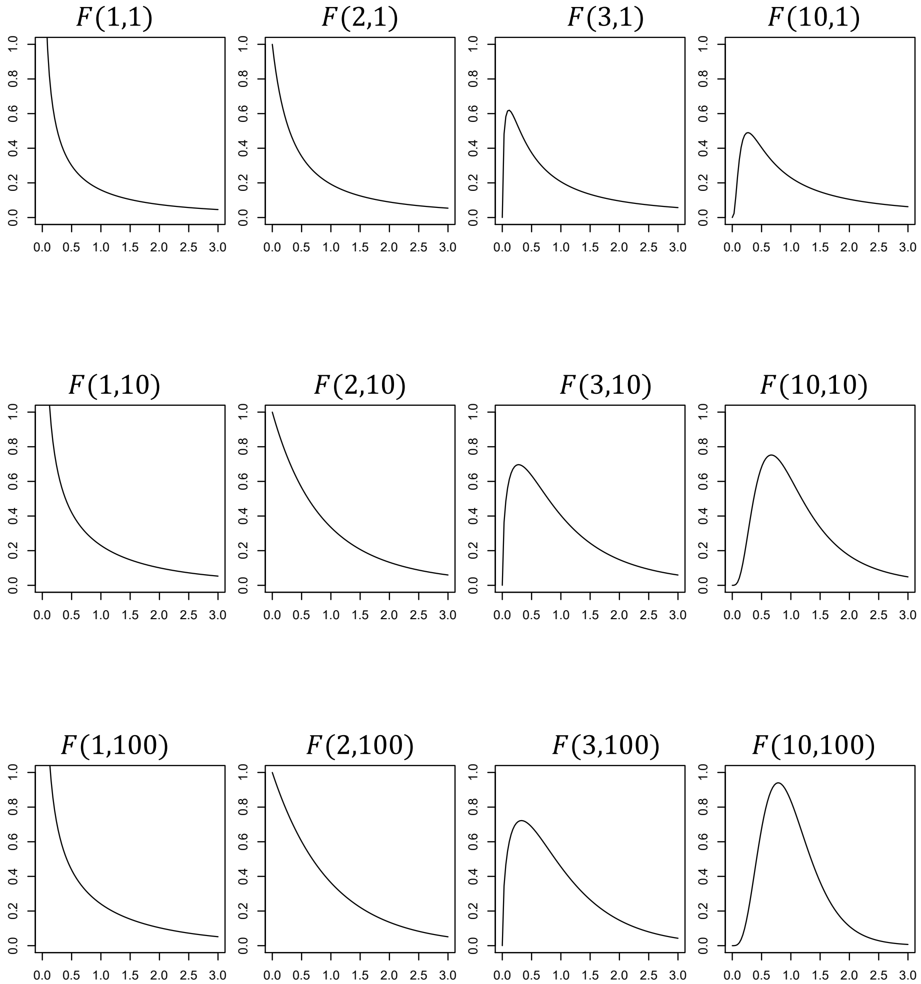
# 資料

## 資料5-1　　F 分布


$F(1,1)$
$F(2,1)$
$F(3,1)$
$F(10,1)$
$F(1,10)$
$F(2,10)$
$F(3,10)$
$F(10,10)$
$F(1,100)$
$F(2,100)$
$F(3,100)$
$F(10,100)$

## ★★★ 演習問題 ☾☀

### 基礎問題

### 1 記法

次の記法がどのようなものを指すのか、簡潔に説明しなさい。

(1) $R^2$
(2) $R_{adj}^2$
(3) $\mathrm{r_{yx_2}}$
(4) $\mathrm{r_{y(2|1)}}$
(5) $\mathrm{r_{y2|1}}$
(6) 単回帰分析における $\beta_1$
(7) 重回帰分析における $\beta_1$

### 2 数式による表現 ［✎ノート1］

重回帰分析で母集団に想定されているモデルについての表現に関して次の問いに答えなさい。ただし、独立変数は$x_i$と$x_2$の二つを考えているとする。次の文章のうち正しいものには○、誤っているものには×と書きなさい。

(1) $i$ 番目の $x$ の値が与えられた時の $y$ の値は、$x$ の値に関係なく同じ正規分布に従うため、次のように表現できる。

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} \quad \text{かつ} \quad y_i \sim N(0, \sigma^2)$$

