第１３回　回帰分析その１

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

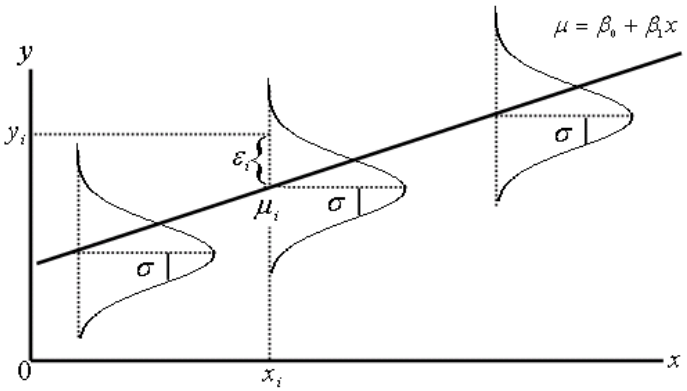
Ａ．回帰

１．回帰

x が指定した値であり，y が指定された x に対して，誤差が正規分布にしたがいながら決定する場合，両変数間の関係を回帰分析する．x と y の関係は直線だけでなく，指数，対数，２次曲線なども考えられる．ここでは直線の場合だけを考える（単回帰分析）．

２．回帰直線

x と y の２つの変数の関係を表した直線的な式を回帰直線という．



回帰直線のモデル（母回帰式）は右の図のようになっている．このことから単回帰分析の適用できる条件は基本的には以下の２つである．

①　x が指定されたときに y は正規分布にしたがって決定する．

②　x の値にかかわらず，y の標準偏差は一定である．

ただしこの条件が満たさないときでも変数変換などで単回帰分析できることも多い．

以上のようにして，標本から求めた標本回帰式に対して，母回帰式を考えることができる．標本回帰式の傾き（回帰係数）と切片について推定・検定や分散分析をすることができる．

３．回帰分析の基本用語

①　（　　）変数　指定できる変数 x のこと．独立変数ともいう．

②　（　　）変数　y のこと．この値をコントロールすることが目的なので目的変数と呼ぶ．x の値が決まると y の母平均が決まるので，従属変数ともいう．

③　母回帰直線　母集団での回帰直線．この傾きを母回帰係数，切片を母切片という．両者をあわせて回帰母数という．

④　予測値　x をある値にしたときに y がどのような値になるかを考えたものをいう．ある信頼区間をつけて予測値を求めることができる．

⑤　残差　目的変数のデータの値と予測値との差.

⑥　寄与率　相関係数を二乗した値である．y の変動のうち回帰による変動の割合を示す.

B．単回帰分析

１．相関係数

回帰の時でも相関係数を同じように計算できる．ただし先週述べた母相関係数に関する推定や検定はできない．相関係数を二乗した値である寄与率の方が回帰では重要な値である．

２．回帰係数

回帰係数は説明変数がある一定量変化したときに，目的変数がどれだけ変化するかを示す値であり，回帰直線で最も重要な値である．

回帰係数が 0 であるときはその両者の値の間に相関がない．

回帰係数が正（負）の時は，両者の値の間には正（負）の相関がある．

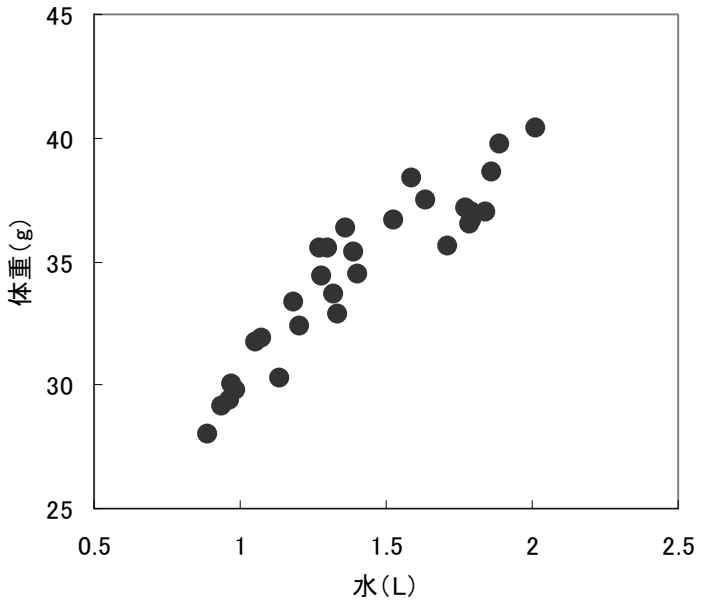
回帰直線 $\hat{y} = A + Bx$ はその傾きである回帰係数 B とその y 切片 A から成り立っている．

