Logistic Regression - Part II

Statistics 149

Spring 2006

Confidence Intervals for $\pi(X)$

Similarly to getting confidence intervals for $\mu(Y|X)$ in linear regression it is also useful to construction confidence intervals of

$$\mu(Y|X) = \pi(X) = \frac{e^{X^T\beta}}{1 + e^{X^T\beta}}$$

One possible approach is to mimic what we did for confidence intervals for odds.

As mentioned last time,

$$\hat{\beta} \overset{approx.}{\sim} N(\beta, \hat{\Sigma})$$

where $\hat{\Sigma}$ is the estimate of the variance based on the inverse of the observed information matrix.

Then

$$\hat{\eta}(X) = X^T\hat{\beta} \overset{approx.}{\sim} N(X^T\beta, \hat{\sigma}_\eta^2) = N(\eta(X), \hat{\sigma}_\eta^2)$$

where

$$\hat{\sigma}_\eta^2 = X^T\hat{\Sigma}X$$

So an approximate CI for the log odds, $\eta(X)$, is given by

$$X^T\hat{\beta} \pm z^*_{\alpha/2}\hat{\sigma}_\eta = \hat{\eta}(X) \pm z^*_{\alpha/2}\hat{\sigma}_\eta = (L_x, U_x)$$

The information for this interval is easily gotten in **R** with the `predict` function.

```
> pred.ages1 <- c(15, 25, 35, 45)
> pred.levels1 <- data.frame(age=pred.ages1)
> birthwt.pred1 <- predict(birthwt.glm, pred.levels1,
+    type="link", se.fit=T)
> birthwt.pred1
$fit
         1          2          3          4
-0.3827122 -0.8942416 -1.4057711 -1.9173005

$se.fit
        1         2         3         4
0.2891894 0.1741429 0.4190995 0.7208235

$residual.scale [1] 1
```

Once you have the fits and se's from here, it is easy to get the CI's.

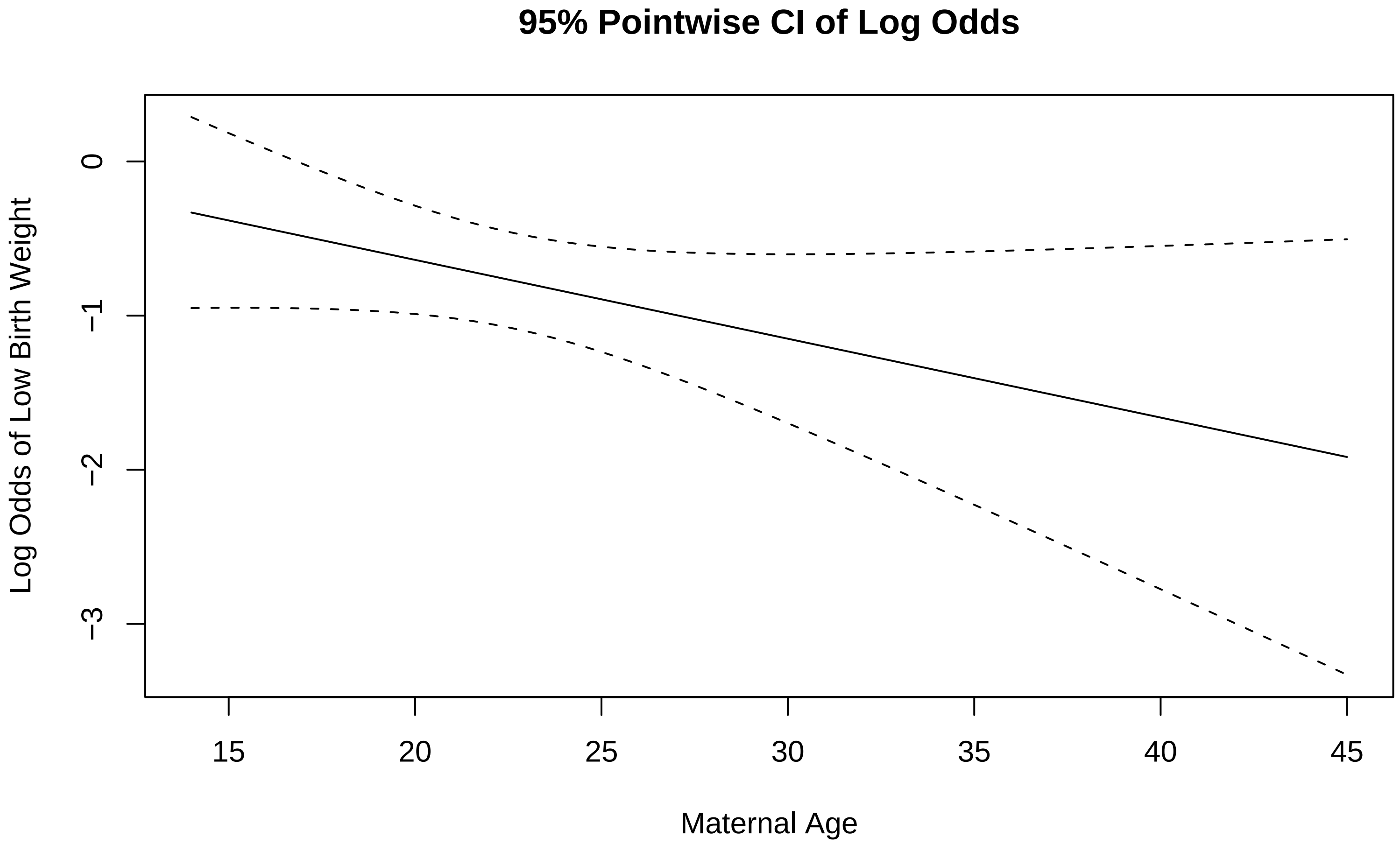
```
> glmlink.ci(birthwt.glm, pred.levels1)
        lower      upper
15 -0.9495131  0.1840887
25 -1.2355554 -0.5529279
35 -2.2271910 -0.5843511
45 -3.3300885 -0.5045124
```

