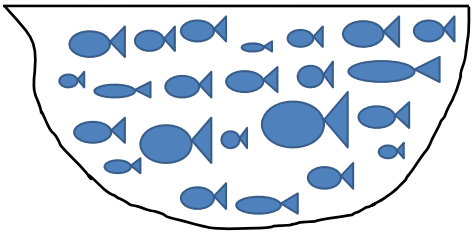
第１４回　回帰分析その２

★　教材「生物統計学__推定と予測 2013」を予習しながら空所を埋めておくこと

Ａ．推定と予測

１．推定と予測の違い

統計を用いて，何かを予想することは現代では頻繁に行われることである．これまでに統計を利用して，（統計的）推定と（統計的）検定を行う方法を学んできた．統計を利用して，予測を行うこともでき，しかも予測もよく利用される統計的手法である．最初に推定と予測の違いを考えよう．



池の魚を 10 匹，無作為に釣り上げて，その体重を測定し，池の魚全体（母集団）の平均体重を推定する方法（ｔ分布を使った母平均の推定）をすでに学んだ．この場合，釣り上げる池の魚の数を増やせば増やすほど，池の魚の母平均の信頼区間の幅（範囲）は
（　どんどん広くなる　・　変わらない　・　どんどん狭くなり，0 に近づいていく）．そのことは母平均を推定する時に利用する標準誤差（＝標準偏差/√（標本数））からもわかる．標本を増やせば，標準誤差は 0 に近づく．

次に池の魚を 10 匹，無作為に釣り上げて，その体重を測定したあとで，その次に釣り上げる 1 匹の魚の体重を予想するとしよう．この場合も，先に無作為標本として釣り上げた 10 匹の魚の平均体重と比べて，その次に釣り上げる 1 匹の魚の予想される体重は，
（　それより大きい　・　同じである　・　それより小さい　）．しかし，この場合，最初に予想するために釣り上げる池の魚の数を増やせしても，次に釣り上げる 1 匹の魚の予想体重の信頼区間の幅は小さくはなるものの，池の魚そのものにばらつきがあるから，ある一定の値（標準偏差によって決まる）よりは小さくはならない．つまり次に釣る魚の重さを予想するために，池の魚をいくらたくさん標本として調査しても，誤差はある一定限度までしか，小さくできない．

次にもしこの池が富栄養化した結果，池の魚の体重が変化したかもしれないという状況を考えよう．池の魚の体重に及ぼす富栄養化の影響を研究したい科学者は池の魚全体（母集団）がどのように変化したかに興味があるだろう．その場合，富栄養化の結果，池の魚の母平均（あるいは母標準偏差などでもかまわない）がどう変化したかを統計的に（推定・予測）することになる．

一方，富栄養化した結果，ある漁師は今日，釣る魚の体重に関心をもつだろう．その場合，富栄養化の結果，今日，釣る池の魚の体重（なお漁獲高は 1 匹だけと限らず，数匹釣り上げた魚の体重でもかまわない）がどう変化したかを統計的に（推定・予測）することになる．

このように池の魚の母集団の性質を統計的に推測することを統計的推定といい，一方，次に釣る魚が実際にどうなるかを推測することは予測という．推定と予測では信頼区間が異なるので，注意が必要である．

推定

標本から母集団の母数（母平均，母分散など）を予想すること．

予測

標本から得られた統計量を用いて，次に取り出す標本の値がどうなるかを予想すること．

予習問題　推定と予測の違いを考えよう．（推定・予測）のうち，自分が当てはまると考える方に○をつけよう．

天気予報は（推定・予測）である．したがって，明日の降水確率は 90%というときは，その明日 1 日だけについて（推定・予測）している．一方，過去 30 年の気象から 11 月 3 日は晴れの特異日であるというときは，降水確率を（推定・予測）していると考えることができる．

製薬会社Ａとしては風邪薬Ｑが従来の薬Ｐより平均して効果があるかを知ろうとするから，新薬Ｑの効果を（推定・予測）する．医師として新薬Ｑが対象となる薬Ｐよりある患者Ｂに効果があるかを知ろうとするから新薬Ｑの効果を（推定・予測）する．

研究者Ｃは温度が上昇すると，みかんの糖度がどうなるかを（推定・予測）した．みかん農家Ｄは温度が上昇すると，今年のみかんの糖度がどうなるかを（推定・予測）した．

いくつかの並行する道路のうち，交通量が多い道路沿いに新規出店を考えた．その場合，数回の交通量調査によって，それぞれの道路の平均交通量を（推定・予測）する．

明日は観光に出かけることにした．観光地に向かういくつかの並行する道路のうち，交通量が少ない道路を選んで，時間の節約を図りたい．その場合，明日におけるそれぞれの道路の交通量を（推定・予測）する．

血圧を下げる成分を 20mg にしたときに，患者全体で平均どの程度，血圧が低下するかを（推定・予測）する．そのことによって，この成分が効果があるか，新薬として申請できるかを調べた．

血圧を下げる成分を 20mg にしたときに，患者Ａさんはどの程度，血圧が低下するかを（推定・予測）する．そのことによって，この成分をＡさんに投与できるかを調べた．

