Poisson Regression for Count Data

Poisson Regression and related Negative Binomial Regression, are generalized linear models typically involving count data as the response variable. As such, non-negative values are not expected (nor allowed), and responses are considered to be integers. The distribution of response values, as in a histogram for each set of independent variables indicated by index i, may look quite different for small versus large expected values (means). For the Poisson distribution, the distribution around small expected values are highly asymmetric, whereas for larger values symmetric. In addition, variance increases with expected value.

Following Zuur et al. 2009 (Za) in *Mixed Effects Models and Extensions in Ecology with R*, a Poisson GLM consists of 3 formal elements: 1 - a statement of the distribution of Y, 2 - a statement of the expected value or mean of Y (exp(Y)) and variance of Y (var(Y)), and 3 - a description of the link function. For count data related to density or scaled to other modifying variables, offsets may be applied that correct scale of the counts without introducing a regression coefficient. After an initial fit of a Poisson GLM, it is important to check for overdispersion. An approximate 1:1 ratio of Residual Deviance (RD) versus residual degrees of freedom (df) found in summary() indicate proper fit. If RD is greater than df, then overdispersion occurs and uncorrected P-values should be considered unreliable. Overdispersion is often the result model mis-specification, such caused by inappropriate or missing independent variables, but data may also be naturally overdispersed. Za recommend use of quasi-Poisson GLM fits as a "quick fix" for moderate RD/df ratios, called dispersion parameter ϕ, (also calculated from Pearson residuals in R). If ϕ is greater than 15-20 then "alternative methods" such as Negative Binomial Regression are recommended by Za. Examples below come from Za and from Crawley (Cr) 2007, *The R Book*.

Poisson Glm:
Assumptions:

Regression depends on specifying in Response & Independent variables in advance:

Y = vector of count response variable, each row of Y indicated by index i.
X = matrix independent variables (columns) with observations of X_i (rows) matched to Y_i (rows of Y).

β = vector of linear coefficients and Xβ is the linear predictor (systematic component) of the model.
Cases Y_i are independent.

Model:

$Y_i \sim$ Poisson(μ_i)   **< Y's are Poisson distributed with single parameter μ_i for each case.**
$E(Y_i) = \mu_i$, and var(Y_i) = μ_i   **< mean and variance defined.**

$\ln(\mu_i) = X\beta$
alternatively: $\mu_i = e^{(X\beta)}$ **< for natural log link to μ_i for Poisson Regression**

Estimation of Regression Coefficients:

Estimation is based on determining the maximum likelihood function given the data. Since a closed-form solution doesn't exit, this requires interative computation, here using glm() in the {nlme} package for Poisson regression and glm.nb() in the {MASS} package for Negative Binomial regression in R.

```
#GLM 020 LOGISTIC REGRESSION
library(nlme)
library(MASS)
setwd("c:/DATA/Models")
```

```
R=read.table("Roadkills.txt",header=T)
R
killed=R$TOT.N
distance=R$D.PARK
cbind(killed,distance)
FM1=glm(killed~distance,family=poisson)
FM1a=glm.nb(killed~distance)


#PLOTTING FIT:
plot(distance,fitted(FM1,type="response"),col='red',pch=19,
        ylab="Number Killed",type='l')
lines(distance,fitted(FM1a,type="response"),col='blue',pch=19)
points(distance,killed,col='black',pch=20)
summary(FM1)
```

Data from Za: Roadkills.txt
killed   = response
distance = independent variable (covariate)