Note: `glmlink.ci` is a function I've written for doing confidence intervals for linear predictor in generalized linear models. For logistic regression, this corresponds to CIs for log odds.

Effectively it is doing

```
lower <- pred.obj$fit - qnorm(alpha/2) * pred.obj$se.fit
upper <- pred.obj$fit + qnorm(alpha/2) * pred.obj$se.fit
```

The CI's for $\eta(X)$ work similarly to the CI's for $\mu(Y|X)$ in linear regression. They are narrowest for Xs around the mean X and get wider as X moves away from the mean X.



To get a confidence interval for the odds at X, $\omega(X) = e^{\eta(X)}$, we just need to exponentiate the CI for $\eta(X)$, giving

$$\begin{aligned} CI(\omega(X)) &= (e^{L_x}, e^{U_x}) \\ &= e^{\hat{\eta}(X) \pm z^*_{\alpha/2}\hat{\sigma}_\eta} \\ &= \hat{\omega}(X) \times (e^{-z^*_{\alpha/2}\hat{\sigma}_\eta}, e^{z^*_{\alpha/2}\hat{\sigma}_\eta}) \end{aligned}$$

Similarly we can get a confidence interval for $\pi(X)$ by

$$CI(\pi(X)) = \left(\frac{e^{L_x}}{1+e^{L_x}}, \frac{e^{U_x}}{1+e^{U_x}}\right)$$

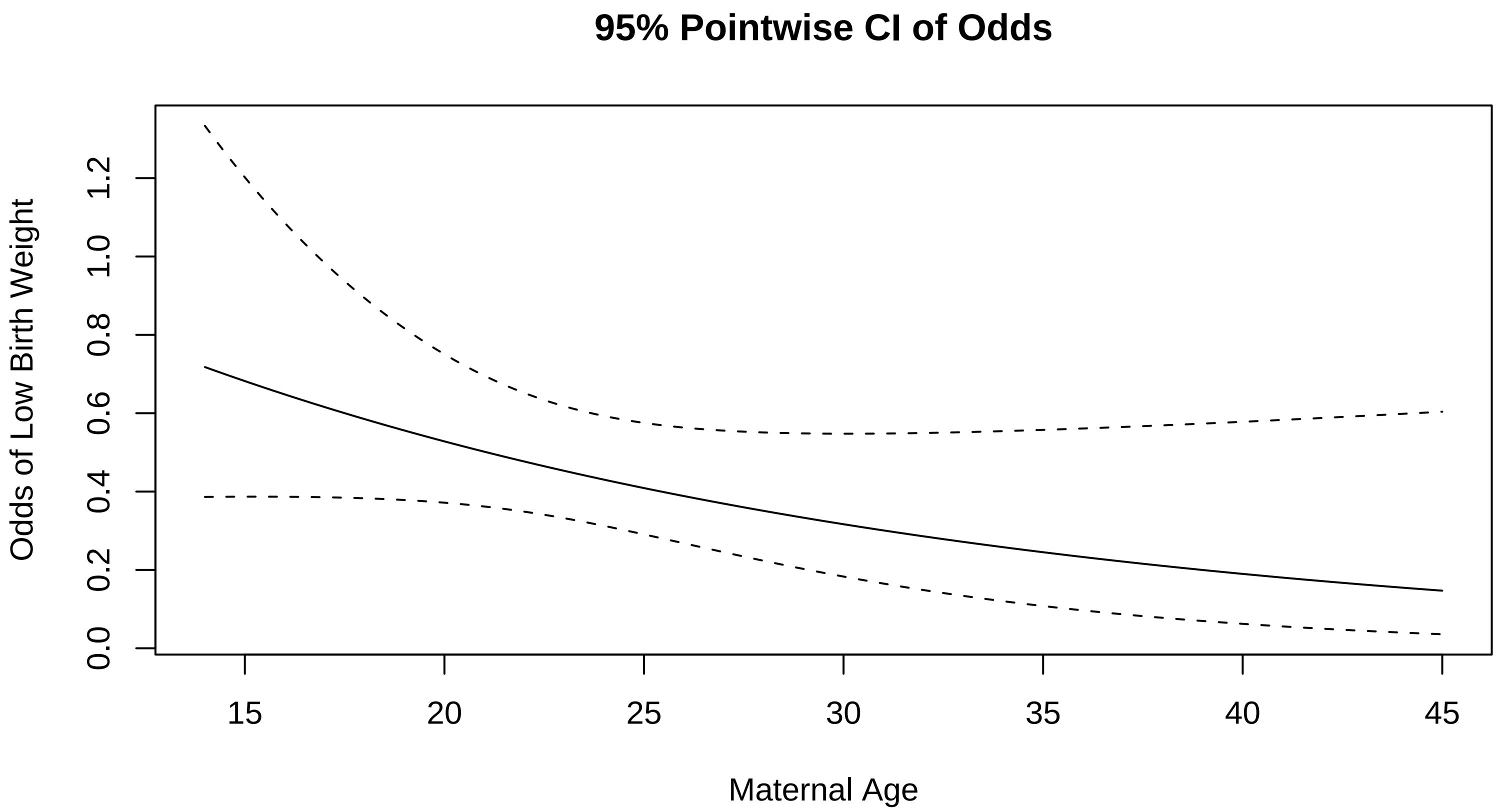
For ages 15, 25, 35, and 45, 95% CIs for the odds and probability of a low birth weight baby are