★　教材「生物統計学__単回帰分析における区間推定 2013」を予習しながら空所を埋めておくこと

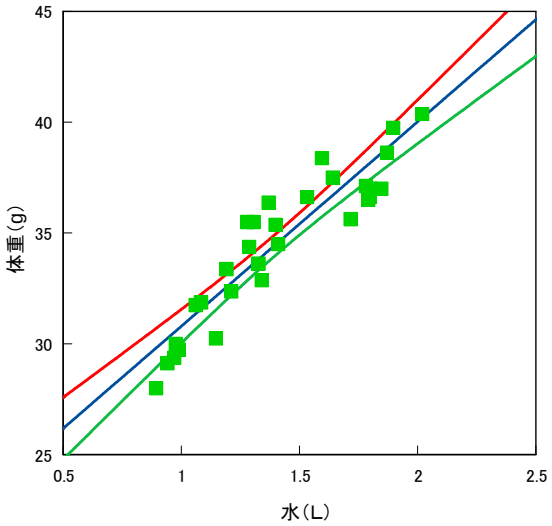
B．単回帰分析における区間推定

1．回帰直線に対して，推定値はどのようにばらつくか？

回帰分析の目的のひとつは説明変数 x を指定したときに y がどのような値をとるかである．例えば，気温と果実糖度の関係を回帰分析したときに気温が 20℃なら，果実糖度がいくらになるかを知りたいということである．この場合，回帰式の x に 20℃を代入して，y（果実糖度）を計算したばあい，その y は点推定である．信頼率 95%で温度 20℃の条件で y（果実糖度）がいくらになるかを区間推定したい場合には回帰式だけでは計算できない．

回帰分析における y（目的変数）の推定に関する信頼区間の計算は面倒であり，しかもパソコンでできるので，ここでは図を見て，データのばらつきがどのようなものかを理解することに重点を置く．

右の図は前回のデータ（飲み水とウズラの体重の関係）から得られた回帰直線と，それぞれの x に対する y の平均値の 95%信頼区間を示している．



この図から次のことがわかる．

① 全体の平均（重心）に近いほど平均の信頼区間は狭くなる．すなわち推定精度が高い．

② 全体の平均から遠いほど，特に回帰式を求めるデータの範囲外にでると，信頼区間は広くなり，精度は落ちる．

③ 以上のことから，回帰分析では説明変数 x はできるだけ広い範囲をカバーすることが望ましい．さらに説明変数 x がカバーしない部分で y を推定するとあまり精度は高くないことがわかる．

2．推定の誤差はどうすれば小さくなるか？

① 独立変数 x の範囲を広く取る

② 標本数を増やす

③ 全体の平均に近いところを推定すると精度が高いから，推定したい x が平均付近に来るようにデータを集める

平均	1.4093	34.50417	
平方和	3.278872	325.0151	
推定値	1.5	35.34068	=C39*G20+G19

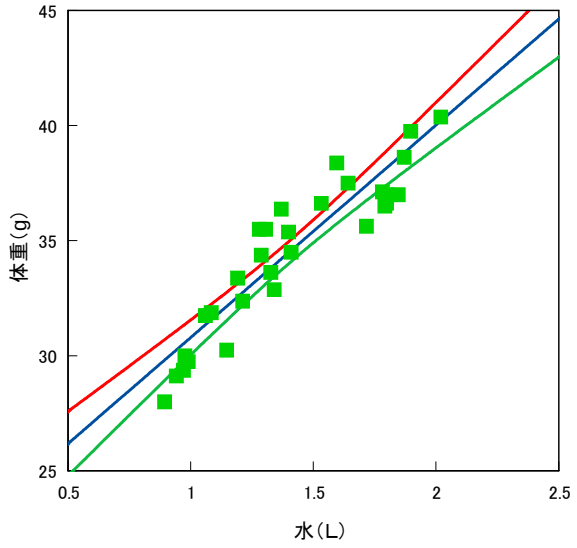
3．回帰による推定と推定値の信頼区間

① η_0（イータとよむ）の点推定

x_0 に対する η_0 の点推定は $\hat{\mu} = \hat{\alpha} + \hat{\beta}x_0$ の式から計算する．例えば，飲み水の量とウズラのヒナの体重の回帰の例で，飲み水 $x_0 = 1.5$ に対するウズラの体重 y の母平均 η_0 はエクセルでは以下のように回帰分析の結果から，回帰係数と切片を代入して計算すればよい．

② η_0 の区間推定

しかし，上の点推定のもととなる回帰式は標本から求めた標本回帰式であり，母回帰式は標本回帰式そのものではない．したがって，y の平均値 η_0 を推定するときには誤差がふくまれるはずである．先述のウズラのヒナの例について，グラフに x_0 に対する η_0 の 95%信頼区間（上の線と下の線の間が信頼区間，真ん中の線は点推定）を計算し，図示すると右下の図のようになる．平均に近いほど推定の精度が高く，標本の範囲を超えると推定の精度が急速に落ちることがわかる．



