## Weighted Least Squares Analysis of Repeated Categorical Data

- The first general approach to the analysis of repeated categorical outcomes
- Makes no assumptions concerning the time dependence
- Inherently nonparametric; based only on the multinomial sampling model
- Useful when:
  - Response has only a few possible outcomes
  - All covariates are categorical
  - Number of measurement times is small
  - Sample sizes are large
- Can accommodate randomly missing data

## The Multinomial Distribution

- Consider a sequence of $n$ independent trials
- On each trial, one of $c$ mutually exclusive and exhaustive events $E_1, \ldots, E_c$ occurs
- $\pi_i = \Pr(E_i)$ is constant across trials
  $0 < \pi_i < 1, \quad \sum_{i=1}^{c} \pi_i = 1$
- The probability that $E_1$ occurs $x_1$ times, …, $E_c$ occurs $x_c$ times is given by

  $$f(x_1, \ldots, x_c) = \frac{n!}{x_1!\, x_2! \cdots x_c!}\, \pi_1^{x_1}\, \pi_2^{x_2} \cdots \pi_c^{x_c},$$

  where $x_i \geq 0$ and $\sum_{i=1}^{c} x_i = n$
- The random vector $x = (x_1, \ldots, x_c)'$ has the $M_c(n, \pi)$ distribution with parameters $n$ and $\pi = (\pi_1, \ldots, \pi_c)'$

## Moments of the Multinomial Distribution

- $\mathrm{E}(x_i) = n\pi_i$, for $i = 1, \ldots, c$
- $\mathrm{Var}(x_i) = n\pi_i(1 - \pi_i)$, for $i = 1, \ldots, c$
- $\mathrm{Cov}(x_i, x_j) = -n\pi_i\pi_j$, for $i \neq j = 1, \ldots, c$
- The variance-covariance matrix of the vector $x = (x_1, \ldots, x_c)'$ is given by

$$\begin{pmatrix} n\pi_1(1-\pi_1) & -n\pi_1\pi_2 & \ldots & -n\pi_1\pi_c \\ -n\pi_1\pi_2 & n\pi_2(1-\pi_2) & \ldots & -n\pi_2\pi_c \\ \vdots & \vdots & \vdots & \vdots \\ -n\pi_1\pi_c & -n\pi_2\pi_c & \ldots & n\pi_c(1-\pi_c) \end{pmatrix}$$

- This variance-covariance matrix can be written as $n(D_\pi - \pi\pi')$, where $D_\pi$ is a diagonal matrix with the vector $\pi = (\pi_1, \ldots, \pi_c)'$ on the main diagonal

## Parameter Estimates

- The maximum likelihood estimators of $\pi_1, \ldots, \pi_c$ are given by $p_i = x_i/n$
- $\mathrm{E}(p_i) = \pi_i$, $\mathrm{Var}(p_i) = \pi_i(1 - \pi_i)/n$
- $\mathrm{Cov}(p_i, p_j) = -\pi_i \pi_j / n$
- $p = (p_1, \ldots, p_c)'$ is an unbiased estimator of $\pi = (\pi_1, \ldots, \pi_c)'$
- The variance-covariance matrix of $p$ is:

$$\frac{1}{n}\begin{pmatrix} \pi_1(1-\pi_1) & -\pi_1\pi_2 & \ldots & -\pi_1\pi_c \\ -\pi_1\pi_2 & \pi_2(1-\pi_2) & \ldots & -\pi_2\pi_c \\ \vdots & \vdots & \vdots & \vdots \\ -\pi_1\pi_c & -\pi_2\pi_c & \ldots & \pi_c(1-\pi_c) \end{pmatrix}$$

$$= \frac{1}{n}(D_\pi - \pi\pi')$$

## Large-Sample Distribution Theory

- As $n \rightarrow \infty$, the asymptotic distribution of $\sqrt{n}(p-\pi)$ is $N_c(0, D_\pi - \pi\pi')$

- A consistent estimator of Var($p$) is

$$V_p = \frac{1}{n}\begin{pmatrix} p_1(1-p_1) & -p_1p_2 & \dots & -p_1p_c \\ -p_1p_2 & p_2(1-p_2) & \dots & -p_2p_c \\ \vdots & \vdots & \vdots & \vdots \\ -p_1p_c & -p_2p_c & \dots & p_c(1-p_c) \end{pmatrix}$$

$$= \frac{1}{n}(D_p - pp')$$

- $p = (p_1, \dots, p_c)'$ has an approximate multivariate normal distribution with mean vector $\pi$ and variance-covariance matrix $V_p$

  $p \approx N_c(\pi, V_p)$

# Linear Models Using Weighted Least Squares

- A generalization of ordinary least squares that permits observations to be correlated and have nonconstant variance

- Let $y = (y_1, \ldots, y_n)'$ be an $n \times 1$ vector of observations

- Suppose that $y \sim N_n(X\beta, V)$

  $X$ is an $n \times p$ design (model) matrix ($p \leq n$)

  $\beta$ is a $p \times 1$ vector of parameters

  $V$ is the $n \times n$ variance-covariance matrix of $y$

- The linear model is $y = X\beta + \epsilon$, where $\epsilon \sim N_n(0, V)$

## Weighted Least Squares

Basic idea:

Transform the observations $y = (y_1, \ldots, y_n)'$ to other variables $y^*$ which satisfy the assumptions of the linear model

$y^* = X^*\beta + \epsilon^*$, where $\epsilon^* \sim N_n(0, I)$

- A unique nonsingular symmetric matrix $V^{1/2}$ exists such that $V^{1/2}V^{1/2} = V$

- Multiplying both sides of the equation

$$y = X\beta + \epsilon$$

by $V^{-1/2}$ yields

$$V^{-1/2}y = V^{-1/2}X\beta + V^{-1/2}\epsilon$$

## Weighted Least Squares

- Thus, we have $y^* = X^*\beta + \epsilon^*$, where

$$y^* = V^{-1/2}y$$

$$X^* = V^{-1/2}X$$

$$\epsilon^* = V^{-1/2}\epsilon$$

$$\mathrm{E}(\epsilon^*) = \mathrm{E}(V^{-1/2}\epsilon) = V^{-1/2}\mathrm{E}(\epsilon) = 0,$$

$$\begin{aligned}\mathrm{Var}(\epsilon^*) &= \mathrm{Var}(V^{-1/2}\epsilon) \\ &= V^{-1/2}\mathrm{Var}(\epsilon)V^{-1/2'} \\ &= V^{-1/2}VV^{-1/2} \\ &= V^{-1/2}(V^{1/2}V^{1/2})V^{-1/2} \\ &= I\end{aligned}$$

- $\widehat{\beta} = b$, the least squares estimator of $\beta$, is found by minimizing the error sum of squares

# Parameter Estimation

$$\begin{aligned}
\mathrm{SSE} &= \sum_{i=1}^{n} e_i^{*2} \\
&= (y^* - X^* b)'(y^* - X^* b) \\
&= (V^{-1/2}y - V^{-1/2}Xb)'(V^{-1/2}y - V^{-1/2}Xb) \\
&= [V^{-1/2}(y - Xb)]'[V^{-1/2}(y - Xb)] \\
&= (y - Xb)'V^{-1/2\prime}V^{-1/2}(y - Xb) \\
&= (y - Xb)'V^{-1}(y - Xb) \\
&= (y' - b'X')V^{-1}(y - Xb) \\
&= y'V^{-1}y - 2b'X'V^{-1}y + b'X'V^{-1}Xb
\end{aligned}$$

$$\frac{\partial \mathrm{SSE}}{\partial b} = -2X'V^{-1}y + 2X'V^{-1}Xb$$

$$\frac{\partial \mathrm{SSE}}{\partial b} = 0 \Longrightarrow X'V^{-1}Xb = X'V^{-1}y$$

## Parameter Estimation

- If $X$ is of full rank, $b = (X'V^{-1}X)^{-1}(X'V^{-1}y)$
- Since $b$ is a linear function of $y$, $b$ is normally-distributed with mean and variance given by:

$$\begin{aligned}
\mathrm{E}(b) &= (X'V^{-1}X)^{-1}X'V^{-1}\mathrm{E}(y) \\
&= (X'V^{-1}X)^{-1}X'V^{-1}X\beta \\
&= \beta
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(b) &= \mathrm{Var}[(X'V^{-1}X)^{-1}X'V^{-1}y] \\
&= (X'V^{-1}X)^{-1}X'V^{-1}\mathrm{Var}(y) \\
&\qquad \times [(X'V^{-1}X)^{-1}X'V^{-1}]' \\
&= (X'V^{-1}X)^{-1}X'V^{-1}V \\
&\qquad \times V^{-1}X(X'V^{-1}X)^{-1} \\
&= (X'V^{-1}X)^{-1}(X'V^{-1}X)(X'V^{-1}X)^{-1} \\
&= (X'V^{-1}X)^{-1}
\end{aligned}$$

## Sums of Squares

$$
\begin{aligned}
\text{SST} &= y^{*\prime} y^{*} \\
&= (V^{-1/2} y)' V^{-1/2} y \\
&= y' V^{-1/2} V^{-1/2} y \\
&= y' V^{-1} y
\end{aligned}
$$

$$
\begin{aligned}
\text{SSE} &= y'V^{-1}y - 2b'X'V^{-1}y + b'X'V^{-1}Xb \\
&= y'V^{-1}y - b'X'V^{-1}y - b'X'V^{-1}y \\
&\qquad + b'X'V^{-1}Xb \\
&= y'V^{-1}y - b'X'V^{-1}y \\
&\qquad - b'(X'V^{-1}y - X'V^{-1}Xb) \\
&= y'V^{-1}y - b'X'V^{-1}y - b' \times \\
&\quad (X'V^{-1}y - X'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}y) \\
&= y'V^{-1}y - b'X'V^{-1}y \\
&\qquad - b'(X'V^{-1}y - X'V^{-1}y) \\
&= y'V^{-1}y - b'X'V^{-1}y
\end{aligned}
$$

## Sums of Squares

- Since $X'V^{-1}y = X'V^{-1}Xb$,

$$\begin{aligned} \text{SSE} &= y'V^{-1}y - b'X'V^{-1}Xb \\ &= y'V^{-1}y - (Xb)'V^{-1}Xb \\ &= \text{SST} - \text{SSR} \end{aligned}$$

  where SSR is the regression sum of squares

- For theoretical purposes, it is useful to express SSE as a quadratic form in $y$:

$$\begin{aligned} \text{SSE} &= y'V^{-1}y - b'X'V^{-1}y \\ &= y'V^{-1}y - [(X'V^{-1}X)^{-1}(X'V^{-1}y)]'X'V^{-1}y \\ &= y'V^{-1}y - y'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}y \\ &= y'Ly \end{aligned}$$

  where $L = V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}$

- It can be shown that the rank of $L$ is $n - p$

## Hypothesis Tests

- The fit of the model can be tested using the minimum value of SSE

$$W = \min \mathrm{SSE} = y'V^{-1}y - (Xb)'V^{-1}Xb$$

- If the model fits, then $W$ has a chi-square distribution with $n - p$ degrees of freedom

  i.e., under $H_0\colon \mathrm{E}(y) = X\beta$, $W \sim \chi^2_{n-p}$

- If the model fits, additional hypotheses of the form $H_0\colon C\beta = 0$ may be tested, where $C$ is a $c \times p$ coefficient matrix

- The Wald statistic

$$W_C = (Cb)'[C(X'V^{-1}X)^{-1}C']^{-1}Cb$$

has a $\chi^2_c$ distribution if $H_0$ is true

## Predicted Values and Residuals

- $\widehat{y} = (\widehat{y}_1, \ldots, \widehat{y}_n)' = Xb = X(X'V^{-1}X)^{-1}X'V^{-1}y$

$$\mathrm{E}(\widehat{y}) = \mathrm{E}(Xb) = X\beta$$

$$\begin{aligned}\mathrm{Var}(\widehat{y}) &= \mathrm{Var}(Xb) = X\mathrm{Var}(b)X' \\ &= X(X'V^{-1}X)^{-1}X'\end{aligned}$$

- $r = y - \widehat{y} = y - X(X'V^{-1}X)^{-1}X'V^{-1}y$

$$\mathrm{E}(r) = \mathrm{E}(y - \widehat{y}) = X\beta - X\beta = 0$$

$$\begin{aligned}\mathrm{Var}(r) &= \mathrm{Var}[(I - X(X'V^{-1}X)^{-1}X'V^{-1})y] \\ &= (I - X(X'V^{-1}X)^{-1}X'V^{-1})V \\ &\qquad \times (I - V^{-1}X(X'V^{-1}X)^{-1}X') \\ &= (V - X(X'V^{-1}X)^{-1}X') \\ &\qquad \times (I - V^{-1}X(X'V^{-1}X)^{-1}X') \\ &= V - X(X'V^{-1}X)^{-1}X'\end{aligned}$$

## Nonlinear Response Functions

- We start from the assumption that the underlying vector of multinomial proportions is approximately normal: $p \approx N_{c^t}(\pi, V_p)$

- For linear response functions $f(p) = Ap$, $f(p) \approx N_u(A\pi, V_f)$, where $V_f = AV_pA'$