(2) $i$ 番目の $y$ の値は、回帰平面からの誤差をはらんで生成されると想定されるため、次のように表現できる。

$$y_i \sim N(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}, \sigma^2)$$

### 基本問題

### 3 図を用いたモデルの表現 I ［✎ノート1］

次の条件に適うモデルを数式で表し、また、パスモデルによって表現しなさい。

(1) 独立変数$x$で従属変数$y$の値を予測する単回帰モデル
(2) 二つの独立変数$x_1$，$x_2$で従属変数$y$の値を予測する重回帰モデル
(3) 三つの独立変数$x_1$，$x_2$，$x_3$で従属変数$y$の値を予測する重回帰モデル

## 4　図を用いたモデルの表現Ⅱ　　　　　　　　[✎ノート１]

次の条件に適うモデルを数式で表し、また、グラフィカルモデリングによって表現しなさい。ただし、標本サイズは $n$、独立変数の値が与えられた時に従属変数が従うばらつきを表した母集団の分散は $\sigma^2$ と表記することとする。

(1) 独立変数 $x$ で従属変数 $y$ の値を予測する単回帰モデル
(2) 二つの独立変数 $x_1$，$x_2$ で従属変数 $y$ の値を予測する重回帰モデル
(3) 三つの独立変数 $x_1$，$x_2$，$x_3$ で従属変数 $y$ の値を予測する重回帰モデル

## 5　発展的なモデルたち　　　　　　　　[✎ノート２]

次の文章を読んで後の問いに答えなさい。なお、文章１から文章４までは一続きの文章であるが、ここでは、設問の関係で四つに分割して提示をしている。

### 文章１

重回帰モデルなど、何らかの統計的なモデルを立てるということは、独立変数と従属変数の関係を考えるということである。だが、独立変数と従属変数との関係とは、独立変数が分かれば、従属変数が簡単に予測できるというようなシンプルなものだけとは限らない。

　例えば、ある独立変数は別の変数を経由して従属変数に影響を与えるということがある。このような間に入って影響を伝える変数のことを[　A　]、そして、[　A　]を経由して伝わる効果のことを[　B　]と呼ぶ。例えば、[　C　]という独立変数が[　D　]という従属変数に影響を与えるときには、[　E　]という[　A　]となり、影響を伝えていると考えることができる。

問５　空欄Ａ，Ｂを埋めなさい。
問６　この文章で説明されている概念を説明するのに適切な例を考え、空欄ＣからＥに入る言葉を答えなさい。なお、授業で説明された例とは異なるものを自分で考えて答えること。

### 文章２

　独立変数と従属変数が取り持つ関係はそれだけではない。例えば、二つの独立変数がモデルに組み込まれている次の式で表されるような関係を考えてみよう。

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$$
$$\varepsilon_i \sim N(0, \sigma^2)$$

このとき、独立変数$x_1$にかかる係数$\beta_1$はこのとき[　F　]を表す。このモデルでは、$\beta_2$の値に関わらず、$\beta_1$の値は一定であり、逆もまたしかりである。ところが、二つの独立変数が交互作用を持つモデルというものも考えることができる。これは、次のような式で表されるような状況を示している。

[　G　]
$$\varepsilon_i \sim N(0, \sigma^2)$$

一般に$x_1$と$y$の関係が第三の変数$x_2$の値によって変化するときに、この第三の変数を[　H　]と呼ぶが、この交互作用を組み入れたモデルはこの[　H　]の存在をモデル化した式であるとも見なすことができる。

問７　空欄Ｆには係数$\beta_1$に対する説明が入る。適切な言葉でこの空欄を補いなさい。
問８　空欄Gに入る数式を答えなさい。
問９　空欄Hに入る言葉を答えなさい。

## 文章３

重回帰分析では、その重要な仮定に標本抽出の[　I　]性が挙げられる。しかし、実際の研究では、標本たちにグループが存在しこの[　I　]性が成り立たない場合が多い。このような欠点を克服するために、むしろこの階層性（グループ化）を念頭に置いたモデルを立てて対応することがあり、そのようなモデルを階層モデルなどと呼んだりする（呼び名には複数のものがある）。例えば、[　J　]という従属変数$y$を、[　K　]という独立変数$x_1$で予測しようとするとき、標本たちには[　L　]によってグループが形成されるであろうからこれをモデルに組み込むとよいであろう。

問１０　空欄Ｉに入れるのに適切な言葉を答えなさい。
問１１　空欄ＪからＬに入れるのに数式を答えなさい。
問１２　この文章で説明されている概念を説明するのに適切な例を考え、空欄ＪからＬに入る言葉を答えなさい。なお、授業で説明された例とは異なるものを自分で考えて答えること。

