Applied Statistical Modelling: GLMs – Logit & Probit

Prof.ssa Chiara Seghieri

Laboratorio di Management e Sanità, Istituto di Management, L'EMbeDS
Scuola Superiore Sant'Anna, Pisa
c.seghieri@santannapisa.it
c.tortu@santannapisa.it

Outline

1. Motivation and Intuition
2. Why the standard regression model is not appropriate with binary outcomes
3. Introduction to GLMs
4. The LOGIT Model
5. Interpretation of coefficients
6. Inference on the coefficients
7. Goodness of fit
8. Other common GLMs: the Probit and Poisson Model

1) Motivation and intuition

Introduction

Sometimes the dependent variable Y can take two mutually exclusive values:

- *Y* = get into college (success), or not (failure); *X* = gender
- *Y* = person smokes, or not; *X* = income
- *Y* = mortgage application is accepted, or not; *X* = income, house characteristics, marital status, race

The goal is to describe the way in which the probability of Y=1 (i.e. get into college) varies by X.

Y is a random bernoulli variable taking, for each observation, two values: 1 with probability p, 0 with probability 1-p.

We know that the mean and the variance of a Bernoulli are:

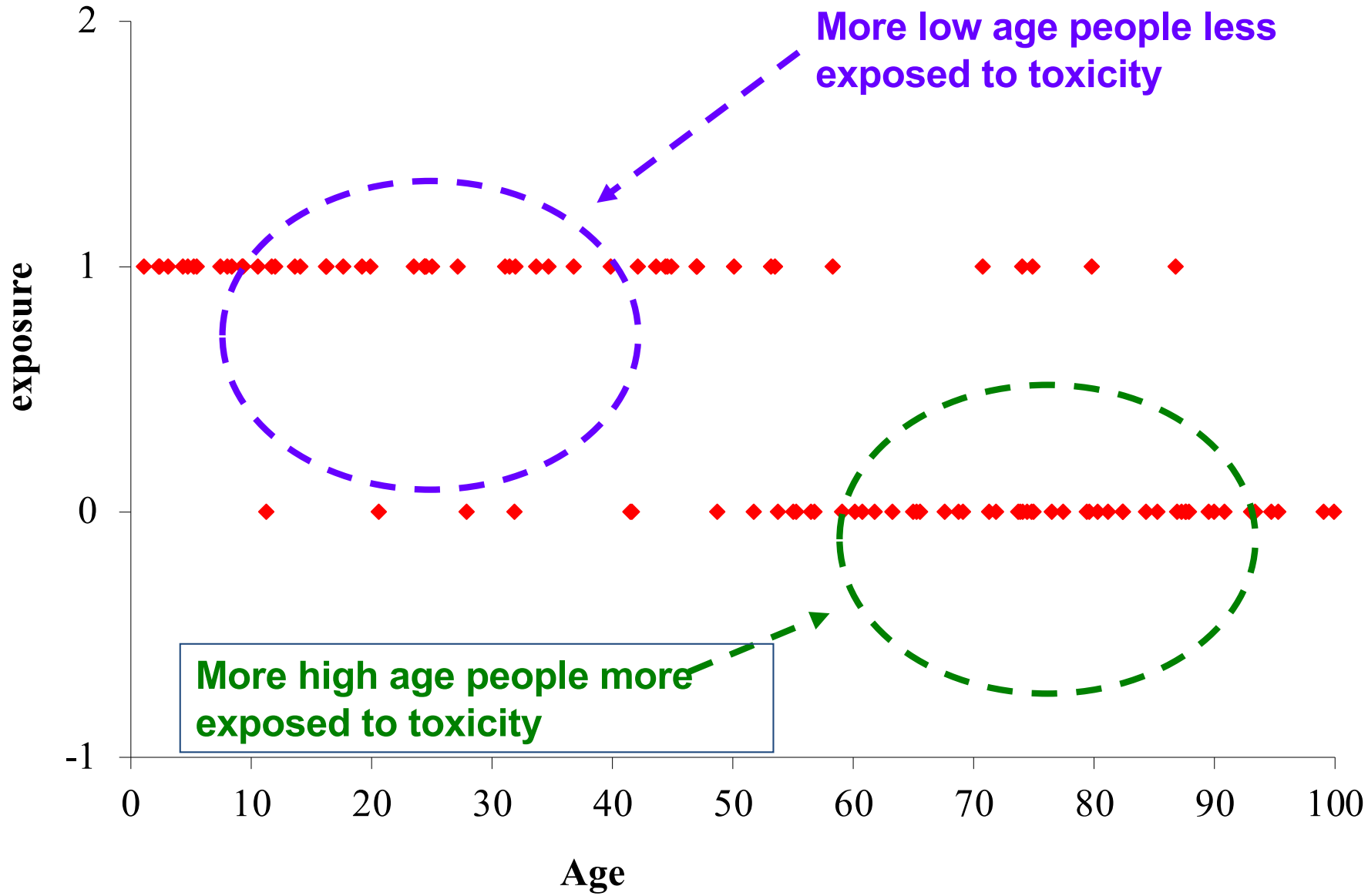
$E(Y)=p$

$var(Y)=p(1-p)$

The general aim is estimating:

$$p_i = P(y_i = j) = F(x_j'\beta) \qquad \text{i=1,2,…,n}$$

where y_i is the dichotomous dependent variable, x_i is a column vector containing the k explanatory variables, and β the k-dimensional vector containing the unknown parameters, p_i represents the probability that the i-th individual makes the j-th choice. In the binary case, j=(0,1)

