

# 回帰分析のための R の関数

Vito Ricci

vito\_ricci@yahoo.com

訳:荒木孝治

arakit@kansai-u.ac.jp

2011年6月21日

回帰分析のための R の関数を目的別に整理している。各項目の最後にある丸括弧 ( ) 内は、パッケージ名。

## 線形モデル

**Anova:** 線形モデル (lm) と一般化線形モデル (glm), multinom, polr に対する分散分析表の作成 (car)

**anova:** 1 つまたは 2 つ以上の推定した線形モデルに対して分散分析表を作成 (stats)

**coef:** モデル作成関数で作られたオブジェクトからモデルの係数を抽出するための総称的関数。coefficients は別名。(stats)

**coeftest:** 回帰係数の  $z$  検定または  $t$  検定 (lmtest)

**confint:** 当てはめたモデルにおける 1 つ以上のパラメータの信頼区間 (confidence interval) を求める。クラス "lm" を継承するオブジェクトに対する方法。(stats)

**deviance:** 推定したモデルオブジェクトの逸脱度 (deviance) を計算 (stats)

**effects:** 当てはめたモデル (通常は線形モデル) から効果を抽出する。これは総称的な関数であるが、現在は、クラス "lm" と "glm" を継承するオブジェクトに対してのみ有効。(stats)

**fitted:** モデル作成関数により作られたオブジェクトから予測値を抽出するための総称的関数。fitted.values はその別名 (stats)

**formula:** 他のオブジェクトに含まれるモデル式を抽出する方法を与える。(stats)

**linear.hypothesis:** 線形仮説の検定 (car)

**lm:** 線形モデルを作成。回帰分析, 単層分散分析, 共分散分析で利用できる。(stats)

**model.matrix:** デザイン行列の生成 (stats)

**predict:** 線形モデルオブジェクトに基づいて予測した値 (stats)

**residuals:** モデル作成関数により作られたオブジェクトから残差を抽出するための総称的関数 (stats)

**summary.lm:** クラス "lm" の要約情報 (stats)

**vcov:** 当てはめたモデルオブジェクトの主要なパラメータの分散共分散行列を返す (stats)

## モデル — 変数選択

**add1:** 説明変数を 1 つずつモデルに追加したり削除したりしてモデルを当てはめ、適合度の変化を計算して表にする (stats)

**AIC:** 1 つまたは複数の当てはめたモデルに対して、赤池情報量基準 (AIC) を求めるための総称的関数。式  $-2 * \log\text{-likelihood} + k * \text{npar}$  に基づいて対数尤度を求める。ここで、npar は当てはめたモデルでのパラメータ数で、通常は AIC に対しては  $k = 2$  を、BIC つまり SBC (Schwarz's Bayesian criterion) に対しては  $k = \log(n)$  ( $n$  は観測値の数) を用いる。(stats)

**Cpplot:**  $C_p$  のプロット (faraway)

**drop1:** 説明変数を 1 つずつモデルに追加したり、モデルから削除したりしてモデルを当てはめ、適合度の変化の表を計算する

( stats )  
**extractAIC**: 当てはめたパラメトリックモデルに対する ( 一般化 ) 赤池情報量基準 ( AIC ) を求める ( stats )  
**leaps**: 総当たり法による変数選択 ( leaps )  
**maxadjr**: 最大自由度調整済み寄与率 ( faraway )  
**offset**: offset は , 一般化線形モデルなどのモデルの説明変数に付け加える項である . その係数は推定するのではなく , 既知の値 1 を持つもの ( stats )  
**step**: AIC を用いた変数選択 ( stats )  
**update.formula**: モデル式のアップデートを行う . 基本は , 変数の取り込みと除去であるが , より一般的なことも実行可能である . ( stats )

#### 診断

**cookd**: 線形モデル , 一般化線形モデルに対する Cook ( クック ) の距離 ( car )  
**cooks.distance**: Cook の距離 ( stats )  
**covratio**: 共分散比 ( stats )  
**dfbeta**: DBETA ( stats )  
**dfbetas**: DBETAS ( stats )  
**dffits**: DFFTITS ( stats )  
**hat**: ハット行列の対角要素 ( stats )  
**hatvalues**: ハット行列の対角要素 ( stats )  
**influence.measures**: 線形モデルと一般化線形モデルに対する回帰診断のための関数の組 ( stats )  
**lm.influence**: 回帰の適合性をチェックするための様々な基本統計量を計算する関数 ( stats )  
**ls.diag**: 回帰に対する基本統計量 ( 標準誤差 , 回帰係数に対する  $t$  値 ,  $p$  値を含む ) を計算 ( stats )  
**outlier.test**: Bonferroni ( ボンフェローニ ) の外れ値の検定 ( car )  
**rstandard**: 標準化残差 ( stats )  
**rstudent**: スチューデント化残差 ( stats )  
**vif**: 分散拡大要因 ( car )