> summary(FM1)
```
Call:
glm(formula = killed ~ distance, family = poisson)

Deviance Residuals:
    Min      1Q   Median      3Q      Max
-8.1100  -1.6950  -0.4708   1.4206   7.3337

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  4.316e+00  4.322e-02   99.87   <2e-16 ***
distance    -1.059e-04  4.387e-06  -24.13   <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: 1071.4  on 51  degrees of freedom
Residual deviance:  390.9  on 50  degrees of freedom
AIC: 634.29

Number of Fisher Scoring iterations: 4
```
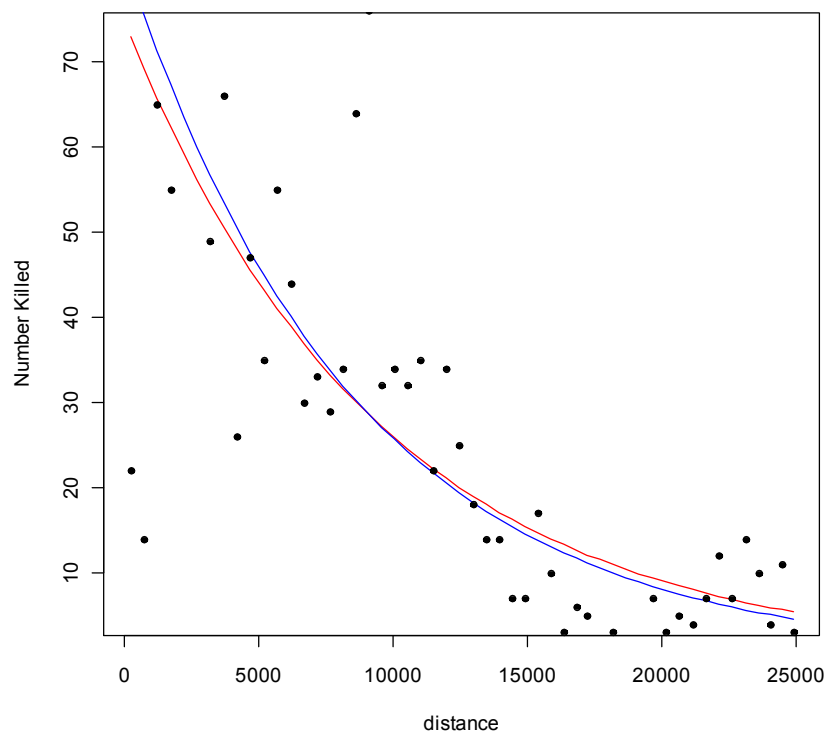
Comparison plot of fitted values of the same data for:

FM1 - Poisson Regression fit (red)
FM1a - Negative Binomial fit (blue)

Simple Poisson Regression:

Fitted Values:

Example for first point $Y_1 = 22$

$X_1 := 250.214$ $Y_1 := 22$ **< observed values of X (covariate) & Y (response)**
                                                   for the first case

$\beta_0 := 4.316$ $\beta_1 := -1.059 \cdot 10^{-4}$ **< regression coefficients from above**

$\mu_1 := e^{\beta_0 + \beta_1 \cdot X_1}$ $\mu_1 = 72.9302$ **< fitted value for the first case μ_1**

Pearson Residuals:

$r_P := \dfrac{Y_1 - \mu_1}{\sqrt{\mu_1}}$ $r_P = -5.9638$ **< pearson residual for the first case**

Deviance Residuals:

$d_1 := 2 \cdot Y_1 \cdot \ln\left(\dfrac{Y_1}{\mu_1}\right) - 2 \cdot (Y_1 - \mu_1)$ $d_1 = 49.128$ **< From Za formula at bottom of p. 217**

$r_{dev1} := \text{sign}(Y_1 - \mu_1) \cdot \sqrt{d_1}$ $r_{dev1} = -7.0091$ **< deviance residual for the first case**

^ Note: slight mis-match is interpreted as rounding.

Dispersion Parameter ϕ:

$D := 390.9$ $n := 52$ $p := 1$ **< Residual Deviance, number of observations, number of**
                                                (slope) parameters

$\phi := \dfrac{D}{n - p}$ $\phi = 7.665$ **< ϕ calculated according to Za Eq. 9.8**

> phi=sum(Pearson^2)/50   **< p = 2 Total number of parameters including intercept.**
> phi   **< Calculation using sum of Pearson residuals squarred Za. p. 323**
[1] 7.629755   **also as calculated by R.**

^ Note: to get values to match here,
I had to adjust (n-p) differently in
the two ways to calculate ϕ.

fit=fitted(FM1,type="response")
Pearson=resid(FM1,type="pearson")
Deviance=resid(FM1,type="deviance")
**cbind(killed,distance,fit,Pearson,Deviance)**