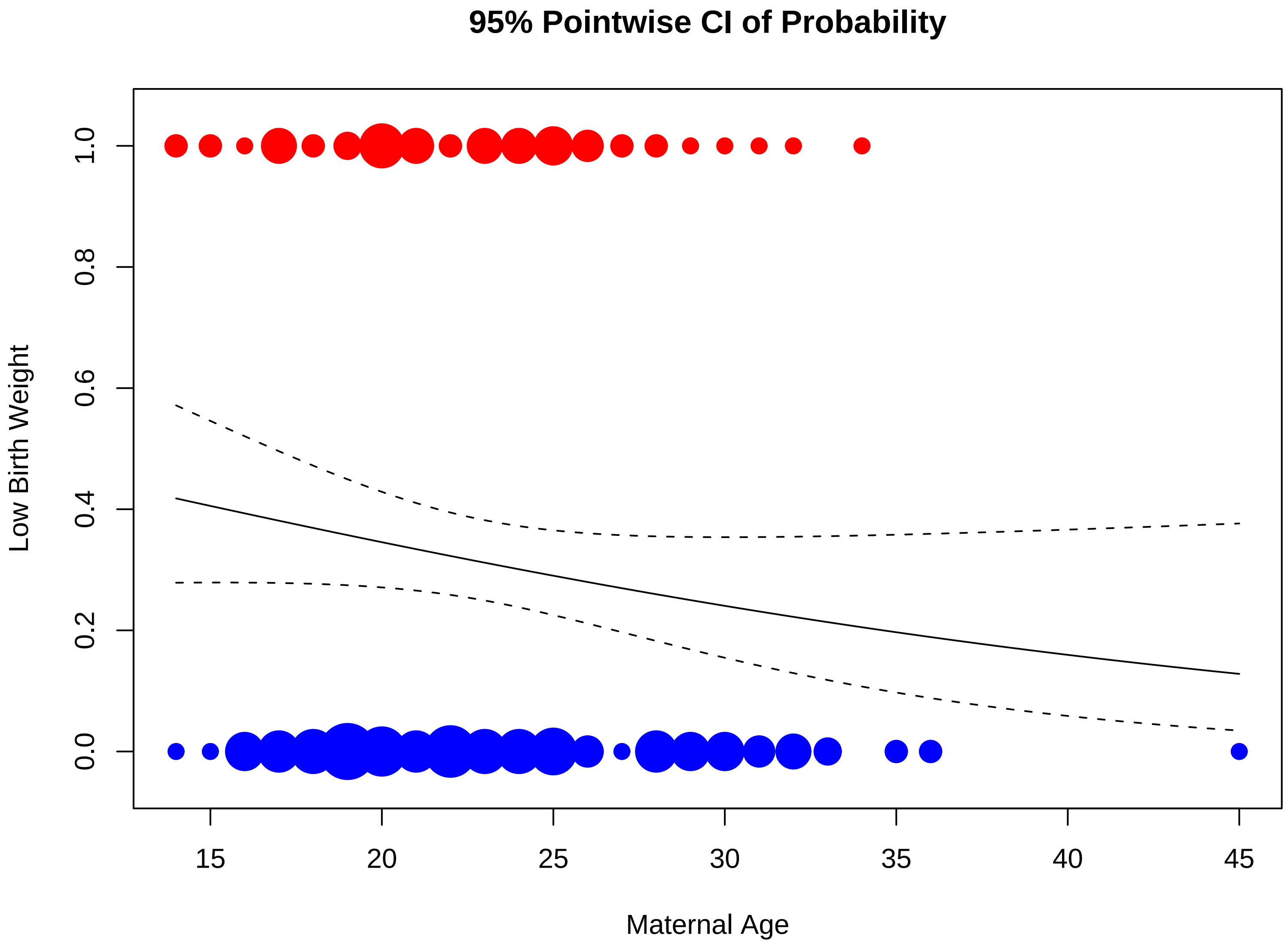
```
# Odds
> exp(glmlink.ci(birthwt.glm, pred.levels1))
        lower     upper
15 0.38692937 1.2021224
25 0.29067329 0.5752630
35 0.10783091 0.5574675
45 0.03578994 0.6037999

# Probability
> glmpred.ci(birthwt.glm, pred.levels1)
        lower     upper
15 0.27898275 0.5458926
25 0.22521059 0.3651854
35 0.09733517 0.3579320
45 0.03455328 0.3764808
```

