

斜面方位を説明変数とする回帰モデルについて

荒瀬輝夫¹・岡野哲郎¹・熊谷真由子¹・内田泰三²

¹信州大学農学部, ²九州産業大学工学部

Recommended Regression Models Using Slope Direction as a Continuous Explanatory Variable

Teruo ARASE¹, Tetsuo OKANO¹, Mayuko KUMAGAI¹ & Taizo UCHIDA²

¹Faculty of Agriculture, Shinshu University

²Faculty of Engineering, Kyushu Sangyo University

Summary: Slope direction is an important environmental factor in various field studies, but it is difficult to employ as an explanatory variable in regression analysis because of its periodicity. In the present study, actual data was used to assess the goodness of fit of three different regression models using slope direction as an explanatory variable. Results show that a non-linear regression of $y = a \sin(\theta + b)$ can be transformed into a linear multiple regression with $\sin \theta$ and $\cos \theta$. A multiple regression, with $\sin \theta$ and $\cos \theta$, should be applied when the two slope directions bearing a maximum or minimum are assumed to be exactly opposite to each other. Conversely, a cubic-curve regression should be applied when the two slope directions are not assumed to be opposite. Presumed errors of estimation were almost equal in the multiple and cubic-curve regressions. Furthermore, p -values were almost equal for both the multiple and cubic-curve regressions when there was little variation in the data, but the p -value for the cubic-curve regression dramatically increased as the variation in the data increased.

キーワード: 斜面方位, 三角関数, 3次曲線, 極値, 精度

Key words: Slope direction, Trigonometric function, Cubic curve, Extremal value, Precision

1. はじめに

斜面方位は,山地における様々なフィールド調査で記録すべき重要な環境要因の1つである。とくに高山帯では,斜面方位は標高よりも植物の分布・生育に影響することが報告されている^{4, 7)}。しかし,斜面方位と他の変量との関連性を知りたいとき,とくに決まった統計的手法はない。理由は,斜面方位を角度で表現すれば比例尺度の連続量にできるものの,角度(ラジアン)は 2π で0に戻るためである。角度 $0 \sim 2\pi$ (北から始めて,時計回りに東,南,西を経て北に戻る)を横軸として描かれた散布図に対し,両端で推定値と変化率が同じになるような推定曲線を当てはめる必

要がある。すなわち,周期的に変化する曲線を想定しなければならない。回帰直線は適用できないので,相関係数は使えないことになる。

斜面方位を要因に取り上げた調査研究は数多くあるが,斜面方位を横軸としたグラフを示して,視覚的に読み解くことが多いようである。説明変数として扱った例として,ア)ダミー変数による重回帰分析³⁾,イ)数量化理論 類および 類といった多変量の予測モデル^{5, 8)},ウ)三角関数を用いた非線形回帰⁶⁾などが挙げられる。ア)とイ)では,斜面方位は名義尺度として扱われている。ア)は,川の左岸と右岸の2種類しかないような場合に有効であろう。イ)は,4方向または8方

