Logistic Regression

Outline

- Review of simple and multiple regression
- Simple Logistic Regression
  - The logistic function
  - Interpretation of coefficients
    - continuous predictor (X)
    - dichotomous categorical predictor (X)
    - categorical predictor with three or more levels (X)
- Multiple Logistic Regression
- Example with Weather dataset in Rattle()

Background

- Odds – like probability. Odds are usually written as “5 to 1 odds” which is equivalent to 1 out of five or .20 probability or 20% chance, etc.
  - The problem with probabilities is that they are non-linear
  - Going from .10 to .20 doubles the probability, but going from .80 to .90 barely increases the probability.

Background

- Odds ratio – the ratio of the odds over 1 – the odds. The probability of winning over the probability of losing. 5 to 1 odds equates to an odds ratio of .20/.80 = .25.

Background

- Logit – this is the natural log of an odds ratio; often called a log odds even though it really is a log odds ratio. The logit scale is linear and functions much like a z-score scale.

Background

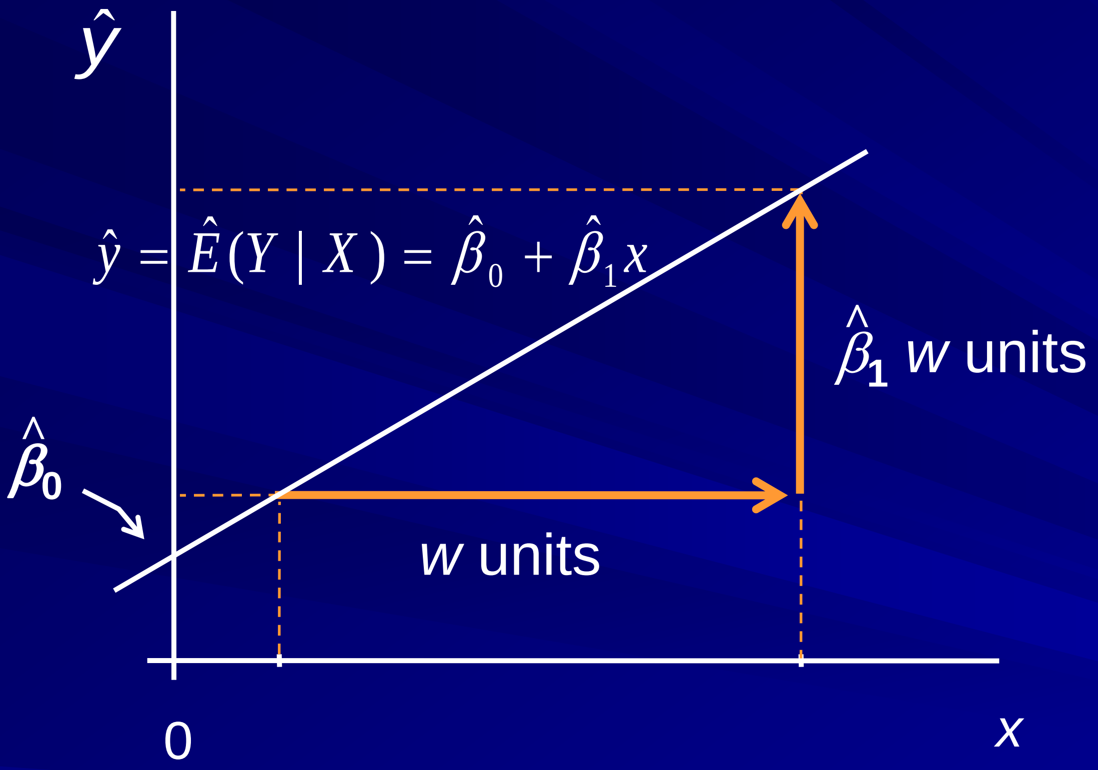
LOGITS ARE CONTINOUS, LIKE Z SCORES

p = 0.50, then logit = 0

p = 0.70, then logit = 0.84

p = 0.30, then logit = -0.84

Simple Linear Regression



$\hat{\beta}_0$ **= Estimated Intercept**

= $\hat{y}$-value at *x* = 0

Interpretable only if x = 0 is a value of particular interest.

$\hat{\beta}_1$ **= Estimated Slope**

= Change in $\hat{y}$ for every unit increase in *x*

= estimated change in the mean of Y for a unit change in X.

Always interpretable!

Multiple Linear Regression

We model the mean of a numeric response as linear combination of the predictors themselves or some functions based on the predictors, i.e.

$$E(Y|\mathbf{X}) = \underbrace{\beta_o + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p}_{\text{Here the terms in the model are the predictors}}$$

$$E(Y|\mathbf{X}) = \beta_o + \underbrace{\beta_1 f_1(\mathbf{X}) + \beta_2 f_2(\mathbf{X}) + \ldots + \beta_k f_k(\mathbf{X})}_{\text{Here the terms in the model are k different functions of the p predictors}}$$

Multiple Linear Regression

For the classic multiple regression model

$E(Y|\mathbf{X}) = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$

the regression coefficients (β_i) represent the estimated change in the mean of the response Y associated with a unit change in X_i while the other predictors are held constant.