回帰直線は両変数の平均を座標とする点を通る．傾き B を**回帰係数**という（相関係数と混同しないこと）． y 切片の単位は目的変数と同じ，回帰係数 B の単位は目的変数／説明変数である．

３．回帰分析の手順

例：与えた飲み水の量（単位：L，独立変数）によって，ウズラのヒナの体重（単位：g，従属変数）がどのように決定されるかを回帰分析した．

① ２つの変量のうちどちらが独立変数であり，どちらが従属変数かを確認し，グラフを書く．



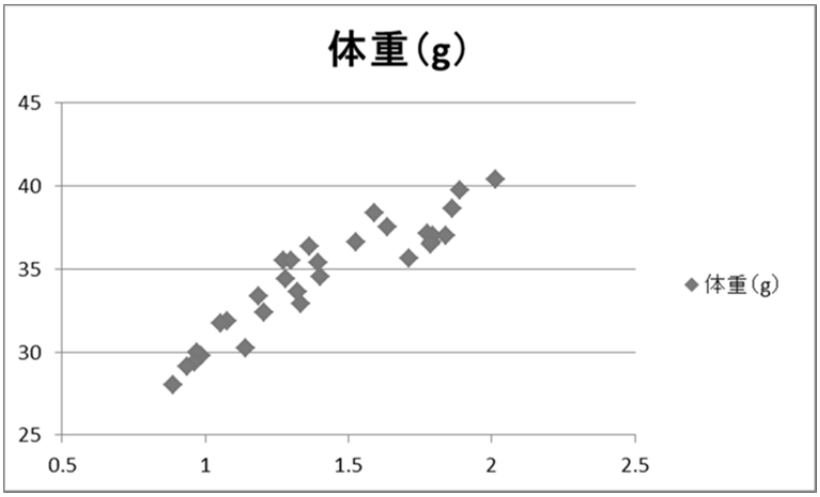
ウズラ	水(L)	体重(g)
1	1.864	38.625
2	1.775	37.125
3	1.794	37
4	1.84	37
5	1.792	36.625
6	1.71	35.625
7	1.392	35.375
8	1.334	32.875
9	1.403	34.5
10	1.32	33.625
11	1.785	36.5
12	1.635	37.5
13	1.89	39.75
14	2.013	40.375
15	1.077	31.875
16	0.984	29.75
17	0.97	30
18	0.964	29.375
19	1.526	36.625
20	1.589	38.375
21	1.363	36.375
22	1.281	34.375
23	0.935	29.125
24	0.887	28
25	1.054	31.75
26	1.185	33.375
27	1.273	35.5
28	1.299	35.5
29	1.14	30.25
30	1.205	32.375

★　グラフ（散布図）の書き方

①　最初に独立変数（ｘ）と従属変数（ｙ）を選択する.

