1 gamma.mixed: Mixed effects gamma regression

Use generalized multi-level linear regression if you have covariates that are grouped according to one or more classification factors. Gamma regression models a continuous, positive dependent variable.

While generally called multi-level models in the social sciences, this class of models is often referred to as mixed-effects models in the statistics literature and as hierarchical models in a Bayesian setting. This general class of models consists of linear models that are expressed as a function of both *fixed effects*, parameters corresponding to an entire population or certain repeatable levels of experimental factors, and *random effects*, parameters corresponding to individual experimental units drawn at random from a population.

1.0.1 Syntax

```
z.out <- zelig(formula= y ~ x1 + x2 + tag(z1 + z2 | g),
               data=mydata, model="gamma.mixed")

z.out <- zelig(formula= list(mu=y ~ xl + x2 + tag(z1, delta | g),
               delta= ~ tag(w1 + w2 | g)), data=mydata, model="gamma.mixed")
```

1.0.2 Inputs

`zelig()` takes the following arguments for `mixed`:

- `formula:` a two-sided linear formula object describing the systematic component of the model, with the response on the left of a ~ operator and the fixed effects terms, separated by `+` operators, on the right. Any random effects terms are included with the notation `tag(z1 + ... + zn | g)` with `z1 + ... + zn` specifying the model for the random effects and `g` the grouping structure. Random intercept terms are included with the notation `tag(1 | g)`.
  Alternatively, `formula` may be a list where the first entry, `mu`, is a two-sided linear formula object describing the systematic component of the model, with the repsonse on the left of a ~ operator and the fixed effects terms, separated by `+` operators, on the right. Any random effects terms are included with the notation `tag(z1, delta | g)` with `z1` specifying the individual level model for the random effects, `g` the grouping structure and `delta` references the second equation in the list. The `delta` equation is one-sided linear formula object with the group level model for the random effects on the right side of a ~ operator. The model is specified with the notation `tag(w1 + ... + wn | g)` with `w1 + ... + wn` specifying the group level model and `g` the grouping structure.

1.0.3 Additional Inputs

In addition, `zelig()` accepts the following additional arguments for model specification:

- `data:` An optional data frame containing the variables named in `formula`. By default, the variables are taken from the environment from which `zelig()` is called.
- `method:` a character string. The criterion is always the log-likelihood but this criterion does not have a closed form expression and must be approximated. The default approximation is `"PQL"` or penalized quasi-likelihood. Alternatives are `"Laplace"` or `"AGQ"` indicating the Laplacian and adaptive Gaussian quadrature approximations respectively.
- `na.action:` A function that indicates what should happen when the data contain `NA`s. The default action (`na.fail`) causes `zelig()` to print an error message and terminate if there are any incomplete observations.

Additionally, users may with to refer to `lmer` in the package `lme4` for more information, including control parameters for the estimation algorithm and their defaults.

1.0.4 Examples

1. Basic Example with First Differences

   Attach sample data:

   ```
   > data(coalition2)
   ```

   Estimate model using optional arguments to specify approximation method for the log-likelihood, and the log link function for the Gamma family:

   ```
   > z.out1 <- zelig(duration ~ invest + fract + polar + numst2 + crisis + tag(1 | country
   ```

   Summarize regression coefficients and estimated variance of random effects:

   ```
   > summary(z.out1)
   ```

   Set the baseline values (with the ruling coalition in the minority) and the alternative values (with the ruling coalition in the majority) for X:

   ```
   > x.high <- setx(z.out1, numst2 = 1)
   > x.low <- setx(z.out1, numst2 = 0)
   ```

Simulate expected values (`qi$ev`) and first differences(`qi$fd`):

```
> s.out1 <- sim(z.out1, x=x.high, x1=x.low)
> summary(s.out1)
```

1.0.5 Mixed effects gamma regression Model

Let Y_{ij} be the continuous, positive dependent variable, realized for observation j in group i as y_{ij}, for $i = 1, \ldots, M$, $j = 1, \ldots, n_i$.

- The *stochastic component* is described by a Gamma model with scale parameter α.

$$Y_{ij} \sim \text{Gamma}(y_{ij}|\lambda_{ij}, \alpha)$$

  where

$$Gamma(y_{ij}|\lambda_{ij}, \alpha) = \frac{1}{\alpha^{\lambda_{ij}}\Gamma\lambda_{ij}} y_{ij}^{\lambda_{ij}-1} \exp(-\{\frac{y_{ij}}{\alpha}\})$$

  for $\alpha,\ \lambda_{ij},\ y_{ij}\ > 0$.