They measure the association between Y and X_i **adjusted** for the other predictors in the model.

Logistic Regression

- Models relationship between set of variables X_i
  - dichotomous (yes/no, smoker/nonsmoker,…)
  - categorical (social class, race, ...)
  - continuous (age, weight, gestational age, ...)

and

  - dichotomous categorical response variable Y

  e.g. Success/Failure, Remission/No Remission Survived/Died, CHD/No CHD, Low Birth Weight/Normal Birth Weight, etc…

What is Logistic Regression?

- Form of regression that allows the prediction of discrete variables by a mix of continuous and discrete predictors.
- Addresses the same questions that discriminant function analysis and multiple regression do but with no distributional assumptions on the predictors (the predictors do not have to be normally distributed, linearly related or have equal variance in each group)

What is Logistic Regression?

- Logistic regression is often used because the relationship between the DV (a discrete variable) and a predictor is non-linear
  - Example from the text: the probability of heart disease changes very little with a ten-point difference among people with low-blood pressure, but a ten point change can mean a drastic change in the probability of heart disease in people with high blood-pressure.

Examples of Binary Outcomes

- Should a bank give a person a loan or not?
- Is an individual transaction fraudulent or not?
- What determines admittance into a school?
- Which people are more likely to vote against a new law?
- Which customers are more likely to buy a new product?

Data for Example: Customers' Subscription

- We have data on 1,000 random customers from a given city. We want to know what determines their decision to subscribe to a magazine.
- *Subscribe*: Indicates if a customer has subscribed to the magazine.
- *Age*: We will start by examining how age influences the likelihood of subscription.
- *Gender*: May also influence likelihood of subscription.

A linear model?

- Aside from being binary, there's really nothing special about our dependent variable (y).
- Its value is higher (from a 0 to a 1) if a customer subscribes, so whatever makes it higher increases the likelihood of subscription.
- We can then run:

$$subscribe = \beta_0 + \beta_1\, age + \varepsilon$$

Result of Linear Model

gretl: model 1

File Edit Tests Save Graphs Analysis LaTeX

```
Model 1: OLS, using observations 1-1000
Dependent variable: subscribe

              coefficient   std. error   t-ratio    p-value
  ---------------------------------------------------------
  const       -1.70073      0.0638035    -26.66    1.20e-118 ***
  age          0.0645433    0.00178736    36.11    2.52e-183 ***

Mean dependent var    0.573000   S.D. dependent var   0.494890
Sum squared resid     106.0736   S.E. of regression   0.326016
R-squared             0.566464   Adjusted R-squared   0.566030
F(1, 998)             1304.002   P-value(F)           2.5e-183
Log-likelihood       -297.1275   Akaike criterion     598.2550
Schwarz criterion     608.0705   Hannan-Quinn         601.9855
```

$$subscribe = -1.700 + 0.064 \; age$$

Interpreting the Result

- If our dependent variable is binary, then we want to see what makes it change from a 0 to 1.
- This can also be interpreted as what increases the likelihood of subscription, or $P(subscribe = 1)$, which we can also simply denote as p.
- The result can be read as:

$$P(subscribe = 1) = p = -1.700 + 0.064\, age$$

Interpreting the Result

- If our dependent variable is binary, then we want to see what makes it change from a 0 to 1.
- This can also be interpreted as what increases the likelihood of subscription, or $\mathrm{P}(subscribe = 1)$, which we can also simply denote as p.
- The result can be read as:

$$\mathrm{P}(subscribe = 1) = p = -1.700 + 0.064\ age$$

- Every additional year of *age* increases the probability of subscription by 6.4%.

Problems with the Linear Approach

- Probabilities are bounded whereby $0 \leq p \leq 1$.
- The range of *age* in the data is such that $20 \leq age \leq 55$.
- The probability that a 35 year-old person subscribes is:

$$p = -1.700 + 0.064 \times 35 = 0.54$$

- What about people with 25 and 45 years of age?

$$p = -1.700 + 0.064 \times 25 = -0.09$$

$$p = -1.700 + 0.064 \times 45 = 1.20$$

Fixing the prior Approach

- We need to somehow constrain p such that $0 \leq p \leq 1$.
- We know $p = f(\ age\ )$, but the linear function didn't work.
- What must $f(\bullet)$ satisfy to always produce reasonable forecasts?

Pause video and think, what must our probability forecasting function satisfy?

Two Steps!

1. It must always be positive (since $p \geq 0$)

$$p = \exp(\beta_0 + \beta_1\, age) = e^{\beta_0 + \beta_1\, age}$$

2. It must be less than 1 (since $p \leq 1$)

$$p = \frac{\exp(\beta_0 + \beta_1\, age)}{\exp(\beta_0 + \beta_1\, age) + 1} = \frac{e^{\beta_0 + \beta_1\, age}}{e^{\beta_0 + \beta_1\, age} + 1}$$

The Linear Thinking is not Completely Gone

- The previous expression (by doing some algebra) can be rewritten as:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1\, age$$

- Even though the probability of a customer subscribing (p) is not a linear function of age, the simple transformation is a linear function of age.

