Exposure to ionizing radiation is recognized as a cancer risk. In the United States, EPA sets guidelines specifying upper limits on the amount of exposure that can be tolerated by humans without incurring a significant increase in cancer risk. The estimation of cancer risks following exposure to ionizing radiation has been the subject of several studies during the past few decades. These estimates have been based largely on data for survivors of the atomic bombings of Hiroshima and Nagasaki, supplemented in some instances by information from studies of medically-exposed groups.

In these studies, the response variable of interest often is a count of rare events such as the number of new cases of lung cancer occurring in a population over a certain period of time. The aim of regression in the analysis of data obtained from these studies is to model the number of outcomes (cancer cases or deaths attributable to cancer) using a number of explanatory variables believed to be related to the incidence of cancer. The data discussed below report the number of cancer deaths among the survivors of the Nagasaki and Hiroshima atomic bombs.

The Figure to the right shows the data. The response variable is the log-ratio of cancer deaths to the person-years (hundreds) at risk, computed by adding (over individuals) the number of years of risk over which the cancer deaths are counted. Responses are grouped according to two variables: their estimated exposure (measured in rads), either 0, 25, 75, 150, 250 or 400, and the number of years after exposure when the response variable was observed.

For the groups with low exposure, there is a positive association between the log-ratio of deaths to number at risk and years since exposure. This apparent linear relationship appears less clear for the higher exposure groups. The log-ratio of deaths to number at risk appears to be higher overall for the high exposure groups than low exposure.

The objective of regression analysis is to estimate the rate of cancer deaths per person per year at a particular exposure level after accounting for years of exposure. Once this rate

1 This is the sum total of all years spent by all persons in the category.
is estimated, then the risk after one year or ten years of exposure can be estimated.

When the response variable was normally distributed, \( E(Y|x) \) could be linked to a set of explanatory variables using a linear function of the form \( E(Y|x) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p \). In the case of binomial regression, the upper and lower bounds on \( \pi \) imposed a constraint that was resolved by modeling \( \logit(\pi) \) as linear function of the form \( \logit(\pi) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p \).

The normality assumption of multiple linear regression was lost, and so also was the assumption of constant variance. Without these assumptions, the \( F \) and \( t \) tests have no basis and the drop-in-deviance test is generally used for significance testing.

The regression coefficients were interpreted as increasing or decreasing the log odds of an event, and \( \exp(\hat{\beta}_i) \) was interpreted as the estimated (multiplicative) increase in the odds of the event for a one unit increase (or decrease) in the explanatory variable \( x_i \).

Another (different) constraint is imposed by count data. Counts are all non-negative integers without an upper bound (in principle). Particularly for rare events, the Poisson distribution is more appropriate than the normal since the Poisson mean is positive and unbounded, and because the variance of the Poisson random variable is not independent of the mean (as it is with normal random variables), but instead increases with the mean. Specifically, \( E(Y|x) = \mu = \sigma^2 = \text{var}(Y|x) \). The natural link between the mean of count data and a linear regression function is the natural logarithm: \( \log[E(Y|x)] = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p \). A typical Poisson regression model expresses the log outcome rate as a linear function of a set of predictors.

**Conditions, properties and assumptions in Poisson regression**

1. The logarithm of the mean (or rate of occurrence of the event) changes linearly with the explanatory variables, i.e.,
   \[
   \log[E(Y|x)] = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p.
   \]

2. Changes in the rate from combined effects of different exposures or risk factors are multiplicative, i.e.,
   \[
   E(Y|x) = \exp[\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p] \\
   = e^{\beta_0}e^{\beta_1 x_1} \times \cdots \times e^{\beta_p x_p}.
   \]

A one-unit change in \( x_i \) results in a multiplicative change in \( E(Y|x) \) of \( e^{\beta_i} \), if all other variables are held fixed.

---

\(^2\text{Variance that increases with the mean is prevalent with count data.}\)
3. The response variable has variance equal to the mean at all levels of the explanatory variables. Residual plots (residuals versus the fitted value) can be used to investigate the conformity of the data to this property.