文章４

独立変数と従属変数の両方に影響を与えている変数のことを[　M　]と呼ぶ。例えば、[　N　]という独立変数と[　O　]という独立変数から[　P　]という従属変数を予測する重回帰モデルを作ることはできる。しかし、[　N　]と[　O　]の間には強い[　Q　]関係が予測されるので、結果として、最小二乗推定によってとりあえずの係数の推定値を出すことはできるのだが、その推定値はサンプルを取り換えるごとに、大きくばらついてしまい、とても不安定なものとなることが予想される。このような状況を[　R　]の問題が生じていると表現する。

このような問題に対処する一つの方法として、この[　N　]と[　O　]に強い[　Q　]を生み出す要因そのものをモデルに組み込んでみるというアプローチが考えられる。例えば今回のケースでは、[　S　]という[　T　]要因を統計モデルに組み込んでみると、適切な推定を行うことができ、解も安定することが予想される。このように、研究を行う際には、独立変数同士の[　Q　]関係を調べ、適切に対応することに極めて大きな意義がある。

問１３　空欄 M と Q、R、T に入れるのに適切な言葉を答えなさい。

問１４　この文章で説明されている概念を説明するのに適切な例を考え、空欄 N から P、そして S に入る言葉を答えなさい。なお、授業で説明された例とは異なるものを自分で考えて答えること。

## ６　偏回帰係数の解釈 I　[✎ノート３]

次の文章を読み後の問に答えなさい。

私たちが行う実際の研究では、多くの場合、複数の独立変数が従属変数と結びついているという状況を想定する。例えば、ある日の大阪市におけるアイスクリームの総売上高を予測したいと思った場合、[　　A　　]といった複数の要因で予測したいという目的があったとする。このような場合に下に示すような統計モデルを立てて推論を行う統計手法を重回帰分析と呼ぶ。

（モデル１）　$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi} + \varepsilon_i$

このモデル１では、各要因（$x_{1i}, x_{2i}, \cdots, x_{pi}$）に重みを付けて$y_i$の値が決まるという想定を置いている。重みを付けていることを理解するために下のようなモデル２と比較してみよう。(B)モデル２では、すべての独立変数の値をそのまま足し合わせると、$y_i$の値が決まるという強い仮定を置いている。そうではなく、$x_{1i}$については$\beta_1$倍して、$x_{2i}$については$\beta_2$倍してから足し合わせる、つまりその独立変数の重要さに応じて、重みを付けてから値を足し合わせているのである。

（モデル２）　$y_i = \beta_0 + x_{1i} + x_{2i} + \cdots + x_{pi} + \varepsilon_i$

問１　空欄 A について、自身の経験や知識に基づいて、ここに入れるのに適切だと思われる要因を 3 つ挙げなさい。

問２　下線部 B について、モデル１とモデル２の関係について述べた下記の文章の中で、最も適切なものを一つ選び記号で答えなさい。

① モデル１が成り立てば、モデル２は成り立たないという意味で両者は互いに背反する関係にある。
② モデル１を式変形すると、モデル２になるため両者は同値関係にある。
③ モデル１では検討されていなかった$p+1$番目の変数を、モデル１に投入するとモデル２になるので、後者は前者を拡張した関係にある。
④ モデル１にある制約を導入した時、モデル２が生まれることから、後者は前者の特殊な場合に相当する。

## ７　偏回帰係数の解釈Ⅱ　[✎ノート３]

いま、$i$番目の従属変数の値$y_i$を、その日の天気$x_{1i}$、湿度$x_{2i}$、雲の量$x_{3i}$、気温$x_{4i}$という独立変数から予測するモデルを立てた。これに対して次の疑問が提示されたとする。適切に応答せよ。

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \varepsilon_i$$

偏回帰係数とは、他の変数の影響を排除した下でのその独立変数が一単位変化したときに、平均してどれくらい従属変数が変化するのかを表している。だが、「湿度、気温、雲量というものを排除した天気」を、もはや天気と言えるのだろうか？人間は、湿度や気温、雲量というものも総合して天気というものを認識して体験しているのではないだろうか？

## 8　部分相関と偏相関 I　　　　　　　　　　[✎ノート3]

母集団に下に示したモデル式を想定し、最小二乗法によって点推定を行った。続く下の三つの散布図は、このモデルを想定したときの、(a)独立変数$x_{1i}$と従属変数の散布図、(b) 独立変数$x_{1i}$と従属変数の部分相関係数を求める際の散布図、(c) 独立変数$x_{1i}$と従属変数の偏相関係数を求める際の散布図のいずれかを表している。

