Slide 1

Reference: Agresti, Section 7.1

Count data and Poisson distribution

- Many response variables have counts as their possible outcome.
- The standard distribution to model count data is the Poisson distribution, which has probability mass function

$$p(Y = y) = \frac{e^{-\mu}\mu^y}{y!}, \quad y = 0, 1, 2, \ldots, \quad \mu > 0.$$

$$E[Y] = Var(Y) = \mu.$$

- Thus, any factor that affects the mean will affect the variance (and vice versa). So the usual assumption of homoscedasticity is not appropriate.

Slide 2

- The Poisson distribution with $\mu = n\pi$ can be derived as a limiting distribution of the binomial(n, π) as $n \to \infty$ and $\pi \to 0$.
- An alternative derivation is in terms of a stochastic process described somewhat informally as follows:
  - the probability of at least one event in a given time interval is proportional to the length of the interval;
  - the probability of two or more events in a very small time interval is negligible;
  - the number of events in disjoint time intervals are mutually independent.

Slide 3

A useful property of the Poisson distribution is that the sum of independent Poisson random variables is also Poisson.

- Thus, we can analyze individual or grouped data with equivalent results.
- If $Y_{ij} \sim$ Poisson(μ_i) for $j = 1, 2, \ldots, n_i$, then the group total $Y_i \sim$ Poisson$(n_i\mu_i)$.
- We obtain exactly the same likelihood function if we work with the individual counts Y_{ij} or the group counts Y_i.

Slide 4

Link function for count data

- Recall that $\eta_i = \log \mu_i$ is the canonical link for Poisson data

$$f(y_i) = \exp\left\{y_i \log \mu_i - \mu_i - \log(y_i!)\right\}.$$

- Thus, the log link is the canonical link for a Poisson GLM

$$\log \mu_i = \boldsymbol{x_i'\beta},$$

  which is often referred to as a **loglinear model**.
- β_j represents the expected change in the *log* of the mean per unit change in x_j.
- Exponentiating, we obtain a multiplicative model for the mean:

$$\mu_i = \exp(\boldsymbol{x_i'\beta}).$$

  So increasing x_j by one unit multiplies the mean by a factor $\exp(\beta_j)$.

Slide 5

Maximum likelihood estimation

- The log-likelihood function for n independent Poisson observations is

$$\log L(\boldsymbol{\beta}) = \sum_{i=1}^{n} (y_i \log \mu_i - \mu_i),$$

  where $\mu_i = \exp(\boldsymbol{x_i'\beta})$.

- The score functions are given by

$$\frac{\partial \log L}{\partial \boldsymbol{\beta}} = \boldsymbol{X'(y-\mu)},$$

- So the ML estimates satisfy the estimating equations

$$\boldsymbol{X'y} = \boldsymbol{X'\hat{\mu}},$$

- Recall that this estimating equation arises more generally in any GLM with canonical link.

Slide 6

Model fitting

- In general, we use the IRWLS algorithm or Fisher scoring (same as Newton-Raphson) procedure for estimation.

- Recall that one iteration of the algorithm is

$$\boldsymbol{\beta}^{(t)} = (\boldsymbol{X'W^{(t)}X})^{-1}\boldsymbol{X'W^{(t)}z^{(t)}},$$

  where the diagonal matrix $\boldsymbol{W}$ of iterative weights is

$$\boldsymbol{W} = \text{ Diag}\left[Var(Y_i)\left(\frac{\partial \eta_i}{\partial \mu_i}\right)^2\right]^{-1} = \text{Diag}(\mu_i)$$

  and the working dependent variate $\boldsymbol{z}$ is

$$\boldsymbol{z}^{(t)} = \boldsymbol{\eta}^{(t)} + \left(\frac{\partial \boldsymbol{\eta}^{(t)}}{\partial \boldsymbol{\mu}^{(t)}}\right)(\boldsymbol{y} - \boldsymbol{\mu^{(t)}}).$$

Slide 7

Quantities of interest

After convergence, we should save and examine the following:

- $Var(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X'WX})^{-1}$
- the Pearson residuals and the Pearson goodness-of-fit statistic

$$\mathcal{X}^2 = \sum_{i=1}^{N} \left(\frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i}} \right)^2$$