- The q-dimensional vector of *random effects*, b_i, is restricted to be mean zero, and therefore is completely characterized by the variance covarance matrix Ψ, a $(q \times q)$ symmetric positive semi-definite matrix.

$$b_i \sim Normal(0, \Psi)$$

- The *systematic component* is

$$\lambda_{ij} \equiv \frac{1}{X_{ij}\beta + Z_{ij}b_i}$$

  where X_{ij} is the $(n_i \times p \times M)$ array of known fixed effects explanatory variables, β is the p-dimensional vector of fixed effects coefficients, Z_{ij} is the $(n_i \times q \times M)$ array of known random effects explanatory variables and b_i is the q-dimensional vector of random effects.

1.0.6 Quantities of Interest

- The predicted values (`qi$pr`) are draws from the gamma distribution for each given set of parameters (α, λ_{ij}), for

$$\lambda_{ij} = \frac{1}{X_{ij}\beta + Z_{ij}b_i}$$

  given X_{ij} and Z_{ij} and simulations of of β and b_i from their posterior distributions. The estimated variance covariance matrices are taken as correct and are themselves not simulated.

- The expected values (qi$ev) are simulations of the mean of the stochastic component given draws of α, β from their posteriors:

$$E(Y_{ij}|X_{ij}) = \alpha\lambda_{ij} = \frac{\alpha}{X_{ij}\beta}.$$

- The first difference (qi$fd) is given by the difference in expected values, conditional on X_{ij} and X'_{ij}, representing different values of the explanatory variables.

$$FD(Y_{ij}|X_{ij}, X'_{ij}) = E(Y_{ij}|X_{ij}) - E(Y_{ij}|X'_{ij})$$

- In conditional prediction models, the average predicted treatment effect (qi$att.pr) for the treatment group is given by

$$\frac{1}{\sum_{i=1}^{M}\sum_{j=1}^{n_i} t_{ij}} \sum_{i=1}^{M} \sum_{j:t_{ij}=1}^{n_i} \{Y_{ij}(t_{ij}=1) - Y_{ij}(\widehat{t_{ij}=0})\},$$

  where t_{ij} is a binary explanatory variable defining the treatment ($t_{ij} = 1$) and control ($t_{ij} = 0$) groups. Variation in the simulations is due to uncertainty in simulating $Y_{ij}(t_{ij} = 0)$, the counterfactual predicted value of Y_{ij} for observations in the treatment group, under the assumption that everything stays the same except that the treatment indicator is switched to $t_{ij} = 0$.

- In conditional prediction models, the average expected treatment effect (qi$att.ev) for the treatment group is given by

$$\frac{1}{\sum_{i=1}^{M}\sum_{j=1}^{n_i} t_{ij}} \sum_{i=1}^{M} \sum_{j:t_{ij}=1}^{n_i} \{Y_{ij}(t_{ij}=1) - E[Y_{ij}(t_{ij}=0)]\},$$

  where t_{ij} is a binary explanatory variable defining the treatment ($t_{ij} = 1$) and control ($t_{ij} = 0$) groups. Variation in the simulations is due to uncertainty in simulating $E[Y_{ij}(t_{ij} = 0)]$, the counterfactual expected value of Y_{ij} for observations in the treatment group, under the assumption that everything stays the same except that the treatment indicator is switched to $t_{ij} = 0$.

1.0.7 Output Values

The output of each Zelig command contains useful information which you may view. You may examine the available information in `z.out` by using `slotNames(z.out)`, see the fixed effect coefficients by using `summary(z.out)@coefs`, and a default summary of information through `summary(z.out)`. Other elements available through the operator are listed below.

- From the `zelig()` output stored in `summary(z.out)`, you may extract:

  - `fixef`: numeric vector containing the conditional estimates of the fixed effects.
  - `ranef`: numeric vector containing the conditional modes of the random effects.
  - `frame`: the model frame for the model.

- From the `sim()` output stored in `s.out`, you may extract quantities of interest stored in a data frame:
  - `qi$pr`: the simulated predicted values drawn from the distributions defined by the expected values.
  - `qi$ev`: the simulated expected values for the specified values of x.
  - `qi$fd`: the simulated first differences in the expected values for the values specified in x and x1.
  - `qi$ate.pr`: the simulated average predicted treatment effect for the treated from conditional prediction models.
  - `qi$ate.ev`: the simulated average expected treatment effect for the treated from conditional prediction models.

How to Cite the Multi-level Gamma Model

How to Cite the Zelig Software Package

To cite Zelig as a whole, please reference these two sources:

> Kosuke Imai, Gary King, and Olivia Lau. 2007. "Zelig: Everyone's Statistical Software," `http://GKing.harvard.edu/zelig`.

> Imai, Kosuke, Gary King, and Olivia Lau. (2008). "Toward A Common Framework for Statistical Analysis and Development." Journal of Computational and Graphical Statistics, Vol. 17, No. 4 (December), pp. 892-913.