4． η_0 の区間推定値の信頼区間の計算方法

ウズラの飲み水のデータを例にして，ウズラに水を 1.5L 飲ませた場合，ヒナの平均体重を 95%信頼区間をつけて推定してみよう．

水(L)	体重(g)		概要								
1.864	38.625										
1.775	37.125		回帰統計								
1.794	37		重相関 R	0.926356							
1.84	37		重決定 R2	0.858136							
1.792	36.625		補正 R2	0.853069							
1.71	35.625		標準誤差	1.283244							
1.392	35.375		観測数	30							
1.334	32.875										
1.403	34.5		分散分析表								
1.32	33.625			自由度	変動	分散	測された分散	有意 F			
1.785	36.5		回帰	1	278.9071	278.9071	169.3718	2.15E-13			
1.635	37.5		残差	28	46.10803	1.646715					
1.89	39.75		合計	29	325.0151						
2.013	40.375										
1.077	31.875			係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 99.0%	上限 99.0%
0.984	29.75		切片	21.50633	1.025847	20.96446	1.18E-18	19.40498	23.60769	18.67165	24.34102
0.97	30		水(L)	9.2229	0.708675	13.01429	2.15E-13	7.771246	10.67455	7.264646	11.18115
0.964	29.375										

生物統計学__授業用データ集 2013 のエクセルファイルの第１４回回帰その２見本タブにウズラの飲み水の計算例がある．実際の計算は第１４回回帰その２計算用を使う．推定で必要なのはエクセルの分散分析表で赤で示した残差分散，黄色で示した切片係数，紫で示した回帰係数である．

下のようにデータを指定する．指定した x $_1$（推定）にここでは飲み水の量 1.5 L を入れる．そうすると 1.5L 水を飲んだときのヒナの平均体重を推定する．

次に残差分散，切片係数，回帰係数を分散分析表から読み取り，代入する．信頼率を指定する．

指定したx1(推定)	1.5	
指定したx2(予測)		
指定したy(逆推定)		
残差分散	1.646715218	分散分析表の残差分散を代入する
切片係数	21.5063332	分散分析表の切片係数を代入する
回帰係数	9.222900355	分散分析表の回帰係数を代入する
信頼率	95	%
推定値の標準誤差	0.242944469	
t値	2.048407142	
yの推定値	35.34068373	← 点推定値
下限	34.84303454	← 区間推定値
上限	35.83833291	

ヒナの体重について推定の点推定値(g)

$$\eta_0 = 35.34$$

95%信頼区間をつけた区間推定値（g）

$$34.84 \le \eta_0 \le 35.84$$

予習問題

右のデータは輪ゴムを伸ばした長さが輪ゴムの飛ぶ距離に及ぼす影響を調べたものである．輪ゴムを 5.3cm 伸ばしたときに輪ゴムは平均で何 cm 飛ぶかを 95%信頼区間をつけて区間推定せよ．

伸ばした長さ(cm)	飛んだ距離(cm)
1	10
1.5	52
2	89
2.5	141
3	152
3.5	163
4	213
4.5	223
5	227
5.5	234
6	275
6.5	335
7	352
7.5	360
8	378
8.5	384
9	399
9.5	428
10	461
10.5	478

★　教材「生物統計学__単回帰分析における区間予測 2013」を予習しながら空所を埋めておくこと

C．回帰分析における予測

１．推定と予測の違い

回帰分析である x に対して y がどんな値になるかを知りたいときに，y の平均値 η_0 を区間推定したいのではなく，ある１回だけ得られる $\hat{y}_0$ を区間推定したいときがある．例えば，薬Ａを飲んで，ある患者の血圧がどの程度下がるかを知りたいときに，母平均（すなわち薬を飲んだすべての人の平均）の区間推定ではなく，そのたった一人の患者では 95%の信頼区間で血圧がどうなるかを知りたいだろう．そういう場合，このような予想をすることを予測という．推定は母数（母平均など）について使い，個々のデータについて統計的に予想するときには予測という言葉を用いる．

例えば，気温と果実糖度の関係を回帰分析したときに気温が 20℃なら，今年とれる果実糖度がいくらになるかを知りたいとする．この場合は推定ではなく，予測となる．予測の場合は推定よりも信頼区間の幅が広くなる．