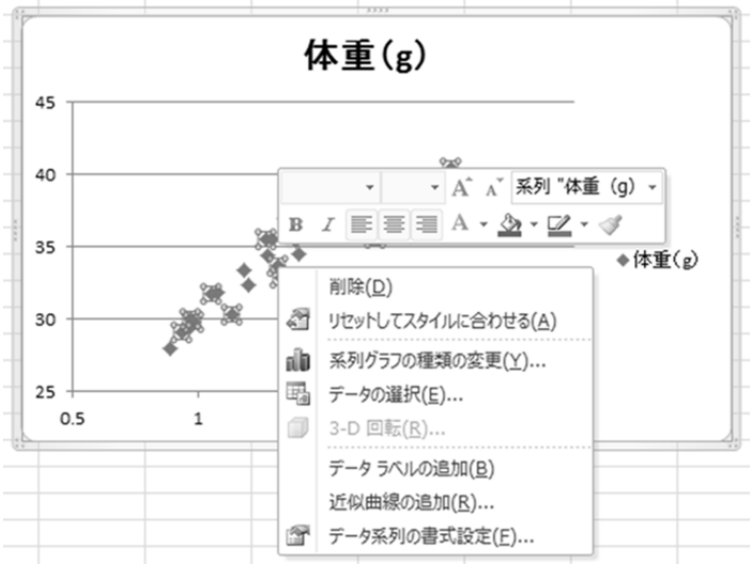
②　挿入→散布図　散布図のメニューから適切なパターンを選ぶ.

③　必要に応じて，できたグラフを加工する．右のグラフでは点がグラフ全体にばらついていないので，傾向が読み取りにくいから，縦軸と横軸の範囲（最大値，最小値）を修正する．（前回の相関分析のときと同じ要領）
右のようなグラフができる.



④　次に回帰直線を入れる．プロット（グラフの点）の上で右クリックして，近似曲線の追加を選ぶ.

　複数のデータを同時にグラフにした場合は近似曲線を描きたいプロット（点）の上をクリックする.

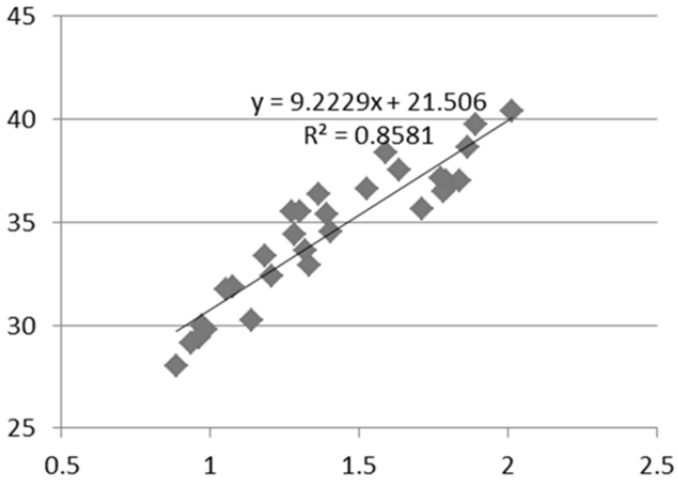


⑤　右の近似曲線の書式設定で線形近似を選ぶとグラフに回帰直線を入れることができる．

直線以外にも指数近似，対数近似，多項式近似（例えば，２次曲線）など曲線を指定することもできる．

下の２つのチェックボックスは必要に応じてチェックする．グラフに数式を表示したいときはチェックを入れると，回帰直線の式をグラフ上に出力する．グラフに R-2 乗値（すなわち寄与率）を表示したいときはこれにチェックを入れる．

⑥　右のように，回帰直線と回帰直線の式（回帰式）と寄与率（R^2=0.8581）をグラフ上に出力した．



予習問題

ある地方特産のにわとりは温度が高くなると産卵数が少なくなるというので実験したところ以下のデータを得た．

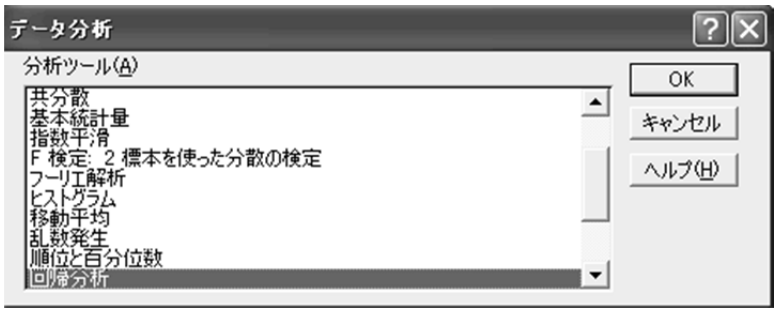
散布図を作成し，その中に，回帰直線，回帰直線の式，寄与率を入れよ．

温度(℃)	産卵数(個/月)
15	28
17.5	24
20	24
22.5	20
25	23
27.5	16
30	19
32.5	20
35	12
37.5	15
40	9