See also

Mixed effects gamma regression is part of `lme4` package by Douglas M. Bates [1].

2 `logit.mixed`: Mixed effects logistic Regression

Use generalized multi-level linear regression if you have covariates that are grouped according to one or more classification factors. The logit model is appropriate when the dependent variable is dichotomous.

While generally called multi-level models in the social sciences, this class of models is often referred to as mixed-effects models in the statistics literature and as hierarchical models in a Bayesian setting. This general class of models consists of linear models that are expressed as a function of both *fixed effects*, parameters corresponding to an entire population or certain repeatable levels of experimental factors, and *random effects*, parameters corresponding to individual experimental units drawn at random from a population.

2.0.8 Syntax

```
z.out <- zelig(formula= y ~ x1 + x2 + tag(z1 + z2 | g),
               data=mydata, model="logit.mixed")

z.out <- zelig(formula= list(mu=y ~ xl + x2 + tag(z1, gamma | g),
               gamma= ~ tag(w1 + w2 | g)), data=mydata, model="logit.mixed")
```

2.0.9 Inputs

`zelig()` takes the following arguments for `mixed`:

- `formula:` a two-sided linear formula object describing the systematic component of the model, with the response on the left of a ˜ operator and the fixed effects terms, separated by + operators, on the right. Any random effects terms are included with the notation `tag(z1 + ... + zn | g)` with `z1 + ... + zn` specifying the model for the random effects and `g` the grouping structure. Random intercept terms are included with the notation `tag(1 | g)`.
  Alternatively, `formula` may be a list where the first entry, `mu`, is a two-sided linear formula object describing the systematic component of the model, with the repsonse on the left of a ˜ operator and the fixed effects terms, separated by + operators, on the right. Any random effects terms are included with the notation `tag(z1, gamma | g)` with `z1` specifying the individual level model for the random effects, `g` the grouping structure and `gamma` references the second equation in the list. The `gamma` equation is one-sided linear formula object with the group level model for the random effects on the right side of a ˜ operator. The model is specified with the notation `tag(w1 + ... + wn | g)` with `w1 + ... + wn` specifying the group level model and `g` the grouping structure.

2.0.10 Additional Inputs

In addition, `zelig()` accepts the following additional arguments for model specification:

- `data:` An optional data frame containing the variables named in `formula`. By default, the variables are taken from the environment from which `zelig()` is called.

- `na.action:` A function that indicates what should happen when the data contain `NA`s. The default action (`na.fail`) causes `zelig()` to print an error message and terminate if there are any incomplete observations.

Additionally, users may with to refer to `lmer` in the package `lme4` for more information, including control parameters for the estimation algorithm and their defaults.

2.0.11 Examples

1. Basic Example with First Differences

   Attach sample data:

   ```
   > data(voteincome)
   ```

   Estimate model:

   ```
   > z.out1 <- zelig(vote ~ education + age + female + tag(1 | state), data=voteincome, mo
   ```

   Summarize regression coefficients and estimated variance of random effects:

   ```
   > summary(z.out1)
   ```

   Set explanatory variables to their default values, with high (80th percentile) and low (20th percentile) values for education:

   ```
   > x.high <- setx(z.out1, education=quantile(voteincome$education, 0.8))
   > x.low <- setx(z.out1, education=quantile(voteincome$education, 0.2))
   ```

   Generate first differences for the effect of high versus low education on voting:

   ```
   > s.out1 <- sim(z.out1, x=x.high, x1=x.low)
   > summary(s.out1)
   ```

2.0.12 Mixed effects Logistic Regression Model

Let Y_{ij} be the binary dependent variable, realized for observation j in group i as y_{ij} which takes the value of either 0 or 1, for $i = 1, \ldots, M$, $j = 1, \ldots, n_i$.

- The *stochastic component* is described by a Bernoulli distribution with mean vector π_{ij}.

  $$Y_{ij} \sim \text{Bernoulli}(y_{ij}|\pi_{ij}) = \pi_{ij}^{y_{ij}}(1 - \pi_{ij})^{1-y_{ij}}$$

  where

  $$\pi_{ij} = \Pr(Y_{ij} = 1)$$

- The q-dimensional vector of *random effects*, b_i, is restricted to be mean zero, and therefore is completely characterized by the variance covarance matrix Ψ, a $(q \times q)$ symmetric positive semi-definite matrix.

$$b_i \sim Normal(0, \Psi)$$

- The *systematic component* is

$$\pi_{ij} \equiv \frac{1}{1 + \exp(-(X_{ij}\beta + Z_{ij}b_i))}$$

where X_{ij} is the $(n_i \times p \times M)$ array of known fixed effects explanatory variables, β is the p-dimensional vector of fixed effects coefficients, Z_{ij} is the $(n_i \times q \times M)$ array of known random effects explanatory variables and b_i is the q-dimensional vector of random effects.