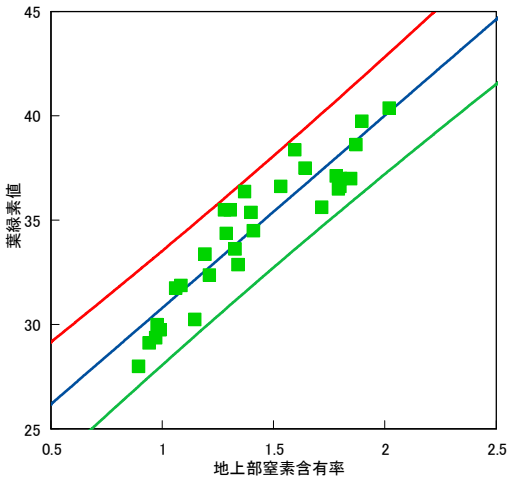
２．予測の信頼区間

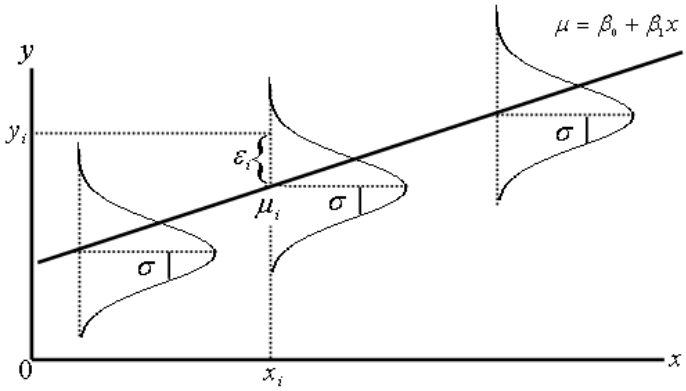
①　y_0 の点予測

x_0 に対する y_0 の点予測は点推定 η_0 の場合と同じである．すなわち $y_0 = \eta_0$ である．

② 予測の信頼区間

予測のときの信頼区間をグラフに示して，推定と予測における信頼区間の違いをみる．先述のウズラのヒナの例について計算した場合，グラフに x_0 に対する y_0 の予測の 95%信頼区間（上の線と下の線の間が信頼区間，真ん中の線は点予測）を図示すると右図のようになる．予測の場合は母平均の推定に比べると，データの平均に近づいても信頼区間の幅はあまり小さくならない．その理由は，回帰分析では独立変数 x に対して，従属変数 y がある幅のあるばらつきをともなって決まるという仮定のため，そのばらつきによって決まる誤差よりは信頼区間は小さくならないこと，さらにそのばらつきは x の値にかかわらず一定であると仮定しているからである．





③ 95%信頼区間をつけた予測値の計算方法

ウズラの飲み水のデータを例にして，ある 1 匹のウズラに水を 1.5L 飲ませた場合，そのヒナの体重がいくらになるかを 95%信頼区間をつけて予測してみよう．

生物統計学__授業用データ集 2013 のエクセルファイルの第 14 回回帰その 2 見本タブにウズラの飲み水の計算例がある．実際の計算は第 １４回回帰その２計算用を使う．推定と計算方法はほとんど一緒で，予測で必要なのはエクセルの分散分析表で赤で示した残差分散，黄色で示した切片係数，紫で示した回帰係数である．

下のようにデータを指定する．指定した x_2（予測）にここでは飲み水の量 1.5 L を入れる．そうすると 1.5L 水を飲んだときのヒナの体重を予測する．

次に残差分散，切片係数，回帰係数を分散分析表から読み取り，代入する．信頼率を指定する．

指定したx1(推定)	1.5				
指定したx2(予測)	1.5				
指定したy(逆推定)					
残差分散	1.646715218	分散分析表の残差分散を代入する			
切片係数	21.5063332	分散分析表の切片係数を代入する			
回帰係数	9.222900355	分散分析表の回帰係数を代入する			
信頼率	95	%			

予測値の標準誤差	1.306038756
t値	2.048407142
yの予測値	35.34068373
下限	32.66538461
上限	38.01598285

予測の点予測値

95%信頼区間をつけた区間予測値　$y_0 = 35.34$

$$32.67 \le y_0 \le 38.02$$

予習問題

右のデータは輪ゴムを伸ばした長さが輪ゴムの飛ぶ距離に及ぼす影響を調べたものである．輪ゴムを 7.2cm 伸ばしたときに次に輪ゴムは何 cm 飛ぶかを 95%信頼区間をつけて区間予測せよ．

伸ばした長さ(cm)	飛んだ距離(cm)
1	10
1.5	52
2	89
2.5	141
3	152
3.5	163
4	213
4.5	223
5	227
5.5	234
6	275
6.5	335
7	352
7.5	360
8	378
8.5	384
9	399
9.5	428
10	461
10.5	478

★　教材「生物統計学__単回帰分析における逆推定 2013」を予習しながら空所を埋めておくこと

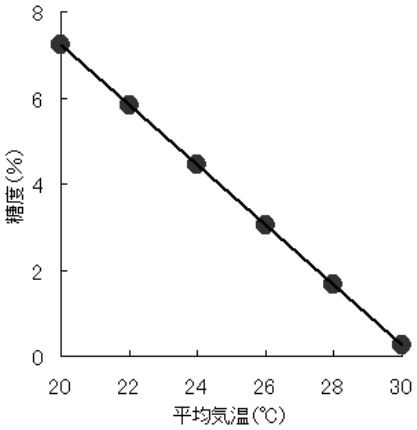
D．回帰の逆推定

1．逆推定とは？

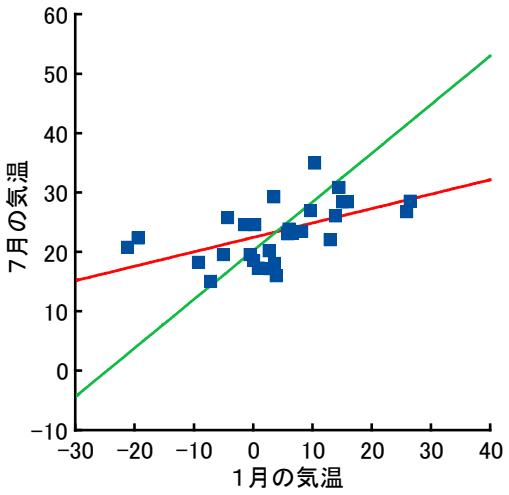
独立変数 x から従属変数 y を推定・予測するのではなく，従属変数 y から独立変数 x を予想したいことがある．そのまえに x と y は２つの変数のどちらでもよいのかということを考える．

