Receiver Operating Characteristic (ROC) Regression

Description

Fit an ROC-GLM regression model for continuous or ordinal disease marker(s) or diagnostic test variables. Bootstrap confidence intervals for estimates are optionally included. Covariate adjustment is also accommodated.

Usage

```
rocreg(dataset = NULL, d, markers, regcov = NULL, sregcov = NULL,
       link = "probit", interval = c(0, 1, 10), ordinal = FALSE,
       pvcmeth = "empirical", tiecorr = FALSE, adjcov = NULL,
       adjmodel = "stratified", nsamp = 1000, nobstrap = FALSE,
       noccsamp = FALSE, nostsamp = FALSE, cluster = NULL,
       resfile = NULL, replace = FALSE, level = 95)
```

Arguments

dataset	optional character string specifying the name of the dataset to be used for analysis.
d	character string specifying the name of the 0/1 outcome vector.
markers	vector of character strings specifying the names of the test markers/variables.
regcov	character string vector specifying variables to be included in the ROC regression model that affect the intercept term of the ROC curve.
sregcov	character string vector specifying variables to be included in the ROC regression model that affect the slope of the ROC curve.
link	character string specifying the ROC-GLM link function as "probit" (default) or "logit". "probit" corresponds to the binormal ROC model, that is, $PHI^{-1}\{ROC(f)\} = intercept + slope * PHI^{-1}(f)$, where PHI is the standard normal cumulative distribution function. "logit" corresponds to the bilogistic ROC model, that is, $logit\{ROC(f)\} = intercept + slope * logit(f)$.
interval	numeric vector (a,b,np) specifying an FPR interval (a,b) in (0,1), and the number of points, np, in the interval over which the ROC-GLM model is to be fit. The default is (0,1,10).
ordinal	logical. If TRUE, the test marker(s) are specified as ordinal-valued ratings, rather than continuous measures. This option affects the fitting algorithm for the ROC-GLM and also affects the covariate adjustment options. Must be TRUE if adjmodel is "ologit" or "oprobit". "linear" model adjustment is not permitted with `ordinal`=TRUE. The default is FALSE.

pvcmeth character string specifying PV calculation method as "empirical" (default) or "normal". "empirical" uses the empirical distribution of the test measure among controls (D=0) as the reference distribution for the calculation of case PVs. The PV for the case measure y_i is the proportion of control measures smaller than y_i. "normal" models the test measure among controls with a normal distribution. The PV for the case measure y_i is the standard normal cumulative distribution function of $(y_i - \text{mean})/\text{sd}$, where the mean and the standard deviation (sd) are calculated by using the control sample.

tiecorr logical. If FALSE (default), no correction for ties. If TRUE, it indicates that a correction for ties between case and control values is included in the empirical PV calculation. The correction is important only in calculating summary indices, such as the area under the ROC curve. The tie-corrected PV for a case with the marker value y_i is the proportion of control values $Y_{Db} < y_i$ plus one half the proportion of control values $Y_{Db} = y_i$, where Y_{Db} denotes controls. By default, the PV calculation includes only the first term, i.e. the proportion of control values $Y_{Db} < y_i$. This option applies only to the empirical PV calculation method.

adjcov character string vector specifying covariates to adjust for.

adjmodel character string specifying how the covariate adjustment is to be done: "stratified" (default), "linear", "oprobit" (ordered probit), or "ologit" (ordered logit). If "stratified", PVs are calculated separately for each stratum defined by `adjcov`. This is the default if `adjmodel` is not specified and `adjcov` is. Each case-containing stratum must include at least two controls. Strata that do not include cases are excluded from calculations. "linear" fits a linear regression of the marker distribution on the adjustment covariates among controls. Standardized residuals based on this fitted linear model are used in place of the marker values for cases and controls. "oprobit" calculates PVs based on the fit of an ordered probit regression model of the marker on the adjustment covariates among controls. "ologit" calculates PVs based on the fit of an ordered logit regression model of the marker on the adjustment covariates among controls. "oprobit" and "ologit" assume that `markers` consists of ordinal-valued marker variables.

nsamp number of bootstrap samples to be drawn for estimating sampling variability of parameter estimates; default is nsamp=1000.

nobstrap logical. If TRUE, omit boostrap sampling and estmation of standard errors and CIs. If `nsamp` is specified, `nobstrap` will override it. Default is FALSE.

noccsamp logical. If TRUE, bootstrap samples are drawn from the combined sample (cohort sampling) rather than sampling separately from cases and controls (case-control sampling); default is FALSE (case-control sampling).

nostsamp logical. If TRUE (default), bootstrap samples are drawn without respect to covariate strata. By default, samples are drawn from within covariate strata when stratified covariate adjustment is requested via the `adjcov` and

	`adjmodel` options.
cluster	character string specifying variables that identify bootstrap resampling clusters.
resfile	character string specifying the filename to save bootstrap results for the ROC-GLM model in. The .txt file is called [filename].txt if a single marker is specified or [filename]\#.txt for the \#th marker if more than 1 marker is included in `markers`.
replace	logical. If TRUE, overwrite existing specified boostrap results file if it already exists; default is FALSE.
level	specify confidence level for CIs as a percentage; default is level=95.

