SS9055B: Generalized Linear Models

Section 6: Poisson Log-Linear Models

1 **Objectives**

2 In this lecture we will review the Poisson log-linear model. By the end of this lecture you
3 should be able to:

4 • show that the Poisson log-linear model falls into the GLM framework,
5 • derive the likelihood equations, covariance matrix, and deviance function,
6 • apply Poisson log-linear models to data in contingency tables and interpret the results,
7 • explain the use of offsets in modeling rates, and
8 • fit Poisson log-linear models in `R`

9 **Introduction**

10 As motivation we will consider the following data which come from a study of malignant
11 melanoma conducted by Roberts et al.~(1981). Records on patients with melanoma were
12 collected for one year and each patient was categorized according to the site (head and neck,
13 trunk, or extremities) and histological type of their tumor. The primary question of interest
14 is whether or not there is an association between the location of the tumor and the type.

15 The original data considers 4 different types of tumours and 3 different sites. Here are the
16 counts that were observed:

17

	Tumour Site		
Tumour Type	Head and Neck	Trunk	Extremities
Hutchinson's melanotic freckle	22	2	10
Superficial spreading melanoma	16	54	115
Nodular	19	33	73
Indeterminate	11	17	28

One way to test for an association is to apply the standard chi-square test. For this data, the test statistic is $\chi^2 = 65.81$ on 6 DF which provides very strong evidence that there is an association between the type of tumour and where on the body it is locate ($p < .0001$). It is clear that tumour site and tumour type are associated. Looking at the table we can see that HMF is more likely to occur on the head and neck whereas the other three types of tumours are more likely to occur on the trunk and extremities. However, to understand this trend fully we need to fit a model. Since we are dealing with count data it makes sense to fit a Poisson model.

Interpretation of Coefficients

One advantage of the Poisson log-linear model over the logistic regression model is that the regression coefficients have a very clear and easy to understand interpretation[1]. Suppose that two individuals, with ids 1 and 2, share the same values of the predictors except for the kth which differs by one unit:

$$x_{2j} = \left\{ \begin{array}{ll} x_{1j} & j \neq k \\ x_{1j} + 1 & j = k \end{array} \right. .$$

If we let μ_1 and μ_2 represent the mean response for the two individuals then

$$\log(\mu_2) - \log(\mu_1) = \beta_k$$

or equivalently

$$\mu_2 = \mu_1 e^{\beta_k}.$$

This shows that covariates in a log-linear model have a multiplicative effect on the mean. Increasing x_k by one unit while adjusting for the other predictors (i.e., holding their values fixed) increases the mean e^{β_k} times. More generally, increasing/decreasing x_k by u units while adjusting for the other predictors in the model increases/decreases the mean $e^{\beta_k u}$ times. We can also construct a $(1 - \alpha)100\%$ confidence interval for this multiplicative effect simply by exponentiating the endpoints of a confidence interval for β_k. If (L_k, U_k) is a $(1 - \alpha)100\%$ confidence interval for β_k then we can have confidence $(1 - \alpha)100\%$ that increasing x_k by u units while holding the predictors fixed will increase the mean by between $e^{L_k u}$ and $e^{U_k u}$ times.

Contingency Tables

If the predictors that we are considering are all categorical, as in the example, then the data can be displayed in a contingency table. We can then apply Poisson log-linear models to study the dependence between the different predictors. However, we are going to change the notation slightly to make it easier to interpret the model coefficients.

---

[1]Don't worry. There will be a few other complications to deal with!

2×2 **Contingency Tables**

We will start by considering the simplest situation in which we have two predictors each with two levels, which I'll label as A and B. So far we have indexed the response with a single number, i. There problem with using this notation in a contingency table is that there are many different ways in which the cells could be numbered. We might label the counts as:

x_1	x_2 A	B		x_1	x_2 A	B		x_1	x_2 A	B
A	y_1	y_2	or	A	y_1	y_3	or	A	y_1	y_3
B	y_3	y_4		B	y_2	y_4		B	y_4	y_2

The latter is decidedly odd, but the point is that if you label the values with one index then you can't easily know which cell in the contingency table y_i belongs to and so you can't know how to interpret the regression coefficients without constantly referring to the indexing.