回帰分析では独立変数 x は指定できる値であり，従属変数 y はあるばらつきをともなう値であるという前提で行う．次の例は先ほど用いた平均気温と果実の糖度の関係である．この関係では平均気温を指定すると果実の糖度が決まるという関係になっている．しかし，果実を生産する現場からすると，果実の糖度を何パーセント以上にするには平均気温を何度以下に設定したらよいかという問題の方が現実的である．では果実の糖度を独立変数 x にして，平均気温を y にして回帰分析したらよいのか？

平均気温	糖度
20	7.23
22	5.83
24	4.44
26	3.04
28	1.65
30	0.25



実は回帰分析では右の図のように x と y を入れ替えると得られる回帰式は平均を共通して通る異なる２つの直線になる．右の例は回帰分析には向かないデータであるが，仮に回帰分析できるとしたら 1 月の気温と 7 月の気温のどちらを独立変数とするかによって，得られる回帰式が異なることを右のグラフは示している．

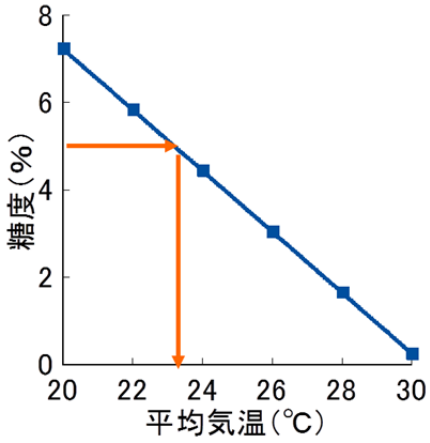


回帰分析の前提条件から独立変数 x は指定できる値でなければならない．すなわち果実の例では独立変数 x は平均気温でなければならない．もし果実の糖度を指定した上で，平均気温を決めたいのであれば，平均気温を独立変数とした回帰式を求めてから， x についてこの式を解けばよい．

すなわち果実の糖度 $y = 21.18286 - 0.69771 \times$平均気温だから

平均気温 $x = -($ 果実の糖度 $- 21.18286) / 0.69771$ である．

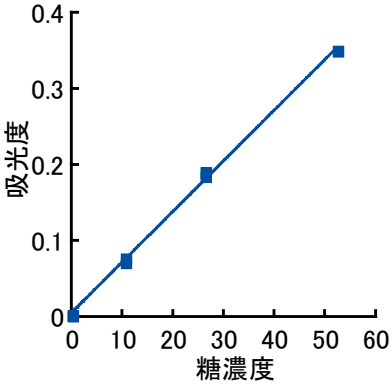
このように従属変数 y から独立変数 x を回帰によって求めることを逆推定という．



このように独立変数 x から従属変数 y を推定・予測するのではなく，従属変数 y から独立変数 x を予想したいことはよくある．

化学分析で標準液をいくつか使って，検量線を書くのも逆推定である．標準液の濃度を指定すると吸光度などの反応量が決まる．未知試料の濃度は反応量から逆推定される．

回帰分析では独立変数 x は指定できる値を使うのが原則である． 2 つの分析法を比較する場合は両変数ともにばらつきがあるが，独立変数 x をよりばらつきの小さい分析法にすれば実用上問題ない．



例：先述のウズラのヒナの例で，ヒナの体重として 35g を得た．飲み水の量を逆推定せよ．

$y = 21.50633 + 9.2229x$

$y = 35$ を代入し， x を求める． $x = (35 - 21.50633) / 9.2229 = 1.463061(\%)$

逆推定	1.463061	=(F37-G19)/G20
35		

エクセル授業用シートでは y の値を代入するだけで， x を逆推定できる．生物統計学__授業用データ集 2013 のエクセルファイルの第１４回回帰その２見本タブにウズラの飲み水の計算例がある．実際の計算は第１４回回帰その２計算用を使う．指定した y (逆推定)に代入するだけで，下の逆推定のところに答えが出る．

指定したx1(推定)	1.5	
指定したx2(予測)	1.5	
指定したy(逆推定)	35	
残差分散	1.646715218	分散分析表の残差分散を代入する
切片係数	21.5063332	分散分析表の切片係数を代入する
回帰係数	9.222900355	分散分析表の回帰係数を代入する
信頼率	95	%
推定値の標準誤差	0.242944469	
t値	2.048407142	
yの推定値	35.34068373	
下限	34.84303454	
上限	35.83833291	
予測値の標準誤差	1.306038756	
t値	2.048407142	
yの予測値	35.34068373	
下限	32.66538461	
上限	38.01598285	
逆推定	1.463061107	

