A sample of 32 pregnant rats were divided into two groups at random, one group getting a control drug treatment and the other a potentially active drug treatment. Each rat gave birth to a litter of between 5 and 13 young. The number $y$ of surviving rats at 21 days was recorded (Collett, page 5).

```
> loc <- url("http://www.stat.umn.edu/~sandy/courses/5421/data/collett005.txt")
> tab14 <- read.table(loc, header = T)
> tab14$frac <- tab14$y/tab14$n
> tab14

   litter trt y  n      frac
  1      1   T 13 13 1.0000000
  2      2   T 12 12 1.0000000
  3      3   T  9  9 1.0000000
  4      4   T  9  9 1.0000000
  5      5   T  8  8 1.0000000
  6      6   T  8  8 1.0000000
  7      7   T 12 13 0.9230769
  8      8   T 11 12 0.9166667
  9      9   T  9 10 0.9000000
 10     10   T  9 10 0.9000000
 11     11   T  8  9 0.8888889
 12     12   T 11 13 0.8461538
 13     13   T  4  5 0.8000000
 14     14   T  5  7 0.7142857
 15     15   T  7 10 0.7000000
 16     16   T  7 10 0.7000000
 17     17   C 12 12 1.0000000
 18     18   C 11 11 1.0000000
 19     19   C 10 10 1.0000000
 20     20   C  9  9 1.0000000
 21     21   C 10 11 0.9090909
 22     22   C  9 10 0.9000000
 23     23   C  9 10 0.9000000
 24     24   C  8  9 0.8888889
 25     25   C  8  9 0.8888889
 26     26   C  4  5 0.8000000
 27     27   C  7  9 0.7777778
 28     28   C  4  7 0.5714286
 29     29   C  5 10 0.5000000
 30     30   C  3  6 0.5000000
 31     31   C  3 10 0.3000000
 32     32   C  0  7 0.0000000
```

```
> library(lattice)
> print(dotplot(frac ~ trt, data = tab14))
```
GLM approach

First, assume no litter effects, and fit usual logistic regression models:

```R
> m1 <- glm(cbind(y, n - y) ~ 1, data = tab14, family = binomial)
> summary(m2 <- update(m1, ~. + trt))$coef

Estimate Std. Error  z value Pr(>|z|)
(Intercept)  1.221991  0.1980694  6.169511  6.850162e-10
trtT         0.961247  0.3297927  2.914700  3.560302e-03
```