2.0.13 Quantities of Interest

- The predicted values (`qi$pr`) are draws from the Binomial distribution with mean equal to the simulated expected value, π_{ij} for

$$\pi_{ij} = \frac{1}{1 + \exp(-(X_{ij}\beta + Z_{ij}b_i))}$$

given X_{ij} and Z_{ij} and simulations of of β and b_i from their posterior distributions. The estimated variance covariance matrices are taken as correct and are themselves not simulated.

- The expected values (`qi$ev`) are simulations of the predicted probability of a success given draws of β from its posterior:

$$E(Y_{ij}|X_{ij}) = \pi_{ij} = \frac{1}{1 + exp(-X_{ij}\beta)}.$$

- The first difference (`qi$fd`) is given by the difference in predicted probabilities, conditional on X_{ij} and X'_{ij}, representing different values of the explanatory variables.

$$FD(Y_{ij}|X_{ij}, X'_{ij}) = Pr(Y_{ij} = 1|X_{ij}) - Pr(Y_{ij} = 1|X'_{ij})$$

- The risk ratio (`qi$rr`) is defined as

$$RR(Y_{ij}|X_{ij}, X'_{ij}) = \frac{Pr(Y_{ij} = 1|X_{ij})}{Pr(Y_{ij} = 1|X'_{ij})}$$

- In conditional prediction models, the average predicted treatment effect (`qi$att.pr`) for the treatment group is given by

$$\frac{1}{\sum_{i=1}^{M}\sum_{j=1}^{n_i} t_{ij}} \sum_{i=1}^{M} \sum_{j:t_{ij}=1}^{n_i} \{Y_{ij}(t_{ij} = 1) - Y_{ij}(\widehat{t_{ij} = 0})\},$$

where t_{ij} is a binary explanatory variable defining the treatment ($t_{ij} = 1$) and control ($t_{ij} = 0$) groups. Variation in the simulations is due to uncertainty in simulating $Y_{ij}(t_{ij} = 0)$, the counterfactual predicted value of Y_{ij} for observations in the treatment group, under the assumption that everything stays the same except that the treatment indicator is switched to $t_{ij} = 0$.

- In conditional prediction models, the average expected treatment effect (`qi$att.ev`) for the treatment group is given by

$$\frac{1}{\sum_{i=1}^{M}\sum_{j=1}^{n_i} t_{ij}} \sum_{i=1}^{M} \sum_{j:t_{ij}=1}^{n_i} \{Y_{ij}(t_{ij} = 1) - E[Y_{ij}(t_{ij} = 0)]\},$$

  where t_{ij} is a binary explanatory variable defining the treatment ($t_{ij} = 1$) and control ($t_{ij} = 0$) groups. Variation in the simulations is due to uncertainty in simulating $E[Y_{ij}(t_{ij} = 0)]$, the counterfactual expected value of Y_{ij} for observations in the treatment group, under the assumption that everything stays the same except that the treatment indicator is switched to $t_{ij} = 0$.

2.0.14 Output Values

The output of each Zelig command contains useful information which you may view. You may examine the available information in `z.out` by using `slotNames(z.out)`, see the fixed effect coefficients by using `summary(z.out)@coefs`, and a default summary of information through `summary(z.out)`. Other elements available through the operator are listed below.

- From the `zelig()` output stored in `summary(z.out)`, you may extract:
  - `fixef`: numeric vector containing the conditional estimates of the fixed effects.
  - `ranef`: numeric vector containing the conditional modes of the random effects.
  - `frame`: the model frame for the model.
- From the `sim()` output stored in `s.out`, you may extract quantities of interest stored in a data frame:
  - `qi$pr`: the simulated predicted values drawn from the distributions defined by the expected values.
  - `qi$ev`: the simulated expected values for the specified values of x.
  - `qi$fd`: the simulated first differences in the expected values for the values specified in x and x1.
  - `qi$ate.pr`: the simulated average predicted treatment effect for the treated from conditional prediction models.
  - `qi$ate.ev`: the simulated average expected treatment effect for the treated from conditional prediction models.

How to Cite the Multi-level Logit Model

How to Cite the Zelig Software Package

To cite Zelig as a whole, please reference these two sources:

> Kosuke Imai, Gary King, and Olivia Lau. 2007. "Zelig: Everyone's Statistical Software," `http://GKing.harvard.edu/zelig`.

> Imai, Kosuke, Gary King, and Olivia Lau. (2008). "Toward A Common Framework for Statistical Analysis and Development." Journal of Computational and Graphical Statistics, Vol. 17, No. 4 (December), pp. 892-913.