Result of Logistic Regression

gretl: model 2

File Edit Tests Save Graphs Analysis LaTeX

```
Model 2: Logit, using observations 1-1000
Dependent variable: subscribe
Standard errors based on Hessian

              coefficient   std. error      z        slope
  ---------------------------------------------------------
  const       -26.5240      1.82819      -14.51
  age           0.781053    0.0535623     14.58    0.154207

Mean dependent var    0.573000   S.D. dependent var    0.494890
McFadden R-squared    0.636613   Adjusted R-squared    0.633683
Log-likelihood      -247.9937    Akaike criterion    499.9873
Schwarz criterion   509.8028     Hannan-Quinn        503.7179

Number of cases 'correctly predicted' = 884 (88.4%)
f(beta'x) at mean of independent vars = 0.197
Likelihood ratio test: Chi-square(1) = 868.915 [0.0000]

           Predicted
              0     1
  Actual 0  350    77
         1   39   534
```

The Estimated Logistic Model

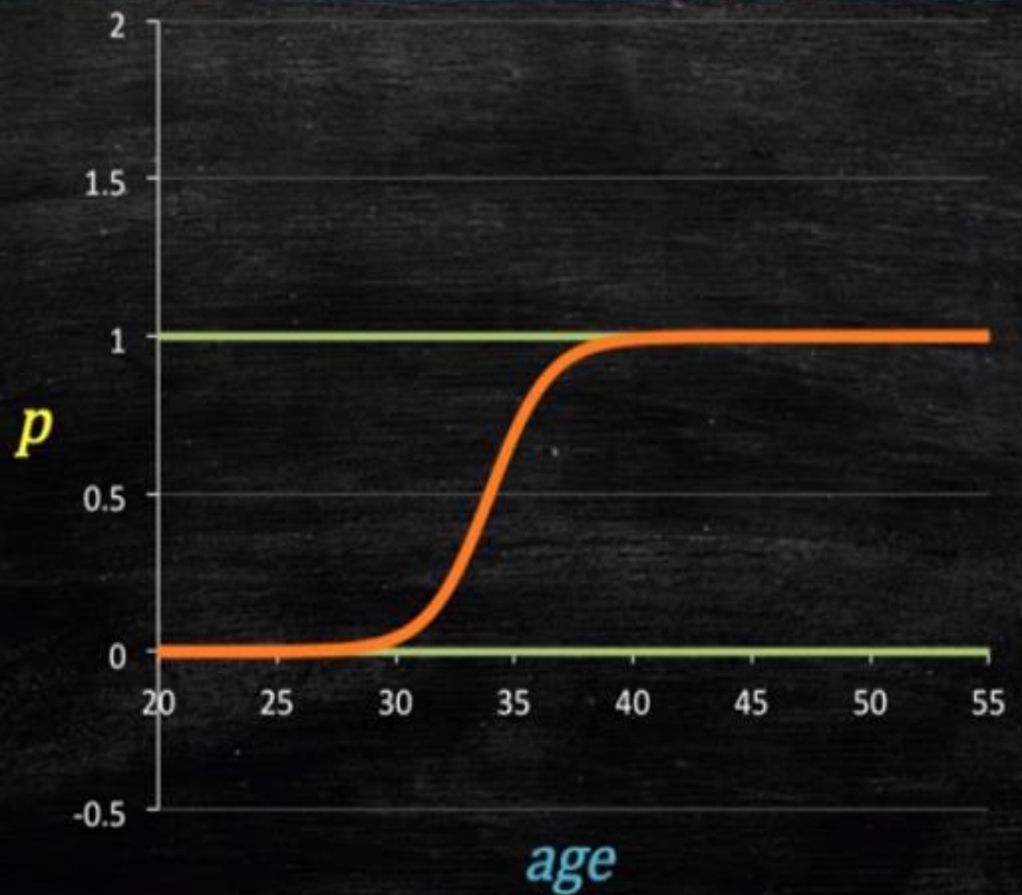
- The estimated model was:

$$\ln\left(\frac{p}{1-p}\right) = -26.52 \; + 0.78 \, age$$

- Or written in terms of the probability p we have:

$$p = \frac{\exp(-26.52 + 0.78 \, age)}{\exp(-26.52 + 0.78 \, age) + 1} = \frac{e^{-26.52 + 0.78 \, age}}{e^{-26.52 + 0.78 \, age} + 1}$$

