interaction Between a Dichotomous Predictor and a Continuous Predictor

Suppose that I had some reason to suspect that the effect of idealism differed between men and women. I can create the interaction term just as shown above.





Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 <sup>a</sup>						
idealism	-.773	.145	28.572	1	.000	.461
gender	-.530	1.441	.135	1	.713	.589
gender by idealism	.268	.223	1.439	1	.230	1.307
Constant	4.107	.921	19.903	1	.000	60.747

a. Variable(s) entered on step 1: idealism, gender, gender \* idealism .

As you can see, the interaction falls short of significance.

## Partially Standardized B Weights and Odds Ratios

The value of a predictor's B (and the associated odds ratio) is highly dependent on the unit of measure. Suppose I am predicting whether or not an archer hits the target. One predictor is distance to the target. Another is how much training the archer has had. Suppose I measure distance in inches and training in years. I would not expect much of an increase in the logit when decreasing distance by an inch, but I would expect a considerable increase when increasing training by a year. Suppose I measured distance in miles and training in seconds. Now I would expect a large B for distance and a small B for training. For purposes of making comparisons between the predictors, it may be helpful to standardize the B weights.

Suppose that a third predictor is the archer's score on a survey of political conservatism and that a photo of Karl Marx appears on the target. The unit of measure here is not intrinsically meaningful – how much is a one point change in score on this survey. Here too it may be helpful to standardize the predictors. Menard (*The American Statistician*, 2004, 58, 218-223) discussed several ways to standardize B weights. I favor simply standardizing the predictor, which can be simply accomplished by converting the predictor scores to z scores or by multiplying the unstandardized B weight by the predictor's standard deviation. While one could also standardize the dichotomous outcome variable (group membership), I prefer to leave that unstandardized.

In research here at ECU, Cathy Hall gathered data that is useful in predicting who will be retained in our engineering program. Among the predictor variables are high school GPA, score on the quantitative section of the SAT, and one of the Big Five personality measures, [openness to experience](#). Here are the results of a binary logistic regression predicting retention from high school GPA, quantitative SAT, and openness (you can find more detail [here](#)).

Predictor	Unstandardized		Standardized	
	B	Odds Ratio	B	Odds Ratio
HS-GPA	1.296	3.656	0.510	1.665
SAT-Q	0.006	1.006	0.440	1.553
Openness	0.100	1.105	0.435	1.545

The novice might look at the unstandardized statistics and conclude that SAT-Q and openness to experience are of little utility, but the standardized coefficients show that not to be true. The three predictors differ little in their unique contributions to predicting retention in the engineering program.

## Practice Your Newly Learned Skills

Now that you know how to do a logistic regression, you should practice those skills. I have presented below three exercises designed to give you a little practice.

### Exercise 1: What is Beautiful is Good, and Vice Versa

Castellow, Wuensch, and Moore (1990, *Journal of Social Behavior and Personality*, 5, 547-562) found that physically attractive litigants are favored by jurors hearing a civil case involving alleged sexual harassment (we manipulated physical attractiveness by controlling the photos of the litigants seen by the mock jurors). Guilty verdicts were more likely when the male defendant was physically unattractive and when the female plaintiff was physically attractive. We also found that jurors rated the physically attractive litigants as more socially desirable than the physically unattractive litigants -- that is, more warm, sincere, intelligent, kind, and so on. Perhaps the jurors treated the physically attractive litigants better because they assumed that physically attractive people are more socially desirable (kinder, more sincere, etc.).

Our next research project (Egbert, Moore, Wuensch, & Castellow, 1992, *Journal of Social Behavior and Personality*, 7, 569-579) involved our manipulating (via character witness testimony) the litigants' social desirability but providing mock jurors with no information on physical attractiveness. The jurors treated litigants described as socially desirable more favorably than they treated those described as socially undesirable. However, these jurors also rated the socially desirable litigants as more physically attractive than the socially undesirable litigants, despite having never seen them! Might our jurors be treating the socially desirable litigants more favorably because they assume that socially desirable people are more physically attractive than are socially undesirable people?

We next conducted research in which we manipulated both the physical attractiveness and the social desirability of the litigants (Moore, Wuensch, Hedges, &

Castellow, 1994, *Journal of Social Behavior and Personality*, 9, 715-730). Data from selected variables from this research project are in the SPSS data file found at <http://core.ecu.edu/psyc/wuenschk/SPSS/Jury94.sav>. Please download that file now.