4. Observations are independent.

**The Poisson distribution**

Let \( Y \) count the number of events occurring in a fixed time period.\(^3\) The possible realizations of \( Y \) are \( 0, 1, 2, \ldots \). The probability of a particular realization can be computed using the following formula.

\[
P(Y = y) = \frac{\mu^y e^{-\mu}}{y!}, \quad y = 0, 1, 2, \ldots,
\]

The expected value and variance of \( Y \) are:

\[
\mu = E(Y) \text{ and } \text{var}(Y) = \mu.
\]

Since \( \mu \) is both the expected value and variance of \( Y \), and \( Y \) is a count, \( \mu > 0 \).

**The log-linear regression model for Poisson responses**

A regression model of \( E(Y|x) = \mu \) for a Poisson random variable should produce fitted values or predictions that are within the range of possible values for \( \mu \), that is, within the interval \((0, \infty)\). A conventional linear regression model of \( E(Y|x) = \mu \) will produce fitted values that are negative. The solution is to model the log transformation of \( \mu \), as there are no upper or lower bounds on the range of \( \log(\mu) \) when \( \mu \in (0, \infty) \). The log-linear regression model is

\[
\log(\mu|x) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p.
\]

It’s often helpful to transform a fitted model back to the original scale on which the counts were measured. The inverse transformation of the natural logarithm is the anti-log \( \exp[\log(\mu)] = \mu = e^{\log(\mu)} \).\(^4\) After transforming back to the original scale, the fitted linear model becomes a multiplicative model:

\[
\log(\mu|x) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p \\
\Rightarrow \mu = \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p)
\]

\[
= e^{\beta_0} e^{\beta_1 x_1} \times \cdots \times e^{\beta_p x_p}
\]

\[
= \gamma_0 \gamma_1^{x_1} \times \cdots \times \gamma_p^{x_p}
\]

\(^3\)The fixed time period can be replaced with an area of fixed size, for example, a circular plot.

\(^4\)\( e \approx 2.71828 \) is called Euler’s number.
where $\gamma_0 = e^{\beta_0}$, $\gamma_1 = e^{\beta_1}$ and $\gamma_p = e^{\beta_p}$. The interpretation of a fitted coefficient $\hat{\beta}_i$ is that a one unit change in $x_i$ changes $\mu$ by a factor of $e^{\hat{\beta}_i}$ given all other variables are held constant. For instance, if $\hat{\beta}_i = .18232$, then a one unit change in $x_i$ increases $\mu$ by a multiplicative factor of $e^{.18232} = 1.2$. If $\hat{\beta}_i = -.182$, then a one unit change in $x_i$ decreases $\mu$ by a multiplicative factor of $e^{-1.18232} = 1/1.2 = .833$.

Parameter estimates are computed using the method of maximum likelihood and model assessment is based on the observed likelihood function or, equivalently, the deviance. Hypothesis testing is much the same as with logistic regression with binomial response variables. The Wald statistic is usually sufficiently accurate to rely upon it for hypothesis tests involving a single parameter, though the drop-in-deviance test should be used if the majority of the counts are small.

The cancer deaths data set requires special attention to the total risk (hundreds of person-years) associated with each observed count. In particular, the objective is to model $\mu_i/r_i$—the expected count per unit of risk. Accounting for the number of person years at risk (the risk set) is critical since it is not the same for every count. The log-linear model that accounts for the risk set is

$$
\log \left( \frac{\mu_i}{r_i} \right) = \beta_0 + \beta_1 x_{i,1} + \cdots + \beta_p x_{i,p}.
$$

The person-years at risk ($r_i$) is a term in the linear portion of the model with a known coefficient, and the value of the coefficient is one. This situation is common enough in log-linear regression to deserve a name—it’s called an offset.

Treating years after exposure as a quantitative variable (alternatively, it might be treated as a factor and avoid the presumption that the relationship between years after exposure and $\log(\mu)$ is linear) yields the fitted model shown in Table 2.