Instead, we will label the counts according the levels of the two predictors such that y_{ij} represents the count from the cell with the ith level of the first predictor and the jth level of the second predictor. Equivalently, i and j refer to the row and column in the tables above. Similarly, we can encode the two binay predictors (often called dummy variables) such that

$$x_{i1} = \begin{cases} 0 & \text{for the first level of the first predictor (A)} \\ 1 & \text{for the second level of the first predictor (B)} \end{cases}$$

and

$$x_{j2} = \begin{cases} 0 & \text{for the first level of the second predictor (A)} \\ 1 & \text{for the second level of the second predictor (B)} \end{cases} .$$

The cell means are now defined by the linear predictor

$$\log(\mu_{ij}) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{j2} + \beta_3 x_{i1} x_{j2}.$$

which can naturally be displayed in a 2×2 table with i indexing the row and j indexing the column:

	$j = 1$	$j = 2$
$i = 1$	β_0	$\beta_0 + \beta_2$
$i = 2$	$\beta_0 + \beta_1$	$\beta_0 + \beta_1 + \beta_2 + \beta_3$

The interpretations of the coefficients int he model now become clear: β_1 represents the row effect (the effect of going from $i = 1$ to $i = 2$ when $j = 1$), β_2 represents the column effect (the effect of going from $j = 1$ to $j = 2$ when $i = 1$), and β_3 represents the interaction (the

68 difference in the row/column effect when we $j = 2/i = 2$). Keep in mind that these effects on
69 the scale of the linear predictor (i.e., the log-scale).

70 The usual question of interest when working with contingency tables is whether or not there
71 is an association. I.e., whether or not the probability an observation falls in one column or
72 another depends on which row it is in, or vice versa. This can be assessed by looking at the
73 ratio of means within each of the rows with the null hypothesis being that the ratios in two
74 rows are equal. The probability that an observation falls within the first column given that it
75 is in row i is

$$p_{i|j=1} = \frac{\mu_{i1}}{\mu_{i1} + \mu_{i2}}$$

76 and the probability that an observation falls within column 2 given that it is in row i is

$$p_{i|j=2}\frac{\mu_{i2}}{\mu_{i1} + \mu_{i2}}.$$

77 The hypothesis of no interaction states that these probabilities do not vary by i. Mathemati-
78 cally,

$$H_0 : p_{1|j=1} = p_{1|j=2} \text{ and } p_{2|j=1} = p_{2|j=2}.$$

79 Note that the first equality implies the second since $p_{1|j=1} = 1 - p_{2|j=1}$, so we can restate the
80 hypothesis simply as

$$H_0 : p_{1|j=1} = p_{1|j=2}$$

81 which is equivalent to

$$H_0 : \frac{\exp(\beta_0)}{\exp(\beta_0) + \exp(\beta_0 + \beta_1)} = \frac{\exp(\beta_0 + \beta_2)}{\exp(\beta_0 + \beta_2) + \exp(\beta_0 + \beta_1 + \beta_2 + \beta_3)}$$

82 which is true if and only if

$$H_0 : \beta_3 = 0.$$

83 Hence, testing for an association in the contingency table is equivalent to testing the hypotheses
84 $H_0 : \beta_3 = 0$ verse $H_a : \beta_3 \neq 0$.