Details

`rocreg` fits an ROC-GLM regression model (Pepe 2003 , sec. 6.4; Alonzo and Pepe) for each of the specified continuous disease markers or diagnostic test variables in `markers`, the 0/1 outcome indicator variable `d`, and optionally, covariates. Covariates specified with `regcov` are included in the intercept of the ROC curve. Covariates specified with `sregcov` are included in the slope of the ROC curve. Bootstrap standard errors and confidence intervals (CIs) for the model parameters are obtained.

ROC calculations are based on percentile values (PVs) of the case measures relative to the corresponding marker distribution among controls.

Adjustment for variables that affect the control distribution of the marker can be achieved either by stratification or with a linear regression approach. Ordinal regression covariate adjustment options are available if the `markers` measures are ordinal.

The ROC-GLM is fit over the FPR range (a,b) using thresholds corresponding to `np` equally spaced FPR points in (a,b) when markers values are considered to be continuous. Alternatively, the observed FPR points serve as thresholds for ordinal valued markers. If an ordered regression model is used for covariate adjustment, the resulting cutpoint estimates are used to calculate thresholds.

Value

List containing parameter estimates from the ROC-GLM curve fit and the corresponding bootstrap covariance matrix. Returned list items include the following:

b	1 x k matrix of ROC-GLM parameter estimates; k = 2 + number of covariates included in the intercept and slope terms. Columns correspond to alpha_0 and alpha_1 parameters plus coefficients for any specified covariates.
V	k x k bootstrap covariance matrix for the k ROC-GLM parameters.
GLMparm	n x k matrix of ROC-GLM parameter estimates. Rows correspond to the

marker variables included in `markers`, and columns are as for `b`. Returned whether bootstrap sampling is specified or not (`nobstrap`).

Author(s)

Aasthaa Bansal, University of Washington, Seattle, WA. abansal@u.washington.edu

Daryl Morris, University of Washington, Seattle, WA. darylm@u.washington.edu

Gary Longton, Fred Hutchinson Cancer Research Center, Seattle, WA. glongton@fhcrc.org

Margaret Pepe, Fred Hutchinson Cancer Research Center and University of Washington, Seattle, WA. mspepe@u.washington.edu

Holly Janes, Fred Hutchinson Cancer Research Center and University of Washington, Seattle, WA. hjanes@fhcrc.org

References

Alonzo, T.A., Pepe, M.S. 2002 Distribution-free ROC analysis using binary regression techniques. *Biostatistics* **3**,421–32.

Janes, H., Pepe, M.S. 2008. Adjusting for covariates in studies of diagnostic, screening, or prognostic markers: an old concept in a new setting. *American Journal of Epidemiology* **168**,89–97.

Janes, H., Longton G, Pepe, M.S. 2009. Accommodating covariates in receiver operating characteristic analysis. *Stata Journal* **9(1)**,17–39.

Morris, D.E., Pepe, M.S., Barlow, W.E. Contrasting Two Frameworks for ROC Analysis of Ordinal Ratings. *Medical Decision Making* (in press)

Pepe, M.S. 2003 *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford University Press.

See Also

comproc, roccurve.

Examples

```
nnhs2 <- read.csv("http://labs.fhcrc.org/pepe/book/data/nnhs2.csv",
                  header = TRUE, sep = ",")
rocreg(dataset="nnhs2", d="d", markers="y1", cluster="id", noccsamp=T)
rocreg(dataset="nnhs2", d="d", markers="y1",
      adjcov="gender", regcov="gender", cluster="id", noccsamp=T, level=90)
```

```
rocreg(dataset="nnhs2", d="d", markers="y1", adjcov="gender",
       regcov="gender", pvcmeth="normal", cluster="id", noccsamp=T)
rocreg(dataset="nnhs2", d="d", markers=c("y1","y2"),
       adjcov=c("currage","gender"), adjmodel="linear", regcov="currage",
       cluster="id", noccsamp=T)
rocreg(dataset="nnhs2", d="d", markers="y1",
       adjcov="gender", regcov="gender", sregcov="gender", link="logit",
       cluster="id", noccsamp=T)
```

[Package *pcvsuite* version 1.0 [Index](#)]