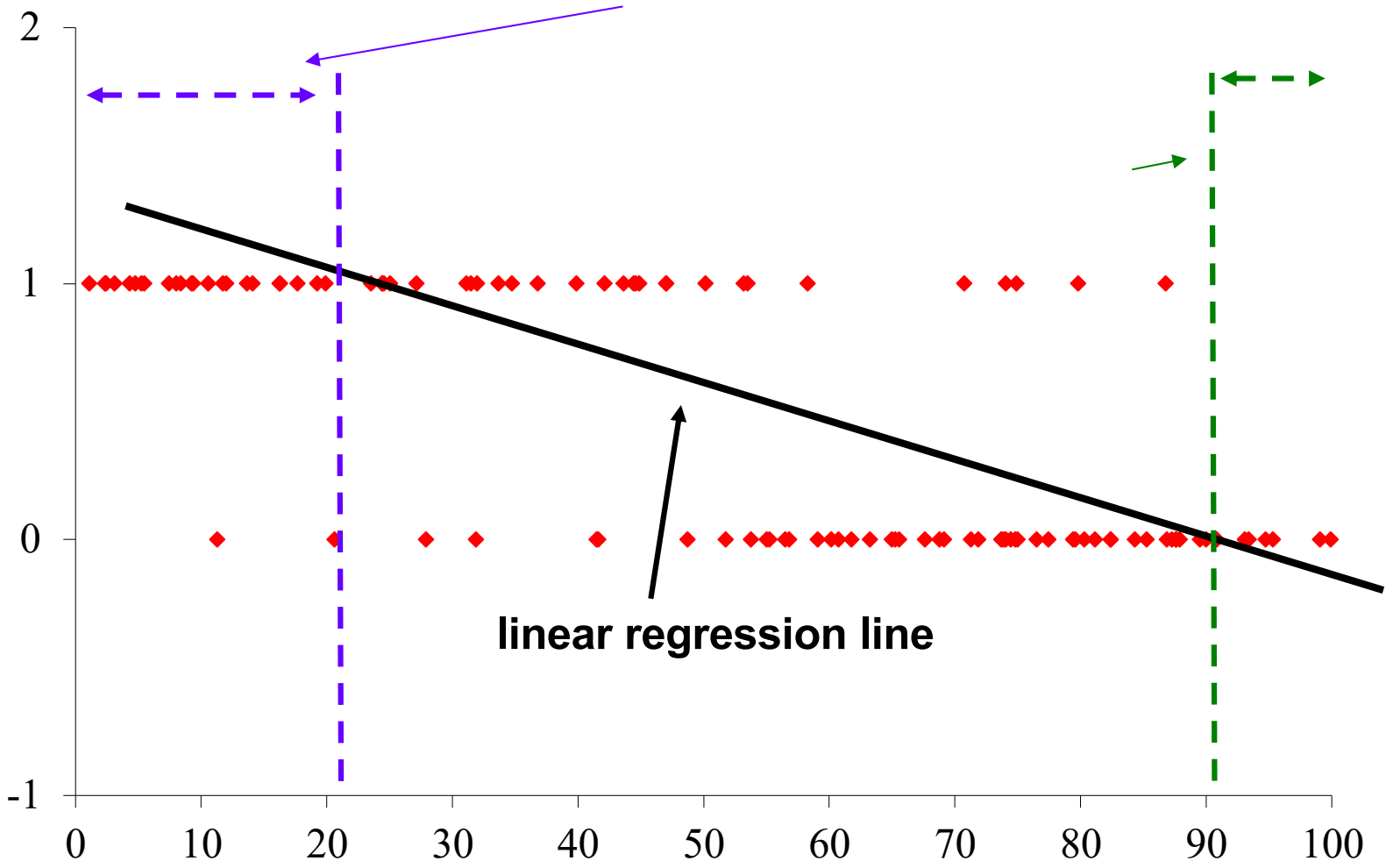

More low age people less exposed to toxicity
More high age people more exposed to toxicity
exposure
Age
2
1
0
-1
0
10
20
30
40
50
60
70
80
90
100

We could estimate the probability using linear regression

Linear Probability Model

The most obvious idea is to let the probability of success be a linear function of x.

$$E(y_i|x) = P(y_i = 1|x = x_i) = \alpha + \beta x \qquad i=1,2,\ldots,n$$


2
1
0
-1
0
10
20
30
40
50
60
70
80
90
100
linear regression line

Linear Probability Model

The most obvious idea is to let the probability of success be a linear function of x.

$$E(y_i|x) = P(y_i = 1|x = x_i) = \alpha + \beta x \qquad i=1,2,\ldots,n$$

- The right hand takes values in the range $(-\infty, +\infty)$
- The left hand ranges between 0 and 1
- $var(y_i)=E(y_i)*(1-E(y_i))$ **the variance of Y is not constant, depend on *i*!**

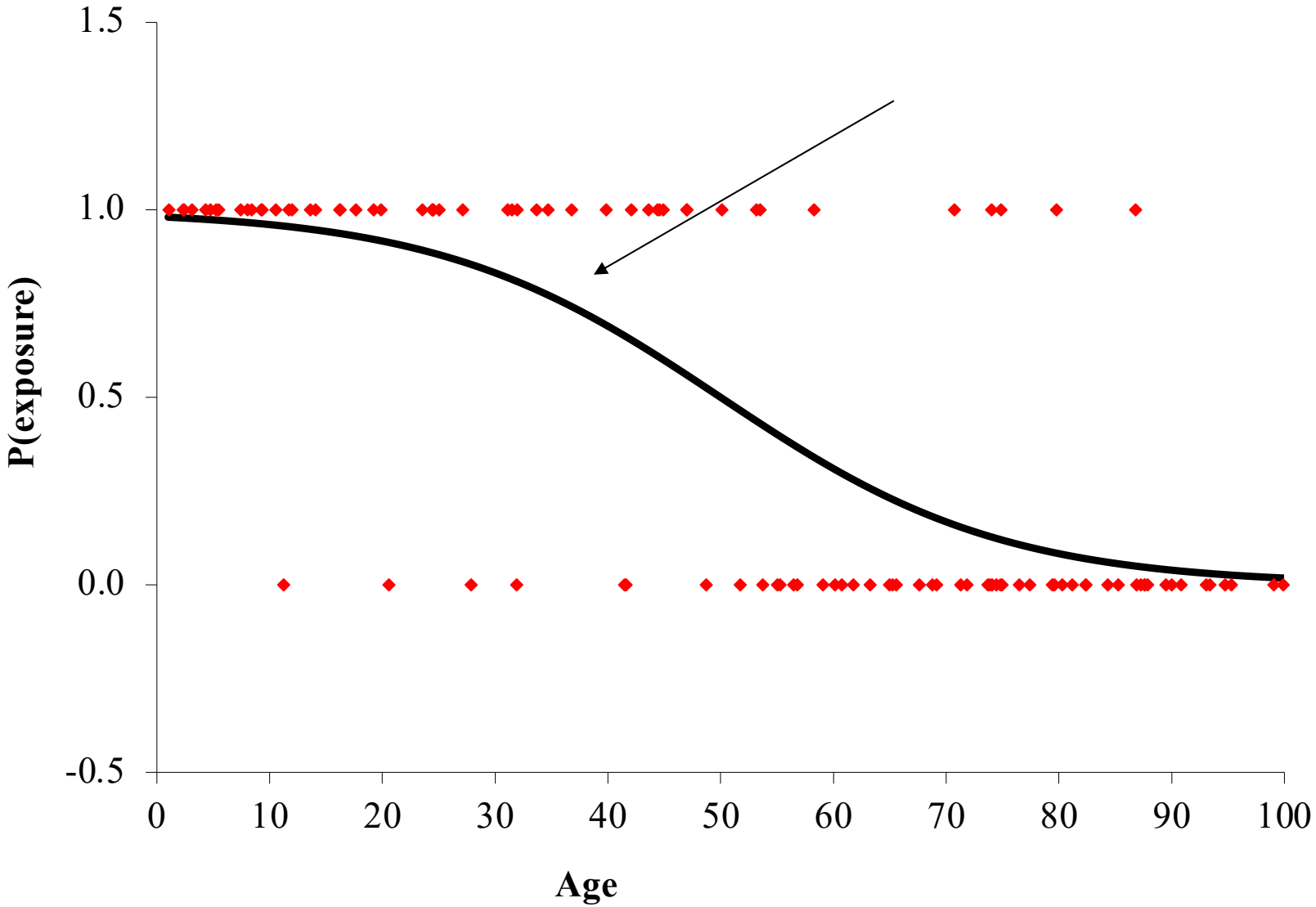
Problems with LPM

Moreover, LPM assumes that probabilities increase linearly with the explanatory variables

- Each unit increase in an X has the same effect on the probability of Y occurring regardless of the level of the X.
- In many situations we empirically see "diminishing returns" — changing p by the same amount requires a bigger change in x when p is already large (or small) than when p is close to 1/2. Linear models can't do this.