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

Analysis of Deviance Table

| Resid. Df | Resid. Dev | Df | Deviance | Pr(>|Chi|) |
|-----------|------------|----|----------|-----------|
| 1         | 31         | 1  | 95.203   |           |
| 2         | 30         | 1  | 86.187   | 9.016     | 0.003     |
```

The change in $G^2$ statistic suggests that treatment and control differ. However, even $m2$ with a treatment effect does not appear to fit the data, as $G^2 = 86.187$ is much larger than its df= 30. The $G^2$ statistic consists in part of the variation between litters. As suggested by the figure above, the lack-of-fit is due to between litter variation in the control group. Modeling between litter variation is likely to be useful here.

GLMM

The generalized linear mixed model is fit using the `lmer` function in the `lme4` library. Unfortunately, there are bugs present in earlier version of this package, and it use it for this example you need to update to R version 2.4.0 or newer (go to http://cran.us.r-project.org/), and a version of `lme4` dated December 3, 2006 or later (inside R, type `update.packages()`, choose a download site, and answer yes to all questions).
> library(lme4)
> c1 <- lmer(cbind(y, n - y) ~ 1 + (1 | litter), data = tab14,
+    family = binomial)
> c2 <- update(c1, ~trt + .)
> anova(c1, c2)

Data: tab14
Models:
  c1: cbind(y, n - y) ~ 1 + (1 | litter)
  c2: cbind(y, n - y) ~ trt + (1 | litter)
          Df  AIC  BIC logLik Chisq Chi Df  Pr(>Chisq)
  c1       2 70.90 73.83 -33.45
  c2       3 69.86 74.26 -31.93  3.04  1    0.08138

The apparent treatment effect as a change in survival rate is now much weaker, with significance level around 0.08. The two models that are fit above are given by:

\[
\log(\text{odds}) = \beta_0 + u
\]

\[
\log(\text{odds}) = \beta_0 + \beta_1 \text{trt} + u
\]

(1)

respectively, where \(u\) is a normal random variable with mean 0 and variance \(\sigma_u^2\), with a different value for each litter, and \(\text{trt}\) has the value 1 for treatment and 0 for control.

> print(c1, corr = FALSE)

Generalized linear mixed model fit using Laplace
Formula: cbind(y, n - y) ~ 1 + (1 | litter)
Data: tab14
Family: binomial(logit link)
  AIC  BIC logLik deviance
  70.9 73.83 -33.45  66.9
Random effects:
  Groups Name        Variance Std.Dev.
  litter (Intercept) 2.0321   1.4255
  number of obs: 32, groups: litter, 32

Estimated scale (compare to 1 ) 0.9135356

Fixed effects:
  Estimate Std. Error    z value  Pr(>|z|)
  (Intercept)  2.1014     0.3182   6.604     4.01e-11

This all seems reasonable, except that, from the initial graph, the excess variance seems to be relevant only for control, not for treatment. Thus suggests fitting separate variances for the two groups. We need to be careful about how this is done. In the command below, the phrase \((0 + \text{trt} | \text{litter})\), means: don’t fit an overall variance term (the zero part), but then allow a separate variance term for each value of \(\text{trt}\). This amounts to fitting the same model as (1) above, except that the variance of \(u\) is now \(\sigma_C^2\), for control litters and \(\sigma_T^2\) for treatment litters. We want to fit a separate variance for each treatment level, but not fit a covariance between them because they are fit to completely separate sets of observations. This requires a separate random effect for a dummy variable for the treatment group, and a dummy variable for the control group.
> trtC <- ifelse(tab14$trt == "C", 1, 0)
> trtT <- ifelse(tab14$trt == "T", 1, 0)
> c3 <- update(c2, ~trt + (0 + trtT | litter) + (0 + trtC |
+ litter))
> anova(c2, c3)

Data: tab14
Models:
c2: cbind(y, n - y) ~ trt + (1 | litter)
c3: cbind(y, n - y) ~ trt + (0 + trtT | litter) + (0 + trtC | litter)

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>Chisq</th>
<th>Chi Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>c2</td>
<td>3</td>
<td>69.865</td>
<td>74.262</td>
<td>-31.932</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c3</td>
<td>4</td>
<td>68.246</td>
<td>74.109</td>
<td>-30.123</td>
<td>3.6189</td>
<td>1</td>
<td>0.05713</td>
</tr>
</tbody>
</table>

> print(c3, corr = FALSE)

Generalized linear mixed model fit using Laplace
Formula: cbind(y, n - y) ~ trt + (0 + trtT | litter) + (0 + trtC | litter)
Data: tab14
Family: binomial(logit link)

AIC  BIC  logLik  deviance
68.25 74.11 -30.12 60.25            

Random effects:
Groups Name   Variance Std.Dev.
litter trtT   0.22819  0.47769
litter trtC   3.08391  1.75611

number of obs: 32, groups: litter, 32; litter, 32

Estimated scale (compare to 1 )  0.948109

Fixed effects:

|            | Estimate | Std. Error | z value | Pr(>|z|) |
|------------|----------|------------|---------|----------|
| (Intercept)| 1.6768   | 0.5125     | 3.272   | 0.00107  |
| trtT       | 0.5878   | 0.5925     | 0.992   | 0.32118  |

Summary

Between-litter heterogeneity is evident, but the evidence, in spite of the graph, that variance is larger for C than for T is almost significant (the p-value is 0.06). After adjusting for heterogeneity, we cannot conclude that the survival rates are different for the two groups. Quite likely, the experiment is simply too small to show differences, given the variation between litters.