> cbind(killed,distance,fit,Pearson,Deviance)

	killed	distance	fit	Pearson	Deviance
1	22	250.214	72.966445	-5.96654261	-7.01275778
2	14	741.179	69.271303	-6.64084014	-8.10995385
3	65	1240.080	65.708070	-0.08735078	-0.08750837
4	55	1739.885	62.322161	-0.92750879	-0.94662240
5	88	2232.130	59.158048	3.74988312	3.49427393
6	104	2724.089	56.156278	6.38448400	5.70026832
7	49	3215.511	53.309852	-0.59028077	-0.59851490
8	66	3709.401	50.594486	2.16583118	2.06789145
9	26	4206.477	48.001238	-3.17556418	-3.48133623
...					
42	3	20141.987	8.885700	-1.97447798	-2.29268710
43	5	20640.729	8.428772	-1.18101781	-1.27883235
44	4	21138.903	7.995821	-1.41310526	-1.56545357
45	7	21631.542	7.589555	-0.21400107	-0.21686539
46	12	22119.102	7.207804	1.78497868	1.62770486
47	7	22613.647	6.840197	0.06110142	0.06086579
48	14	23113.450	6.487726	2.94934240	2.55176792
49	10	23606.088	6.158086	1.54819184	1.41864500
50	4	24046.886	5.877358	-0.77438323	-0.82232103
51	11	24444.874	5.634903	2.26013467	1.99648262
52	3	24884.803	5.378520	-1.02559385	-1.11992599

```
FM2=update(FM1,family=quasipoisson)
#PLOTTING FITS & 95%CI:
PredFM1=predict(FM1,type="link",se=T)
PredFM2=predict(FM2,type="link",se=T)
U.FM1=exp(PredFM1$fit+1.96*PredFM1$se.fit)
L.FM1=exp(PredFM1$fit-1.96*PredFM1$se.fit)
U.FM2=exp(PredFM2$fit+1.96*PredFM2$se.fit)
L.FM2=exp(PredFM2$fit-1.96*PredFM2$se.fit)
plot(distance,killed,col='black',pch=20)
lines(distance,fitted(FM2,type="response"),col='red',
    ylab="Number Killed")
lines(distance,U.FM1,lty=2,col='red')
lines(distance,L.FM1,lty=2,col='red')
lines(distance,U.FM2,lty=2,col='blue')
lines(distance,L.FM2,lty=2,col='blue')
```

Note: quasi-poisson glm provides the same fit, but widens the 95% Confidence Interval (blue dashed lines compared with red dashed lines) due to the fact that variance is now estimated as $\phi \cdot \mu_i$ Results of all tests are similarly modified making it harder (but more accurate) to reject the Null hypothesis.
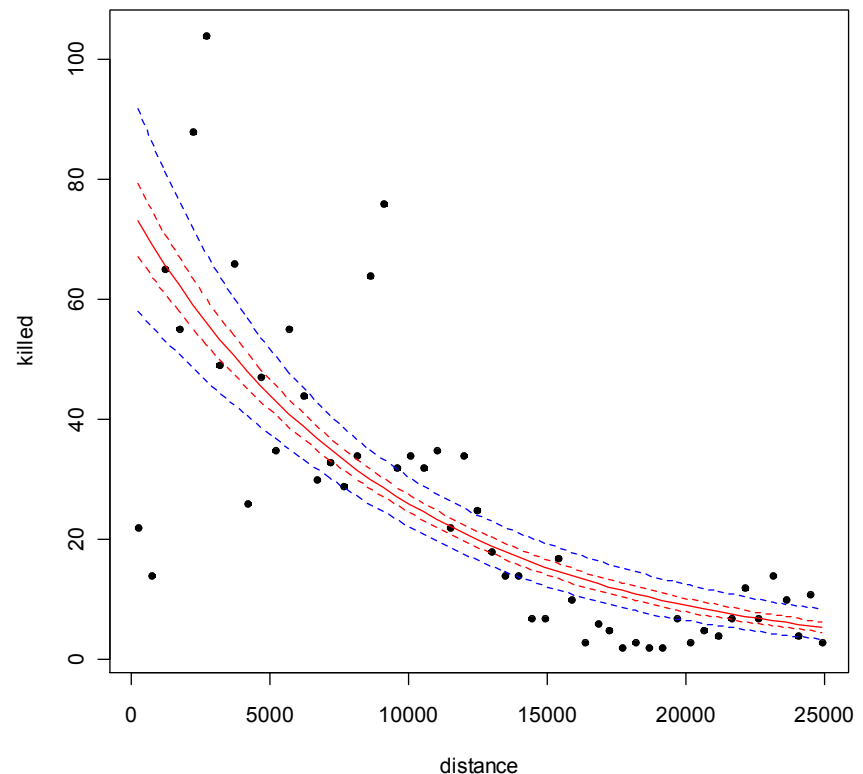
summary(FM2)