General Two-Way Contingency Tables

86 More generally, suppose that we have two categorical predictors with $I > 1$ and $J > 1$ levels,
87 respectively, so that the data can be put into an $I \times J$ contingency table. Let Y_{ij} denote the
88 number of observations that fall into cell i, j, and μ_{ij} be the mean of the value so that so that
89 $Y_{ij} \sim \text{Poisson}(\mu_{ij})$ $i = 1, \ldots, I$ and $j = 1, \ldots, J$. We will then write the linear predictor as

$$\log(\mu_{ij}) = \beta_0 + \beta_{1i} + \beta_{2j} + \beta_{3ij}.$$

90 Note that this model is overparametrized (there are $1 + I + J + IJ$ parameters and only IJ
91 cells in the table) and so we need to impose some restrictions on the model. The most common
92 solution is to assume that $\beta_{11} = 0$, $\beta_{21} = 0$, and $\beta_{3ij} = 0$ whenever $i = 1$ or $j = 1$. With these
93 restrictions, the table of linear predictors (the log of the means) becomes:

	1	2	$\cdots$	J
1	β_0	$\beta_0 + \beta_{22}$	$\cdots$	$\beta_0 + \beta_{2J}$
2	$\beta_0 + \beta_{12}$	$\beta_0 + \beta_{12} + \beta_{22} + \beta_{322}$		$\beta_0 + \beta_{12} + \beta_{2J} + \beta_{32J}$
$\vdots$				
I	$\beta_0 + \beta_{1I}$	$\beta_0 + \beta_{1I} + \beta_{22} + \beta_{3I2}$		$\beta_0 + \beta_{12} + \beta_{2J} + \beta_{3IJ}$

The coefficients β_{1i} model the row effects – the differences in the log-means between cell $i, 1$ and cell $1, 1$; the coefficients β_{2j} models the column effects – the difference in the log-means between cell $1, j$ and cell $1, 1$; and the parameter β_{3ij} models the interaction within the 2×2 subtable formed by the 4 cells in rows 1 and i and columns 1 and j. Again, the two variables are associated if the ratios of probabilities across the columns differ down the rows (or vice versa). Following the same argument as above shows that this occurs if $\beta_{3ij} \neq 0$ for any i and j. Hence, the null hypothesis for the test of association is

$$H_0 : \beta_{3ij} = 0 \quad \text{for all} \quad i = 1, \ldots, I; j = 1, \ldots, J.$$

If the null hypothesis is true then the variables are independent and otherwise the variables are dependent.

Three-Way Contingency Tables

Poisson log-linear models can be used to assess associations between any number of categorical variables with any number of categories. However, the indexing gets complicated quickly. Here we will consider classification by three categorical variables, which I will call X, Y, and Z, so that the counts form a three-dimensional table with cells y_{ijk}, $i = 1, \ldots, I$, $j = 1, \ldots, J$, and $k = 1, \ldots, K$. We must now model the cell means in terms of the main effects, two-way, and three-way interactions so that

$$\log(\mu_{ijk}) = \beta_0 + \beta_{1i} + \beta_{2j} + \beta_{3k} + \beta_{3ij} + \beta_{4ik} + \beta_{5jk} + \beta_{6ijk}.$$

This model is overparametrized again and so we need to apply constraints. Following the method for two-way tables, we can obtain an identifiable model by setting any coefficient equal to zero whenever any of the indexes, i, j, or k, is equal to 1.

Relationships in a three-way table are more complicated than in a two-way table because the association between two factors may depend on the value of the third. Suppose that we are interested in the association between X and Y. If the relationship between X and Y does not depend on the value of Z then we say that these two variables have a uniform (or homogeneous) association. The hypothesis of uniform association is equivalent to testing whether or not the three way interaction is significant

$$H_0 : \beta_{6ijk} = 0, \text{ for all } i, j, k.$$

120 If X and Y are independent for a specific level of Z, say level k^*, then we say that X and Y are
121 conditionally independent given $Z = k^*$. Conditional independence occurs if $\beta_{ij} + \beta_{ijk^*} = 0$
122 for all i and j. If all interactions involving X and Y are equal to zero, i.e., $\beta_{ijk} = 0$ and $\beta_{ij} = 0$
123 for all i, j, k, then X and Y are independent.

