Regression Models for a Probability of Response Outcome

- To now we have talked about regression models where the response variable Y was continuous and (approximately) normally distributed, or was an integer count. We now consider the case where the outcome of interest is a binary discrete variable, taking on only values 0 or 1. **For example:**,
  - infection event: if $Y_i = 1$ if with infection(case), 0 otherwise (unaffected).
  - disease recurrence: if $Y_i = 1$ if with disease recurrence, 0 otherwise (free of recurrence).
  - school drop-out: $Y_i = 1$ if dropped out, $Y = 0$ if not

Regression Models for Probability of Response

- Each individual's realization of their response can be thought of as a 'trial' in a binary outcome experiment, where the probability of $Y_i = 1$ is some value p. For a series of $i =, 1, 2, 3, ...n$ independent individuals, Y is a *binomially distributed random variable*, with probability of success $\mathrm{P}(Y = 1) = p$ per trial.

- For this series of independent individuals, we write $Y \sim Bin(n, p)$, where n is the total number of trials (individuals) and p is the probability of success ($Y = 1$) for each trial. We use this to calculate, for example, the probability of so many successes k out of n tries

Regression Models for Probability of Response

- Can we model this $\mathrm{E}(Y)$ as a linear function of predictors X? Identifying factors related to the probability of an event ($y = 1$) or predicting the probability would be of great value.
- Note that the sum of the successes (coded 1) and failures (coded 0) divided by n is the mean of Y, and also equals the proportion of 1's, equaling the $\mathrm{P}(Y = 1)$ Thus, $E(Y) = p$
- Thus, we should be able express the mean of Y, or the probability of an event, as a function of, or conditional on, some covariate X's. This is what we call regression modeling.

Regression Models for Probability of Response

Note that if we try to model $\mathrm{E}(Y)$ as a linear function of covariates X directly:

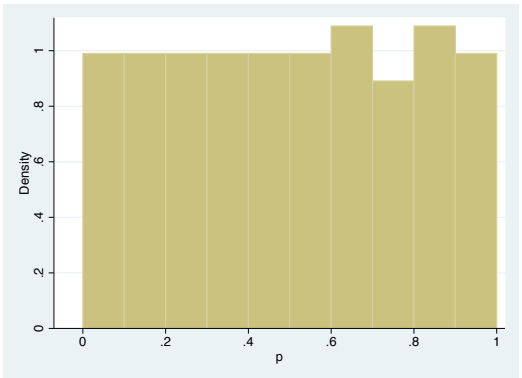
$$\mathrm{E}(Y|\mathbf{X}) = \mathrm{P}(Y = 1|\mathbf{X}) = \beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p$$

Several problems arise:

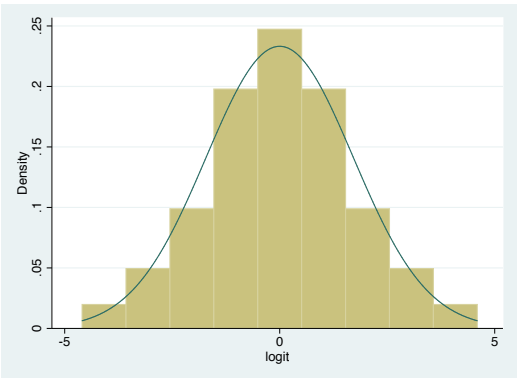
- The model as written above can readily produce probability estimates less than 0 or greater than 1.
- Y is not approx. normally distributed (individual Y itself and aggregated Y over n trials are not even continuous), as required for least squares model approach and theory to work.
- The variance of Y or the error term is not constant over X This arises from the fact that $Y \sim \mathrm{Bin}(n, p)$. For binomial random variables, $\mathrm{E}(Y) = np$ and $\mathrm{var}(Y) = np(1 - p)$. So if p varies conditional on X, so does the variance.