> summary(FM2)

```
Call:
glm(formula = killed ~ distance, family = quasipoisson)

Deviance Residuals:
    Min       1Q    Median       3Q       Max
-8.1100  -1.6950  -0.4708   1.4206    7.3337

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 4.316e+00  1.194e-01  36.156  < 2e-16 ***
distance    -1.058e-04  1.212e-05  -8.735 1.24e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasipoisson family taken to be 7.630148)

    Null deviance: 1071.4  on 51  degrees of freedom
Residual deviance:  390.9  on 50  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 4
```

Multiple Poisson Regression:

Cr Ch. 14 - species.txt data for ANCOVA study:
Species = response
Biomass = numeric covariate
pH = factor with three levels

```
S=read.table("species.txt",header=T)
S
FM3=glm(Species~Biomass*pH,family=poisson,data=S)
summary(FM3)
```

```
> S
     pH    Biomass  Species
1  high 0.46929722      30
2  high 1.73087043      39
3  high 2.08977848      44
4  high 3.92578714      35
5  high 4.36679265      25
6  high 5.48197468      29
7  high 6.68468591      23
8  high 7.51165063      18
9  high 8.13220251      19
...
81  low 0.28972266      17
82  low 0.07756009      14
83  low 1.42902041      15
84  low 1.12074092      17
85  low 1.50795384       9
86  low 2.32596318       8
87  low 2.99570582      12
88  low 3.53819909      14
89  low 4.36454121       7
90  low 4.87050789       3
```

```
> summary(FM3)
Call:
glm(formula = Species ~ Biomass * pH, family = poisson, data = S)

Deviance Residuals:
    Min      1Q    Median      3Q      Max
-2.4978  -0.7485  -0.0402   0.5575   3.2297

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    3.76812    0.06153  61.240  < 2e-16 ***
Biomass       -0.10713    0.01249  -8.577  < 2e-16 ***
pHlow         -0.81557    0.10284  -7.931 2.18e-15 ***
pHmid         -0.33146    0.09217  -3.596 0.000323 ***
Biomass:pHlow -0.15503    0.04003  -3.873 0.000108 ***
Biomass:pHmid -0.03189    0.02308  -1.382 0.166954
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 452.346  on 89  degrees of freedom
Residual deviance:  83.201  on 84  degrees of freedom
AIC: 514.39

Number of Fisher Scoring iterations: 4
```

^ Overdispersion is not detected so one proceeds with standard Poisson GLM.

Wald Test of single β:

Hypotheses:

H_0: $\beta = 0$
H_1: $\beta \neq 0$

Test Statistic:

z = Estimate/std.error

Example calculations:

$\beta_1 := -0.10713$ $s_{\beta_1} := 0.01249$ **< values from table above**

$z_1 := \dfrac{\beta_1}{s_{\beta_1}}$ $z_1 = -8.577$

$\beta_5 := -0.03189$ $s_{\beta_5} := 0.02308$ **< values from table above**

$z_5 := \dfrac{\beta_5}{s_{\beta_5}}$ $z_5 = -1.382$

Sampling Distribution:

If Assumptions hold and H_0 is true,

then z ~ N(0,1)