#### グラフィックス

**ceres.plots**: Ceres プロット ( car )  
**cr.plots**: 要素 + 残差 ( 偏残差 ) プロット ( car )  
**influence.plot**: 影響プロット ( car )  
**leverage.plots**: レベレッジプロット ( car )  
**panel.car**: 条件付き散布図 ( car )  
**plot.lm**: 回帰診断のための 4 つの図の作成 ( 選択可能 ) . 予測値と残差の散布図 , 予測値に対する残差の絶対値の平方根の散布図 , 正規 QQ プロット , クックの距離のプロット ( stats )  
**prplot**: 偏残差プロット ( faraway )  
**qq.plot**: QQ プロット ( car )  
**qqline**: 正規 QQ プロットに , 第 1 四分位数と第 3 四分位数を通る直線を付加 ( stats )  
**qqnorm**: 正規 QQ プロットを作成するための総称的な関数 ( stats )  
**reg.line**: 回帰直線のプロット ( car )  
**scatterplot.matrix**: 散布図行列 ( car )  
**scatterplot**: 箱ひげ図付きの散布図 ( car )  
**spread.level.plot**: Spread-Level プロット ( car )

#### 検定

**ad.test**: Anderson-Darling ( アンダーソン - ダーリン ) の正規性の検定 ( nortest )  
**bartlett.test**: Bartlett ( バートレット ) の分散の一様性の検定 ( stats )  
**bgtest**: Breusch-Godfrey 検定 ( lmtest )  
**bptest**: Breusch-Pagan ( ブルーシュ - ペーガン ) 検定 ( lmtest )  
**cvm.test**: Cramer-von Mises の正規性の検定 ( nortest )  
**durbin.watson**: Durbin-Watson ( ダービン - ワトソン ) の自己相関の検定 ( car )  
**dwtest**: Durbin-Watson ( ダービン - ワトソン ) 検定 ( lmtest )  
**levene.test**: Levene ( ルビーン ) の検定 ( car )  
**lillie.test**: Lilliefors ( Kolmogorov-Smirnov ) の正規性の検定 ( nortest )  
**ncv.test**: 分散の非均一性に対するスコア検定 ( car )

**pearson.test:** 正規性に対する Pearson (ピアソン) の  $\chi^2$  検定 (nortest)  
**sf.test:** Shapiro-Francia の正規性の検定 (nortest)  
**shapiro.test:** Shapiro-Wilk (シャピロ - ウィルク) の正規性の検定 (stats)

#### 変数変換

**box.cox:** Box-Cox (ボックス - コックス) 変換 (car)  
**boxcox:** 線形モデルに対する Box-Cox 変換 (MASS)  
**box.cox.powers:** 多変量非条件付き Box-Cox 変換 (car)  
**box.tidwell:** Box-Tidwell 変換 (car)  
**box.cox.var:** Box-Cox 変換の Constructed Variable を求める (car)

#### リッジ回帰

**lm.ridge:** リッジ回帰 (MASS)

#### 区分回帰

**segmented:** 回帰モデルにおける区分回帰関係 (segmented)  
**slope.segmented:** 区分回帰関係の傾きに対する要約情報 (segmented)

#### 一般化最小 2 乗法 (GLS)

**ACF.gls:** gls 残差の自己相関関数 (nlme)  
**anova.gls:** 推定したオブジェクトの尤度を比較する (nlme)  
**gl:** 一般化最小 2 乗法を用いて線形モデルを推定する (nlme)  
**intervals.gls:** gls パラメータの信頼区間 (nlme)  
**lm.gls:** 一般化最小 2 乗法による線形モデルの推定 (MASS)  
**plot.gls:** オブジェクトのプロット (nlme)  
**predict.gls:** gls オブジェクトを用いた予測 (nlme)  
**qqnorm.gls:** gls オブジェクトからの正規 QQ プロット (nlme)  
**residuals.gls:** gls 残差の抽出 (nlme)

**summary.gls:** gls オブジェクトの要約 (nlme)

#### 一般化線形回帰 (GLM)

**family:** glm といった関数で用いられるモデルの詳細を指定するための分布族オブジェクト (stats)  
**glm.nb:** 負の 2 項分布に対する一般化線形モデルの当てはめ (MASS)  
**glm:** 線形予測子の記号による記述と誤差分布の記述を与えることによって特定化された一般化線形モデルの当てはめに用いる (stats)  
**polr:** 比例オッズロジスティック回帰 (MASS)

#### 非線形最小 2 乗 (NLS)

**nlm:** Newton (ニュートン) 型のアルゴリズムを用いて関数  $f$  の最小化を実行する関数 (stats)  
**nls:** 非線形モデルのパラメータを非線形最小 2 乗法により決定し, nls オブジェクトを返す (stats)  
**nls.control:** nls 非線形最小 2 乗アルゴリズムのパラメータを指定 (stats)