Example

125 The code in the file `melanoma.R` fits a Poisson log-linear model to the melanoma data and
126 tests for an interaction:

127 The results indicate that there is a significant association between the type and location of
128 the tumour ($p < .001$). To understand the nature of this association we need to look at the
129 estimates for the individual terms in the interaction:

```
130 # A tibble: 6 x 7
131   term              estimate std.error statistic        p.value conf.low conf.high
132   <chr>                <dbl>     <dbl>     <dbl>          <dbl>    <dbl>     <dbl>
133 1 typeInd:siteH&N      -1.72     0.522     -3.30 0.000957         -2.79    -0.730
134 2 typeNo:siteH&N       -2.13     0.460     -4.64 0.00000352       -3.07     -1.26
135 3 typeSSM:siteH&N      -2.76     0.465     -5.93 0.00000000300    -3.71     -1.88
136 4 typeInd:siteTru       1.11     0.833      1.33 0.183            -0.367     3.06
137 5 typeNo:siteTru       0.815     0.802      1.02 0.310            -0.587     2.72
138 6 typeSSM:siteTru      0.853     0.792      1.08 0.281            -0.523     2.75
```

139 The results show that the interaction term, β_{3ij}, is signficant, and negative, for all types of
140 tumours when the site is Head and Neck ($p < .001$ for all three terms), but is not significant
141 when the site is Trunk ($p > .15$ for all three terms.

142 To interpret these values we need to note that the base leves of the type of tumour are
143 Hutchinson's Melanotic Freckle (HMF) and the base level of the location is Extremeties. These
144 levels have been chosen because they come first alphabetically. The fact that the interactions
145 including Head and Neck are all negative means that the relative rates of occurence of the
146 SSM, No, and Ind types of tumour are all significanlty lower on the Head and Neck than on
147 the Extremeties, which must mean that the relative rate of occurrence of HMF is higher on
148 the Head and Neck than on the Extremeties. On the other hand, the fact that the interaction
149 terms involving Trunk are not significant indicates that there is no evidence of a difference
150 in the relative rates of occurence for any of the types of tumour between the Trunk and the
151 Extremeties. Keep in mind that the interaction terms tell us only about the relative rates –
152 not the absolute rates. Based on these tests we can say nothing about how many tumours
153 occurred at a given site, only that proportions varied or not. In order to compare the absolute
154 rates we would need to look at the lower order effects, but most often this is not of interest.

Another way to see the differences is to examine the fitted values, which are shown in Figure Figure 1. The fitted values do represent the absolute rates, and you can see that there seem to be many more cases of tumours on the Extremeties than on the Trunk than on the Head and Neck. However, t is clear from this plot that the relative rates of the different types of cancer are similar between the Extremities and Trunk, but show a very different pattern for the head and neck. The code for creating this plot is included in the R script.
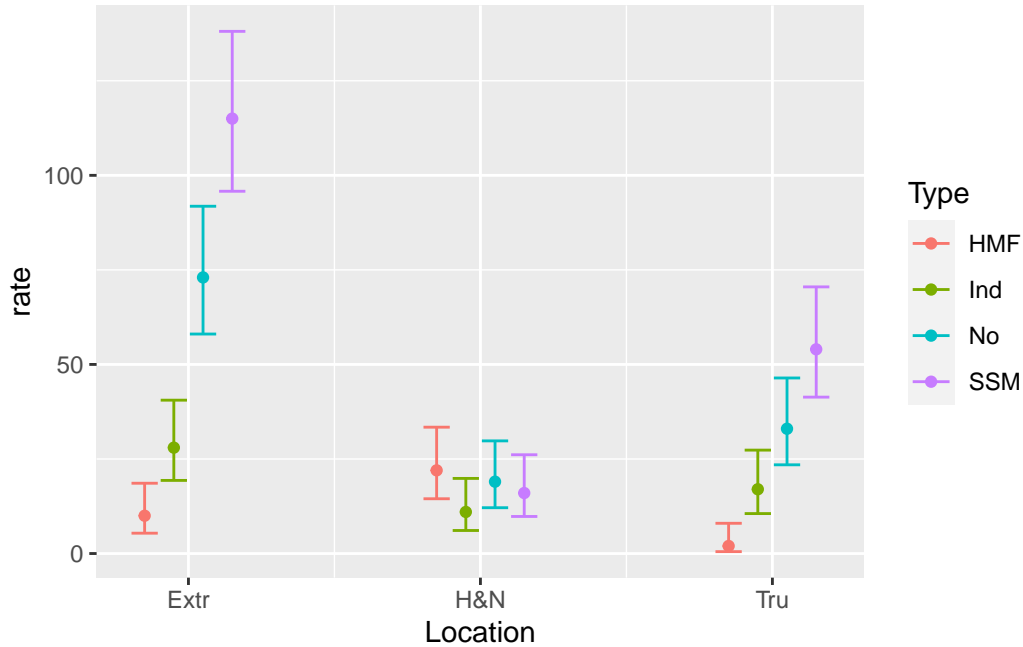
Figure 1: Fitted rates for each type of cancer by site (Extremities, Head and Neck, and Trunk).

