



| Method                      | Description   |
|-----------------------------|---|
| <code>print()</code>        | Simple printed display with coefficients  |
| <code>summary()</code>      | Standard regression summary; returns <code>summary.negbin1</code> object (with <code>print()</code> method) |
| <code>coef()</code>         | Extract coefficients  |
| <code>vcov()</code>         | Associated covariance matrix  |
| <code>predict()</code>      | (Different types of) predictions for new data   |
| <code>fitted()</code>       | Fitted values for observed data   |
| <code>residuals()</code>    | Extract (different types of) residuals  |
| <code>terms()</code>        | Extract terms   |
| <code>model.matrix()</code> | Extract model matrix (or matrices)  |
| <code>nobs()</code>         | Extract number of observations  |
| <code>logLik()</code>       | Extract fitted log-likelihood   |
| <code>bread()</code>        | Extract bread for <b>sandwich</b> covariance  |
| <code>estfun()</code>       | Extract estimating functions (= gradient contributions) for <b>sandwich</b> covariances                     |
| <code>getSummary()</code>   | Extract summary statistics for <code>mtable()</code>  |

Table 1: S3 methods provided in **negbin1**.

A number of standard S3 methods are provided, see Table 1. Especially the `predict` method deserves greater attention:

```
predict.negbin1 <- function(object, newdata = NULL,
                           type = c("response", "location",
                                     "probability", "quantile"),
                           na.action = na.pass, at = 0.5, ...)
```

Due to these methods a number of useful utilities work automatically, e.g., `AIC()`, `BIC()`, `coefstest()` (**lmtest**), `lrtest()` (**lmtest**), `waldtest()` (**lmtest**), `linearHypothesis()` (**car**), `mtable()` (**memisc**), etc.

## Predictions

It is also possible to obtain predictions from an object of class **negbin**. For `type = "response"` (as well as for `type = "location"`), the conditional mean, the inverse link applied to the linear predictor, is calculated. `type = "probability"` computes the expected probabilities for each count `at = 0, 1, 2, 3, ...`, whereas `type = "quantile"` gives the quantile function for probabilities `at`.

```
R> mod <- negbin1(y ~ x1 + x2, data = d)
R> newdata <- data.frame(x1 = c(0, 0.4, 0.1, 2), x2 = c(0.1, 0, -0.1, 0.1))
R> predict(mod, newdata, type = "response")
```

```
      1      2      3      4
0.2373843 0.5994933 0.2776455 2.1915466
```

```
R> predict(mod, newdata, type = "quantile", at = 0.95)
```

```
1 2 3 4
1 3 2 10
```

```
R> predict(mod, newdata, type = "probability", at = 0)
```

```
          1          2          3          4
0.8292225 0.6911623 0.8089649 0.4688839
```

### Computation time

Below, the computation time of function `negbin1()` is compared to NB1 implementations from other packages, as `vglm` from **VGAM** (??) and `gamlss` from package **gamlss** (?). The simulated dataset comprises 10,000 observations.

|                                    | user.self | sys.self | elapsed |
|------------------------------------|-----------|----------|---------|
| <code>negbin1(grad = FALSE)</code> | 3.596     | 0.016    | 3.611   |
| <code>negbin1(grad = TRUE)</code>  | 1.336     | 0.000    | 1.336   |
| <code>vglm</code>                  | 5.456     | 0.000    | 5.459   |
| <code>gamlss</code>                | 0.720     | 0.004    | 0.724   |

The computation time of `negbin1()` does quite well compared to `vglm()`, but only in case that `grad = TRUE`. `gamlss()` is even faster than `negbin1()` with `grad = TRUE`.

## 3. Example

Data on the number of publications produced by Ph.D. biochemists taken from [Long \(1990\)](#) are employed to test the `negbin1` function (see also [Scott Long 1997](#)). The data are loaded from the **negbin1** package.

```
R> data("publications", package = "negbin1")
```

In a next step, a benchmark Poisson model is estimated, where the number of publications produced by Ph.D. biochemist appears as a dependent variable. Explanatory variables are `gender`, marriage status (`married`), number of children under age six (`kids`), prestige of Ph.D. program (`prestige`) and articles by mentor in the last three years (`mentor`).

```
R> poi <- glm(articles ~ gender + married + kids + prestige + mentor,
+           family = "poisson", data = publications)
R> summary(poi)
```

Call:

```
glm(formula = articles ~ gender + married + kids + prestige +
    mentor, family = "poisson", data = publications)
```

Deviance Residuals:

| Min     | 1Q      | Median  | 3Q     | Max    |
|---------|---------|---------|--------|--------|
| -3.5672 | -1.5398 | -0.3660 | 0.5722 | 5.4467 |

