Binary Logistic Regression

The coefficients of the multiple regression model are estimated using sample data with k independent variables

Estimated (or predicted) value of Y

Estimated intercept

Estimated slope coefficients

$$\hat{Y}_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + \ldots + b_k X_{ki}$$

- Interpretation of the Slopes: (referred to as a Net Regression Coefficient)
  - b_1=The change in the mean of Y per unit change in X_1, taking into account the effect of X_2 (or net of X_2)
  - b_0 Y intercept. It is the same as simple regression.

Binary Logistic Regression

- Binary logistic regression is a type of regression analysis where the dependent variable is a dummy variable (coded 0, 1)
- Why not just use ordinary least squares?

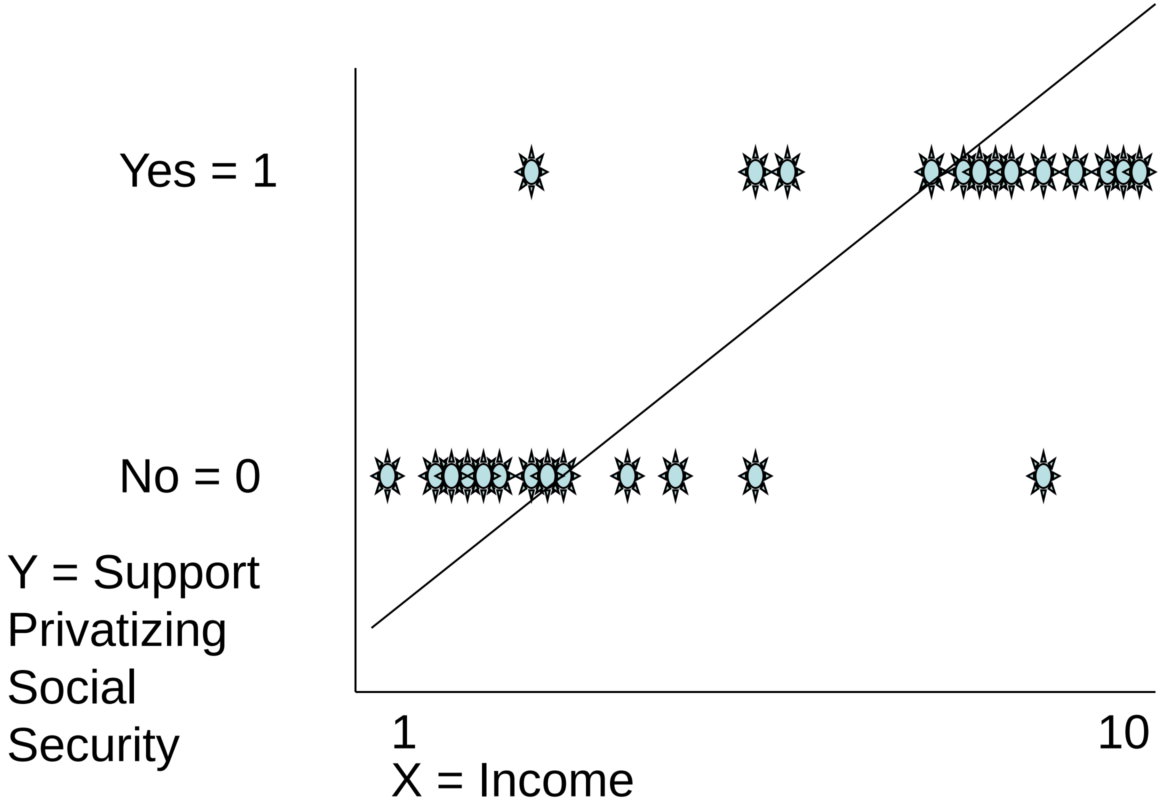
  $\hat{Y} = a + bx$

  – You would typically get the correct answers in terms of the sign and significance of coefficients

  – However, there are three problems

Binary Logistic Regression

OLS on a dichotomous dependent variable:



Binary Logistic Regression

– However, there are three problems

1. The error terms are heteroskedastic (variance of the dependent variable is different with different values of the independent variables
2. The error terms are not normally distributed
3. And ***most importantly***, for purpose of interpretation, the predicted probabilities can be greater than 1 or less than 0, which can be a problem for subsequent analysis.