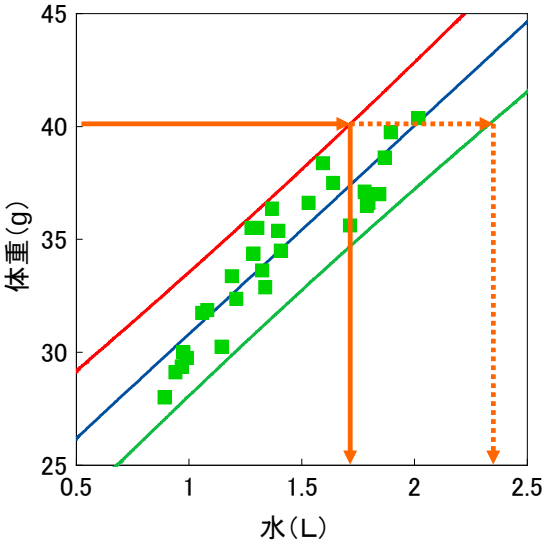
ここに従属変数yを代入

$x = 1.46$

逆推定値

２．逆推定での信頼区間

逆推定での信頼区間の計算も煩雑である．ここでは専門的すぎるので，逆推定でも右の図のようにして信頼区間を計算できることだけを紹介するにとどめる．現実問題としては，果実の糖度の例でも果実の糖度が望む糖度より高くなる確率を 95%以上にしたいということがあるなら，信頼区間を計算しなければならないことになる．



予習問題

右のデータは輪ゴムを伸ばした長さが輪ゴムの飛ぶ距離に及ぼす影響を調べたものである．輪ゴムが 300cm 飛んだとする．この輪ゴムは何 cm 伸ばしたのかを逆推定(点推定でよい)せよ．

伸ばした長さ(cm)	飛んだ距離(cm)
1	10
1.5	52
2	89
2.5	141
3	152
3.5	163
4	213
4.5	223
5	227
5.5	234
6	275
6.5	335
7	352
7.5	360
8	378
8.5	384
9	399
9.5	428
10	461
10.5	478

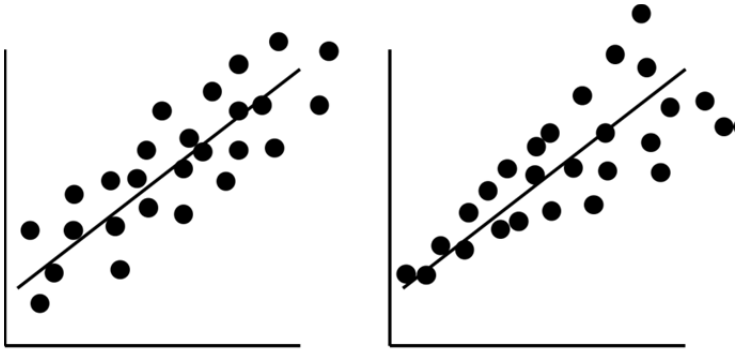
★　教材「生物統計学__発展した回帰分析 2013」を予習しながら空所を埋めておくこと

Ｅ．発展した回帰分析

未知の条件であっても，回帰式に x を代入して，推定・予測ができる，あるいは y を代入して逆推定できる回帰分析は現代においてもっともよく使われる統計的解析手法の 1 つとなっている．そのためにさまざまな応用・発展した回帰分析がある．ここではその一部を紹介する．実際に行うには(重回帰分析を除くと)専用の統計ソフトが必要である．

１．重み付きの回帰

基本的な回帰分析では説明変数 x の値に関わりなく，目的変数 y の誤差が一定であるという前提であった．しかし，時間にともない，生物の体重などの成長量が増える，加熱した水の温度が上昇する，このような現象では誤差は時間の経過にともない大きくなる．したがって，この場合，x の大きさが大きくなるにつれて，y の誤差を大きく見積もったうえで，回帰式を計算する方がよりあてはまりのよい式を求めることができる．このように誤差を x の大きさによって重みを付けてやる方法を重み付き回帰という．

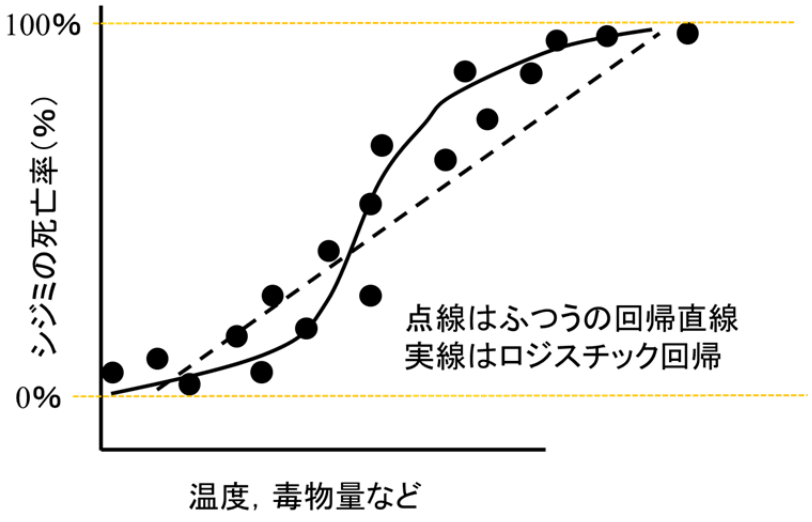


２．ロジスチック回帰

温度，毒物の量など説明変数 x に対して，目的変数 y が頻度データ（パーセントで示されるデータ，第 7 回の授業で取り上げた）であるとき，そのまま回帰分析をしようとすると困った問題が２つある．

