Logistic Regression

Introduction to Binary Outcomes

Continuous vs. Categorical Variables

- General linear regression model:
  $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$
- Independent variables (x's):
  - Continuous: age, income, height → use numerical value.
  - Categorical: gender, city, ethnicity → use dummies.
- Dependent variable (y):
  - Continuous: consumption, time spent → use numerical value.
  - Categorical: yes/no → use dummies.

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?

Representing the Binary Outcomes

- There are two outcomes: Yes and No
- We will create a dummy variable to indicate if an observation is a Yes or a No:
  - $y = 1$ if Yes
  - $y = 0$ if No
- If we code the variable the other way around, our coefficients will have the same magnitudes but opposite signs.

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.

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.

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

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 $0 \leq p \leq 1$.

Problems with the Linear Approach

- Probabilities are bounded whereby $0 \le p \le 1$.
- The range of age in the data is such that $20 \le age \le 55$.

Problems with the Linear Approach

- Probabilities are bounded whereby $0 \le p \le 1$.
- The range of age in the data is such that $20 \le age \le 55$.
- The probability that a 35 year-old person subscribes is:
  $p = -1.700 + 0.064 \times 35 = 0.54$

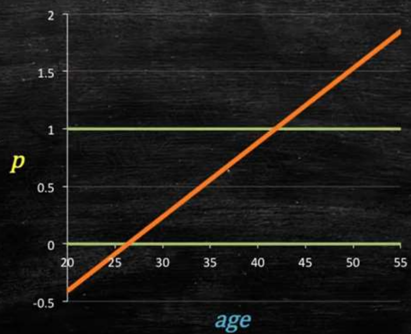
Problems with the Linear Approach

- Probabilities are bounded whereby $0 \le p \le 1$.
- The range of age in the data is such that $20 \le age \le 55$.
- 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?
  $p = -1.700 + 0.064 \times 25 = -0.09$

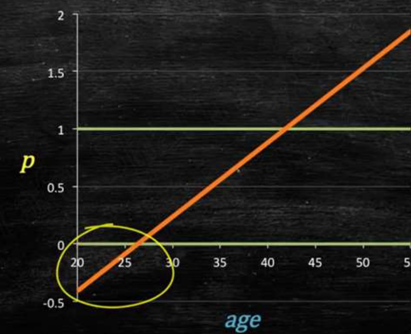
Problems with the Linear Approach

- Probabilities are bounded whereby $0 \le p \le 1$.
- The range of age in the data is such that $20 \le age \le 55$.
- 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?
  $p = -1.700 + 0.064 \times 25 = -0.09$
  $p = -1.700 + 0.064 \times 45 = 1.20$

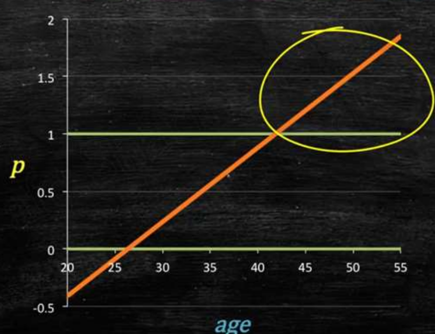
Linear Model Plot



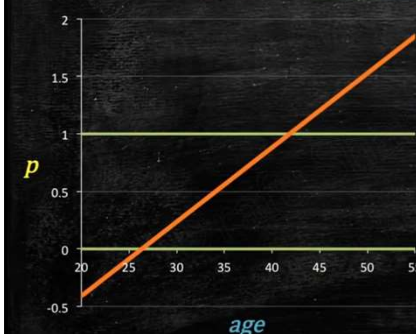
Linear Model Plot

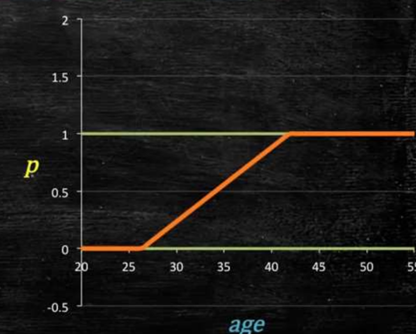


Linear Model Plot



Linear Model Plot





Fixing the prior Approach

- We need to somehow constrain p such that $0 \leq p \leq 1$.

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.

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?

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?
- $f(\bullet)$ must satisfy two things:
  - It must always be positive (since $p \geq 0$)
  - It must be less than 1 (since $p \leq 1$)

Two Steps!

Two Steps!

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

Two Steps!

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

f(x)=abs(x)=|x|

Two Steps!

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

f(x)=x^2

Two Steps!

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

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

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 $p \leq 1$)

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 Linear Thinking is not Completely Gone

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

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

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.

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.
- The above equation is the one used in **logistic regressions**.