Binary Logistic Regression

- The "logit" model solves these problems:
  - $\ln[p/(1-p)] = a + BX$

  or

  - $p/(1-p) = e^{a + BX}$
  - $p/(1-p) = e^{a} (e^{B})^{X}$

  Where:

  "ln" is the natural logarithm, $\log_{exp}$, where e=2.71828

  "p" is the probability that Y for cases equals 1, p (Y=1)

  "1-p" is the probability that Y for cases equals 0,
  1 – p(Y=1)

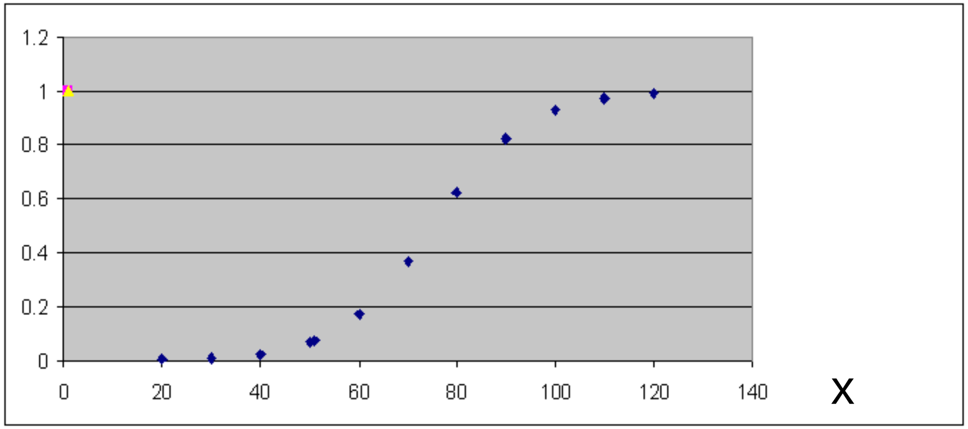
  "p/(1-p)" is the odds

  ln[p/1-p] is the log odds, or "logit"

Binary Logistic Regression

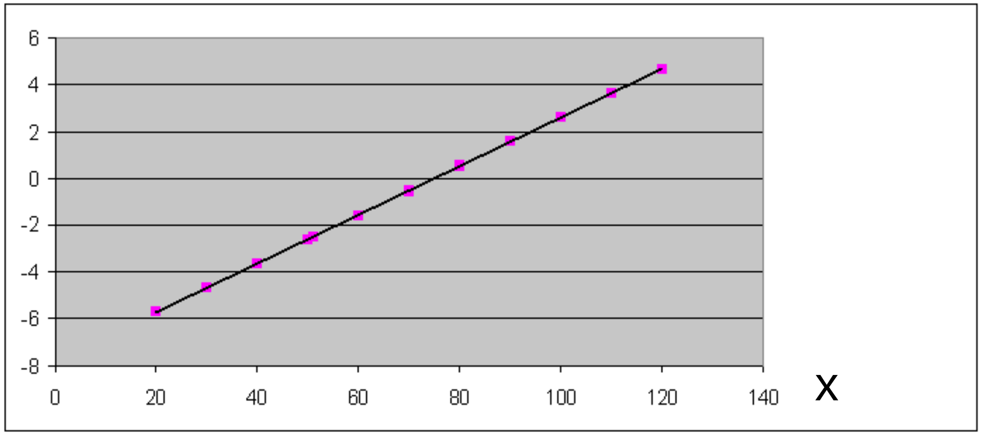
- Logistic Distribution

P (Y=1)