Note: `glmpred.ci` is a function I've written for doing confidence intervals for "response" predictor in generalized linear models. For logistic regression, this corresponds to CIs for the probabilities.

These are asymmetric intervals as the transformations e^x and $\frac{e^x}{1+e^x}$ are non-linear. The asymmetry can be seen in the following to plots.




95% Pointwise CI of Probability
Low Birth Weight
0.0
0.2
0.4
0.6
0.8
1.0
15
20
25
30
35
40
45
Maternal Age

As $\hat{\omega}(X)$ and $\hat{\pi}(X)$ are MLEs, they have asymptotic normal distributions. By the delta rule

$$\begin{aligned}\widehat{\mathrm{Var}}(\hat{\omega}(X)) &= e^{2\hat{\eta}(X)}\hat{\sigma}_\eta^2 \\ &= (\hat{\omega}(X))^2\hat{\sigma}_\eta^2\end{aligned}$$

$$\begin{aligned}\widehat{\mathrm{Var}}(\hat{\pi}(X)) &= \left(\frac{e^{\hat{\eta}(X)}}{(1+e^{\hat{\eta}(X)})^2}\right)^2 \hat{\sigma}_\eta^2 \\ &= \left(\hat{\pi}(X)(1-\hat{\pi}(X))\right)^2 \hat{\sigma}_\eta^2\end{aligned}$$

So an alternative approach to confidence intervals is to use

$$CI_s(\omega(X)) = \hat{\omega}(X) \pm z^*_{\alpha/2} SE(\hat{\omega}(X))$$

$$CI_s(\pi(X)) = \hat{\pi}(X) \pm z^*_{\alpha/2} SE(\hat{\pi}(X))$$

This interval for $\pi(X)$ is also easy to get in **R** as the `predict` function can be made to give $\hat{\pi}(X)$ and $SE(\hat{\pi}(X))$ by

```
> birthwt.rpred1 <- predict(birthwt.glm, pred.levels1,
+    type="response", se.fit=T)
> birthwt.rpred1
$fit
        1         2         3         4
0.4054729 0.2902353 0.1969019 0.1281629

$se.fit
         1          2          3          4
0.06971335 0.03587321 0.06627286 0.08054278

$residual.scale [1] 1
```

The option `type` determines what predictor is used. `type="link"` does things on the "link" (linear predictor - $X^T\beta$) scale and `type="response"` does things on the response scale, in this case giving probabilities.