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
```

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:

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

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:

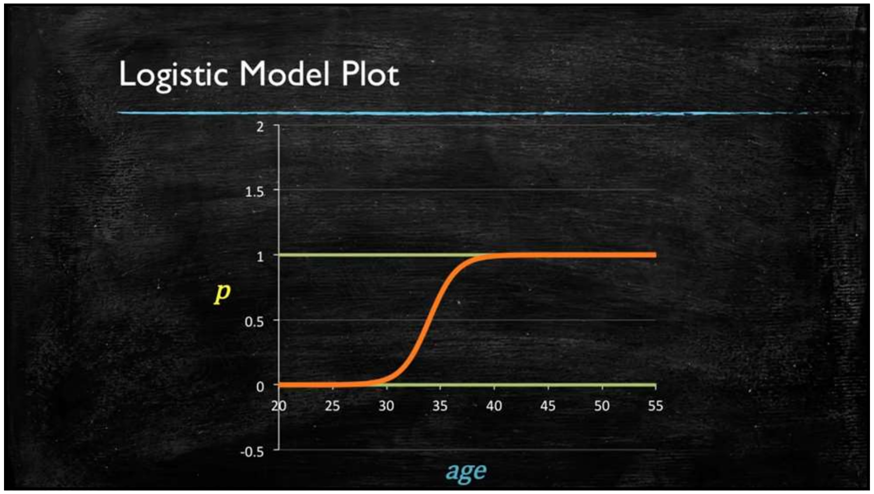
The Estimated Logistic Model

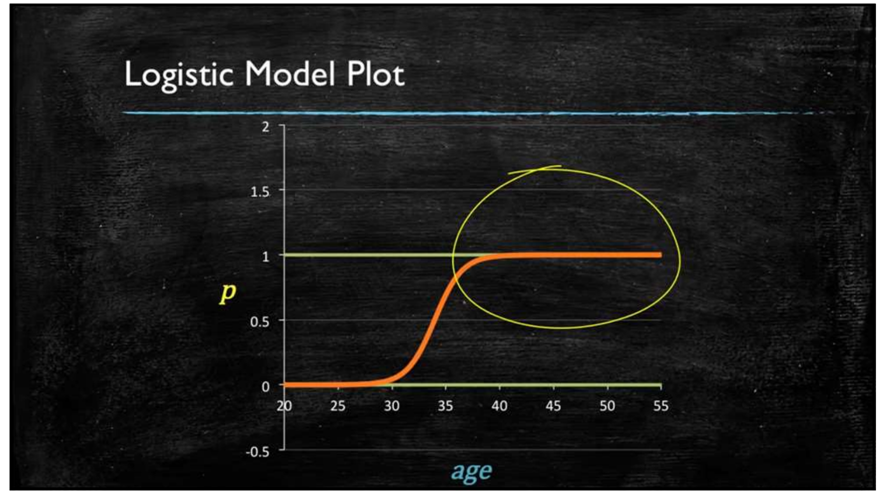
- The estimated model was:

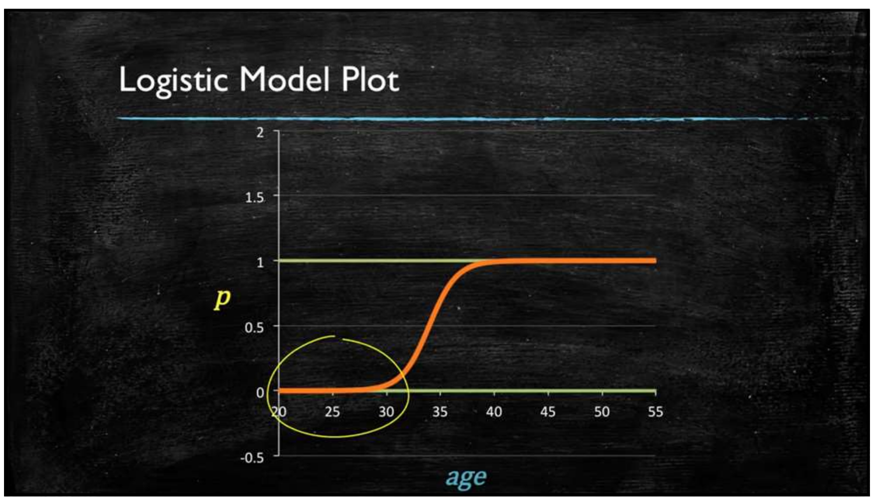
$$\ln\left(\frac{p}{1-p}\right) = -26.52 + 0.78\ age$$

- 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 Regression

- Supervised learning method for classification.
- "logit" = "log odds"

$$odds = \frac{P(event)}{1 - P(event)}$$

- Let $\Pr(y = 1 | X) = p(X)$

Logistic Regression

- Supervised learning method for classification.
- "logit" = "log odds"

$$odds = \frac{P(event)}{1 - P(event)} \qquad X \epsilon \boldsymbol{R} \qquad p(X) \epsilon [0, 1]$$

- Let $\Pr(y = 1 | X) = p(X)$

Logistic Regression

- Supervised learning method for classification.
- "logit" = "log odds"

$$odds = \frac{P(event)}{1 - P(event)} \qquad X \epsilon \boldsymbol{R} \qquad p(X) \epsilon [0, 1]$$

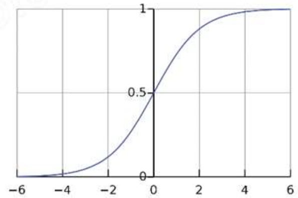
- Let $\Pr(y = 1 | X) = p(X)$
- Sigmoid Function: $p(X) = \frac{1}{1 + e^{-\beta X}}$

Logistic Regression

- Supervised learning method for classification.
- "logit" = "log odds"

$$odds = \frac{P(event)}{1 - P(event)} \qquad X \epsilon \boldsymbol{R} \qquad p(X) \epsilon [0, 1]$$

- Let $\Pr(y = 1 | X) = p(X)$
- Sigmoid Function: $p(X) = \frac{1}{1 + e^{-\beta X}}$



Parameter Estimation

- Goal of learning is to estimate parameter vector $\hat{\beta}$

Parameter Estimation

- Goal of learning is to estimate parameter vector $\widehat{\beta}$
- Logistic Regression uses Maximum Likelihood for parameter estimation. How does this work?

Why not OLS?

Parameter Estimation

Why not OLS?

What is the problem with this, if any? For purely predictive purposes, this actually is not a crazy idea—it tends to give decent predictions. But there are two drawbacks:

1. We cannot use any of the well-established routines for statistical inference with least squares (e.g., confidence intervals, etc.), because these are based on a model in which the outcome is continuously distributed. At an even more basic level, it is hard to precisely interpret $\hat{\beta}$
2. We cannot use this method when the number of classes exceeds 2.

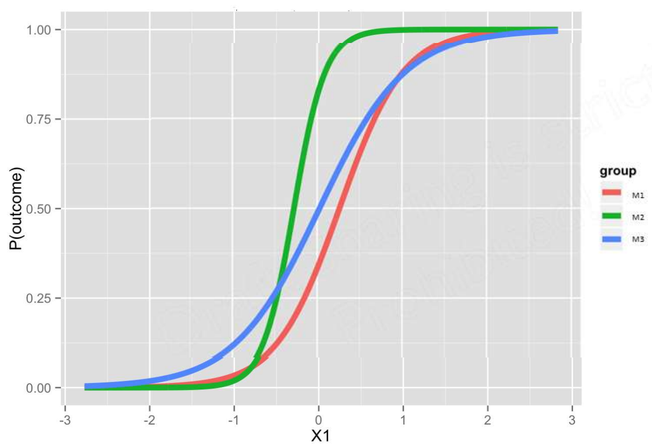
Parameter Estimation

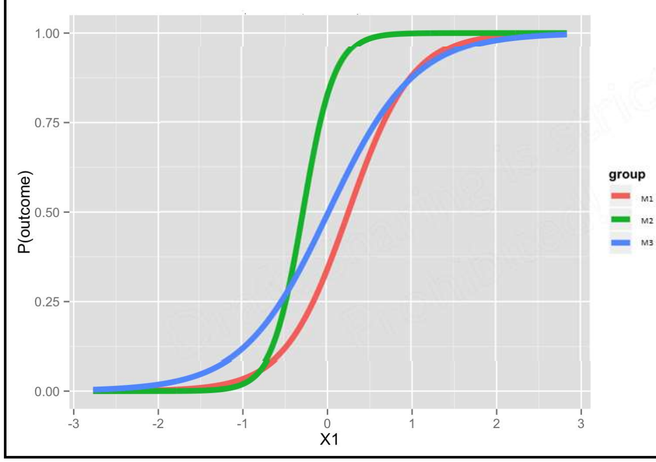
- Goal of learning is to estimate parameter vector $\widehat{\beta}$
- Logistic Regression uses Maximum Likelihood for parameter estimation. How does this work?
  - Consider N samples with labels either 0 or 1

Parameter Estimation

- Goal of learning is to estimate parameter vector $\widehat{\beta}$
- Logistic Regression uses Maximum Likelihood for parameter estimation. How does this work?