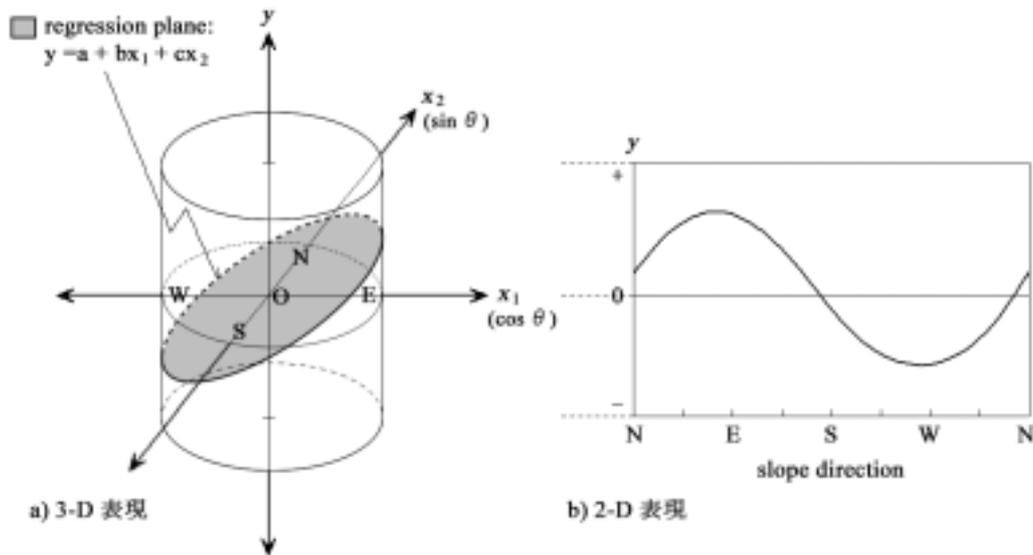


図-1 $\cos \theta$ と $\sin \theta$ による重回帰イメージ

向といった水準数であれば有効であるが、カテゴリ数の増加は説明変数の増加を意味し、信頼性を保つために多くのデータが必要となる。ウ)では斜面方位は比例尺度の連続量として扱われ、周期関数を適用するので理に叶っているが、非線形の最適化なので手計算での解析は困難になる。このように、斜面方位を重回帰の説明変数にする簡便で理に叶った方法は、現在のところ確定されていない。

そこで本報では、斜面方位を説明変数とする重回帰モデルについて、いくつかの方法を取り上げて比較することにした。

2. 重回帰モデル

斜面方位を説明変数とする重回帰モデルとして、 $\sin \theta$ 、 $\sin \theta$ と $\cos \theta$ による重回帰、3次多項式による曲線重回帰の3種類を取り上げることとした。いずれも、斜面方位を角度で表現した θ の範囲は、 $0 \leq \theta < 2\pi$ である。

まず $\sin \theta$ は、周期 2π になる関数として選んだもので、重回帰モデルは以下のようなになる。

$$y = a \sin(\theta + b) \cdots (1)$$

ここで a, b は定数である。周期 2π なので、 a, b にかかわらず θ の範囲の両端で重回帰曲線は連続する。

$\sin \theta$ と $\cos \theta$ による重回帰は、斜面方位の角度を \sin 成分と \cos 成分に分解して説明変数としたもので、重回帰モデルは

$$y = a + b \sin \theta + c \cos \theta \cdots (2)$$

となる。ここで a, b, c は定数である。この場合の空間的なイメージを図1に示した。 y が最大、最小となるときの θ は、原点 O を中心として対称になるので、互いに正反対の方向となるのは自明である。また、説明変数を $x_1 = \cos \theta, x_2 = \sin \theta$ とおくと、点 (x_1, x_2) は図中の円筒の側面上にあるという制約で動いており、(2)式は重回帰平面上の楕円となる。そのため、(2)式は、非線形ではなく線形の重回帰として推定することができる。

3次多項式による曲線重回帰は、 $\sin \theta$ を Taylor 展開すると $\theta, \theta^3, \theta^5, \dots$ の和として表現されることから取り上げた。

$$y = a + b\theta + c\theta^2 + d\theta^3$$

これに、 $\theta = 0$ のときと $\theta = 2\pi$ のときの y, y' が一致するという条件を加えて解くと、 $a_1 = 2a_3\pi^2, b = -3d\pi$ となり、整理すると