- In many applications, the response functions of interest are nonlinear functions of $p$

  marginal logits for binary responses

  generalized logits for polytomous responses

  cumulative logits for ordinal responses

- Nonlinear response functions also arise when there are missing data

## Univariate Taylor Series Approximations

- Let $X$ be a random variable with known mean and variance

  $$E(X) = \mu, \quad \text{Var}(X) = E[(X-\mu)^2] = \sigma^2$$

- Let $Y = g(X)$, where the function $g(x)$ has first and second derivatives

- Suppose that exact calculation of $E(Y)$ and $\text{Var}(Y)$ is difficult

- We will expand $g(X)$ in a Taylor series about $\mu$ and use this series representation to approximate $E(Y)$ and $\text{Var}(Y)$

## Univariate Taylor Series Approximations

- The first three terms are

$$g(X) = g(\mu) + g'(\mu)(X - \mu) + \frac{1}{2}g''(\mu)(X - \mu)^2$$

- The approximation for the mean of $Y$ is

$$\begin{aligned} \mathrm{E}(Y) &\doteq \mathrm{E}\big[g(\mu) + g'(\mu)(X - \mu) + \frac{1}{2}g''(\mu)(X - \mu)^2\big] \\ &= g(\mu) + \frac{1}{2}g''(\mu)\mathrm{Var}(X) \end{aligned}$$

- Using the linear term only, $\mathrm{E}(Y) \doteq g(\mu)$

- The approximation for the variance of $Y$ is

$$\begin{aligned} \mathrm{Var}(Y) &= \mathrm{E}\Big[\big(g(X) - \mathrm{E}[g(X)]\big)^2\Big] \\ &\doteq \mathrm{E}\Big[\big(g(\mu) + g'(\mu)(X - \mu) - g(\mu)\big)^2\Big] \\ &= \mathrm{E}\Big[\big(g'(\mu)(X - \mu)\big)^2\Big] \\ &= \big(g'(\mu)\big)^2\,\mathrm{E}\big[(X - \mu)^2\big] \\ &= \big(g'(\mu)\big)^2\,\mathrm{Var}(X) \end{aligned}$$

## The $\delta$ Method

- Let $X_n$ be a random variable such that the asymptotic distribution of $\sqrt{n}(X_n - \mu)$ is $N\big(0, \sigma^2(\mu)\big)$

- Let $g(x)$ be a function that can be differentiated at $x{=}\mu$ so that it has the following expansion as $x \to \mu$:

$$g(x) = g(\mu) + g'(\mu)(x - \mu) + o(|x - \mu|)$$

- Then the asymptotic distribution of $\sqrt{n}\big(g(X_n) - g(\mu)\big)$ is $N\Big(0, \sigma^2(\mu)\big(g'(\mu)\big)^2\Big)$

### Definition of $o(|x - \mu|)$

If $x_n$ is any sequence such that $x_n \to \mu$ and if

$$a_n = g(x_n) - g(\mu) - g'(\mu)(x_n - \mu),$$
$$b_n = x_n - \mu,$$

then for any $\epsilon > 0$, there exists $n(\epsilon)$ such that if $n > n(\epsilon)$, then $|a_n| < \epsilon|b_n|$

## Multivariate Taylor Series Approximations

- Let $X = (X_1, \ldots, X_n)'$ be a random vector with known mean vector and covariance matrix
  $\mathrm{E}(X) = \mu, \quad \mathrm{Var}(X) = \Sigma$

- Let $Y = g(X_1, \ldots, X_n)$, where $g(x_1, \ldots, x_n)$ is a continuous function with first and second partial derivatives

- Expand $g(X)$ in a Taylor series about $\mu$:

$$g(x) = g(\mu) + \sum_{i=1}^{n} \frac{\partial g}{\partial \mu_i}(x_i - \mu_i) + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 g}{\partial \mu_i \partial \mu_j}(x_i - \mu_i)(x_j - \mu_j),$$

  where

$$\frac{\partial g}{\partial \mu_i} = \frac{\partial g}{\partial x_i}\Big|_{x=\mu},$$

$$\frac{\partial^2 g}{\partial \mu_i \partial \mu_j} = \frac{\partial^2 g}{\partial x_i \partial x_j}\Big|_{x=\mu}$$

## Multivariate Taylor Series Approximations

- Let $g^{(1)}(\mu) = \left(\dfrac{\partial g}{\partial \mu_1}, \ldots, \dfrac{\partial g}{\partial \mu_n}\right)$ be the row vector of first partial derivatives

- $y = g(x_1, \ldots, x_n) \doteq g(\mu) + \big(g^{(1)}(\mu)\big)(x - \mu)$

- The approximate mean and variance of $Y$ are

$$\begin{aligned}
\mathrm{E}(Y) &\doteq \mathrm{E}[g(\mu)] + \big(g^{(1)}(\mu)\big)\mathrm{E}(X - \mu) \\
&= g(\mu)
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(Y) &= \mathrm{E}\Big[\big(g(X) - \mathrm{E}[g(X)]\big)^2\Big] \\
&\doteq \mathrm{E}\Big[\big(g(\mu) + \big(g^{(1)}(\mu)\big)(X - \mu) - g(\mu)\big)^2\Big] \\
&= \mathrm{E}\Big[\Big(\big(g^{(1)}(\mu)\big)(X - \mu)\Big)^2\Big] \\
&= \mathrm{E}\Big[\big(g^{(1)}(\mu)\big)(X - \mu)(X - \mu)'\big(g^{(1)}(\mu)\big)'\Big] \\
&= \big(g^{(1)}(\mu)\big)\mathrm{E}\big[(X - \mu)(X - \mu)'\big]\big(g^{(1)}(\mu)\big)' \\
&= \big(g^{(1)}(\mu)\big)\,\Sigma\,\big(g^{(1)}(\mu)\big)'
\end{aligned}$$

## Taylor Series Approximations for Multiple Functions of a Random Vector

- Let $X = (X_1, \ldots, X_n)'$ be a random vector with known mean vector and covariance matrix $\mathrm{E}(X) = \mu, \quad \mathrm{Var}(X) = \Sigma$

- Let $Y = (Y_1, \ldots, Y_m)'$, where

$$Y_i = g_i(X_1, \ldots, X_n),$$

  for $i = 1, \ldots, m$

- From the results for a univariate function of a random vector:

$$\mathrm{E}(Y_i) \doteq g_i(\mu)$$

$$\mathrm{Var}(Y_i) \doteq \left(g_i^{(1)}(\mu)\right) \Sigma \left(g_i^{(1)}(\mu)\right)'$$

- We would now like to approximate $\mathrm{Cov}(Y_i, Y_j)$

## Taylor Series Approximations for Multiple Functions of a Random Vector

$$\begin{aligned}
\mathrm{Cov}(Y_i, Y_j) &= \mathrm{E}\big[\big(Y_i - \mathrm{E}(Y_i)\big)\big(Y_j - \mathrm{E}(Y_j)\big)\big] \\
&\doteq \mathrm{E}\big[[g_i^{(1)}(\mu)](X-\mu)[g_j^{(1)}(\mu)](X-\mu)\big] \\
&= \mathrm{E}\big[[g_i^{(1)}(\mu)](X-\mu)(X-\mu)'[g_j^{(1)}(\mu)]'\big] \\
&= [g_i^{(1)}(\mu)]\mathrm{E}\big[(X-\mu)(X-\mu)'\big][g_j^{(1)}(\mu)]' \\
&= [g_i^{(1)}(\mu)]\Sigma[g_j^{(1)}(\mu)]'\big]
\end{aligned}$$

- Now let $\left(\dfrac{\partial g}{\partial \mu}\right)$ denote the $m \times n$ matrix whose $i$th row is $g_i^{(1)}(\mu)$
- The $(i, j)$ element of $\left(\dfrac{\partial g}{\partial \mu}\right)$ is $\left.\dfrac{\partial g_i}{\partial x_j}\right|_{x=\mu}$
- The approximate mean and covariance matrix of $Y$ are:

$$\mathrm{E}(Y) \doteq g(\mu)$$

$$\mathrm{Var}(Y) \doteq \left(\frac{\partial g}{\partial \mu}\right)\Sigma\left(\frac{\partial g}{\partial \mu}\right)'$$

## The Multivariate $\delta$ Method

- Let $X_n$ be a $p$-dimensional random variable such that the asymptotic distribution of $\sqrt{n}(X_n - \mu)$ is $N_p\big(0, \Sigma(\mu)\big)$

- Let $g(X_n) = \big(g_1(X_n), \ldots, g_u(X_n)\big)'$ be a function with the following expansions as $X_n \to \mu$:

$$g_i(x) = g_i(\mu) + \sum_{j=1}^{p}(x_j - \mu_j)\frac{\partial g_i}{\partial x_j}\Big|_{x=\mu} + o(\|x - \mu\|)$$

- Let $\dfrac{\partial g}{\partial \mu}$ be the $u \times p$ matrix whose $(i, j)$ entry is $\dfrac{\partial g_i}{\partial x_j}\Big|_{x=\mu}$

- Then the asymptotic distribution of $\sqrt{n}\big(g(X_n) - g(\mu)\big)$ is

$$N_u\left(0, \left(\frac{\partial g}{\partial \mu}\right)\Sigma(\mu)\left(\frac{\partial g}{\partial \mu}\right)'\right)$$

## Variances of Linear Functions

- Let $x_{(n\times 1)}$ have mean $\mu$ and covariance matrix $\Sigma$

- Let $y_{(m\times 1)} = f(x) = A_{(m\times n)}x$

$$= \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{j=1}^n a_{1j}x_j \\ \vdots \\ \sum_{j=1}^n a_{mj}x_j \end{pmatrix} = \begin{pmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{pmatrix}$$

- $\dfrac{\partial f_1}{\partial x} = (a_{11}, \dots, a_{1n}), \dots, \dfrac{\partial f_m}{\partial x} = (a_{m1}, \dots, a_{mn})$,
and

$$\left(\frac{\partial f}{\partial x}\right) = \begin{bmatrix} \dfrac{\partial f_1}{\partial x} \\ \vdots \\ \dfrac{\partial f_m}{\partial x} \end{bmatrix} = A$$

- Therefore, $\mathrm{Var}(y) = \left(\dfrac{\partial f}{\partial x}\right) \Sigma \left(\dfrac{\partial f}{\partial x}\right)' = A\,\Sigma\,A'$

## Variance Approximations for Logarithmic Functions

- Let $x$ be a $n \times 1$ vector with mean $\mu$ and covariance matrix $\Sigma$

- Let $y_{(n\times 1)} = f(x) = \log x$

$$\begin{aligned} \text{i.e.,}\ \ y_1 &= f_1(x) = \log(x_1) \\ &\vdots \\ y_n &= f_n(x) = \log(x_n) \end{aligned}$$

- The partial derivatives are given by:

$$\begin{aligned} \frac{\partial f_1}{\partial x} &= (1/x_1, 0, \ldots, 0) \\ \frac{\partial f_2}{\partial x} &= (0, 1/x_2, 0, \ldots, 0) \\ &\vdots \\ \frac{\partial f_n}{\partial x} &= (0, \ldots, 0, 1/x_n) \end{aligned}$$

## Variance Approximations for Logarithmic Functions

$$\frac{\partial f}{\partial x} = \begin{pmatrix} \dfrac{\partial f_1}{\partial x} \\ \dfrac{\partial f_2}{\partial x} \\ \vdots \\ \dfrac{\partial f_n}{\partial x} \end{pmatrix} = \begin{pmatrix} 1/x_1 & 0 & \cdots & 0 \\ 0 & 1/x_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & 1/x_n \end{pmatrix}$$

$$\frac{\partial f}{\partial x}\Big|_{x=\mu} = \begin{pmatrix} 1/\mu_1 & 0 & \cdots & 0 \\ 0 & 1/\mu_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & 1/\mu_n \end{pmatrix}$$

$$= D_\mu^{-1},$$

$$\mathrm{Var}(y) \doteq \left(\frac{\partial f}{\partial x}\Big|_{x=\mu}\right) \Sigma \left(\frac{\partial f}{\partial x}\Big|_{x=\mu}\right)'$$

$$= D_\mu^{-1}\, \Sigma\, D_\mu^{-1}$$

## Variance Approximations for Exponential Functions

- Let $x$ be a $n \times 1$ vector with mean $\mu$ and covariance matrix $\Sigma$

- Let $y_{(n\times 1)} = f(x) = \exp(x)$

$$\begin{aligned} \text{i.e., } \; y_1 &= f_1(x) = e^{x_1} \\ &\vdots \\ y_n &= f_n(x) = e^{x_n} \end{aligned}$$

- The partial derivatives are given by:

$$\begin{aligned} \frac{\partial f_1}{\partial x} &= (e^{x_1}, 0, \ldots, 0) \\ \frac{\partial f_2}{\partial x} &= (0, e^{x_2}, 0, \ldots, 0) \\ &\vdots \\ \frac{\partial f_n}{\partial x} &= (0, \ldots, 0, e^{x_n}) \end{aligned}$$

# Variance Approximations for Exponential Functions

$$\frac{\partial f}{\partial x} = \begin{pmatrix} \dfrac{\partial f_1}{\partial x} \\ \dfrac{\partial f_2}{\partial x} \\ \vdots \\ \dfrac{\partial f_n}{\partial x} \end{pmatrix} = \begin{pmatrix} e^{x_1} & 0 & \cdots & 0 \\ 0 & e^{x_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & e^{x_n} \end{pmatrix}$$

$$\frac{\partial f}{\partial x}\Big|_{x=\mu} = \begin{pmatrix} e^{\mu_1} & 0 & \cdots & 0 \\ 0 & e^{\mu_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & e^{\mu_n} \end{pmatrix}$$

$$= D_{e^{\mu}},$$

$$\mathrm{Var}(y) \doteq \left(\frac{\partial f}{\partial x}\Big|_{x=\mu}\right) \Sigma \left(\frac{\partial f}{\partial x}\Big|_{x=\mu}\right)'$$

$$= D_{e^{\mu}}\, \Sigma\, D_{e^{\mu}}$$

## Variance Approximations for Compound Functions

- Two types of compound functions are commonly used:

  $f(p) = A_2 \log(A_1 p)$

  $g(p) = \exp\big(A_2 \log(A_1 p)\big) = \exp\big(f(p)\big)$

- We wish to approximate $V_f = \mathrm{Var}\big(f(p)\big)$ and $V_g = \mathrm{Var}\big(g(p)\big)$

- First, let $f_1(p) = A_1 p$

$$V_{f_1} = \mathrm{Var}\big(f_1(p)\big) = A_1\, V_p\, A_1'$$

- Now, let $f_2(p) = \log\big(f_1(p)\big) = \log(A_1 p)$

$$\begin{aligned} V_{f_2} &= \mathrm{Var}\big(f_2(p)\big) = \mathrm{Var}\Big(\log\big(f_1(p)\big)\Big) \\ &\doteq D_{f_1}^{-1} V_{f_1} D_{f_1}^{-1} = D_{f_1}^{-1} A_1 V_p A_1' D_{f_1}^{-1} \end{aligned}$$

## Variance Approximations for Compound Functions

- Finally, let $f(p) = A_2 f_2(p) = A_2 \log(A_1 p)$

$$\begin{aligned} V_f &= \mathrm{Var}\big(f(p)\big) \\ &= \mathrm{Var}\big(A_2 f_2(p)\big) \\ &= A_2 V_{f_2} A_2' \\ &\doteq A_2 D_{f_1}^{-1} A_1 V_p A_1' D_{f_1}^{-1} A_2' \end{aligned}$$

- Now, let $g(p) = \exp\big(f(p)\big) = \exp\big(A_2 \log(A_1 p)\big)$

$$\begin{aligned} V_g &= \mathrm{Var}\big(g(p)\big) \\ &= \mathrm{Var}\big(\exp(f(p))\big) \\ &\doteq D_{e^f} V_f D_{e^f} \\ &= D_{e^f} A_2 D_{f_1}^{-1} A_1 V_p A_1' D_{f_1}^{-1} A_2' D_{e^f} \end{aligned}$$

## One-Sample Repeated Measures

- The data are as follows:

| | Time Point | | | | |
|---|---|---|---|---|---|
| Subject | 1 | $\ldots$ | $j$ | $\ldots$ | $t$ |
| 1 | $y_{11}$ | $\ldots$ | $y_{1j}$ | $\ldots$ | $y_{1t}$ |
| $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ | $\ddots$ | $\vdots$ |
| $i$ | $y_{i1}$ | $\ldots$ | $y_{ij}$ | $\ldots$ | $y_{it}$ |
| $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ | $\ddots$ | $\vdots$ |
| $n$ | $y_{n1}$ | $\ldots$ | $y_{nj}$ | $\ldots$ | $y_{nt}$ |

- $y_{ij}$ is a categorical response variable with $c$ possible outcomes

- These data correspond to a sample of size $n$ from a multinomial distribution with $c^t$ potential outcomes

## Example

- 46 subjects were treated with each of three drugs (A, B, and C)

- The response to each drug was recorded as favorable (F) or unfavorable (U)

- We wish to test if the drugs have similar response profiles

- Since the response is dichotomous, the null hypothesis can be written as

$$H_0: \ \pi_A = \pi_B = \pi_C,$$

where $\pi_X$=Pr(favorable response with drug $X$)

## Reference

Grizzle, J. E., Starmer, C. F., and Koch, G. G. (1969). Analysis of categorical data by linear models. *Biometrics* **25**, 489–504.

# Drug Response Data

| | Drug | | | | Drug | | |
|---|---|---|---|---|---|---|---|
| Subject | A | B | C | Subject | A | B | C |
| 1 | F | F | U | 24 | U | F | U |
| 2 | U | U | U | 25 | F | F | U |
| 3 | U | U | F | 26 | U | U | U |
| 4 | F | F | U | 27 | F | U | U |
| 5 | U | U | U | 28 | U | U | F |
| 6 | F | F | U | 29 | U | U | U |
| 7 | F | F | F | 30 | F | F | U |
| 8 | F | F | U | 31 | F | F | F |
| 9 | F | U | U | 32 | F | U | F |
| 10 | U | U | F | 33 | F | F | U |
| 11 | F | F | U | 34 | U | F | F |
| 12 | U | F | U | 35 | U | F | U |
| 13 | F | F | F | 36 | F | F | U |
| 14 | F | F | U | 37 | F | F | U |
| 15 | U | F | F | 38 | F | F | F |
| 16 | F | U | F | 39 | F | U | U |
| 17 | U | U | U | 40 | U | U | F |
| 18 | F | F | U | 41 | F | F | U |
| 19 | F | U | U | 42 | U | U | U |
| 20 | U | U | F | 43 | U | U | F |
| 21 | F | F | F | 44 | F | F | U |
| 22 | F | F | U | 45 | F | F | F |
| 23 | F | F | U | 46 | U | F | U |

## Multinomial Structure of the Example

- There are $c = 2$ possible outcomes at each of $t = 3$ time points
- Thus, there are $c^t = 2^3 = 8$ potential response profiles
- The observed data can be displayed as follows:

| Response Profile | Drug A | Drug B | Drug C | Frequency |
|---|---|---|---|---|
| 1 | F | F | F | 6 |
| 2 | F | F | U | 16 |
| 3 | F | U | F | 2 |
| 4 | F | U | U | 4 |
| 5 | U | F | F | 2 |
| 6 | U | F | U | 4 |
| 7 | U | U | F | 6 |
| 8 | U | U | U | 6 |
| Total | | | | 46 |

## WLS Approach

- The underlying multinomial probability vector will be denoted by $\pi = (\pi_1, \ldots, \pi_{c^t})'$

- The corresponding vector of sample proportions $p$ is an unbiased estimator of the probability vector $\pi$

- A consistent estimator of the variance-covariance matrix of $p$ is given by the $c^t \times c^t$ matrix
$$V_p = \frac{1}{n}(D_p - pp')$$

- Since the elements of $\pi$ (and $p$) are linearly dependent, we must restrict consideration to a set of linearly independent functions $f(p)$

## Models for Linear Functions of the Response Proportions

- Consider models of the form $f(p) = X\beta$

  $f(p) = Ap$, where $A$ is a $u \times c^t$ matrix of rank $u$

  $X$ is an $u \times v$ model matrix

  $\beta$ is a $v \times 1$ vector of unknown parameters

- $p \approx N_{c^t}(\pi, V_p)$

- $f(p) = Ap \approx N_u(A\pi, V_f)$, where $V_f = AV_pA'$ has rank $u$

- Using the vector of observed functions $f(p)$ and the consistent estimator of its covariance matrix $V_f$, linear models of the form

$$f(p) = X\beta$$

  can be fit using weighted least squares

## Application to the Example

- The $8 \times 1$ vector of proportions $p$ is defined as follows:

| Response Profile | Drug A | Drug B | Drug C | Component of $p$ |
|---|---|---|---|---|
| 1 | F | F | F | $p_1$ |
| 2 | F | F | U | $p_2$ |
| 3 | F | U | F | $p_3$ |
| 4 | F | U | U | $p_4$ |
| 5 | U | F | F | $p_5$ |
| 6 | U | F | U | $p_6$ |
| 7 | U | U | F | $p_7$ |
| 8 | U | U | U | $p_8$ |

- Let $A$ be the $3 \times 8$ matrix given by

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}$$

## Application to the Example

- The transformation $f(p) = Ap$ defines the response functions

$$f(p) = (p_A, p_B, p_C)',$$

where $p_X$ is the observed proportion with a favorable response to drug $X$

- A consistent estimator of the $3 \times 3$ covariance matrix of $f(p)$ is given by

$$V_f = AV_pA' = \frac{1}{n}A(D_p - pp')A'$$

- Note that the elements of $V_f$ are consistent estimates of

$$\begin{pmatrix} \text{Var}(p_A) & \text{Cov}(p_A, p_B) & \text{Cov}(p_A, p_C) \\ \text{Cov}(p_A, p_B) & \text{Var}(p_B) & \text{Cov}(p_B, p_C) \\ \text{Cov}(p_A, p_C) & \text{Cov}(p_B, p_C) & \text{Var}(p_C) \end{pmatrix}$$

## Model Fitting and Test of Marginal Homogeneity

The null hypothesis is $H_0$: $\pi_A = \pi_B = \pi_C$

<u>Method 1</u>:

- Fit the model $f(p) = X_1\beta$, where $X_1 = I_3$

  a saturated model with 0 df for lack-of-fit

- Test marginal homogeneity as $H_0$: $C\beta = 0$, where

$$C = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix}$$

- If $H_0$ is rejected, pairwise comparisons between drugs can be tested using:

| $H_0$ | C |
|---|---|
| $\pi_A = \pi_B$ | $(1 \quad -1 \quad 0)$ |
| $\pi_A = \pi_C$ | $(1 \quad 0 \quad -1)$ |
| $\pi_B = \pi_C$ | $(0 \quad 1 \quad -1)$ |

## SAS Statements for Method 1

```
data a;
input subject a $ b $ c $;
cards;
 1 F F U
 2 U U U
   ...
45 F F F
46 U F U
 ;
proc catmod;
response 1 1 1 1 0 0 0 0,
         1 1 0 0 1 1 0 0,
         1 0 1 0 1 0 1 0;
model a*b*c=(1 0 0,
             0 1 0,
             0 0 1) / noprofile;
contrast 'A=B=C' all_parms 1 -1  0,
                 all_parms 1  0 -1;
contrast 'A=B' all_parms 1 -1  0;
contrast 'A=C' all_parms 1  0 -1;
contrast 'B=C' all_parms 0  1 -1;
```

# Model Fitting and Test of Marginal Homogeneity

Method 2:

- Fit the model $f(p) = X_2\beta$, where

$$X_2 = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \end{pmatrix}$$

  This model includes an overall intercept and two parameters for drug differences

$$\pi_A = \mu + \alpha_1$$

$$\pi_B = \mu + \alpha_2$$

$$\pi_C = \mu - \alpha_1 - \alpha_2$$

  "sum-to-zero" parameterization

- Test marginal homogeneity as $H_0$: $C\beta = 0$, where

$$C = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

# Model Fitting and Test of Marginal Homogeneity

Method 2 (continued):

- If $H_0$ is rejected, pairwise comparisons between drugs can be tested using:

| $H_0$ | C |
|---|---|
| $\pi_A = \pi_B$ | $(0 \quad 1 \quad -1)$ |
| $\pi_A = \pi_C$ | $(0 \quad 2 \quad 1)$ |
| $\pi_B = \pi_C$ | $(0 \quad 1 \quad 2)$ |

Method 3:

- Fit the model $f(p) = X_3\beta$, where

$$X_3' = (1 \quad 1 \quad 1)$$

- Marginal homogeneity is tested by the lack-of-fit statistic
- Pairwise drug comparisons can not be tested

## SAS Statements for Method 2

```
proc catmod;
response marginals;
model a*b*c=_response_ / noprofile;
repeated drug 3;
contrast 'A=B' all_parms 0 1 -1;
contrast 'A=C' all_parms 0 2  1;
contrast 'B=C' all_parms 0 1  2;
```

- For a $k$-level dependent variable, `response marginals` computes the first $k-1$ independent marginal proportions
- The `_response_` effect indicates that variation among the repeated measures will be modelled
- The `repeated` statement specifies:
  1. the names of the repeated measures effect(s)
  2. the number of levels of the effects(s)
  3. the effects to be included in the model

  This statement could have been written:

  ```
  repeated drug 3 / _response_=drug;
  ```

## SAS Statements for Method 3

```
proc catmod;
response marginals;
model a*b*c=(1,
             1,
             1) / noprofile;
```

- When the model matrix is specified explicitly, the default ANOVA table contains effects labelled

```
MODEL:MEAN
RESIDUAL
```

- The `MODEL:MEAN` effect tests the significance of all sources of variation other than an overall intercept

- In this example, the only effect is the overall intercept

## Reduced Model

- At $\alpha$=.05, the marginal homogeneity hypothesis $H_0$: $\pi_A$=$\pi_B$=$\pi_C$ is rejected ($p = .037$)

- The pairwise comparisons indicate that drug C differs from drugs A and B

- A reduced model, which includes an overall intercept and an incremental effect for drug C, can be fit as follows:

```
proc catmod;
response marginals;
model a*b*c=(1 0,
             1 0,
             1 1)
      (1='Intercept',
       2='Drug C Effect') / noprofile;
```

- Tests that the labelled parameters are equal to zero will be included in the ANOVA table

## Example

- In a longitudinal study of health effects of air pollution, 1019 children were examined annually at ages 9, 10, 11, and 12

- At each examination, the response variable was the presence or absence of wheezing

- The questions of interest include:
    - Does the prevalence of wheezing change with age?
    - Is there a quantifiable trend in the age-specific prevalence rates?

**References**

Agresti, A. (1990). *Categorical Data Analysis.* New York: John Wiley and Sons, p. 408.

Ware, J. H., Lipsitz, S., and Speizer, F. E. (1988). Issues in the analysis of repeated categorical outcomes. *Statist Med* **7**, 95–107.