See also

Mixed effects logistic regression is part of `lme4` package by Douglas M. Bates [1].

3 `ls.mixed`: Mixed effects Linear Regression

Use multi-level linear regression if you have covariates that are grouped according to one or more classification factors and a continuous dependent variable.

While generally called multi-level models in the social sciences, this class of models is often referred to as mixed-effects models in the statistics literature and as hierarchical models in a Bayesian setting. This general class of models consists of linear models that are expressed as a function of both *fixed effects*, parameters corresponding to an entire population or certain repeatable levels of experimental factors, and *random effects*, parameters corresponding to individual experimental units drawn at random from a population.

3.0.15 Syntax

```
z.out <- zelig(formula= y ~ x1 + x2 + tag(z1 + z2 | g),
               data=mydata, model="lm.multi")

z.out <- zelig(formula= list(mu=y ~ xl + x2 + tag(z1, gamma | g),
               gamma= ~ tag(w1 + w2 | g)), data=mydata, model="lm.multi")
```

3.0.16 Inputs

`zelig()` takes the following arguments for `multi`:

- `formula:` a two-sided linear formula object describing the systematic component of the model, with the response on the left of a ~ operator and the fixed effects terms, separated by + operators, on the right. Any random effects terms are included with the notation `tag(z1 + ... + zn | g)`

with `z1 + ... + zn` specifying the model for the random effects and `g` the grouping structure. Random intercept terms are included with the notation `tag(1 | g)`.
Alternatively, `formula` may be a list where the first entry, `mu`, is a two-sided linear formula object describing the systematic component of the model, with the repsonse on the left of a ˜ operator and the fixed effects terms, separated by `+` operators, on the right. Any random effects terms are included with the notation `tag(z1, gamma | g)` with `z1` specifying the individual level model for the random effects, `g` the grouping structure and `gamma` references the second equation in the list. The `gamma` equation is one-sided linear formula object with the group level model for the random effects on the right side of a ˜ operator. The model is specified with the notation `tag(w1 + ... + wn | g)` with `w1 + ... + wn` specifying the group level model and `g` the grouping structure.

3.0.17 Additional Inputs

In addition, `zelig()` accepts the following additional arguments for model specification:

- `data:` An optional data frame containing the variables named in `formula`. By default, the variables are taken from the environment from which `zelig()` is called.
- `family:` A GLM family, see `glm` and `family` in the `stats` package. If `family` is missing then a linear mixed model is fit; otherwise a generalized linear mixed model is fit. In the later case only `gaussian` family with `"log"` link is supported at the moment.
- `na.action:` A function that indicates what should happen when the data contain `NAs`. The default action (`na.fail`) causes `zelig()` to print an error message and terminate if there are any incomplete observations.

Additionally, users may wish to refer to `lmer` in the package `lme4` for more information, including control parameters for the estimation algorithm and their defaults.

3.0.18 Examples

1. Basic Example with First Differences

   Attach sample data:

   ```
   > data(voteincome)
   ```

   Estimate model:

   ```
   > z.out1 <- zelig(income ~ education + age + female + tag(1 | state), data=voteincome,
   ```

Summarize regression coefficients and estimated variance of random effects:

```
> summary(z.out1)
```

Set explanatory variables to their default values, with high (80th percentile) and low (20th percentile) values for education:

```
> x.high <- setx(z.out1, education=quantile(voteincome$education, 0.8))
> x.low <- setx(z.out1, education=quantile(voteincome$education, 0.2))
```

Generate first differences for the effect of high versus low education on income:

```
> s.out1 <- sim(z.out1, x=x.high, x1=x.low)
> summary(s.out1)

> plot(s.out1)
```

3.0.19 Mixed effects linear regression model

Let Y_{ij} be the continuous dependent variable, realized for observation j in group i as y_{ij}, for $i = 1, \ldots, M$, $j = 1, \ldots, n_i$.

- The *stochastic component* is described by a univariate normal model with a vector of means μ_{ij} and scalar variance σ^2.

$$Y_{ij} \sim \text{Normal}(y_{ij}|\mu_{ij}, \sigma^2)$$

- The q-dimensional vector of *random effects*, b_i, is restricted to be mean zero, and therefore is completely characterized by the variance covarance matrix Ψ, a $(q \times q)$ symmetric positive semi-definite matrix.

$$b_i \sim Normal(0, \Psi)$$

- The *systematic component* is

$$\mu_{ij} \equiv X_{ij}\beta + Z_{ij}b_i$$

  where X_{ij} is the $(n_i \times p \times M)$ array of known fixed effects explanatory variables, β is the p-dimensional vector of fixed effects coefficients, Z_{ij} is the $(n_i \times q \times M)$ array of known random effects explanatory variables and b_i is the q-dimensional vector of random effects.

