

## TRIM Q&A

→TRIMのHELPにあるQ&Aの一部を日本語に訳しました（原文とは文章の順番が変わっているところがあります）。本原稿に誤訳等がありましたらお知らせください。

→求めている情報が見つからない時、より詳しい情報を確認したいときは「Q&A TRIM HELP」ファイルもしくはTRIM内のHELPをご覧ください。それでも解決しない場合には [Statistics Netherlands \(http://www.cbs.nl/en-GB/menu/themas/natuur-milieu/methoden/trim/default.htm?language\\_switch=on\)](http://www.cbs.nl/en-GB/menu/themas/natuur-milieu/methoden/trim/default.htm?language_switch=on) 内にある、「Contact information」などでソフト作成元に直接お問い合わせください。

\*TRIMはStatistics Netherlandsがフリーソフトとして公開しているWindows用の統計ソフトです。

### \*データについて

**Q1**：長期間の調査の途中で追加された調査地について、調査を行っていないなかった年の値（ex. カウント数）は欠損値にしてよいのですか？

**A1**：大丈夫です。そのような調査地の調査を行っていないなかった年の値は欠損値を表す数値（ex. -1）に置き換えて分析を行ってください。

**Q2**：全ての種のindexを1回のrunで分析することは可能ですか？

**A2**：TCF fileを作成し(マニュアル page 23)、Shift keyを押しながら全てを選択して分析を行うか、AccessやDelphiといったソフトウェアを土台とした自動化システムにTRIMを組み込むことで可能です。

**Q3**：データ内の欠損値の割合はどこまで許されますか？

**A3**：50%以上の欠損値のあるデータは多くの場合推奨されません。また、20%以上の欠損のあるデータもあまり推奨しません。しかし、統計的なモデルのfitに依存する部分もあるため、欠損値の割合についてのルールを示すことは難しいのが実際です。データにより多くの欠損値が含まれる場合、統計的モデルは推定された値により依存します。もしもモデルがfitし、推定された数値が実際の数値に近い数値として予想された場合は、欠損値を相当高い割合で含むデータも許容されます。

**Q4**：欠損値はモデルが推定される前に補完されるのですか？

**A4**：いいえ、そうではありません。観察値を用いてモデルを推定し、それからそのモデルを用いて欠損値を予測しています。欠損値が予測値に置き換えられた完全なデータセットに基づいてindexが計算されています。

### \*Modelについて

**Q5**：Liner trend modelとはなんですか？

**A5**：1年もしくはそれ以上の年に値が少ないと、Time effect modelではデータが扱えない場合が生じます。その場合は、Liner trend modelを選択し、問題となった年を除いた全ての年をchange pointとして選択することでTime effect modelと同様の分析を行うことができます。また、Liner trend modelでは注目したい年の前後の変化を検証することや、Stepwise法を用いて大きな変化を示した年を探索することができます。ただし、change pointとしていずれの年も選択しなかった場

合（初期設定ではいずれの年も選択されていません）は直線的な変化を前提に分析が行われてしまうので、注意が必要です。

**Q6** : No time effects model とはなんですか？

**A6** : このモデルでは基準とする年と等しい index を全ての年に作成します。モデル選択をするとき以外は通常使用しません。

**Q7** : Covariate (共変量)とはなんですか？

**A7** : 例えば、森林や農地といったハビタットタイプなど、調査地の性質を表したものです。共変量を含んだ状態で分析を行うと、欠損値は共変量のカテゴリごとにそれぞれ推定されます。傾向などもカテゴリごとに示されます。

**Q8** : Overdispersion (過分散)とはなんですか？また、index に対してどのような影響があるのですか？

**A8** : TRIM では、データが Poisson 分布に当てはまることを前提としています。過分散は Poisson 分布の逸脱の程度を示しており、index そのものではなく、index の標準誤差と他のパラメーターに影響を及ぼします。モデルが fit しないことに起因して過分散が高くなりますが、このことはよりよいモデルでは過分散が減少することを暗に意味しています。しかし、群れをなして現れる種の研究の特性として過分散は高くなる可能性があるものです。その場合、よりよいモデルを探すことによっても過分散が減少することは望めません。

**Q9** : もしもデータが Poisson 分布をしていなかったら？

**A9** : これについては問題にする必要はありません。TRIM のモデルでは Poisson 分布からの逸脱は過分散として考慮されています。最も良い戦略としては、推定の過程に常に過分散を組み込んでおくことです；さもなければ標準誤差が深刻に過小評価されてしまうかもしれません。

**Q10** : “unexpected value site...”というエラーメッセージが出たのですが...？

**A10** : データの形式が TRIM で扱えるものとは異なっていませんか？また、TRIM の操作中に表示されるダイアログボックスに与える年や共変量の数値などに打ち間違いはありませんか？

**Q11** : “model error: zero observations for change point at time 3”というというエラーメッセージが出たのですが...？

**A11** : TRIM では分析に使用する各年の値が正の値をとっていることが必要です。この理由でモデルが推定できなかった場合にエラーメッセージが表示されます。また、同様のデータで Liner trend model を行い、change point として time 3 を選択していた場合にも同様のエラーメッセージが表示されます。この3番目の年のデータに欠損値が多いもしくは数値が少ない時は、この年を除いて分析を行ってみてください。