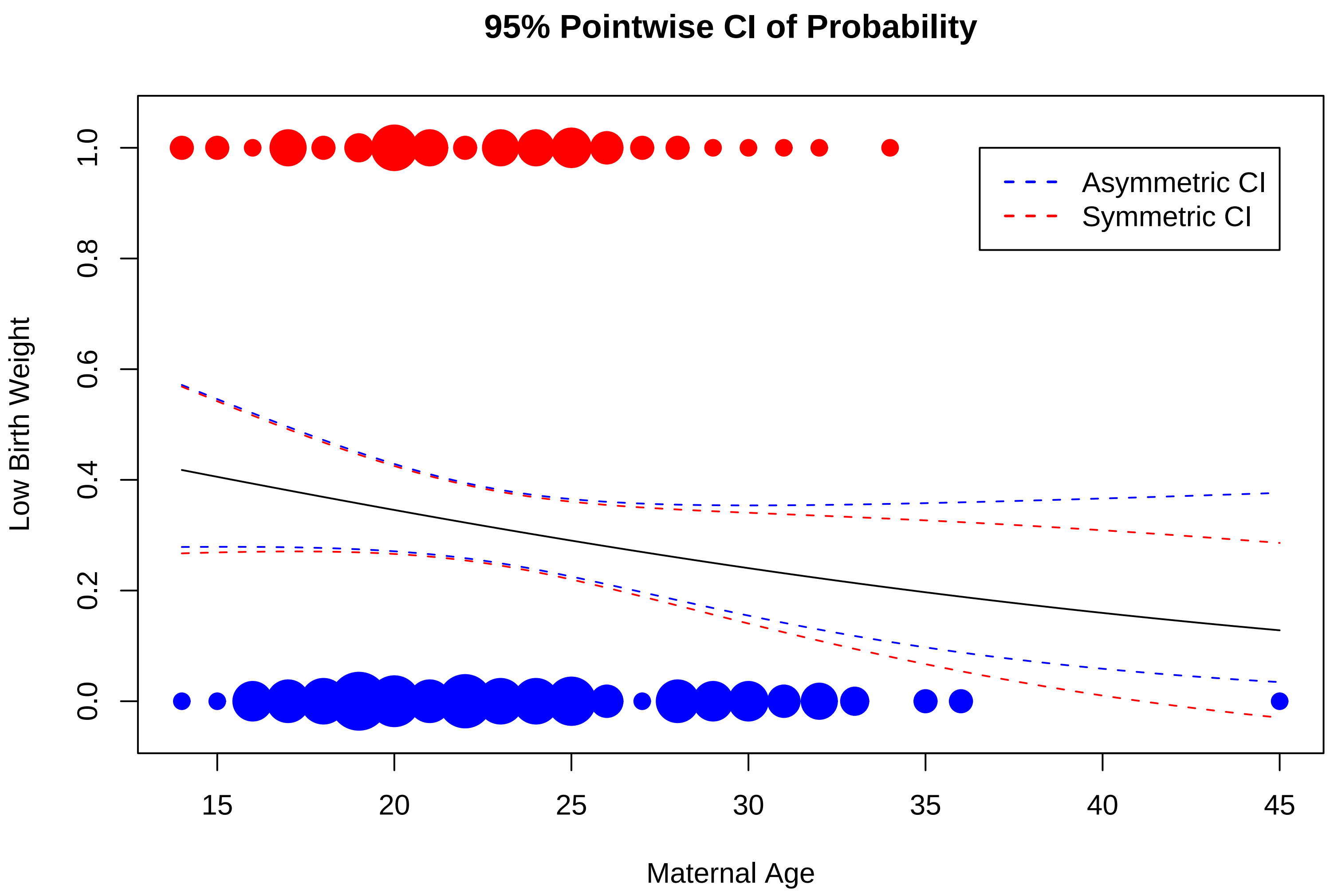
These intervals are generally suboptimal.

The asymptotic normality works better on the “link” (linear predictor) scale.

So the coverage properties of these intervals is not as good as the transformed based intervals.

One other property of these intervals is that they are symmetric. This can lead to intervals going outside the interval (0,1). This will tend to when $\hat{\pi}(X)$ approaches 0 or 1.

```
> good.ci <- glmpred.ci(birthwt.glm, pred.levels1)
> bad.ci <- glmpred.ci.bad(birthwt.glm, pred.levels1)
> cbind(good.ci, bad.ci)
        lower     upper       lower     upper
15 0.27898275 0.5458926  0.26883726 0.5421086
25 0.22521059 0.3651854  0.21992507 0.3605455
35 0.09733517 0.3579320  0.06700952 0.3267943
45 0.03455328 0.3764808 -0.02969804 0.2860238
```



Because of these problems, the first intervals discussed are normally used.

CI for Odds Ratios

A related problem is constructing confidence intervals for odds ratios of the form

$$\frac{\omega(x_1)}{\omega(x_2)} = e^{\beta_1(x_1 - x_2)}$$

(assuming only a single predictor variable).

For example, we might want to compare the odds of a low birth weight birth for 15 and 25 year old women.

Since the odds ratio is just a simple function of the parameters, we can calculate a CI for the log odds ratio and exponentiate it back up.

