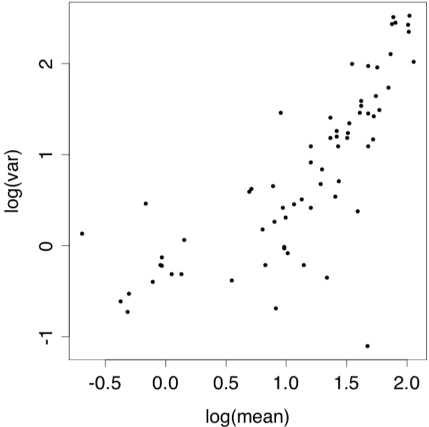
Outline

Example: Poisson regression

- Number of Children Ever Born (CEB) to Women of Indian Race By Marital Duration, Type of Place of Residence and Educational Level
- Each cell shows the mean, variance, and sample size.

Marr.	Suva				Urban				Rural			
Dur.	N	LP	UP	S+	N	LP	UP	S+	N	LP	UP	S+
0–4	0.50	1.14	0.90	0.73	1.17	0.85	1.05	0.69	0.97	0.96	0.97	0.74
	1.14	0.73	0.67	0.48	1.06	1.59	0.73	0.54	0.88	0.81	0.80	0.59
	8	21	42	51	12	27	39	51	62	102	107	47
5–9	3.10	2.67	2.04	1.73	4.54	2.65	2.68	2.29	2.44	2.71	2.47	2.24
	1.66	0.99	1.87	0.68	3.44	1.51	0.97	0.81	1.93	1.36	1.30	1.19
	10	30	24	22	13	37	44	21	70	117	81	21
10–14	4.08	3.67	2.90	2.00	4.17	3.33	3.62	3.33	4.14	4.14	3.94	3.33
	1.72	2.31	1.57	1.82	2.97	2.99	1.96	1.52	3.52	3.31	3.28	2.50
	12	27	20	12	18	43	29	15	88	132	50	9
15–19	4.21	4.94	3.15	2.75	4.70	5.36	4.60	3.80	5.06	5.59	4.50	2.00
	2.03	1.46	0.81	0.92	7.40	2.97	3.83	0.70	4.91	3.23	3.29	–
	14	31	13	4	23	42	20	5	114	86	30	1
20–24	5.62	5.06	3.92	2.60	5.36	5.88	5.00	5.33	6.46	6.34	5.74	2.50
	4.15	4.64	4.08	4.30	7.19	4.44	4.33	0.33	8.20	5.72	5.20	0.50
	21	18	12	5	22	25	13	3	117	68	23	2
25–29	6.60	6.74	5.38	2.00	6.52	7.51	7.54	–	7.48	7.81	5.80	–
	12.40	11.66	4.27	–	11.45	10.53	12.60	–	11.34	7.57	7.07	–
	47	27	8	1	46	45	13	–	195	59	10	–

- Sample unit: the individual women
- Response: the number of children she has borne.
- Predictors: the duration since her first marriage, the type of place where she resides, and education level, classified in 4 categories
- The mean-variance relationship for the data:



Offset in Poisson model

- Y_{ijkl}; the number of children borne by the l-th women in the (i, j, k)-th group, where i denotes marital duration, j residence, and k education
- $Y_{ijk} = \sum_l Y_{ijkl}$ denotes the group total.
- If Y_{ijkl} i.i.d. $\sim$ Poi(μ_{ijk}), then $Y_{ijk} \sim$ Poi$(n_{ijk}\mu_{ijk})$ where n_{ijk} is the number of observations in the (i, j, k)-th cell.