★　教材「生物統計学__単回帰分析の実際 2013」を予習しながら空所を埋めておくこと

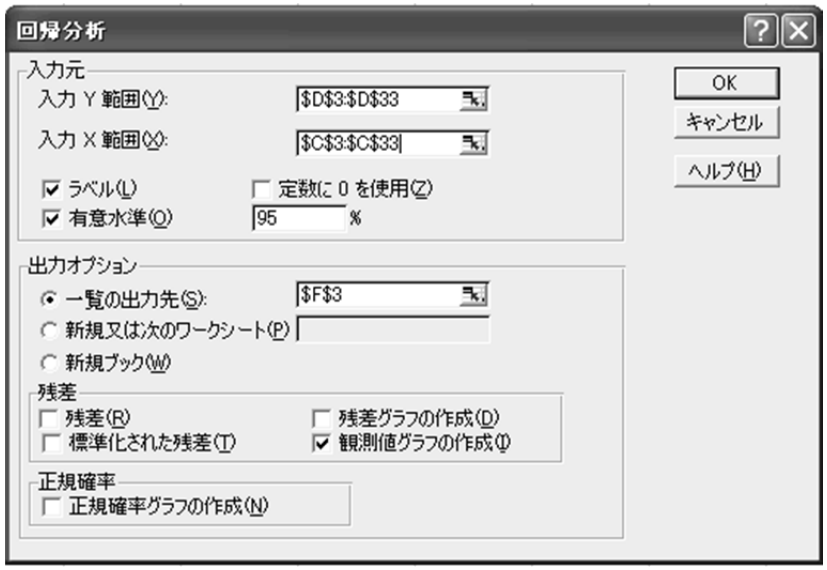
②　回帰係数，回帰直線の式を求める→③　分散分析を行う

エクセルの分析ツールを使えば回帰直線の計算，分散分析などが一度にできる．



入力 y 範囲には目的変数 y を指定する．x と y の順序を間違えないようにする．入力 x 範囲には説明変数 x を指定する．

有意水準は 99%を入れておくと便利である．



概要

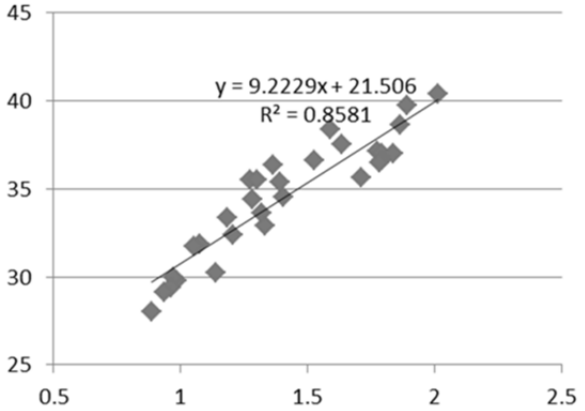
回帰統計	
重相関 R	0.926356
重決定 R2	0.858136
補正 R2	0.853069
標準誤差	1.283244
観測数	30

分散分析表

	自由度	変動	分散	測された分散	有意 F
回帰	1	278.9071	278.9071	169.3718	2.15E-13
残差	28	46.10803	1.646715		
合計	29	325.0151			

	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	21.50633	1.025847	20.96446	1.18E-18	19.40498	23.60769	18.67165	24.34102
水(L)	9.2229	0.708675	13.01429	2.15E-13	7.771246	10.67455	7.264646	11.18115

エクセルの表から $\hat{y} = 21.51 + 9.22x$ という回帰直線の式が得られた．



③　分散分析

この例の場合，

帰無仮説：ウズラの体重は飲み水の量によって変わらない

対立仮説：ウズラの体重は飲み水の量によって変わる

である．

回帰の p-値は（　　）なので，飲み水の量の効果は 5%の有意水準で（　有意である　・　有意でない　）．したがって，飲み水の量の効果は（　ある（認められる）　・　あるとはいえない（認められない）　）