Logistic Model Plot

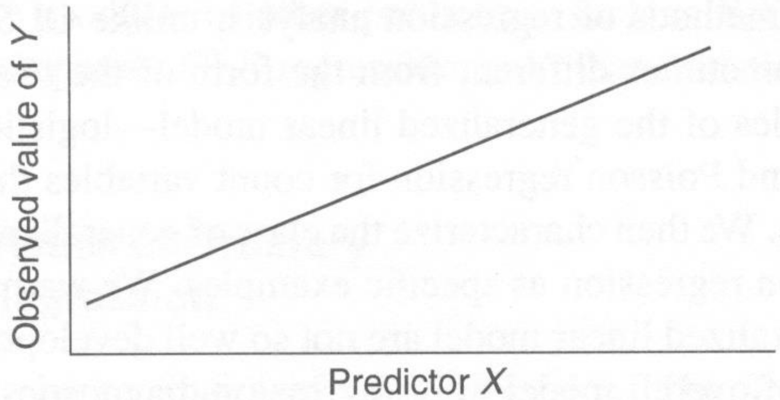


Logistic Regression Assumptions

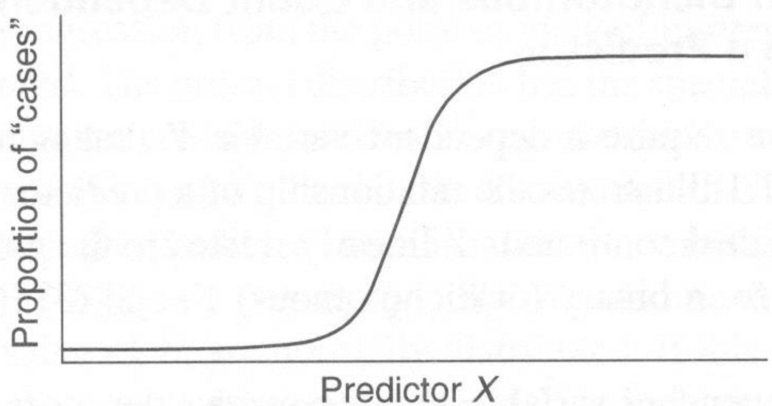
1. Binary logistic regression requires the dependent variable to be binary.
2. For a binary regression, the factor level 1 of the dependent variable should represent the desired outcome.
3. Only the meaningful variables should be included.
4. The independent variables should be independent of each other. That is, the model should have little or no multi-collinearity.
5. The independent variables are linearly related to the log odds.
6. Logistic regression requires quite large sample sizes.

The logistic function

(A) For a continuous outcome variable Y, the numerical value of Y at each value of X.