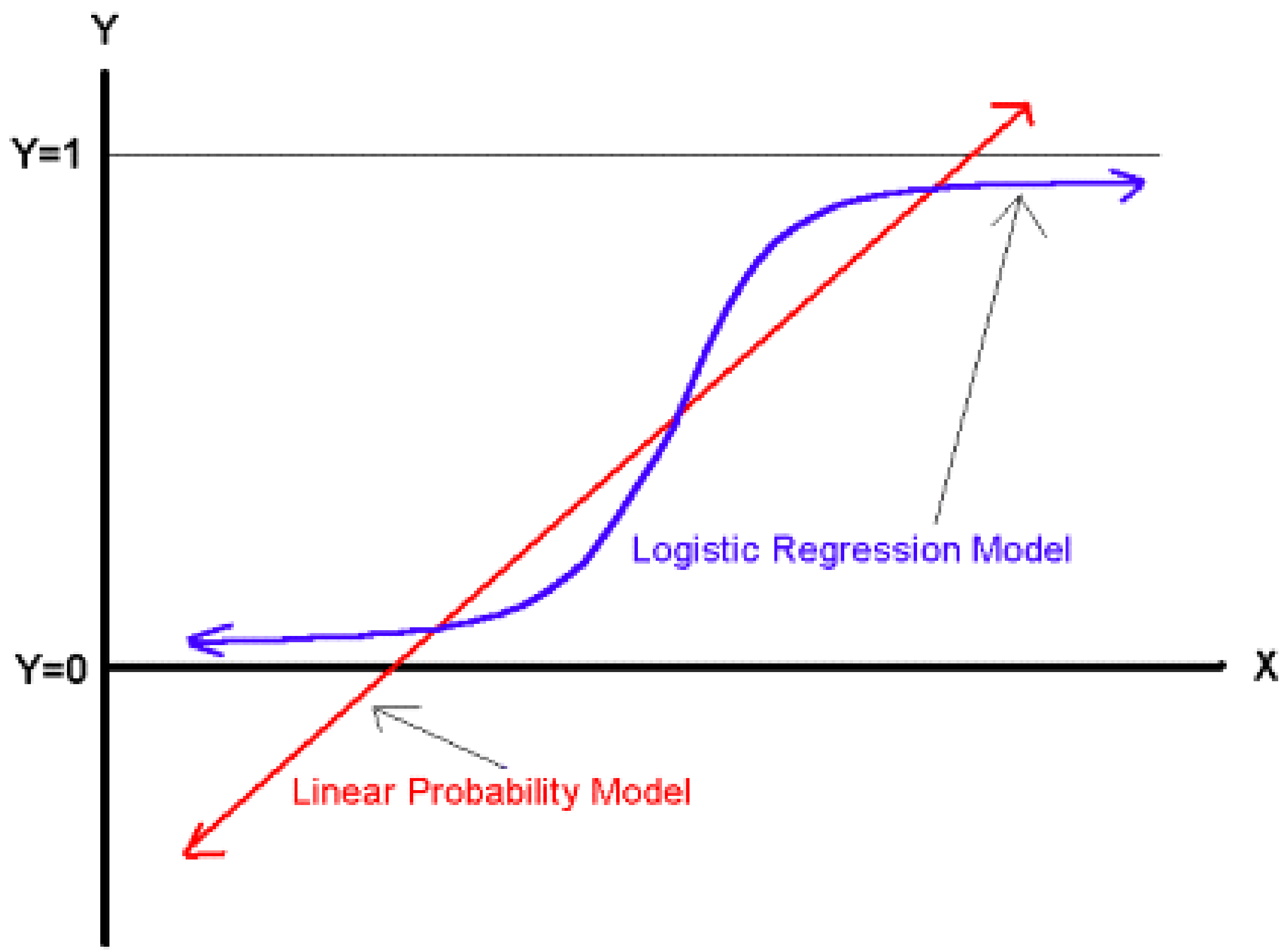


- Transformed, however, the “log odds” are linear.

ln[p/(1-p)]



Comparing the LP and Logit Models



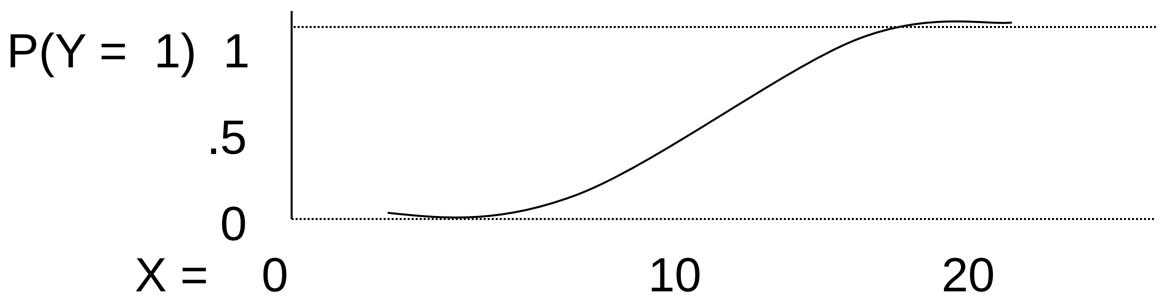
Binary Logistic Regression

- The logistic regression model is simply a non-linear transformation of the linear regression.
- The logistic distribution is an S-shaped distribution function (cumulative density function) which is similar to the standard normal distribution and constrains the estimated probabilities to lie between 0 and 1.

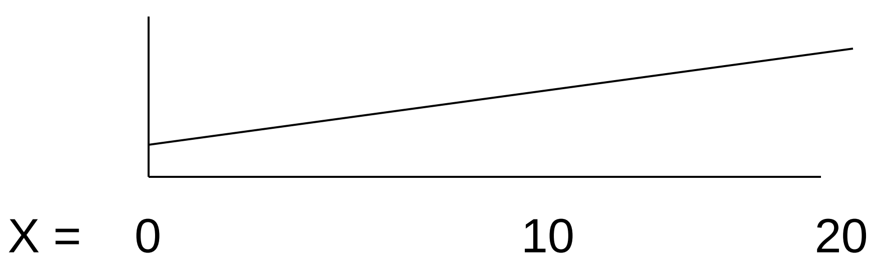
Binary Logistic Regression

- Logistic Distribution

With the logistic transformation, we're fitting the "model" to the data better.



- Transformed, however, the "log odds" are linear.



Binary Logistic Regression

- You're likely feeling overwhelmed, perhaps anxious about understanding this.
- Don't worry, coherence is gained when you see similarity to OLS regression:
  1. Model fit
  2. Interpreting coefficients
  3. Inferential statistics
  4. Predicting Y for values of the independent variables (the most difficult, but we'll make it easy)

Review & Summary

- In logistic regression, we predict Z, not p, because of Z's convenient mathematical properties

- Z is a linear function of the predictors, and we can translate that prediction into a probability.