Lets look at the CI for $\log \frac{\omega(x_1)}{\omega(x_2)} = \beta_1(x_1 - x_2)$.

$$
\begin{aligned}
CI(\beta_1(x_1 - x_2)) &= \hat{\beta}(x_1 - x_2) \pm z^*_{\alpha/2} SE(\hat{\beta}(x_1 - x_2)) \\
&= \hat{\beta}(x_1 - x_2) \pm z^*_{\alpha/2}(x_1 - x_2) SE(\hat{\beta}) \\
&= (x_1 - x_2)\hat{\beta} \pm z^*_{\alpha/2} SE(\hat{\beta}) \\
&= (x_1 - x_2) CI(\beta_1) = (L_{lo}, U_{lo})
\end{aligned}
$$

Exponentiating back gives

$$
\begin{aligned}
CI\left(\frac{\omega(x_1)}{\omega(x_2)}\right) &= (e^{L_{lo}}, e^{U_{lo}}) \\
&= e^{CI(\beta_1(x_1 - x_2))} \\
&= e^{(x_1 - x_2) CI(\beta_1)} \\
&= (CI(e^{\beta_1}))^{(x_1 - x_2)}
\end{aligned}
$$

So to compare the odds for 15 and 25 year women,

```
> birthwt.coef <- coef(birthwt.glm)
> birthwt.coef[2]
        age
-0.05115294
> birthwt.se <- sqrt(diag(vcov(birthwt.glm)))
> birthwt.se[2]
       age
0.03151376
> est <- (15 - 25) * birthwt.coef[2]
> est
      age
0.5115294
> se <- abs((15 - 25) * birthwt.se[2])
> se
      age
0.3151376
```

A 95% confidence interval for the log odds ratio is

$$\begin{aligned} CI\left(\log\frac{\omega(x_1)}{\omega(x_2)}\right) &= -0.5515 \pm 1.96 \times 0.3151 \\ &= -0.5515 \pm 0.6177 \\ &= (-0.106, 1.129) \end{aligned}$$

yielding a 95% CI for the odds ratio of

$$CI\left(\frac{\omega(x_1)}{\omega(x_2)}\right) = (e^{-0.106}, e^{1.129}) = (0.899, 3.093)$$

The estimated odds ratio is

$$\begin{aligned} \frac{\hat{\omega}(15)}{\hat{\omega}(25)} &= e^{\hat{\beta}_1(15-25)} \\ &= e^{-0.5115} = 1.668 \end{aligned}$$

```
> logodds.ci <- c(est - qnorm(0.025, lower.tail=F)*se ,
+                 est + qnorm(0.025, lower.tail=F)*se)
> logodds.ci
       age        age
-0.1061290  1.1291878
> odds.ci <- exp(logodds.ci)
> odds.ci
      age       age
0.8993086 3.0931433
```

While the estimated odds ratio is greater than 1, since the CI contains 1, it is possible that odds are the same for 15 and 25 year old women. This should happen in this case since the CI for β contains 0 (age may not have an effect).

Comparing Models

Now lets consider running the model with all the possible predictor variables in it

```
> attach(birthwt)
>       race <- factor(race, labels = c("white", "black", "other"))
>       ptd <- factor(ptl > 0)
>       ftv <- factor(ftv)
>       levels(ftv)[-(1:2)] <- "2+"
>       bwt <- data.frame(low = factor(low), age, lwt, race,
+           smoke = (smoke > 0), ptd, ht = (ht > 0),
+       ui = (ui > 0), ftv)
> detach("birthwt")
> options(contrasts = c("contr.treatment", "contr.poly"))
> birthwtall.glm <- glm(low ~ ., binomial, bwt)
```

```
> summary(birthwtall.glm)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7038  -0.8068  -0.5008   0.8836   2.2152

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.82302    1.24471   0.661  0.50848
age         -0.03723    0.03870  -0.962  0.33602
lwt         -0.01565    0.00708  -2.211  0.02705 *
raceblack    1.19241    0.53597   2.225  0.02609 *
raceother    0.74069    0.46174   1.604  0.10869
smokeTRUE    0.75553    0.42502   1.778  0.07546 .
ptdTRUE      1.34376    0.48062   2.796  0.00518 **
htTRUE       1.91317    0.72074   2.654  0.00794 **
uiTRUE       0.68019    0.46434   1.465  0.14296
ftv1        -0.43638    0.47939  -0.910  0.36268
ftv2+        0.17901    0.45638   0.392  0.69488
```

```
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 234.67  on 188  degrees of freedom
Residual deviance: 195.48  on 178  degrees of freedom
AIC: 217.48

Number of Fisher Scoring iterations: 4
```

Suppose we want to examine whether age (β_1) or `lwt` - weight at last menstrual period (β_2) are useful in predicting the probability of low birth weight babies, i.e.

$$H_0 : \beta_1 = \beta_2 = 0 \qquad \text{vs} \qquad H_A : \beta_1 \neq 0 \text{ or } \beta_2 \neq 0$$

From the Wald tests it appears that `lwt` might be important. However we want a single test statistic to examine this hypothesis, similar to the F-test in linear regression.