Probability:

$$P_1 := \min\left[\left(2 \cdot \text{pnorm}\left(z_1, 0, 1\right)\right), 2 \cdot \left(1 - \text{pnorm}\left(z_1, 0, 1\right)\right)\right] \qquad P_1 = 0$$

$$P_5 := \min\left[\left(2 \cdot \text{pnorm}\left(z_5, 0, 1\right)\right), 2 \cdot \left(1 - \text{pnorm}\left(z_5, 0, 1\right)\right)\right] \qquad P_5 = 0.1671$$

Decision Rule:

IF $P < \alpha$ THEN REJECT H_0, OTHERWISE ACCEPT H_0

The Wald test is a marginal test of single regression coefficients having a function similar to t-test in standard Linear Regression.

Note: Regression coefficients and standard errors are obtained from the maximum likelihood fit as first and second partial derivatives of the likelihood function. Summary information is obtained from the summary() wrapper of an object made by glm(). No attempt has been made to perform those calculations here.

Confidence Interval for β:

$\alpha := 0.05$

$$C := \left| \text{qnorm}\left(1 - \frac{\alpha}{2}, 0, 1\right) \right| \qquad\qquad C = 1.96$$

	estimate	confidence interval
$CI_1 := \left(\beta_1 - C \cdot s_{\beta_1} \quad \beta_1 + C \cdot s_{\beta_1}\right)$	$\beta_1 = -0.107$	$CI_1 = (-0.132 \quad -0.083)$
$CI_5 := \left(\beta_5 - C \cdot s_{\beta_5} \quad \beta_5 + C \cdot s_{\beta_5}\right)$	$\beta_5 = -0.03189$	$CI_5 = (-0.077 \quad 0.013)$

Likelihood Ratio Test:

The Likelihood Ratio Test in GLM models serves the same function as the General F Ratio test of Full Model (FM) versus Reduced Model (RM) in linear regression. Since model likelihoods are not reported by summary.glm or anova.glm, the Likelihood Ratio test of is conducted by consulting the Analysis of Deviance Table in anova(RM,FM), or anova(FM) (for serial testing), or drop1 (for marginal testing). Note the use of the χ^2 distribution for anova() here.

```
anova(FM3,RM1,test="Chisq")
drop1(FM3,test="Chisq")
```

Hypotheses:

H_0: a specified subset of β's = 0

H_1: same subset of β's <> 0

Test Statistic:

G^2 = -2ln[L(R)/L(F)] = -2ln(L(R) - 2ln(L(F)) = dev_F - dev_R

< where dev_F & dev_R are residual deviances calculated from the Maximum Likelihood fits of FM & RM

Sampling Distribution:

If Assumptions hold and H_0 is true,

then $G^2 \sim \chi^2(df_F - df_R)$ **< where df_F & df_R are degrees of freedom of FM & RM**

Probability:

P = 1 - pchisq(Δdef,Δdf) < where Δdev & Δdf are differences in values between FM & RM

Decision Rule:

IF P < α THEN REJECT H_0, OTHERWISE ACCEPT H_0

Example Calculations:

For Interaction term (blue in table):

$dev_F := 83.201$ $dev_R := 99.242$

$\Delta dev := dev_R - dev_F$ $\Delta dev = 16.041$

$df_F := 84$ $df_R := 86$

$\Delta df := df_R - df_F$ $\Delta df = 2$

$P := 1 - pchisq(\Delta dev, \Delta df)$ $P = 0.0003287$

Note: in this case, the interaction term is significant, so no further reduction in the model involving factors is attempted.

Slight differences in result are interpreted as rounding in the hand calculation.

P-values reported by Wald Test and Likelihood Ratio test for a single β will not necessarily be equal.