3.0.20 Quantities of Interest

- The predicted values (`qi$pr`) are draws from the normal distribution defined by mean μ_{ij} and variance σ^2,

$$\mu_{ij} = X_{ij}\beta + Z_{ij}b_i$$

  given X_{ij} and Z_{ij} and simulations of β and b_i from their posterior distributions. The estimated variance covariance matrices are taken as correct and are themselves not simulated.

- The expected values (`qi$ev`) are averaged over the stochastic components and are given by

$$E(Y_{ij}|X_{ij}) = X_{ij}\beta.$$

- The first difference (`qi$fd`) is given by the difference in expected values, conditional on X_{ij} and X'_{ij}, representing different values of the explanatory variables.

$$FD(Y_{ij}|X_{ij}, X'_{ij}) = E(Y_{ij}|X_{ij}) - E(Y_{ij}|X'_{ij})$$

- In conditional prediction models, the average predicted treatment effect (`qi$att.pr`) for the treatment group is given by

$$\frac{1}{\sum_{i=1}^{M}\sum_{j=1}^{n_i} t_{ij}} \sum_{i=1}^{M} \sum_{j:t_{ij}=1}^{n_i} \{Y_{ij}(t_{ij}=1) - Y_{ij}(\widehat{t_{ij}=0})\},$$

  where t_{ij} is a binary explanatory variable defining the treatment ($t_{ij} = 1$) and control ($t_{ij} = 0$) groups. Variation in the simulations is due to uncertainty in simulating $Y_{ij}(t_{ij} = 0)$, the counterfactual predicted value of Y_{ij} for observations in the treatment group, under the assumption that everything stays the same except that the treatment indicator is switched to $t_{ij} = 0$.

- In conditional prediction models, the average expected treatment effect (`qi$att.ev`) for the treatment group is given by

$$\frac{1}{\sum_{i=1}^{M}\sum_{j=1}^{n_i} t_{ij}} \sum_{i=1}^{M} \sum_{j:t_{ij}=1}^{n_i} \{Y_{ij}(t_{ij}=1) - E[Y_{ij}(t_{ij}=0)]\},$$

  where t_{ij} is a binary explanatory variable defining the treatment ($t_{ij} = 1$) and control ($t_{ij} = 0$) groups. Variation in the simulations is due to uncertainty in simulating $E[Y_{ij}(t_{ij} = 0)]$, the counterfactual expected value of Y_{ij} for observations in the treatment group, under the assumption that everything stays the same except that the treatment indicator is switched to $t_{ij} = 0$.

- If `"log"` link is used, expected values are computed as above and then exponentiated, while predicted values are draws from the log-normal distribution whose logarithm has mean and variance equal to μ_{ij} and σ^2, respectively.

3.0.21 Output Values

The output of each Zelig command contains useful information which you may view. You may examine the available information in `z.out` by using `slotNames(z.out)`, see the fixed effect coefficients by using `summary(z.out)@coefs`, and a default summary of information through `summary(z.out)`. Other elements available through the operator are listed below.

- From the `zelig()` output stored in `summary(z.out)`, you may extract:
  - `fixef`: numeric vector containing the conditional estimates of the fixed effects.
  - `ranef`: numeric vector containing the conditional modes of the random effects.
  - `frame`: the model frame for the model.
- From the `sim()` output stored in `s.out`, you may extract quantities of interest stored in a data frame:
  - `qi$pr`: the simulated predicted values drawn from the distributions defined by the expected values.
  - `qi$ev`: the simulated expected values for the specified values of x.
  - `qi$fd`: the simulated first differences in the expected values for the values specified in x and x1.
  - `qi$ate.pr`: the simulated average predicted treatment effect for the treated from conditional prediction models.
  - `qi$ate.ev`: the simulated average expected treatment effect for the treated from conditional prediction models.

How to Cite the Multi-level Least Squares Model

How to Cite the Zelig Software Package

To cite Zelig as a whole, please reference these two sources:

Kosuke Imai, Gary King, and Olivia Lau. 2007. "Zelig: Everyone's Statistical Software," `http://GKing.harvard.edu/zelig`.

Imai, Kosuke, Gary King, and Olivia Lau. (2008). "Toward A Common Framework for Statistical Analysis and Development." Journal of Computational and Graphical Statistics, Vol. 17, No. 4 (December), pp. 892-913.

See also

Mixed effects linear regression is part of `lme4` package by Douglas M. Bates [1].

4 `poisson.mixed`: Mixed effects poisson Regression

Use generalized multi-level linear regression if you have covariates that are grouped according to one or more classification factors. Poisson regression applies to dependent variables that represent the number of independent events that occur during a fixed period of time.