Table 1: Coefficients and standard errors obtained from a log-linear regression of number of cancer death on years after exposure and exposure. Person-years at risk was treated as an offset. Residual deviance is 56.044 on 39 degrees of freedom, $n = 42$.

| Variable         | Coefficient | Std. Error | Wald statistic ($Z$) | $P(Z > |z|)$ |
|------------------|-------------|------------|----------------------|-------------|
| Intercept        | -3.60       | .151       | -23.7                | <.0001      |
| Years after exposure | .0829   | .00643     | 12.9                 | <.0001      |
| Exposure         | .00183      | .000514    | 3.56                 | .001        |

No coefficient appears in the Table 2 for the risk set though the fitted model is associated with each observed count. The log-linear model that accounts for the risk set $r_1, r_2, \ldots, r_{42}$
is

\[
\log (\mu_i | x) = \log(r_i) - 3.60 + .0829 x_{i, \text{years}} + .00183 x_{i, \text{exposure}} \quad (1)
\]

\[
\Rightarrow \log \left( \frac{\mu_i}{r_i} \right) = -3.60 + .0829 x_{i, \text{years}} + .00183 x_{i, \text{exposure}}. \quad (2)
\]

Formula (1) shows that setting the number of person-years at risk as an offset produces a model of the cancer risk after adjusting for the person-years at risk.

If years of exposure are fixed, the estimated increase in the one-year cancer rate given a 10 rad increase in exposure is estimated to be \(\exp(.00183 \times 10) = 1.0185\) (or 1.85%) since

\[
\frac{\mu(\text{exposure} = x + 10, r_i = 1)}{\mu(\text{exposure} = x, r_i = 1)} = \frac{\exp[\log(1) - 3.60 + .0829 x_{\text{years}} + .00183 (x_{\text{exposure}} + 10)]}{\exp[\log(1) - 3.60 + .0829 x_{\text{years}} + .00183 x_{\text{exposure}}]} = \exp(.00183 \times 10).
\]

The risk is assumed to be 1 (I am considering the risk for a single person over the course of a single year). An approximate 95% confidence interval for the increase in the cancer rate given a 10 rad increase in exposure is

\[
[\exp\{10 \times (.00183) - 10 \times 1.96\times.000514\}, \exp\{10 \times (.00183) + 10 \times 1.96\times.000514\}] = [1.00826, 1.0288].
\]

The analysis above is tentative as there is some question of whether the effect of exposure is best modeled as linear term that does not interact with years of exposure. A simple alternative is to consider a polynomial effect of exposure on risk by introducing the square of exposure into the model. The drop-in-deviance test for the significance of the quadratic term yields \(D = 3.42\) on 1 degree of freedom, and p-value = .0645. A fuller assessment of the linearity assumption can be obtained by treating years since exposure as a factor with seven levels. The fitted model is shown below. The goodness of fit test yielded \(D = 50.106\) on 34 degrees of freedom and p-value = .0369. The principal benefit of this model is the Figure above showing that there is a slight quadratic trend in the estimated effect of exposure on cancer risk.
Table 2: Coefficients and standard errors obtained from a log-linear regression of number of cancer death on years after exposure and exposure. Person-years at risk was treated as an offset. Residual deviance is 50.106 on 34 degrees of freedom, \( n = 42 \).

| Variable     | Coefficient | Std. Error | Wald statistic (Z) | \( P(Z > |z|) \) |
|--------------|-------------|------------|--------------------|-----------------|
| Intercept    | -3.21       | .186       | -17.20             | < .0001         |
| Years(8–11)  | .233        | .253       | 0.923              | .356            |
| Years(12–15) | .551        | .237       | 2.32               | .020            |
| Years(16–19) | 1.24        | .213       | 5.854              | < .0001         |
| Years(20–23) | 1.40        | .210       | 6.685              | < .0001         |
| Years(24–27) | 1.73        | .204       | 8.52               | < .0001         |
| Years(28–31) | 2.03        | .200       | 10.17              | < .0001         |
| Exposure     | .00183      | .000439    | 4.17               | < .0001         |

The residual deviance is 50.106 (df= 34) and a test for goodness of fit supports the alternative hypothesis stating that the fitted model fit is worse than the saturated model fit is supported by the data since \( p=\text{value}= .0369 \). It appears that all of the \( p \)-values in Table 2 are biased downwards (because of lack of model fit). This topic is discussed in more detail below.