Poisson Models for Rates

In some problems the data may represent rates or densities not counts. This often occurs if the counts we have were obtained from time periods of different lengths or from regions of different size. We can't compare the counts directly in these cases because we'd likeley expect that the count will be proportional to the size of the measurement unit, but this can easily be accounted for in a Poisson log-linear model. Suppose that we obtain counts $Y_1, \ldots, Y_N$ from units with size $a_1, \ldots, a_N$. A reasonable model might assume that the mean count is proportional to the size of the unit, after accounting for the other predictors in the model. That is, we might assume that $Y_i \sim \text{Poisson}(\mu_i)$ with

$$\mu_i = a_i \exp(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip})$$

170　which is equivalent to saying that

$$\log(\mu_i) = \log(a_i) + \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}.$$

171　This is equivalent to our log-linear Poisson model except that we have included a new predictor,
172　$\log(a_i)$, which has a fixed coefficient of 1. This new predictor is called an offset, and we can
173　model the data by including this predictor and fixing the value of the coefficient. This can be
174　done in `R` by specifying the log of the predictor with the `offset` argument in your call to the
175　`glm()` function.

176　**Example: Ducks**

177　As an example of Poisson log-linear models we will analyse data from **?**. The aim of this
178　study was to determine if the number of ducks seen on a lake was determined by the overall
179　productivity of the lake. In particular, Dr. E. Fudd wanted to know if the number of ducks
180　on the lake was accurately predicted by his faviourite measure of productivity, the size of the
181　biggest fish he'd ever caught in the lake. As an avid outdoorsman, Dr.~Fudd was able to
182　compile data on fifty lakes near his home in Orlando, Florida. The data is contained in the
183　file `ducks\_on\_lakes.csv`.

184　The following plot illustrates the relationship between the size of the biggest fish he caught in
185　each lake (in kilograms) and the number of ducks he counted on the lake.
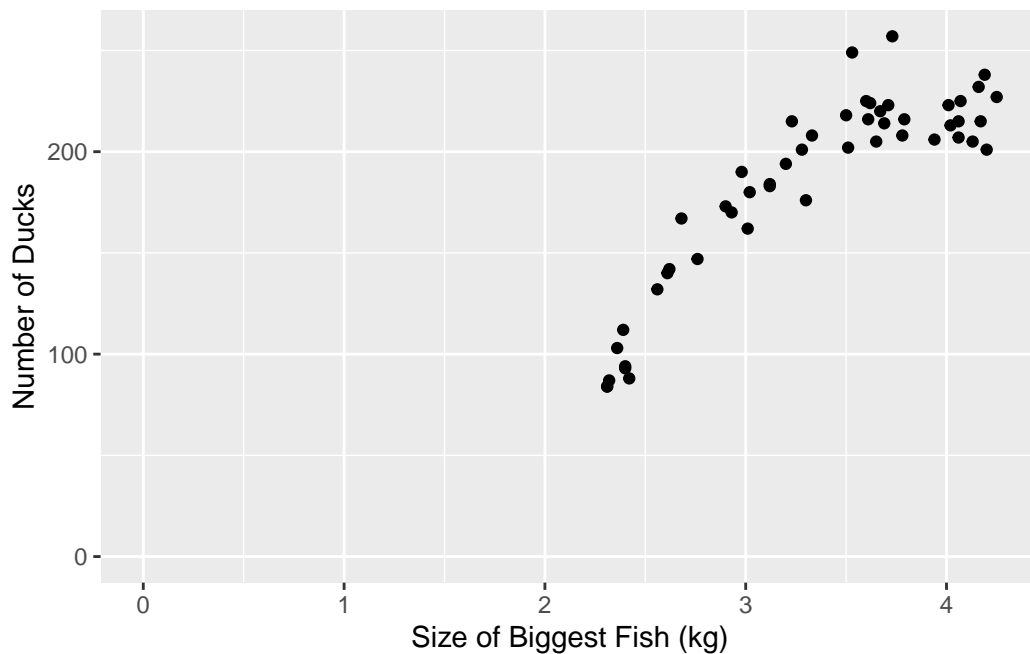
Model 1