Regression Models for Probability of Response
A Transformation

- probability p has range $[0, 1]$. (Natural) logarithm of p has range $(-\infty, 0)$
- Define $\frac{p}{1-p}$ as the *odds*. Has range $(0, \infty)$
- then log(odds) has range $(-\infty, \infty)$ and looks like this over a range of p



(a) raw p



(b) logit transform

Regression Models for Probability of Response

- The following transform does permit us to formulate $\mathrm{P}(Y = 1)$ (actually, a function of it) as a linear function of predictor variables X

$$= \log_e \left(\frac{\mathrm{P}(Y = 1|\mathbf{X})}{1 - \mathrm{P}(Y = 1|\mathbf{X})} \right) = \beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p.$$

From here, after some algebra

$$\mathrm{P}(Y = 1|\mathbf{X}) = \frac{\exp(\beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p)}{1 + \exp(\beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p)}$$

This is a model relating the probability of $Y = 1$ to predictor variables. Observations with same values of Xs will have the same probability of 'success' ($\mathrm{P}(Y = 1)$) under this model. Note that on this probability scale, it is not a linear model.

Regression Models for Probability of Response

The **logit transform** on response variable Y coming from a binomial distribution is defined as

$$\text{logit}(p) = \log_e \left(\frac{p}{1-p} \right).$$

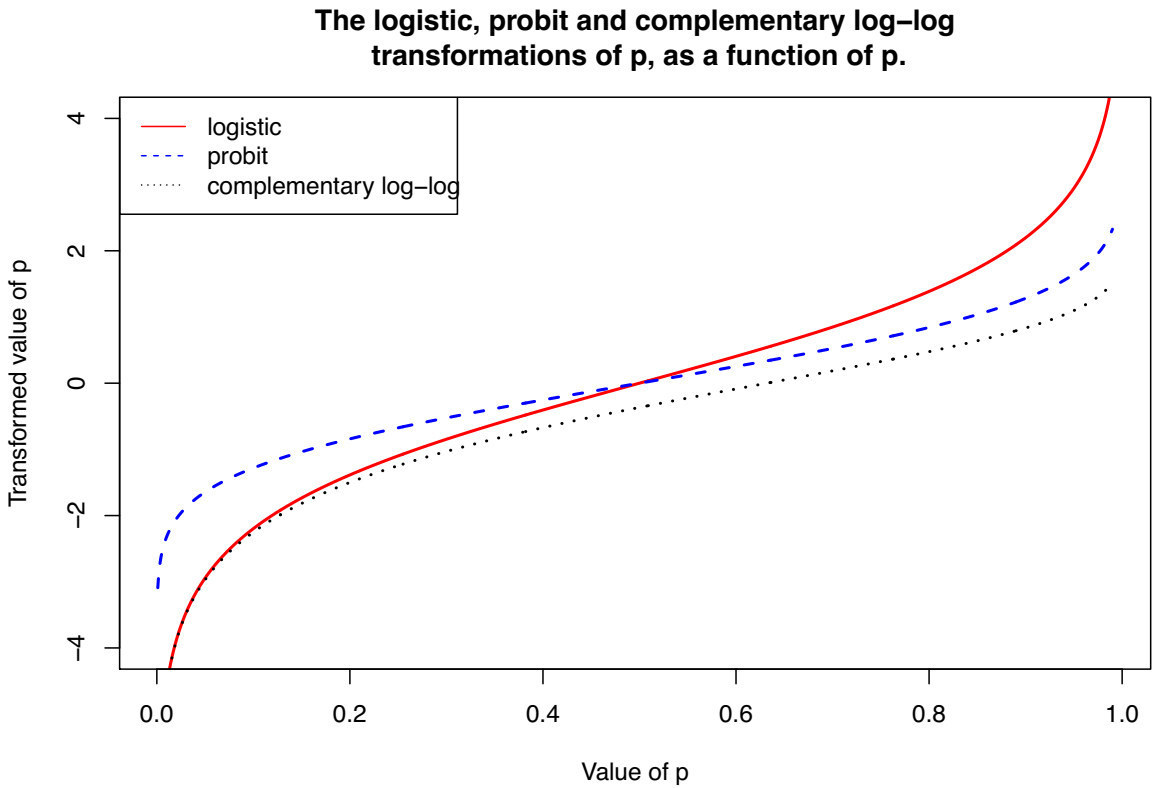
Models predicting the logit as a linear function of X are known as **logistic regression** models

- The name derives from the fact that under this model, we can write

$$p = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

and the right-hand side is called a logistic function (it is also a probability distribution). Note that it produces valid probability values for any values of X

Transformations of probability p to a Real Line Scale



Analysis of Probability of Response Data

- Taking a step back, how do we typically summarize and analyze data with binary response? We should have analogous basic methods in logistic regression, just as in linear regression, where we can reproduce the two-sample t-test for comparison of means via SLR.