**Residuals**

As with all generalized linear models, a Pearson residual is the difference between an observed response value \( y_i \) and its fitted model (on the original scale of the data) divided by its estimated standard deviation. In this case of Poisson regression, the Pearson residuals are

\[
r_i^{P} = \frac{y_i - \hat{\mu}_i}{\hat{\sigma}(\hat{\mu}_i)} = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i}}.
\]

The deviance residual is

\[
r_i^{D} = \text{sign}(y_i - m_i) \sqrt{2 \left[ y_i \log \left( \frac{y_i}{\hat{\mu}_i} \right) - y_i + \hat{\mu}_i \right]}.
\]
The figure to the right shows the deviance residuals. All but one residual is between $-1$ and $2$ and there is no clear deviation from a random scatter of points (allowing for the fact that most fitted values are small). As there are more residuals associated with small values, there is greater variation among the small values. The visual evidence of non-constant variance is limited, and so the distribution of the residuals support the use of the Poisson model.

*Goodness-of-fit*

As with binomial regression, the fit of the model can be assessed using the residual deviance. The residual deviance provides a goodness-of-fit test comparing the fitted model (or constrained model) to the saturated model. The alternative hypothesis states that at least one of the parameters in the saturated model (and absent from the constrained model) is non-zero, and the null hypothesis states that all additional parameters are zero, and hence the model of interest fits the data as well as any other model. The hope is that there is insufficient data to conclude that the saturated model fit is better than the constrained model fit.

As with logistic regression, the theoretical saturated model contains one parameter for every observation and yields the maximal model fit. Specifically,

$$\text{Residual deviance} = -2 \left[ \log(L_{\text{constrained}}) - \log(L_{\text{saturated}}) \right]$$

where $L_{\text{constrained}}$ is the likelihood function evaluated using the model of interest with parameters $\beta_0, \ldots, \beta_p$ to estimate $\mu_i$, $i = 1, \ldots, n$, and $L_{\text{saturated}}$ is the likelihood function computed using one parameter for every observation (and hence fitting the data exactly). In practice, since the saturated model fits the data exactly, $\hat{\mu}_i = y_i$, $i = 1, \ldots, n$, and

$$\log(L_{\text{saturated}}) = \sum_{i=1}^{n} y_i \log(\hat{\mu}_i) - \hat{\mu}_i - \log(y!)$$

$$= \sum_{i=1}^{n} y_i \log(y_i) - y_i - \log(y!)$$

Returning to the radiation data, the residual deviance, 50.106 on 34 degrees of freedom, is larger than expected, and the p-value is obtained by determining the probability of observ-
ing a chi-square random variable with 34 degrees of freedom greater than 50.106. Using the function call \(1-pchisq(50.1,34)\), the p-value is approximately .037. It cannot be said that the constrained Poisson model fits the data.

Evidence of lack-of-fit is commonplace when using Poisson regression, in part because when the expected value of the response \(\mu_i\) is inaccurately modeled, then so are the variances of the responses. Unmeasured effects, clustering of events, and other unidentified contaminating influences can introduce extra variation (more variation than predicted by the Poisson distribution and the relation \(E(Y) = \text{var}(Y)\)). Three consequence result:

1. Parameter estimates remain approximately unbiased.
2. The maximum likelihood standard errors (e.g., \(\hat{\sigma}(\hat{\beta}_i)\)) tend to be biased downwards.
3. Tests tend to produce p-values that are smaller than they ought to be. This consequence is closely related to 2, since the Wald statistic is \(z = \hat{\beta}_i/\hat{\sigma}(\hat{\beta}_i)\). The numerator is approximately correct but the denominator tends to be too small. The Wald statistic tends to be too large in magnitude resulting in a p-value that is too small.

The problem extends to the drop-in-deviance test since the drop-in-deviance test is supposed to be comparing the likelihood of a model that fits well (the unconstrained model) to a constrained version of the model. If the unconstrained model does not fit well, then the drop-in-deviance test may be inaccurate. The degree of inaccuracy depends (in part) on the extent of lack-of-fit.