More precisely, if the individual mean $\mathbb{E}(Y_{ijkl}) = \mu_{ijk}$ follows a log-linear poisson model,

$$\log(\mathbb{E}(Y_{ijkl})) = X'_{ijk}\boldsymbol{\beta},$$

then the group totals follows a log-linear model with exactly the same coefficients $\boldsymbol{\beta}$:

$$\log(\mathbb{E}(Y_{ijk})) = \log(n_{ijk}\mu_{ijk}) = \underbrace{\log(n_{ijk})}_{\text{off set}} + X'_{ijk}\boldsymbol{\beta}$$

We can analyze the data by fitting poisson model to either individual counts, or to the group totals.

Poisson regression with offset

We consider additive log-linear model on the group total counts Y_{ijk}:

$$Y_{ijk} \sim \text{Poi}(n_{ijk}\mu_{ijk}) \text{ independently,}$$

where

$$\log(\mathbb{E}(Y_{ijk})) = \underbrace{\log(n_{ijk})}_{\text{offset}} + \underbrace{\beta_0}_{\text{intercept}} + \underbrace{\sum_{i=1}^{5} \beta_i \text{Duration}_i}_{\text{Duration}} + \underbrace{\sum_{j=1}^{2} \alpha_j \text{Residence}_j}_{\text{Residence}} + \underbrace{\sum_{k=1}^{3} \gamma_k \text{Education}_k}_{\text{Education}}$$

Additive model

TABLE 4.4: Estimates for Additive Log-Linear Model of Children Ever Born by Marital Duration, Type of Place of Residence and Educational Level

Parameter		Estimate	Std. Error	z-ratio
Constant		−0.1173	0.0549	−2.14
Duration	0–4	–		
	5–9	0.9977	0.0528	18.91
	10–14	1.3705	0.0511	26.83
	15–19	1.6142	0.0512	31.52
	20–24	1.7855	0.0512	34.86
	25–29	1.9768	0.0500	39.50
Residence	Suva	–		
	Urban	0.1123	0.0325	3.46
	Rural	0.1512	0.0283	5.34
Education	None	–		
	Lower	0.0231	0.0227	1.02
	Upper	-0.1017	0.0310	-3.28
	Sec+	-0.3096	0.0552	-5.61

Outline

Exponential family of distributions

- Exponential family of distributions has a density of the form

$$f(y) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right) \tag{1}$$

- $\mathbb{E}(Y) = b'(\theta)$ and $Var(Y) = a(\phi)b''(\theta)$.
- θ is commonly called the canonical parameter.
- ϕ is a fixed (known) scale parameter, also called dispersion parameter.
- A canonical link function is a function $g()$ such that

$$g(\mu) = \theta, \quad \text{where } \mu = \mathbb{E}(Y)$$

- Variance function: $V(\mu) = b''(\theta) = b''(g(\mu))$.

Example: normal distribution

Normal distribution: $Y \sim N(\mu, \sigma^2)$. The probability distribution of Y is

$$
\begin{aligned}
f(y) =& (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}(y-\mu)^2\right\} \\
=& \exp\left\{\frac{y\mu - \mu^2/2}{\sigma^2} - \frac{1}{2}\left(y^2/\sigma^2 + \log(2\pi\sigma^2)\right)\right\}
\end{aligned}
$$

1. $\theta = \mu = \mathbb{E}(Y)$; $\phi = \sigma^2$;
2. $b(\theta) = \mu^2/2 = \theta^2/2$; $b'(\theta) = \theta$; $b''(\theta) = 1$;
3. $g(\mu) = \mu$; the canonical link is identity;
4. Note that $Var(Y) = a(\phi)b''(\theta) = a(\phi)$. We have
$$V(\mu) = 1.$$
5. $a(\phi) = \phi$.