  - Consider N samples with labels either 0 or 1

Intuition?



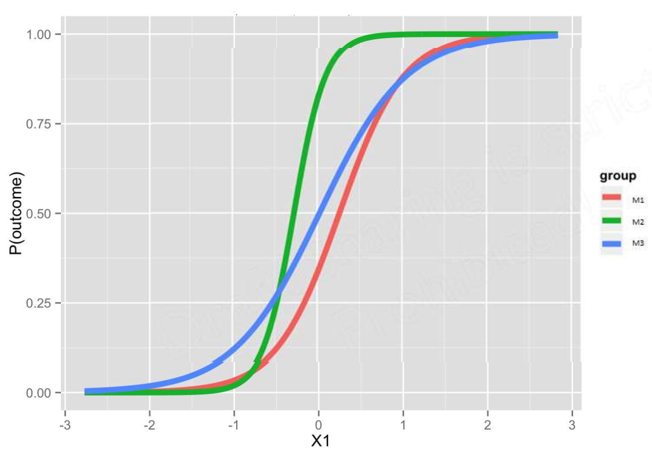


Data:
Students = {A, B, C, D}
A = Pass
B = Fail
C = Fail
D = Pass

M1:
P(A = Pass) = .85
P(B = Pass) = .25
P(C = Pass) = .45
P(D = Pass) = .76

M2:
P(A = Pass) = .94
P(B = Pass) = .23
P(C = Pass) = .10
P(D = Pass) = .91

M3:
P(A = Pass) = .75
P(B = Pass) = .64
P(C = Pass) = .39
P(D = Pass) = .47

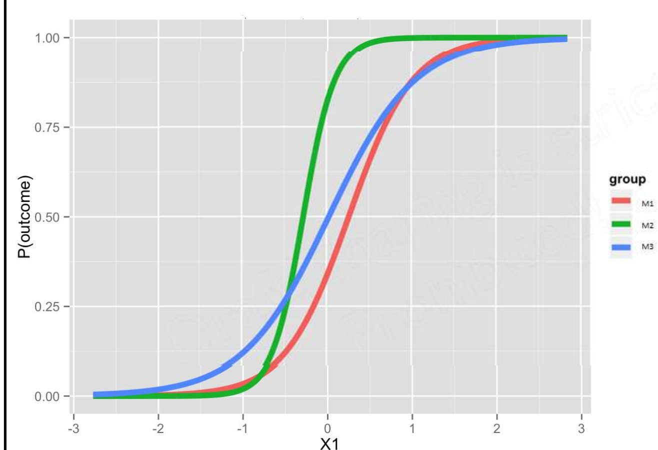


Data:
Students = {A, B, C, D}
A = Pass
B = Fail
C = Fail
D = Pass

M1:
P(A = Pass) = .85
P(B = Pass) = .25
P(C = Pass) = .45
P(D = Pass) = .76

M2:
P(A = Pass) = .94
P(B = Pass) = .23
P(C = Pass) = .10
P(D = Pass) = .91

M3:
P(A = Pass) = .75
P(B = Pass) = .64
P(C = Pass) = .39
P(D = Pass) = .47

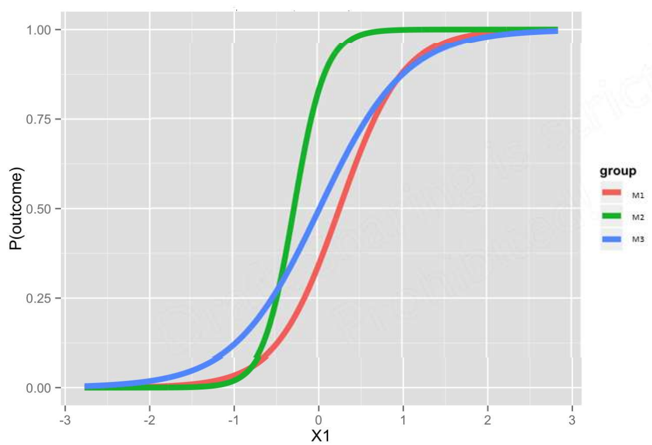


Data:
Students = {A, B, C, D}
A = Pass
B = Fail
C = Fail
D = Pass

M1:
P(A = Pass) = .85
P(B = Pass) = .25
P(C = Pass) = .45
P(D = Pass) = .76

M2:
P(A = Pass) = .94
P(B = Pass) = .23
P(C = Pass) = .10
P(D = Pass) = .91

M3:
P(A = Pass) = .75
P(B = Pass) = .64
P(C = Pass) = .39
P(D = Pass) = .47

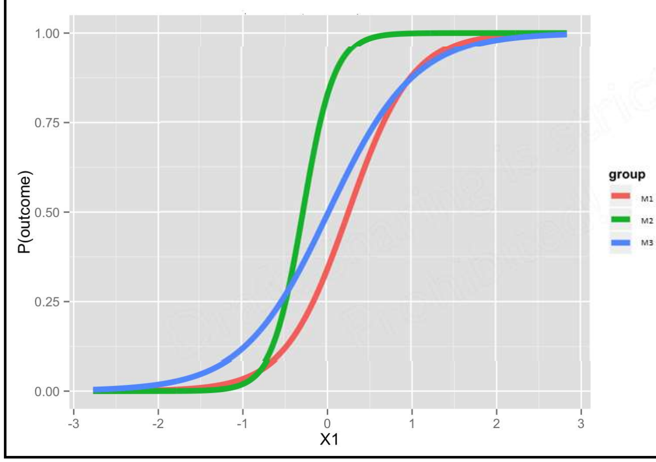


Data:
Students = {A, B, C, D}
A = Pass
B = Fail
C = Fail
D = Pass

M1:
P(A = Pass) = .85
P(B = Pass) = .25
P(C = Pass) = .45
P(D = Pass) = .76

M2:
P(A = Pass) = .94
P(B = Pass) = .23
P(C = Pass) = .10
P(D = Pass) = .91

M3:
P(A = Pass) = .75
P(B = Pass) = .64
P(C = Pass) = .39
P(D = Pass) = .47



Data:
Students = {A, B, C, D}
A = Pass
B = Fail
C = Fail
D = Pass

M1:
P(A = Pass) = .85
P(B = Pass) = .25
P(C = Pass) = .45
P(D = Pass) = .76

M2:
P(A = Pass) = .94
P(B = Pass) = .23
P(C = Pass) = .10
P(D = Pass) = .91

M3:
P(A = Pass) = .75
P(B = Pass) = .64
P(C = Pass) = .39
P(D = Pass) = .47

Parameter Estimation

- Goal of learning is to estimate parameter vector $\widehat{\beta}$
- Logistic Regression uses Maximum Likelihood for parameter estimation. How does this work?
  - Consider N samples with labels either 0 or 1
  - For samples labelled "1": Estimate $\widehat{\beta}$ such that $\widehat{p(X)}$ is as close to 1 as possible

Parameter Estimation

- Goal of learning is to estimate parameter vector $\widehat{\beta}$
- Logistic Regression uses Maximum Likelihood for parameter estimation. How does this work?
  - Consider N samples with labels either 0 or 1
  - For samples labelled "1": Estimate $\widehat{\beta}$ such that $\widehat{p(X)}$ is as close to 1 as possible
  - For samples labelled "0": Estimate $\widehat{\beta}$ such that $1 - \widehat{p(X)}$ is as close to 1 as possible

Parameter Estimation

- Goal of learning is to estimate parameter vector $\widehat{\beta}$
- Logistic Regression uses Maximum Likelihood for parameter estimation. How does this work?
  - Consider N samples with labels either 0 or 1
  - For samples labelled "1": Estimate $\widehat{\beta}$ such that $\widehat{p(X)}$ is as close to 1 as possible