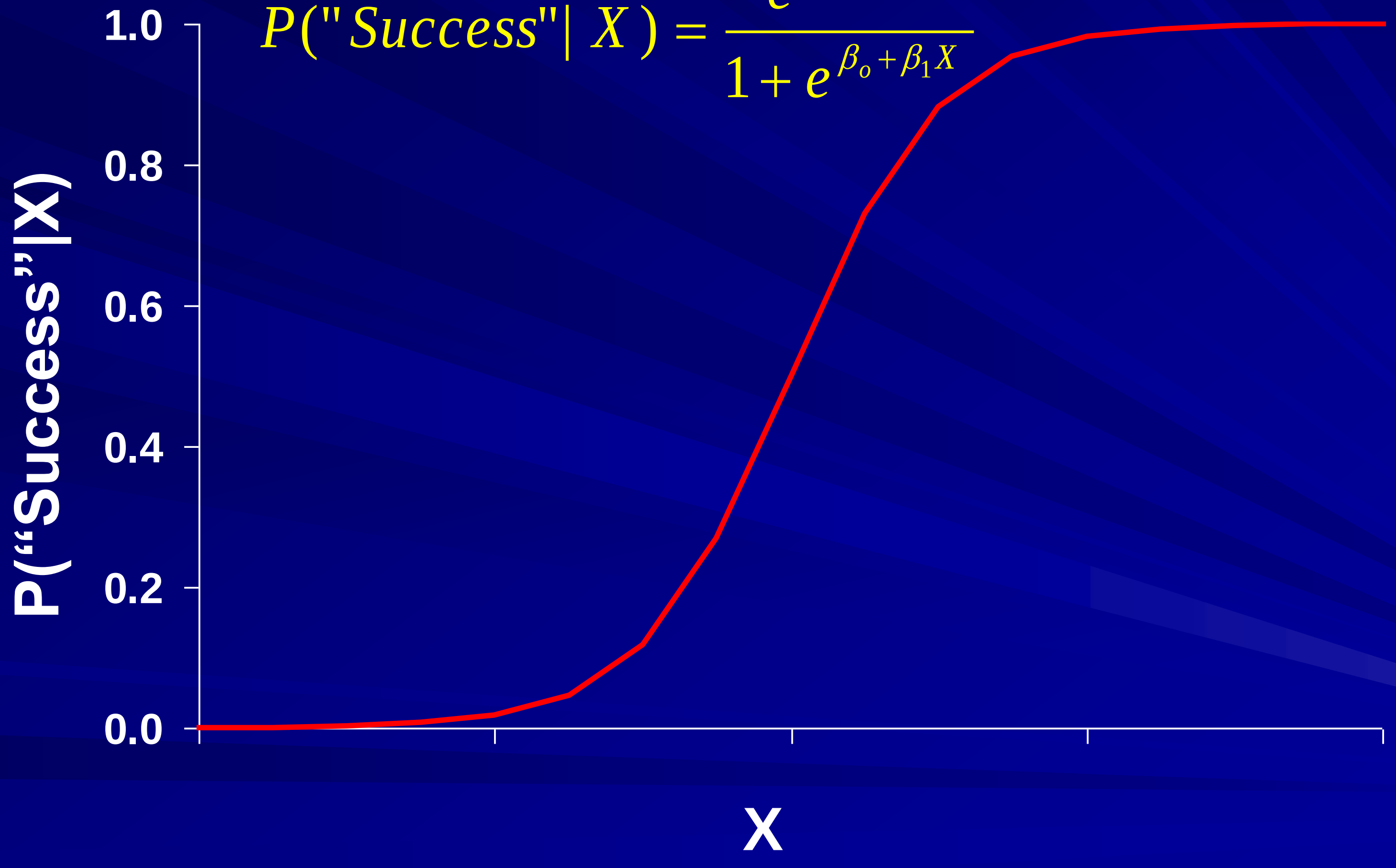


(B) For a binary outcome variable, the proportion of individuals who are "cases" (exhibit a particular outcome property) at each value of X.



Logistic Function

$$P(\text{"}Success\text{"} \mid X) = \frac{e^{\beta_o + \beta_1 X}}{1 + e^{\beta_o + \beta_1 X}}$$



The logistic function

$$\widehat{Y_i} = \frac{e^u}{1+e^u}$$

- Where Y-hat is the estimated probability that the ith case is in a category and u is the regular linear regression equation:

$$u = A + B_1 X_1 + B_2 X_2 + \cdots + B_K X_K$$

The logistic function

$$\hat{\pi}_i = \frac{e^{b_0 + b_1 X_1}}{1 + e^{b_0 + b_1 X_1}}$$

The logistic function

- Change in probability is not constant (linear) with constant changes in X
- This means that the probability of a success (Y = 1) given the predictor variable (X) is a non-linear function, specifically a logistic function

The Logit

- By algebraic manipulation, the logistic regression equation can be written in terms of an <u>odds ratio for success</u>:

$$\left[\frac{P(Y=1|X_i)}{(1-P(Y=1|X_i))}\right]=\left[\frac{\hat{\pi}}{(1-\hat{\pi})}\right]=\exp(b_0+b_1X_{1i})$$

The Logit

- Odds ratios range from 0 to positive infinity
- Odds ratio: P/Q is an odds ratio; less than 1 = less than .50 probability, greater than 1 means greater than .50 probability

The Logit

- Finally, taking the natural log of both sides, we can write the equation in terms of logits (log-odds):

$$\ln\left[\frac{P(Y=1|X)}{(1-P(Y=1|X))}\right]=\ln\left[\frac{\hat{\pi}}{(1-\hat{\pi})}\right]=b_0+b_1X_1$$

For a single predictor

The Logit

$$\ln\left[\frac{\hat{\pi}}{(1-\hat{\pi})}\right] = b_0 + b_1 X_1 + b_2 X_2 \ldots + b_k X_k$$

- For multiple predictors

The Logit

- Log-odds are a linear function of the predictors
- The regression coefficients go back to their old interpretation (kind of)
  - The expected value of the logit (log-odds) when X = 0
  - Called a 'logit difference'; The amount the logit (log-odds) changes, with a one unit change in X; the amount the logit changes in going from X to X + 1

Conversion

- EXP(logit) or = odds ratio
- Probability = odd ratio / (1 + odd ratio)

Logit is Directly Related to Odds

The logistic model can be written

$$\ln\left(\frac{P(Y \mid X)}{1-P(Y \mid X)}\right) = \ln\left(\frac{P}{1-P}\right) = \beta_o + \beta_1 X$$

This implies that the odds for success can be expressed as

$$\frac{P}{1-P} = e^{\beta_o + \beta_1 X}$$

This relationship is the key to interpreting the coefficients in a logistic regression model !!

Dichotomous Predictor (+1/-1 coding)

Consider a dichotomous predictor (X) which represents the presence of risk (1 = present)

	Risk Factor (X)	
Disease (Y)	**Present** (X = 1)	**Absent** (X = -1)
Yes (Y = 1)	$P(Y=1 \mid X=1)$	$P(Y=1 \mid X=-1)$
No (Y = 0)	$1-P(Y=1 \mid X=1)$	$1-P(Y=1 \mid X=-1)$

$$\frac{P}{1-P} = e^{\beta_o + \beta_1 X} \begin{cases} \text{Odds for Disease with Risk Present} = \dfrac{\mathrm{P}(\mathrm{Y}=1 \mid \mathrm{X}=1)}{1-\mathrm{P}(\mathrm{Y}=1 \mid \mathrm{X}=1)} = e^{\beta_o + \beta_1} \\ \text{Odds for Disease with Risk Absent} = \dfrac{\mathrm{P}(\mathrm{Y}=1 \mid \mathrm{X}=-1)}{1-\mathrm{P}(\mathrm{Y}=1 \mid \mathrm{X}=-1)} = e^{\beta_o - \beta_1} \end{cases}$$