Example: Binomial distribution

Binomial distribution: $Y \sim Bin(n, \pi)$. The probability distribution is

$$
\begin{aligned}
f(y) &= \binom{n}{y} \pi^y (1-\pi)^{n-y} \\
&= \exp\left(y \log(\pi/(1-\pi)) + n \log(1-\pi) + \log(\binom{n}{y}) \right)
\end{aligned}
$$

1. $\theta = \log(\pi/(1-\pi))$; $\phi = 1$;
2. $b(\theta) = -n\log(1-\pi) = n\log(1+\exp(\theta))$;
3. $\mu = \mathbb{E}(Y) = b'(\theta) = n\frac{\exp(\theta)}{1+\exp(\theta)} = n\pi$
4. Thus $g(\mu) = \log(\mu/(n-\mu))$; the canonical link is logit
5. Because $Var(Y) = a(\phi)b''(\theta) = a(\phi)n\pi(1-\pi)$, we have

$$V(\mu) = n\pi(1-\pi) = n^{-1}\mu(1-\mu)$$

Example: Poisson distribution

Poisson distribution: $Y \sim Poi(\lambda)$. The probability distribution of Y is

$$f(y) = \frac{\lambda^y \exp(-\lambda)}{y!} = \exp\left(y \log(\lambda) - \lambda - \log(y!)\right).$$

1. $\theta = \log(\lambda)$; $\phi = 1$;
2. $b(\theta) = \lambda = \exp(\theta)$;
3. $\mu = \mathbb{E}(Y) = b'(\theta) = \exp(\theta) = \lambda$
4. Note that $\theta = \log(\mu)$. Then $g(\mu) = \log(\mu)$; the canonical link is log
5. Note that $Var(Y) = a(\phi)b''(\theta) = a(\phi)\exp(\theta)$. We have

$$V(\mu) = \exp(\theta) = \mu = \lambda.$$

Generalized Linear Models

- Generalized linear models (GLM) is a class of models including linear and nonlinear regression.
- A GLM is formulated as follows.
  1. $Y_1, \ldots, Y_n$ are n independent responses that follow a probability distribution in the exponential family of distributions with expected value $\mathbb{E}(Y_i) = \mu_i$
  2. A linear predictor based on the predictor variables $X_{i1}, \ldots, X_{i,p-1}$ is utilized, denoted by $\boldsymbol{X}_i'\boldsymbol{\beta}$:

     $$\boldsymbol{X}_i'\boldsymbol{\beta} = \beta_0 + \beta_1 X_{i1} + \ldots + \beta_{p-1} X_{i,p-1}$$

  3. A link function g relates the linear predictor to the mean response

     $$g(\mu_i) = \boldsymbol{X}_i'\beta$$

Generalized Linear Models

- In the linear regression, $g(\mu_i) = \mu_i$ (identity link).
- In the logistic regression, $\mu_i = \pi_i$, $g(\pi_i) = log(\pi_i/(1 - \pi_i))$ (logit link).
- In the Poisson regression, $g(\mu_i) = log(\mu_i)$ (log link).
- MLEs of model parameters are often obtained by iteratively reweighted least squares.
- Hypothesis testing of model parameters are often based on likelihood ratio test.

Measuring the goodness of fit

- The discrepancy of a fit is proportional to twice the difference betwen the maximum log likelihood achievable and that achieved by the model under investigation.

$$-2\{\log \mathcal{L}(R) - \log \mathcal{L}(F)\} = 2\sum_{i=1}^{n} \frac{\omega_i}{\phi}\left(Y_i(\tilde{\theta}_i - \hat{\theta}_i) - b(\tilde{\theta}_i) + b(\hat{\theta}_i)\right) \quad (2)$$