単回帰分析の分散分析では要因を回帰と残差（偶然誤差だけでなく，回帰で採用した因子以外のすべての因子の影響も含んでいる）の２つに分け，回帰による変動が残差変動より十分に大きいかをＦ検定する．

単回帰分析の分散分析表は以下のようにする．

要因	平方和（変動）	自由度	平均平方(分散)	分散比	p-値
回帰	S_R	$\phi_R = 1$	$V_R = S_R / \phi_R$	V_R / V_e	pR
残差	S_e	$\phi_e = n - 2$	$V_e = S_e / \phi_e$		
総変動	S_T	$\phi_T = n - 1$			

エクセルで出力された分散分析表から回帰の項を見て，有意Ｆ（p-値：回帰がないという帰無仮説が成り立つとした場合，今回のようなデータの得られる確率である）が 0.05 や 0.01 より小さいと，５％あるいは１％の有意水準で回帰は有意であると結論できる．

④　分散分析の結果，回帰が有意であれば，必要に応じて，回帰係数などの推定や検定を行う

予習問題

以下のデータは以下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．単回帰分析せよ．

帰無仮説：

対立仮説：

p 値

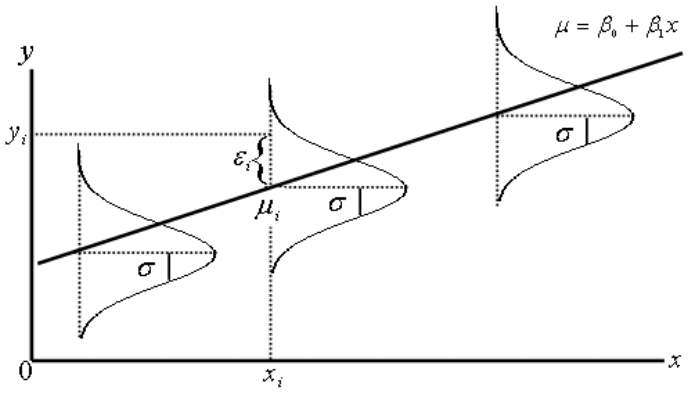
結論

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推定と検定 2013」を予習しながら空所を埋めておくこと

C．回帰分析における推定と検定

回帰分析では独立変数 x を与えたときに従属変数 y がどんな値となるかを決めることが主な目的である．回帰係数，回帰直線などはそのために算出するのである．回帰係数，切片，独立変数 x に対する y の値について推定・検定ができる．



1．母回帰係数の推定

エクセルでは母回帰係数および母切片の区間推定ができる．

概要

回帰統計	
重相関 R	0.926356
重決定 R2	0.858136
補正 R2	0.853069
標準誤差	1.283244
観測数	30

ここに区間推定に必要な値が表示される

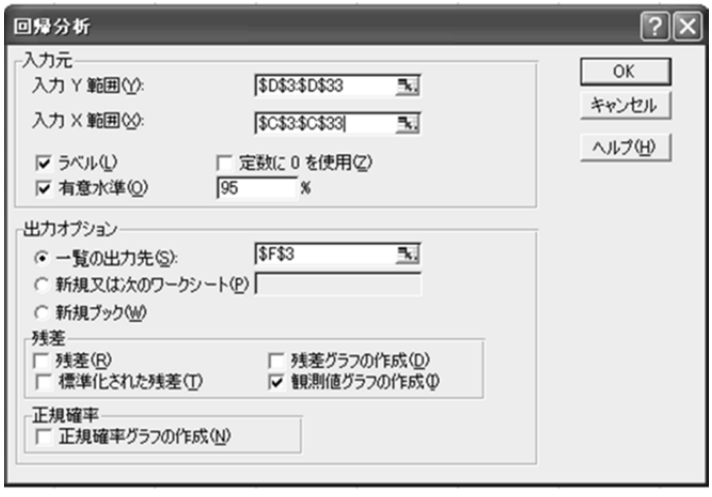
分散分析表

	自由度	変動	分散	測された分散	有意 F
回帰	1	278.9071	278.9071	169.3718	2.15E-13
残差	28	46.10803	1.646715		
合計	29	325.0151			

