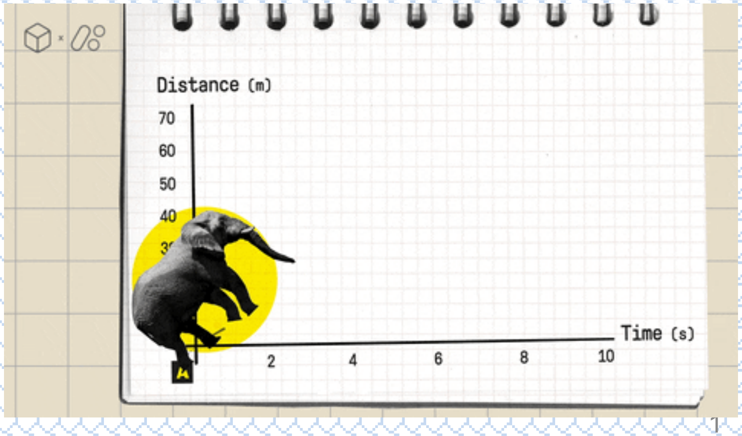
Biostatistics 200

Linear and Logistic Models
Multiple Regression & Logistic Regression



Continuous explanatory variable

```
.
. . logistic peth_audit_50 age_b

Logistic regression                               Number of obs   =        324
                                                  LR chi2(1)      =       5.32
                                                  Prob > chi2     =     0.0210
Log likelihood = -221.81862                       Pseudo R2       =     0.0119

------------------------------------------------------------------------------
peth_audit_50 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        age_b |   1.029989   .0133876     2.27   0.023     1.004081    1.056565
        _cons |   .3782632   .1590054    -2.31   0.021     .1659534    .8621885
------------------------------------------------------------------------------
```

- Interpretation of the coefficients: The odds ratio is for a one unit change in the predictor
- For this example 1.03 is the odds ratio for a one-year difference in age
- When the explanatory variable is a continuous variable, there is no equivalent tabular method without dividing the variable into groups (e.g. age groups)
- You need to use logistic regression when you want to model the relationship between a dichotomous outcome and a continuous explanatory variable

Continuous explanatory variable

- To find the OR for a 10-year change in age

```
.
. logistic peth_audit_50 age_b, coef

Logistic regression                               Number of obs   =        324
                                                  LR chi2(1)      =       5.32
                                                  Prob > chi2     =     0.0210
Log likelihood = -221.81862                       Pseudo R2       =     0.0119

-------------------------------------------------------------------------------
peth_audit_50 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
--------------+----------------------------------------------------------------
        age_b |   .0295481   .0129978     2.27   0.023     .0040729    .0550234
        _cons |   -.972165   .4203565    -2.31   0.021    -1.796049   -.1482814
-------------------------------------------------------------------------------
```

OR for a 10-year change in age = exp(10*0.029) = 1.34

.

Continuous explanatory variable

- To find the OR for a 10-year change in age
-

```
. lincom age_b*10

 ( 1)  10*[peth_audit_50]age_b = 0

------------------------------------------------------------------------------
peth_audi~50 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   1.343773   .1746612     2.27   0.023      1.04157    1.733658
------------------------------------------------------------------------------
```

- Another way to find the OR for a 10-year change in age
- gen age_10 = age_b/10

```
. logistic peth_audit_50 age_10
Logistic regression                             Number of obs   =        324
                                                LR chi2(1)      =       5.32
                                                Prob > chi2     =     0.0210
Log likelihood = -221.81862                     Pseudo R2       =     0.0119

-------------------------------------------------------------------------------
peth_audit_50 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
--------------+----------------------------------------------------------------
       age_10 |   1.343773   .1746612     2.27   0.023      1.04157    1.733658
        _cons |   .3782632   .1590054    -2.31   0.021     .1659534    .8621885
-------------------------------------------------------------------------------
```

Checking logit-linearity

```
. tab agecat

     agecat |      Freq.     Percent        Cum.
------------+-----------------------------------
         15 |         80       24.69       24.69
         25 |        136       41.98       66.67
         35 |         80       24.69       91.36
         45 |         28        8.64      100.00
------------+-----------------------------------
      Total |        324      100.00
```

```
 logistic peth_audit_50 i.agecat

Logistic regression                                 Number of obs   =        324
                                                    LR chi2(3)      =       8.22
                                                    Prob > chi2     =     0.0416
Log likelihood =  -220.3693                         Pseudo R2       =     0.0183

-------------------------------------------------------------------------------
peth_audit_50 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
--------------+----------------------------------------------------------------
       agecat |
          25  |   1.525822   .4390615     1.47   0.142     .8680944    2.681887
          35  |   2.254902    .728875     2.52   0.012     1.196699    4.248838
          45  |   2.575758    1.16071     2.10   0.036      1.06495    6.229893
              |
        _cons |         .6   .1385641    -2.21   0.027     .3815704    .9434694
-------------------------------------------------------------------------------
```

Logistic regression with >1 explanatory variable

- Example: 2 independent variables, X_1 a numerical variable, X_2 a dichotomous exposure variable
- logit(p) = log(p/(1-p)) = $\alpha + \beta_1 x_1 + \beta_2 x_2$

- When x_2=1 (exposed), holding x_1 constant,

  $\text{logit}(p_1) = \alpha + \beta_1 x_1 + \beta_2$

- When x_2=0 (not exposed), holding x_1 constant,

  $\text{logit}(p_0) = \alpha + \beta_1 x_1$

```
. logistic peth_audit_50 age_10 i.studyarm_n

Logistic regression                             Number of obs     =        324
                                                LR chi2(2)        =       6.00
                                                Prob > chi2       =     0.0498
Log likelihood = -221.48068                     Pseudo R2         =     0.0134

------------------------------------------------------------------------------
peth_audit_50 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
--------------+---------------------------------------------------------------
       age_10 |   1.328877   .1737573     2.17   0.030     1.028457    1.717053
 1.studyarm_n |   .8295213   .1886524    -0.82   0.411     .5311835     1.29542
        _cons |   .4248297    .188248    -1.93   0.053     .1782524    1.012498
------------------------------------------------------------------------------
```

- 0.83 is the odds ratio for PEth>=50 for being in the minimally assessed vs. standard study arm when you hold age constant

To sum up…

<u>OLS-Linear Regression</u>

- Response variable = Y
- Link function between the response variable and what is being modeled =1
- Eq. Y=a+bX
- Distr. = Normal
- Var. = Constant
- Y=a+bX
- Interpretation of b
- Omnibus test =F
- Testing of b = t
- Goodness of fit=R^2

<u>Logistic</u>

- Response variable = Y
- Link function between the response variable and what is being modeled= logit
- $Ln(\pi/1-\pi)=a+bX$
- Distr. = Binomial
- Var = Not Constant
- $Ln(\pi/1-\pi)=a+bX$
- Interpretation of b
- Omnibus test =-2LL
- Testing of b = Wald
- Goodness of fit= -2LL, pseudo-R^2

To sum up… assumptions

OLS

- Linearity
- Additivity
- Normal distribution
- Errors terms are independent
- OLS relies on the central limit theorem which states that ‘as we increase the sample size of the statistics that make up the sampling distribution,.., the form of the distribution becomes increasingly symmetrical and bell-shaped’

Logistic

- Non-Linearity … however assumes a linear relationship between the logit of the dependent and the independent variables.
- Additivity
- Non-normal distribution
- Errors terms are independent
- Maximum likelihood estimation relies on large-sample asymptotic normality which means that reliability of estimates decline when there are few cases for each observed combination of X variables.