EX: Let p(x) denote the probability of buying a new house when annual family income is *x*. An increase of $50,000 in annual income would have less effect when *x i*s $1,000,000 (for which p(x) is near 1) than when *x i*s $50,000.

What to do?



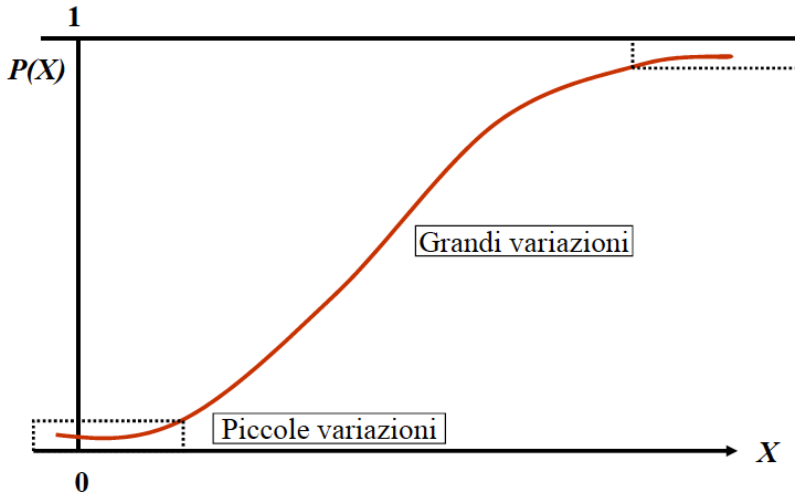
We therefore need to find a functional form for $F(x_i'\beta)$ for which:

$$\begin{cases} \lim_{x_i'\beta \to -\infty} F(x_i'\beta) = 0 \\ \lim_{x_i'\beta \to +\infty} F(x_i'\beta) = 1 \end{cases}$$

➤ Any distribution function (CDF) of a continuous random variable is suitable for this objective. Since $F(x_i'\beta)$:R→[0,1]

The logistic distribution and standard normal distribution are two candidates. The logistic gives rise to the *logit* model; the standard normal gives rise to the *probit* model

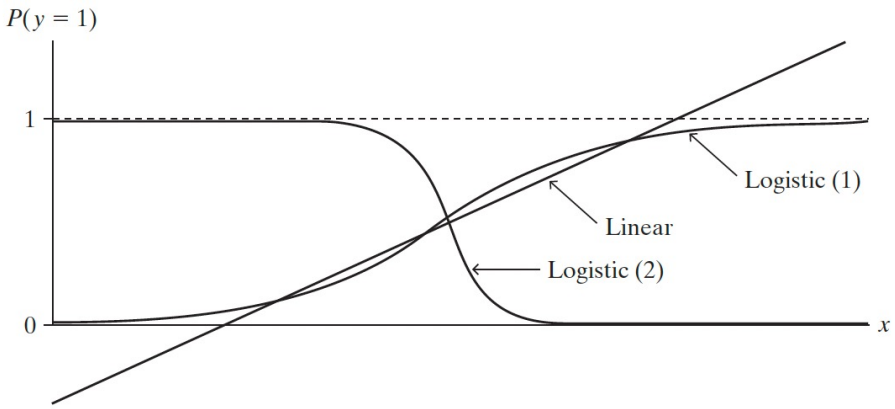
The logistic function



$$p_i = pr(y_i = 1) = \frac{e^{\alpha+\beta x_i}}{1 + e^{\alpha+\beta x_i}}$$

$$\begin{cases} \lim_{x_i'\beta \to -\infty} F(x_i'\beta) = 0 \\ \lim_{x_i'\beta \to +\infty} F(x_i'\beta) = 1. \end{cases}$$

β indicates if [P(y = 1)] increase (β > 0) or decrease (β < 0) for an increase x



3) Extension to GLMs

Motivation

Linear model

$$E[y|x] = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$

In the standard linear model we look for a linear predictor for expected value of the response variable.

BUT - as we have hinted -linear models have limitations

- Hypotheses are not always valid
- Necessity to study non – linear relationships
- Problems with heteroskedastic data

Solution: jump to the Generalized Linear Models (GLMs) where the linear predictor models a function of the Expected Value of the Y.

A generalized linear model (GLM) is a flexible generalization of ordinary linear regression.

What characterizes a GLM

The three elements that fully characterize a GLM:

- Specify the statistical distribution of the responses Y: f(Y|X) and mean μ
- The link function $g(\cdot)$
- A linear predictor $X\beta$

$$E(Y|X) = \mu = g^{-1}(X\beta)$$

Extend linear regression to a broader family of outcome variables. LMs can be regarded as a particular GLM

Family	*Default "Link"*	*Range of y_i*
gaussian	identity	$(-\infty, +\infty)$
binomial	logit	$\frac{0,1,\ldots,n_i}{n_i}$
poisson	log	$0, 1, 2, \ldots$
Gamma	inverse	$(0, \infty)$
inverse gaussian	$1/\mu^2$	$(0, \infty)$

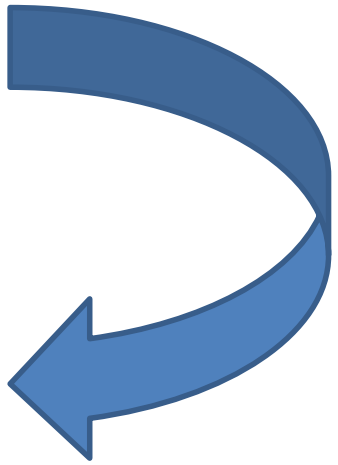
4) The LOGIT Model

The logit model

Y~Bin(n,p)

Probability form – the logistic function

$$p(Y=1)=\frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}}$$



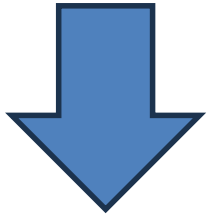
After a bit of manipulation we find:

$$\frac{p(Y=1)}{1-p(Y=1)}=e^{\alpha+\beta x}$$

ODDS (0, ∞)

By taking the logarithm of both sides we obtain:

$$log\left(\frac{p(Y=1)}{1-P(Y=1)}\right) = \alpha + \beta X$$



Log odds or logit, which is linear in X

In a logistic regression model, increasing X by one unit changes the log odds by β, or equivalently it multiplies the odds by exp(β). However, because the relationship between p(Y=1) and X is not a straight line, β does not correspond to the change in p(X) associated with a one-unit increase in X. The amount that p(X) changes due to a one-unit change in X will depend on the current value of X. Regardless of the value of X, if β is positive then increasing X will be associated with increasing p, if β is negative then increasing X will be associated with decreasing p.

Maximum likelihood estimates

- Logit and Probit models are nonlinear in the coefficients β therefore these models can't be estimated by the standard OLS (you could use non linear OLS but not efficient!)

- MLE is an alternative to OLS. It consists of finding the parameters values which is the most consistent with the data we have.

- In Statistics, the likelihood is defined as the joint probability to observe a given sample, given the parameters involved in the generating function. It is the joint probability distribution of the data, treated as a function of the unknown coefficients.