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i$$

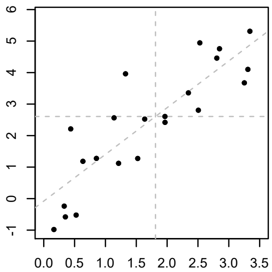


図 1

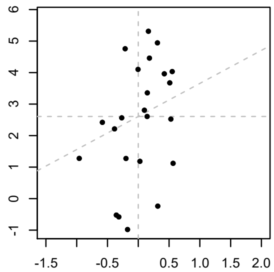


図 2

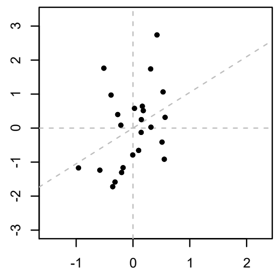


図 3

問１　図 1～3 の中で、偏相関係数を求めるときの散布図を表しているものはどれか、一つ選び記号で答えなさい。

問２　下の三つの表は、それぞれ図 1～3 の散布図に対して引いた直線の切片と傾きを表している。これをもとに上記モデル式に表された$\beta_1$の推定値である$\hat{\beta}_1$を求めなさい。

| 図 1 | 点推定値 | 標準誤差 | t 値 | p 値 |
|---|---|---|---|---|
| 切片 | -0.08 | 0.44 | -0.18 | 0.859 |
| 傾き | 1.48 | 0.21 | 7.13 | 3.72e-07 *** |

| 図 2 | 点推定値 | 標準誤差 | t 値 | p 値 |
|---|---|---|---|---|
| 切片 | 2.60 | 0.41 | 6.42 | 1.86e-06 *** |
| 傾き | 1.05 | 1.03 | 1.02 | 0.32 |

| 図 3 | 点推定値 | 標準誤差 | t 値 | p 値 |
|---|---|---|---|---|
| 切片 | 0.00 | 0.22 | 0.00 | 1.00 |
| 傾き | 1.05 | 0.57 | 1.851 | 0.08 . |

## 9　偏回帰係数の解釈 III　　　　　　　　　　[✎ノート3]

重回帰モデルに含まれるパラメータの点推定値には、非標準偏回帰係数と標準偏回帰係数と呼ばれる二つタイプが存在する。それぞれどのようなものか説明し、どのようなときにどちらのタイプを用いるべきか、具体例を挙げながら説明しなさい。

## 10 決定係数 [✎ノート4]

決定係数について説明した次の文章の中で正しいものには○と、誤りを含むものについては×と記した上で適切に直しなさい。

(1) 決定係数は、そのモデルがどのくらい標本データにフィットしているかを表し指標であり、その値が大きければ大きいほどデータへより適合しているということを表している。
(2) 決定係数は、予測値と観測値の相関係数として定義することもできる。
(3) モデルと観測データとの適合度を示すので、決定係数は異なるモデルたちを比較する基準として用いられる。
(4) 観測値$y$の平方和を$ss_y$、予測値$\hat{y}$の平方和を$ss_{\hat{y}}$とすると、決定係数は$ss_{\hat{y}}/ss_y$と表すことができる。
(5) 観測値$y$の分散を$s_y^2$、予測値$\hat{y}$の分散を$s_{\hat{y}}^2$とすると、決定係数は$s_{\hat{y}}^2/s_y^2$と表すことができる。

## 11 モデル比較 [✎ノート5]

統計モデルを立てた研究を行う際には、モデル比較が求められる。なぜモデル比較を行う必要があるのか、次の質問に対して適切に答えなさい。

> Model B より Model A のほうがより多くの独立変数を含むので、より細かい変数間の関係の検証を可能にしてくれる。このため、わざわざより少ない独立変数を持つモデルを立て、比較をするのは意味がないのではないか？最初から一番たくさんの変数を含むモデルを採択し研究を行ったほうが良いのではないか？

## 12 重回帰分析の運用Ⅰ

従属変数$y$を説明するために、4つの独立変数を用意した。交互作用などは考えないものとする。これらの独立変数をすべて含むモデルは、次のように表される。

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \varepsilon_i$$

この四つの独立変数の一部を含むモデルについても構築し、これらの中で最も良いモデルを選択したい。これを踏まえて次の問いに答えなさい。

問１ 下線部(a)について、モデル比較を行うための指標として各種情報量基準を計算した。合計 16 個のモデルそれぞれの値を示したのが下の図である。X と Y は Mallos's $C_p$ と自由度調整済み決定係数のいずれかを表している。Y が示しているのはどちらか答えなさい。

