Review of Tuesday

The Poisson distn

- applied to counts - X

$\mu = E(X)$ $\quad P(X=x) = \dfrac{\mu^x e^{-\mu}}{x!}$

$Var(X) = \mu$ $\quad \hat{\mu} = \bar{X}$ - sample n

$s.e.(\hat{\mu}) = \sqrt{\dfrac{\bar{X}}{n}}$ note sometimes $n = 1$

usual asymptotic procedures

C.I. $\hat{\mu} \pm 1.96\, se(\hat{\mu})$

Large sample Test : $Z = \dfrac{\hat{\mu} - \mu_0}{se(\hat{\mu})}$ → use μ_0 not $\bar{X}$

→ need total count $= 5$

exact. fla. CI

exact test ....

Inference about rates -

- usually case that rates are more interpretable scientifically

- denominator
  - span of time
  - count of persons
  - person. time

Goodness-of-fit

→ set of counts $X_1 \dots X_k$

$p_1 \dots p_k = 1$ sum

inference is compare $\hat{p}_i = \frac{X_i}{n}$ $n = \sum X_i$

or X_i to np_i =

obs'd exp'd

frequencies

discrepancy measure $X^2 = \sum \frac{(X_i - np_i)^2}{(np_i)}$ obs - exp / exp

- Pearson's G.o.f.

If p's are "right", X^2

follows χ^2 dstn

- $k - \nu - 1$ df

* parameters estimated to form p_i's

to apply to

Poisson or variable

Poisson with infinite range ⇒ collapse the tail to single category

$G^2 = 2\Sigma$ obs x $\log\left(\frac{obs}{exp}\right)$

- likelihood ratio test

Comparing Rates

In example - $j = 1..4$ age categories,

rates $\lambda_j : 1..4$

? $\lambda_2 > \lambda_3$ - consider λ_2 / λ_3

- relative risk . R.R.

$\lambda_2 = \mu_2 / T_2$ - person years

5×5.15

$\lambda_3 : \mu_3 / T_3$ 5×2.67

$$R.R. = \mu_2/\mu_3 \times \left(T_3 / T_2 \right)$$

$X_2 / X_3 - 47/30$

convenient to switch to logs

$\theta = \log(\mu_2/\mu_3) = \log \mu_2 - \log \mu_3$

$\hat{\Theta} = \log X_2 - \log X_3$ $\quad Var(\hat{\Theta})$

$= Var(\log(X_2)) + Var(\log(X_3))$

non-linear

$Var(f(y))$? ← approximation in terms of $Var(y)$

$Var(a + by) = b^2 Var(y)$

$f(y) \approx f(y_0) + f'(y_0) \times (y - y_0)$

Let $y_0 = E(y)$

$$Var(f(y)) \approx \left[\frac{df}{dy}\Big|_{y=E(y)} \right]^2 Var(y)$$

– delta method

$$\frac{d \log x}{dx} = \frac{1}{x} \quad \therefore Var(\log(X_2)) \approx \frac{\mu_2}{(\mu_2)^2} = \frac{1}{\mu_2}$$

est'd by $\frac{1}{X_2}$

$$Var(\hat{\Theta}) \approx \frac{1}{X_2} + \frac{1}{X_3}$$

est'd by

95% CI : $(\log 47 - \log 30) \pm 1.96 \sqrt{\frac{1}{47} + \frac{1}{30}}$

$+.45$ $.46$

95% CI for θ = $-.01$ to $.91$

95% CI for $\frac{\mu_2}{\mu_3}$ take $e^{\cdot}$ = $.991$ to 2.47

adjust for denominators

- multiply by T_3/T_2 : $\frac{5 \times 2.28}{5 \times 1.82}$

Conf. Int. for RR ~ $1.\underline{24}$ to 3.1

indicates the risk is at least 25% higher in <4 & possibly 3x higher

point estimate $e^{\hat{\theta}} \times T_3/T_2$

$= 1.96$

- neglect → hypothesis testing Test RR = 1

exact inference for RR's

I need to give references

Random Events in a Human Population

The following data gives incidence of leukemia in males 0—14 during in a five year period in Birmingham, U.K. (source: Breslow and Day, Vol. I).

T

$X = 2$ → denominator - Person years

Ages	Cases	Pop. (100k)	Years	Rate (per 100k person-years)
<1	2	0.45	5	0.88
1 thru 4	47	1.82	5	5.15
5 thru 9	30	2.28	5	2.63
10 thru 14	13	2.03	5	1.28

5x.4[illegible]

2.2[illegible]

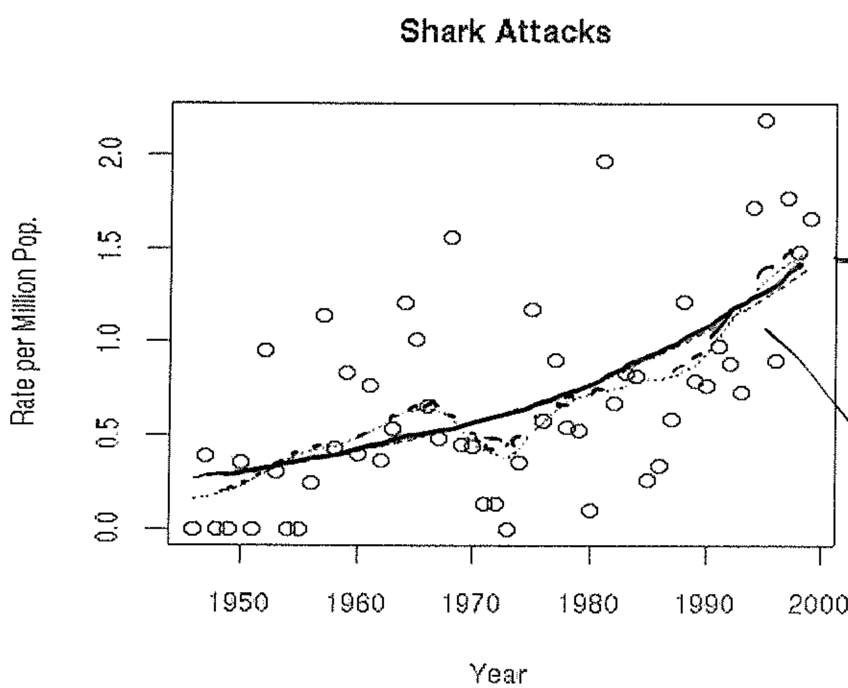
count

$.88 = \frac{X}{T}$

Poisson Regression

The following plot portrays the rate of shark attacks (per million population) recorded in Florida from 1945—2000. A non-parametric smoother has been added, as well as the fitted line from a Poisson Regression model.

$\lambda = \frac{\mu}{T}$



→ scatterplot smoother - supsmu

→ poisson regression line

→ describes a pattern of
Consistent pattern of proportional
or (%) increase year to year
(exponential)