- the deviance residuals and the deviance statistic

$$G^2 = 2 \sum_{i=1}^{N} \left\{ y_i \log \frac{y_i}{\hat{\mu}_i} - (y_i - \hat{\mu}_i) \right\}$$

- the leverage values, i.e., the diagonal elements of

$$\boldsymbol{H} = \boldsymbol{W}^{1/2} \boldsymbol{X} (\boldsymbol{X'WX})^{-1} \boldsymbol{X'W}^{1/2}.$$

Slide 8

Tests of hypotheses

- Likelihood ratio tests for log-linear models can easily be constructed, as we did for the other GLMs.
- In large samples, the difference in deviances between two nested models is approximately $\chi^2_{df_1 - df_2}$.
- One can also construct Wald tests and confidence intervals as we have done before, based on the asymptotic distribution

$$\hat{\boldsymbol{\beta}} \sim N\left(\boldsymbol{\beta}, (\boldsymbol{X'WX})^{-1}\right).$$

Slide 9

Example

The data below reports fatalities in the Prussian army due to horse kicks from 1875 to 1894 across 14 army corps:

Year	75	76	77	78	79	80	81	82	83	84
Deaths	3	5	7	9	10	18	6	14	11	9
Year	85	86	87	88	89	90	91	92	93	94
Deaths	5	11	15	6	11	17	12	15	8	4

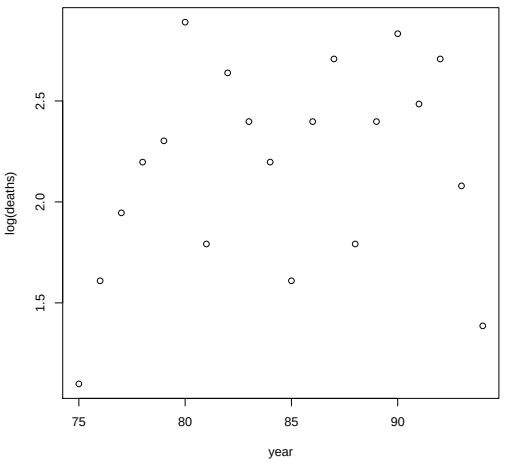
For illustration, we will look at the data by year to see if there appears to be any trend over time.

Slide 10

```
year = 75:94
deaths = c(3, 5, 7, 9, 10, 18, 6, 14, 11, 9, 5, 11, 15, 6, 11, 17, 12, 15, 8, 4)
plot(year, log(deaths))
fit.loess = loess(log(deaths) ~ year)
lines(year, predict(fit.loess))
```

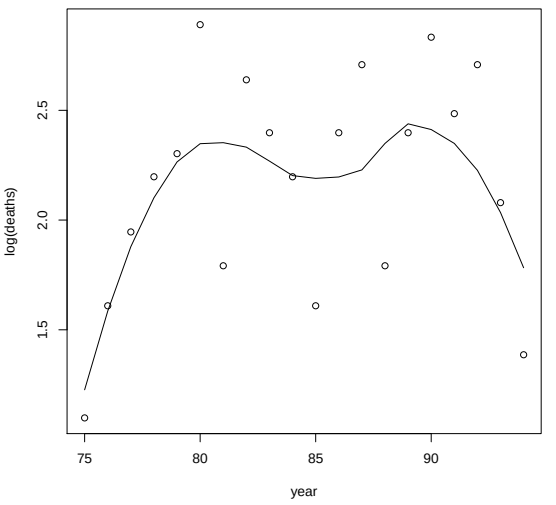
Slide 11

Scatter plot of log(death) versus year:



Slide 12

Apply `loess( )` to look for trends:



May need a 4th-degree polynomial to capture the trend over time.

Slide 13

```
fit.poisson = glm(deaths ~ year, family=poisson)

> summary(fit.poisson)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.69095    1.06056   0.651    0.515
year         0.01876    0.01243   1.509    0.131

    Null deviance: 38.503  on 19  degrees of freedom
Residual deviance: 36.216  on 18  degrees of freedom
AIC: 120.87

> 1-pchisq(36.216, 18)
[1] 0.006619696
```

The model does not fit the data well.

