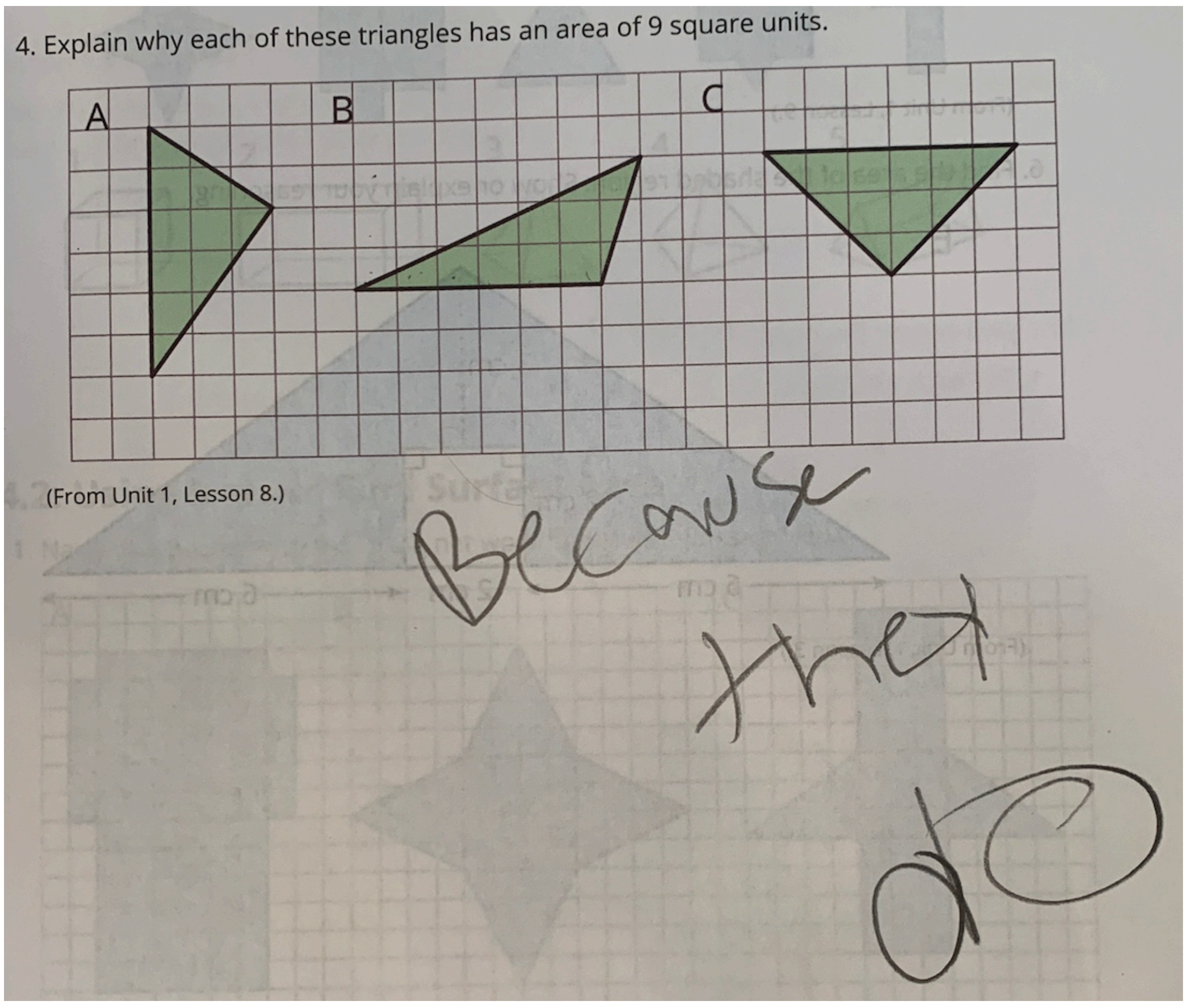
BST 210
Applied Regression Analysis



Tales of a 6th grade math student

Lecture 12
Plan for Today

- Review binary outcomes and logistic regression thus far
- Assumptions of the logistic regression model

- Interpretation, interpretation!
- More interpretation

- Confounding in logistic regression
- Multiple logistic regression

- Lot’s of examples, some code

Recall: Binary Outcomes -> Probabilities and Proportions

- Probabilities *can be estimated using sample proportions p*

- The probabilities of the outcome are actually conditional probabilities,

  p_1 = P(outcome | exposure) and

  p_2 = P(outcome | no exposure)

- Comparisons of these two conditional probabilities above are called *measures of effect*

- One measure of effect we will focus on in logistic regression is the *odds ratio* of an outcome for exposed versus unexposed subjects, which is defined as:

$$\text{OR} = \frac{p_1 / (1 - p_1)}{p_2 / (1 - p_2)}$$

Recall: Logistic Regression Model

- In order to model a binary outcome Y with covariate X and draw inference regarding the resulting proportions, or ORs, we could use the logistic regression model defined as

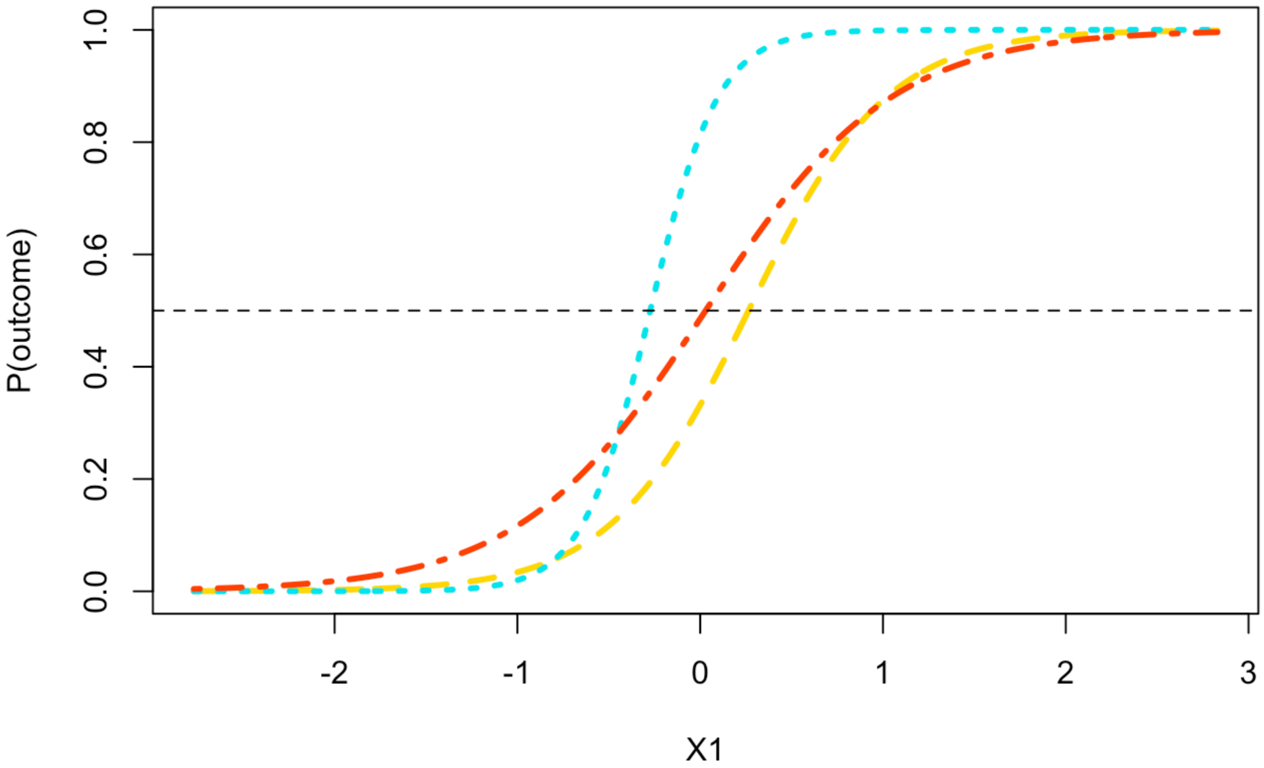
$$\log[p/(1-p)] = \alpha + \beta X$$

- Here $p = P(Y=1|X=x)$, and we assume that the relationship between $\log[p/(1-p)] = \text{logit}(p)$ and x is linear (linear on logit scale)
- Solving for p, we obtain:

$$p = \frac{\exp^{(\alpha + \beta X)}}{1 + \exp^{(\alpha + \beta X)}}$$

and the estimated probabilities fall between (0,1), as on next plot ->

Recall: Logistic Regression Model



New! Assumptions of Logistic Regression Model

- Y follows a binomial distribution
- E[Y|X] = P(x) is defined as the logistic function
  (ie linear in the logit)

  $$\text{logit}(p_i) = \log[p_i/(1 - p_i)] = \alpha + \beta\, x_i$$

- Y are independent
- 10-20 observations for each covariate
- Little to no collinearity

Recall: Logistic Regression: single binary covariate

- **What do our logistic model coefficients represent?**
- Let the subscript *i* represent the i^{th} subject in a sample
- Let *X* be a binary covariate such that

  $x_i = 1$ if the subject is exposed and

  $x_i = 0$ if unexposed

  (by the way X could be continuous, indicator, etc)
- p_i = probability of disease (or 'success') for the i^{th} subject
- We then fit the logistic regression model

$$\mathbf{logit}(p_i) = \log[p_i/(1 - p_i)] = \alpha + \beta x_i$$

Recall: Logistic Regression: single binary covariate

- **Interpret 'intercept' α**

  → if $x_i = 0$, $\mathbf{e}^{\alpha}$ **=** odds of disease (or 'success') for unexposed
  → …or…. **α =** log(odds of disease for unexposed)

- **Interpret 'slope' β**

  → $\mathbf{e}^{\beta}$ **=** OR of disease (or 'success') for exposed versus unexposed
  → β = log(OR of disease for exposed versus unexposed)

- **In general**

  Exposed = **α + β**
  Unexposed = **α**

New! Logistic Regression: single continuous covariate

- **What do our logistic model coefficients represent?**
- Let the subscript *i* represent the i^{th} subject in a sample
- Let *X* be a <u>continuous covariate rather than binary</u>
- p_i = probability of disease for the i^{th} subject
- We fit the logistic regression model

$$\mathbf{logit}(p_i) = \mathbf{log}[p_i/(1 - p_i)] = \alpha + \beta\, x_i$$

Logistic Regression: single continuous covariate

- We compare two people, A and B, with levels of exposure (x + 1) and x *(say person A who is age (x +1 year) compared with person B who is age x)*

- $\text{logit}_A = \alpha + \beta\,(x + 1) = \alpha + \beta\,x + \beta$

  $\text{logit}_B = \alpha + \beta\,x$

- $\beta = \text{logit}_A - \text{logit}_B = \log(\text{odds}_A) - \log(\text{odds}_B) = \log \dfrac{\text{odds}_A}{\text{odds}_B} = \log \text{OR}_{\text{A versus B}}$

  $e^{\beta} = \text{odds}_A \,/\, \text{odds}_B = \text{OR}_{\text{A versus B}}$

- **Interpret 'slope' β**

  $\boldsymbol{\beta = \log \text{OR}_{\text{A versus B}}}$

  $\boldsymbol{e^{\beta} = \text{OR}_{\text{A versus B}}}$

Logistic Regression: single continuous covariate

More specifically,

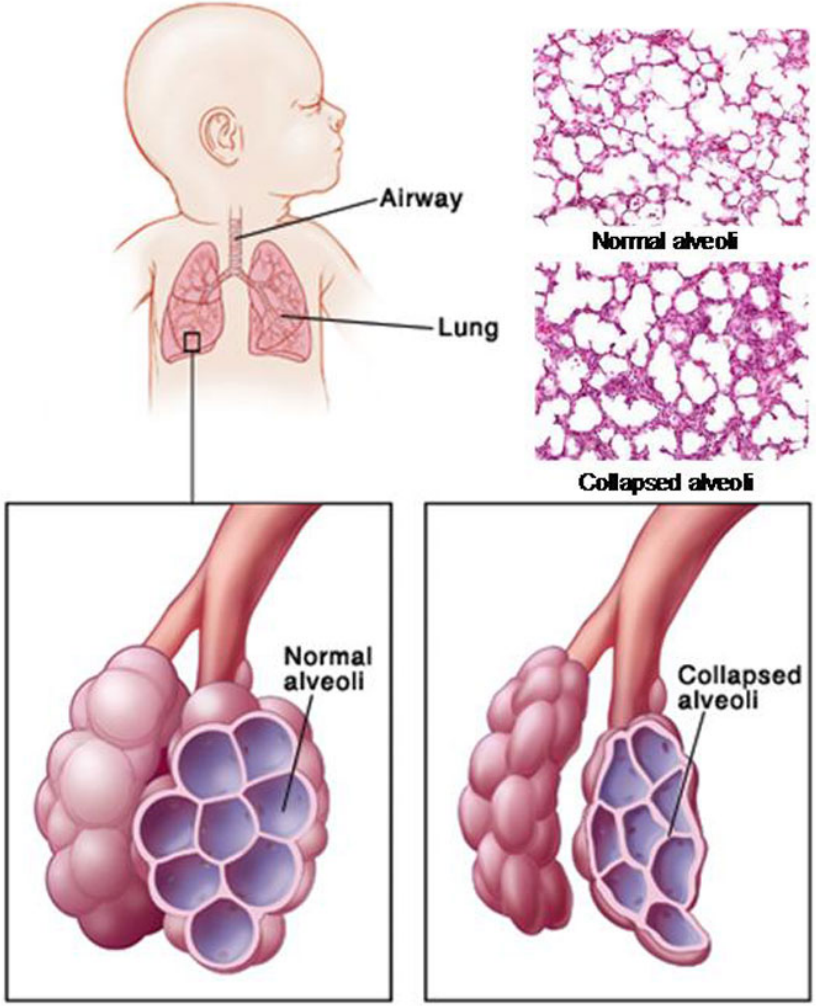
- The odds in favor of disease for person A (with exposure $x + 1$) versus person B (with exposure x) are $\mathbf{e^{\beta}}$

- $\mathbf{e^{\beta}} =$ $OR_{A\ \text{versus}\ B}$ is the OR associated with a 1 unit increase in the continuous covariate X

- We extend this to say that the

  OR associated with a *10* unit increase in X is $e^{10\beta}$, and the
  OR associated with a *20* unit increase in X is $e^{20\beta}$

Recall Example: Surfactant Use



Recall Example: Surfactant Use

- A study was performed comparing in-hospital mortality (0/1 variable) in low birth weight infants in 14 hospitals before and after the start of surfactant use
- Before: 3922 births with weight 500-1500 g, of which 960 died in the hospital (group 2)
- After: 1707 births with weight 500-1500 g, of which 335 died in the hospital (group 1)

$$\hat{p}_2 = 0.245, \hat{p}_1 = 0.196.$$

Recall Example: Surfactant Use

		In-Hospital Mortality		
		Yes	No	Total
Surfactant Use	Yes	335 (19.6%)	1372	1707
	No	960 (24.5%)	2962	3922
	Total	1295 (23.0%)	4334	5629

Remember: **OR** = (335)(2962)/(960)(1372) = 0.75

Recall Example: Surfactant Use

- We can use most stat packages to fit the logistic regression model and get estimates for α and β

$$\text{logit}(p_i) = \alpha + \beta\, x_i$$

- p_i = P(Y=death) for the i^{th} subject

- $x_i = 1$ if surfactant use = yes, and
  $x_i = 0$ if surfactant use = no

- We expect to get the same estimates for OR as the 2x2 table whenever the covariate is binary

Recall Example: Surfactant Use

```
> dat.surfactant
   birthwt surfactant death freq
1        1          1     1  177
2        1          1     0  153
3        1          0     1  479
4        1          0     0  290
5        2          1     1   78
6        2          1     0  336
7        2          0     1  257
8        2          0     0  646
9        3          1     1   40
10       3          1     0  375
11       3          0     1  142
12       3          0     0  885
13       4          1     1   40
14       4          1     0  508
15       4          0     1   82
16       4          0     0 1141
```

(Need to expand these rows (by freq) in our programming to create a line for each subject.)

Single Binary Covariate: Surfactant Use

R code:

```
> # Fit logistic regression model and find coefficients and CIs
>
> surf.glm <- glm(dat.surfactant$death ~ dat.surfactant$surfactant,
 family = binomial(), weights = freq, data = dat.surfactant)
> summary(surf.glm)

Call:
glm(formula = dat.surfactant$death ~ dat.surfactant$surfactant,
    family = binomial(), data = dat.surfactant, weights = freq)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-25.311  -13.325    1.619   16.952   36.719

Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)               -1.12669    0.03714 -30.337  < 2e-16
dat.surfactant$surfactant -0.28321    0.07137  -3.968 7.24e-05


> coefficients(surf.glm)
              (Intercept) dat.surfactant$surfactant
               -1.1266867                -0.2832076
> exp(coefficients(surf.glm))
              (Intercept) dat.surfactant$surfactant
                0.3241053                 0.7533634
> exp(confint(surf.glm))
Waiting for profiling to be done...
                              2.5 %    97.5 %
(Intercept)               0.3012131 0.3484226
dat.surfactant$surfactant 0.6543888 0.8656901
```

Single Binary Covariate: Surfactant Use

- So our fitted model is:

$$\text{logit}\,(p_{death}) = -1.127 + -0.283 * \textit{surfactant}$$

- Assume 'exposed' = (X=1) = surfactant-use group = p_1 (treated)
  'unexposed' = (X=0) = no surfactant-use group = p_2 (untreated)

- Then the estimated **odds of death** (or 'disease') in the 'unexposed/no surfactant' group is estimated to be

  $$e^{(-1.127)} = 0.324$$

  and in the 'exposed/surfactant' group is estimated to be

  $$e^{(-1.127 + -0.283)} = 0.244$$

- And the estimated **OR of death** in the 'exposed' (surfactant) group versus the 'unexposed' (non-surfactant) group is

  $$e^{(-0.283)} = 0.753, \quad \text{with 95\% CI: } e^{(-.283 - 1.96*0.07\ ,\ -.283 + 1.96*0.07)} = (0.654, 0.866)$$

Single Binary Covariate: Surfactant Use

- We estimate that the odds of death in the exposed group is ~ 25% lower than the odds of death in the unexposed group.

- That is, the odds of death after introduction of surfactant is ~ 25% lower than the odds of death before surfactant use. Another way of saying this is that the odds of death in the surfactant group versus the non-surfactant group are about 3:4.

- With 95% confidence, the odds of death in the surfactant use group is between 13% and 35% lower than that of the non-surfactant use group (CI doesn't contain 1)

- That is, $\widehat{OR} = 0.7533634, \;\; z = -3.97, \; p < 0.001,$

$$95\%\; CI = \;(0.6550237, 0.866467)$$

- Thus, the use of surfactant shows statistical significance and appears protective against in-hospital mortality in this lower birth weight cohort

Next - Continuous Covariate: Birth weight

- A potential covariate of interest is birth weight, which is divided into the following categories:

500-749 g	(group 1)
750-999 g	(group 2)
1000-1249 g	(group 3)
1250-1500 g	(group 4)

Continuous Covariate: Birth weight

- We will first treat birth weight as 'continuous' (even though it's coded ordinal – we'll try ordinal after)

- We can run the model

  $$\text{logit}(p) = \alpha + \beta x, \quad \text{where } x = 1, 2, 3, \text{ or } 4$$

- Here a 1 unit increase in x corresponds to a ~250 g increase in birth weight

- (Another alternative is to use a continuous birth weight for *each subject*, rather than just ordinally categorized)

Continuous Covariate: Birth weight

logistic death birthwt

```
Logistic regression                               Number of obs   =       5629
                                                  LR chi2(1)      =    1079.76
                                                  Prob > chi2     =     0.0000
Log likelihood = -2496.1053                       Pseudo R2       =     0.1778

------------------------------------------------------------------------------
       death | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     birthwt |   .3539453   .0126528   -29.05   0.000      .329995    .3796338
------------------------------------------------------------------------------
```

- β = log(OR) = log(.354) = -1.04
- logit(p) = α + -1.04*birthwt
- The estimated odds ratio is 0.354, with 95% confidence interval (0.329, 0.379) and $p < 0.001$

Continuous Covariate: Birth weight

- This means that if we compare two infants A and B such that infant A weighs 250 g more than infant B (infant A is one ordinal category higher than B), then the odds ratio in favor of death for infant A versus infant B is 0.35

- Infant A has a lower odds of death (65% lower)

- This would be true for category 4 versus 3, 3 versus 2, and 2 versus 1 since we are treating birth weight as continuous

- Pros and Cons to modeling a ‘continuous’ covariate this way

Next - Categorical Covariate: Birth weight

- We next consider birth weight as a categorical variable divided into four mutually exclusive categories:

500-749 g	(group 1)
750-999 g	(group 2)
1000-1249 g	(group 3)
1250-1500 g	(group 4)

- We select one category as a reference or baseline group, and create indicator variables (dummy variables) I(x) for the other categories
- If we choose 500-749 g (group 1) as the reference group x_1, then we create three indicator variables as follows ->

Categorical Covariate: Birth weight

x_2 = 1 if birth weight = group 2, 0 otherwise

x_3 = 1 if birth weight = group 3, 0 otherwise

x_4 = 1 if birth weight = group 4, 0 otherwise

- And we run the model

$$\log[p/(1-p)] = \alpha + \beta_2\, x_2 + \beta_3\, x_3 + \beta_4\, x_4$$

Categorical Covariate: Birth weight

Then,

$e^{\beta 2} = OR_2$

= odds of death for birth weight group 2 versus odds of death for birth weight group 1

ie group 2 versus group 1

$e^{\beta 3} = OR_3$ group 3 versus group 1

$e^{\beta 4} = OR_4$ group 4 versus group 1

Statistical packages will provide estimates of the three odds ratios OR_2, OR_3, and OR_4 (or the three β coefficients), where the reference group is birth weight group 1 (500-749 g)

Categorical Covariate: Birth weight

Tabulation:

Birth weight	Frequency	Percent
500-749 g	1,099	19.5
750-999 g	1,317	23.4
1000-1249 g	1,442	25.6
1250-1500 g	1,771	31.5

Categorical Covariate: Birth weight

Tabulation:

Birth weight	Freq.	# Died	Percent Died
500-749 g	1,099	656	59.7
750-999 g	1,317	335	25.4
1000-1249 g	1,442	182	12.6
1250-1500 g	1,771	122	6.7

Categorical Covariate: Birth weight

Getting a look at the breakdown, and labeling in Stata:

```
. label variable birthwt "Birth weight (g)"

. label define bwcat 1 "500-749" 2 "750-999" 3 "1000-1249" 4 "1250-1500"

. label values birthwt bwcat

. tabulate birthwt

      Birth |
 weight (g) |      Freq.     Percent        Cum.
------------+-----------------------------------
    500-749 |      1,099       19.52       19.52
    750-999 |      1,317       23.40       42.92
  1000-1249 |      1,442       25.62       68.54
  1250-1500 |      1,771       31.46      100.00
------------+-----------------------------------
      Total |      5,629      100.00
```

Categorical Covariate: Birth weight

Making I(x) with group =1 as reference category:

```
. generate bwt1 = 0

. replace bwt1 = 1 if birthwt == 1
(1099 real changes made)

. generate bwt2 = 0

. replace bwt2 = 1 if birthwt == 2
(1317 real changes made)

. generate bwt3 = 0

. replace bwt3 = 1 if birthwt == 3
(1442 real changes made)

. generate bwt4 = 0

. replace bwt4 = 1 if birthwt == 4
(1771 real changes made)

. logistic death bwt2 bwt3 bwt4
```

```
. * Stata can also create the indicator variables for you with the i. command
. logistic death i.birthwt
```

Categorical Covariate: Birth weight

Running logistic regression model, all comparisons to group 1:

```
Logistic regression                               Number of obs   =       5629
                                                  LR chi2(3)      =    1114.73
                                                  Prob > chi2     =     0.0000
Log likelihood = -2478.6223                       Pseudo R2       =     0.1836

------------------------------------------------------------------------------
       death | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     birthwt |
    750-999  |   .2303739   .0203267   -16.64   0.000     .1937889    .2738657
  1000-1249  |    .097544   .0097884   -23.19   0.000     .0801279    .1187457
  1250-1500  |   .0499619   .0056048   -26.71   0.000     .0401005    .0622485
             |
       _cons |   1.480813   .0910637     6.38   0.000     1.312668    1.670496
------------------------------------------------------------------------------

. logit

Logistic regression                               Number of obs   =       5629
                                                  LR chi2(3)      =    1114.73
                                                  Prob > chi2     =     0.0000
Log likelihood = -2478.6223                       Pseudo R2       =     0.1836

------------------------------------------------------------------------------
       death |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     birthwt |
    750-999  |  -1.468052   .0882334   -16.64   0.000    -1.640986   -1.295118
  1000-1249  |  -2.327451   .1003488   -23.19   0.000    -2.524131   -2.130771
  1250-1500  |  -2.996494   .1121823   -26.71   0.000    -3.216368   -2.776621
             |
       _cons |    .392591   .0614957     6.38   0.000     .2720616    .5131205
------------------------------------------------------------------------------

. * Without exactly telling you, Stata is using the first/lowest category as
. * baseline group here (500-749 g), so these results match with using
. * bwt2, bwt3, and bwt4 as indicator variables we included above
```

Categorical Covariate: Birth weight

Logistic regression results:

Coef	$\hat{\beta}$	$s.e.(\hat{\beta})$	$\widehat{OR}$	P
x_2	-1.468	0.088	0.230	<0.001
x_3	-2.327	0.100	0.098	<0.001
x_4	-2.996	0.112	0.050	<0.001
Const	0.392	0.061	---	<0.001

- Here we are comparing with the baseline category 1, namely 500-749 g

Categorical Covariate: Birth weight

- The estimated odds ratios for groups 2, 3, and 4 versus group 1, are 0.23, 0.10, and 0.05 respectively

- All have $p < 0.001$, meaning that we would reject each of the null hypotheses (separately):

$$H_0: \beta_2 = 0 \text{ vs. } H_1: \beta_2 \neq 0$$
$$H_0: \beta_3 = 0 \text{ vs. } H_1: \beta_3 \neq 0$$
$$H_0: \beta_4 = 0 \text{ vs. } H_1: \beta_4 \neq 0$$

Categorical Covariate: Birth weight

- We conclude that each of the three β coefficients are not equal to 0, and therefore that all three odds ratios are not equal to 1
- They are all less than 1 in fact
- These are called *Wald tests* -- we will discuss hypothesis testing next week
- The odds of death are significantly lower for birth weights greater than 500-749 g

Categorical Covariate: Birth weight

- We may also wish to compare to group = 4 rather than group = 1:

Logistic regression results:

Coef	$\hat{\beta}$	$s.e.(\hat{\beta})$	$\widehat{OR}$	P
x_1	2.996	0.112	20.0	<0.001
x_2	1.528	0.113	4.6	<0.001
x_3	0.669	0.123	2.0	<0.001
Const	-2.604	0.094	---	<0.001

- Here we are comparing with the baseline category 4, namely 1250-1500 g, and the $\widehat{OR}$ estimates are > 1

Next, Estimating the p's (probabilities): Birth weight

- We may want to estimate the probabilities of disease p_i (or death in this example) for given values of X

$$\hat{p} = \frac{\exp(a + b_2x_2 + b_3x_3 + b_4x_4)}{1 + \exp(a + b_2x_2 + b_3x_3 + b_4x_4)}$$

- Each of the indicator variables in the expression above takes the value 0 or 1

- For a child in birth weight category 3, for example, $x_2 = 0$, $x_3 = 1$, and $x_4 = 0$

- You would simply insert coefficients from the fitted logistic regression model

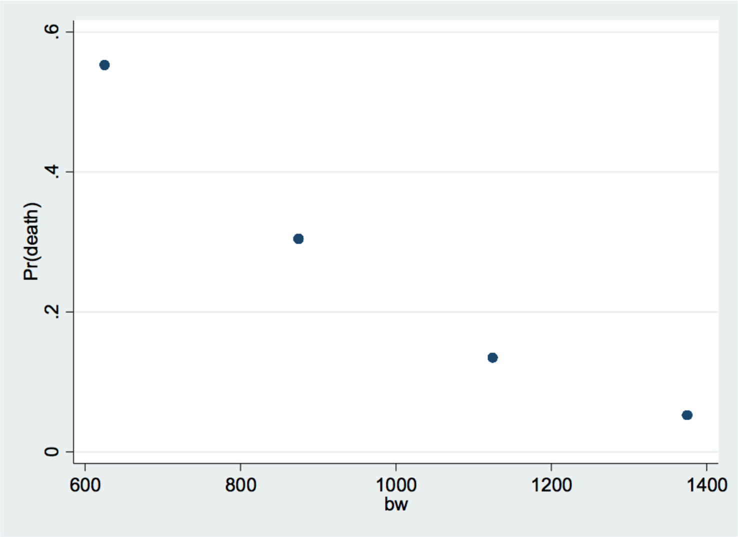
Estimating the p's (probabilities): Birth weight

- We are able to predict the probability of in-hospital mortality for each birth weight category

```
. predict phat_bwt

. list birthwt phat_bwt

     +----------------------+
     |   birthwt   phat_bwt |
     |----------------------|
  1. |   500-749   .5969063 |

  5. |   750-999    .254366 |

  9. | 1000-1249   .1262136 |

 13. | 1250-1500   .0688876 |
```

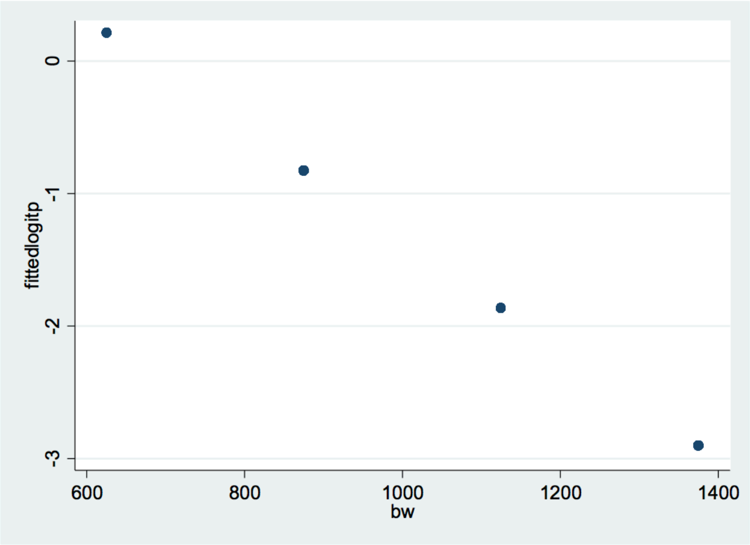
Estimating the p's (probabilities): Birth weight

Estimated p



More sigmoidal

Estimated logit(p)



More linear in the logit

Changing reference group: Birth weight

- Returning to the model with birth weight coded as a categorical variable, suppose that the odds ratio for group 2 versus group 1 is not significantly different from 1…

```
. logistic death bwt2 bwt3 bwt4
```

```
Logistic regression                               Number of obs   =       5629
                                                  LR chi2(3)      =    1114.73
                                                  Prob > chi2     =     0.0000
Log likelihood = -2478.6223                       Pseudo R2       =     0.1836

------------------------------------------------------------------------------
       death | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        bwt2 |   .2303739   .0203267   -16.64   0.000     .1937889    .2738657
        bwt3 |    .097544   .0097884   -23.19   0.000     .0801279    .1187457
        bwt4 |   .0499619   .0056048   -26.71   0.000     .0401005    .0622485
------------------------------------------------------------------------------
```

Changing reference group: Birth weight

- We might choose to drop **bwt2** from the model

```
. logistic death bwt3 bwt4

Logistic regression                               Number of obs   =       5629
                                                  LR chi2(2)      =     819.51
                                                  Prob > chi2     =     0.0000
Log likelihood = -2626.2303                       Pseudo R2       =     0.1350

------------------------------------------------------------------------------
       death | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        bwt3 |   .2077027   .0185763   -17.57   0.000     .1743063    .2474976
        bwt4 |    .106385   .0109085   -21.85   0.000     .0870162    .1300651
------------------------------------------------------------------------------
```

- If we do this, we are changing the reference group ->

Changing reference group: Birth weight

- We are now comparing group 3 to groups (1 + 2), and group 4 to groups (1 + 2)
- Not only do the odds ratios change, but their interpretations change as well
- We must be cautious when removing indicator variables from models (e.g., suppose it had been the odds ratio for group 4 versus group 1 that was not statistically significant)

Summary: Surfactant Data

- We have shown that among low birth weight infants with respiratory distress syndrome, the odds ratio for in-hospital mortality after surfactant use versus before surfactant use is 0.75 ($p < 0.001$)

- Also, birth weight is significantly associated with mortality, with lower odds of mortality as birth weight increases

Other relationships: Surfactant Data

- If there are trends in birth weight over time, it is possible that the surfactant effect might be due (at least in part) to birth weight
- Birth weight might be a *confounder* of the relationship between surfactant use and mortality
- We would like to examine the relationship between surfactant use and mortality, controlling for birth weight

Other relationships: Surfactant Data

- A *confounder* is a variable that is related to both the exposure and the disease
- It is important to control for confounding variables when studying disease-exposure associations, to avoid possible misinterpretation of the relationship
- Confounders can make the relationship seem either stronger or weaker than it really is (and occasionally even make it go in the opposite direction)

Confounding: Surfactant Data

- Do we always want to control for ‘confounding-like’ variables?
- In some cases, a ‘confounder-like’ factor may be an *intermediate variable* between exposure and disease
- In particular, such a variable may be on the “causal pathway” between exposure and outcome, and adjusting for this factor may ***not*** be appropriate
- Intermediate variables should usually not be controlled for when studying the association between exposure and disease

Confounding: Surfactant Data

- The decision about whether a variable is in the causal pathway between exposure and disease should be based on biological rather than statistical considerations
- What are your thoughts on the relationship between birth weight, surfactant and mortality?
- If we do not think a confounder is an intermediate variable, how do we control for its effect?

Controlling for Confounding

There are three methods commonly used:

- Stratification (stratified analysis, performing separate analyses for each stratum; simpler but more inefficient)
- Adjustment after stratification (Mantel-Haenszel methods; appropriate for categorical confounding factors)
- Multivariable analysis (logistic regression; appropriate for categorical or continuous confounding factors)

Controlling for Confounding

- Stratification: A stratified analysis is the simplest approach analytically, but it is inefficient if there are many strata and sample sizes within strata are small
- This means there is a loss of statistical power, and usually more complication since samples are not always big enough across strata!

Controlling for Confounding

- <u>Adjustment after stratification:</u>

```
. cc death surfactant, by(birthwt) woolf

         birthwt |       OR       [95% Conf. Interval]   M-H Weight
-----------------+-------------------------------------------------
               1 |   .7003971      .5395481    .909198     66.68517 (Woolf)
               2 |   .5835186      .4385443   .7764187      65.5672 (Woolf)
               3 |   .6647887      .4587133   .9634428     36.92788 (Woolf)
               4 |    1.09564      .7401889   1.621786     23.52117 (Woolf)
-----------------+-------------------------------------------------
           Crude |   .7533634      .6550237    .866467              (Woolf)
    M-H combined |   .7020486      .6002899    .821057
-------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(3) =     6.64  Pr>chi2 = 0.0843

                   Test that combined OR = 1:
                                Mantel-Haenszel chi2(1) =     19.81
                                                Pr>chi2 =    0.0000
```

Controlling for Confounding

- Multiple Logistic Regression:

Another approach is to consider the model

$$\text{logit}(p_i) = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i}$$

where

x_{1i} = surfactant use (1 = yes, 0 = no)

x_{2i} = 1 if birth weight = 750-999 g, 0 otherwise

x_{3i} = 1 if birth weight = 1000-1249 g, 0 otherwise

x_{4i} = 1 if birth weight = 1250-1500g, 0 otherwise

Ahead:

- Multiple Logistic Regression and examples
- Hypothesis testing and confidence intervals
- Effect modification
- Compare and contrast to linear regression
- Model building and model assessment

…and more!