$$y = a + d \cdot \theta(\theta - \pi)(\theta - 2\pi)$$

となる。ただしこれでは極値をとる θ が常に一定になってしまうので、角度 α (α は定数) の分ずらせるように、次の変数

$$X = \theta - \alpha \quad (\alpha < \theta \text{ のとき})$$

$$2\pi - \alpha + \theta \quad (\theta < \alpha \text{ のとき})$$

を導入すると、重回帰モデルは

$$y = a + d \cdot X(X - \pi)(X - 2\pi) \cdots (3)$$

となる。 y が極値をとるときの X は、 $y'' = 0$ を解くと、

$$X = \left(1 \pm \frac{\sqrt{3}}{3}\right)\pi$$

となる。すなわち、 y が最大・最小となる方向は

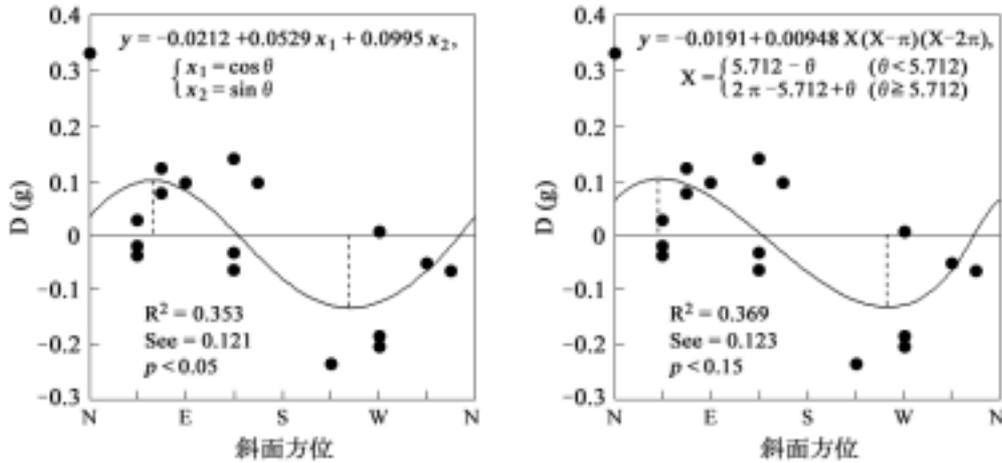


図-2 斜面方位を説明変数とする回帰モデルを検証データに適用した分析結果

$(2\sqrt{3}/3)\pi$ (約 207.8°) ずれるので, 正反対の方向にはならない。

ここで, (1) 式を変形すると,

$$y = a(\sin \theta \cos b + \cos \theta \sin b) \\ = a \cos b \cdot \sin \theta + a \sin b \cdot \cos \theta \quad \cdots (1')$$

このとき $a \cos b$, $a \sin b$ は定数なので, (1) 式による回帰モデルは (2) 式による回帰モデルで表現できる。したがって, 以下では, $\sin \theta$ と $\cos \theta$ による重回帰: (2) 式と, 3 次多項式による曲線回帰: (3) 式による回帰モデルを比較することとした。

3. 検証のためのデータ

回帰モデルの検証のために, 長野県内における木本性つる植物マツブサの果粒重についてのデータ¹⁾を用いた。平均果粒重に対し, 緯度, 標高および斜面方位 (\sin 成分と \cos 成分に分解) との関係が有意であった¹⁾。そこで, 緯度と標高の影響を取り除いた平均果粒重 (D) を求め, これを斜面方位により (2) 式と (3) 式の回帰モデルで推定することとした。なお, D は, 緯度と標高を説明変数とする重回帰式による平均果粒重の推定値 \hat{y} と実測値 y との差 ($D = y - \hat{y}$) である。

回帰モデルの係数の推定は, 正規方程式を立てて行列計算を行なうことによって求めた。有意性の判定は, 残差平方和をもとにした F 検定により行なった。演算は表計算ソフト (Microsoft Office Excel 2003) によった。

4. 結果と考察

$\sin \theta$ と $\cos \theta$ による重回帰: (2) 式と, 3 次多

項式による曲線回帰: (3) 式による分析結果は, 図 2 のようになった。決定係数は (2) 式で $R^2 = 0.353$, (3) 式で 0.369 と後者のほうがやや高い値であった。推定誤差は (2) 式で $See = 0.121$, (3) 式で 0.123 とほぼ同等であった p 値は (2) 式で 0.05 以下となったが, (3) 式では 0.10 を越えていた。

また, グラフを比較すると, (2) 式による回帰曲線は極値となる斜面方位 (NEE, SWW) が正反対の方向にあるのに対し, (3) 式による回帰曲線では極値となる斜面方位がともに北よりにずれ, 互いに正反対の方向にならないことが確認された。したがって, 最大値・最小値となる斜面方位が正反対の方向になることが想定される場合には $\sin \theta$ と $\cos \theta$ による重回帰: (2) 式が妥当であり, 正反対にならない場合には 3 次多項式による曲線回帰: (3) 式を適用することが妥当であるといえる。

ここで, 検証データ (D) のみでは (2) 式と (3) 式による回帰の特性が捉えきれないので, データのばらつきがより小さいまたは大きい場合について検討することにした。 D について, (2) 式を当てはめたときの残差とすると, D は推定値と残差の和で表現される。残差を拡大または縮小した仮想データ (残差 100%: 元データ, 残差を 0~200% に調整) を作成して, (2) 式および (3) 式による回帰分析を行った。その結果, 推定精度を示す See と p 値の変化は図 3 のようになった。残差の大きさが変化しても, See は (2) 式と (3) 式とでほぼ近接した値となるが常に (2) 式 < (3) 式であった。また, p 値も常に (2) 式 < (3) 式