  - For samples labelled "0": Estimate $\widehat{\beta}$ such that $1 - \widehat{p(X)}$ is as close to 1 as possible

$$\prod_{s \text{ in } yi=1} p(x_i)$$

Parameter Estimation

- Goal of learning is to estimate parameter vector $\widehat{\beta}$
- Logistic Regression uses Maximum Likelihood for parameter estimation. How does this work?
  - Consider N samples with labels either 0 or 1
  - For samples labelled "1": Estimate $\widehat{\beta}$ such that $\widehat{p(X)}$ is as close to 1 as possible
  - For samples labelled "0": Estimate $\widehat{\beta}$ such that $1 - \widehat{p(X)}$ is as close to 1 as possible

$$\prod_{s \text{ in } yi=1} p(x_i) \qquad \prod_{s \text{ in } yi=0} (1 - p(x_i))$$

Parameter Estimation

$$L(\beta) = \prod_{s \text{ in } yi=1} p(x_i) \times \prod_{s \text{ in } yi=0} (1 - p(x_i))$$

Parameter Estimation

$$L(\beta) = \prod_{s \text{ in } yi=1} p(x_i) \times \prod_{s \text{ in } yi=0} (1 - p(x_i))$$

$$L(\beta) = \prod_{s} p(x_i)^{y_i} \times \left(1 - p(x_i)\right)^{1-y_i}$$

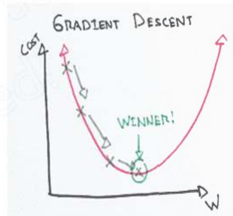
Parameter Estimation

$$L(\beta) = \prod_{s\ in\ yi=1} p(x_i) \times \prod_{s\ in\ yi=0} (1 - p(x_i))$$

$$L(\beta) = \prod_{s} p(x_i)^{y_i} \times (1 - p(x_i))^{1-y_i}$$

$$l(\beta) = \sum_{i=1}^{n} y_i \log(p(x_i)) + (1 - y_i) \log(1 - p(x_i))$$

The method most commonly used for logistic regression to estimate parameters is gradient descent.

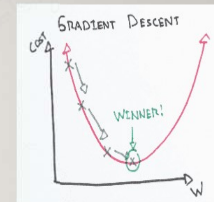


But, for gradient descent, we try to minimize the cost function. So, we can use a simple trick. maximizing $l(\beta)$ is equivalent to minimizing $-l(\beta)$. Therefore, our job turns into:

Minimize $$\sum_{i=1}^{n} -y_i \log(p(x_i)) - (1 - y_i) \log(1 - p(x_i))$$

Sum up: Convexity, Gradient Descent, and Log-Likelihood

- The error function is the function through which we optimize the parameters of a machine learning model
- This optimization takes place through algorithmic methods for learning
- The method most commonly used for logistic regression is gradient descent
- Gradient descent requires convex cost functions
- Mean Squared Error, commonly used for linear regression models, isn't convex for logistic regression. This is because the logistic function isn't always convex (another reason for not using OLS)
- The logarithm of the likelihood function is however always convex



We, therefore, elect to use the log-likelihood function as a cost function for logistic regression.

Logistic Regression

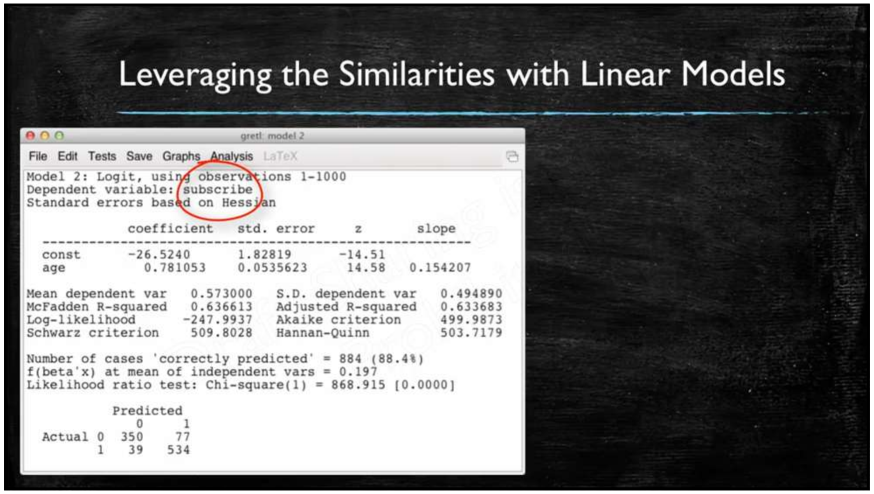
Interpretation of Coefficients and Forecasting

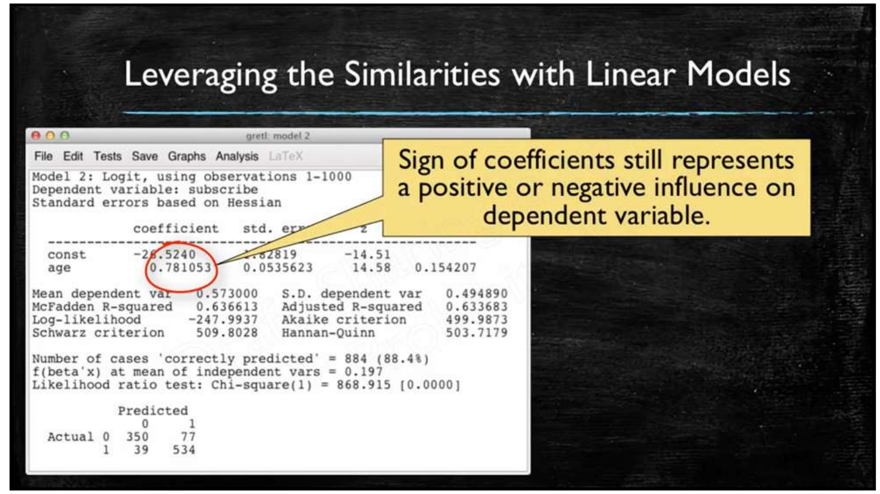
Leveraging the Similarities with Linear Models

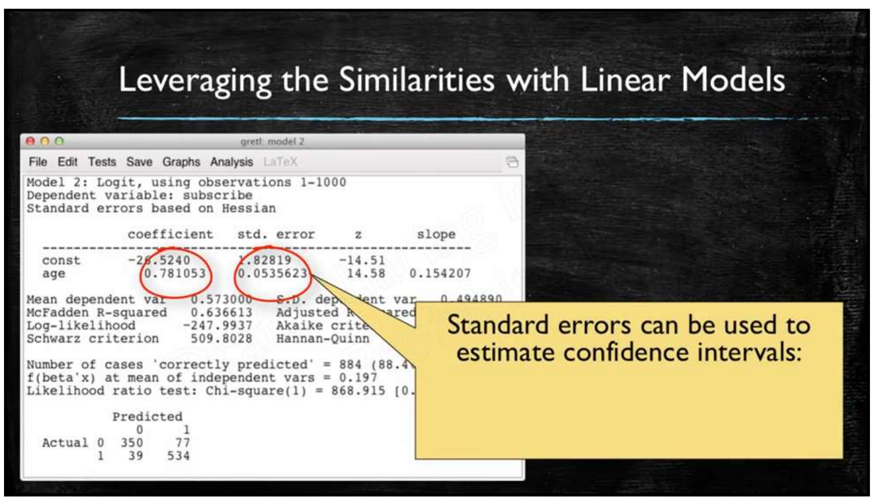
gretl: model 2

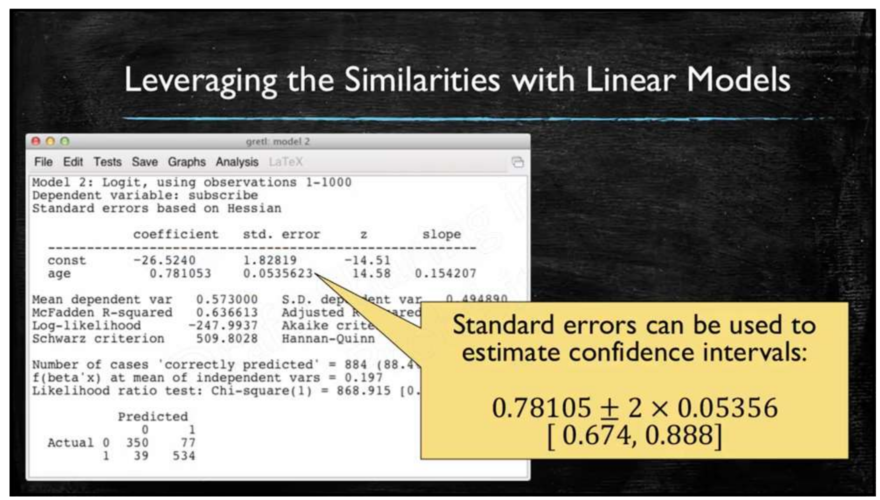
File Edit Tests Save Graphs Analysis LaTeX