Slide 14

Let's include the effects of time as a fourth degree polynomial

```
yr2 = year^2; yr3 = year^3; yr4 = year^4;
fit.polynom = glm(deaths ~ year+yr2+yr3+yr4, family=poisson)

> summary(fit.polynom)
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.921e+04  5.880e+03  -3.266  0.00109 **
year         9.096e+02  2.789e+02   3.262  0.00111 **
yr2         -1.613e+01  4.952e+00  -3.258  0.00112 **
yr3          1.270e-01  3.902e-02   3.254  0.00114 **
yr4         -3.743e-04  1.151e-04  -3.251  0.00115 **
---
    Null deviance: 38.503  on 19  degrees of freedom
Residual deviance: 17.669  on 15  degrees of freedom
AIC: 108.32
```

There seems to be a problem of collinearity.

Slide 15

Let's center the variable year and refit the model:

```
yr = (year-mean(year)); yr2 = yr^2; yr3 = yr^3; yr4 = yr^4;
fit.center = glm(deaths ~ yr+yr2+yr3+yr4, family=poisson)

> summary(fit.center)
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.2142611  0.1365518  16.216  < 2e-16 ***
yr          -0.0006781  0.0315795  -0.021  0.98287
yr2          0.0222576  0.0091181   2.441  0.01465 *
yr3          0.0004726  0.0005855   0.807  0.41950
yr4         -0.0003743  0.0001151  -3.251  0.00115 **
---
    Null deviance: 38.503  on 19  degrees of freedom
Residual deviance: 17.669  on 15  degrees of freedom
AIC: 108.32
```

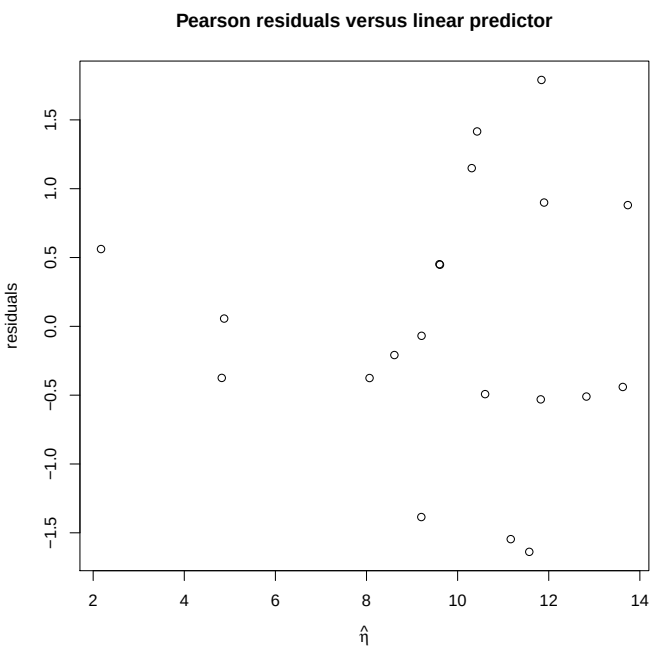
Slide 16

```
> 1-pchisq(17.669, 15)
[1] 0.2804668
```

Let's do a few diagnostic plots:

```
resid.fit = residuals(fit.center, type="pearson")
plot(year, fit.center$fitted.values, ylab=expression(hat(mu)),
        main="Estimated Mean by Year")
plot(fit.center$fitted.values, resid.fit, xlab=expression(hat(eta)), ylab="residuals",
      main = "Pearson residuals versus linear predictor")
plot(fit.center$fitted.values, abs(resid.fit), xlab=expression(hat(eta)), ylab="|residuals|",
      main = "Absolute value of residuals versus linear predictor")
```

Slide 17



Slide 18

- There's an increasing trend, suggesting that the variance function $Var(Y) = \mu$ may be wrong.
- It's possible that the overall size of the Prussian army varied over the twenty-year period, which could then cause the mean number of deaths to vary.
- Next, we talk about how to include a measure of size when it is available, i.e., consider rates rather than counts.

Slide 19

Poisson regression for rates

- When events occur over some index of size (time or space) that varies among observations, it is more relevant to model the **rate** of events.
- When a response count Y_i has index (such as population size) equal to t_i, the sample rate of outcomes is Y_i/t_i.
- A log-linear model for the expected rate is given by

$$\log(\mu_i/t_i) = \sum_{j=1}^{p} \beta_j x_{ij} \quad \Rightarrow \quad \log \mu_i = \log t_i + \sum_{j=1}^{p} \beta_j x_{ij}.$$

- The adjustment term, $\log t_i$, is called an **offset**.
- The fit corresponds to using $\log t_i$ as a predictor and forcing its coefficient to equal 1.0.