# Respiratory Illness Data

| Wheeze (1=present, 2=absent) | | | | No. of |
|---|---|---|---|---|
| Age 9 | Age 10 | Age 11 | Age 12 | Children |
| 1 | 1 | 1 | 1 | 94 |
| 1 | 1 | 1 | 2 | 30 |
| 1 | 1 | 2 | 1 | 15 |
| 1 | 1 | 2 | 2 | 28 |
| 1 | 2 | 1 | 1 | 14 |
| 1 | 2 | 1 | 2 | 9 |
| 1 | 2 | 2 | 1 | 12 |
| 1 | 2 | 2 | 2 | 63 |
| 2 | 1 | 1 | 1 | 19 |
| 2 | 1 | 1 | 2 | 15 |
| 2 | 1 | 2 | 1 | 10 |
| 2 | 1 | 2 | 2 | 44 |
| 2 | 2 | 1 | 1 | 17 |
| 2 | 2 | 1 | 2 | 42 |
| 2 | 2 | 2 | 1 | 35 |
| 2 | 2 | 2 | 2 | 572 |
| Total | | | | 1019 |

## Response Functions

- $c$=2 outcomes, $t$=4 time points, $c^t$=$2^4$=16 response profiles

- Let $p$ denote the $16 \times 1$ vector of proportions corresponding to the multi-way cross-classification of response at the four ages (ordered as shown)

- Let $P_x$ denote the marginal prevalence of wheezing at age $x$, for $x = 9, \ldots, 12$

- The response functions of interest are given by $f(p) = (P_9, P_{10}, P_{11}, P_{12})' = Ap$, where $A$ is the $4 \times 16$ matrix

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}$$

## Model 1

- Test $H_0$: $\Pi_9 = \Pi_{10} = \Pi_{11} = \Pi_{12}$

- Instead of one observation per child, there is one observation per combination of response categories

- In this situation, the `weight` statement is used

```
data a;
input w9 w10 w11 w12 count;
cards;
1 1 1 1  94
1 1 1 2  30
   ...
2 2 2 1  35
2 2 2 2 572
 ;
proc catmod; weight count;
response marginals;
model w9*w10*w11*w12=_response_
      / noprofile;
repeated age 4;
```

## Model 2

- Test linear and nonlinear components of the age effect

- Since the observations are equally spaced, orthogonal polynomial coefficients can be used

```
proc catmod; weight count;
response marginals;
model w9*w10*w11*w12=(1 -3  1 -1,
                      1 -1 -1  3,
                      1  1 -1 -3,
                      1  3  1  1)
      (1='Intercept',
   2 3 4='Age',
       2='  Linear',
       3='  Quadratic',
       4='  Cubic',
     3 4='  Nonlinear') / noprofile;
```

- Note that tests of multiple df effects can be specified in the model statement

## Model 3

- Fit the linear model $\Pi_x = \alpha + \beta(x - 9)$

```
proc catmod; weight count;
response marginals;
model w9*w10*w11*w12=(1  0,
                      1  1,
                      1  2,
                      1  3)
        (1='Intercept',
         2='Linear Age')
        / noprofile p;
```

- The `p` option prints observed and predicted response functions

- Standard errors of the observed and predicted response functions are also printed

## Results

- This model provides a good fit to the observed data

  $W = .54$ with 2 df, $p = .762$

- The predicted prevalence ($\pm$ SE) at age 9 is $.263 \pm .013$

- The linear effect of age is highly significant

  $W_{\text{age}} = 12.31$, $p < .001$

- The probability of wheezing is estimated to decrease by $.0161 \pm .0046$ per year of age

**Example**

- In a longitudinal study of health effects of air pollution, 1019 children were examined annually at ages 9, 10, 11, and 12

- At each examination, the response variable was the presence or absence of wheezing

- The previous model was $\Pi_x = \alpha + \beta(x - 9)$, where $\Pi_x =$ marginal prob. of wheezing at age $x$

- It may also be of interest to analyze these data on the logit scale

  - age effects are multiplicative (instead of additive)

  - predicted probabilities are constrained to (0,1)

  - parameters have odds ratio interpretations

## Response Functions

- $c = 2$ outcomes, $t = 4$ time points, $c^t = 2^4 = 16$ response profiles

- The marginal logit response functions can be defined as $f(p) = A_2 \log(A_1 p)$, where $p$ is the $16 \times 1$ vector of multinomial proportions, $A_1$ is the $8 \times 16$ matrix

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix},$$

and $A_2 = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}$

## Logit Model

- The model is $L_x = \alpha + \beta(x - 9)$, where $L_x = \log\big(\Pi_x/(1 - \Pi_x)\big)$

```
proc catmod; weight count;
response 1 -1  0  0  0  0  0  0,
         0  0  1 -1  0  0  0  0,
         0  0  0  0  1 -1  0  0,
         0  0  0  0  0  0  1 -1
log
1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0,
0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1,
1 1 1 1 0 0 0 0 1 1 1 1 0 0 0 0,
0 0 0 0 1 1 1 1 0 0 0 0 1 1 1 1,
1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0,
0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1,
1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0,
0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1;
model w9*w10*w11*w12=(1  0,
                      1  1,
                      1  2,
                      1  3)
      (1='Intercept',
       2='Linear Age') / noprofile p;
```

## Interpretation of Logit Model

- The model provides a good fit to the observed data ($W = .67$ with 2 df, $p = .72$)
- The age effect is highly significant ($W_{\text{age}} = 11.77$, $p < .001$)
- The predicted model is $\widehat{L}_x = \widehat{\alpha} + \widehat{\beta}(x - 9)$, where $\widehat{\alpha} = -1.0276$ and $\widehat{\beta} = -.0879$
- The log-odds in favor of wheezing are estimated to decrease by .0879 per year
- The estimated odds ratio is $e^{-.0879} = .916$
- The odds of wheezing are 0.916 times as great at age $x$ than at age $x - 1$
- The odds against wheezing are 1.092 (1/.916) times as high at age $x$ than at age $x - 1$

## Polytomous Response Variables

- When the response is dichotomous:

  There is one response function per time point

  The test of marginal homogeneity has $t - 1$ df

- If $y_{ij}$ has $c$ possible outcomes, there are at most $c - 1$ linearly independent response functions per time point

  Test of marginal homogeneity has $(c-1)(t-1)$ df

- For nominal response variables, the natural linear response functions are marginal proportions

- For ordinal response variables, cumulative marginal proportions and mean scores can also be considered

## Example

- In the Iowa 65+ Rural Health Survey, 1926 elderly individuals were followed over a six-year period

- Each individual was surveyed at years 0, 3, and 6

- One of the variables of interest was the number of friends reported

  An ordinal categorical response:

  no friends

  1–2 friends

  3 or more friends

- We wish to determine if the distribution of the number of reported friends is changing over time

**Subject Classification, by No. of Friends**

| Year 0 | Year 3 | Year 6 | Count |
|---|---|---|---|
| 0 | 0 | 0 | 31 |
| 0 | 0 | 1–2 | 22 |
| 0 | 0 | 3+ | 54 |
| 0 | 1–2 | 0 | 15 |
| 0 | 1–2 | 1–2 | 25 |
| 0 | 1–2 | 3+ | 50 |
| 0 | 3+ | 0 | 22 |
| 0 | 3+ | 1–2 | 20 |
| 0 | 3+ | 3+ | 139 |
| 1–2 | 0 | 0 | 11 |
| 1–2 | 0 | 1–2 | 13 |
| 1–2 | 0 | 3+ | 30 |
| 1–2 | 1–2 | 0 | 12 |
| 1–2 | 1–2 | 1–2 | 64 |
| 1–2 | 1–2 | 3+ | 82 |
| 1–2 | 3+ | 0 | 13 |
| 1–2 | 3+ | 1–2 | 44 |
| 1–2 | 3+ | 3+ | 189 |
| 3+ | 0 | 0 | 9 |
| 3+ | 0 | 1–2 | 21 |
| 3+ | 0 | 3+ | 44 |
| 3+ | 1–2 | 0 | 18 |
| 3+ | 1–2 | 1–2 | 55 |
| 3+ | 1–2 | 3+ | 121 |
| 3+ | 3+ | 0 | 31 |
| 3+ | 3+ | 1–2 | 85 |
| 3+ | 3+ | 3+ | 706 |
| Total | | | 1926 |

## Models for Marginal Proportions

- $c = 3, \quad t = 3, \quad c^t = 3^3 = 27$ response profiles

- Let $p$ denote the $27 \times 1$ vector of proportions corresponding to the multi-way cross-classification of response at the three time points (ordered as shown)

- Let $p_{ij}$ denote the marginal proportion of subjects at year $i$ in response category $j$

  $i = 0,\ 3,\ 6, \quad j = 0,\ 1\{–2\},\ 3\{+\}$

- Linearly independent response functions are given by

$$f(p) = (p_{00}, p_{01}, p_{30}, p_{31}, p_{60}, p_{61})'$$

## Models for Marginal Proportions

- In matrix notation, $f(p) = Ap$, where $A$ is the $6 \times 27$ matrix

$$\begin{pmatrix}
1&1&1&1&1&1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&1&1&1&1&1&1&1&1&1&0&0&0&0&0&0&0&0&0\\
1&1&1&0&0&0&0&0&0&1&1&1&0&0&0&0&0&0&1&1&1&0&0&0&0&0&0\\
0&0&0&1&1&1&0&0&0&0&0&0&1&1&1&0&0&0&0&0&0&1&1&1&0&0&0\\
1&0&0&1&0&0&1&0&0&1&0&0&1&0&0&1&0&0&1&0&0&1&0&0&1&0&0\\
0&1&0&0&1&0&0&1&0&0&1&0&0&1&0&0&1&0&0&1&0&0&1&0&0&1&0
\end{pmatrix}$$

- In PROC CATMOD, these response functions can be defined by:

```
response
1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0,
0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0,
1 1 1 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0,
0 0 0 1 1 1 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 1 1 1 0 0 0,
1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0,
0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0;
```

  or more simply by:

```
response marginals;
```

## Model 1: Analysis of Marginal Proportions

- For each of the marginal responses of interest [Pr(no friends), Pr(1–2 friends)], the model includes an intercept, and two parameters for linear and nonlinear time effects

- Since the measurements were equally spaced (0, 3, 6 years), orthogonal polynomial coefficients will be used in the model matrix

- The SAS statements creating the data file are:

```
data a;
input (n0 n3 n6 count)
($char3. +1 $char3. +1 $char3. 4.);
cards;
 0   0   0   31
      . . .
3+  3+  3+  706
 ;
```

## CATMOD Statements for Model 1

```
proc catmod; weight count;
response marginals;
model n0*n3*n6=(1 -1  1  0  0  0,
                0  0  0  1 -1  1,
                1  0 -2  0  0  0,
                0  0  0  1  0 -2,
                1  1  1  0  0  0,
                0  0  0  1  1  1)
      (1='Pr(0):   Intercept',
        2='        Linear',
        3='        Quadratic',
      2 3='        Overall',
        4='Pr(1-2): Intercept',
        5='        Linear',
        6='        Quadratic',
      5 6='        Overall',
      2 5='Both:   Linear',
      3 6='        Quadratic',
  2 3 5 6='Homogeneity') / noprofile;
```

## Alternate Form of Model 1

The same results can be obtained using CATMOD's default repeated measures capabilities

Advantage: model matrix need not be specified

Disadvantage: contrasts must be specified

```
proc catmod; weight count;
response marginals;
model n0*n3*n6=_response_ / noprofile;
repeated time 3;
contrast '0:    L' all_parms 0 0 2 0 1 0;
contrast '      Q' all_parms 0 0 0 0 1 0;
contrast '    L&Q' all_parms 0 0 1 0 0 0,
                   all_parms 0 0 0 0 1 0;
contrast '1-2:  L' all_parms 0 0 0 2 0 1;
contrast '      Q' all_parms 0 0 0 0 0 1;
contrast '    L&Q' all_parms 0 0 0 1 0 0,
                   all_parms 0 0 0 0 0 1;
contrast 'Both: L' all_parms 0 0 2 0 1 0,
                   all_parms 0 0 0 2 0 1;
contrast '      Q' all_parms 0 0 0 0 1 0,
                   all_parms 0 0 0 0 0 1;
```

## Mean Score Response Functions

- Applicable for discrete numeric or ordinal response variables
- Examples:
  - Number of times married (0, 1, ...)
  - Litter size (0, 1, ...)
  - Pain severity (none, mild, moderate, severe)
- If "reasonable" scores can be assigned to the levels of the response, we can analyze the change in the mean over time (rather than the change in the entire distribution)
- We now have a single response function per time point
- The hypothesis of homogeneity of means over time has $t-1$ df

**Formulation of Marginal Mean Scores**

- Consider an ordinal or discrete numeric response variable with $c$ categories
- Let $a_j$ denote the score assigned to the $j$th level of the response, for $j = 1, \ldots, c$
- Let $f_1(p) = A_1 p$ denote the $ct \times 1$ vector of marginal proportions

  $p$ is the $c^t \times 1$ vector of multinomial proportions

  $A_1$ is a $ct \times c^t$ matrix
- Let $f_2(p) = A_2\big(f_1(p)\big)$ denote the $t \times 1$ vector of marginal mean scores, where $A_2$ is the $t \times ct$ matrix

$$\begin{pmatrix} a_1 \cdots a_c & 0 \cdots 0 & \cdots & 0 \cdots 0 \\ 0 \cdots 0 & a_1 \cdots a_c & \vdots & \vdots \\ \multicolumn{4}{c}{\dotfill} \\ 0 \cdots 0 & 0 \cdots 0 & \cdots & a_1 \cdots a_c \end{pmatrix}$$