```
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
```


Leveraging the Similarities with Linear Models
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


Leveraging the Similarities with Linear Models
Sign of coefficients still represents a positive or negative influence on dependent variable.
Model 2: Logit, using observations 1-1000
Dependent variable: subscribe
Standard errors based on Hessian
const -26.5240 1.82819 -14.51
age 0.781053 0.0535623 14.58 0.154207


Leveraging the Similarities with Linear Models
Standard errors can be used to estimate confidence intervals:
Model 2: Logit, using observations 1-1000
Dependent variable: subscribe
const -26.5240 1.82819 -14.51
age 0.781053 0.0535623 14.58 0.154207


Leveraging the Similarities with Linear Models
Standard errors can be used to estimate confidence intervals:
0.78105 ± 2 × 0.05356
[0.674, 0.888]
Model 2: Logit, using observations 1-1000
Dependent variable: subscribe
const -26.5240 1.82819 -14.51
age 0.781053 0.0535623 14.58 0.154207

What changed?

What changed?

- We can no longer interpret the (magnitude of) the coefficients as we did before.

What changed?

- We can no longer interpret the (magnitude of) the coefficients as we did before.
- What is the meaning of 0.78 in our estimated model?

What changed?

- We can no longer interpret the (magnitude of) the coefficients as we did before.
- What is the meaning of 0.78 in our estimated model?

$$\ln\left(\frac{p}{1-p}\right) = -26.524 + 0.781\ age$$

What changed?

- We can no longer interpret the (magnitude of) the coefficients as we did before.
- What is the meaning of 0.78 in our estimated model?

$$\ln\left(\frac{p}{1-p}\right) = -26.524 + 0.781\ age$$

- For every unit increase of age, $\ln\left(\frac{p}{1-p}\right)$ increases 0.78 units.

What changed?

- We can no longer interpret the (magnitude of) the coefficients as we did before.
- What is the meaning of 0.78 in our estimated model?

$$\ln\left(\frac{p}{1-p}\right) = -26.524 + 0.781\ age$$

- For every unit increase of age, $\ln\left(\frac{p}{1-p}\right)$ increases 0.78 units.

- But what is $\ln\left(\frac{p}{1-p}\right)$? Let us call it y^*

From y^* to p

- If we have: $y^* = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1\ age$

From y^* to p

- If we have: $y^* = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1\ age$

- Then: $p = \dfrac{\exp(\beta_0 + \beta_1\ age)}{\exp(\beta_0 + \beta_1\ age) + 1} = \dfrac{e^{\beta_0 + \beta_1\ age}}{e^{\beta_0 + \beta_1\ age} + 1}$

From y^* to p

- If we have: $y^* = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1\, age$
- Then: $p = \dfrac{\exp(\beta_0 + \beta_1\, age)}{\exp(\beta_0 + \beta_1\, age) + 1} = \dfrac{e^{\beta_0 + \beta_1\, age}}{e^{\beta_0 + \beta_1\, age} + 1}$
- Or simply put: $p = \dfrac{\exp(y^*)}{\exp(y^*) + 1} = \dfrac{e^{y^*}}{e^{y^*} + 1}$

From y^* to p in Excel

Variable	Coefficients	35y
Constant	-26.524000	1
Age	0.781053	35
y* = ln(p/(1-p))		0.813
p = exp(y*)/(exp(y*)+1)		0.693

From y^* to p in Excel

Variable	Coefficients	35y	36y
Constant	-26.524000	1	1
Age	0.781053	35	36
y* = ln(p/(1-p))		0.813	1.594
p = exp(y*)/(exp(y*)+1)		0.693	0.831

From y^* to p in Excel

Variable	Coefficients	35y	36y
Constant	-26.524000	1	1
Age	0.781053	35	36
y* = ln(p/(1-p))		0.813	1.594
p = exp(y*)/(exp(y*)+1)		0.693	0.831
Change			0.138

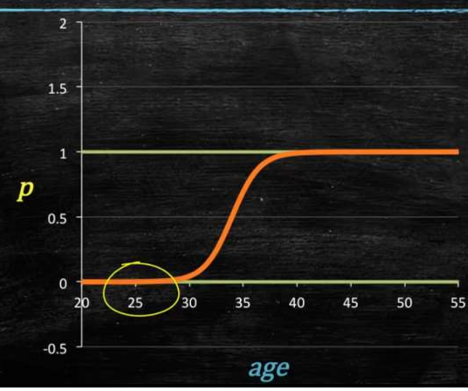
From y^* to p in Excel

Variable	Coefficients	35y	36y	25y	26y
Constant	-26.524000	1	1	1	1
Age	0.781053	35	36	25	26
y* = ln(p/(1-p))		0.813	1.594	-7	-6.22
p = exp(y*)/(exp(y*)+1)		0.693	0.831	9E-04	0.002
Change			0.138		0.001

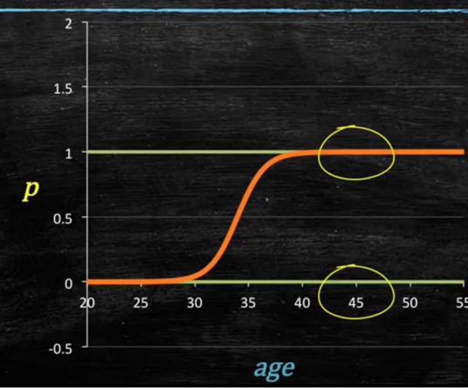
From y^* to p in Excel

Variable	Coefficients	35y	36y	25y	26y	45y	46y
Constant	-26.524000	1	1	1	1	1	1
Age	0.781053	35	36	25	26	45	46
y* = ln(p/(1-p))		0.813	1.594	-7	-6.22	8.623	9.404
p = exp(y*)/(exp(y*)+1)		0.693	0.831	9E-04	0.002	1	1
Change			0.138		0.001		1E-04

Logistic Model Plot



Logistic Model Plot



Multiple Logistic Regression

- Moving from a single regression with a single independent variable (age) to a multiple regression model with more than one (age and gender) is very simple.

Multiple Logistic Regression

- Moving from a single regression with a single independent variable (age) to a multiple regression model with more than one (age and gender) is very simple.

- Let's add a dummy variable for gender and run

$$y^* = \beta_0 + \beta_1\, age + \beta_2\, woman$$

Result of Multiple Logistic Regression

gretl: model 3

File Edit Tests Save Graphs Analysis LaTeX