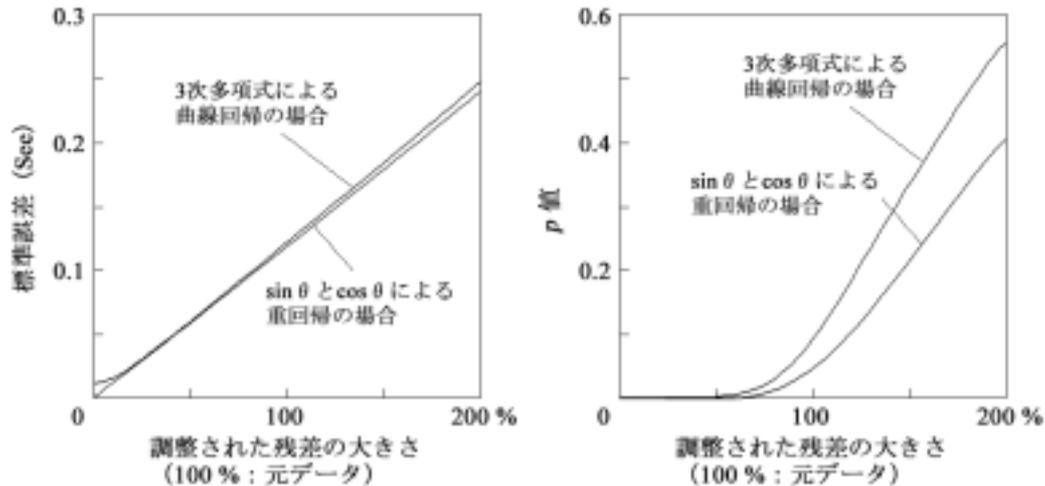


図-3 残差の大きさを变化させたときの推定精度の変化

であり、残差が大きくなると(3)式ではより急に p 値が増大し、残差が小さくなると(3)式のほうが急に減少して、(2)式の場合とほとんど変わらなくなることが読み取れた。

この原因として、回帰モデルのパラメータ数が(2)式では3、(3)式では4(未知数が2つだけなのは両端で連続という条件を付けているため)であることが考えられる。Seeは残差分散に基づくので、残差の自由度がより大きい(2)式のほうが有利であろう。また、 p 値については、3次多項式のほうがモデル式の自由度が1つ多いことでより柔軟な曲線近似ができるものの、データのばらつきが大きくなると、高次の項を含むためにルンゲの現象²⁾のような当てはまりの悪さを露呈するのではと考えられる。

5. まとめ

本報では、斜面方位を説明変数とする回帰モデルについて、検証データを用いて比較・検討した。

1. $\sin \theta$ による非線形回帰は、 $\sin \theta$ と $\cos \theta$ による線形の重回帰に変形できる。
2. $\sin \theta$ と $\cos \theta$ による重回帰と、3次多項式による曲線回帰をモデル式として取り上げた。前者は極値となる方位が正反対の方向になる場合、後者は正反対にならない場合に適用すべきである。
3. 推定誤差は両者でほぼ同等であった。データのばらつきが小さいときには両者で p 値はほぼ同等であったが、ばらつきが大きくなると、後者では p 値が急に増加した。

引用文献

- 1) 荒瀬輝夫・熊谷真由子・内田泰三(2012)長野県におけるマツブサ(*Schisandra repanda* (Sieb. et Zucc.) Radlk.)の結実と自生地の立地環境との関係. 信州大学農学部AFC報告 10:印刷中
- 2) de Boor, C. (2001) A practical guide to splines (revised edition). Springer-Verlag, Now York. 346pp.
- 3) 福島慶太郎・徳地直子(2008)皆伐・再造林施業が渓流水質に与える影響:集水域単位で林齢の異なるスギ人工林を用いて. 日本森林学会誌 90:6-16
- 4) 菊地多賀夫(2003)周氷河地形と高山植生のかたち. 遺伝 57:44-47
- 5) 三浦 覚・松本陽介・真下育久(1986)台風15号(昭和56年)により東京大学北海道演習林に発生した風害の要因解析. 東京大学農学部演習林報告 73:139-160
- 6) 岡野哲郎・古賀信也(1995)アカエゾマツの生育と立地環境について 九大北海道演習林のアカエゾマツ人工林における解析. 九州大学演習林報告 72:45-52
- 7) Raffl, C., Mallaun, M., Mayer, R. and Erschbamer, B. (2006) Vegetation succession pattern and diversity changes in a glacier valley, Central Alps, Austria. Arctic, Antarctic, and Alpine Research 38: 421 - 428
- 8) 山本幸一(1985)木造校舎の木製窓枠の腐朽要因. 木材保存 11:97-102

(原稿受付 2012.2.29)