Logistic regression predicts the natural logarithm of the odds

- The natural log of the odds is called the "logit" ="Z"

- Formula:

  $Z = \log (p/1\text{-}p) = B_0 + B_1.X_1 + B_2.X_2 + B_3.X_3 \ldots e$

- B’s in logistic regression are analogous to b’s in OLS

- B_1 is the average change in Z per one unit increase in X_1, controlling for the other predictors

- We calculate changes in the log odds of the dependent variable, not changes in the dependent variable (as in OLS).

Interpreting logistic regression results

- In SPSS output, look for:

  1) Model chi-square (equivalent to F)

  2) WALD statistics and “Sig.” for each B

  3) Logistic regression coefficients (B’s)

  4) Exp(B) = odds ratio

Interpreting logistic coefficients

- Identify which predictors are significant by looking at “Sig.”
- Look at the sign of B_1

  * If B_1 is positive, a unit change in x_1 is raising the odds of the event happening, after controlling for the other predictors

  * If B_1 is negative, the odds of the event decrease with a unit increase in x_1.

Interpreting the odds ratio

- Look at the column labeled Exp(B)
  - Exp(B) means "e to the power B" or e^B
  - Called the "odds ratio" (Gr. symbol: Ψ)
  - e is a mathematical constant used as the "base" for natural logarithms

- In logistic regression, e^B is the factor by which the odds change when X increases by one unit.

Interpreting the odds ratio

- New odds / Old odds = e^B = odds ratio
- e.g. if the odds-ratio for EDUC is 1.05, that means that for every year of education, the odds of the outcome (e.g. voting) increase by a factor of 1.05.
- Odds ratios > 1 indicate a positive relationship between IV and DV (event likely to occur)
- Odds ratios < 1 indicate a negative relationship between IV and DV (event less likely to occur)

Let’s come up with an example … run it … and interpret it …

Binary Logistic Regression

- A researcher is interested in the likelihood of gun ownership in the US, and what would predict that.
- She uses the GSS to test the following research hypotheses:

1. Men are more likely to own guns than are women
2. Older people are more likely to own guns
3. White people are more likely to own guns than are those of other races
4. More educated people are less likely to own guns

Binary Logistic Regression

- Variables are measured as such:

Dependent:

Havegun: no gun = 0, own gun(s) = 1

Independent:

1. Sex: men = 0, women = 1
2. Age: entered as number of years
3. White: all other races = 0, white =1
4. Education: entered as number of years

SPSS: Analyze → Regression → Binary Logistic

Enter your variables and for output below, under options, I checked “iteration history”

Binary Logistic Regression

SPSS Output: Some descriptive information first…

Logistic Regression

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	1325	65.5
	Missing Cases	698	34.5
	Total	2023	100.0
Unselected Cases		0	.0
Total		2023	100.0

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

Dependent Variable Encoding

Original Value	Internal Value
.00	0
1.00	1

Binary Logistic Regression

Goodness-of-fit statistics for new model come next…

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	180.810	4	.000
	Block	180.810	4	.000
	Model	180.810	4	.000

Test of new model vs. intercept-only model (the null model), based on difference of -2LL of each. The difference has a X^2 distribution. Is new -2LL significantly smaller?

$-2(\sum(Yi * \ln[P(Yi)] + (1-Yi) \ln[1-P(Yi)])$

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	1532.747[a]	.128	.176

These are attempts to replicate R^2 using information based on -2 log likelihood, (C&S cannot equal 1)

The -2LL number is "ungrounded," but it has a χ^2 distribution. Smaller is better. In a perfect model, -2 log likelihood would equal 0.

Assessment of new model's predictions

Classification Table[a]

Observed			Predicted: R own's a gun =1 .00	Predicted: R own's a gun =1 1.00	Percentage Correct
Step 1	R own's a gun =1	.00	740	123	85.7
		1.00	301	161	34.8
	Overall Percentage				68.0

a. The cut value is .500

Binary Logistic Regression

Goodness-of-fit statistics for new model come next…

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	180.810	4	.000
	Block	180.810	4	.000
	Model	180.810	4	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	1532.747[a]	.128	.176

17.6% of the variance in gun ownership is explained by gender, age, race, and education

Classification Table[a]

Observed			Predicted		
			R own's a gun =1		Percentage Correct
			.00	1.00	
Step 1	R own's a gun =1	.00	740	123	85.7
		1.00	301	161	34.8
	Overall Percentage				68.0

a. The cut value is .500

Remember When Assessing Predictors, The Odds Ratio or Exp(*b*)…

$$Exp(b) = \frac{\textit{Odds after a unit change in the predictor}}{\textit{Odds before a unit change in the predictor}}$$

- Indicates the change in odds resulting from a unit change in the predictor.
  - OR > 1: Predictor $\uparrow$, Probability of outcome occurring $\uparrow$.
  - OR < 1: Predictor $\uparrow$, Probability of outcome occurring $\downarrow$.