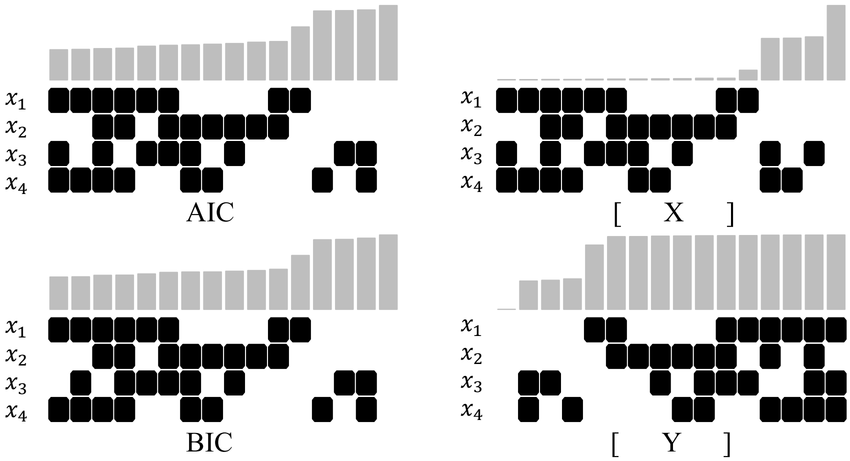



図１ 16 個のモデルと四つの情報量基準の大きさ

※各パネルにおいて、上部の棒グラフは、その情報量基準で測定した 16 個のモデルの情報量基準の大きさを表し、下部のバーコードプロットは、そのモデルに含まれている独立変数の有無を表している（黒い塗りつぶしは、当該変数が含まれていることを示す）。

問２ 図１に示された各種情報量基準は。どのようなモデルを選ぶことを推奨しているか、それぞれ答えなさい。

問３ 上記図１で行ったように考察対象の全てのモデルを比較して、最良のモデルを選択する方法を何と呼ぶか、答えなさい。

## 13 重回帰分析の運用Ⅱ

従属変数$y$を説明するために、4 つの独立変数を用意し、モデル比較の結果、次のモデルが最良のモデルであるという結論を出した。

$$y_i = \beta_0 + \beta_1 x_{1i} \qquad\qquad + \beta_4 x_{4i} + \varepsilon_i$$

この選ばれたモデルのパラメータの点推定値に関する情報は下の表にまとめられている。これをもとに以下の問いに答えなさい。

| | 点推定値 | 標準誤差 | t 値 | p 値 |
|---|---|---|---|---|
| $\beta_0$ | 0.20 | 0.32 | [ A ] | 0.54 |
| $\beta_1$ | 2.04 | 0.19 | [ B ] | 6.98e-09 |
| $\beta_4$ | -3.58 | 0.16 | [ C ] | 6.19e-14 |

| | |
|---|---|
| 残差の標準誤差: | 0.79 |
| サンプルサイズ: | [ E ] |
| 決定係数: | 0.981 |
| 自由度調整済み決定係数: | 0.979 |
| F 値: | 433.6 |
| | （自由度 [ D ], 17） |
| F 値の当該 F 分布における p 値: | 2.59e-15 |

問１ 表中の A から C の値を計算して答えなさい。

問２ 表の下の最後の三行は、重回帰式全体の検定、すなわち「すべての偏回帰係数がゼロである」という帰無仮説を検証するために用いられる検定統計量 F とその結果を表している。この検定統計量 F 値は、自由度が[ D ]と 17 の F 分布に従う。この D に入る数字を答えなさい。

問３ この分析に用いたデータは全部で何個存在しているか、[ E ]に当てはまる数を答えなさい。

問４ 図 1（左）は推定したモデルに基づく(i) 回帰平面、(ii) 95%信頼区間、(iii) 95%予測区間を表している。$x_4 = -0.6$の平面（図中の(a)）に描かれる断面図を右図に示している。信頼区間を表している線を(b)から(f)すべて選び記号で答えなさい。