The following code fits a Poisson log-linear model to the number of ducks on each lake as a function of the size of the biggest fish.

```
## Poisson Model
glm1 <- glm(Ducks ~ Fish,family=poisson(),data=ducks)
summary(glm1)
```

```
Call:
glm(formula = Ducks ~ Fish, family = poisson(), data = ducks)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.91026    0.06158   63.50   <2e-16 ***
Fish         0.38387    0.01750   21.93   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 690.04  on 49  degrees of freedom
Residual deviance: 192.28  on 48  degrees of freedom
AIC: 546.76

Number of Fisher Scoring iterations: 4
```

```
tidy(glm1, conf.int = TRUE)
```

```
# A tibble: 2 x 7
  term          estimate std.error statistic   p.value conf.low conf.high
  <chr>            <dbl>     <dbl>     <dbl>     <dbl>    <dbl>     <dbl>
1 (Intercept)      3.91     0.0616      63.5 0            3.79      4.03
2 Fish             0.384    0.0175      21.9 1.34e-106    0.350     0.418
```

The results indicate that the the natural (base e) log of the mean number of ducks seen on a lake increases by .38 (95%CI=.35,.42) for every 1~kg increase in the size of the biggest fish that Dr.~Fudd had ever caught on the lake. The effect is highly significant ($p < .001$). However, the residual deviance of 192.28 on 48 degrees of freedom raises some concern about the fit of the model ($p < .001$). Looking at the residual plot below there seems to be some sort of a trend and it seems that the variance may change from left to right, though it is somewhat hard to tell.