- One way to distinguish between OLS and MLE is as follows:

OLS adapts the model to the data you have: you only have one model derived from your data. MLE instead supposes there is an infinity of models and chooses the model most likely to explain your data.

The method of maximum likelihood yields values for the unknown parameters that maximize the probability of obtaining the observed set of data. MLE's are the parameter values "most likely" to have produced the data.

ML: we want estimates for β_0 and β_1 such that the predicted probabilities of y_i=1 for each individual i using the logistic function corresponds as closely as possible to the individual's observed probability.

The likelihood function is the probability that we get $y_1, y_2, \ldots, y_n$ from n draws. Since each draw is independent we use the multiplicative rule to calculate the joint probability (that is the likelihood function):

$$L(\beta_0, \beta_1) = \prod_{i=1}^{n} p(x_i)^{y_i} (1 - p(x_i))^{1-y_i}$$

In large samples, the MLE is:

- ✓ consistent
- ✓ normally distributed
- ✓ efficient (has the smallest variance of all estimators)

In large n, the sampling distribution of $\hat{p}^{MLE}$ is normally distributed.

coefficient interpretation

Relationship between age and probability of been exposed to an infection

Variable	**Coefficient value**	***p*-value**
Age	-0.05	0.00
Intercept	2.60	0.00

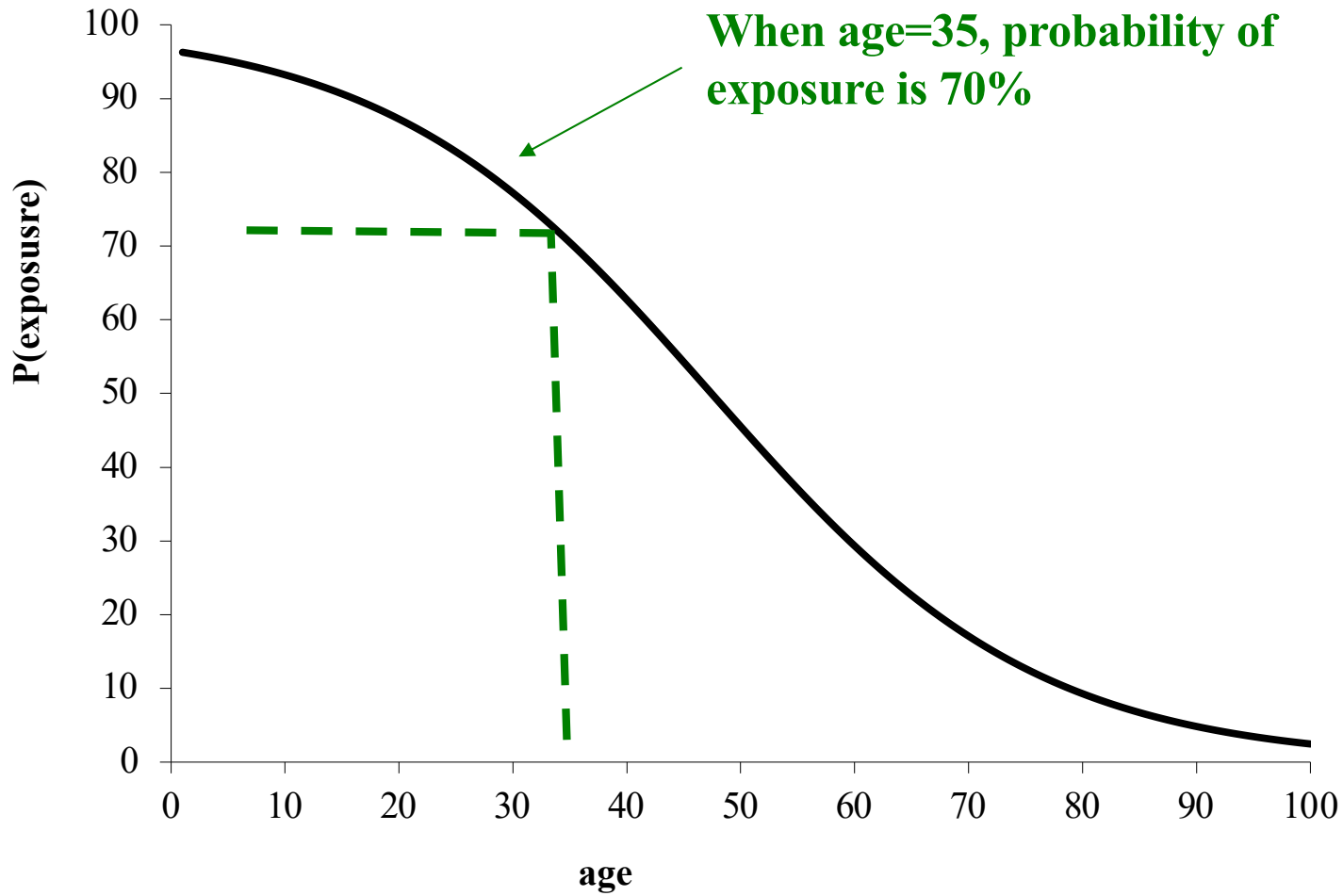
- It’s clear that age has a negative (and statistically significant) effect on exposure.
- But what does the -0.05 actually mean?

coefficient interpretation

- The other way of thinking about things is in terms of probabilities.
- If we rearrange the 'antilogged' equation then we work out what the probability (for a particular value of *X*) would be.
- The probability of a person with age=35 of exposure is thus 70%.

$$p(Y=1) = \frac{e^{2.60-0.05X}}{1+e^{2.60-0.05X}}$$

$$p(Y=1) = \frac{e^{2.60-0.05*35}}{1+e^{2.60-0.05*35}} = 70\%$$


When age=35, probability of exposure is 70%
P(exposusre)
100
90
80
70
60
50
40
30
20
10
0
0
10
20
30
40
50
60
70
80
90
100
age

Odds, Odds Ratios

- **The odds** of "success" is the ratio: $\omega = \dfrac{p}{1-p}$

- consider two groups with success probabilities:

$$p_1 \text{ and } p_2$$

- **The odds ratio** (OR) is a measure of the odds of success in group 1 *relative* to group 2

$$\theta = \frac{\omega_1}{\omega_2} = \frac{p_1/(1-p_1)}{p_2/(1-p_2)}$$

Odds for Independent Variable Groups

- We can compute the **odds of receiving a death penalty** by race:

	Blacks	Nonblacks	Total
Death sentence	28	22	50
Life imprisonment	45	52	97
Total	73	74	147

- The odds of receiving a death sentence if the defendant was Black =p/1-p= (28/73)/(1-(28/73)) = 0.6222
- The odds of receiving a death sentence if the defendant was not Black = 0.4231

The Odds Ratio Measures the Effect

- The impact of being black on receiving a death penalty is measured by the odds ratio which equals:

  = the odds if black / the odds if not black

  = 0.6222 / 0.4231 = 1.47

- Which can be interpreted as:
  - Blacks are 1.47 times more likely to receive a death sentence as non blacks
  - The risk of receiving a death sentence are 1.47 times greater for blacks than non blacks
  - The odds of a death sentence for blacks are 47% higher than the odds of a death sentence for non blacks. (1.47 - 1.00)
  - A one unit change in the independent variable race (nonblack to black) increases the odds of receiving a death penalty by a factor of 1.47.