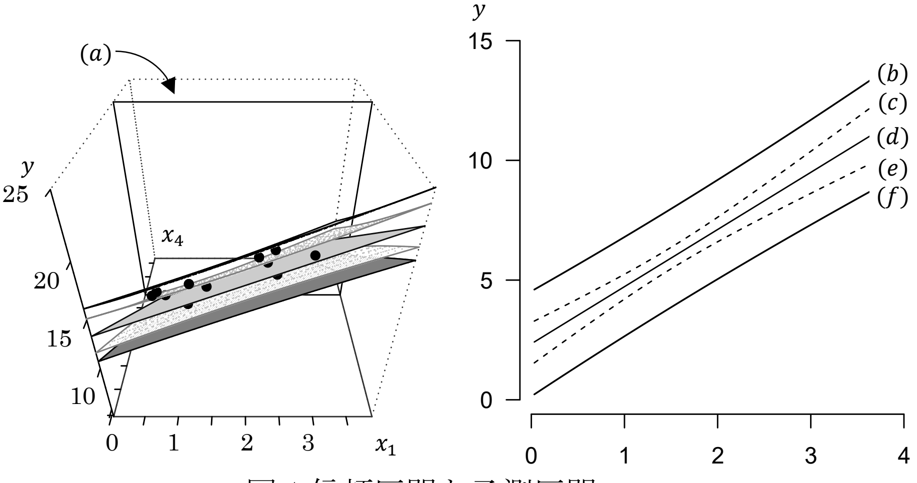



図 1 信頼区間と予測区間

問５ 図 1 の情報を基にすると、$x_1 = 0.5$ かつ $x_4 = -0.6$のとき、$y$の値が 5 を取ることは予測されうることか答えなさい。

問６　有意水準を 0.05 とするとき、$\beta_1$と$\beta_4$の係数の大きさがゼロであるか否かに対して統計的仮説検定を行うとき、どのような結論を出すことになるか、説明しなさい。

問７　今回作成した回帰モデルはデータにフィットしていると言えると言えるかどうか、適切に応答せよ。

問８　回帰分析を行う際には、母集団についていくつかの仮定を設けている。しかし、それが満たされているか否かを検討する必要は当然存在する。下に示されているのは、回帰診断に用いられる様々な統計量を図示したものである。ここからどのようなことが読み取れ、誠実な研究者としてどのような検討を行う必要があるか、分かりやすく説明しなさい

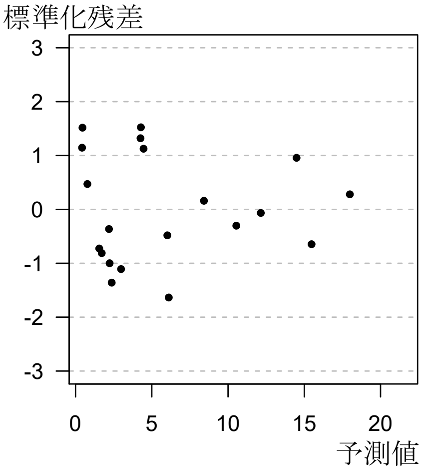


図３残差プロット

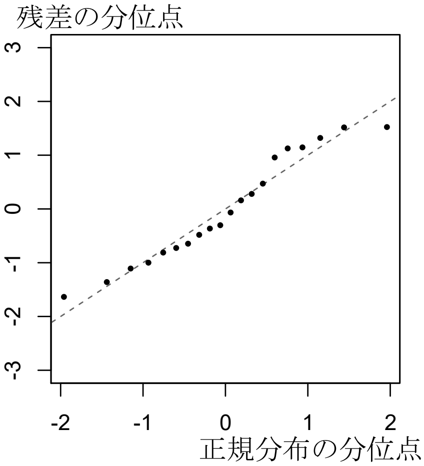


図４　正規 QQ プロット

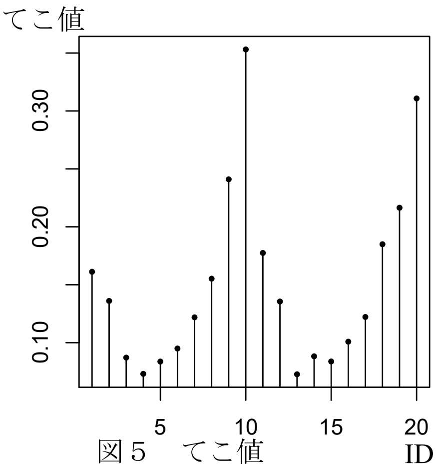


図５　てこ値

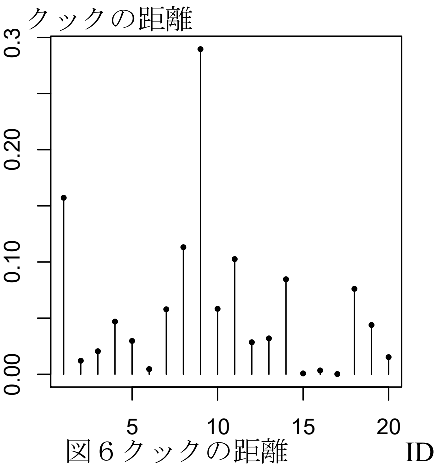


図６クックの距離

※データにおける数字は、データの値につけられた通し番号（ID）を表している（$i$ 番目のデータと呼ぶときの $i$ の値である。なお、一部の図では作図の都合上、小さい値は●で表し数字を省略している。