## Direct Product Notation

- $f_2(p) = A_2\big(f_1(p)\big)$, where

$$A_2 = I_t \otimes (a_1, \ldots, a_c)$$

- In general, the direct (Kronecker) product of a $p \times q$ matrix $A$ and a $m \times n$ matrix $B$ is given by

$$A \otimes B = \begin{pmatrix} a_{11}B & \ldots & a_{1q}B \\ \vdots & & \vdots \\ a_{p1}B & \ldots & a_{pq}B \end{pmatrix}$$

- $A \otimes B$ is a $pm \times qn$ matrix consisting of all possible products of an element of $A$ multiplied by an element of $B$

- This definition is sometimes referred to as the right direct product

# Choice of Scores

- Whenever possible, choose scores based on substantive considerations

- Different choices of scores lead to different assumed spacings between categories

- Types of scores include:

  1. Natural scores

     numeric value of the categorical variable

     midpoint or median of class interval

  2. Binary partition scores

     $a_j = 0$ for $j < d$, $a_j = 1$ for $j \geq d$

     dichotomizes the response variable

  3. Rank scores

  4. Integer scores ($a_j = j$)

## Model 2: Analysis of Mean Scores

- One choice of scores for the response categories 0, 1–2, 3+ is $a_1$=0, $a_2$=1.5, $a_3$=4
- The following SAS statements use the default model matrix and test for linear and nonlinear time effects using `contrast` statements

```
proc catmod; weight count;
response
1.5 4   0  0   0  0,
 0  0  1.5 4   0  0,
 0  0   0  0  1.5 4 *
0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0,
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1,
0 0 0 1 1 1 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 1 1 1 0 0 0,
0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 1 1 1,
0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0,
0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1;
model n0*n3*n6=_response_ / noprofile;
repeated time 3;
contrast 'Linear'    all_parms 0 2 1;
contrast 'Quadratic' all_parms 0 0 1;
```

## Linear and Quadratic Contrasts

- In terms of the default "sum-to-zero" parameterization, we have:

| | | Orthogonal Coefficients | |
|---|---|---|---|
| Year | Parameterization | Linear | Quadratic |
| 0 | $\mu + \alpha_1$ | $-1$ | 1 |
| 3 | $\mu + \alpha_2$ | 0 | $-2$ |
| 6 | $\mu - \alpha_1 - \alpha_2$ | 1 | 1 |

- The test of linearity is:

$$0 = -(\mu + \alpha_1) + 0(\mu + \alpha_2) + (\mu - \alpha_1 - \alpha_2)$$

$$= -2\alpha_1 - \alpha_2$$

- The test of nonlinearity is:

$$0 = (\mu + \alpha_1) - 2(\mu + \alpha_2) + (\mu - \alpha_1 - \alpha_2)$$

$$= -3\alpha_2$$

## Alternate Method for Model 2

- If the response variable is numeric, the statement

  ```
  response means;
  ```

  computes marginal mean scores

- First, create numeric variables:

  ```
  data b; set a;
  s0=1.5*(n0>' 0 ')+2.5*(n0='3+ ');
  s3=1.5*(n3>' 0 ')+2.5*(n3='3+ ');
  s6=1.5*(n6>' 0 ')+2.5*(n6='3+ ');
  ```

- Model 2 can be then be fit as follows:

  ```
  proc catmod; weight count;
  response means;
  model s0*s3*s6=_response_ / noprofile;
  repeated time 3;
  contrast 'Linear'    all_parms 0 2 1;
  contrast 'Quadratic' all_parms 0 0 1;
  ```

## Model 3: Reduced Mean Score Model

- The nonlinear time effect is nonsignificant ($W_{\text{nonlinearity}} = .42$, df $= 1$, $p = .51$)

- Thus, we may wish to fit the reduced model $\mu_x = \alpha + \beta x$, where $\mu_x$ is the marginal mean at year $x$

```
proc catmod; weight count;
response means;
model s0*s3*s6=(1 0,
                1 3,
                1 6)
      (1='Intercept',
       2='Linear Time')
      / noprofile p;
```

- The resulting model is $\widehat{\mu}_x = 2.629 + .0978x$

  The average number of friends is estimated to increase by .0978 per year

# Multi-Sample Problems

| | | Time Point | | | | |
|---|---|---|---|---|---|---|
| Group | Subject | 1 | $\ldots$ | $j$ | $\ldots$ | $t$ |
| 1 | 1 | $y_{111}$ | $\cdots$ | $y_{11j}$ | $\cdots$ | $y_{11t}$ |
| | $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ | $\ddots$ | $\vdots$ |
| | $i$ | $y_{1i1}$ | $\cdots$ | $y_{1ij}$ | $\cdots$ | $y_{1it}$ |
| | $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ | $\ddots$ | $\vdots$ |
| | $n_1$ | $y_{1n_1 1}$ | $\cdots$ | $y_{1n_1 j}$ | $\cdots$ | $y_{1n_1 t}$ |
| $h$ | 1 | $y_{h11}$ | $\cdots$ | $y_{h1j}$ | $\cdots$ | $y_{h1t}$ |
| | $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ | $\ddots$ | $\vdots$ |
| | $i$ | $y_{hi1}$ | $\cdots$ | $y_{hij}$ | $\cdots$ | $y_{hit}$ |
| | $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ | $\ddots$ | $\vdots$ |
| | $n_h$ | $y_{hn_h 1}$ | $\cdots$ | $y_{hn_h j}$ | $\cdots$ | $y_{hn_h t}$ |
| $s$ | 1 | $y_{s11}$ | $\cdots$ | $y_{s1j}$ | $\cdots$ | $y_{s1t}$ |
| | $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ | $\ddots$ | $\vdots$ |
| | $i$ | $y_{si1}$ | $\cdots$ | $y_{sij}$ | $\cdots$ | $y_{sit}$ |
| | $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ | $\ddots$ | $\vdots$ |
| | $n_s$ | $y_{sn_s 1}$ | $\cdots$ | $y_{sn_s j}$ | $\cdots$ | $y_{sn_s t}$ |

- $y_{hij}$ is categorical with $c$ possible outcomes

## Multinomial Structure

- The data correspond to $s$ independent multinomial samples of size $n_h$ $(h = 1, \ldots, s)$

- The probability vector $\pi_h$ for each of these independent multinomial distributions has $r = c^t$ potential outcomes (response profiles)

  If "missing" is a possible outcome, $r = (c+1)^t - 1$

- Let $p_h$ denote the $r \times 1$ vector of sample proportions in the $h$th subpopulation

- $p_h$ is an unbiased estimator of $\pi_h$

- The variance-covariance matrix of $p_h$ can be consistently estimated by

$$V_{p_h} = \frac{1}{n_h}(D_{p_h} - p_h p_h')$$

## Multinomial Structure

- Now let $p = (p_1, \ldots, p_s)'$ denote the $sr \times 1$ vector of observed proportions from all $s$ subpopulations

- $p$ is an unbiased estimator of $\pi = (\pi_1, \ldots, \pi_s)'$

- $\mathrm{Var}(p)$ can be consistently estimated by the block diagonal matrix

$$V_p = \begin{pmatrix} V_{p_1} & 0 & \cdots & 0 \\ 0 & V_{p_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & V_{p_s} \end{pmatrix}$$

- $V_p$ is a $sr \times sr$ matrix

- Since the elements of $\pi_h$ (and $p_h$) are linearly dependent, $V_p$ is a singular covariance matrix

## Model Fitting Using WLS

- Consider models of the form $f(p) = X\beta$
  - $f(p)$ is a $u \times 1$ vector of response functions

    (defined in terms of a sequence of linear, exponential, and logarithmic operations)
  - $X$ is a $u \times v$ model matrix of rank $v \leq u$
  - $\beta$ is a $v \times 1$ vector of unknown parameters
- Since $p \approx N_{sr}(\pi, V_p)$, $f(p) \approx N_u(f(\pi), V_f)$, where $V_f$ is the Taylor series estimate of $\mathrm{Var}(f(p))$
  - $V_f$ is assumed to have rank $u$
- The WLS estimate of $\beta$ is

$$\widehat{\beta} = (X'V_f^{-1}X)^{-1}(X'V_f^{-1}f(p))$$

## Model Fitting Using WLS

- The goodness of fit of the model can be tested using the statistic

$$W = \big(f(p) - X\widehat{\beta}\big)' V_f^{-1} \big(f(p) - X\widehat{\beta}\big)$$

  - If the model $X$ fits, then $W \approx \chi^2_{u-v}$

- Additional hypotheses of the form $H_0$: $C\beta = 0$ may then be tested

  - $C$ is a $c \times v$ coefficient matrix

  - Since $\widehat{\beta} \approx N_v\big(\beta, (X'V_f^{-1}X)^{-1}\big)$,
    $C\widehat{\beta} \approx N_c\big(C\beta, C(X'V_f^{-1}X)^{-1}C'\big)$

  - The Wald statistic

$$W_C = (C\widehat{\beta})'[C(X'V_f^{-1}X)^{-1}C']^{-1}C\widehat{\beta}$$

    is approximately $\chi^2_c$ if $H_0$ is true

**Example**

- The Iowa 65+ Rural Health Study

  sponsored by National Institute on Aging as part of the Established Populations for Epidemiologic Study of the Elderly (EPESE) project

- Cohort of elderly individuals was followed over a six-year period

- At each of three surveys (years 0, 3, 6), extensive demographic and social support data were obtained from each respondent

- One variable of interest was church attendance

| | |
|---|---|
| Yes: | subject is a regular church attender |
| No: | subject does not regularly attend |
| Missing: | subject did not answer this question or did not participate in this survey |

## Example (continued)

- Data were obtained from 3085 individuals
  - 1935 females, 1150 males

- There were a substantial number of missing values
  - most occur at the end of a sequence of nonmissing responses
  - due largely to deaths or losses to follow-up

- Questions to be answered:
  - Do church attendance rates differ between females and males?
  - Are attendance rates changing over time?
  - Are the observed patterns of change the same for females and males?

- Thus, interest focuses on modeling the marginal probability of regular church attendance as a function of gender and survey year.

## Church Attendance Data

| Regular Attender at: | | | Frequency | | |
|---|---|---|---|---|---|
| Year 0 | Year 3 | Year 6 | Female | Male | Total |
| Missing | No | Missing | 3 | 2 | 5 |
| Missing | No | No | 1 | 3 | 4 |
| Missing | No | Yes | 1 | 1 | 2 |
| Missing | Yes | Missing | 2 | 2 | 4 |
| Missing | Yes | Yes | 2 | 0 | 2 |
| No | Missing | Missing | 101 | 122 | 223 |
| No | Missing | No | 11 | 5 | 16 |
| No | Missing | Yes | 3 | 2 | 5 |
| No | No | Missing | 71 | 86 | 157 |
| No | No | No | 158 | 143 | 301 |
| No | No | Yes | 30 | 18 | 48 |
| No | Yes | Missing | 14 | 5 | 19 |
| No | Yes | No | 22 | 21 | 43 |
| No | Yes | Yes | 33 | 15 | 48 |
| Yes | Missing | Missing | 195 | 125 | 320 |
| Yes | Missing | No | 4 | 0 | 4 |
| Yes | Missing | Yes | 18 | 9 | 27 |
| Yes | No | Missing | 28 | 16 | 44 |
| Yes | No | No | 51 | 26 | 77 |
| Yes | No | Yes | 25 | 12 | 37 |
| Yes | Yes | Missing | 170 | 110 | 280 |
| Yes | Yes | No | 88 | 36 | 124 |
| Yes | Yes | Yes | 904 | 391 | 1295 |

## Complete Data

- We first analyze the data from 1973 individuals who responded at all three surveys:

| | Regular Church Attender | | | |
|---|---|---|---|---|
| Gender | Year 0 | Year 3 | Year 6 | Count |
| Females | No | No | No | 158 |
| | No | No | Yes | 30 |
| | No | Yes | No | 22 |
| | No | Yes | Yes | 33 |
| | Yes | No | No | 51 |
| | Yes | No | Yes | 25 |
| | Yes | Yes | No | 88 |
| | Yes | Yes | Yes | 904 |
| | | | Total | 1311 |
| Males | No | No | No | 143 |
| | No | No | Yes | 18 |
| | No | Yes | No | 21 |
| | No | Yes | Yes | 15 |
| | Yes | No | No | 26 |
| | Yes | No | Yes | 12 |
| | Yes | Yes | No | 36 |
| | Yes | Yes | Yes | 391 |
| | | | Total | 662 |

## Analysis of Marginal Proportions

- Let $p_h$ denote the underlying $8 \times 1$ proportion vector in subpopulation $h$
  ($h = 1$ for females, $h = 2$ for males)

- In each subpopulation, let $f(p_h)$ denote the $3 \times 1$ vector of marginal proportions:

$$f(p_h) = (p_{h0}, p_{h3}, p_{h6})',$$

  where $p_{hj}$ is the marginal proportion of subjects in group $h$ who regularly attend church at year $j$

- $f(p_h) = Ap_h$, where

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

## Analysis of Marginal Proportions

- In this example,

$$f(p_1) = \begin{pmatrix} 0.815 \\ 0.799 \\ 0.757 \end{pmatrix}, \qquad f(p_2) = \begin{pmatrix} 0.702 \\ 0.699 \\ 0.659 \end{pmatrix}$$