#### 一般化非線形最小 2 乗法 (GNLS)

**coef.gnls:** gnls 係数の抽出 (nlme)  
**gnls:** 一般化最小 2 乗法を用いた非線形モデルの推定 (nlme)  
**predict.gnls:** gnls オブジェクトからの予測 (nlme)

#### loess (locally weighted regression: 局所重み付け回帰)

**loess:** 局所当てはめを用いて, 1 つ以上の数値説明変数により多項式表面を決定する (stats)  
**loess.control:** loess の当てはめに対するコントロールパラメータの設定 (stats)  
**predict.loess:** loess の当てはめからの予測, オプションとして標準誤差を計算 (stats)

**scatter.smooth**: loess で求めた平滑線を散布図に付加 ( stats )

#### スプライン回帰

**bs**: 多項式スプラインに対する B-スプライン基底 ( splines )

**ns**: 自然スプラインに対する基底行列の生成 ( splines )

**periodicSpline**: 周期補完スプラインの当てはめ ( splines )

**polySpline**: 区分的多項式スプライン表現 ( splines )

**predict.bSpline**: 新しい  $X$  値に対するスプラインの評価 ( splines )

**predict.bs**: スプラインの基底の評価 ( splines )

**splineDesign**: B-スプラインに対する計画行列 ( splines )

**splineKnots**: スプラインからの節点ベクトル ( splines )

**splineOrder**: スプラインの次数の決定 ( splines )

#### ロバスト回帰

**lqs**: 耐性回帰 ( Resistant Regression ) ( MASS )

**rlm**: 線形モデルのロバスト適合 ( MASS )

#### 構造方程式モデル

**sem**: 一般構造方程式モデル ( sem )

**tsls**: 2 段階最小 2 乗法 ( sem )

#### 同時方程式推定

**systemfit**: 最小 2 乗法 ( OLS ) や重み付き最小 2 乗法 ( WLS ), 見かけ上無関係な回帰 ( SUR : Seemingly Unrelated Regression ), 2 段階最小 2 乗法 ( 2SLS ), 重み付き 2 段階最小 2 乗法 ( W2SLS ), 3 段階最小 2 乗法 ( 3SLS ) を用いた線形構造方程式の当てはめ ( systemfit )