Direct Report of Likelihood Ratio Tests:

```
> anova(FM3,RM1,test="Chisq")
Analysis of Deviance Table

Model 1: Species ~ Biomass * pH
Model 2: Species ~ Biomass + pH
  Resid. Df Resid. Dev Df Deviance P(>|Chi|)
1        84      83.201
2        86      99.242 -2   -16.04 0.0003288 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Marginal report of Likelihood Ratio Tests:

```
> drop1(FM3,test="Chisq")
Single term deletions

Model:
Species ~ Biomass * pH
           Df Deviance    AIC    LRT   Pr(Chi)
<none>           83.201 514.39
Biomass:pH  2    99.242 526.43 16.04 0.0003288 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Plotting Fits:

```
#PLOTTING FITS:
plot(S$Biomass,S$Species,pch=20,col='black',
        ylab="Number of Species",
        xlab="Biomass")
points(S$Biomass[S$pH=="high"],S$Species[S$pH=="high"],pch=20,col="violet")
points(S$Biomass[S$pH=="mid"],S$Species[S$pH=="mid"],pch=20,col="blue")
points(S$Biomass[S$pH=="low"],S$Species[S$pH=="low"],pch=20,col="green")

FM1a=glm(Species[pH=="high"]~Biomass[pH=="high"],family=poisson,data=S)
FM1b=glm(Species[pH=="mid"]~Biomass[pH=="mid"],family=poisson,data=S)
FM1c=glm(Species[pH=="low"]~Biomass[pH=="low"],family=poisson,data=S)

points(S$Biomass[S$pH=="high"],fitted(FM1a),col="violet",pch=4)
points(S$Biomass[S$pH=="mid"],fitted(FM1b),col="blue",pch=4)
points(S$Biomass[S$pH=="low"],fitted(FM1c),col="green",pch=4)
```
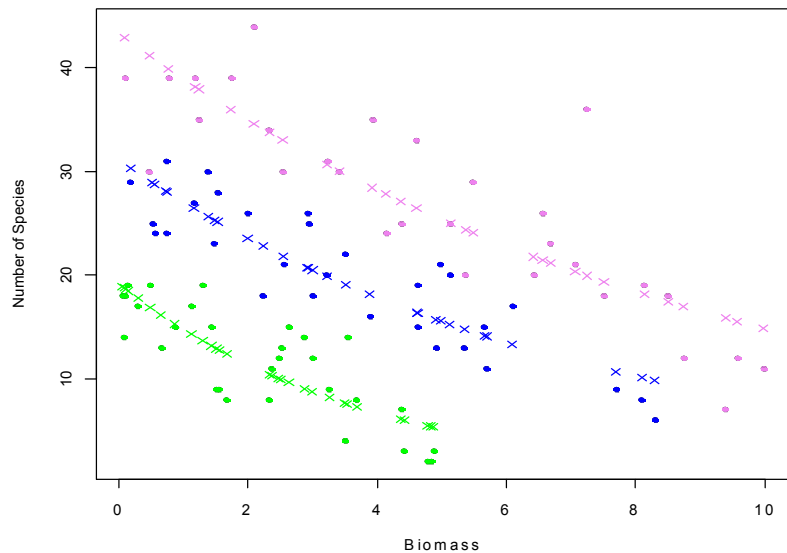
Points color coded with Fitted values + for each level of factor pH.

Multiple Poisson Regression and Model Simplification with Dispersion:

```
openl=R$OPEN.L
monts=R$MONT.S
polic=sqrt(R$POLIC)
dpark=distance
shrub=sqrt(R$SHRUB)
watres=sqrt(R$WAT.RES)
lwatc=R$L.WAT.C
lproad=sqrt(R$L.P.ROAD)
dwatcour=sqrt(R$D.WAT.COUR)
T=data.frame(killed,openl,monts,polic,shrub,watres,lwatc,lproad,dwatcour,dpark)
T
FM4=glm(killed~openl+monts+polic+shrub+watres+lwatc+lproad+dwatcour+dpark,
        family=poisson,data=T)
summary(FM4)
```