	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	21.50633	1.025847	20.96446	1.18E-18	19.40498	23.60769	18.67165	24.34102
N%(T)	9.2229	0.708675	13.01429	2.15E-13	7.771244	10.67456	7.264645	11.18116

ここに95%信頼区間が表示される　ここに指定した確率における信頼区間が表示される

信頼率の指定は右下のようにする



右上の表から

母回帰係数の 95%信頼区間を付けた区間推定値

$7.77 \le \beta \le 10.67$　（単位　g/L）

99%の信頼率では　$7.26 \leqq \beta \leqq 11.18$（単位　g/L）

すなわち飲み水を 1L 増やすとウズラの体重は 95%の確率で 7.77～10.67g 増える

母切片の推定もできる．エクセルでは 95%信頼区間をつけた母切片区間推定（さらに p%信頼区間をつけた母切片の区間推定）が計算できる．なお，切片は $x = 0$ のときの y の値であるから，$x = 0$ があり得ないデータについて，母切片の推定をしても意味はない．例えば父と子の身長についての回帰式では，父の身長が 0 であることはあり得ない．この場合，切片について推定・検定はしても意味はない．

母切片の 95%信頼区間を付けた区間推定値　$19.40 \le \alpha \le 23.61$　（単位　g）

予習問題

以下のデータは風力発電における風速と発電量の関係である．母回帰係数を信頼率 95%で区間推定せよ．

風速（m/s）	発電量（MWh/y）
7.5	179.1
6.8	82.7
6.2	47.9
6.4	27.4
7.6	43.5
7.9	81.5
8.4	77.8
5.1	14.3
4.2	4.4
3.8	1.2
5.3	7.5
6.3	54.7
7.9	81.5
7.6	43.5
5.1	14.3
6.8	82.7
7.5	179.1
6.2	47.9
5.1	14.3
6.8	82.7

母回帰係数は信頼率 95%で

（　　）≦β≦（　　）

単位（　　）

２．母回帰係数と母切片の検定

回帰分析で独立変数 x が従属変数 y に対して何らかの効果を与えていることを検定するには母回帰係数が 0 であるという帰無仮説を立て，これを棄却できるかを検定すればよい．母回帰係数が 0 であるとは独立変数 x とは無関係に，従属変数 y が決まることを示す．