```
Model 3: Logit, using observations 1-1000
Dependent variable: subscribe
Standard errors based on Hessian

             coefficient   std. error      z        slope
  ---------------------------------------------------------
  const      -26.4653      1.84246      -14.36
  age          0.787213    0.0542546     14.51     0.155652
  woman       -0.557795    0.231171     -2.413    -0.110022

Mean dependent var   0.573000   S.D. dependent var   0.494890
McFadden R-squared   0.640941   Adjusted R-squared   0.636545
Log-likelihood      -245.0401   Akaike criterion     496.0803
Schwarz criterion    510.8035   Hannan-Quinn         501.6762

Number of cases 'correctly predicted' = 886 (88.6%)
f(beta'x) at mean of independent vars = 0.198
Likelihood ratio test: Chi-square(2) = 874.822 [0.0000]

           Predicted
              0     1
  Actual 0  367    60
         1   54   519
```

- Estimated model is:

$$y^* = -26.47 + 0.79 \times age - 0.56 \times woman$$

Result of Multiple Logistic Regression

gretl: model 3

File Edit Tests Save Graphs Analysis LaTeX

```
Model 3: Logit, using observations 1-1000
Dependent variable: subscribe
Standard errors based on Hessian

             coefficient   std. error      z        slope
  ---------------------------------------------------------
  const      -26.4653      1.84246      -14.36
  age          0.787213    0.0542546     14.51     0.155652
  woman       -0.557795    0.231171     -2.413    -0.110022

Mean dependent var   0.573000   S.D. dependent var   0.494890
McFadden R-squared   0.640941   Adjusted R-squared   0.636545
Log-likelihood      -245.0401   Akaike criterion     496.0803
Schwarz criterion    510.8035   Hannan-Quinn         501.6762

Number of cases 'correctly predicted' = 886 (88.6%)
f(beta'x) at mean of independent vars = 0.198
Likelihood ratio test: Chi-square(2) = 874.822 [0.0000]

           Predicted
              0     1
  Actual 0  367    60
         1   54   519
```

- Estimated model is:

$$y^* = -26.47 + 0.79 \times age - 0.56 \times woman$$

- We can interpret signs and compute confidence intervals.

Result of Multiple Logistic Regression

```
gretl: model 3
File  Edit  Tests  Save  Graphs  Analysis  LaTeX
Model 3: Logit, using observations 1-1000
Dependent variable: subscribe
Standard errors based on Hessian

             coefficient   std. error      z        slope
  ---------------------------------------------------------
  const      -26.4653      1.84246      -14.36
  age          0.787213    0.0542546     14.51     0.155652
  woman       -0.557795    0.231171      -2.413   -0.110022

Mean dependent var   0.573000   S.D. dependent var   0.494890
McFadden R-squared   0.640941   Adjusted R-squared   0.636545
Log-likelihood      -245.0401   Akaike criterion     496.0803
Schwarz criterion    510.8035   Hannan-Quinn         501.6762

Number of cases 'correctly predicted' = 886 (88.6%)
f(beta'x) at mean of independent vars = 0.198
Likelihood ratio test: Chi-square(2) = 874.822 [0.0000]