Plot residuals
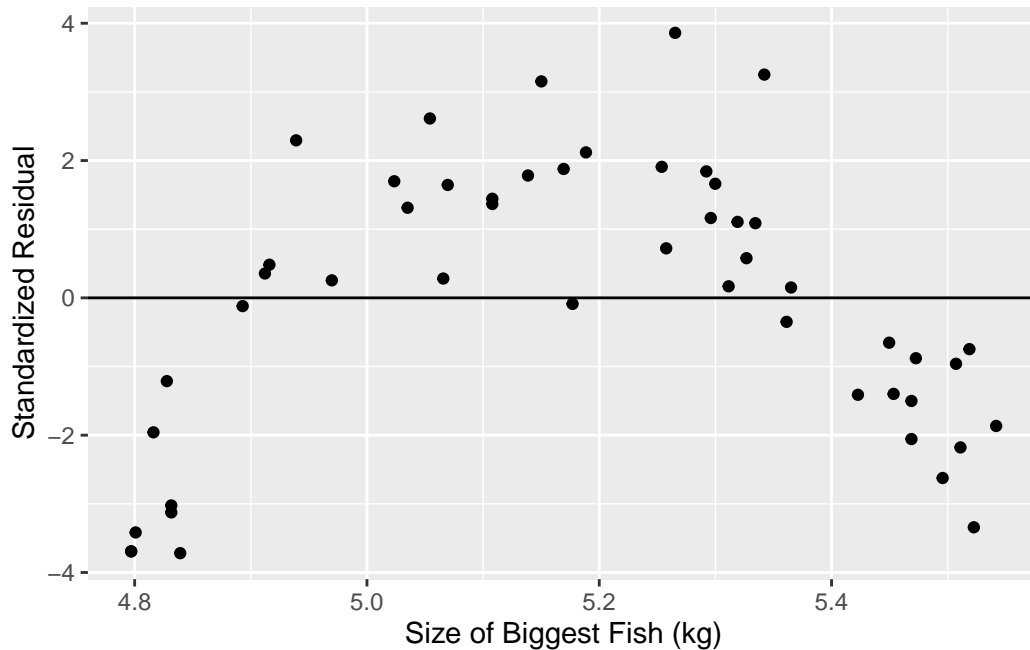
Model 2

The reason for the lack of fit is that our analysis has not accounted for the size of the lake. Suppose that two lakes are the same in all aspects except that one is twice as big. We'd expect about twice as many ducks on the bigger one. The following plots illustrates the relationship between the size of the biggest fish and the density of ducks (i.e., ducks per unit area). This plot now seems to show the opposite effect with more ducks on lakes with smaller fish (and, hence, lower productivity according to Dr.~Fudd's criterion).

After seeing this plot it make sense to model the density of ducks on each lake as a function of the size of the biggest fish. Unfortunately, the density of ducks is no longer integer valued and so we can't consider this quantity as the response in a Poisson log-linear model. However, we can achieve exactly the same thing by including the log of the lakes' area as an offset:

```
## Analysis with offset
glm2 <- glm(Ducks ~ Fish,family=poisson(),offset=log(Area),data=ducks)
summary(glm2)
```

Call:
glm(formula = Ducks ~ Fish, family = poisson(), data = ducks,
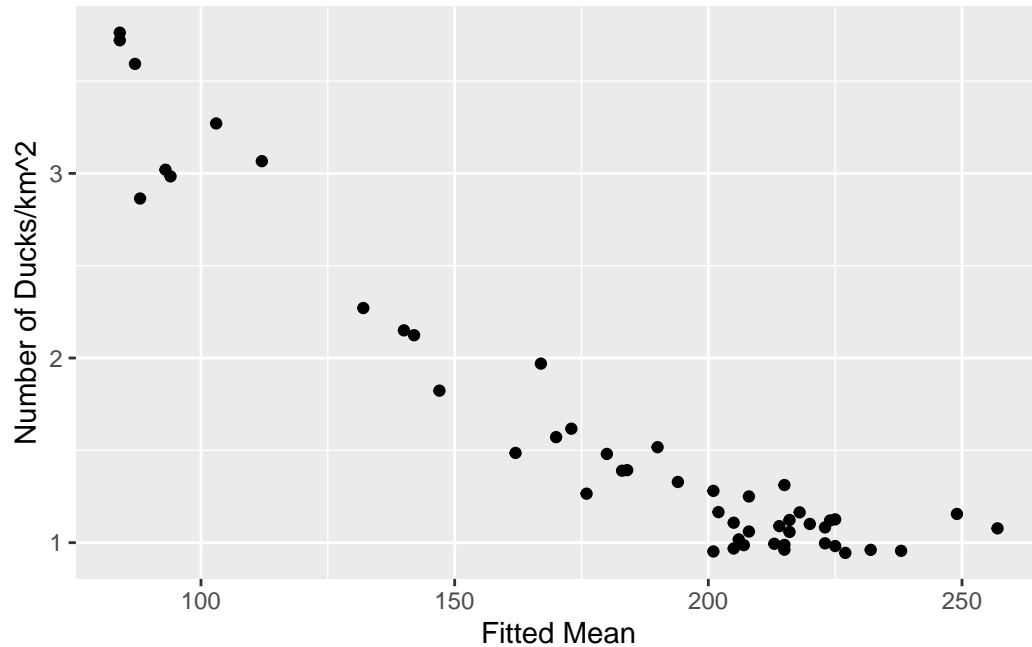
```
235        offset = log(Area))
236
237   Coefficients:
238                Estimate Std. Error z value Pr(>|z|)
239   (Intercept)  6.06150    0.06536   92.75   <2e-16 ***
240   Fish        -0.51398    0.01861  -27.62   <2e-16 ***
241   ---
242   Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
243
244   (Dispersion parameter for poisson family taken to be 1)
245
246       Null deviance: 767.572  on 49  degrees of freedom
247   Residual deviance:  43.218  on 48  degrees of freedom
248   AIC: 397.7
249
250   Number of Fisher Scoring iterations: 3
```

```
tidy(glm2, conf.int = TRUE)
```

```
252   # A tibble: 2 x 7
253     term         estimate std.error statistic   p.value conf.low conf.high
```

```
254    <chr>            <dbl>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>
255  1 (Intercept)      6.06      0.0654      92.7 0                 5.93       6.19
256  2 Fish            -0.514     0.0186     -27.6 6.13e-168        -0.550     -0.477
```