  If the questions is whether the probability of a binary event (equal to the proportion responding) differs between, say two groups, we have

  - Test for difference of proportions between two independent groups - normal-based test, exact binomial test, etc
  - Test for independence in a 2x2 contingency table - χ^2 test, Fisher's exact test

- We also have associated effect measures - difference or ratio of proportions, and another summary that we will use

Analysis of Probability of Response Data via Table Summary

- Recall the basic set-up for the **2 $\times$ 2 Table**: This is for an epidemiological study of disease in relation to some exposure

	Exposure		Total
Disease Status	yes	no	
case/affected	a	b	a+b
non-case/unaffected	c	d	c+d
Total	a+c	b+d	a+b+c+d=N

Table Summary - Measures

	Exposure		Total
Disease Status	yes	no	
case/affected	a	b	a+b
non-case/unaffected	c	d	c+d
Total	a+c	b+d	a+b+c+d=N

- We can summarize effects via:
  - **Difference of Proportions, Absolute Risk Difference**:

$$\text{risk difference} = \frac{a}{a+c} - \frac{b}{b+d}$$

  - **Relative Risk or Risk Ratio**:

$$\text{relative risk} = \frac{a/(a+c)}{b/(b+d)}$$

– **Another measure - the Odds Ratio**: Ratio of odds of event $(Y = 1)$ (alternative way of expressing probabilities) in each exposure group

Note that in the *exposed* group

$$\text{odds} = \frac{p}{1-p} = \frac{a/(a+c)}{c/(a+c)} = a/c$$

Case non-case

and in the *unexposed* group

$$\text{odds} = \frac{p}{1-p} = \frac{b/(b+d)}{d/(b+d)} = b/d$$

Case non-case

Then

$$\text{ratio} = \frac{a/c}{b/d} = \frac{ad}{bc}$$

Exposed

unexposed

Is the **odds ratio** or relative odds. It is the cross-product of the table cell values.

Odds, Odds Ratio

- **What are odds?** Just another metric for expressing probabilities, expressed as

Sum=1

$$O = \frac{\text{proportion of successes}}{\text{proportion of failures}} = \frac{p}{1-p}$$

- Example. What were the overall odds of survival on the Titanic? 38% of passengers survived. Therefore:

$$\text{Odds of survival} = \frac{0.38}{1-.38} = \frac{0.38}{0.62} = 0.61$$

- In betting, 1:1 (1/2) odds means 50% probability, 2:1 odds (2/3) is 67% probability, 1:5 odds (1/6) is 16.7%, etc

Logistic Regression
The Log Odds

- **Example** To motivate the example, suppose we have mortality (outcome) and hospital admission type (exposure, X variable)
- **Define:**
  $p_0 =$ risk of death in the population of elective admission patients
  $p_1 =$ risk of death in the population of emergency admission patients
- For the elective admission population, the "odds" of death is:

$$\frac{p_0}{1-p_0} = \frac{b/(b+d)}{d/(b+d)} = \frac{b}{d}$$

  and

$$\log\left\{\frac{p_0}{1-p_0}\right\}$$

  is the "log odds" of death, or the "logit" of p_0

Logistic Regression
The Log Odds

- For the emergency admission population, the "odds" of death is:

$$\frac{p_1}{1-p_1} = \frac{a/(a+c)}{c/(a+c)} = \frac{a}{c}$$

  and

$$\log\left\{\frac{p_1}{1-p_1}\right\}$$

  is the "log odds" of death, or the "logit" of p_1

Logistic Regression
The Log Odds Ratio

- **Why log odds?** Rather than model on either the probability or odds ratio scale (restricted valid range), we rely on equivalent unconstrained scale, and that has better properties
- It turns out that we can take the difference of log odds (in the same way that we work with a difference of means in linear regression) as a basic effect measure

$$\log(p_1/(1-p_1)) - \log(p_0/(1-p_0))$$

but note that

$$\log(p_1/(1-p_1)) - \log(p_0/(1-p_0)) = \log\left\{\frac{p_1/(1-p_1)}{p_0/(1-p_0)}\right\}$$

from basic rules of logarithms

Logistic Regression
The Log Odds

Then

$$\log\left\{\frac{p_1/(1-p_1)}{p_0/(1-p_0)}\right\} = \log(OR)$$

Exponentiate for OR.

is the population "log odds ratio" of death, comparing the emergency admissions to the elective admissions. We label it β_1 and want our model to estimate it.