- Now let $f(p)$ be the $6 \times 1$ vector

$$\begin{pmatrix} f(p_1) \\ f(p_2) \end{pmatrix}$$

- The estimated covariance matrix of $f(p)$ is the $6 \times 6$ matrix

$$V_f = \begin{pmatrix} V_{f_1} & 0 \\ 0 & V_{f_2} \end{pmatrix},$$

  where $V_{f_h}$ is the estimated covariance matrix of $f(p_h)$

- We can now use weighted least squares to fit models of the form $f(p) = X\beta$

## Analysis Strategy

- We will first fit a saturated model with separate parameters for females and males

- The model is $f(p) = X_1\beta$, where $X_1$ is a $6 \times 6$ design matrix and

$$\beta = (\beta_{FI}, \beta_{FL}, \beta_{FQ}, \beta_{MI}, \beta_{ML}, \beta_{MQ})'$$

  - $\beta_{GI}$, $\beta_{GL}$, $\beta_{GQ}$ are the intercept, linear time effect, and quadratic time effect for gender $G$

  - Since the surveys were equally spaced, orthogonal polynomial coefficients will be used for the time effects

- Based on results of hypothesis tests concerning the parameters of the saturated model, we will then fit an appropriate reduced model

# SAS Statements for
# Analysis of Complete Data

```
data church;
* gender: F=female, M=male;
* attend: M=missing, N=no, Y=yes;
input gender $ attend0 $ attend3 $
      attend6 $ count;
cards;
F M N M   3
F M N N   1
F M N Y   1
F M Y M   2
F M Y Y   2
  ...
M Y N N  26
M Y N Y  12
M Y Y M 110
M Y Y N  36
M Y Y Y 391
;
data complete; set church;
if attend0='M' or attend3='M' or
   attend6='M' then delete;
title2 'Analysis of Complete Data';
```

# Model 1: Saturated

```
proc catmod; weight count; population gender;
response 0 0 0 0 1 1 1 1,
         0 0 1 1 0 0 1 1,
         0 1 0 1 0 1 0 1;
model attend0*attend3*attend6=
      (1 -1  1  0  0  0,
       1  0 -2  0  0  0,
       1  1  1  0  0  0,
       0  0  0  1 -1  1,
       0  0  0  1  0 -2,
       0  0  0  1  1  1)
      (1='Females: Intercept',
       2='        Linear',
       3='        Quadratic',
     2 3='        Lin & Quad',
       4='Males:   Intercept',
       5='        Linear',
       6='        Quadratic',
     5 6='        Lin & Quad',
     2 5='Both:    Linear',
     3 6='        Quadratic',
 2 3 5 6='        Lin & Quad') / noprofile;
contrast 'Sex Eq.' all_parms 1 0 0 -1  0  0,
                   all_parms 0 1 0  0 -1  0,
                   all_parms 0 0 1  0  0 -1;
contrast 'Int Eq.' all_parms 1 0 0 -1  0  0;
contrast 'Lin Eq.' all_parms 0 1 0  0 -1  0;
contrast 'Quad Eq' all_parms 0 0 1  0  0 -1;
```

## Results from Model 1

- Linear time effect is significant in females and in males

- Nonlinear time effect is nearly significant in females and in males

- The joint test of nonlinearity in females and males is nearly significant

- The intercepts for females and males differ significantly

- The linear time effects in females and males are not significantly different

- The nonlinear time effects in females and males are not significantly different

## Model 2: Reduced

- It seems sensible to fit a reduced model with common linear and nonlinear time effects for females and males

```
proc catmod; weight count;
population gender;
response 0 0 0 0 1 1 1 1,
         0 0 1 1 0 0 1 1,
         0 1 0 1 0 1 0 1;
model attend0*attend3*attend6=
      (1  0 -1  1,
       1  0  0 -2,
       1  0  1  1,
       0  1 -1  1,
       0  1  0 -2,
       0  1  1  1)
      (1='Intercept: Females',
       2='          Males',
       3='Linear Time',
       4='Quadratic Time')
      / noprofile p;
contrast 'Intercept Equality'
         all_parms 1 -1  0  0;
```

## Summary of Analysis of Complete Data

- Model 2 provides a good fit to the observed data
  - the residual chi-square is 0.87 with 2 df

- Each of the model parameters is significantly different from zero
  - intercept for females, intercept for males, linear time effect, quadratic time effect

- Conclusions:
  - probability of regular church attendance is decreasing over time
  - the change is nonlinear
    (decrease from year 0 to year 3 is less than the decrease from year 3 to year 6)
  - at each time point, females are more likely than males to regularly attend church
  - the estimated difference between females and males is $0.7905 - 0.6865 = 0.104$

## Reparameterization of Model 2

- In order to produce results that are more easily interpretable, Model 2 will be re-fit on the natural time scale (years)
  - (instead of using orthogonal polynomials)

```
model attend0*attend3*attend6=
      (1  0  0  0,
       1  0  3  9,
       1  0  6 36,
       0  1  0  0,
       0  1  3  9,
       0  1  6 36)
      (1='Intercept: Females',
       2='           Males',
       3='Linear Time',
       4='Quadratic Time')
      / noprofile p;
contrast 'Intercept Equality'
        all_parms 1 -1  0  0;
```

- All other statements are unchanged

## Comments

- Both parameterizations give the same:
  - lack of fit statistic
  - test of the quadratic time effect
  - predicted values
- The parameter estimates differ, as do the tests of all other effects
- In particular, the reparameterized test of the linear time effect is nonsignificant (due to correlation with the quadratic time effect)
- The time effect parameters are both negative when orthogonal polynomial coefficients are used
- On the natural time scale, the linear time parameter is positive

## Analysis of Marginal Logits

- Let $f^*(p_h)$ denote the $3 \times 1$ vector of marginal logits in group $h$

- $f^*(p_h) = A_2 \log(A_1 p_h)$, where

$$A_1 = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{pmatrix},$$

$$A_2 = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}$$

- The response statement is replaced with:

```
response 1 -1  0  0  0  0,
         0  0  1 -1  0  0,
         0  0  0  0  1 -1
     log 0  0  0  0  1  1  1  1,
         1  1  1  1  0  0  0  0,
         0  0  1  1  0  0  1  1,
         1  1  0  0  1  1  0  0,
         0  1  0  1  0  1  0  1,
         1  0  1  0  1  0  1  0;
```

## Summary of Analysis of Marginal Logits

- The model with
  - separate intercepts for females and males
  - common linear and quadratic time effects

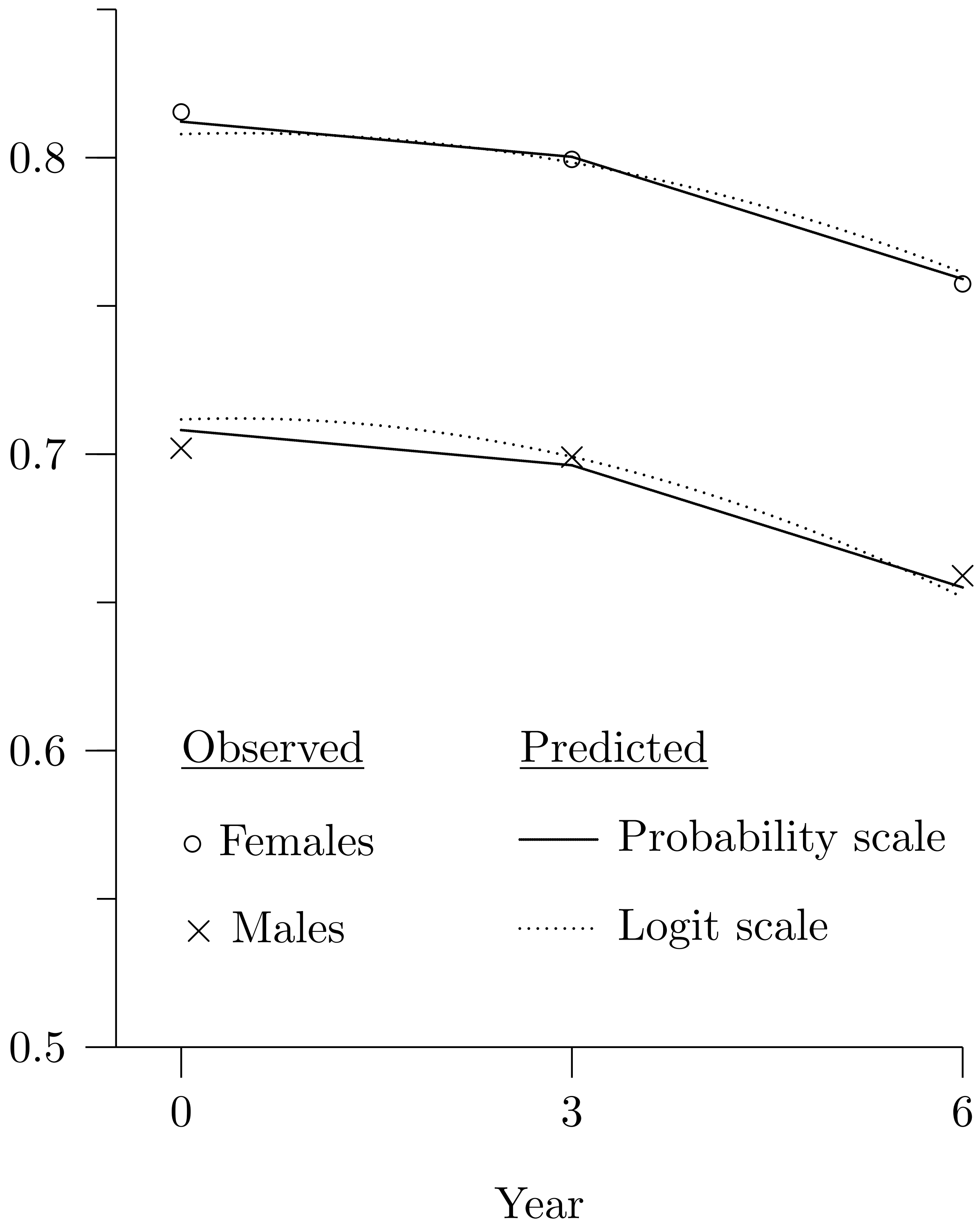
  provides an adequate fit ($p = 0.29$)

- Each of the model parameters is significantly different from zero

- At each time point, the odds of regularly attending church are estimated to be $e^{1.3241-0.7911} = e^{0.533} = 1.7$ times higher for females than for males

- The fit of the marginal logit model is not quite as good as the fit of the corresponding model on the marginal probability scale
  - logit scale: residual chi-square $= 2.47$
  - probability scale: residual chi-square $= 0.87$

**Observed and Predicted Probabilities of Church Attendance**



## Missing Data

- Data collected in longitudinal studies are often incomplete
- Generally, some of the individuals who are intended to be followed over time will fail to provide information at one or more of the scheduled follow-up times
- An observation may be missing:
  - by design
  - at random
  - due to characteristics of the subject
- Most of the standard methods of analysis require complete data
- In a longitudinal study, the analysis of complete cases can lead to a substantial reduction in sample size

## Ratio Estimation for Proportions

- Consider a one-sample repeated measures study with a categorical response
  - $n$ subjects
  - $t$ time points
  - $y_{ij}$ is a categorical response variable with $c$ possible outcomes, for $i = 1, \ldots, n$ and $j = 1, \ldots, t$

- If there are no missing data, there are $c^t$ potential response profiles

- If "missing" is also considered to be a category of response, there are $c + 1$ potential outcomes at each time point

- In this case, the number of response profiles is $(c + 1)^t - 1$

## Ratio Estimation for Proportions

- Let $\pi_{jl}$ denote the marginal probability of response category $l$ at time $j$, for $j = 1, \ldots, t$, $l = 1, \ldots, c$

- $\pi_{jl}$ can be estimated by $\widehat{\pi}_{jl}$, where

$$\widehat{\pi}_{jl} = \frac{\text{no. of subjects in category } l \text{ at time } j}{\text{no. of subjects with a response at time } j}$$

- The $tc \times 1$ vector

$$\widehat{\pi} = (\widehat{\pi}_{11}, \ldots, \widehat{\pi}_{1c}, \ldots, \widehat{\pi}_{t1}, \ldots, \widehat{\pi}_{tc})'$$

  can be calculated as

$$\widehat{\pi} = \exp\bigl(A_2 \log(A_1 p)\bigr)$$

- $p$ is the $((c{+}1)^t\text{-}1) \times 1$ vector of proportions corresponding to the multi-way cross-classification of response at the $t$ time points

# Ratio Estimation for Proportions

- $A_1$ is a $t(c+1) \times ((c+1)^t - 1)$ matrix

| Row | Proportion of subjects with: |
|---|---|
| 1 | response category 1 at time 1 |
| 2 | response category 2 at time 1 |
| $\vdots$ | $\vdots$ |
| $c$ | response category $c$ at time 1 |
| $c+1$ | non-missing response at time 1 |
| ⋯ | ⋯ |
| $(t\text{-}1)(c{+}1){+}1$ | response category 1 at time $t$ |
| $(t\text{-}1)(c{+}1){+}2$ | response category 2 at time $t$ |
| $\vdots$ | $\vdots$ |
| $(t\text{-}1)(c{+}1){+}c$ | response category $c$ at time $t$ |
| $(t\text{-}1)(c{+}1){+}c{+}1$ $[= t(c+1)]$ | non-missing response at time $t$ |

## Ratio Estimation for Proportions

- $A_2$ is the $tc \times t(c+1)$ matrix $I_t \otimes [I_c, -1_c]$, where $I_k$ is the $k \times k$ identity matrix and $1_k$ is the $k \times 1$ vector $(1, \ldots, 1)'$

- Since the elements of $\widehat{\pi}$ are linearly dependent, additional transformations can then be used to compute:
  - marginal proportions
  - marginal cumulative proportions
  - marginal mean scores
  - marginal logits

- In practice, the matrices $A_1$ and $A_2$ can often be simplified (to compute only the specific marginal proportions of interest)

## Example

- The Muscatine Coronary Risk Factor Study

  A longitudinal study of coronary risk factors in school children

- From 1971–1981, six biennial cross-sectional school screens were completed

  Data from 1977, 1979, and 1981 were reported

- Only children currently enrolled in school were eligible to participate, and about 70% of eligible children were screened

- Height and weight were measured on each participating child, from which relative weight was computed

  (ratio of child's weight to the median weight in the sex-age-height group)

## Example

- The outcome of interest was dichotomous (obese, not obese)

  Children with relative weight greater than 110% of the median weight were classified as obese

- In this example, we consider the cohort of males who were 7–9 years old in 1977

- This group consists of 522 children

  Only 225 children participated in all three surveys (356 in 1977, 375 in 1979, 380 in 1981)

## Reference

Woolson, R. F. and Clarke, W. R. (1984). Analysis of categorical incomplete longitudinal data. *J. Roy. Statist. Soc A* **147**, 87–99.