           Predicted
             0      1
  Actual 0  367     60
         1   54    519
```

- Estimated model is:
  $y^* = -26.47 + 0.79 \times age - 0.56 \times woman$
- We can interpret signs and compute confidence intervals.
- We want p.

From y^* to p in Excel

Variable	Coefficients
Constant	-26.465300
Age	0.787213
Woman	-0.557795

y* = ln(p/(1-p))

p = exp(y*)/(exp(y*)+1)

Change

From y^* to p in Excel

Variable	Coefficients	35y W	35y M
Constant	-26.465300	1	1
Age	0.787213	35	35
Woman	-0.557795	1	0
y* = ln(p/(1-p))		0.529	1.087
p = exp(y*)/(exp(y*)+1)		0.629	0.748
Change			0.119

From y^* to p in Excel

Variable	Coefficients	35y W	35y M	36y W	36 M
Constant	-26.465300	1	1	1	1
Age	0.787213	35	35	36	36
Woman	-0.557795	1	0	1	0
y* = ln(p/(1-p))		0.529	1.087	1.317	1.874
p = exp(y*)/(exp(y*)+1)		0.629	0.748	0.789	0.867
Change			0.119		

From y^* to p in Excel

Variable	Coefficients	35y W	35y M	36y W	36 M
Constant	-26.465300	1	1	1	1
Age	0.787213	35	35	36	36
Woman	-0.557795	1	0	1	0
y* = ln(p/(1-p))		0.529	1.087	1.317	1.874
p = exp(y*)/(exp(y*)+1)		0.629	0.748	0.789	0.867
Change			0.119		

From y^* to p in Excel

Variable	Coefficients	35y W	35y M	36y W	36 M
Constant	-26.465300	1	1	1	1
Age	0.787213	35	35	36	36
Woman	-0.557795	1	0	1	0
y* = ln(p/(1-p))		0.529	1.087	1.317	1.874
p = exp(y*)/(exp(y*)+1)		0.629	0.748	0.789	0.867
Change			0.119		0.078

Example : continuation of previously shown linear regression example

Categorical dependent variable

What if Y is categorical?

$$\underbrace{Ln\left(\frac{P_i}{1-P_i}\right) = \beta_0 + \beta_1(Price_i) + \beta_2(Pink\ Slip_i) + \varepsilon_i}$$

BINOMIAL LOGISTIC REGRESSION

Categorical dependent variable

What if Y is categorical?

REMEMBER!
"Chance" = "Probability"
≠ "Odds"

If the probability of rain tomorrow is 20%,
what are the odds of rain tomorrow?

Categorical dependent variable

What if Y is categorical?

If the probability of rain tomorrow is 20%,
what are the odds of rain tomorrow?

Categorical dependent variable

What if Y is categorical?

If the probability of rain tomorrow is 20%,
what are the odds of rain tomorrow?

$$Odds_i = \frac{P}{1-P} = \frac{0.2}{0.8} = 0.25$$

Categorical dependent variable

What if Y is categorical?

If the probability of rain tomorrow is 20%,
what are the odds of rain tomorrow?

$$Odds_i = \frac{P}{1-P} = \frac{0.2}{0.8} = 0.25 \quad OR \quad 1:4$$

Categorical dependent variable

What if Y is categorical?

$$Ln\left(\frac{P_i}{1-P_i}\right) = \beta_0 + \beta_1(Price_i) + \beta_2(Pink\ Slip_i) + \varepsilon_i$$

DV: Log Odds (Sold)	Coef	SE	Z	P-value	OR	95% LCL	95% UCL
Intercept	0.396	0.480	0.82	0.4097			
Price ($000s)	-0.173	0.057	-3.04	0.0023	0.84	0.75	0.94
Pink slip	1.555	0.531	2.93	0.0034	4.73	1.67	13.41

Categorical dependent variable

What if Y is categorical?

$$Ln\left(\frac{P_i}{1-P_i}\right) = \beta_0 + \beta_1(Price_i) + \beta_2(Pink\ Slip_i) + \varepsilon_i$$

DV: Log Odds (Sold)	Coef	SE	Z	P-value	OR	95% LCL	95% UCL
Intercept	0.396	0.480	0.82	0.4097			
Price ($000s)	-0.173	0.057	-3.04	0.0023	0.84	0.75	0.94
Pink slip	1.555	0.531	2.93	0.0034	4.73	1.67	13.41

$$Ln\widehat{\left(\frac{P_i}{1-P_i}\right)} = 0.396 - 0.173(Price_i) + 1.555(Pink\ Slip_i)$$

Categorical dependent variable

What if Y is categorical?

$$Ln\widehat{\left(\frac{P_i}{1-P_i}\right)} = 0.396 - 0.173(Price_i) + 1.555(Pink\ Slip_i)$$

DV: Log Odds (Sold)	Coef	SE	Z	P-value	OR	95% LCL	95% UCL
Intercept	0.396	0.480	0.82	0.4097			
Price ($000s)	-0.173	0.057	-3.04	0.0023	0.84	0.75	0.94
Pink slip	1.555	0.531	2.93	0.0034	4.73	1.67	13.41

For a $1000 increase in price, the log-odds of selling the car decreases by 0.173, on average, holding all else constant

Categorical dependent variable

What if Y is categorical?

$$Ln\widehat{\left(\frac{P_i}{1-P_i}\right)} = 0.396 - 0.173(Price_i) + 1.555(Pink\ Slip_i)$$

DV: Log Odds (Sold)	Coef	SE	Z	P-value	OR	95% LCL	95% UCL
Intercept	0.396	0.480	0.82	0.4097			
Price ($000s)	-0.173	0.057	-3.04	0.0023	0.84	0.75	0.94
Pink slip	1.555	0.531	2.93	0.0034	4.73	1.67	13.41

For a $1000 increase in price, the odds of selling the car decreases by 16%, on average, holding all else constant.

Categorical dependent variable

What if Y is categorical?

$$Ln\left(\widehat{\frac{P_i}{1-P_i}}\right) = 0.396 - 0.173(Price_i) + 1.555(Pink\ Slip_i)$$

DV: Log Odds (Sold)	Coef	SE	Z	P-value	OR	95% LCL	95% UCL
Intercept	0.396	0.480	0.82	0.4097			
Price ($000s)	-0.173	0.057	-3.04	0.0023	0.84	0.75	0.94
Pink slip	1.555	0.531	2.93	0.0034	4.73	1.67	13.41

If the car has a pink slip, the log odds of sale increases by 1.555, on average, holding price constant.

Categorical dependent variable

What if Y is categorical?

$$Ln\left(\widehat{\frac{P_i}{1-P_i}}\right) = 0.396 - 0.173(Price_i) + 1.555(Pink\ Slip_i)$$

DV: Log Odds (Sold)	Coef	SE	Z	P-value	OR	95% LCL	95% UCL
Intercept	0.396	0.480	0.82	0.4097			
Price ($000s)	-0.173	0.057	-3.04	0.0023	0.84	0.75	0.94
Pink slip	1.555	0.531	2.93	0.0034	4.73	1.67	13.41

Cars with a pink slip have 4.73 times the odds of being sold compared to cars without a pink slip, on average, holding all else constant.

Categorical dependent variable

What if Y is categorical?

DV: Log Odds (Sold)	Coef	SE	Z	P-value	OR	95% LCL	95% UCL
Intercept	0.396	0.480	0.82	0.4097			