The Quasi-likelihood approach A solution to the problem of lack-of-fit sometimes is possible by introducing an additional parameter into the model that accounts for extra-Poisson variation (essentially variation in the response variable that is not explained by the fitted model and the fundamental Poisson property \(E(Y) = \text{var}(Y)\)).

The quasi-likelihood approach defines an extra-Poisson variation model:

\[
\begin{align*}
\text{var}(Y|x) &= \psi E(Y|x) \\
\log[E(Y|x)] &= \log(\mu) \\
&= \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p
\end{align*}
\]

The parameter \(\psi\) (pronounced ”sigh”) is the dispersion parameter. If \(\text{var}(Y|x) = \mu\), then \(\psi = 1\); otherwise extra-Poisson variation is present.

---

5In addition, the Poisson model may only approximate the true distribution (unknown) of the response variables.
Extra-Poisson variation should be expected if there are variables that are believed to affect the response variable but have not been observed, and hence, are missing from the data set. The ionizing radiation data set is missing many variables that may affect the risk of cancer death such as genetics, diet and exposure to carcinogens. If the observed counts tend to be small (say less than 5), extra-Poisson variation tends not to be a problem.

Extra-Poisson variation is detected by comparing sample variances to sample means whenever there are multiple observations \( y_1, \ldots, y_j \) obtained at a particular combination of explanatory variable values. Such observations are usually called replicate observations when the data are obtained from a designed experiment. If replicates are available in groups \( 1, 2, \ldots, g \), then a plot of the sample variances \( s_1^2, \ldots, s_g^2 \) against the sample means \( \bar{y}_1, \ldots, \bar{y}_g \) will indicate extra-Poisson variation if the points do not follow a line with slope 1 and intercept 0 (allowing for the inevitable variation about this line).

Information can be gleaned from the residual deviance associated with a rich model (containing all reasonable explanatory variables without particular attention paid to significance). A large residual deviance comparing the rich model to the saturated model indicates that extra-Poisson variation is present. Finally, examining the residuals from the rich model will indicate whether there are a few outliers that are responsible for the lack of fit.

If there is some evidence of extra-Poisson variation, then it is best to proceed assuming that extra-Poisson variation is present and then follow the quasi-likelihood approach.

**Inferences when extra-Poisson variation is present**

1. The estimator \( \hat{\psi} \) of \( \psi \) is the sum of the squared Pearson residuals\(^6\) divided by the residual degrees of freedom\(^7\). If the model fit is adequate, then the residual deviance is approximately \( \chi^2_{df} \) in distribution (recall \( E(\chi^2_{df}) = df \)).

2. Let \( \hat{\sigma}_{ML}(\hat{\beta}_i) \) denote the usual maximum likelihood standard error (reported by \( R \) using the `summary(glm.obj)` function call). The quasi-likelihood standard errors are \( \hat{\sigma}_{QL}(\hat{\beta}_i) = \sqrt{\hat{\psi} \hat{\sigma}_{ML}(\hat{\beta}_i)} \).

3. All \( t \)-tests and confidence intervals are based on the coefficient estimates (\( \hat{\beta}_i \)) and the quasi-likelihood standard errors. The convention is to use the \( t \)-distribution to obtain critical values for confidence intervals and to compute \( p \)-values (using the degrees of freedom \( n-p-1 \) associated with the model from which \( \hat{\psi} \) was computed.) A significance

---

\(^6\)Sometimes the deviance residuals are used instead

\(^7\)This is the same estimator of \( \sigma^2 \) in the case of multiple regression with normally distributed residuals.
test for $\beta_i$ uses the statistic
\[
t = \frac{\hat{\beta}_i}{\hat{\sigma}_{QL}(\hat{\beta}_i)} \sim t_{df}.
\]
A 100$(1 - 2\alpha)$ confidence interval for $\beta_i$ is
\[
\hat{\beta}_i \pm t^*_{df} \hat{\sigma}_{QL}(\hat{\beta}_i).
\]
where $t^*_{df}$ is $\alpha$ quantile from the $t_{df}$ distribution.