Coefficients:

|                | Estimate  | Std. Error | z value | Pr(> z ) |     |
|----------------|-----------|------------|---------|----------|-----|
| (Intercept)    | 0.304617  | 0.102981   | 2.958   | 0.0031   | **  |
| genderWomen    | -0.224594 | 0.054613   | -4.112  | 3.92e-05 | *** |
| marriedMarried | 0.155243  | 0.061374   | 2.529   | 0.0114   | *   |
| kids           | -0.184883 | 0.040127   | -4.607  | 4.08e-06 | *** |
| prestige       | 0.012823  | 0.026397   | 0.486   | 0.6271   |     |
| mentor         | 0.025543  | 0.002006   | 12.733  | < 2e-16  | *** |

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 1817.4 on 914 degrees of freedom  
 Residual deviance: 1634.4 on 909 degrees of freedom  
 AIC: 3314.1

Number of Fisher Scoring iterations: 5

However, the variance of the number of publications is more than twice the mean.

```
R> var(publications$articles) / mean(publications$articles)
```

[1] 2.191358

As the data exhibit overdispersion (of course without taking any covariates into account), a negative binomial 1 regression offers some remedy in such a situation.

```
R> nb1 <- negbin1(articles ~ gender + married + kids + prestige + mentor,
+               data = publications, grad = TRUE)
R> summary(nb1)
```

Call:

```
negbin1(formula = articles ~ gender + married + kids + prestige + mentor,
        data = publications, grad = TRUE)
```

Standardized residuals:

| Min       | 1Q      | Median | 3Q     | Max       |
|-----------|---------|--------|--------|-----------|
| -235.9945 | -1.0000 | 1.5878 | 4.7260 | 1702.3317 |

Coefficients:

|             | Estimate | Std. Error | z value | Pr(> z ) |   |
|-------------|----------|------------|---------|----------|---|
| (Intercept) | 0.23797  | 0.13222    | 1.800   | 0.071882 | . |

```

genderWomen    -0.18268    0.06985   -2.615  0.008917 **
marriedMarried  0.15667    0.07874    1.990  0.046621 *
kids           -0.17297    0.05108   -3.386  0.000709 ***
prestige       0.03154    0.03400    0.928  0.353513
mentor         0.02416    0.00260    9.296  < 2e-16 ***
log(alpha)     -0.23473    0.12278   -1.912  0.055904 .

```

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Log-likelihood: -1565 on 7 Df

Number of iterations in BFGS optimization: 16

While both sets of parameter estimates would lead to the same conclusions, the standard errors reported reveal that ordinary Poisson regression underestimates the standard errors.

```
R> cbind("Poisson" = sqrt(diag(vcov(poi))), "NB1" = sqrt(diag(vcov(nb1)))[-7])
```

|                | Poisson     | NB1         |
|----------------|-------------|-------------|
| (Intercept)    | 0.102981443 | 0.132217871 |
| genderWomen    | 0.054613488 | 0.069853861 |
| marriedMarried | 0.061374395 | 0.078740071 |
| kids           | 0.040126898 | 0.051082507 |
| prestige       | 0.026397045 | 0.033999414 |
| mentor         | 0.002006073 | 0.002599541 |

```
R> library("memisc")
```

```
R> mtable("Poisson" = poi, "NB1(negbin1)" = nb1,
+        summary.stats = c("Log-likelihood", "AIC", "BIC", "N"))
```

|                         | Poisson              | NB1                  |
|-------------------------|----------------------|----------------------|
| (Intercept)             | 0.305**<br>(0.103)   | 0.238<br>(0.132)     |
| gender: Women/Men       | -0.225***<br>(0.055) | -0.183**<br>(0.070)  |
| married: Married/Single | 0.155*<br>(0.061)    | 0.157*<br>(0.079)    |
| kids                    | -0.185***<br>(0.040) | -0.173***<br>(0.051) |
| prestige                | 0.013<br>(0.026)     | 0.032<br>(0.034)     |
| mentor                  | 0.026***<br>(0.002)  | 0.024***<br>(0.003)  |
| log(alpha)              |                      | -0.235<br>(0.123)    |
| Log-likelihood          | -1651.1              | -1564.7              |
| AIC                     | 3314.1               | 3143.4               |
| BIC                     | 3343.0               | 3177.1               |
| N                       | 915                  | 915                  |

Table 2: Comparing Poisson results with NB1 from `negbin1`.

## References

- Cameron AC, Trivedi PK (2013). *Regression Analysis of Count Data*, volume 53. Cambridge university press.
- Long JS (1990). "The Origins of Sex Differences in Science." *Social Forces*, pp. 1297–1316.
- Scott Long J (1997). "Regression Models for Categorical and Limited Dependent Variables." *Advanced Quantitative Techniques in the Social Sciences*, 7.
- Winkelmann R (2013). *Econometric Analysis of Count Data*. Springer Science & Business Media.

### Affiliation:

Susanne Berger  
 Department of Statistics  
 Faculty of Economics and Statistics  
 Universität Innsbruck  
 Universitätsstr. 15  
 6020 Innsbruck, Austria  
 Telephone: +43/512/507-7113  
 E-mail: [susanne.berger@uibk.ac.at](mailto:susanne.berger@uibk.ac.at)