Same response as above, with extended list of covariates. Several variables treated with square-root transformation to control absolute size ranges, as shown by Za.

```
> summary(FM4)
Call:
glm(formula = killed ~ openl + monts + polic + shrub + watres +
    lwatc + lproad + dwatcour + dpark, family = poisson, data = T)

Deviance Residuals:
    Min      1Q   Median      3Q      Max
-6.8398  -1.3965  -0.1409   1.4641   4.3749

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.749e+00  1.567e-01  23.935  < 2e-16 ***
openl       -3.025e-03  1.580e-03  -1.915 0.055531 .
monts        8.697e-02  1.359e-02   6.398 1.57e-10 ***
polic       -1.787e-01  4.676e-02  -3.822 0.000133 ***
shrub       -6.112e-01  1.176e-01  -5.197 2.02e-07 ***
watres       2.243e-01  7.050e-02   3.181 0.001468 **
lwatc        3.355e-01  4.127e-02   8.128 4.36e-16 ***
lproad       4.517e-01  1.348e-01   3.351 0.000804 ***
dwatcour     7.355e-03  4.879e-03   1.508 0.131629
dpark       -1.301e-04  5.936e-06 -21.923  < 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: 1071.44  on 51  degrees of freedom
Residual deviance:  270.23  on 42  degrees of freedom
AIC: 529.62

Number of Fisher Scoring iterations: 5
```

Overdispersion clearly evident here > making quasi-Poisson refit necessary.

FM5=update(FM4,family=quasipoisson)
summary(FM5)

t-Test of single β:

Hypotheses:

H_0: β = 0

H_1: β ≠ 0

Test Statistic:

t = Estimate/std.error

Example calculations:

$$\beta_1 := -3.025 \cdot 10^{-3} \quad s_{\beta_1} := 3.847 \cdot 10^{-3}$$

$$t_1 := \frac{\beta_1}{s_{\beta_1}} \qquad t_1 = -0.786$$

$$\beta_8 := 7.355 \cdot 10^{-3} \quad s_{\beta_8} := 1.188 \cdot 10^{-2}$$

$$t_8 := \frac{\beta_8}{s_{\beta_8}} \qquad t_8 = 0.619$$

Sampling Distribution:

If Assumptions hold and H_0 is true,

then t ~ t(n-k)

Probability:

n := 52   **< total number**

k := 9   **< number of parameters**

df := n − k   df = 43

$$P_1 := \min\left[\left(2 \cdot pt(t_1, df)\right), 2 \cdot \left(1 - pt(t_1, df)\right)\right] \qquad P_1 = 0.43599$$

$$P_8 := \min\left[\left(2 \cdot pt(t_8, df)\right), 2 \cdot \left(1 - pt(t_8, df)\right)\right] \qquad P_8 = 0.53911$$

Decision Rule:

IF P < α THEN REJECT H_0, OTHERWISE ACCEPT H_0

The t-Test is conducted by summary() instead of the Wald test when using quasi-Poisson GLM instead of Poisson GLM. As in other situtations, the result is a marginal test of single regression coefficients having a function similar to t-test in standard Linear Regression.

Note: Regression coefficients and standard errors are obtained from the maximum likelihood fit as first and second partial derivatives of the likelihood function. Summary information is obtained from the summary() wrapper of an object made by glm(). No attempt has been made to perform those calculations here.

> summary(FM5)
```
Call:
glm(formula = killed ~ openl + monts + polic + shrub + watres
+
    lwatc + lproad + dwatcour + dpark, family = quasipoisson,
    data = T)

Deviance Residuals:
    Min      1Q   Median      3Q      Max
-6.8398  -1.3965  -0.1409   1.4641   4.3749

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.749e+00  3.814e-01   9.830 1.86e-12 ***
openl      -3.025e-03  3.847e-03  -0.786  0.43604
monts       8.697e-02  3.309e-02   2.628  0.01194 *
polic      -1.787e-01  1.139e-01  -1.570  0.12400
shrub      -6.112e-01  2.863e-01  -2.135  0.03867 *
watres      2.243e-01  1.717e-01   1.306  0.19851
lwatc       3.355e-01  1.005e-01   3.338  0.00177 **
lproad      4.517e-01  3.282e-01   1.376  0.17597
dwatcour    7.355e-03  1.188e-02   0.619  0.53910
dpark      -1.301e-04  1.445e-05  -9.004 2.33e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasipoisson family taken to be
5.928003)

    Null deviance: 1071.44  on 51  degrees of freedom
Residual deviance:  270.23  on 42  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 5
```

< formulation same as in multiple linear regression t-test.