## 実践問題

### 14 交絡 ［✎ノート2］

専門分野を問わず、研究者は変数（要因）間の関係を議論する。実際には、これまでの研究で解明されてきた知見を組み込んで独立変数を決めて研究を行うわけであるが、このとき重要なことは何か(a)交絡要因となりえるものを見落としていないか、ということを念入りに問い続けることである。(b) 学会や論文の査読においては、このような重要な要因について研究者がきちんと考察を加えていたのかが、精査されることとなる。考えずに査読に出してしまっても、結局は、交絡要因への検討が甘いと突き返されてしまう。せっかくの時間が無駄になってしまうのだから、きちんと準備をしておくことが必要だ。

だが、交絡要因の見落としが致命的な推論へのダメージを与えるのは何も研究だけに限ったことではない。(c)日常生活の推論においても交絡変数を念頭に置かなかったため、ゆがんだ解釈が生まれるということは頻発しているのである。

問１ 下線部(a)について、交絡変数（要因）とはいったいどのようなものか、わかりやすく説明しなさい。
問２ 下線部(b)について、自分の専門分野において交絡が生じるであろう事例を 3 つ考え、簡単に説明しなさい。
問３ 下線部(c)について、日常生活の推論において、本当は交絡が生じてしまっており正確な結論が出せないはずであるにもかかわらず、無理やり結論が導かれているという事例を 3 つ考え、簡単に説明しなさい。

### 15 †部分相関と偏相関Ⅱ ［✎ノート3］

日本語の「彼が来た」という文が He came. と He has come. の二通りに訳し分けられることからも分かる通り、英語の現在完了形と過去形は、その意味するところがよく似ていると言える。このため、「現在完了形は過去形とどのように違うのか」という問いは多くの意味論・語用論の研究者の関心を集め、例えば、現在完了形は「現在において成立する完了状態について述べている（Perfect state theory）」「出来事がいつ起こったかには無関心だ (Indefinite past theory)」「現在より前の存在する時間幅の中に出来事が起きている（Extended now theory）」などこれまでに様々な理論が提案されてきた（Portner 2011）。

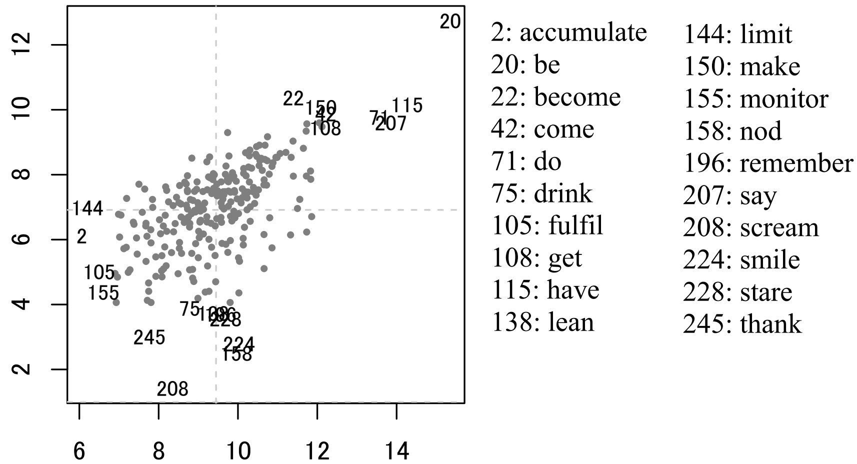
さて、このような先行研究では、「現在完了形」あるいは「過去

形」という構文レベルでの特徴付けに重きが置かれていたが、この構文レベルでの記述・理論は、すべての動詞に当てはまると言えるのだろうか。すなわち、構文レベル（全体）の傾向には還元できない動詞（語彙）間における特殊性（個別性）というものはあるのだろうか。このリサーチクエスチョンに答えるために、頻度の多かった動詞たちがどのような傾向で時制・アスペクトを取っているのかを、コーパスを用いて調査を行った。

まず、コーパスから各動詞について、現在形、未来形、過去形、現在完了形、過去完了形、未来完了形、現在進行形、過去進行形、未来進行形という 9 つの構文で生起した頻度を調査した。そのうち 11 個の動詞についての結果を抜き出して、以下の表にまとめている。これをもとに続く問いに答えなさい。