**Q12** : “invalid floating point operation”, “singular matrix”, “diverging parameters”というというエラーメッセージが出たのですが...？

**A12** : モデルが推定できなかった理由として、データが少ない、モデルが複雑すぎる、などが考えられます。共変量を用いている場合はその数を減らす、もしくは change point を減らす(Time effects model で行っていた分析は Liner trend model でやり直す)、Serial correlation のチェックを外すなど、よりシンプルなモデルにしてから分析をやり直してください。

**Q13** : どの年を base time year (基準の年で index が 1 となる) とすべきでしょうか？

**A13** : 通常は調査期間の最初の年を base year として選択します。もしも最初の年には対象の種がいなかったのであれば、他の年、例えば調査の最終年を base year とすべきでしょう。同様に、最初の年のデータがわずかしかない場合も他の年を base year とした方がいいかもしれません。Base year の index の標準誤差を 0 と定義することによって、base year index に関連した全ての誤差は他の年の index の標準誤差にゆだねられます。Base year のデータがわずかである場合は、すべての index の標準誤差が大きく予想されます。

**Q14** : データに新しく年を加えると最初の頃の年の index が変化します。そうですね？

**A14** : はい。新たな年を加えたデータは全ての年のモデルの推定に影響を及ぼします。一般的には、最初のころの年の index は変化しにくいものです(時には相当に変化することもあります)。

**Q15** : もしもモデルが fit しなかったら？

**A15** : モデルが棄却された場合、調査地間の違いを適切に表現する共変量を組み入れて分析をやり直してみてください。新たに作成されたモデルそれ以前のモデルの AIC を比較することによって fit の程度が向上したかを判断することができます(AIC が小さい方がより fit したモデルです)。

**Q16** : モデルが fit しない場合でも index は信頼できるのですか？

**A16** : 概して問題はありません。Fit していない場合、推定値と index の質は制限されます。しかし、大抵の場合、index の制限された質は index の大きな標準誤差として表されます。Index はたいての場合信頼出来ますが、結果の適切な解釈には index の標準誤差も考慮する必要があるでしょう。同様に、モデルが fit していなくても Overall slope と Wald test には信頼性が残されます。

**Q17** : 系列相関が index に与える影響はなんですか？

**A17** : 系列相関は連続した time-point (年)におけるカウントの独立性を表し、正にも負にもなりません。データがとても少ない時を除いては、系列相関が index に対して持つ影響は小さいといえます。系列相関を考慮することによって高頻度に大きな過分散が生じます。このことはモデルに含めるべきであり、そうしないと過小評価された標準誤差が予想されます。

**Q18** : 分析の前にデータを変換する理由は何ですか？

**A18** : 通常の統計的分析、例えば線形回帰のような分析の場合、データは正規分布することが前提となります。この前提に対応するためには対数変換が必要です。0 の log をとることを避けるために、1 つの方法として全ての数値に定数 (例えば 1) を加えます。しかし、残念なことにかくさんの 0 を含むデータに対しては、対数変換は正規分布の前提に対応する手段として十分なものではなく、加えた定数の規模に依存するでしょう。GLM-model は通常のアプローチに対してよりよい代替案を提供し、カウントデータの分析において標準的な方法となってきました。正規分布の前提は GLM-model ではユーザーの選択、たとえば Poisson や Multinomial などに置き換えられます。これらのモデルの適応に対しては、素データの変換はもはや必要ありません。

## \*Index などについて

**Q19** : Model indices と Imputed indices はどちらが好ましいのですか？

**A19** : Model-based indices は全てのサイトにおけるモデル予測の総和から計算されています。それぞれのサイトのモデル予測は統計モデルに基づいています。サイトごとの Imputed value は観察値、さらに欠損値、モデル予測です。これらの数値の全てのサイトにおける総和から Imputed

indices が算出されます。多くの場合、Model-based indices と Imputed indices はほとんど異なりませんし、Imputed indices の標準誤差は Model indices の標準誤差に近い値をとります。しかし時々 Model-based indices と Imputed indices は異なります。Model-based indices は、モデルが適度に fit した場合には Imputed count よりもいくらか安定的であり、その場合は好ましいと考えられます。Imputed indices はカウントにより近く、とりわけ線形モデルに関しては時間においてより現実的です。我々は Imputed indices を推奨しています。

**Q20 :** Additive と Multiplicative パラメーターはどちらが好ましいのですか？

**A20 :** これらは、同一の推定に対する異なる表記です。Additive パラメーターは Multiplicative の自然対数です。したがって、どちらを使っても問題はありませんが、Multiplicative パラメーターの方が理解はしやすいでしょう。Multiplicative の傾向は年間のパーセント変化の平均を反映しています。もしも傾向が 1 と等しかったならば、傾向はありませんし、もしも傾向が 1.08 であったならば、1 年に 8% 増加していることとなります。もしも傾向が 0.93 であったならば、1 年に 7% 減少していることとなります。

**Q21 :** Overall slope が複数示されますが、どれを用いたらよいのですか？

**A21 :** TRIM version 3.5 では、2つの異なる傾向を算出します。1つ目は Model indices に基づいた回帰線の傾き、2つ目は Imputed indices に基づいた回帰直線の傾きです。両方の回帰線共に index と year の間の切片を伴った直線的な関係を示しています。2つの indices については Q19 とその回答を参照してください。