Binary Logistic Regression

Interpreting Coefficients…

$\ln[p/(1-p)] = a + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4$

Variables in the Equation

e^b

			B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a] X1	sexnew	b1	-.780	.124	39.624	1	.000	.458
X2	age	b2	.020	.004	32.650	1	.000	1.020
X3	White	b3	1.618	.197	67.534	1	.000	5.044
X4	educ	b4	-.023	.020	1.370	1	.242	.977
1	Constant	a	-2.246	.363	38.224	1	.000	.106

a. Variable(s) entered on step 1: sexnew, age, White, educ.

Which b's are significant?

Being male, getting older, and being white have a positive effect on likelihood of owning a gun. On the other hand, education does not affect owning a gun.

We'll discuss the Wald test in a moment…

Binary Logistic Regression

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	sexnew	-.780	.124	39.624	1	.000	.458
	age	.020	.004	32.650	1	.000	1.020
	White	1.618	.197	67.534	1	.000	5.044
	educ	-.023	.020	1.370	1	.242	.977
	Constant	-2.246	.363	38.224	1	.000	.106

a. Variable(s) entered on step 1: sexnew, age, White, educ.

Each coefficient increases the odds by a multiplicative amount, the amount is e^b. "Every unit increase in X increases the odds by e^b."

In the example above, e^b = Exp(B) in the last column.

New odds / Old odds = e^b = odds ratio

For Female: $e^{-.780}$ = .458 …females are less likely to own a gun by a factor of .458.

Age: $e^{.020}$=1.020 … for every year of age, the odds of owning a gun increases by a factor of 1.020.

White: $e^{1.618}$ = 5.044 …Whites are more likely to own a gun by a factor of 5.044.

Educ: $e^{-.023}$ = .977 …Not significant

Binary Logistic Regression

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	sexnew	-.780	.124	39.624	1	.000	.458
	age	.020	.004	32.650	1	.000	1.020
	White	1.618	.197	67.534	1	.000	5.044
	educ	-.023	.020	1.370	1	.242	.977
	Constant	-2.246	.363	38.224	1	.000	.106

a. Variable(s) entered on step 1: sexnew, age, White, educ.

Each coefficient increases the odds by a multiplicative amount, the amount is e^b. "Every unit increase in X increases the odds by e^b."

In the example above, e^b = Exp(B) in the last column.

For Sex: $e^{-.780} = .458$ … If you subtract 1 from this value, you get the proportion increase (or decrease) in the odds caused by being male, -.542. In percent terms, odds of owning a gun decrease 54.2% for women.

Age: $e^{.020}=1.020$ … A year increase in age increases the odds of owning a gun by 2%.

White: $e^{1.618} = 5.044$ …Being white increases the odd of owning a gun by 404%

Educ: $e^{-.023} = .977$ …Not significant

Here's another example … and another way to interpret the results

Equation for Step 1

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Intervention(1)	1.229	.400	9.447	1	.002	3.417	1.561	7.480
	Constant	-.288	.270	1.135	1	.287	.750		

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

$$P(Y) = \frac{1}{1 + e^{-(-.288 + 1.229(Intervention))}}$$

See p 288 for an Example of using equation to compute Odds ratio.

We can say that the odds of a patient who is treated being cured are 3.41 times higher than those of a patient who is not treated, with a 95% CI of 1.561 to 7.480.

The important thing about this confidence interval is that it doesn't cross 1 (both values are greater than 1). This is important because values greater than 1 mean that as the predictor variable(s) increase, so do the odds of (in this case) being cured. Values less than 1 mean the opposite: as the predictor increases, the odds of being cured decreases.

Output: Step 1

Model if Term Removed

Variable		Model Log Likelihood	Change in -2 Log Likelihood	df	Sig. of the Change
Step 1	Intervention	-77.042	9.926	1	.002

Removing Intervention from the model would have a significant effect on the predictive ability of the model, in other words, it would be very bad to remove it.

Variables not in the Equation

			Score	df	Sig.
Step 1	Variables	Duration	.002	1	.964
		Duration by Intervention (1)	.043	1	.835
	Overall Statistics		.063	2	.969

Binary Logistic Regression

Binary Logistic Regression

The test you choose depends on level of measurement:

Independent Variable	Dependent Variable	Test
Dichotomous	Interval-Ratio Dichotomous	Independent Samples t-test
Nominal Dichotomous	Nominal Dichotomous	Cross Tabs
Nominal Dichotomous	Interval-Ratio Dichotomous	ANOVA
Interval-Ratio Dichotomous	Interval-Ratio	Bivariate Regression/Correlation
Two or More… Interval-Ratio Dichotomous	Interval-Ratio	Multiple Regression
Interval-Ratio Dichotomous	Dichotomous	Binary Logistic Regression

The Multiple Regression Model building

Idea: Examine the linear relationship between
1 dependent (Y) & 2 or more independent variables (X_i)