To reiterate: the *difference of log odds* is same as the log of the odds ratio, called *log odds ratio* for short. This is the response variable and the metric on which covariate effects will be expressed.

Logistic Regression Model

- Define an indicator variable for admission type:

$$x = \texttt{typ} \;\; = \begin{cases} 1 \text{ if emergency} \\ 0 \text{ if elective} \end{cases}$$

- Since the log odds is simply a function of the probability of interest ($\Pr$(Y=1)), a model that predicts the log odds (logit) is equivalent to a model that predicts the probability.

- Here is the logistic regression model for the risk of death in ICU as a function of admission type:

$$\log \left\{ \frac{p}{1-p} \mid x \right\} = \beta_0 + \beta_1 x$$

Logistic Regression Model (cont.)

- Recall that a linear regression model describes the mean of a random variable (Y) as a linear combination of some other variables (X's)

- A logistic regression model describes the log odds of a probability (that a binary variable Y takes on the value 1) as a linear combination of some other variables (X's –covariates or predictors)

- Because error structure is Binomial, maximum likelihood estimation estimation is used - this is a GLM

Logistic Regression - Estimation and Model Interpretation

- To fit this model in stata:

```
.* the data: (admission type: 0=elective, 1 = emergency
.* vital status: 0 = alive, 1 = died)
. tab sta typ

     Vital |    Admission Type
    Status |         0          1 |     Total
-----------+----------------------+----------
         0 |        51        109 |       160
         1 |         2         38 |        40
-----------+----------------------+----------
     Total |        53        147 |       200

. ** Fit logistic regression model
. logit sta typ
```

$y \quad x$

```
Iteration 0:   log likelihood = -100.08048
Iteration 1:   log likelihood = -93.425171
. . .
Iteration 4:   log likelihood = -92.524467
Iteration 5:   log likelihood = -92.524467

Logit estimates                   Number of obs =    200
                                  LR chi2(1)    =  15.11
                                  Prob > chi2   = 0.0001
Log likelihood = -92.524467       Pseudo R2     = 0.0755
------------------------------------------------------------------------------
         sta |      Coef.   Std.Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         typ |      2.1849   .74504     2.93   0.003     .72464    3.64519
       _cons |     -3.2386   .72083    -4.49   0.000    -4.6514   -1.82586
------------------------------------------------------------------------------
```

Logistic Regression Model (cont.)

- In R, we use the GLM module. For the logit model, we specify the distribution family as binomial

```
> logitm = glm(sta ~ typ, data=icu, family="binomial")
> summary(logitm)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.7734  -0.7734  -0.7734  -0.2774   2.5601

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -3.2387     0.7206  -4.495 6.97e-06 ***
typ           2.1849     0.7448   2.934  0.00335 **
---
(Dispersion parameter for binomial family taken to be 1)
```

```
    Null deviance: 200.16  on 199  degrees of freedom
Residual deviance: 185.05  on 198  degrees of freedom
AIC: 189.05

Number of Fisher Scoring iterations: 5
```

- **Note:** results (βs) in Stata and R both given in *log odds ratio metric* by default

Logistic Regression Model Interpretation (cont.)

- Interpretation of regression coefficients: Here,

$$\hat{\beta}_0 = -3.24 \quad \text{and} \quad \hat{\beta}_1 = 2.18$$

The model

$$\log\left\{\frac{p}{1-p}\right\} = -3.24 + 2.18(typ)$$

predicts the log odds (and odds) for each exposure group:

when X=0 (elective group): log odds or logit $= -3.24$, then

when typ=0

$$\text{odds} = e^{-3.24} = 0.039$$

when X=1 (emergency group):

when typ=1

$$\text{odds} = e^{-3.24+2.18} = 0.346$$

Logistic Regression Model Interpretation (cont.)