Confidence Interval for β:

$$\alpha := 0.05$$

$$C := \left| qt\left(1 - \frac{\alpha}{2}, n - p\right) \right| \qquad\qquad C = 2.008$$

estimate   confidence interval

$$CI_1 := \left(\beta_1 - C \cdot s\beta_1 \quad \beta_1 + C \cdot s\beta_1\right) \qquad\qquad \beta_1 = -0.00303 \qquad CI_1 = (-0.011 \quad 0.005)$$

$$CI_5 := \left(\beta_5 - C \cdot s\beta_5 \quad \beta_5 + C \cdot s\beta_5\right) \qquad\qquad \beta_5 = -0.03189 \qquad CI_5 = (-0.078 \quad 0.014)$$

F-Ratio Test:

When employing quasi-Poisson (and quasi-Binomial) GLM, F-tests based on deviance are employed to compare Full Model (FM) versus Reducd Model (RM) in a way analogous to that seen in standard linear models. Deviance values can be found in Analysis of Deviance Tables reported by anova(RM,FM), or anova() for serial testing, or drop1 for marginal testing. Note that the test must be called explicitly in anova() and drop1().

RM2=glm(killed~dpark,family=poisson,data=T)
anova(FM5,RM2,test="F")
drop1(FM5,test="F")

< Reduced model RM2 is the "best fit" model reported by Za p. 228, after several rounds of model simplification, and used for comparison here.

Hypotheses:

H_0: a specified subset of β's = 0
H_1: same subset of β's <> 0

Test Statistic:

$F = (dev_R - dev_F) / \phi(p_R - p_F)$

< where dev_R & dev_F are residual deviances calculated from the Maximum Likelihood fits of FM & RM, $(p_R - p_F)$ is the difference in number of parameters of the two models, and ϕ is the dispersion parameter.

Sampling Distribution:

If Assumptions hold and H_0 is true, then $F \sim F(\Delta df, df_F)$

< where $\Delta df = df_F$ & df_R, df_F are degrees of freedom of FM.

Probability:

$P = 1 - pF(\Delta dev, \Delta df, df_F)$

< where Δdev is the difference in deviance between FM & RM

Decision Rule:

IF P < α THEN REJECT H_0, OTHERWISE ACCEPT H_0

Example Calculation:

$$\Delta dev := 120.66 \qquad\qquad \phi := 5.928003$$

$$\Delta df := 8 \qquad\qquad df_F := 42$$

$$F := \frac{\Delta dev}{\phi \cdot \Delta df} \qquad\qquad F = 2.544$$

$$P := 1 - pF(F, \Delta df, df_F) \qquad\qquad P = 0.0234154$$

```
> anova(FM5,RM2,test="F")
Analysis of Deviance Table

Model 1: killed ~ openl + monts + polic + shrub +
watres + lwatc + lproad +
    dwatcour + dpark
Model 2: killed ~ dpark
  Resid. Df Resid. Dev Df Deviance      F  Pr(>F)
1        42     270.23
2        50     390.90 -8  -120.66 2.5444 0.02341 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
0.1 ' ' 1
> drop1(FM5,test="F")
Single term deletions
```

Note: in this case, the drop1() results indicate significant reduction of the FM remains possible ultimately resulting in RM2 above, according to Za.

Marginal report of Likelihood Ratio Tests:

```
> drop1(FM5,test="F")
Single term deletions

Model:
killed ~ openl + monts + polic + shrub + watres + lwatc + lproad
+
    dwatcour + dpark
         Df Deviance F value    Pr(F)
<none>        270.23
openl     1   273.93  0.5739 0.452926
monts     1   306.89  5.6970 0.021574 *
polic     1   285.53  2.3776 0.130585
shrub     1   298.31  4.3635 0.042814 *
watres    1   280.02  1.5217 0.224221
lwatc     1   335.47 10.1389 0.002735 **
lproad    1   281.25  1.7129 0.197727
dwatcour  1   272.50  0.3526 0.555802
dpark     1   838.09 88.2569    7e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```