| | VERB | PRESENT | FUT (WILL) | **PAST** | **PRS PRF** | PST PRF | FUT PRF | PRS PROG | PST PROG | FUT PROG | TOTAL |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A | BECOME | 32,211 | 6,043 | **89,105** | **31,547** | 10,340 | 55 | 6,203 | 2,208 | 2 | 291,488 |
| | EVOLVE | 942 | 209 | **1,813** | **2,222** | 336 | 7 | 374 | 84 | 3 | 12,421 |
| | CHANGE | 14,003 | 2,531 | **16,976** | **10,903** | 3,303 | 38 | 3,090 | 768 | 22 | 139,217 |
| | INCREASE | 6,993 | 2,050 | **6,883** | **4,962** | 676 | 31 | 1,245 | 248 | 13 | 69,476 |
| B | SCREAM | 1,202 | 42 | **4,212** | **4** | 100 | 0 | 367 | 862 | 4 | 18,699 |
| | NOD | 1,229 | 21 | **20,951** | **12** | 79 | 0 | 75 | 219 | 1 | 30,482 |
| | SMILE | 2,498 | 60 | **22,302** | **16** | 165 | 0 | 421 | 1,212 | 11 | 45,951 |
| | STARE | 3,088 | 40 | **15,996** | **35** | 140 | 1 | 644 | 1,612 | 2 | 43,096 |
| C | REMEMBER | 48,731 | 970 | **14,108** | **40** | 242 | 2 | 141 | 224 | 5 | 121,157 |
| | LIKE | 71,705 | 385 | **22,536** | **78** | 259 | 0 | 28 | 18 | 0 | 207,626 |
| | NEED | 135,622 | 5,366 | **42,551** | **165** | 229 | 0 | 26 | 19 | 15 | 323,740 |
| | Total | 19,915,160 | 571,588 | **14,262,925** | **966,351** | 531,979 | 2,922 | 508,233 | 406,085 | 10,578 | 60,877,297 |

過去形の対数頻度を横軸に、現在完了形の対数頻度を縦軸に散布図を描いたものが下記の図 1 である。一つ一つの丸が動詞を表し、値の大きい／小さいものに関しては、その動詞の通し番号が付され、横にそれがどの動詞であるか表示している。

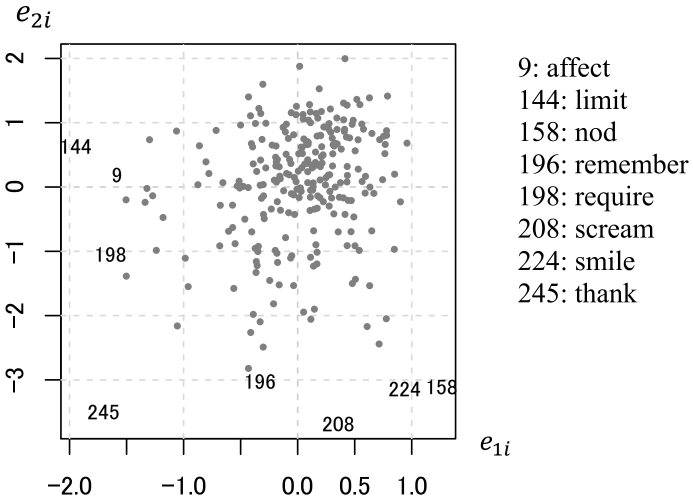



問 1　この図 1 をもとに「過去形でも現在完了形でも使いづらい動詞に monitor, thank などがあることが分かった」あるいは

「be 動詞が最も過去形とも現在完了形とも共起しやすい動詞だ」という結論を引き出してよいか、適切に答えなさい。

$i$ 番目の動詞の過去形の対数頻度を$x_{1i}$、$i$ 番目の動詞の現在完了形の対数頻度を$x_{2i}$、$i$ 番目の動詞のコーパスにおける総頻度に対数を取ったものを$x_{3i}$とする。そのうえで、下記のモデルを立て、その係数を推定した。この推定されたモデルからの残差$e_{1i} = x_{1i} - \hat{x}_{1i}$と$e_{2i} = x_{2i} - \hat{x}_{2i}$を、それぞれ横軸と縦軸にして散布図を描いたものが、下記の図２である。

$$x_{1i} = \beta_{10} + \beta_{11} x_{3_i} + \varepsilon_{1i}$$
$$x_{2i} = \beta_{20} + \beta_{21} x_{3_i} + \varepsilon_{2i}$$




問２ コーパスにおける総頻度の大きさでは予測／説明できないほど、現在完了形での使用量が低い動詞にはどのようなものがあるか答えなさい。

問３ † limit という動詞が図 2 の散布図において特殊なふるまいを見せるのは、もちろん、この動詞の意味特性の持つ現在完了形・過去形に対する結びつきやすさ／結びつきにくさが反映されているからなのであるが、実は、さらに「現在完了形として採取された例文」の一部に本来、現在完了形としてはカウントするべきではない事例が多数含まれているからである可能性が見つかった。「現在完了形」を採取するのにつかわれた検索式は以下のものだったのだが、なぜこの検索式では現在完了形ではない事例まで含まれてしまったのか、考えられる例を挙げながら説明しなさい。

検索式：動詞の後ろに過去分詞形が続く表現