**1977 Cohort of 7–9 Year Old Males**

| Classified as Obese in: | | | No. of |
| --- | --- | --- | --- |
| 1977 | 1979 | 1981 | Children |
| Yes | Yes | Yes | 20 |
| Yes | Yes | No | 7 |
| Yes | Yes | Unk | 11 |
| Yes | No | Yes | 9 |
| Yes | No | No | 8 |
| Yes | No | Unk | 1 |
| Yes | Unk | Yes | 3 |
| Yes | Unk | No | 1 |
| Yes | Unk | Unk | 7 |
| No | Yes | Yes | 8 |
| No | Yes | No | 8 |
| No | Yes | Unk | 3 |
| No | No | Yes | 15 |
| No | No | No | 150 |
| No | No | Unk | 38 |
| No | Unk | Yes | 6 |
| No | Unk | No | 16 |
| No | Unk | Unk | 45 |
| Unk | Yes | Yes | 13 |
| Unk | Yes | No | 3 |
| Unk | Yes | Unk | 4 |
| Unk | No | Yes | 2 |
| Unk | No | No | 42 |
| Unk | No | Unk | 33 |
| Unk | Unk | Yes | 14 |
| Unk | Unk | No | 55 |

# Definition of Response Functions

- Let $\pi_x$ denote the marginal probability of being classified as obese at year $x$, for $x = 77, 79, 81$

- $\pi_x$ is estimated by $p_x$, the observed proportion classified as obese

- The response functions $f(p) = (p_{77}, p_{79}, p_{81})'$ can be computed as $f(p) = \exp\big(A_2 \log(A_1 p)\big)$, where $p$ is the $26 \times 1$ vector of multinomial proportions, $A_1$ is the $6 \times 26$ matrix

$$\begin{pmatrix} 1&1&1&1&1&1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&0&0&0&0&0&0&0&0 \\ 1&1&1&0&0&0&0&0&0&1&1&1&0&0&0&0&0&0&1&1&1&0&0&0&0&0 \\ 1&1&1&1&1&1&0&0&0&1&1&1&1&1&1&0&0&0&1&1&1&1&1&1&0&0 \\ 1&0&0&1&0&0&1&0&0&1&0&0&1&0&0&1&0&0&1&0&0&1&0&0&1&0 \\ 1&1&0&1&1&0&1&1&0&1&1&0&1&1&0&1&1&0&1&1&0&1&1&0&1&1 \end{pmatrix},$$

and $A_2 = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}$

## SAS Statements

```
data a;
* 1=obese, 2=not obese, 3=missing;
input o77 o79 o81 count;
cards;
1 1 1  20
  ...
3 3 2  55
;
proc catmod data=a; weight count;
response exp 1 -1  0  0  0  0,
             0  0  1 -1  0  0,
             0  0  0  0  1 -1
log
1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0,
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0,
1 1 1 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0,
1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 0 0 0,
1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0,
1 1 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 1;
model o77*o79*o81=(1 0,
                   1 2,
                   1 4)
      (1='Intercept',
       2='Linear Age')
      / noprofile p;
```

## Comments on the Example

- It is wise to check the components of the response function vector

  (`PROC FREQ` is often useful in this regard)

- The model provides a very good fit to the observed data

  $W = 0.15$, $p = 0.70$

  (The 1 df residual tests non-linearity)

- The predicted model is

$$\widehat{\pi}_x = 0.186 + 0.0120(x - 77)$$

  (The marginal probability of obesity is estimated to increase by 0.0120 per year)

*How would the results differ if the analysis had been carried out using only the data from the 225 children who participated in all three surveys?*

## Analysis of Complete Data

- Model with linear age effect

```
data b; set a;
if o77 ne 3 & o79 ne 3 & o81 ne 3;
proc catmod; weight count;
response marginals;
model o77*o79*o81=(1 0,
                   1 2,
                   1 4)
      (1='Intercept',
       2='Linear Age') / noprofile;
```

The linear effect of age is not significant

- Model with only an intercept

```
proc catmod; weight count;
response marginals;
model o77*o79*o81=(1,
                   1,
                   1)
      (1='Intercept') / noprofile;
```

- This model fits well to the complete data

## Church Attendance Example
## Analysis of All Data

- Excluding subjects with incomplete data results in a substantially reduced sample size

- To use all available data at each time point, the marginal proportions are estimated as ratios of sums of underlying multinomial proportions

- These are computed using a series of linear, logarithmic, and exponential transformations

- This approach (and extensions) has been discussed by:
  - Stanish, Gillings & Koch (1978, *Biometrics*)
  - Woolson and Clarke (1984, *JRSS A*)
  - Landis et al. (1988, *Statistics in Medicine*)
  - Park and Davis (1993, *Biometrics*)

## Analysis of All Data

- Let $p_1' = (p_{f0}, p_{f3}, p_{f6})$ and $p_2' = (p_{m0}, p_{m3}, p_{m6})$
- In general, $p_i$ is computed as $\exp\big(A_2 \log(A_1 p_i^*)\big)$, where
  - $p_i^*$ is the vector of underlying multinomial proportions in each subpopulation
  - $A_1$ has $ct$ rows and as many columns as there are observed response profiles (a maximum of $(c+1)^t - 1$ columns)
  - $A_2$ is a $(c-1)t \times ct$ matrix
  - $c$ is the number of possible outcomes of the response (excluding the “missing” category)
  - $t$ is the number of time points
- In this example, $c = 2$, $t = 3$, $A_1$ is $6 \times 23$ and $A_2$ is $3 \times 6$

## Analysis of All Data

- It is convenient to define the rows of $A_1$ as:

| Row | Proportion of subjects with: |
|---|---|
| 1 | response category 1 at time 1 |
| $\vdots$ | $\vdots$ |
| $c-1$ | response category $c-1$ at time 1 |
| $c$ | non-missing response at time 1 |
| ... | ... |
| $c(t–1)+1$ | response category 1 at time $t$ |
| $\vdots$ | $\vdots$ |
| $c(t–1)+(c–1)$ | response category $c-1$ at time $t$ |
| $c(t–1)+c$ | non-missing response at time $t$ |

- With this defn. of $A_1$, $A_2 = I_t \otimes [I_{(c-1)}, -e_{(c-1)}]$
  - $I_k$ is the $k \times k$ identity matrix
  - $e_k$ is the $k \times 1$ vector $(1, \ldots, 1)'$
  - $\otimes$ denotes the Kronecker product

# Analysis of All Data

- In this example,

$$p_1' = (p_{f0}, p_{f3}, p_{f6}), \qquad p_2' = (p_{m0}, p_{m3}, p_{m6})$$

$$A_1 = \begin{pmatrix} 0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1&1&1&1&1&1 \\ 0&0&0&0&0&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1 \\ 0&0&0&1&1&0&0&0&0&0&0&1&1&1&0&0&0&0&0&0&1&1&1 \\ 1&1&1&1&1&0&0&0&1&1&1&1&1&1&0&0&0&1&1&1&1&1&1 \\ 0&0&1&0&1&0&0&1&0&0&1&0&0&1&0&0&1&0&0&1&0&0&1 \\ 0&1&1&0&1&0&1&1&0&1&1&0&1&1&0&1&1&0&1&1&0&1&1 \end{pmatrix}$$

$$A_2 = I_3 \otimes (1, -1) = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}$$

- Rows 1, 3, and 5 of $A_1$ compute the proportion of individuals who regularly attended church at years 0, 3, and 6, respectively

- Rows 2, 4, and 6 calculate the corresponding proportions who responded to this question at the three surveys

## Model 1: Saturated

```
proc catmod data=church; weight count;
population gender;
response exp 1 -1  0  0  0  0,
             0  0  1 -1  0  0,
             0  0  0  0  1 -1  log
0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1,
0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1,
0 0 0 1 1 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 1 1 1,
1 1 1 1 1 0 0 0 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1,
0 0 1 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1,
0 1 1 0 1 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1;
model attend0*attend3*attend6=
      (1 -1  1  0  0  0,
       1  0 -2  0  0  0,
       1  1  1  0  0  0,
       0  0  0  1 -1  1,
       0  0  0  1  0 -2,
       0  0  0  1  1  1)
      (1='Females: Intercept',
       2='         Linear',
       3='         Quadratic',
     2 3='         Lin & Quad',
       4='Males:   Intercept',
       5='         Linear',
       6='         Quadratic',
     5 6='         Lin & Quad',
     2 5='Both:    Linear',
     3 6='         Quadratic',
 2 3 5 6='         Lin & Quad') / noprofile;
```

## Model 2: Reduced

- In model 1, the time effects are not significantly different from zero
- This suggests a model with separate intercepts for females and males

```
proc catmod data=church; weight count;
population gender;
response exp 1 -1  0  0  0  0,
             0  0  1 -1  0  0,
             0  0  0  0  1 -1  log
0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1,
0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1,
0 0 0 1 1 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 1 1 1,
1 1 1 1 1 0 0 0 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1,
0 0 1 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1,
0 1 1 0 1 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1;
model attend0*attend3*attend6=(1  0,
                               1  0,
                               1  0,
                               0  1,
                               0  1,
                               0  1)
    (1='Females: Intercept',
     2='Males:   Intercept') / noprofile;
contrast 'Int. Equality' all_parms 1 -1;
```

## Summary of Analysis of All Data

- Model 2 provides a good fit to the observed data
  - residual chi-square is 6.0 with 4 df ($p = 0.2$)
- The two parameters of the model are both statistically significant
- Conclusions:
  - Females are more likely than males to regularly attend church
  - the estimated difference between females and males is $0.7660 - 0.6434 = 0.1226$
  - The estimated probability of regular church attendance does not change over time
- Estimated sex difference is similar to the estimate from the analysis of the complete data
- The conclusions concerning time trends differ

## Alternative Response Functions

- This methodology can be extended to more complicated response functions
  e.g., generalized logits or cumulative logits

- Simplest approach is to first compute the marginal proportions

- In this case, usually necessary to calculate all $c$ proportions
  (Instead of computing only the first $c-1$ linearly independent marginal proportions)

- Thus, $A_1$ has $(c+1)t$ rows and $A_2$ is the $ct \times (c+1)t$ matrix $I_t \otimes [I_c, -e_c]$

- Additional matrix, exponential, and logarithmic operations can then be applied to obtain the response functions of interest

## Logit Response Functions

- Suppose we prefer to analyze the data on the logit scale rather than on the probability scale

- Consider models $l = X\beta$, where $l = (l_1, l_2)'$ and

$$l_1 = \left(\log\left(\frac{p_{f0}}{1-p_{f0}}\right), \log\left(\frac{p_{f3}}{1-p_{f3}}\right), \log\left(\frac{p_{f6}}{1-p_{f6}}\right)\right)$$

$$l_2 = \left(\log\left(\frac{p_{m0}}{1-p_{m0}}\right), \log\left(\frac{p_{m3}}{1-p_{m3}}\right), \log\left(\frac{p_{m6}}{1-p_{m6}}\right)\right)$$

- $l_i' = B_2 \log(B_1 p_i^*)$, where $B_1$ is

$$\begin{pmatrix}
0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1&1&1&1&1&1\\
0&0&0&0&0&1&1&1&1&1&1&1&1&1&0&0&0&0&0&0&0&0&0\\
0&0&0&1&1&0&0&0&0&0&0&1&1&1&0&0&0&0&0&0&1&1&1\\
1&1&1&0&0&0&0&0&1&1&1&0&0&0&0&0&0&1&1&1&0&0&0\\
0&0&1&0&1&0&0&1&0&0&1&0&0&1&0&0&1&0&0&1&0&0&1\\
0&1&0&0&0&0&1&0&0&1&0&0&1&0&0&1&0&0&1&0&0&1&0
\end{pmatrix}$$

and $B_2 = I_3 \otimes (1, -1) = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}$

## A Test of the Missing Data Mechanism

- Fit a single model to complete & incomplete data

- Test if parameter estimates for complete data are significantly different (individually & jointly) from parameter estimates for incomplete data