Parameter interpretation

- Exponentiating both sides of the logit link function we get the following:

$$\left(\frac{p_i}{1-p_i}\right) = \text{odds} = \exp(\beta_0 + \beta_1 X_1) = e^{\beta 0}\, e^{\beta 1 X 1}$$

- The odds increase **multiplicatively** by $e^{\beta 1}$ for every 1-unit increase in x, x continuous.

- The odds at X = x+1 are e^{β} times the odds at X = x, furthermore, $(e^{\beta} -1)*100$ gives the percent increase in the odds of a success for each 1-unit increase in x if estimate of $\beta > 0$. $(1-e^{\beta})*100$ estimate of $\beta < 0$

Parameter interpretation

for *x* (*i.e. income in thousands of dollars-continuos*) and Y=buying a house or not. The estimate of β is equal to 0.012. If income is increased by \$10, this increases the odds of buying a house by about 13%

$$(e^{10\times 0.012}-1)\times 100\% = 12.75\%$$

- if estimate of $\beta > 0$ then percent increase in odds for a unit change in *x* is

$$(e^{\hat{\beta}}-1)\times 100\%$$

- if estimate of $\beta < 0$ then percent decrease in odds for a unit change in *x* is

$$(1-e^{\hat{\beta}})\times 100\%$$

Parameter interpretation

Example: for *x* (***dummy variable coded 1 for female, 0 male***), y satisfaction. The odds ratio is

$$\theta = \frac{p_f / (1 - p_f)}{p_m / (1 - p_m)} = \frac{\omega_f}{\omega_m} = \exp(\hat{\beta}_2) = \exp(0.67) = 1.95$$

- holding the other variable constants, women's odds of buying a house is nearly twice those of men.

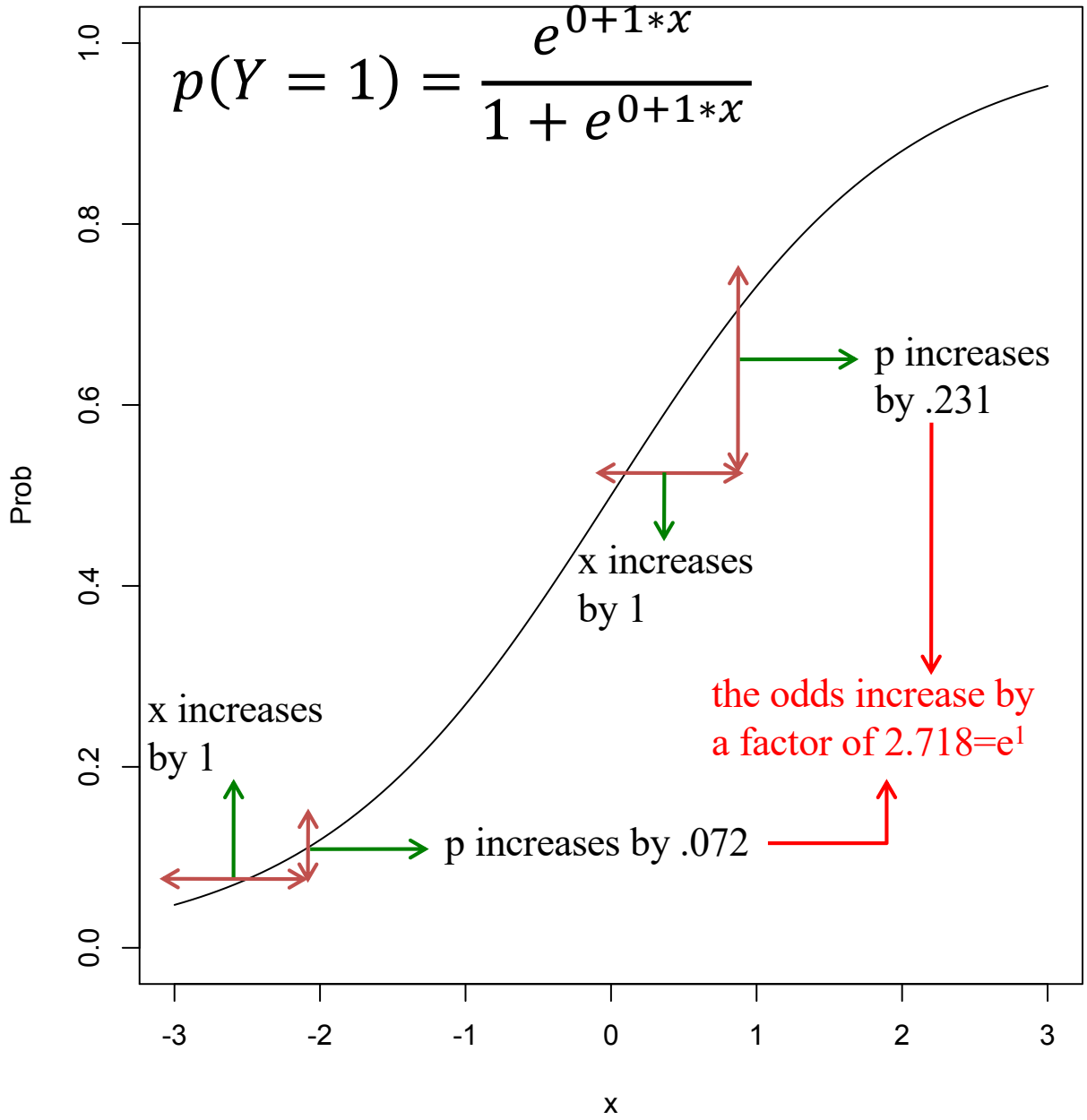
coefficient interpretation

- The probability of a person with age=36 of exposure decrease to 68%.

$$p(Y=1) = \frac{e^{2.60-0.05*36}}{1+e^{2.60-0.05*36}} = 68\%$$

- The probability of a person with age=45 of exposure decrease to 59%.

In terms of odds, the odds increase **multiplicatively** by e^{β} =1.1


$p(Y = 1) = \frac{e^{0+1*x}}{1 + e^{0+1*x}}$
Prob
1.0
0.8
0.6
0.4
0.2
0.0
-3
-2
-1
0
1
2
3
x
p increases by .231
x increases by 1
the odds increase by a factor of 2.718=e1
x increases by 1
p increases by .072

The model equation

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-1.6590	0.2880	33.1855	<.0001
AGE	1	0.0285	0.00838	11.5255	0.0007

$$P(y=1) = \frac{\exp(-1.659 + .0285x_1)}{1 + \exp(-1.659 + .0285x_1)}$$

$$\ln\left(\frac{\pi}{1-\pi}\right) = -1.659 + .0285x_1$$

$$\text{logit}(y) = -1.659 + .0285x_1$$

Ex: Y=1 not winner (of election)

logit nowin leader age scandal

```
Logit estimates                                   Number of obs   =       5036
                                                  LR chi2(3)      =     265.97
                                                  Prob > chi2     =     0.0000
Log likelihood = -1214.2961                       Pseudo R2       =     0.0987

------------------------------------------------------------------------------
       nowin |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      leader |  -1.029759    .516245    -1.99   0.046     -2.04158   -.0179374
         age |  -.0420528   .0035401   -11.88   0.000    -.0489913   -.0351143
     scandal |   2.839299   .3194128     8.89   0.000     2.213261    3.465337
       _cons |  -1.487179   .0859136   -17.31   0.000    -1.655566   -1.318791
------------------------------------------------------------------------------
```