$$= \frac{D(\boldsymbol{Y};\hat{\mu})}{\phi} \quad (3)$$

  where $\tilde{\theta}_i$ and $\hat{\theta}_i$ are the estimates of the canonical parameters of the full and reduced models.
- Write $\log \mathcal{L}(R)$ as $\ell(\hat{\mu}, \phi; \boldsymbol{Y})$; then $\log \mathcal{L}(F) = \ell(\boldsymbol{Y}, \phi; \boldsymbol{Y})$
- $D(\boldsymbol{Y};\hat{\mu})$ is known as the deviance for the current model and is a function of the data only.
- Normal: $D(\boldsymbol{Y};\hat{\mu}) = \sum_{i=1}^{n}(Y_i - \hat{\mu}_i)^2$
- Binomial: $D(\boldsymbol{Y};\hat{\mu}) = 2\sum_{i=1}^{n}\{Y_i \log(Y_i/\hat{\mu}_i) + (n_i - Y_i)\log(\frac{n_i - Y_i}{n_i - \hat{\mu}_i})\}$
- Poisson: $D(\boldsymbol{Y};\hat{\mu}) = 2\sum_{i=1}^{n}\{Y_i \log(Y_i/\hat{\mu}_i) - (Y_i - \hat{\mu}_i)\}$

Outline

Overdispersion in logistic regression

- Overdispersion: the variace of the response exceeds the nominal variance.
- For the binomial response $Y_i \sim Bin(n_i, \pi_i)$,
  - Overdispersion means that the data show evidence that the variance of Y_i is greater than $n_i\pi_i(1-\pi_i)$.
- Overdispersion occurs when the data display more variability than is predicted by the variance-mean relationship for the assumed model.

- Underdispersion is also theoretically possible.

Overdispersion in logistic regression

- To correct for overdispersion in a logit model with binomial response, we assume that

$$\mathbb{E}(Y_i) = n_i\pi_i, \quad Var(Y_i) = \sigma^2 n_i\pi_i(1-\pi_i)$$

where logit(π_i)= $\boldsymbol{X}_i'\beta$ and σ^2 is a scale parameter.

1. If $\sigma^2 \neq 1$, the model is not binomial
2. If $\sigma^2 > 1$, overdispersion

- The Fisher-scoring procedure for estimating β does not change, but its variance-covariance matrix changes

$$Var(\hat{\beta}) \approx \sigma^2(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}$$

- The most popular method for adjusting for overdispersion comes from the theory of quasi-likelihood.

```
glm(formula, family=quasibinomial("logit"), ...)
```

Overdispersion in logistic regression

- Reasons for dispersion:
  1. variation among the probabilities of success
  2. correlation between the responses
- Suppose that $Y_i|\pi_i \sim Bin(n_i, \pi_i)$, for $i = 1, \ldots, c$.
- Assume a prior distribution on π_i such that $\mathbb{E}(\pi_i) = p_i$ and $Var(\pi_i) = \sigma^2 p_i(1 - p_i)$.
- Then $\mathbb{E}(Y_i) = n_i p_i$ and $Var(Y_i) = n_i p_i(1 - p_i)[1 + (n_i - 1)\sigma^2]$.
- Example: $\pi_i \sim Beta(a_i, b_i)$, with pdf

$$f(\pi_i) \propto \pi_i^{a_i - 1}(1 - \pi_i)^{b_i - 1}$$

  This yields Beta-Binomial model.

Overdispersion in Poisson regression

- One of the key features of the Poisson distribution is that the variance equals the mean.
- If $Y_i \sim Poi(\mu_i)$, then $\mathbb{E}(Y_i) = Var(Y_i) = \mu_i$.
- Overdispersion means that the data show evidence that the variance of Y_i is greater than μ_i.
- Quasi-likelihood: $Var(Y_i) = \sigma^2 \mathbb{E}(Y_i)$

```
glm(formula, family=quasipoisson, ...)
```

- Assume that $Y_i|Z_i \sim Poi(Z_i)$ and a prior distribution on Z_i such that $\mathbb{E}(Z_i) = \mu_i$. Then $\mathbb{E}(Y_i) = \mu_i$ and $Var(Y_i) = Var(Z_i) + \mu_i$.
- Negative-Binomial regression model:
  - Z_i is Gamma with mean μ_i and index $\psi\mu_i$; $Var(Z_i) = \mu_i^2/(\psi\mu_i)$.
  - $Var(Y_i) = \mu_i(1 + 1/\psi)$