- Proposed by Park and Davis (1993, *Biometrics*)

- The $12 \times 1$ vector of response functions is now $p = (p_1, p_2)'$, where

$$p_1 = \left(p_{f0c}, p_{f3c}, p_{f6c}, p_{f0i}, p_{f3i}, p_{f6i}\right)$$

$$p_2 = \left(p_{m0c}, p_{m3c}, p_{m6c}, p_{m0i}, p_{m3i}, p_{m6i}\right)$$

- The third subscript is "$c$" for subjects with complete data and "$i$" for subjects with one or more missing responses

- $p_i = \exp\left(A_2 \log(A_1 p_i^*)\right)$

## A Test of the Missing Data Mechanism

$$A_1 = \begin{pmatrix}
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&0&1&1\\
0&0&0&0&0&0&0&0&0&0&0&0&1&1&0&0&0&0&0&0&0&1&1\\
0&0&0&0&0&0&0&0&0&0&1&0&0&1&0&0&0&0&0&1&0&0&1\\
0&0&0&0&0&0&0&0&0&1&1&0&1&1&0&0&0&0&1&1&0&1&1\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1&0&0&1&0&0\\
0&0&0&0&0&1&1&1&1&0&0&1&0&0&1&1&1&1&0&0&1&0&0\\
0&0&0&1&1&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&1&0&0\\
1&1&1&1&1&0&0&0&1&0&0&1&0&0&0&0&0&1&0&0&1&0&0\\
0&0&1&0&1&0&0&1&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0\\
0&1&1&0&1&0&1&1&0&0&0&0&0&0&0&1&1&0&0&0&0&0&0
\end{pmatrix}$$

- Subjects who responded to all three surveys:
  - rows 1–3 compute proportion who attended regularly at years 0, 3, and 6, respectively
  - row 4 computes the proportion who responded at all three surveys
- Subjects who had at least one missing response:
  - rows 5–6 compute numerator (attend regularly) and denominator (responded) for year 0
  - rows 7–8 and 9–10 calculate the corresponding quantities for year 3 and year 6

## A Test of the Missing Data Mechanism

$$A_2 = \begin{pmatrix} 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}$$

- Rows 1–3 pertain to subjects who responded to all three surveys
  - difference between the log of the proportion who attended church regularly at year $i$ and the log of the proportion who responded to all three surveys, for $i = 0, 3, 6$, respectively

- Rows 4–6 pertain to subjects who had at least one missing response
  - difference between the log of the proportion who attended church regularly at year $i$ and the log of the proportion who responded at year $i$, for $i = 0, 3, 6$, respectively

## Model 1: Saturated

- Procedure invocation, defn. of response functions:

```
proc catmod data=church;
weight count;
population gender;
response exp
   1  0  0 -1  0  0  0  0  0  0,
   0  1  0 -1  0  0  0  0  0  0,
   0  0  1 -1  0  0  0  0  0  0,
   0  0  0  0  1 -1  0  0  0  0,
   0  0  0  0  0  0  1 -1  0  0,
   0  0  0  0  0  0  0  0  1 -1
  log
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 1 1,
0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 1 1,
0 0 0 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 1 0 0 1,
0 0 0 0 0 0 0 0 0 1 1 0 1 1 0 0 0 0 1 1 0 1 1,
0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 0 0 1 0 0,
0 0 0 0 0 1 1 1 1 0 0 1 0 0 1 1 1 1 0 0 1 0 0,
0 0 0 1 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0,
1 1 1 1 1 0 0 0 1 0 0 1 0 0 0 0 0 1 0 0 1 0 0,
0 0 1 0 1 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0,
0 1 1 0 1 0 1 1 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0;
```

## Model 1: Saturated (continued)

- MODEL statement:

```
model attend0*attend3*attend6=
   (1 -1  1  0  0  0  0  0  0  0  0  0,
    1  0 -2  0  0  0  0  0  0  0  0  0,
    1  1  1  0  0  0  0  0  0  0  0  0,
    0  0  0  0  0  0  1 -1  1  0  0  0,
    0  0  0  0  0  0  1  0 -2  0  0  0,
    0  0  0  0  0  0  1  1  1  0  0  0,
    0  0  0  1 -1  1  0  0  0  0  0  0,
    0  0  0  1  0 -2  0  0  0  0  0  0,
    0  0  0  1  1  1  0  0  0  0  0  0,
    0  0  0  0  0  0  0  0  0  1 -1  1,
    0  0  0  0  0  0  0  0  0  1  0 -2,
    0  0  0  0  0  0  0  0  0  1  1  1)
   (1='C:  F-Int.',
    2='      Lin.',
    3='      Quad',
    4='    M-Int.',
    5='      Lin.',
    6='      Quad',
    7='IC: F-Int.',
    8='      Lin.',
    9='      Quad',
   10='    M-Int.',
   11='      Lin.',
   12='      Quad') / noprofile;
```

## Model 1: Saturated (continued)

- CONTRAST statements for testing equality of parameters for subjects with complete and incomplete data

```
contrast 'C=IC'
all_parms 1  0  0  0  0  0 -1  0  0  0  0  0,
all_parms 0  1  0  0  0  0  0 -1  0  0  0  0,
all_parms 0  0  1  0  0  0  0  0 -1  0  0  0,
all_parms 0  0  0  1  0  0  0  0  0 -1  0  0,
all_parms 0  0  0  0  1  0  0  0  0  0 -1  0,
all_parms 0  0  0  0  0  1  0  0  0  0  0 -1;
contrast 'C=IC: Int.'
all_parms 1  0  0  0  0  0 -1  0  0  0  0  0,
all_parms 0  0  0  1  0  0  0  0  0 -1  0  0;
contrast 'C=IC: L & Q'
all_parms 0  1  0  0  0  0  0 -1  0  0  0  0,
all_parms 0  0  1  0  0  0  0  0 -1  0  0  0,
all_parms 0  0  0  0  1  0  0  0  0  0 -1  0,
all_parms 0  0  0  0  0  1  0  0  0  0  0 -1;
```

## Results from Model 1

- Highly significant difference between complete and incomplete cases
  - chi-square= 82.7, df= 6, $p < 0.001$
- Time effects for complete and incomplete cases are not significantly different
  - chi-square= 1.58, df= 4, $p = 0.8$
- This motivates fitting a reduced model with:
  - separate intercepts for complete and incomplete females and males (4 parameters)
  - common linear and quadratic time effects for females (2 parameters) and for males (2 parameters)
- The PROC, WEIGHT, POPULATION, and RESPONSE statements are identical to those from Model 1

## Model 2: Reduced

- MODEL statement:

```
model attend0*attend3*attend6=
        (1  0 -1  1  0  0  0  0,
         1  0  0 -2  0  0  0  0,
         1  0  1  1  0  0  0  0,
         0  1 -1  1  0  0  0  0,
         0  1  0 -2  0  0  0  0,
         0  1  1  1  0  0  0  0,
         0  0  0  0  1  0 -1  1,
         0  0  0  0  1  0  0 -2,
         0  0  0  0  1  0  1  1,
         0  0  0  0  0  1 -1  1,
         0  0  0  0  0  1  0 -2,
         0  0  0  0  0  1  1  1)
        (1='F: C Int.',
         2='   IC Int.',
         3='   Linear',
         4='   Quadratic',
         5='M: C Int.',
         6='   IC Int.',
         7='   Linear',
         8='   Quadratic') / noprofile;
```

## Model 2: Reduced (continued)

- CONTRAST statements:

```
contrast 'C=IC: Int.  '
   all_parms 1 -1  0  0  0  0  0  0,
   all_parms 0  0  0  0  1 -1  0  0;
contrast '      F Int.'
   all_parms 1 -1  0  0  0  0  0  0;
contrast '      M Int.'
   all_parms 0  0  0  0  1 -1  0  0;
contrast 'M=F:  L & Q'
   all_parms 0  0  1  0  0  0 -1  0,
   all_parms 0  0  0  1  0  0  0 -1;
contrast '      Lin.'
   all_parms 0  0  1  0  0  0 -1  0;
contrast '      Quad.'
   all_parms 0  0  0  1  0  0  0 -1;
contrast 'Int.: M=F'
   all_parms 1  0  0  0 -1  0  0  0,
   all_parms 0  1  0  0  0 -1  0  0;
contrast 'C  Int.: M=F'
   all_parms 1  0  0  0 -1  0  0  0;
contrast 'IC Int.: M=F'
   all_parms 0  1  0  0  0 -1  0  0;
```

## Results from Model 2

- Provides a good fit to the data
  - residual chi-square is 1.58 with 4 df
- All tests comparing intercepts are statistically significant
  - complete versus incomplete cases (joint, in females, in males)
  - males versus females (joint, in complete cases, in incomplete cases)
- The time effects in females and males are not significantly different
- Motivates a further reduced model with:
  - four intercepts (complete and incomplete females and males)
  - common linear and quadratic time effects (2 parameters)

## Model 3: Further Reduced

- MODEL statement:

```
model attend0*attend3*attend6=
      (1  0  0  0 -1  1,
       1  0  0  0  0 -2,
       1  0  0  0  1  1,
       0  1  0  0 -1  1,
       0  1  0  0  0 -2,
       0  1  0  0  1  1,
       0  0  1  0 -1  1,
       0  0  1  0  0 -2,
       0  0  1  0  1  1,
       0  0  0  1 -1  1,
       0  0  0  1  0 -2,
       0  0  0  1  1  1)
      (1='F: C Int.',
       2='   IC Int.',
       3='M: C Int.',
       4='   IC Int.',
       5='Linear Time',
       6='Quadratic Time')
          / noprofile p;
```

## Model 3: Further Reduced

- CONTRAST statements:

```
contrast 'Int.: Equal.'
   all_parms 1  0  0 -1  0  0,
   all_parms 0  1  0 -1  0  0,
   all_parms 0  0  1 -1  0  0;
contrast '      F=M'
   all_parms 1  0 -1  0  0  0,
   all_parms 0  1  0 -1  0  0;
contrast '      F=M:C'
   all_parms 1  0 -1  0  0  0;
contrast '          IC'
   all_parms 0  1  0 -1  0  0;
contrast '      C=IC'
   all_parms 1 -1  0  0  0  0,
   all_parms 0  0  1 -1  0  0;
contrast '      C=IC:F'
   all_parms 1 -1  0  0  0  0;
contrast '           M'
   all_parms 0  0  1 -1  0  0;
contrast '    C F=IC M'
   all_parms 1  0  0 -1  0  0;
contrast '    IC F=C M'
   all_parms 0  1 -1  0  0  0;
```

## Results of Model 3

- Provides a good fit to the data
  - residual chi-square is 2.58 with 6 df
- Linear and nonlinear time effects are both statistically significant
- The following tests comparing intercepts are statistically significant
  - joint equality (3 df)
  - females versus males
    (joint, in complete cases, in incomplete cases)
  - complete versus incomplete
    (joint, in females, in males)
- The only nonsignificant intercept comparison is incomplete females versus complete males

## Model 4 (Final)

- In order to produce results that are more easily interpretable, Model 3 will be re-fit on the natural time scale (years)
  - (instead of using orthogonal polynomials)

```
model attend0*attend3*attend6=
        (1  0  0  0  0  0,
         1  0  0  0  3  9,
         1  0  0  0  6 36,
         0  1  0  0  0  0,
         0  1  0  0  3  9,
         0  1  0  0  6 36,
         0  0  1  0  0  0,
         0  0  1  0  3  9,
         0  0  1  0  6 36,
         0  0  0  1  0  0,
         0  0  0  1  3  9,
         0  0  0  1  6 36)
```

- All other statements are unchanged

## Results of Model 4 (Final)

- The estimated intercepts are:

| Subpopulation | Estimate |
|---|---|
| Complete females | 0.8128 |
| Incomplete females | 0.6710 |
| Complete males | 0.7089 |
| Incomplete males | 0.5426 |

- The profiles over time are parallel across the four groups

- The estimated probability of regular attendance:
  - decreases nonlinearly over time
  - the decrease from year 0 to year 3 is greater than the decrease from year 3 to year 6
  - is highest for complete females, followed by complete males, incomplete females, incomplete males

## Summary of Approaches

*Analysis of Complete Data Only*

| Parameter | Estimate |
|---|---|
| Female intercept | 0.8122 |
| Male intercept | 0.7081 |
| Linear year | 0.0009 |
| Quadratic year | −0.0016 |

*Analysis of All Data*

| Parameter | Estimate |
|---|---|
| Female intercept | 0.7660 |
| Male intercept | 0.6434 |

*Combined Analysis of Complete & Incomplete Data*

| Parameter | Estimate |
|---|---|
| Complete female intercept | 0.8128 |
| Incomplete female intercept | 0.6710 |
| Complete male intercept | 0.7089 |
| Incomplete male intercept | 0.5426 |
| Linear year | −0.0001 |
| Quadratic year | −0.0015 |

## Comments on the WLS Approach

- A flexible methodology that can handle a wide variety of response functions and accommodate missing data

  (test of missing data mechanism is also possible)

- The main disadvantages are that the:
  - Sample size must be large
  - Number of time points must be small
  - Covariates must be categorical and time-independent

- Some key references:
  - Koch *et al.* (1977, *Biometrics*)
  - Stanish *et al.* (1978, *Biometrics*)
  - Woolson and Clarke (1984, *JRSS A*)
  - Landis *et al.* (1988, *Statistics in Medicine*)
  - Park and Davis (1993, *Biometrics*)