The general framework here is similar. We need to consider 2 models

- Full model: the model with all the predictors of interest. Describes H_A.

$$\text{logit}\pi(X) = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p$$

- Reduced model: the model describing H_0.

$$\text{logit}\pi(X) = \beta_0 + \beta_3 x_3 + \ldots + \beta_p x_p$$

  This model must be a special case of the full model, with some parameter values in the full model being fixed (often by setting them to 0 as above). Sometimes referred to as a nested model.

So we want to see if the full model gives a much better fit than the reduced model. Lets fit the data under both models, giving parameter estimates $\hat{\beta}_f$ and $\hat{\beta}_r$ for the full and reduced models.

Let

$$LMAX_{full} = L(\hat{\beta}_f) \qquad \text{and} \qquad LMAX_{reduced} = L(\hat{\beta}_r)$$

be the values of the likelihood under both parameter estimates.

Then we can use

$$\begin{aligned} LRT &= 2 \log \frac{LMAX_{full}}{LMAX_{reduced}} \\ &= 2 \log(LMAX_{full}) - 2 \log(LMAX_{reduced}) \end{aligned}$$

as a test statistic to compare the fits. Then LRT is compared to a χ_k^2 distribution where k is the difference in the number of parameters of the two models (often the number of βs discussed in H_0.

This is known as the Likelihood Ratio Test.

The p-value satisfies

$$p\text{-value} = P[\chi^2_k \geq LRT]$$

Sometime you will see this written as

$$LRT = 2(l(\hat{\beta}_f) - l(\hat{\beta}_r))$$

As mentioned in the book, some software packages will give a value known as the deviance instead of the value of the likelihood or log likelihood.

$$deviance = constant - 2\log(LMAX)$$

where the constant is the same for both models. Then

$$LRT = deviance_{reduced} - deviance_{full}$$

So lets use this test to examine whether `age` or `lwt` have an effect here.

```
> summary(birthwtall.glm)

    Null deviance: 234.67  on 188  degrees of freedom
Residual deviance: 195.48  on 178  degrees of freedom
AIC: 217.48

> birthwt2.glm <- update(birthwtall.glm, . ~ . - age - lwt)
> summary(birthwt2.glm)

Call:
glm(formula = low ~ race + smoke + ptd + ht + ui + ftv,
    family = binomial, data = bwt)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6472  -0.7661  -0.5860   1.0250   2.0819
```

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.0183     0.4551  -4.435 9.23e-06 ***
raceblack     1.0177     0.5096   1.997   0.0458 *
raceother     0.9427     0.4360   2.162   0.0306 *
smokeTRUE     0.8362     0.4046   2.067   0.0387 *
ptdTRUE       1.2983     0.4592   2.827   0.0047 **
htTRUE        1.3702     0.6513   2.104   0.0354 *
uiTRUE        0.8041     0.4562   1.763   0.0780 .
ftv1         -0.4606     0.4678  -0.985   0.3247
ftv2+        -0.0272     0.4338  -0.063   0.9500
    Null deviance: 234.67  on 188  degrees of freedom
Residual deviance: 202.76  on 180  degrees of freedom
AIC: 220.76
```

So

$$LRT = 202.76 - 195.48 = 7.287$$

$$p\text{-value} = P[\chi^2_2 \geq 7.287] = 0.026$$

So it appears that at least one of the two variables is useful in prediction low birth weight probabilities.

The likelihood ratio test is easily implemented in **R** using the `anova` function. When dealing with with generalized linear models, such as logistic regression, this function gives an Analysis of Deviance table instead of an Analysis of Variance table.

```
> anova(birthwt2.glm, birthwtall.glm, test="Chisq")
Analysis of Deviance Table

Model 1: low ~ race + smoke + ptd + ht + ui + ftv
Model 2: low ~ age + lwt + race + smoke + ptd + ht + ui + ftv
  Resid. Df Resid. Dev  Df Deviance P(>|Chi|)
1       180    202.763
2       178    195.476   2    7.287     0.026
```

The option `test` takes possible values `"Chisq"`, `"F"`, and `"Cp"`. For most generalized linear models, `"Chisq"`, as used here, is the correct choice. Further discussion of this will come later.

The likelihood ratio test can be used to test for single parameter, such as in

$$H_0 : \beta_2 = 0 \qquad \text{vs} \qquad H_A : \beta_2 \neq 0$$

For example, testing whether `lwt` is interesting, assuming all other variable will be in the model, can be done by

```
> birthwt3.glm <- update(birthwtall.glm, . ~ . - lwt)
>
> anova(birthwt3.glm, birthwtall.glm, test="Chisq")
Analysis of Deviance Table