You should use SPSS to predict verdict from all of the other variables. The variables in the file are as follows:

- VERDICT -- whether the mock juror recommended a not guilty (0) or a guilty (1) verdict -- that is, finding in favor of the defendant (0) or the plaintiff (1)
- ATTRACT -- whether the photos of the defendant were physically unattractive (0) or physically attractive (1)
- GENDER -- whether the mock juror was female (0) or male (1)
- SOCIABLE -- the mock juror's rating of the sociability of the defendant, on a 9-point scale, with higher representing greater sociability
- WARMTH -- ratings of the defendant's warmth, 9-point scale
- KIND -- ratings of defendant's kindness
- SENSITIV -- ratings of defendant's sensitivity
- INTELLIG -- ratings of defendant's intelligence

You should also conduct bivariate analysis (Pearson Chi-Square and independent samples t-tests) to test the significance of the association between each predictor and the criterion variable (verdict). You will find that some of the predictors have significant zero-order associations with the criterion but are not significant in the full model logistic regression. Why is that?

You should find that the sociability predictor has an odds ratio that indicates that the odds of a guilty verdict increase as the rated sociability of the defendant increases -- but one would expect that greater sociability would be associated with a reduced probability of being found guilty, and the univariate analysis indicates exactly that (mean sociability was significantly higher with those who were found not guilty). How is it possible for our multivariate (partial) effect to be opposite in direction to that indicated by our univariate analysis? You may wish to consult the following documents to help understand this:

[Redundancy and Suppression](#)

[Simpson's Paradox](#)

## **Exercise 2: Predicting Whether or Not Sexual Harassment Will Be Reported**

Download the SPSS data file found at <http://core.ecu.edu/psyc/wuenschk/SPSS/Harass-Howell.sav>. This file was obtained from David Howell's site, <http://www.uvm.edu/~dhowell/StatPages/Methods/DataMethods5/Harass.dat>. I have added value labels to a couple of the variables. You should use SPSS to conduct a logistic regression predicting the variable "reported" from all of the other variables. Here is a brief description for each variable:

- REPORTED -- whether (1) or not (0) an incident of sexual harassment was reported

- AGE -- age of the victim
- MARSTAT -- marital status of the victim -- 1 = married, 2 = single
- FEMinist ideology -- the higher the score, the more feminist the victim
- OFFENSUV -- offensiveness of the harassment -- higher = more offensive

I suggest that you obtain, in addition to the multivariate analysis, some bivariate statistics, including independent samples t-tests, a Pearson chi-square contingency table analysis, and simple Pearson correlation coefficients for all pairs of variables.

### Exercise 3: Predicting Who Will Drop-Out of School

Download the SPSS data file found at <http://core.ecu.edu/psyc/wuenschk/SPSS/Dropout.sav>. I simulated these data based on the results of research by David Howell and H. R. Huessy (*Pediatrics*, 76, 185-190). You should use SPSS to predict the variable "dropout" from all of the other variables. Here is a brief description for each variable:

- DROPOUT -- whether the student dropped out of high school before graduating - 0 = No, 1 = Yes.
- ADDSC -- a measure of the extent to which each child had exhibited behaviors associated with attention deficit disorder. These data were collected while the children were in the 2<sup>nd</sup>, 4<sup>th</sup>, and 5<sup>th</sup> grades combined into one variable in the present data set.
- REPEAT -- did the child ever repeat a grade -- 0 = No, 1 = Yes.
- SOCPROB -- was the child considered to have had more than the usual number of social problems in the 9<sup>th</sup> grade -- 0 = No, 1 = Yes.

I suggest that you obtain, in addition to the multivariate analysis, some bivariate statistics, including independent samples t-tests, a Pearson chi-square contingency table analysis, and simple Pearson correlation coefficients for all pairs of variables. Imagine that you were actually going to use the results of your analysis to decide which children to select as "at risk" for dropping out before graduation. Your intention is, after identifying those children, to intervene in a way designed to make it less likely that they will drop out. What cutoff would you employ in your decision rule?

Copyright 2009, Karl L. Wuensch - All rights reserved.



- [Logistic Regression With SAS](#)
- [Why Not Let SPSS Do The Dummy Coding of Categorical Predictors?](#)
- [Statistics Lessons](#)
- [SPSS Lessons](#)
- [More Links](#)

- [Letters From Former Students](#) -- some continue to use my online lessons when they go on to doctoral programs