Therefore the odds ratio (OR) $= \dfrac{\text{Odds for Disease with Risk Present}}{\text{Odds for Disease with Risk Absent}} = \dfrac{e^{\beta_o + \beta_1}}{e^{\beta_o - \beta_1}} = e^{2\beta_1}$

Dichotomous Predictor

- Therefore, for the odds ratio associated with risk presence we have $OR = e^{2\beta_1}$

- Taking the natural logarithm we have

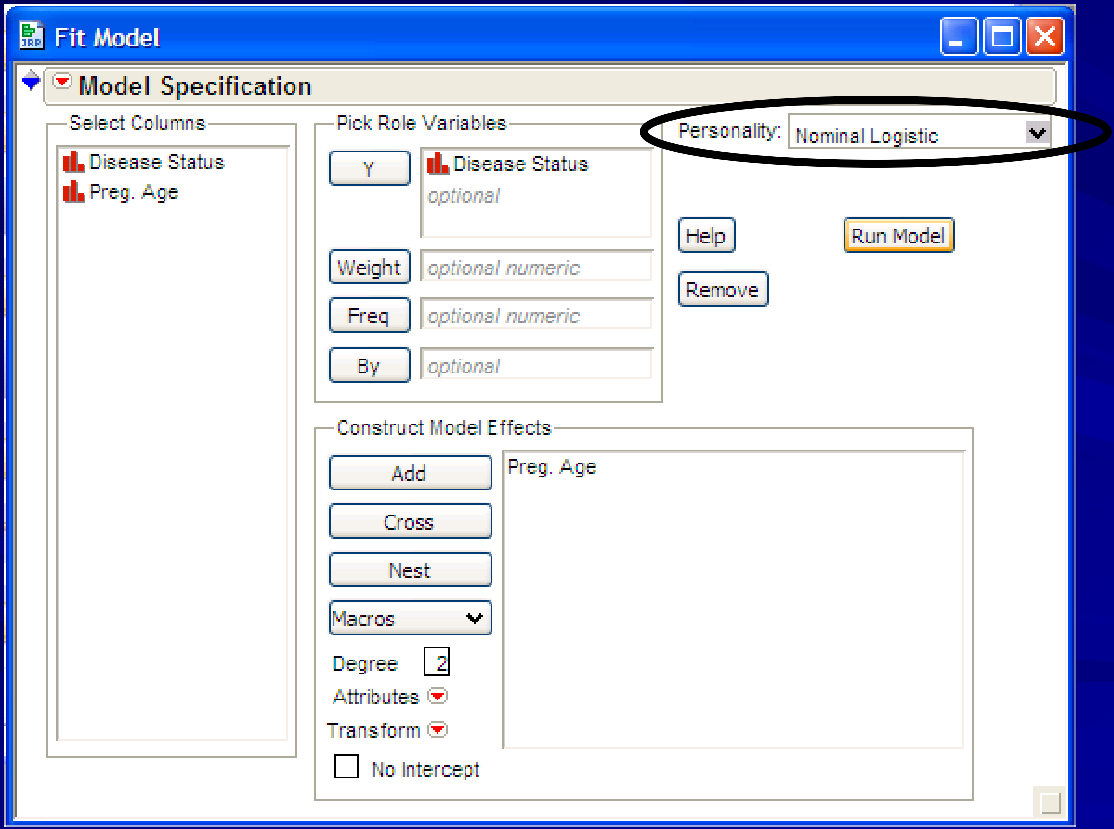
$$\ln(OR) = 2\beta_1$$

thus twice the estimated regression coefficient associated with a +1 / -1 coded dichotomous predictor is the natural log of the OR associated with risk presence!!!

Example: Age at 1st Pregnancy and Cervical Cancer

Use Fit Model   Y = Disease Status

X = Risk Factor Status



When the response Y is a dichotomous categorical variable the Personality box will automatically change to Nominal Logistic, i.e. Logistic Regression will be used.

Remember when a dichotomous categorical predictor is used JMP uses +1/-1 coding. If you want you can code them as 0-1 and treat is as numeric.

Example: Age at 1st Pregnancy and Cervical Cancer

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	-2.1829122	0.2123468	105.68	<.0001*
Preg. Age[<= 25]	0.60737587	0.2123468	8.18	0.0042*

For log odds of Cervical/Control

$$\hat{\beta}_o = -2.183$$
$$\hat{\beta}_1 = 0.607$$

Thus the estimated odds ratio is

$$\ln(OR) = 2\hat{\beta}_1 = 2(.607) = 1.214$$

$$OR = e^{1.214} = 3.37$$

Women whose first pregnancy is at or before age 25 have 3.37 times the odds for developing cervical cancer than women whose 1st pregnancy occurs after age 25.