While generally called multi-level models in the social sciences, this class of models is often referred to as mixed-effects models in the statistics literature and as hierarchical models in a Bayesian setting. This general class of models consists of linear models that are expressed as a function of both *fixed effects*, parameters corresponding to an entire population or certain repeatable levels of experimental factors, and *random effects*, parameters corresponding to individual experimental units drawn at random from a population.

4.0.22 Syntax

```
z.out <- zelig(formula= y ~ x1 + x2 + tag(z1 + z2 | g),
               data=mydata, model="poisson.mixed")

z.out <- zelig(formula= list(mu=y ~ xl + x2 + tag(z1, gamma | g),
               gamma= ~ tag(w1 + w2 | g)), data=mydata, model="poisson.mixed")
```

4.0.23 Inputs

`zelig()` takes the following arguments for `mixed`:

- `formula:` a two-sided linear formula object describing the systematic component of the model, with the response on the left of a ˜ operator and the fixed effects terms, separated by `+` operators, on the right. Any random effects terms are included with the notation `tag(z1 + ... + zn | g)` with `z1 + ... + zn` specifying the model for the random effects and `g` the grouping structure. Random intercept terms are included with the notation `tag(1 | g)`.
  Alternatively, `formula` may be a list where the first entry, `mu`, is a two-sided linear formula object describing the systematic component of the model, with the repsonse on the left of a ˜ operator and the fixed effects terms, separated by `+` operators, on the right. Any random effects terms are included with the notation `tag(z1, gamma | g)` with `z1` specifying the individual level model for the random effects, `g` the grouping structure and `gamma` references the second equation in the list. The `gamma` equation is one-sided linear formula object with the group level model for the random effects on the right side of a ˜ operator. The model is specified with the notation `tag(w1 + ... + wn | g)` with `w1 + ... + wn` specifying the group level model and `g` the grouping structure.

4.0.24 Additional Inputs

In addition, `zelig()` accepts the following additional arguments for model specification:

- `data:` An optional data frame containing the variables named in `formula`. By default, the variables are taken from the environment from which `zelig()` is called.
- `na.action:` A function that indicates what should happen when the data contain `NAs`. The default action (`na.fail`) causes `zelig()` to print an error message and terminate if there are any incomplete observations.

Additionally, users may with to refer to `lmer` in the package `lme4` for more information, including control parameters for the estimation algorithm and their defaults.

4.0.25 Examples

1. Basic Example

   Attach sample data:

   ```
   > data(homerun)
   ```

   Estimate model:

   ```
   > z.out1 <- zelig(homeruns ~ player + tag(player - 1 | month), data=homerun, model="poi
   ```

   Summarize regression coefficients and estimated variance of random effects:

   ```
   > summary(z.out1)
   ```

   Set explanatory variables to their default values:

   ```
   > x.out <- setx(z.out1)
   ```

   Simulate draws using the default bootstrap method and view simulated quantities of interest:

   ```
   > s.out1 <- sim(z.out1, x=x.out)
   > summary(s.out1)
   ```

4.0.26 Mixed effects Poisson Regression Model

Let Y_{ij} be the number of independent events that occur during a fixed time period, realized for observation j in group i as y_{ij}, which takes any non-negative integer as its value, for $i = 1, \ldots, M$, $j = 1, \ldots, n_i$.

- The *stochastic component* is described by a Poisson distribution with mean and variance parameter λ_{ij}.

  $$Y_{ij} \sim \text{Poisson}(y_{ij}|\lambda_{ij}) = \frac{\exp(-\lambda_{ij})\lambda_{ij}^{y_{ij}}}{y_{ij}!}$$

  where

  $$y_{ij} = 0, 1, \ldots$$

- The q-dimensional vector of *random effects*, b_i, is restricted to be mean zero, and therefore is completely characterized by the variance covarance matrix Ψ, a $(q \times q)$ symmetric positive semi-definite matrix.

  $$b_i \sim Normal(0, \Psi)$$

- The *systematic component* is

  $$\lambda_{ij} \equiv \exp(X_{ij}\beta + Z_{ij}b_i)$$

  where X_{ij} is the $(n_i \times p \times M)$ array of known fixed effects explanatory variables, β is the p-dimensional vector of fixed effects coefficients, Z_{ij} is the $(n_i \times q \times M)$ array of known random effects explanatory variables and b_i is the q-dimensional vector of random effects.

4.0.27 Quantities of Interest

- The predicted values (`qi$pr`) are draws from the poisson distribution defined by mean λ_{ij}, for

  $$\lambda_{ij} = \exp(X_{ij}\beta + Z_{ij}b_i)$$

  given X_{ij} and Z_{ij} and simulations of of β and b_i from their posterior distributions. The estimated variance covariance matrices are taken as correct and are themselves not simulated.