１つめは，頻度データでは y は 0～100%の間しかとらないのに，ふつうの回帰分析では−∞から+∞をとるということである．つまり回帰式にある x を代入したら，y が 100%以上になったり，負数（マイナス）になる可能性があるということである．すなわちある温度におけるシジミの生存率を調べようとしたら，生存率が 100%を超えるとか，−50%になるということで現実にはありえない結果が出てくる．2 つめは，頻度データの場合は調べた標本数が多くても，少なくてもパーセント表示にしてしまうとその違いがなくなってしまうことである．すなわちもし２つのデータがどちらも 30%であったとしても，片方が 100 個体から得たデータ（すなわち 30 個体が該当した），片方が 10 個体から得たデータ（すなわち 3 個体が該当した）ではデータの意味が全然違うからである．

この問題を解決する方法として，ロジスチック回帰（と最尤法，さいゆうほうと読む）がある．この計算はコンピューターがないとできなかったので，かつてはほとんど利用されなかったが，近年ではごく一般的に利用されるようになった．

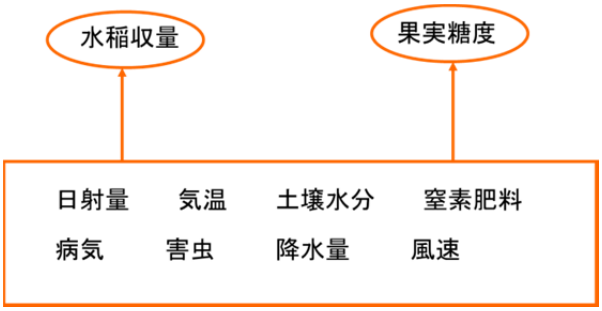


３．多変量解析とは？

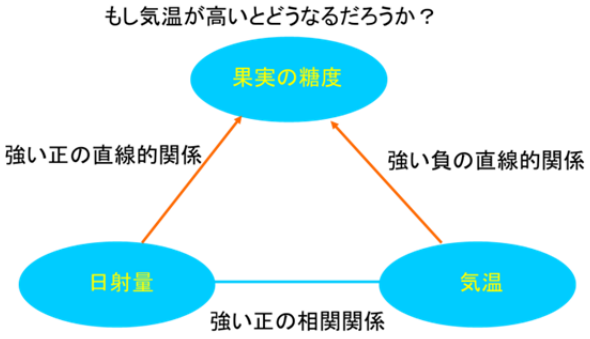
現実の現象は１つだけの要因だけに支配され，それから説明されるということはむしろまれなことである．相関分析や単回帰分析では２つの変量間の関係を調べる．しかし，３つ以上の変量間の関係を調べる必要のあることは多い．このような多数の変量間の関係を調べる方法をまとめて多変量解析という．多変量解析には目的に応じて，多数の手法がある．ここではその一つである重回帰分析を紹介する．

４．重回帰分析

果実の糖度やイネの収量の例で考えよう．果実の糖度やイネの収量は実際には平均気温以外に日射量，土壌水分，肥料などさまざまな要因で決まると考えられる．目的変数である果実の糖度やイネの収量に対して，説明変数（独立変数）を２つ以上考えたいということがある．単回帰分析は１つの説明変数であったが，これを複数に拡張したものが重回帰分析である．



説明変数間に相関があるとき，単回帰分析の式を足しあわせただけでは正しい重回帰式は得られない．例えば，平均気温と日射量にはおそらくかなり強い正の相関があるだろう．重回帰分析を行うとそのような説明変数間の相関を除いて，各説明変数単独の効果を評価できる．