Example: Age at 1st Pregnancy and Cervical Cancer

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	-2.1829122	0.2123468	105.68	<.0001*
Preg. Age[<= 25]	0.60737587	0.2123468	8.18	0.0042*

For log odds of Cervical/Control

$$\hat{\beta}_o = -2.183$$
$$\hat{\beta}_1 = 0.607$$

Thus the estimated odds ratio is

Odds Ratios

For Disease Status odds of Cervical versus Control

Odds Ratios for Preg. Age

Level1	/Level2	Odds Ratio	Reciprocal
> 25	<= 25	0.2967837	3.3694575

Risk Present → Odds Ratio for disease associated with risk presence

Example: Smoking and Low Birth Weight

Use Fit Model Y = Low Birth Weight (**L**ow, **N**orm)

X = Smoking Status (**C**ig, **N**oCig)

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	-2.0608189	0.0127482	26133	0.0000*
Smoking Status[Cig]	0.33493469	0.0127482	690.28	<.0001*

For log odds of Low/Norm

$$\hat{\beta}_1 = .335$$

$$OR = e^{2\hat{\beta}_1} = e^{.670} = 1.954$$

Odds Ratios

For Low Birth odds of Low versus Norm

Odds Ratios for Smoking Status

Level1	/Level2	Odds Ratio	Reciprocal
NoCig	Cig	0.5117754	1.9539821

We estimate that women who smoker during pregnancy have 1.95 times higher odds for having a child with low birth weight than women who do not smoke cigarettes during pregnancy.

Example: Smoking and Low Birth Weight

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	-2.0608189	0.0127482	26133	0.0000*
Smoking Status[Cig]	0.33493469	0.0127482	690.28	<.0001*

For log odds of Low/Norm

$$\hat{\beta}_1 = .335$$

$$OR = e^{2\hat{\beta}_1} = e^{.670} = 1.954$$

Find a 95% CI for OR

1st Find a 95% CI for β_1

$$\hat{\beta}_1 \pm 1.96 SE(\hat{\beta}_1) = .335 \pm 1.96 \cdot (.013) = .335 \pm .025 = \underbrace{(.310, .360)}_{(LCL,UCL)}$$

2nd Compute CI for OR = (e^{2LCL}, e^{2UCL})

$$(e^{2\times 310}, e^{2\times .360}) = (1.86\,,\,2.05)$$

We estimate that the odds for having a low birth weight infant are between 1.86 and 2.05 times higher for smokers than non-smokers, with 95% confidence.

Example: Smoking and Low Birth Weight

We might want to adjust for other potential confounding factors in our analysis of the risk associated with smoking during pregnancy. This is accomplished by simply adding these covariates to our model.

Multiple Logistic Regression Model

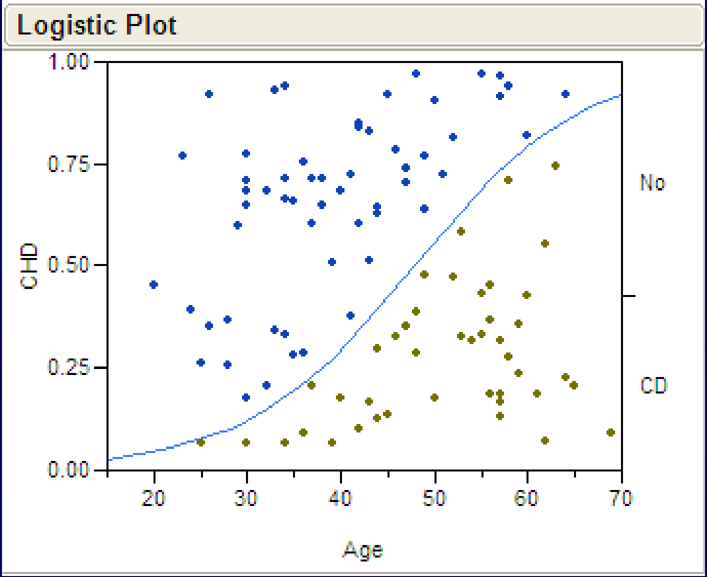
$$\ln\left(\frac{p}{1-p}\right) = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$$

Before looking at some multiple logistic regression examples we need to look at how continuous predictors and categorical variables with 3 or levels are handled in these models and how associated OR's are calculated.

Example 2: Signs of CD and Age

Fit Model Y = CD (CD if signs present, No otherwise)

X = Age (years)



Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	-5.309453	1.1336546	21.94	<.0001*
Age	0.11092114	0.0240598	21.25	<.0001*

For log odds of CD/No

$$\ln\left(\frac{p}{1-p}\right) = -5.31 + .111 \cdot Age$$

$$\text{Odds} = \frac{p}{1-p} = e^{-5.31+.111Age}$$

Consider the risk associated with a **c** year increase in age.

$$\text{Odds Ratio (OR)} = \frac{\text{Odds for Age = x + c}}{\text{Odds for Age = x}} = \frac{e^{\beta_o + \beta_1 (x+c)}}{e^{\beta_o + \beta_1 x}} = e^{c\beta_1}$$

Example 2: Signs of CD and Age

For example consider a 10 year increase in age, find the associated OR for showing signs of CD, i.e. $c = 10$

$OR = e^{c\beta} = e^{10*.111} = 3.03$

Thus we estimate that the odds for exhibiting signs of CD increase threefold for each 10 years of age. Similar calculations could be done for other increments as well.