Multiple Regression Model with k Independent Variables:

Y-intercept

Population slopes

Random Error

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + \varepsilon$$

Binary Logistic Regression

- So what are natural logs and exponents?
  - If you didn't learn about them before this class, you obviously don't need to know it to get your degree … so don't worry about it.
  - But, for those who did learn it, ln(x)=y is the same as: $x=e^y$

READ THE ABOVE

LIKE THIS:

when you see "ln(x)" say "the value after the equal sign is the power to which I need to take e to get x"

so…

y is the power to which you would take e to get x

Binary Logistic Regression

- So… $\ln[p/(1-p)] = y$ is same as: $p/(1-p) = e^y$

READ THE ABOVE

LIKE THIS: when you see “$\ln[p/(1-P)]$” say “the value after the equal sign is the power to which I need to take e to get $p/(1-p)$”

so…

y is the power to which you would take e to get $p/(1-p)$

Binary Logistic Regression

- So… $\ln[p/(1-p)] = a + bX$ is same as: $p/(1-p) = e^{a + bX}$

READ THE ABOVE

LIKE THIS: when you see "ln[p/(1-P)]" say "the value after the equal sign is the power to which I need to take e to get p/(1-p)"

so…

a + bX is the power to which you would take e to get p/(1-p)

Binary Logistic Regression

- Recall that OLS Regression used an "ordinary least squares" formula to create the "linear model" we used.
- The Logistic Regression model will be constructed by **an iterative maximum likelihood procedure**.
- This is a computer dependent program that:
  1. starts with arbitrary values of the regression coefficients and constructs an initial model for predicting the observed data.
  2. then evaluates errors in such prediction and changes the regression coefficients so as make the likelihood of the observed data greater under the new model.
  3. repeats until the model converges, meaning the differences between the newest model and the previous model are trivial.
- The idea is that you "find and report as statistics" the parameters that are most likely to have produced your data.
- Model and inferential statistics will be different from OLS because of using this technique and because of the nature of the dependent variable. (Remember how we used chi-squared with classification?)

Binary Logistic Regression

- So in logistic regression, we will take the “twisted” concept of a transformed dependent variable equaling a line and manipulate the equation to “untwist” the interpretation.
- We will focus on:
  1. Model fit
  2. Interpreting coefficients
  3. Inferential statistics
  4. Predicting Y for values of the independent variables (the most difficult)—the prediction of probability, appropriately, will be an S-shape
- Let’s start with a research example and SPSS output…

Binary Logistic Regression

SPSS Output: Some descriptive information first…

Block 0: Beginning Block

Iteration History[a,b,c]

Iteration		-2 Log likelihood	Coefficients
			Constant
Step 0	1	1713.672	-.605
	2	1713.557	-.625
	3	1713.557	-.625

a. Constant is included in the model.

b. Initial -2 Log Likelihood: 1713.557

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

Maximum likelihood process stops at third iteration and yields an intercept (-.625) for a model with no predictors.

A measure of fit, -2 Log likelihood is generated. The equation producing this:

$-2(\sum(Y_i * \ln[P(Y_i)] + (1-Y_i) \ln[1-P(Y_i)]))$

This is simply the relationship between observed values for each case in your data and the model's prediction for each case. The "negative 2" makes this number distribute as a X^2 distribution.

In a perfect model, -2 log likelihood would equal 0. Therefore, lower numbers imply better model fit.

Binary Logistic Regression

Classification Table[a,b]

Observed			Predicted: R own's a gun =1 .00	Predicted: R own's a gun =1 1.00	Percentage Correct
Step 0	R own's a gun =1	.00	863	0	100.0
		1.00	462	0	.0
	Overall Percentage				65.1

a. Constant is included in the model.

b. The cut value is .500

Originally, the "best guess" for each person in the data set is 0, have no gun!

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-.625	.058	117.487	1	.000	.535

This is the model for log odds when any other potential variable equals zero (null model). It predicts : P = .651, like above. $1/1+e^a$ or $1/1+.535$

Real P = .349

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	sexnew	37.789	1	.000
		age	54.909	1	.000
		White	92.626	1	.000
		educ	.029	1	.866
	Overall Statistics		160.887	4	.000

If you added each…

Binary Logistic Regression

Next are iterations for our full model…

Block 1: Method = Enter

Iteration History[a,b,c,d]

Iteration		-2 Log likelihood	Coefficients				
			Constant	sexnew	age	White	educ
Step 1	1	1546.511	-1.626	-.629	.017	1.070	-.019
	2	1533.086	-2.140	-.765	.020	1.518	-.023
	3	1532.748	-2.242	-.780	.020	1.614	-.023
	4	1532.747	-2.246	-.780	.020	1.618	-.023
	5	1532.747	-2.246	-.780	.020	1.618	-.023

a. Method: Enter

b. Constant is included in the model.

c. Initial -2 Log Likelihood: 1713.557

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

Binary Logistic Regression

- $\ln[p/(1-p)] = a + b_1X_1 + \ldots + b_kX_k$, the power to which you need to take e to get:

$$\frac{P}{1-P} \quad \text{So...} \quad \frac{P}{1-P} = e^{a + b1X1+\ldots+bkXk}$$

- Ergo, plug in values of x to get the odds (= p/1-p).