Price ($000s)	-0.173	0.057	-3.04	0.0023	0.84	0.75	0.94
Pink slip	1.555	0.531	2.93	0.0034	4.73	1.67	13.41

(a) Find the probability that a $4,500 car with a pink slip will sell.

(b) Find the probability that a $4,500 car <u>without</u> a pink slip will sell.

(c) calculate the odds in parts (a) and (b) and find their ratio.

Categorical dependent variable

What if Y is categorical?

DV: Log Odds (Sold)	Coef	SE	Z	P-value	OR	95% LCL	95% UCL
Intercept	0.396	0.480	0.82	0.4097			
Price ($000s)	-0.173	0.057	-3.04	0.0023	0.84	0.75	0.94
Pink slip	1.555	0.531	2.93	0.0034	4.73	1.67	13.41

(a) Find the probability that a $4,500 car with a pink slip will sell

Categorical dependent variable

What if Y is categorical?

DV: Log Odds (Sold)	Coef	SE	Z	P-value	OR	95% LCL	95% UCL
Intercept	0.396	0.480	0.82	0.4097			
Price ($000s)	-0.173	0.057	-3.04	0.0023	0.84	0.75	0.94
Pink slip	1.555	0.531	2.93	0.0034	4.73	1.67	13.41

(a) Find the probability that a $4,500 car with a pink slip will sell

$$Ln\left(\frac{P}{1-P}\right) = 0.396 - 0.173(Price_i) + 1.555(Pink\ Slip_i)$$

Categorical dependent variable

What if Y is categorical?

DV: Log Odds (Sold)	Coef	SE	Z	P-value	OR	95% LCL	95% UCL
Intercept	0.396	0.480	0.82	0.4097			
Price ($000s)	-0.173	0.057	-3.04	0.0023	0.84	0.75	0.94
Pink slip	1.555	0.531	2.93	0.0034	4.73	1.67	13.41

(a) Find the probability that a $4,500 car with a pink slip will sell

$$Ln\left(\frac{P}{1-P}\right) = 0.396 - 0.173(4.5) + 1.555(1)$$

Categorical dependent variable

What if Y is categorical?

DV: Log Odds (Sold)	Coef	SE	Z	P-value	OR	95% LCL	95% UCL
Intercept	0.396	0.480	0.82	0.4097			
Price ($000s)	-0.173	0.057	-3.04	0.0023	0.84	0.75	0.94
Pink slip	1.555	0.531	2.93	0.0034	4.73	1.67	13.41

(a) Find the probability that a $4,500 car with a pink slip will sell

$$Ln\left(\frac{P}{1-P}\right) = 0.396 - 0.173(4.5) + 1.555(1)$$

$$Ln\left(\frac{P}{1-P}\right) = 1.1725$$

Categorical dependent variable

What if Y is categorical?

(a) Find the probability that a $4,500 car with a pink slip will sell.

$$Ln\left(\frac{P}{1-P}\right) = 1.1725$$

Categorical dependent variable

What if Y is categorical?

(a) Find the probability that a $4,500 car with a pink slip will sell.

$$Ln\left(\frac{P}{1-P}\right) = 1.1725$$

$$\frac{P}{1-P} = e^{1.1725} = 3.23$$

Categorical dependent variable

What if Y is categorical?

(a) Find the probability that a $4,500 car with a pink slip will sell.

$$Ln\left(\frac{P}{1-P}\right) = 1.1725$$

$$\frac{P}{1-P} = e^{1.1725} = 3.23$$

$$P = 3.23(1 - P)$$

$$P + 3.23P = 3.23$$

$$P = \frac{3.23}{4.23} = 0.764$$

Categorical dependent variable

What if Y is categorical?

(a) Find the probability that a $4,500 car with a pink slip will sell.

ANSWER:

A car with a sale price of $4,500 and a pink slip has a probability of sale of 76.4%

Categorical dependent variable

What if Y is categorical?

(a) Find the probability that a $4,500 car with a pink slip will sell.

ANSWER:

A car with a sale price of $4,500 and a pink slip has a probability of sale of 76.4%

(b) Find the probability that a $4,500 car without a pink slip will sell.

Categorical dependent variable

What if Y is categorical?

(a) Find the probability that a $4,500 car with a pink slip will sell.

ANSWER:

A car with a sale price of $4,500 and a pink slip has a probability of sale of 76.4%

(b) Find the probability that a $4,500 car without a pink slip will sell.

$$Ln\left(\frac{P}{1-P}\right) = 0.396 - 0.173(4.5) + 1.555(0)$$
$$= 40.6$$

Categorical dependent variable

What if Y is categorical?

(a) Find the probability that a $4,500 car with a pink slip will sell.

ANSWER:

A car with a sale price of $4,500 and a pink slip has a probability of sale of 76.4%

(b) Find the probability that a $4,500 car without a pink slip will sell.

A car with a sale price of $4,500 without a pink slip has probability of sale of 40.6%

Categorical dependent variable

What if Y is categorical?

(c) calculate the odds in parts (a) and (b) and find the odds ratio

Categorical dependent variable

What if Y is categorical?

(c) calculate the odds in parts (a) and (b) and find the odds ratio

	(a)	(b)
Car	$4500, pink slip	$4500, no pink slip

Categorical dependent variable

What if Y is categorical?

(c) calculate the odds in parts (a) and (b) and find the odds ratio

	(a)	(b)
Car	$4500, pink slip	$4500, no pink slip
Prob	76.4%	40.6%

Categorical dependent variable

What if Y is categorical?

(c) calculate the odds in parts (a) and (b) and find the odds ratio

	(a)	(b)
Car	$4500, pink slip	$4500, no pink slip
Prob	76.4%	40.6%
Odds	76.4%/23.6% = 3.23	40.6%/59.4% = 0.68

Categorical dependent variable

What if Y is categorical?

(c) calculate the odds in parts (a) and (b) and find the odds ratio

	(a)	(b)
Car	$4500, pink slip	$4500, no pink slip
Prob	76.4%	40.6%
Odds	76.4%/23.6% = 3.23	40.6%/59.4% = 0.68

Odds Ratio: 3.23/0.68 = 4.73

Categorical dependent variable

What if Y is categorical?

(c) calculate the odds in parts (a) and (b) and find the odds ratio

DV: Log Odds (Sold)	Coef	SE	Z	P-value	OR	95% LCL	95% UCL
Intercept	0.396	0.480	0.82	0.4097			
Price ($000s)	-0.173	0.057	-3.04	0.0023	0.84	0.75	0.94
Pink slip	1.555	0.531	2.93	0.0034	4.73	1.67	13.41

Odds Ratio: 3.23/0.68 = 4.73

THANK YOU!