Model 1: low ~ age + race + smoke + ptd + ht + ui + ftv
Model 2: low ~ age + lwt + race + smoke + ptd + ht + ui + ftv
  Resid. Df Resid. Dev  Df Deviance P(>|Chi|)
1       179    200.949
2       178    195.476   1    5.474     0.019
```

So it appears that `lwt` is important with this test.

We now have 2 tests for examining hypotheses involving single βs, the Wald (z) test and the likelihood ratio test. So this is analogous to the `t` and `F` tests in linear regression.

However unlike the linear regression case, these two tests are not the exactly the same. For example

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.82302    1.24471   0.661  0.50848
age         -0.03723    0.03870  -0.962  0.33602
lwt         -0.01565    0.00708  -2.211  0.02705 *
raceblack    1.19241    0.53597   2.225  0.02609 *
raceother    0.74069    0.46174   1.604  0.10869
smokeTRUE    0.75553    0.42502   1.778  0.07546 .
ptdTRUE      1.34376    0.48062   2.796  0.00518 **
htTRUE       1.91317    0.72074   2.654  0.00794 **
uiTRUE       0.68019    0.46434   1.465  0.14296
ftv1        -0.43638    0.47939  -0.910  0.36268
ftv2+        0.17901    0.45638   0.392  0.69488
```

the p-values for `lwt` don't match up here, though they are similar.

Generally this will be the case. The 2 tests will give similar answers. If there is a big difference, usually the likelihood ratio will give a better answer, as the asymptotic approximation to the χ^2distribution will be better.

R will give an Analysis of Deviance table for any model, such as

```
> birthwt.al.glm <- glm(low ~ age + lwt, family=binomial,
     data=bwt)

> anova(birthwt.al.glm, test="Chisq")
Analysis of Deviance Table

Model: binomial, link: logit

Response: low

Terms added sequentially (first to last)


      Df Deviance Resid. Df Resid. Dev P(>|Chi|)
NULL                        188    234.672
age    1    2.760        187    231.912     0.097
lwt    1    4.789        186    227.123     0.029
```

This table examines the set of nested models where variables are added sequentially. Its an analogue to the sequential sums of squares in linear regression.

Similarly the order variables get entered matters.

Due to this fact, usually only the last line in the output is useful.

```
> birthwt.al.glm <- glm(low ~ lwt + age, family=binomial,
     data=bwt)

> anova(birthwt.la.glm, test="Chisq")
Analysis of Deviance Table

Model: binomial, link: logit

Response: low

Terms added sequentially (first to last)


      Df Deviance Resid. Df Resid. Dev P(>|Chi|)
NULL                      188    234.672
lwt    1    5.981         187    228.691      0.014
age    1    1.567         186    227.123      0.211
```

Aside: Likelihood Ratio Test in Linear Regression

The likelihood ratio test is a general procedure so it could be applied to a normal based regression model.

Remember that the framework for the F-test and the LRT are the same, i.e. the same pairs of models are being compared, so the two tests are trying to get at the same thing.

In fact, the F-test is a special case of the LRT. However instead of using an asymptotic approximation to get p-values, an exact probability result is used instead.

```
> bodyfat.full.lm <- lm(Bodyfat ~ Tricep + Thigh, data=bodyfat)
> bodyfat.red.lm <- lm(Bodyfat ~ 1, data=bodyfat)
> bodyfat.full.glm <- glm(Bodyfat ~ Tricep + Thigh, data=bodyfat,
+   family=gaussian)
> bodyfat.red.glm <- glm(Bodyfat ~ 1, data=bodyfat,
+   family=gaussian)
```

```
> anova(bodyfat.red.lm, bodyfat.full.lm, test="F")
Analysis of Variance Table

Model 1: Bodyfat ~ 1
Model 2: Bodyfat ~ Tricep + Thigh
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1     19 495.39
2     17 109.95  2    385.44 29.797 2.774e-06 ***

> anova(bodyfat.red.glm, bodyfat.full.glm, test="F")
Analysis of Deviance Table

Model 1: Bodyfat ~ 1
Model 2: Bodyfat ~ Tricep + Thigh
  Resid. Df Resid. Dev Df Deviance      F    Pr(>F)
1        19     495.39
2        17     109.95  2   385.44 29.797 2.774e-06 ***
```

```
> anova(bodyfat.red.glm, bodyfat.full.glm, test="Chisq")
Analysis of Deviance Table

Model 1: Bodyfat ~ 1
Model 2: Bodyfat ~ Tricep + Thigh
  Resid. Df Resid. Dev Df Deviance P(>|Chi|)
1        19     495.39
2        17     109.95  2   385.44 1.146e-13
```

The equivalence between the two setting comes from the fact that $Deviance = SSE$ in normal based models.