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	sexnew	-.780	.124	39.624	1	.000	.458
	age	.020	.004	32.650	1	.000	1.020
	White	1.618	.197	67.534	1	.000	5.044
	educ	-.023	.020	1.370	1	.242	.977
	Constant	-2.246	.363	38.224	1	.000	.106

a. Variable(s) entered on step 1: sexnew, age, White, educ.

The coefficients can be manipulated as follows:

Odds = $p/(1-p) = e^{a+b1X1+b2X2+b3X3+b4X4} = e^{a}(e^{b1})^{X1}(e^{b2})^{X2}(e^{b3})^{X3}(e^{b4})^{X4}$

Odds = $p/(1-p) = e^{a+.898X1+.008X2+1.249X3-.056X4} = e^{-1.864}(e^{.898})^{X1}(e^{.008})^{X2}(e^{1.249})^{X3}(e^{-.056})^{X4}$

Binary Logistic Regression

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	sexnew	-.780	.124	39.624	1	.000	.458
	age	.020	.004	32.650	1	.000	1.020
	White	1.618	.197	67.534	1	.000	5.044
	educ	-.023	.020	1.370	1	.242	.977
	Constant	-2.246	.363	38.224	1	.000	.106

a. Variable(s) entered on step 1: sexnew, age, White, educ.

The coefficients can be manipulated as follows:

$$\text{Odds} = p/(1-p) = e^{a+b1X1+b2X2+b3X3+b4X4} = e^{a}(e^{b1})^{X1}(e^{b2})^{X2}(e^{b3})^{X3}(e^{b4})^{X4}$$

$$\text{Odds} = p/(1-p) = e^{-2.246-.780X1+.020X2+1.618X3-.023X4} = e^{-2.246}(e^{-.780})^{X1}(e^{.020})^{X2}(e^{1.618})^{X3}(e^{-.023})^{X4}$$

Each coefficient increases the odds by a multiplicative amount, the amount is e^b. “Every unit increase in X increases the odds by e^b.”

In the example above, e^b = Exp(B) in the last column.

Binary Logistic Regression

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	sexnew	-.780	.124	39.624	1	.000	.458
	age	.020	.004	32.650	1	.000	1.020
	White	1.618	.197	67.534	1	.000	5.044
	educ	-.023	.020	1.370	1	.242	.977
	Constant	-2.246	.363	38.224	1	.000	.106

a. Variable(s) entered on step 1: sexnew, age, White, educ.

Age: $e^{.020} = 1.020$ … A year increase in age increases the odds of owning a gun by 2%.

How would 10 years' increase in age affect the odds? Recall $(e^b)^X$ is the equation component for a variable. For 10 years, $(1.020)^{10} = 1.219$. The odds jump by 22% for ten years' increase in age.

Note: You'd have to know the current prediction level for the dependent variable to know if this percent change is actually making a big difference or not!

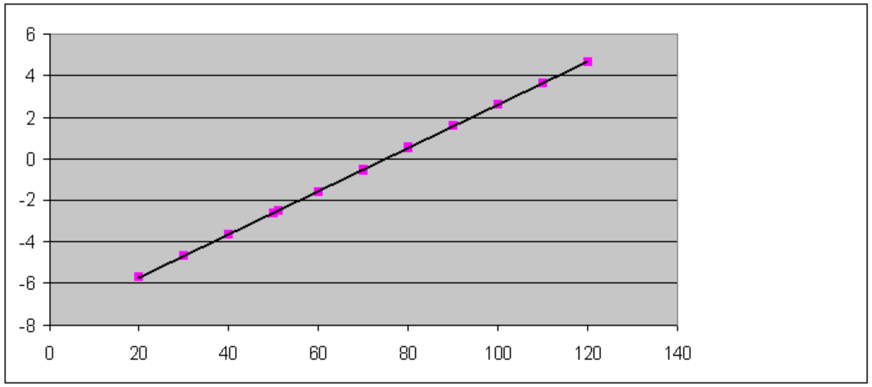
Binary Logistic Regression

Note: You'd have to know the current prediction level for the dependent variable to know if this percent change is actually making a big difference or not!

Recall that the logistic regression tells us two things at once.

- Transformed, the "log odds" are linear.