- Back to the data - table entries are n, row prob., column prob

Vital Status	Admission Type 0	1	Total
0	51	109	160
	31.87	68.12	100.00
	96.23	74.15	80.00
1	2	38	40
	5.00	95.00	100.00
	3.77	25.85	20.00
Total	53	147	200
	26.50	73.50	100.00
	100.00	100.00	100.00

Pearson chi2(1) = 11.8663 Pr = 0.001

- **From this table, note that odds are**

Unexposed elective: $2/51 = 0.039$

Exposed emergency $38/109 = 0.346$

Logistic Regression - Estimation and Model Interpretation

1. • **Note also:** from the table OR $= \frac{51x38}{2x109} = 8.89$

2. • **From the model**, the ratio of predicted odds (i.e odds ratio) is

Exposed

$$\frac{0.346}{0.039} = 8.89 = \text{OR}$$

Unexposed

3. • But also note that the difference of log odds $\equiv$ log odds ratio is **directly given** by the model. An estimate of the odds ratio is therefore:

$$\widehat{OR} = e^{\hat{\beta}_1} = e^{2.18} = 8.89$$

So, the β coefficient for `typ` gives the log odds ratio for type of admission. $e^{\hat{\beta}}$ gives the OR estimate (relative odds of death, emergency vs. elective)

Logistic Regression - Estimation and Model Interpretation - probabilities

- The model was intended to predict probabilities. Does it do this? Use the logistic function to obtain

$$\Pr(Y = 1|X = 0) = \frac{e^{-3.24}}{1 + e^{-3.24}} = 0.0377$$

$$\Pr(Y = 1|X = 1) =$$

check these probabilities against the 2x2 table . . .

Logistic Regression - Estimation and Model Interpretation - probabilities

Vital Status	Admission Type 0	1	Total
0	51	109	160
	31.87	68.12	100.00
	96.23	74.15	80.00
1	2	38	40
	5.00	95.00	100.00
	3.77	25.85	20.00
Total	53	147	200
	26.50	73.50	100.00
	100.00	100.00	100.00

Pearson chi2(1) = 11.8663 Pr = 0.001

. The model predicts $\mathrm{P}(Y = 1)$ by admission type, equaling the proportion of deaths in each group.

Model Interpretation (cont.)

- The **odds ratio** calculation is directly given in STATA using the or option:

```
. logit sta typ , or
. . .
Logit estimates                   Number of obs = 200
                                  LR chi2(1)    = 15.11
                                  Prob > chi2   =  0.0001
Log likelihood = -92.524467       Pseudo R2     =  0.0755
------------------------------------------------------
sta | Odds Ratio  Std.Err.  z   P>|z| [95% Conf.Int.]
----+-------------------------------------------------
typ | 8.8899      6.6234  2.93  0.003  2.0640 38.2899
------------------------------------------------------
```

***note:** unless predicting probabilities, or want the baseline odds, we don't need β_0. Output requesting 'or' only gives odds ratios

Logistic Regression Model Interpretation (cont.)

- Test of the hypothesis that ICU death is associated with type of admission:

$$H_0 : OR = e^{\beta_1} = 1$$

is same as

$$H_0 : \log(OR) = \beta_1 = 0$$

The test statistic reported in the `stata` output is :

$$Z = \frac{\hat{\beta}_1}{\hat{\text{se}}(\hat{\beta}_1)} = \frac{2.185}{.745} = 2.93$$

and the P-value is

$$P\text{-value} = \Pr\{|Z| > 2.93\} = 0.003$$

Model Interpretation (cont.)

```
------------------------------------------------------------------------------
         sta |      Coef.   Std.Err.      z    P>|z|     [95% Conf. Int.]
-------------+----------------------------------------------------------------
         typ |      2.1849   .74504     2.93   0.003     .72464   3.64519
       _cons |     -3.2386   .72083    -4.49   0.000    -4.6514  -1.82586
------------------------------------------------------------------------------
```

- From the model run, a confidence interval for β_1 is constructed as:

$(0.725, 3.65)$

Exponentiating the boundaries of the CI gives a CI for the OR:

$(e^{0.725}, e^{3.65})$

`stata` does this when using "`logit . . . , or`" option

or "`logistic`"

Logistic Regression Model - Global Model Test and Fit

```
. logit sta
...
Logistic regression                                     Number of obs     =        200
                                                        LR chi2(0)        =       0.00
                                                        Prob > chi2       =          .
Log likelihood = -100.08048                             Pseudo R2         =     0.0000
------------------------------------------------------------------------------
         sta |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |  -1.386294   .1767767    -7.84   0.000     -1.73277   -1.039818
------------------------------------------------------------------------------
```

This model produces the odds of death overall (exp(-1.386) $=$ .25 $=$ 40/160). Contrast of the log likelihood here and for the model with admission type gives a global model test:

$$D = -2(-100.08 - (-92.52)) = 15.12$$

This value appears in model runs with admission type as predictor. This test is obviously more useful with multiple covariates.

Logistic Regression Model Interpretation (cont.)

- Overall conclusion: ICU death is significantly and positively associated with emergency (versus elective) admission to the ICU (P-value $= 0.003$). $\widehat{OR} = 8.89$, 95% CI: $[2.06, 38.3]$
- Note: test on β_1 in this model is equivalent to this test:

  H_0 : independence of row and column frequencies in 2x2 table

  vs

  H_A : association of row and column frequencies in 2x2 table

  via the χ^2 test. Advantage here is that we have an effect measure estimate with associated CI
- Logistic regression also can be readily used with predictors that are not binary or categorical, where tables cannot be readily formed. Of course, multiple predictor variables are permitted

Logistic Regression
Multiple and Continuous Covariates

- We can expand the analysis to account for multiple predictors. Suppose that we were interested in the association between both age ('continuous') and admission type, and ICU mortality
- Then as before we have a (0,1) indicator for admission type

$$X_1 = \texttt{typ} \;\; = \begin{cases} 1 \text{ if emergency} \\ 0 \text{ if elective} \end{cases}$$

and define

$$X_2 = \text{age (yrs)}$$

and propose a logistic regression model for death as a function of these two covariates:

$$\log\left\{\frac{p}{1-p} \mid \mathbf{X}\right\} = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

- What is the interpretation of β_2?
- To answer this, lets compare the log odds of death of a 50 year old to a 49 year old subject, both of whom were admitted on an emergency basis

Therefore: β_2 is the log odds ratio for death for a 50 year old compared to a 49 year old, both of whom had emergency admissions to the ICU. Calculation would be same for two elective admitted patients.

Logistic Regression
Multiple and Continuous Covariates

```
. ** Fit multiple logistic regression model
. logit sta age typ , or
Logit estimates                        Number of obs   =   200
                                       LR chi2(2)      =   27.09
                                       Prob > chi2     =   0.0000
Log likelihood = -86.53782             Pseudo R2       =   0.1353

----------------------------------------------------------------
  sta | Odds Ratio  Std.Err.   z    P>|z|   [95% Conf.Int.]
------+---------------------------------------------------------
  age |   1.03460  .0110644  3.18  0.001   1.0131 1.0565
  typ |  11.62939  8.751927  3.26  0.001   2.6605 50.8329
----------------------------------------------------------------
```

Logistic Regression
Multiple and Continuous Covariates

- Same model in STATA, on the log odds ratio scale . . .

```
. logit sta age typ
Logit estimates                   Number of obs =    200
                                  LR chi2(2)    =  27.09
                                  Prob > chi2   = 0.0000
Log likelihood = -86.537821       Pseudo R2     = 0.1353
------------------------------------------------------------------------------
         sta |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |    .034016   .01069     3.18   0.001     .013055    .05497
         typ |   2.45354    .75257     3.26   0.001     .978525   3.92854
       _cons |  -5.50876   1.03351    -5.33   0.000    -7.53440   -3.48311
------------------------------------------------------------------------------
```

Logistic Regression - Multiple Covariates

- Some important notes about β_2:
  - As in linear regression, β_2 represents the increment in Y for a one unit change in X. Here, it is the log odds ratio for death comparing two subjects who differ in age at admission by one year
  - Also as in linear regression, we can say that β_2 is the log odds ratio for death comparing two subjects who differ in age at admission by one year **and** who are of the same admission type (be it emergency or elective)
  - This is what we mean by an **adjusted effect**. β_2 is the log odds ratio comparing two subjects who differ in age at admission by one year, **adjusted for admission type**
- What is the interpretation of β_1?