The new results indicate that if the size of two lakes are the same then the log of the mean number of ducks decreases by .51 (95%CI=.48,.55) for each 1~kg increase in the size of the biggest fish. I.e., if two lakes, call them Lakes 1 and 2, have the same size and the biggest fish in Lake 2 weighs 1~kg more than the biggest fish in Lake 1 then we expect there to be $e^{-.51} = ..60$ times as many ducks on Lake 2 (95%CI=.58,.62). The residual deviance of 43.22 on 48 degrees of freedom provides absolutely no concern about the fit of the model ($p = .33$), and there are no issues visible in the residuals plot below[2].
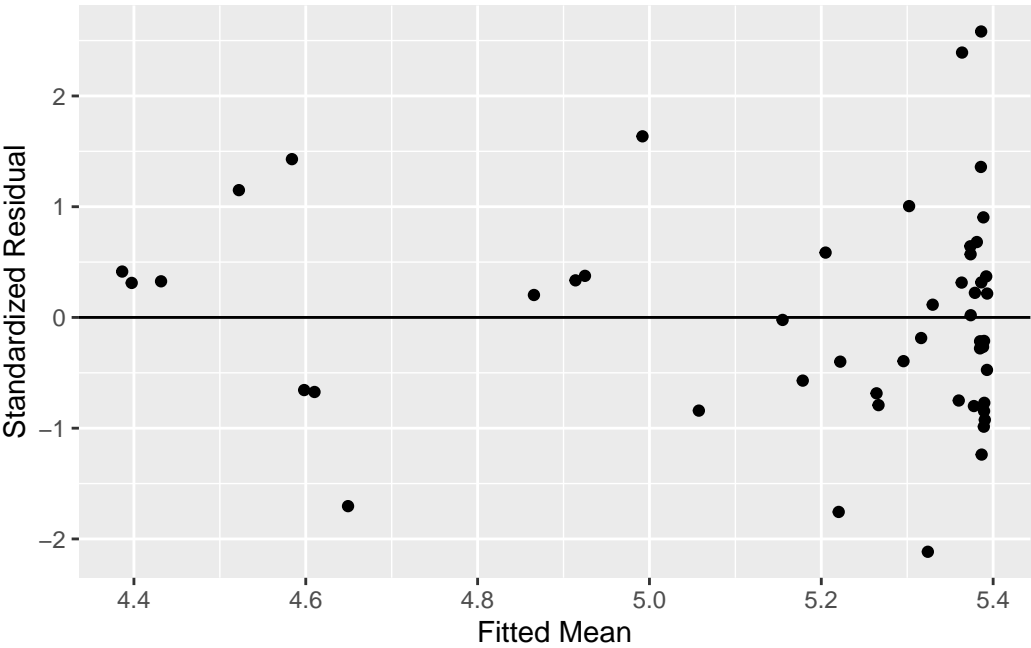
Contrary to Dr.~Fudd's hypothesis, it seems that the density of ducks on a lake is inversely proportional to the size of the biggest fish head ever caught on the lake. For two lakes of the same size, the one with the bigger fish will have fewer ducks. Why would this be true? Big fish eat ducks!

---

[2]I know this because I know the model fits the data perfectly.