<u>For example for a *c = 1* year increase</u>

$OR = e^{\beta} = e^{.111} = 1.18$ or an 18% increase in odds per year

Example 2: Signs of CD and Age

- Can we assume that the increase in risk associated with a c unit increase is constant throughout one’s life?
- Is the increase going from 20 → 30 years of age the same as going from 50 → 60 years?
- If that assumption is not reasonable then one must be careful when discussing risk associated with a continuous predictor.

Example 3: Race and Low Birth Weight

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	-2.1979794	0.0165809	17572	0.0000*
Race[Black]	0.41029325	0.0190908	461.89	<.0001*
Race[Other]	-0.0890288	0.030963	8.27	0.0040*

For log odds of Low/Norm

$$Race[Black] = \begin{cases} +1 & \text{for race} = \text{black} \\ -1 & \text{for race} = \text{white} \end{cases}$$

$$Race[Other] = \begin{cases} +1 & \text{for race} = \text{other} \\ -1 & \text{for race} = \text{white} \end{cases}$$

Calculate the odds for low birth weight for each race (**L**ow, **N**orm)

White Infants (reference group, missing in parameters)

$$e^{-2.198+.410(-1)-.089(-1)} = e^{-2.198-.410+.089} = .0805$$

Black Infants

$$e^{-2.198+.410(+1)-.089(0)} = .167$$

Other Infants

$$e^{-2.198+.410(0)-.089(+1)} = .102$$

OR for Blacks vs. Whites
= .167/.0805 = 2.075

OR for Others vs. Whites
= .102/.0805 = 1.267

OR for Black vs. Others
= .167/.102 = 1.637

Example 3: Race and Low Birth Weight

Finding these directly using the estimated parameters is cumbersome. JMP will compute the Odds Ratio for each possible comparison and their reciprocals in case those are of interest as well.

Odds Ratios

For Low Birth odds of Low versus Norm

Odds Ratios for Race

Level1	/Level2	Odds Ratio	Reciprocal
Other	Black	0.606942	1.6476038
White	Black	0.4811589	2.0783155
White	Other	0.7927592	1.261417

Odds Ratio column is odds for Low for Level 1 vs. Level 2.

Reciprocal is odds for Low for Level 2 vs. Level 1. These are the easiest to interpret here as they represent increased risk.

Putting it all together

Now that we have seen how to interpret each of the variable types in a logistic model we can consider multiple logistic regression models with all these variable types included in the model.

We can then look at risk associated with certain factors adjusted for the other covariates included in the model.

Example 3: Smoking and Low Birth Weight

- Consider again the risk associated with smoking but this time adjusting for the potential confounding effects of education level and age of the mother & father, race of the child, total number of prior pregnancies, number children born alive that are now dead, and gestational age of the infant.

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	24.3444117	0.2557917	9057.9	0.0000*
Gender of child[1]	-0.1858494	0.01419	171.54	<.0001*
Age of father	-0.0012934	0.0030945	0.17	0.6760
Age of mother	0.00221874	0.003829	0.34	0.5623
Education of father (years)	0.00367121	0.0073201	0.25	0.6160
Education of mother (years)	-0.0047079	0.0074148	0.40	0.5255
Total Preg	-0.0444083	0.010651	17.38	<.0001*
BDead	0.15801032	0.0882007	3.21	0.0732
Smoker[Cigs]	0.38427179	0.0207736	342.18	<.0001*
Race[Black]	0.20104655	0.0289675	48.17	<.0001*
Race[Other]	0.07387835	0.0423638	3.04	0.0812
Gest Age	-0.7030361	0.0065632	11474	0.0000*

For log odds of Low/Norm

Several terms are not statistically significant and could consider using backwards elimination to simplify the model.

Example 3: Race and Low Birth Weight

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	22.8038037	0.2054094	12325	0.0000*
Gender of child[1]	-0.1784798	0.0122621	211.86	<.0001*
Total Preg	-0.0335729	0.0079688	17.75	<.0001*
BDead	0.1724931	0.0730957	5.57	0.0183*
Smoker[Cigs]	0.38081327	0.0163453	542.80	<.0001*
Race[Black]	0.21258524	0.0240098	78.40	<.0001*
Race[Other]	0.07032665	0.0378297	3.46	0.0630
Gest Age	-0.6610415	0.0054805	14548	0.0000*

For log odds of Low/Norm

None of the mother and farther related covariates entered into the final model.

Adjusting for the included covariates we find smoking is statistically significant ($p < .0001$)

Odds Ratios for Smoker

Level1	/Level2	Odds Ratio	Reciprocal
No	Cigs	0.4669064	2.141757

Adjusting for the included covariates we find the odds ratio for low birth weight associated with smoking during pregnancy is 2.142.

Odds Ratios for the other factors in the model can be computed as well. All of which can be prefaced by the "adjusting for…" statement.

Summary

- In logistic regression the response (Y) is a dichotomous categorical variable.
- The parameter estimates give the odds ratio associated the variables in the model.
- These odds ratios are adjusted for the other variables in the model.
- One can also calculate $P(Y|\mathbf{X})$ if that is of interest, e.g. given demographics of the mother what is the estimated probability of her having a child with low birth weight.