Slide 20

- The expected number of outcomes satisfies

$$\mu_i = t_i \exp\left(\sum_{j=1}^{p} \beta_j x_{ij}\right).$$

- A one-unit increase in the j-th element of $\boldsymbol{x}_i$ multiplies the incidence rate by $\exp(\beta_j)$.
- The mean is proportional to the index t_i, with proportionality constant depending on the value of the explanatory variable, x_i.
- For example, for a fixed value of x_i, doubling the population size t_i also doubles the expected number μ_i.

Slide 21

Example

The data below reports the survival of patients after heart-valve replacement surgery (Agresti, p. 129, Table 4.5). A sample of 109 patients were classified by type of heart valve (aortic or mitral) and by age ($< 55, \geq 55$). Follow-up observations occurred until the patient died or the study ended, and covered lengths of time varying from 3 to 97 months.

Age	Type	Time at risk (months)	Deaths
Under 55	aortic	1,259	4
	mitral	2,082	1
55+	aortic	1,417	7
	mitral	1,647	9

Slide 22

- We need to account for the fact that each group involves a different total time at risk (i.e., number of person-years).
- Thus, we model the *rate* of death

$$\log(\mu_i/t_i) = \boldsymbol{x}_i'\boldsymbol{\beta},$$

which is equivalent to

$$\eta_i = \log \mu_i = \log t_i + \boldsymbol{x}_i'\boldsymbol{\beta}.$$

Slide 23

Let's first fit the saturated model with log-exposure as an offset:

```
heart = data.frame(Age=as.factor(rep(c("<55","55+"),rep(2,2))),
              Type=as.factor(rep(c("Aortic","Mitral"),2)),
              Deaths=c(4,1,7,9), Exposure=c(1259,2082,1417,1647))

heart.fit = glm(Deaths ~ Age*Type , offset = log(Exposure),
                         family=poisson, data=heart)
> summary(heart.fit)
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)        -5.3104     0.3780 -14.050   <2e-16 ***
Age<55             -0.4414     0.6268  -0.704    0.481
TypeMitral          0.1009     0.5040   0.200    0.841
Age<55:TypeMitral  -1.9902     1.2264  -1.623    0.105
```

Make sure the offset is the logarithm of the exposure, not the exposure itself.

Slide 24

Note that none of the regression coefficients are statistically significant, so they should not be interpreted. If we went ahead and did it nonetheless, we would state:

- When type = aortic, the death rate for the younger group is estimated to be 0.644 times that in the older group ($\exp(-0.44) = 0.644$).
- For the older group, the death rate for mitral valve replacement is 1.11 times that for aortic ($\exp(1.11) = 1.106$).
- For the younger group, the effect of mitral versus aortic replacement is $\beta_2 + \beta_3$, which is estimated to be $0.10009 - 1.9902 = -1.889$. Since $\exp(-1.889) = 0.15 = 1/6.67$, we estimate that among younger patients, the death rate for mitral valve replacement is 6 to 7 times lower than for aortic.

Slide 25

Repeating the model-fitting for various sets of predictors, we obtain the following analysis-of-deviance table:

Model	G^2	*df*	*p*-value
Saturated**	0.00	0	–
Age+Type	3.223	1	0.073
Type	9.886	2	0.007
Age	3.790	2	0.150
Null	10.841	3	0.013

The model with Age is the best fit we get for these data.

Slide 26

```
> summary(glm(Deaths ~ Age , offset = log(Exposure),
+                           family=poisson, data=heart))

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -5.2549     0.2500 -21.020   <2e-16 ***
Age<55       -1.2497     0.5123  -2.439   0.0147 *
---
    Null deviance: 10.8405  on 3  degrees of freedom
Residual deviance:  3.7897  on 2  degrees of freedom
AIC: 20.917
```

Slide 27

Overdispersion

- Under the Poisson model

$$E[Y_i] = Var(Y_i) = \mu_i.$$

- Real data, however, often exhibit more variation than allowed by the Poisson model.
- One approach to deal with overdispersion is to change the response distribution to **negative binomial**, which is more dispersed than the Poisson.