- The expected values (`qi$ev`) is the mean of simulations of the stochastic component given draws of β from its posterior:

  $$E(Y_{ij}|X_{ij}) = \lambda_{ij} = \exp(X_{ij}\beta).$$

- The first difference (`qi$fd`) is given by the difference in expected values, conditional on X_{ij} and X'_{ij}, representing different values of the explanatory variables.

  $$FD(Y_{ij}|X_{ij}, X'_{ij}) = E(Y_{ij}|X_{ij}) - E(Y_{ij}|X'_{ij})$$

- In conditional prediction models, the average predicted treatment effect (`qi$att.pr`) for the treatment group is given by

$$\frac{1}{\sum_{i=1}^{M}\sum_{j=1}^{n_i} t_{ij}} \sum_{i=1}^{M} \sum_{j:t_{ij}=1}^{n_i} \{Y_{ij}(t_{ij}=1) - Y_{ij}(\widehat{t_{ij}=0})\},$$

  where t_{ij} is a binary explanatory variable defining the treatment ($t_{ij} = 1$) and control ($t_{ij} = 0$) groups. Variation in the simulations is due to uncertainty in simulating $Y_{ij}(t_{ij} = 0)$, the counterfactual predicted value of Y_{ij} for observations in the treatment group, under the assumption that everything stays the same except that the treatment indicator is switched to $t_{ij} = 0$.

- In conditional prediction models, the average expected treatment effect (`qi$att.ev`) for the treatment group is given by

$$\frac{1}{\sum_{i=1}^{M}\sum_{j=1}^{n_i} t_{ij}} \sum_{i=1}^{M} \sum_{j:t_{ij}=1}^{n_i} \{Y_{ij}(t_{ij}=1) - E[Y_{ij}(t_{ij}=0)]\},$$

  where t_{ij} is a binary explanatory variable defining the treatment ($t_{ij} = 1$) and control ($t_{ij} = 0$) groups. Variation in the simulations is due to uncertainty in simulating $E[Y_{ij}(t_{ij} = 0)]$, the counterfactual expected value of Y_{ij} for observations in the treatment group, under the assumption that everything stays the same except that the treatment indicator is switched to $t_{ij} = 0$.

4.0.28 Output Values

The output of each Zelig command contains useful information which you may view. You may examine the available information in `z.out` by using `slotNames(z.out)`, see the fixed effect coefficients by using `summary(z.out)@coefs`, and a default summary of information through `summary(z.out)`. Other elements available through the operator are listed below.

- From the `zelig()` output stored in `summary(z.out)`, you may extract:
  - `fixef`: numeric vector containing the conditional estimates of the fixed effects.
  - `ranef`: numeric vector containing the conditional modes of the random effects.
  - `frame`: the model frame for the model.
- From the `sim()` output stored in `s.out`, you may extract quantities of interest stored in a data frame:
  - `qi$pr`: the simulated predicted values drawn from the distributions defined by the expected values.

- `qi$ev`: the simulated expected values for the specified values of x.
- `qi$fd`: the simulated first differences in the expected values for the values specified in x and x1.
- `qi$ate.pr`: the simulated average predicted treatment effect for the treated from conditional prediction models.
- `qi$ate.ev`: the simulated average expected treatment effect for the treated from conditional prediction models.

How to Cite the Multi-level Poisson Model

How to Cite the Zelig Software Package

To cite Zelig as a whole, please reference these two sources:

> Kosuke Imai, Gary King, and Olivia Lau. 2007. "Zelig: Everyone's Statistical Software," `http://GKing.harvard.edu/zelig`.

> Imai, Kosuke, Gary King, and Olivia Lau. (2008). "Toward A Common Framework for Statistical Analysis and Development." Journal of Computational and Graphical Statistics, Vol. 17, No. 4 (December), pp. 892-913.

See also

Mixed effects poisson regression is part of `lme4` package by Douglas M. Bates [1].

References

[1] Douglas Bates. *lme4: Fit linear and generalized linear mixed-effects models*, 2007.

[2] Matthew Owen, Olivia Lau, Kosuke Imai, and Gary King. *gamma.mixed: Mixed Effects Gamma Regression*, 2011.

[3] Matthew Owen, Olivia Lau, Kosuke Imai, and Gary King. *logit.mixed: Mixed Effects Logit Regression*, 2011.

[4] Matthew Owen, Olivia Lau, Kosuke Imai, and Gary King. *ls.mixed: Mixed Effects Least Squares Regression*, 2011.

[5] Matthew Owen, Olivia Lau, Kosuke Imai, and Gary King. *poisson.mixed: Mixed Effects Poisson Regression*, 2011.