#### Partial Least Squares Regression ( PLSR ) と主成分回帰 ( PCR )

**biplot.mvr**: PLSR と PCR モデルに対するバイプロット ( pls )

**coefplot**: PLSR と PCR モデルの回帰係数のプロット ( pls )

**crossval**: PLSR と PCR モデルの交差検証法 ( pls )

**cvsegments**: 交差検証のためのデータのセグメントを作成 ( pls )

**kernelpls.fit**: カーネル PLS ( Dayal and MacGregor ) ( pls )

**mnc**: Multiplicative Scatter Correction ( pls )

**mvr**: Partial Least Squares と主成分回帰 ( pls )

**mvrCv**: 交差検証法 ( pls )

**oscorespls.fit**: 直交スコア PLSR ( pls )

**predplot**: 予測プロット ( pls )

**scoreplot**: スコアと負荷量のプロット ( pls )

**scores**: PLSR と PCR モデルからスコアと負荷量の抽出 ( pls )

**svdpc.fit**: 主成分回帰 ( pls )

**validationplot**: 検証プロット ( pls )

#### 分位点回帰

**anova.rq**: 分位点回帰に対する Anova 関数 ( quantreg )

**boot.rq**: 分位点回帰のブートストラッピング ( quantreg )

**lprq**: 局所多項分位点回帰 ( quantreg )

**nlrq**: 非線形分位点回帰を実行するための関数 ( quantreg )

**qss**: rqss 当てはめに対する加法ノンパラメトリック項 ( quantreg )

**ranks**: 分位点回帰ランク ( quantreg )

**rq**: 分位点回帰 ( quantreg )

**rqss**: 加法分位点回帰平滑化 ( quantreg )

**rrs.test**: 分位点回帰の順位スコア検定 ( quantreg )

**standardize**: 分位点回帰過程を標準化する関数 ( quantreg )

#### 線形, 非線形混合モデル

**ACF**: 自己相関関数 (nlme)  
**ACF.lme**: lme の残差に対する自己相関関数 (nlme)  
**anova.lme**: 推定したオブジェクトの尤度の比較 (nlme)  
**fitted.lme**: lme による推定値の抽出 (nlme)  
**fixed.effects**: 母数効果の抽出 (nlme)  
**intervals**: 係数の信頼区間 (nlme)  
**intervals.lme**: lme パラメータの信頼区間 (nlme)  
**lme**: 線形混合モデル (nlme)  
**nlme**: 非線形混合モデル (nlme)  
**predict.lme**: lme オブジェクトを用いた予測 (nlme)  
**predict.nlme**: nlme オブジェクトを用いた予測 (nlme)  
**qqnorm.lme**: lme オブジェクトからの残差または変量効果の正規 QQ プロット (nlme)  
**random.effects**: 変量効果の抽出 (nlme)  
**ranef.lme**: lme の変量効果の抽出 (nlme)  
**residuals.lme**: lme の残差の抽出 (nlme)  
**simulate.lme**: lme モデルのシミュレート (nlme)  
**summary.lme**: lme オブジェクトの要約情報 (nlme)  
**glmmPQL**: PQL (Penalized Quasi-Likelihood) を用いた一般化線形混合モデルの推定 (MASS)

#### 一般加法モデル (GAM: Generalized Additive Models)

**anova.gam**: 複数の gam モデルの比較 (Analysis of Deviance 表) (gam)  
**gam.control**: gam モデルの推定のためのコントロールパラメータ (gam)  
**gam**: gam モデルの当てはめ (gam)  
**na.gam.replace**: gam における欠測値の取り扱いの指定 (gam)  
**plot.gam**: gam のための作図 (gam)  
**predict.gam**: gam オブジェクトを用いた予測 (gam)

**preplot.gam**: 直接プロットできる形で gam から要素を抽出 (gam)  
**step.gam**: gam におけるステップワイズモデル選択 (gam)  
**summary.gam**: gam に対する要約情報 (gam)

#### 生存分析

**anova.survreg**: survreg オブジェクトに対する分散分析表 (survival)  
**clogit**: 条件付きロジスティック回帰 (survival)  
**cox.zph**: Cox (コックス) 回帰に対する比例ハザード仮定の検定 (survival)  
**coxph**: 比例ハザード回帰 (survival)  
**coxph.detail**: Cox モデル当てはめの詳細 (survival)  
**coxph.rvar**: Cox モデルに対するロバストな分散 (survival)  
**ridge**: リッジ回帰 (survival)  
**survdiff**: 生存曲線の違いの検定 (survival)  
**survexp**: 期待サバイバルの計算 (survival)  
**survfit**: 打ち切りデータに対する生存曲線の計算 (survival)  
**survreg**: パラメトリック生存モデルのための回帰 (survival)

#### 決定木・回帰木

**cv.tree**: ツリーの複雑性を決定するための交差検証 (tree)  
**deviance.tree**: tree オブジェクトから逸脱度を抽出 (tree)  
**labels.rpart**: rpart オブジェクトのラベル情報 (rpart)  
**meanvar.rpart**: rpart オブジェクトの平均 - 分散プロット (rpart)  
**misclass.tree**: 決定木における誤分類数 (tree)  
**na.rpart**: rpart オブジェクトにおける欠測値の取り扱い方の指定 (rpart)  
**partition.tree**: ツリーモデルの分割図の作成 (tree)  
**path.rpart**: rpart オブジェクトの指定し

たノードへのパスの表示 ( rpart )  
**plotcp**: rpart による推定に対する複雑性の指標 ( cp ) の表示 ( rpart )  
**printcp**: rpart オブジェクトに対する cp 値の表の表示 ( rpart )  
**prune.misclass**: 誤り率に基づくコスト - 複雑性剪定 ( tree )  
**prune.rpart**: rpart オブジェクトのコスト - 複雑性 ( Cost-complexity ) 剪定 ( rpart )  
**prune.tree**: ツリーオブジェクトのコスト - 複雑性に基づく剪定 ( tree )  
**rpart**: 再帰的分割と回帰木モデルの当てはめ ( rpart )  
**rpconvert**: rpart オブジェクトをアップデート ( rpart )  
**rsq.rpart**: 異なるスプリット数に対する近似  $R^2$  のプロット ( rpart )  
**snip.rpart**: rpart オブジェクトの部分木の刈り取り ( rpart )  
**solder**: プリント回路基板への部品のはんだ付けデータ ( rpart )  
**text.tree**: 樹状図にテキスト情報を付与 ( tree )  
**tile.tree**: 樹状図に棒グラフを付与 ( tree )  
**tree.control**: ツリーのためのパラメータを選択 ( tree )  
**tree.screens**: 樹状図を表示するためにスクリーンを分割 ( tree )  
**tree**: 決定木・回帰木の当てはめ ( tree )

#### ベータ回帰

**betareg**: ベータ回帰モデルの推定 ( betareg )  
**plot.betareg**: ベータ回帰モデルの診断プロット ( betareg )  
**predict.betareg**: ベータ回帰モデルによる予測 ( betareg )  
**residuals.betareg**: ベータ回帰モデルに対する残差 ( betareg )  
**summary.betareg**: ベータ回帰の要約情報 ( betareg )