Logistic Regression Lab

Use the data **cardiorehab.sav** from the web site.

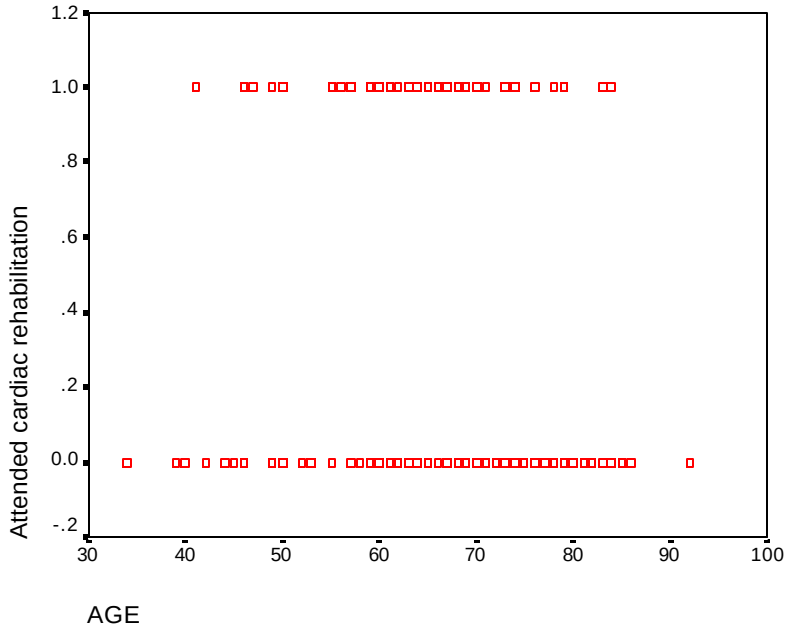
I. Simple Logistic Regression

Used when *Y* is a binary variable – only two outcomes, usually denoted with 0 (absent), 1 (present). The predictor variable can be continuous or categorical. Graphics are not helpful anymore.

According to Gallagher et al., cardiac rehabilitation programs offer "information, support, and monitoring for return to activities, symptom management, and risk factor modification." The researchers conducted a study to identify factors among women that are associated with participation in such programs. The data in Table 11.4.3 are the ages of 185 women discharged from a hospital in Australia who met eligibility criteria involving discharge for Myocardial infarction, artery bypass surgery, angioplasty, or stent. We wish to use these data to obtain information regarding the relationship between age (years) and participation in a cardiac rehabilitation program (ATT=1, if participated, and ATT=0, if not). We wish also to know if we may use the results of our analysis to predict the likelihood of participation by a woman if we know her age.

SOURCE: Robyn Gallagher, Sharon McKinley, Kathleen Dracup, "Predictor's of Women's Attendance at Cardiac Rehabilitation Programs," *Progress in Cardiovascular Nursing, 18* (2003), 121-126.

Here is a scatterplot of the outcome variable (**att**) vs. the predictor variable (**age**).



Not very exciting.

Model:

Click on **Analyze\Regression\Binary Logistic** and then click **att** into the **Dependent** box and then click **age** over to the **Covariate** box. Also click the **Save** button and put a check in the **Probabilities** box for the **Predicted Values** section. Then hit **Continue** and then **OK**.

Model Summary

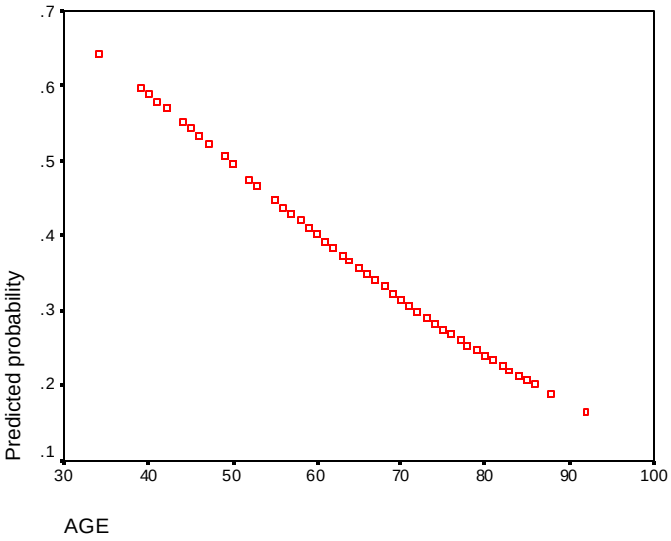
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	229.520	.037	.051

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	AGE	-.038	.015	6.710	1	.010	.963
	Constant	1.875	.981	3.653	1	.056	6.519

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

To look at our "logistic curve" of predicted probabilities, we click on **Graphs\Scatter\Simple\Define** and then place **pre_1** into the **Y-axis** box and then put **age** in the **X-axis** box.



Prediction: Predicted Probabilities have the form: $\frac{e^{\hat{b}_0+\hat{b}_1 x}}{1+e^{\hat{b}_0+\hat{b}_1 x}}$.

Estimate the probability of attendance for a woman age 50.

1. Calculate 1.875-.038(50)=_____

2. Calculate e^{ANS} =_____

3. Calculate $\frac{ANS}{1+ANS}$ =_____

II. Binary Predictor

We can also use binary variables or categorical variables to predict the binary variable outcome variable. For example, we could use employment status (0=Unemployed or Retired, 1=Employed) to predict attendance. In this case, the **Dependent** variable is again **att**, and the covariate variable is **employed**. Following the same steps as above, we get:

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	EMPLOYED	1.154	.448	6.633	1	.010	3.171
	Constant	-.818	.172	22.730	1	.000	.441

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

Calculate:

$P(Att = 1 | Employed = 1) =$

$P(Att = 1 | Employed = 0) =$

Estimate the prevalence ratio and interpret its meaning:

III. Quadratic Logistic Regression

First construct the variable **agesq** using **Transform\Compute**. Then click on **Analyze\Regression\Binary Logistic** and then click **att** into the **Dependent** box and then click **age** and **agesq** over to the **Covariates** box. (Make sure **Method** says **Enter**). Then hit **OK**.

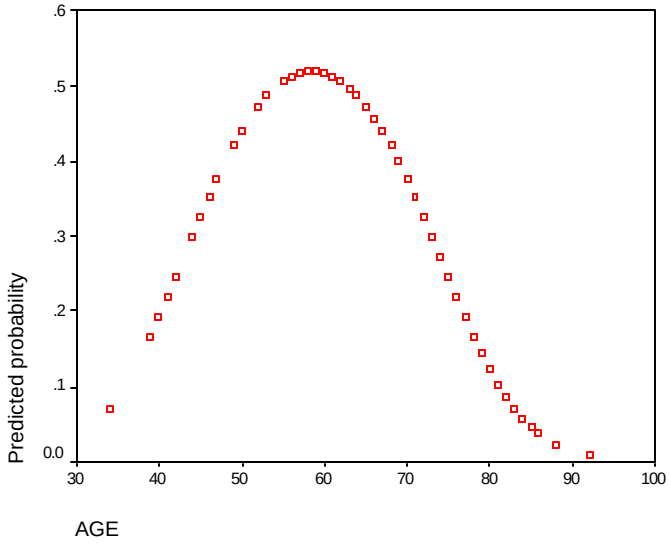
Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	216.192	.104	.144

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	AGE	.517	.173	8.954	1	.003	1.677
	AGESQ	-.004	.001	10.358	1	.001	.996
	Constant	-15.046	5.371	7.849	1	.005	.000

a. Variable(s) entered on step 1: AGE, AGESQ.



Prediction for woman age 50: