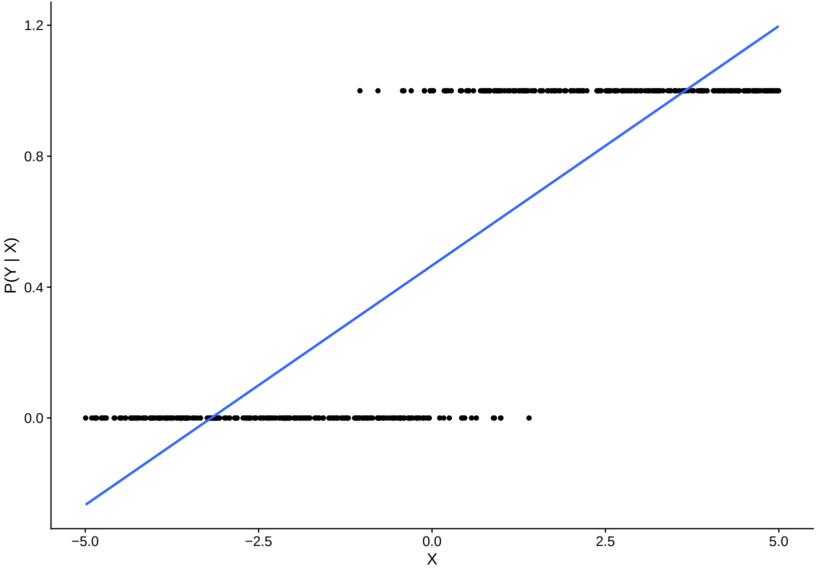
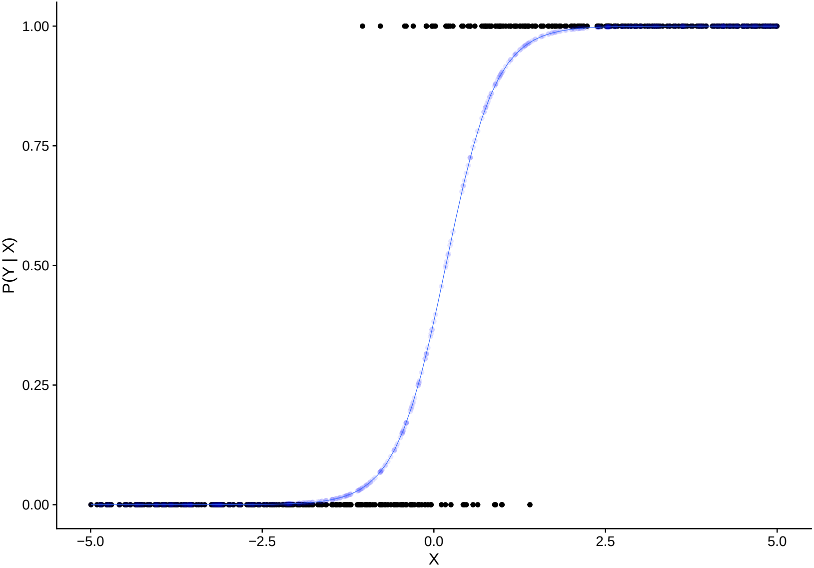
Logistic Regression

Randy Johnson

Modeling binary outcomes



Modeling binary outcomes



Logistic Regression

$$\log odds(Y|X) = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n$$

$$\beta_1 = \log OR(Y|X)$$

- Caveats:
  - β_0 doesn't have any real-world interpretation for case-control studies.
  - OR $>$ RR

OR vs RR

Given A equal to the number of events and B equal to the number of non-events,

$$\begin{aligned} odds(Y|X) &= \frac{\mathrm{P}(Y|X)}{1 - \mathrm{P}(Y|X)} \\ &= \frac{\frac{A}{\cancel{A+B}}}{\frac{B}{\cancel{A+B}}} \\ &= \frac{A}{B} \end{aligned}$$

$$\begin{aligned} RR(Y|X) &= \mathrm{P}(Y|X) \\ &= \frac{A}{A+B} \end{aligned}$$

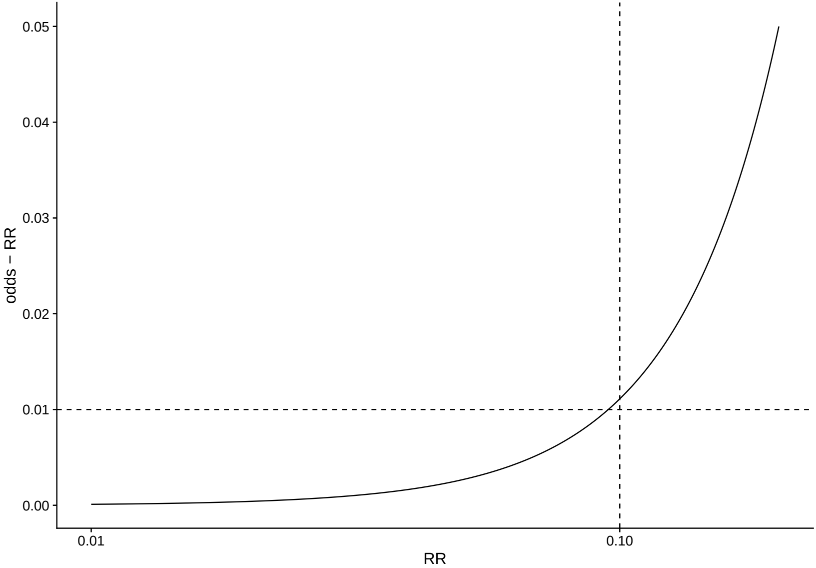
OR vs RR

The odds will always overestimate the relative risk,

$$
\begin{gathered}
odds(Y|X) > RR(Y|X) \\
\frac{A}{B} > \frac{A}{A+B} \\
\frac{A}{B} \approx \frac{A}{A+B}
\end{gathered}
$$

but they will be approximately equal if the number of events, A, is small relative to B (i.e. when the event is rare).

OR vs RR



Example

```
##
## Call:
## glm(formula = d ~ as.factor(grade), family = "binomial", data = gbsg)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -1.162  -1.102  -0.709   1.255   1.734
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -1.2528     0.2673  -4.687 2.77e-06 ***
## as.factor(grade)2   1.0721     0.2837   3.778 0.000158 ***
## as.factor(grade)3   1.2155     0.3103   3.917 8.96e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 939.68  on 685  degrees of freedom
## Residual deviance: 920.86  on 683  degrees of freedom
## AIC: 926.86
##
## Number of Fisher Scoring iterations: 4
```

Example

```r
ci(model)[2:3,c('Estimate', 'CI lower', 'CI upper')] %>%
    exp()
```

```
##                      Estimate CI lower CI upper
## as.factor(grade)2 2.921488 1.673598 5.099845
## as.factor(grade)3 3.371951 1.833540 6.201149
```

Example

$$\log odds(d|grade = 3) - \log odds(d|grade = 2):$$

$$\begin{array}{r} 1 * \beta_0 + 0 * \beta_1 + 1 * \beta_2 \\ -(1 * \beta_0 + 1 * \beta_1 + 0 * \beta_2) \\ \hline = 0 * \beta_0 - 1 * \beta_1 + 1 * \beta_2 \end{array}$$

Example

```
estimable(model, cm = c(0,-1,1), conf.int = 0.95)[,c("Esti
    exp()
```

```
##             Estimate  Lower.CI Upper.CI
## (0 -1 1)  1.15419 0.8008484 1.663429
```

Log binomial regression

```
##
## Call:
## glm(formula = d ~ as.factor(grade), family = binomial(link = log),
##     data = gbsg)
##
## Deviance Residuals:
##    Min       1Q   Median       3Q      Max
## -1.162  -1.102  -0.709   1.255   1.734
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -1.5041     0.2079  -7.236 4.63e-13 ***
## as.factor(grade)2   0.7165     0.2143   3.344 0.000825 ***
## as.factor(grade)3   0.7921     0.2228   3.555 0.000378 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 939.68  on 685  degrees of freedom
## Residual deviance: 920.86  on 683  degrees of freedom
## AIC: 926.86
##
## Number of Fisher Scoring iterations: 6
```

Log binomial regression

```
estimable(cm = matrix(c(0,1,0,
                        0,0,1,
                        0,-1,1), nrow = 3, byrow = TRUE),
          obj = model)[,c("Estimate"), drop = FALSE] %>%
    exp()
```

```
##            Estimate
## (0 1 0)  2.047297
## (0 0 1)  2.208075
## (0 -1 1) 1.078531
```