4. The drop-in-deviance statistic is scaled to reflect extra-Poisson variation. The distribution of the scaled statistic is approximately $F$. The numerator degrees of freedom is $k$, the difference in degrees of freedom between the unconstrained and constrained models. The denominator degrees of freedom is $n - p - 1$, the degrees of freedom associated with the unconstrained model. This $F$ statistic is analogous to the extra-sums-of-squares $F$-statistic used in linear regression except that the deviance residuals are used rather than the ordinary residuals. The statistic is
\[
F = \frac{\log(L_{\text{unconstrained}}) - \log(L_{\text{constrained}})}{k \hat{\psi}}
\]
where $L_{\text{constrained}}$ and $L_{\text{unconstrained}}$ are the likelihood functions for two nested models and $k$ is the difference in degrees of freedom between the constrained and unconstrained models. The unconstrained model contains all parameters (and variables) in the constrained model and an additional $k$ variables (and hence $k$ additional parameters). $H_0$ states that all of the parameters associated with the additional variables are zero and $H_a$ states that at least one of the additional parameters is not zero.

Extra-Poisson variation is accommodated in R by changing the family to quasipoisson. For example, the significance tests in Table 2 are of dubious accuracy since there is evidence of over-dispersion in the fitted model shown in the table. The function call
\[
\text{glm(formula = y ~ D$Years + Exposure + offset(log(D$Risk)), family = quasipoisson())}
\]
produces a fitted model shown in Table 3 below.

For example, from Table 2, the conventional Poisson regression model yields $\hat{\beta}_{\text{exposure}} = 0.00183, \hat{\sigma}(\hat{\beta}_{\text{exposure}}) = 0.000439$. Hence, $z = 4.17 \Rightarrow$ p-value $< .0001$. Since there is evidence of extra-Poisson variation, the test is overly liberal and ought to be replaced by a test that accounts for over-dispersion.

The dispersion parameter is the sum of Pearson residuals divided by the degrees of freedom (same idea as in linear regression except that the measure of lack-of-fit is the residual
deviance instead of the sum of the squared Pearson residuals). The Pearson residuals are extracted using the function call `residuals(glm.2,"pearson")` where `glm.2` is the fitted unconstrained model. The result is

\[
\hat{\psi} = \frac{\sum_{i=1}^{n}(r_i^P)^2}{n - p - 1} = \frac{48.326}{34} = 1.421.
\]

The test of significance for radiation exposure, corrected for extra-Poisson variation is

\[
t = \frac{\hat{\beta}_i}{\hat{\sigma}_{QL}(\hat{\beta}_i)} = \frac{.00183}{.000524} = 3.50.
\]

The standard error has been adjusted for over-dispersion (see Table 3). The p-value is \(P(t_{34} \geq |3.5|) = 2P(t_{34} \leq -3.5) = .00132\) (verified using the function call `2*pt(-3.498,34)`).

The drop-in-deviance test produces a slightly different p-value. The test statistic is

\[
F = \frac{\log(L_{unconstrained}) - \log(L_{constrained})}{k\hat{\psi}} = \frac{65.008 - 50.106}{1 \times 1.4213} = 10.48.
\]

The p-value is \(P(F_{1,34} \geq 10.484) = .00268\). The function call `anova(glm.1,glm.2,test="F")` will compute the F-statistic and p-value provided that both `glm.1` and `glm.2` were fit with `family=quasipoisson()`.

Table 3 shows the fitted coefficients from the quasi-Poisson model fit.

| Variable    | Coefficient | Std. Error | Wald statistic (Z) | \(P(Z > |z|)\) |
|-------------|-------------|------------|--------------------|-----------------|
| Intercept   | -3.21       | .223       | -14.4              | < .0001         |
| Years(8–11) | .233        | .301       | .774               | .444            |
| Years(12–15)| .551        | .282       | 1.95               | .0592           |
| Years(16–19)| 1.24        | .254       | 4.91               | < .0001         |
| Years(20–23)| 1.40        | .250       | 5.60               | < .0001         |
| Years(24–27)| 1.73        | .242       | 7.14               | < .0001         |
| Years(28–31)| 2.03        | .238       | 8.53               | < .0001         |
| Exposure    | .00183      | .000524    | 3.50               | .0013           |