A positive coefficient: the log-odds of not winning are decreasing as a function of being in a leadership position and with increasing of age and increase as a function of being involved in a scandal.

logistic nowin leader age scandal

```
Logit estimates                                   Number of obs   =       5036
                                                  LR chi2(3)      =     265.97
                                                  Prob > chi2     =     0.0000
Log likelihood = -1214.2961                       Pseudo R2       =     0.0987

------------------------------------------------------------------------------
       Nowin | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      leader |    .357093   .1843475    -1.99   0.046     .1298234    .9822225
         age |   .9588192   .0033943   -11.88   0.000     .9521895     .965495
     scandal |   17.10377   5.463164     8.89   0.000     9.145496    31.98723
------------------------------------------------------------------------------
```

Scandal: the odds of not winning of candidate who is involved in a scandal are about 17 times higher than a candidate who is not involved.

Age: (1-(.9588))*100=4% decrease in the odds of not winning for a unit increase in age

7) Goodness of fit

Goodness of Fit Measures

- In ML estimations, there is no such measure as the R^2

- However, the log likelihood measure can be used to assess the goodness of fit.
  - *NOTE* Given the number of observations, the better the fit, the higher the LL measures
- The philosophy is to compare two models looking at their LL values. One is meant to be the constrained model, the other one is the unconstrained model.

Goodness of Fit Measures

- A model is said to be constrained when the parameters associated with some variable are set to zero.
- A model is said to be unconstrained when the parameters associated with some variable are allowed to be different from zero.

- For example, we can compare two models, one with no explanatory variables, one with all our explanatory variables. The one with no explanatory variables implicitly assume that all parameters are equal to zero. Hence it is the constrained model because we (implicitly) constrain the parameters to be null.

The likelihood ratio test (LR test)

- The most used measure of goodness of fit in ML estimations is the likelihood ratio. The likelihood ratio is the difference between the unconstrained model and the constrained model. This difference is distributed χ^2.
- If the difference in the LL values is (no) important, it is because the set of explanatory variables brings in (un)significant information. The null hypothesis H_0 is that the model brings no significant information as follows:

$$LR = 2\left[\ln L_{unc} - \ln L_c\right]$$

- High LR values will lead the observer to reject hypothesis H_0 and accept the alternative hypothesis H_a that the set of explanatory variables does significantly explain the outcome.

Other usage of the LR test

- The LR test can also be generalized to compare any two models, the unconstrained one being *nested* in the constrained one.
- Any variable which is added to a model can be tested for its explanatory power as follows :
  - `logit [model contraint]`
  - `est store [name1]`
  - `logit [model non contraint]`
  - `est store [name2]`
  - `lrtest name2 name1`

The McFadden Pseudo R^2

- We also use the McFadden Pseudo R^2 (1973). Its interpretation is analogous to the OLS R^2. However it remains generally low.

- pseudo-R^2 also compares The likelihood ratio is the difference between the unconstrained model and the constrained model and is comprised between 0 and 1.

$$\text{Pseudo } R^2_{MF} = \frac{[\ln L_c - \ln L_{unc}]}{\ln L_{unc}} = 1 - \frac{\ln L_{unc}}{\ln L_c}$$

Model Fit Statistics

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	238.329	228.316
SC	241.607	234.873
-2 Log L	236.329	224.316

AIC Akaike Information Criterion. Used to compare non-nested models. Smaller is better. AIC is only meaningful in relation to another model's AIC value.