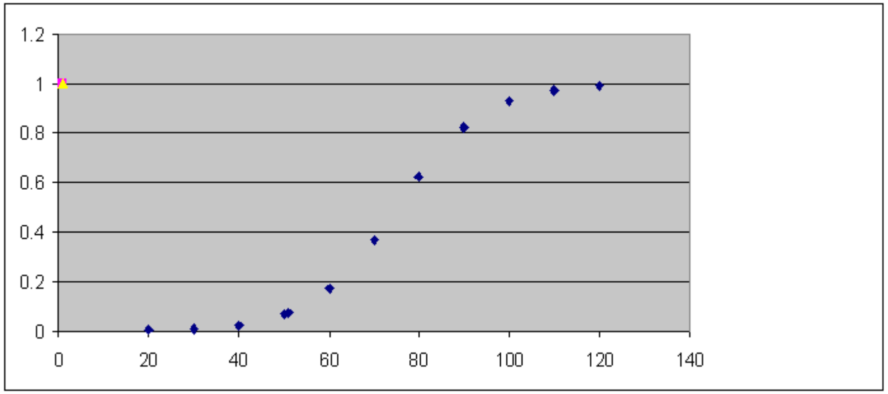
ln[p/(1-p)]



x

- Logistic Distribution

P (Y=1)



x

Binary Logistic Regression

We can also get p(y=1) for particular folks.

Odds = p/(1-p); p = P(Y=1)

With algebra…

Odds(1-p) = p … Odds-p(odds) = p …

Odds = p+p(odds) … Odds = p(1+odds)

… Odds/1+odds = p or

p = Odds/(1+odds)

Ln(odds) = a + bx and odds = $e^{a + bx}$

so…

$P = e^{a+bX}/(1+ e^{a+bX})$

We can therefore plug in numbers for X to get P

If a + BX = 0, then p = .5 As a + BX gets really big, p approaches 1

As a + BX gets really small, p approaches 0 (our model is an S curve)

Binary Logistic Regression

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Male	.898	.151	35.312	1	.000	2.454
	age	.008	.004	3.405	1	.065	1.008
	white	1.249	.233	28.658	1	.000	3.487
	educ	-.056	.023	5.658	1	.017	.946
	Constant	-1.864	.425	19.221	1	.000	.155

a. Variable(s) entered on step 1: Male, age, white, educ.

For our problem, $P = \dfrac{e^{-2.246-.780X1+.020X2+1.618X3-.023X4}}{1 + e^{-2.246-.780X1+.020X2+1.618X3-.023X4}}$

For, a man, 30, Latino, and 12 years of education, the P equals?

Let's solve for $e^{-2.246-.780X1+.020X2+1.618X3-.023X4} = e^{-2.246-.780(0)+.020(30)+1.618(0)-.023(12)}$

$e^{-2.246 - 0 + .6 + 0 - .276} = e^{-1.922} = 2.71828^{-1.922} = .146$

Therefore,

$P = \dfrac{.146}{1.146} = .127$ The probability that the 30 year-old, Latino with 12 years of education will own a gun is .127!!! Or you could say there is a 12.7% chance.

Binary Logistic Regression

Inferential statistics are as before:

- In model fit, if χ^2 test is significant, the expanded model (with your variables), improves prediction.
- This Chi-squared test tells us that as a set, the variables improve classification.

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	180.810	4	.000
	Block	180.810	4	.000
	Model	180.810	4	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	1532.747[a]	.128	.176

Classification Table[a]

Observed			Predicted		
			R own's a gun =1		Percentage Correct
			.00	1.00	
Step 1	R own's a gun =1	.00	740	123	85.7
		1.00	301	161	34.8
	Overall Percentage				68.0

a. The cut value is .500

Binary Logistic Regression

Inferential statistics are as before:

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	sexnew	-.780	.124	39.624	1	.000	.458
	age	.020	.004	32.650	1	.000	1.020
	White	1.618	.197	67.534	1	.000	5.044
	educ	-.023	.020	1.370	1	.242	.977
	Constant	-2.246	.363	38.224	1	.000	.106

a. Variable(s) entered on step 1: sexnew, age, White, educ.

- The significance of the coefficients is determined by a “wald test.” Wald is χ^2 with 1 df and equals a two-tailed t^2 with p-value exactly the same.

Binary Logistic Regression

So how would I do hypothesis testing? An Example:

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	sexnew	-.780	.124	39.624	1	.000	.458
	age	.020	.004	32.650	1	.000	1.020
	White	1.618	.197	67.534	1	.000	5.044
	educ	-.023	.020	1.370	1	.242	.977
	Constant	-2.246	.363	38.224	1	.000	.106

a. Variable(s) entered on step 1: sexnew, age, White, educ.

1. Significance test for α-level = .05
2. Critical $X^2_{df=1} = 3.84$
3. To find if there is a significant slope in the population,

   $H_o: \beta = 0$

   $H_a: \beta \neq 0$
4. Collect Data
5. Calculate Wald, like t (z): $t = \frac{b - \beta_o}{s.e.}$ (1.96 * 1.96 = 3.84)
6. Make decision about the null hypothesis
7. Find P-value

Reject the null for Male, age, and white. Fail to reject the null for education. There is a 24.2% chance that the sample came from a population where the education coefficient equals 0.