帰無仮説：母回帰係数は 0 である（x は y を変化させない）

対立仮説：母回帰係数は 0 でない（x は y を変える）

帰無仮説が棄却された場合

回帰直線の傾きが 0 でない．

独立変数 x を変化させると， y が変化することを示す．

独立変数 x で表される因子は y に影響を与えることを示す．

相関の場合と同じく，回帰が有意であるからといって現実に独立変数 x が y に直接的に効果を与えているとは限らない．

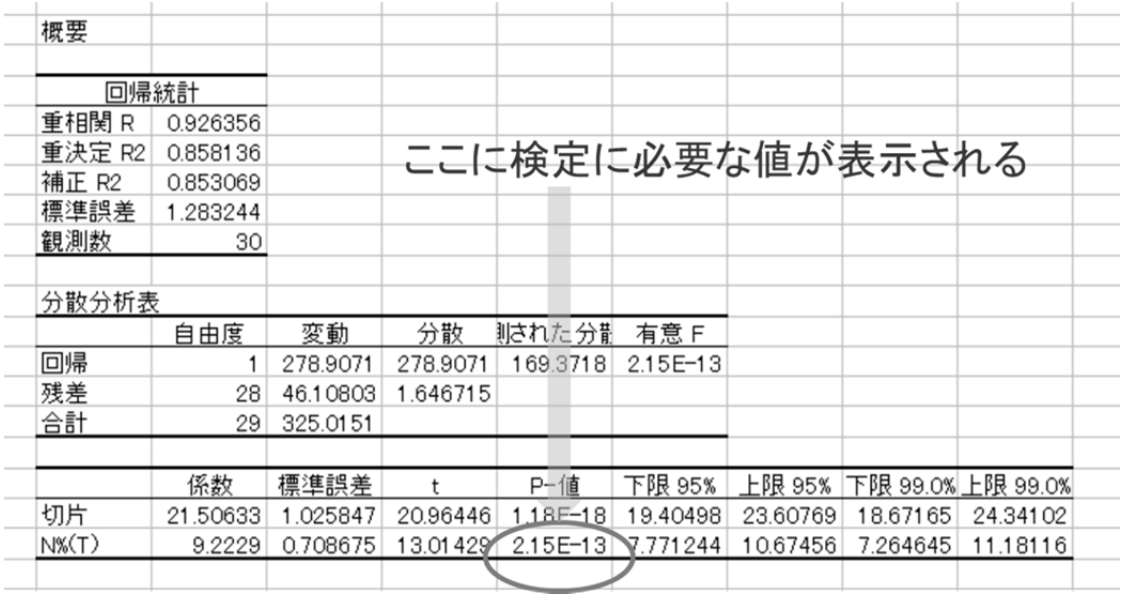
また，p-値が小さいことと両変数間の関係が強いこととは関係がない．寄与率＝$\frac{S_R}{S_T}$（エクセルでは重決定 R2）によって独立変数がどれだけ従属変数を決定しているかを示すことができる．

帰無仮説が棄却できなかった場合

回帰直線の傾きが 0 でないとはいえない

独立変数 x で表される因子は y に影響を与えたとはいえない．

エクセルでは回帰および回帰係数の検定は有意Ｆ（p-値と同じである）をみて，検定する．



概要

回帰統計	
重相関 R	0.926356
重決定 R2	0.858136
補正 R2	0.853069
標準誤差	1.283244
観測数	30

分散分析表

	自由度	変動	分散	測された分散	有意 F
回帰	1	278.9071	278.9071	169.3718	2.15E-13
残差	28	46.10803	1.646715		
合計	29	325.0151			

	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	21.50633	1.025847	20.96446	1.18E-18	19.40498	23.60769	18.67165	24.34102
N%(T)	9.2229	0.708675	13.01429	2.15E-13	7.771244	10.67456	7.264645	11.18116

帰無仮説：母回帰係数は 0 である．ウズラの体重は飲み水の量によって変わらない
対立仮説：母回帰係数は 0 でない．ウズラの体重は飲み水の量によって変わる
である．

回帰係数の p-値は（　　）なので，5%の有意水準で（　有意である　・　有意でない　）．したがって，（　母回帰係数は 0 でない，すなわちウズラの体重は飲み水によって変わる　・　母回帰係数は 0 でないとはいえない，すなわちウズラの体重は飲み水によって変わるとはいえない　）

単回帰分析のときは，実際のところは回帰の分散分析の結果（このプリント p6）と母回帰係数の検定の結果は一致する．どちらも飲み水とウズラの体重の回帰と母回帰係数の例でどちらも p 値は 2.15×10^{-13} となり，同じ結果となる．

次回，学ぶ重回帰分析では説明変数 x が 2 個以上となり，その場合は，それぞれの説明変数について母回帰係数を検定できるので，回帰と母回帰係数の p 値が異なる．

予習問題

以下のデータは風力発電における風速と発電量の関係である．母回帰係数が 0 であるという帰無仮説を検定せよ．

帰無仮説：

対立仮説：

母回帰係数に関する p-値は（　　）なので，風速の効果は 5%の有意水準で（　有意である　・　有意でない　）．したがって，母回帰係数は　（　0 である　・　0 でない　・　0 でないとはいえない）

風速（m/s）	発電量（MWh/y）
7.5	179.1
6.8	82.7
6.2	47.9
6.4	27.4
7.6	43.5
7.9	81.5
8.4	77.8
5.1	14.3
4.2	4.4
3.8	1.2
5.3	7.5
6.3	54.7
7.9	81.5
7.6	43.5
5.1	14.3
6.8	82.7
7.5	179.1
6.2	47.9
5.1	14.3
6.8	82.7

D．宿題

宿題は https://moodle.cerd.shimane-u.ac.jp/moodle/を見てください．