Checking assumptions

0. Independent data points

Problem: likelihood function is wrong otherwise + confidence intervals too small

1. Influential data points

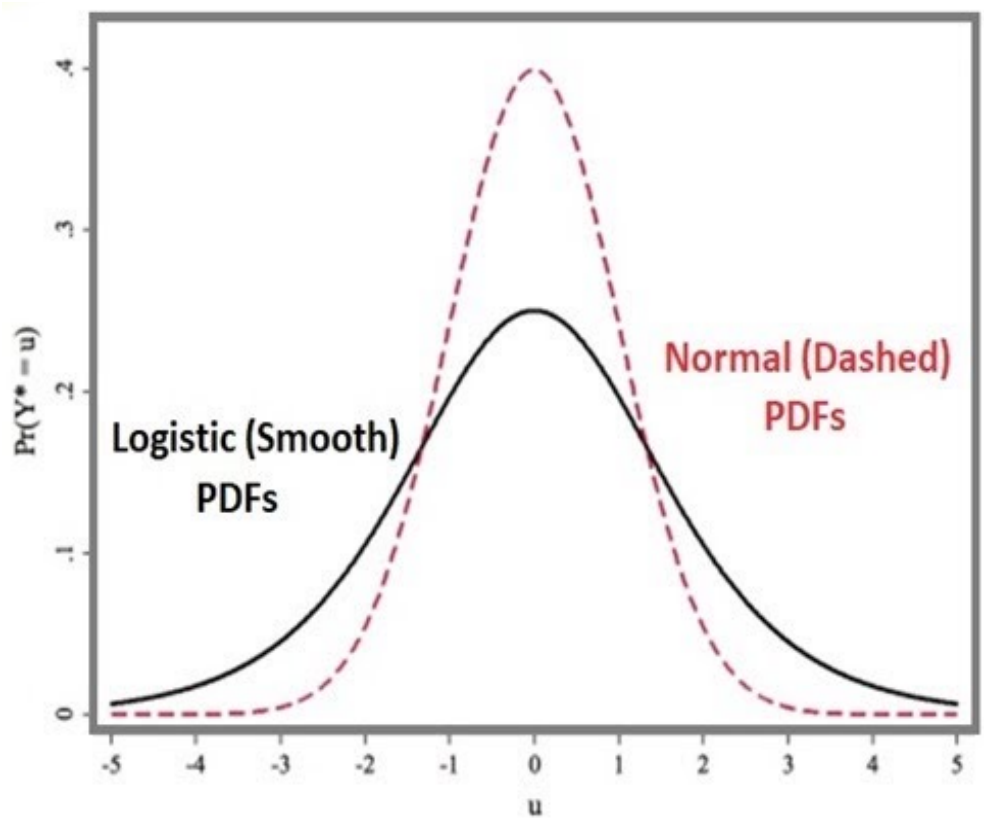
2. No multi-collinearity

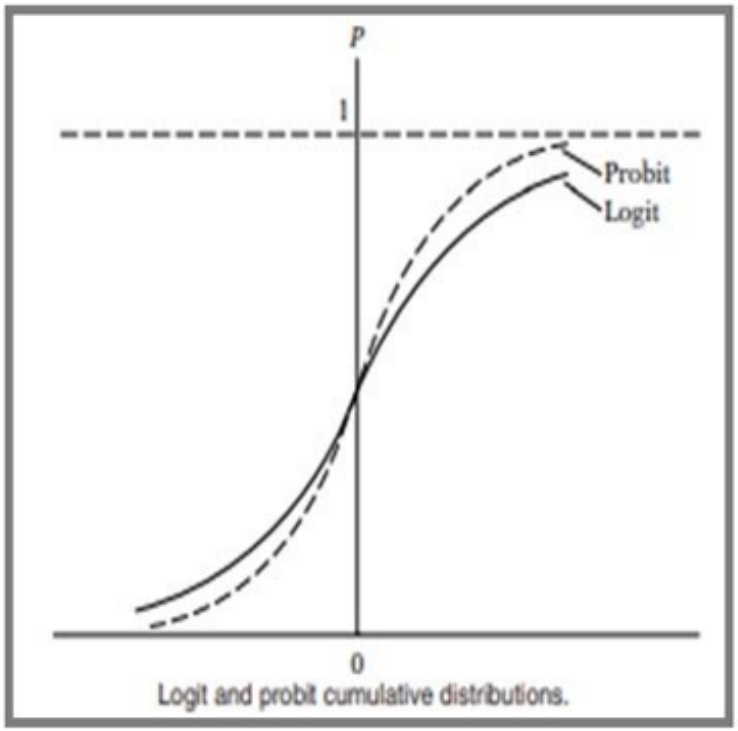
3. All relevant variables included

8) Other GLMs: the Probit model

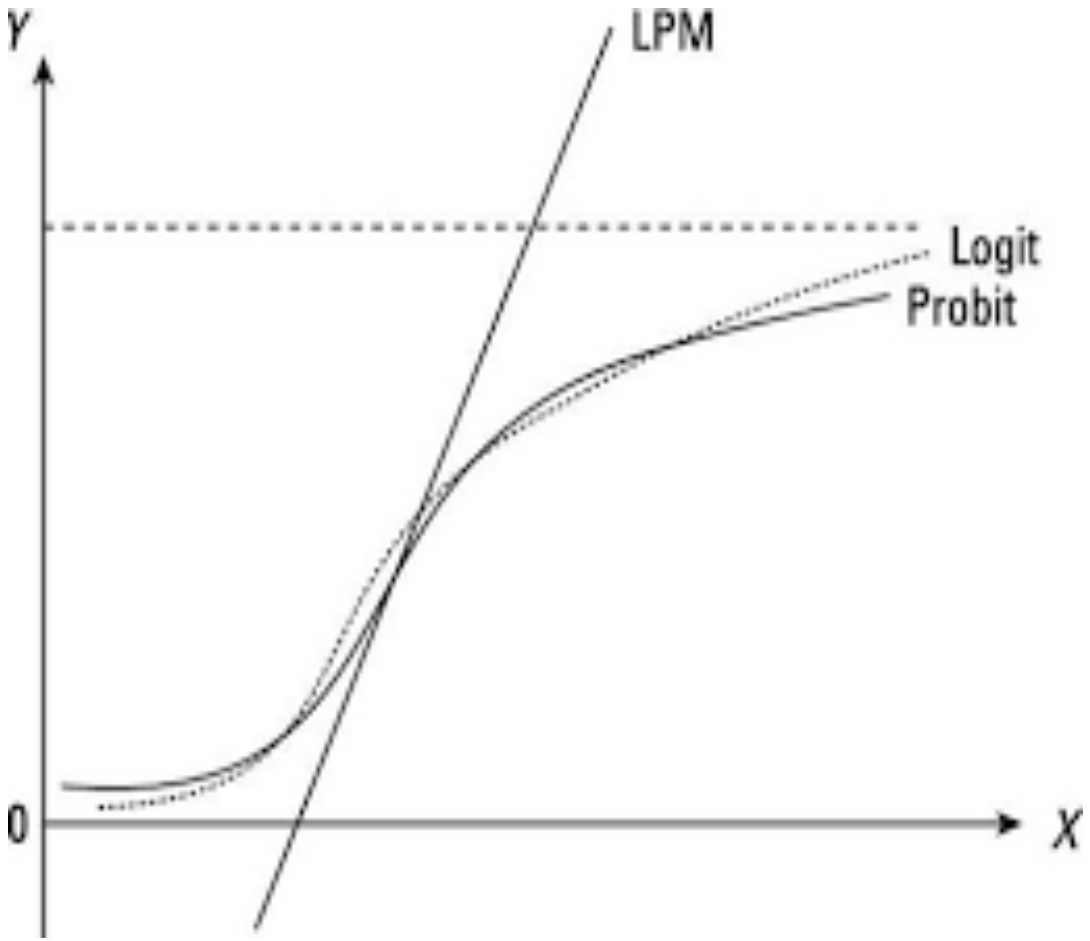
1 – Probit Model

An alternative model for binary outcomes:
Logit model: uses the CDF of the logistic distribution
Probit model: uses the CDF of the normal distribution --- > z-scores





Probit Model vs Logit Model vs LPM



- In the Logit model the interpretation focuses on probabilities, while in the Probit model coefficients are interpreted in terms of their effect on the change of z-scores.

- The logit model has heavier tails

- As a consequence, the logit model is more robust with respect to outliers.

Let Φ( ) be the cdf for standard normal distribution. Then for the i-th observation we have:

$$p_i = \Phi(\beta_0 + \beta_1 x_i)$$

Although there are a variety of ways to describe the effect of the independent variables, we focus on the calculation of marginal effects and discrete differences which are easier to interpret. To find the marginal effect of a continuous variable, *xk*, we take the derivative of the above equation with respect to *xk:*

$$\frac{dp}{dx_k} = \frac{d\Phi}{dx_k} = \phi(x\beta)\beta_k$$

where ϕ is the probability density function (pdf) of standard normal distribution, the derivative of Φ.

- This expression depends on not just β_k, but on the value of x_i and all other variables in the equation
- So to even calculate the impact of x_k on Y you have to choose values for all other covariates.
- Typical options are to set all variables to their means or their medians
- Another approach is to fix the xj and let x_k vary across its observed range of values

It is usually described as the “instantaneous rate of change of y with respect to x.” The visualization of this idea is to plot the slope of a tangent line at y. The slope of the line might be interpreted as the effect of “x” on “y” at that instant in time.

If the predictors are factors, then we calculate what are called "first-differences". This means we calculate the predicted outcome when a factor is set to a given level and then subtract the predicted outcome when the factor is set to its baseline level. We do this for all subjects and take the mean.

As example, if X is binary, the marginal effect then measures the probability difference between X = 1 and X = 0.

$$\begin{aligned} & \Pr(Y = 1|X = 1) - \Pr(Y = 1|X = 0) \\ = \; & \Phi(\hat{\beta}_0 + \hat{\beta}_1) - \Phi(\hat{\beta}_0) \end{aligned}$$

In STATA, the command **dprobit** reports the marginal effect, instead of $\hat{\beta}$.

Coefficients and marginal effects

- The marginal effect depends on the value of the predictors.
- Therefore, there exists an individual marginal effect for each person of the sample
- Different types of marginal effects can be calculated, one of the most used is:
  - Margin effect at the mean (*margins* in stata)

- Interpretation of marginal effects:
  - Continuous variables:
    A marginal (infinitesimal) change of the covariate changes the probability that the dependent variable takes the value one by b%, after "controlling"* for the other variables in the model. What the ME more or less tells you is that, if, say, X increased by some very small amount (e.g. .001). There is no guarantee that the instantaneous rate of change is similar to the change in P(Y=1) as X increases by one
  - Dummy variable like:
    The ME for categorical variables shows how P(Y=1) changes as the categorical variable changes from 0 to 1, after "controlling"* for the other variables in the model. With a dichotomous independent variable, the marginal effect is the difference in the adjusted predictions for the two groups, e.g. for blacks and whites.