  A: difference of log odds (emergency vs. elective) for two subjects of a given age. e^{β_1} is the odds ratio

Logistic Regression
Prediction from the Model

We can predict the probabilities after executing

```
. predict prob_d
(option pr assumed; Pr(sta))

. tab prob_d

    Pr(sta) |      Freq.     Percent        Cum.
------------+-----------------------------------
   .0074175 |          2        1.00        1.00
   .0076722 |          1        0.50        1.50
   .0103916 |          1        0.50        2.00
   .0190021 |          1        0.50        2.50
   .0196467 |          1        0.50        3.00
   .0203127 |          1        0.50        3.50
   .0232047 |          1        0.50        4.00
   .0239884 |          1        0.50        4.50
   .0247978 |          1        0.50        5.00
   .0256339 |          4        2.00        7.00
. . . .
    .442287 |          1        0.50       94.50
```

```
   .4506935 |          1        0.50       95.00
   .4591281 |          1        0.50       95.50
   .4760631 |          2        1.00       96.50
   .4845538 |          3        1.50       98.00
   .4930533 |          1        0.50       98.50
   .5100596 |          2        1.00       99.50
   .5185565 |          1        0.50      100.00
------------+-----------------------------------
      Total |        200      100.00
```

Note: There will be a unique probability predicted at each combination of age and admission type (90 of these). Model does not output 0's and 1's for each patient, but rather a probability of being a 1 (event)

Logistic Regression
Multiple and Continuous Covariates

```
. logit sta age typ , or
Logit estimates                      Number of obs   =   200
                                     LR chi2(2)      =   27.09
                                     Prob > chi2     =   0.0000
Log likelihood = -86.53782           Pseudo R2       =   0.1353
---------------------------------------------------------------
  sta | Odds Ratio  Std.Err.  z   P>|z|  [95% Conf.Int.]
------+--------------------------------------------------------
  age |  1.03460  .0110644  3.18  0.001  1.0131 1.0565
  typ | 11.62939  8.751927  3.26  0.001  2.6605 50.8329
---------------------------------------------------------------
```

- Again, global test against null model (statistic = 27.09) is given, as is pseudo-R^2:

$$pseudo - R^2 = 1 - \frac{-86.53782}{-100.08048} = 0.1353$$

Logistic Regression
Multiple and Continuous Covariates

- Note that admission type effect is even larger taking age into account (OR = 11.6)
  - This may be because older individuals are more likely to die under either admission type (since they are older), whereas among younger individuals, emergency admissions may be more strongly associated with death than elective admissions.
  - At a given age, emergency admissions have much higher (11.6-fold) greater risk of death

- The age effect is to increase odds of death by 1.035, or about 3.5%, per year of age increase

Multiple Logistic Regression
Inference, uses, etc

In multiple logistic regression analysis:

- **Likelihood ratio (LR) test** - we can contrast any model with a smaller nested model. Statistic has a χ^2 distribution with degrees of freedom equal to the difference in number of model parameters ($X's$).

- As already discussed, we have Z–tests for individual β coefficients ($H_0 : \beta = 0$) for the $X's$, which are in fact tests for corresponding ORs ($H_0 : \text{OR} = 1.0$), which will help indicate which predictors are useful.

Multiple Logistic Regression
Inference, uses, etc

- **Predicted probabilities for individuals.** Model can be used to develop classification algorithms, predict probability of response prospectively, etc.

  Ex: For all data records, Y's are either 1 (case, event, etc) or 0 (non-event), but the model predicts Prob(Y=1) for each record.

  - Set a cut-point for predicting/assigning a case to be an event, how many do you get right?
  - Adjust the cut-point based on # right/wrong to optimize classification. This may be different according to goals

- Model selection and diagnostics in logistic regression: Similar tools to OLS regression.