Stat 5421, Fall 2006: Drinking and Birth defects

This example concerns a prospective study of 17,114 pregnant women, classified by their alcohol consumption (determined by survey at approximately 12 weeks of pregnancy) and whether or no congenital malformation is present or absent in the resulting birth. The data are discussed in both Chapter 3 and in Chapter 5, around page 179.

```r
> aa179 <- data.frame(AlchConsumpt = c("0", "<1", "1-2", "3-5", "6+"), alcohol = c(0, 0.5, 1.5, 4, 7), Present = c(48, 38, 5, 1, 1), Absent = c(17066, 14464, 788, 126, 37))
> aa179$frac <- aa179$Present/(aa179$Absent + aa179$Present)
> aa179$logit <- log(aa179$frac/(1 - aa179$frac))
> print(aa179)

<table>
<thead>
<tr>
<th>AlchConsumpt</th>
<th>alcohol</th>
<th>Present</th>
<th>Absent</th>
<th>frac</th>
<th>logit</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0</td>
<td>48</td>
<td>17066</td>
<td>0.002804721</td>
<td>-5.873642</td>
</tr>
<tr>
<td>&lt;1</td>
<td>0.5</td>
<td>38</td>
<td>14464</td>
<td>0.002620328</td>
<td>-5.941832</td>
</tr>
<tr>
<td>1-2</td>
<td>1.5</td>
<td>5</td>
<td>788</td>
<td>0.006305170</td>
<td>-5.060060</td>
</tr>
<tr>
<td>3-5</td>
<td>4.0</td>
<td>1</td>
<td>126</td>
<td>0.007874016</td>
<td>-4.836282</td>
</tr>
<tr>
<td>6+</td>
<td>7.0</td>
<td>1</td>
<td>37</td>
<td>0.026315789</td>
<td>-3.610918</td>
</tr>
</tbody>
</table>
```

We see that malformation is extremely rare in all alcohol consumption categories, with only one event in the highest drinking category. This will mean that the distribution of test statistics will be suspect.

**Null model**  We will look at a series of models for these data. We start first with the null model that the probability of “present” is independent of number of drinks.

```r
> m1 <- glm(cbind(Present, Absent) ~ 1, data = aa179, family = binomial())
> summary(m1)

Call:
  glm(formula = cbind(Present, Absent) ~ 1, family = binomial(), data = aa179)

Deviance Residuals:
  1       2       3       4       5
 -0.1237 -0.5372  1.5686  0.8703  1.6372

Coefficients:          Estimate  Std. Error   z value  Pr(>|z|)
 (Intercept)   -5.8558     0.1038    -56.39    <2e-16

(Dispersion parameter for binomial family taken to be 1)
Null deviance: 6.202 on 4 degrees of freedom
Residual deviance: 6.202 on 4 degrees of freedom
AIC: 26.829

Number of Fisher Scoring iterations: 4

> (estimated.odds <- exp(coef(m1)))

(Intercept)  
0.002863212

> (pearsons.X2 <- sum(resid(m1, type = "pearson")^2))

[1] 12.08205

> 1 - pchisq(c(m1$deviance, pearsons.X2), m1$df.residual)

[1] 0.18456227 0.01675140

The estimated intercept is −5.8558, so this is estimated log-odds of malformation, assuming the women are a sample from a population. Exponentiating this gives the odds, of about 0.00286.

The value of $G^2$ for this model is 6.202 with 4 df, while the value of $X^2 = 12.08$, also with 4 df. These two statistics differ greatly because of the wildly different cell sizes, and lead to wildly different conclusions. Which should you believe?

An alternative is the bootstrap we learned earlier:

> chisq.test(aa179[, 3:4], simulate = T)

Pearson's Chi-squared test with simulated p-value
(based on 2000 replicates)

data:  aa179[, 3:4]
X-squared = 12.0821, df = NA, p-value = 0.03448

This gives a simulated p-value of about 0.03, which is more likely to be believable. Thus we have only very weak evidence against the null hypothesis that drinking is associated with birth defects. This is contrary to “common knowledge.” So, what’s up? Read on...
Linear function of alcohol  The next model we fit is more complex than \( m_1 \), and assumes that log-odds of defects increase linearly with number of drinks. Following Agresti, a number of drinks has been assigned to each of the five categories (this can be criticized, but not by me), and then fit:

```
> summary(m2 <- update(m1, ~. + alcohol))
```

Call:
\[
\text{glm(formula = cbind(Present, Absent) ~ alcohol, family = binomial(),}
\]
\[
data = aa179)
\]

Deviance Residuals:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>0.5921</td>
<td>-0.8801</td>
<td>0.8865</td>
<td>-0.1449</td>
<td>0.1291</td>
</tr>
</tbody>
</table>

Coefficients:

|                 | Estimate | Std. Error | z value | Pr(>|z|) |
|-----------------|----------|------------|---------|----------|
| (Intercept)     | -5.9605  | 0.1154     | -51.637 | <2e-16   |
| alcohol         | 0.3166   | 0.1254     | 2.523   | 0.0116   |

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 6.2020  on 4 degrees of freedom
Residual deviance: 1.9487 on 3 degrees of freedom
AIC: 24.576

Number of Fisher Scoring iterations: 4

```
> exp(coef(m2))
```

(Intercept)  alcohol
0.002578723  1.372399123

First, look at \( G^2 = 1.95 \) with 3 df. This measures lack of fit, so since \( G^2 \) is less than its df, then the \( p \)-value is at least 0.5, and there is no evidence that this model is inadequate. The coefficient for alcohol os 0.3166, with a (Wald) \( p \)-value of 0.0116, suggesting that defects are not independent of drinking, but rather increase with drinking. The \( p \)-value here is smaller than the \( p \)-value for \( m_1 \) because the test for \( m_1 \) is omnibus, testing for any type of dependence, whereas the test for \( m_2 \) is testing for exclusively for dependence that can be modeled as a linear increase or possibly decrease with the number of drinks. Indeed, one could justify dividing the \( p \)-value by 2 to get 0.0055 because a one-sided test is of interest here (we do not believe that incidence could decrease with drinking).

How large is the effect? The coefficient for \( \text{alcohol} \) is 0.3166. The “+” sign means incidence increases with drinking, and \( \exp(0.3166) = 1.37 \) is the relative risk: increasing intake by one drink multiplies the odds of defect by 1.37, or a 37% increase. We can get confidence intervals, too:
> confint(m2)

Waiting for profiling to be done...

2.5 %  97.5 %

(Intercept) -6.19301781 -5.7397025
alcohol 0.01862779 0.5234339

> exp(confint(m2, parm = c("alcohol")))

Waiting for profiling to be done...

2.5 %  97.5 %
1.018802 1.687813

The confidence interval suggests the relative risk is between a 1% increase and a 69% increase.

**Note on confidence intervals (primarily for Stat 8421)**

The textbook describes primarily the **Wald confidence intervals** which are computed from the familiar pattern of

\[
\text{estimate} \pm \text{mult} \times \text{standard error}
\]

where the value of \text{mult} is 1.96 for normal-based intervals at the 95% level, as is appropriate in fitting binomial regression. The \text{confint} method in R does something quite different, and computes **likelihood-ratio based confidence intervals**, based on computing a **profile log-likelihood**. Here is a (very) brief outline of the method for the model \text{m2}.

1. let \( \ell(\beta_0, \beta_1) \) be the log-likelihood function for model \text{m2}. A 95% joint confidence **region** for \((\beta_0, \beta_1)\) simultaneously is given by the set of all \((\beta_0, \beta_1)\) such that:

\[
2 \left\{ \ell(\hat{\beta}_0, \hat{\beta}_1) - \ell(\beta_0, \beta_1) \right\} \leq \chi^2(2, .95)
\]

where \(\chi^2(2, .95)\) is the 95% point of the \(\chi^2\) distribution with 2 df.

2. We are interested in a confidence interval for say \(\beta_1\), not a joint interval for the two parameters. For this, suppose we fix a value of \(\beta_1 = b_1\) and define

\[
\ell_1(b_1) = \max_{\beta_0} \ell(\beta_0, b_1)
\]

This is called the profile log-likelihood for \(b_1\), as it “removes” \(\beta_0\) by maximizing over it to get a one-dimensional “profile.”

3. The likelihood interval is the set of all \(b_1\) such that

\[
2 \left\{ \ell(\hat{\beta}_0, \hat{\beta}_1) - \ell_1(b_1) \right\} \leq \chi^2(1, .95)
\]

This interval can’t in general be found using a formula, but rather require using an iterative computational method that in general requires a lot more computations than just fitting a glm.
**Saturated model**  The final model we fit is called the *saturated model*, which allows each value of alcohol to have its own parameter (since we have five levels of alcohol, we will fit five parameters. To fit this model, we need to create a *factor*, which is a *variable with categories rather than numeric values*. R automatically assumes that a text variable is a factor, so we can fit this last model as follows:

```r
> class(aa179$AlchConsumpt)
[1] "factor"
> levels(aa179$AlchConsumpt)
[1] "0" "<1" "1-2" "3-5" "6+
> summary(m3 <- update(m1, ~AlchConsumpt))
```

**Call:**
```
glm(formula = cbind(Present, Absent) ~ AlchConsumpt, family = binomial(),
     data = aa179)
```

**Deviance Residuals:**

```
[1] 0 0 0 0 0 0
```

**Coefficients:**

```
                      Estimate Std. Error  z value  Pr(>|z|)
(Intercept)          -5.87364   0.14454 -40.637 <2e-16
AlchConsumpt<1        -0.06819   0.21743  -0.314  0.7538
AlchConsumpt1-2       0.81358   0.47134   1.726  0.0843
AlchConsumpt3-5       1.03736   1.01431   1.023  0.3064
AlchConsumpt6+        2.26272   1.02368   2.210  0.0271
```

(Dispersion parameter for binomial family taken to be 1)

- Null deviance: 6.2020e+00 on 4 degrees of freedom
- Residual deviance: 1.8163e-13 on 0 degrees of freedom
- AIC: 28.627

**Number of Fisher Scoring iterations:** 4

This model fits exactly—it is saturated—so all the residuals are exactly equal to zero. We rarely fit this model, but I’ve done so to (1) illustrate using factors, and (2) have this model available for model comparison, in the next section.
**Model comparison**  Model comparison is done in an *analysis of deviance table*, which parallels the analysis of variance table you learned when using linear models. Indeed, the function in R is called `anova`.

```
> anova(m1, m2, m3, test = "Chisq")
```

Analysis of Deviance Table

| Model  | cbind(Present, Absent) ~ | Resid. Df | Resid. Dev | Df | Deviance | P(>|Chi|) |
|--------|-------------------------|-----------|------------|----|----------|----------|
| 1      | 1                       | 4         | 6.2020     |    |          |          |
| 2      | alcohol                 | 3         | 1.9487     | 1  | 4.2533   | 0.0392   |
| 3      | AlchConsumpt            | 0         | 1.816e-13  | 3  | 1.9487   | 0.5831   |

These three models are *nested* with m1 the smallest, m2 using part of the information in alcohol, and m3 using all the information. The difference between the residual deviance for m2 and m1 tests the hypothesis NH: m1 versus AH: m2. It has an approximate $\chi^2$ distribution with df equal to the change in df. We get a $p$-value of about 0.04 for this test. This is the likelihood ratio test that is equivalent to the Wald test described in the section on m2. The large $p$-value for comparing m3 to m2 suggests that m2 is an adequate model.

**Closing comments**  (1) If pregnant, don’t drink. (2) What would happen if we pooled the 6+ category with the 4–5 category? (3) What would happen if the one defect in the 6+ category with in fact not a defect, or almost equivalently, if that pregnancy were deleted from the sample, reducing the sample size to 17,113? How comfortable would you be in setting public policy based on the outcome of one pregnancy?