* Usually at the average

Same syntax as REG but with probit

```
. * run probit model;
. probit smoker age incomel male black hispanic
> hsgrad somecol college worka;

Iteration 0:   log likelihood =  -9171.443
Iteration 1:   log likelihood =  -8764.068
Iteration 2:   log likelihood = -8761.7211
Iteration 3:   log likelihood = -8761.7208
```

Converges rapidly for most problems

Test that all non-constant Terms are 0

```
Probit estimates                  Number of obs   =      16258
                                  LR chi2(9)      =     819.44
                                  Prob > chi2     =     0.0000
Log likelihood = -8761.7208       Pseudo R2       =     0.0447

------------------------------------------------------------------------------
      smoker |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.0012684   .0009316    -1.36   0.173    -.0030943    .0005574
     incomel |   -.092812   .0151496    -6.13   0.000    -.1225047   -.0631193
        male |   .0533213   .0229297     2.33   0.020     .0083799    .0982627
       black |  -.1060518    .034918    -3.04   0.002      -.17449   -.0376137
    hispanic |  -.2281468   .0475128    -4.80   0.000    -.3212701   -.1350235
      hsgrad |  -.1748765   .0436392    -4.01   0.000    -.2604078   -.0893453
     somecol |   -.363869   .0451757    -8.05   0.000    -.4524118   -.2753262
     college |  -.7689528   .0466418   -16.49   0.000     -.860369   -.6775366
       worka |  -.2093287   .0231425    -9.05   0.000    -.2546873   -.1639702
       _cons |    .870543    .154056     5.65   0.000     .5685989    1.172487
------------------------------------------------------------------------------
```

Report z-statistics Instead of t-stats

Males are 1.7 percentage points more likely to smoke

Those w/ college degree 21.5 % points Less likely to smoke

margins, dydx(*) atmean

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
age	-.0003951	.00029	-1.36	0.173	-.000964	.000174	38.5474
incomel	-.0289139	.00472	-6.13	0.000	-.03816	-.019668	10.421
male*	.0166757	.0072	2.32	0.021	.002568	.030783	.39476
black*	-.0320621	.01023	-3.13	0.002	-.052111	-.012013	.111945
hispanic*	-.0658551	.01259	-5.23	0.000	-.090536	-.041174	.060709
hsgrad*	-.053335	.01302	-4.10	0.000	-.07885	-.02782	.335527
somecol*	-.1062358	.01228	-8.65	0.000	-.130308	-.082164	.268545
college*	-.2149199	.01146	-18.76	0.000	-.237378	-.192462	.329376
worka*	-.0668959	.00756	-8.84	0.000	-.08172	-.052072	.68514

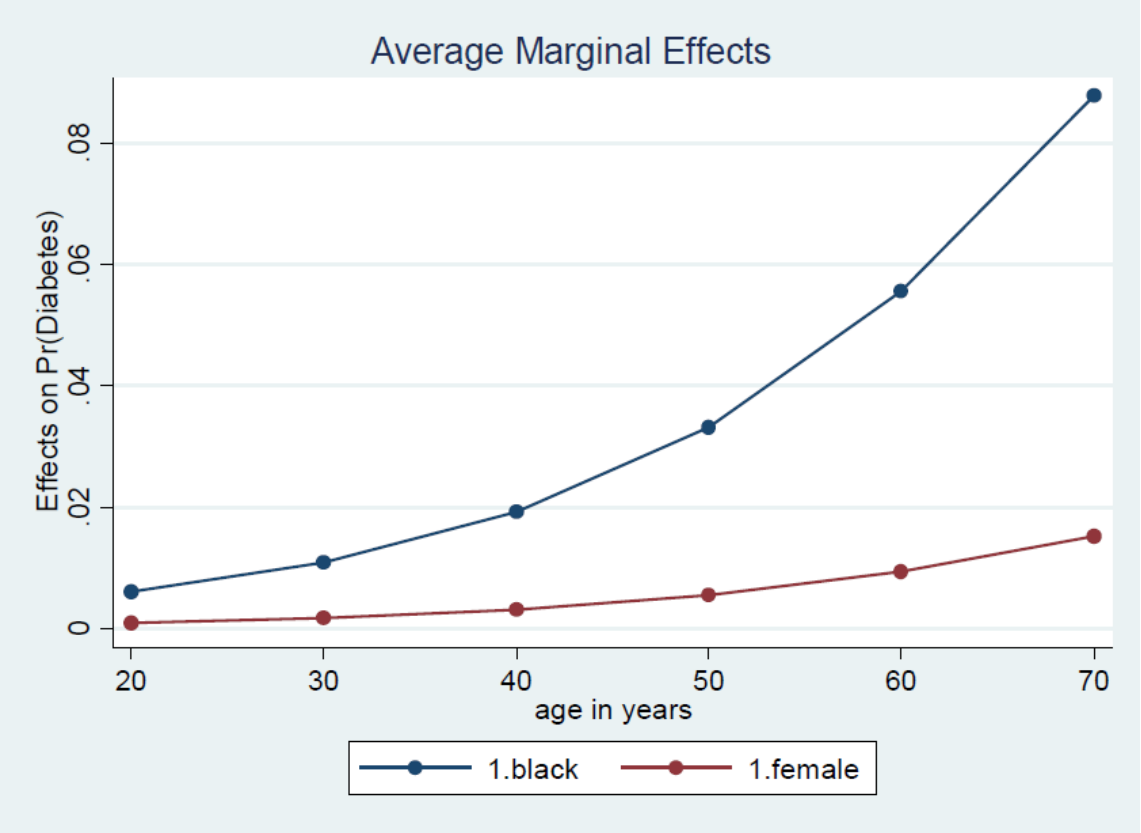
(*) dy/dx is for discrete change of dummy variable from 0 to 1

10 years of age reduces smoking rates by 4 tenths of a percentage point

10 percent increase in income will reduce smoking By .29 percentage points

Alternatively, choose ranges of values for one variable, and then see how the marginal effects differ across that range. EX: probability of having diabetes by age and race.

```
probit diabetes i.black i.female i.agegrp
margins black, at(age=(20 30 40 50 60 70))
marginsplot
```



Marginal effects - LOGIT

Marginal effects: change in the probability of the event occurring (e.g., the probability of success) due to a one-unit change in an independent variable, while holding all other variables constant.

How to compute marginal effects?

1.Fit the logit model model to your data.
2.Once you have estimated the model parameters (the β coefficients), you can calculate the marginal effects for each independent variable. The marginal effect for an independent variable **Xj** is calculated as:
3.Marginal Effect (Xj) = βj * p * (1 - p)
Where:
•**βj** is the coefficient of the independent variable **Xj**.
•**p** is the predicted probability of the event happening, given the values of all the independent variables. You can obtain this by plugging the values of your predictors into the logistic regression equation.