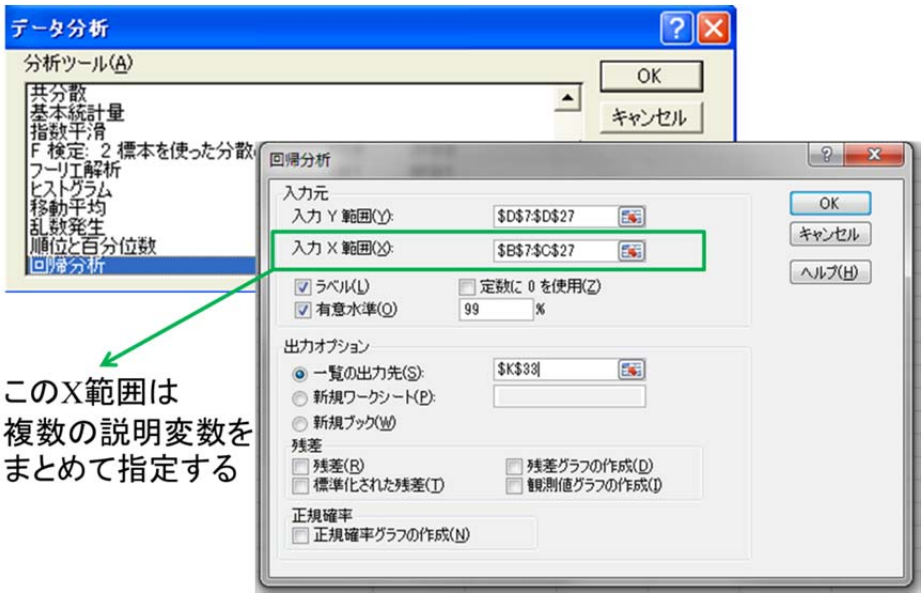


右の図のような関係があるとする．つまり，果実糖度は日射量が高いほど高くなる正の直線的関係があり，気温が高くなるほど低くなる負の直線的関係がある．しかし，日射量が高いほど，気温が高くなる強い正の相関関係がある．その場合，日射量が高くなると気温も高くなるので，単に気温と糖度の関係を調べても，気温が高いほど糖度が高くなるとか，気温も日射量も糖度に影響しないという結論が出てくる可能性があります．

例：気温と日射量がイネの収量にどのような関係があるかを調べた結果が以下の結果である．単回帰分析と重回帰分析を行ってみるとどうなるか？

気温（℃）	日射量（MJ/m2・日）	収量（kg/10a）
23.3	19.4	610.1
28.9	34.5	463.2
25.4	27.5	515.7
19.3	14.2	371.6
20.3	9.5	425.3
24.7	16.1	471.8
28.8	33.7	662.6
23.9	13.9	355.2
28.1	22.3	358.6
28.0	18.8	358.2
19.3	3.6	369.1
25.5	27.7	410.8
20.0	9.4	369.4
19.0	6.9	498.6
19.6	14.8	536.0
24.8	17.9	318.6
22.9	19.1	619.1
26.0	17.0	476.8
23.9	16.7	323.0
20.5	3.8	180.4

エクセルの重回帰分析は単回帰分析とほとんど同じ手順でできる．エクセルの分析ツールから回帰分析を選択する．入力Y範囲は目的変数を選ぶ．例題では収量の列を選びます．入力X範囲で複数の説明変数をまとめて選びます．つまり，例題では気温と日射量をあわせて，入力X範囲にまとめて指定します．したがって，エクセルで重回帰分析をするときは複数の目的変数を必ず隣になるようにまとめておく必要があります．



概要								
回帰統計								
重相関 R	0.614286							
重決定 R2	0.377347							
補正 R2	0.304093							
標準誤差	98.21606							
観測数	20							
分散分析表								
	自由度	変動	分散	測された分散	有意 F			
回帰	2	99382.13103	49691.07	5.151258	0.017828			
残差	17	163988.6985	9646.394					
合計	19	263370.8295						
	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 99.0%	上限 99.0%
切片	745.9386	219.9066504	3.392069	0.003467	281.9761	1209.901	108.5984	1383.279
気温（℃）	−23.5567	11.88640697	−1.98182	0.063907	−48.6348	1.521432	−58.0062	10.89285
日射量（MJ/m2・日）	14.12572	4.624010789	3.054863	0.007164	4.36991	23.88153	0.72427	27.52717

重回帰式は係数のところを読む．

収量 y=745.9386+(-23.5567)×気温+14.12572×日射量という式が得られた．

すなわち収量を推定あるいは予測するにはこの式に気温と日射量の 2 つの変数を代入するとよい．式から温度の項の係数はマイナスなので，温度が高いほど，糖度は低くなること，日射量の項は係数がプラスなので，日射量が多いほど糖度が高くなることがわかる．

５．単回帰分析と重回帰分析の適用現場の違い

重回帰分析はある目的変数を説明する要因がいくつも考えられるときに，説明変数を絞り込むために用い，単回帰分析は絞り込んだ説明変数が他の重要な要因は一定になるように制御された条件で目的変数にどのように影響するかを定量化する場合に用いることが多いようである．

重回帰分析でとりあげる説明変数は数が多ければ多いほどよいというものでもない．説明変数の絞り込みは難しい問題なので，ここではとりあげない．よく知られた問題として多重共線性と呼ばれる問題がある．多重共線性とは説明変数間の相関が高すぎると，重回帰式が不安定になり，少しの説明変数の変化で式が大きく変わることをさす．したがって，重回帰分析をするときはやみくもに説明変数を増やさないで，相関が強すぎない説明変数をいくつか絞り込むのがよい．

G．定期試験について

日時　2 月 12 日（水）　午後 12 時 45 分から 2 時 15 分まで（90 分間）

場所　マルチメディア演習室 1

すべて持ち込み可．パソコンは必ず持ってくること．ただし通信機能の使用は不可．

問題の入ったエクセルファイルをあらかじめ Moodle からダウンロードしてください（2 月 7 日ごろ Moodle に載せる予定）．問題の入ったファイルはパスワードを入力しないと開けられません．パスワードは試験当日に発表します．答案は紙に書きます．Moodle でファイルを提出する必要はありません．

出題範囲　全部

成績評価：定期試験＋レポート（宿題と予習）＋授業

宿題未提出・再提出分の最終締め切り　2 月 11 日（火）午後 5 時

宿題を期限までに全部出せば，期末試験の結果が悪くても追試を必ず受けることができます．