

クラス担当教員名※ <sup>1</sup>	学部・学科名	学籍番号※ <sup>2</sup>	氏名※ <sup>2</sup>	提出日

※1:履修登録したクラスの担当教員名を書く ※2:学籍番号及び氏名が未記入のもの、また授業終了後に提出されたものは採点しないので、注意すること。

### 3 単回帰分析

#### Tips

前回講義では、データの散らばり具合を表す偏差平方和、分散や標準偏差、また2変数の関係を表す相関係数を、Excelで数回のステップに分けて求めました。考え方を学ぶといううえでは計算手順を確認することは必要なことですが、毎回、同じように計算をしているのでは時間ももったいないですね。Excelには便利な関数があって、前回講義で学んだ4つの指標を簡単に計算することができます。詳しくは、講義中に習ったExcelのヘルプで調べることができます。

表 3-1 便利な Excel 関数

	関数	簡単な使い方
偏差平方和	DEVSQ	“=DEVSQ(C4:C10)” と入力すると、C4 から C10 のセルにあるデータの偏差平方和が計算される。
分散	VARP	“=VARP C4:C10)” と入力すると、C4 から C10 のセルにあるデータの分散が計算される。
標準偏差	STDEVP	“=STDEVP(C4:C10)” と入力すると、C4 から C10 のセルにあるデータの標準偏差が計算される。
相関係数	CORREL	“=CORREL(C5:C28,D5:D28)” と入力すると、C5 から C28 のセルにあるデータと D5 から D28 のセルにあるデータの相関係数が計算される。

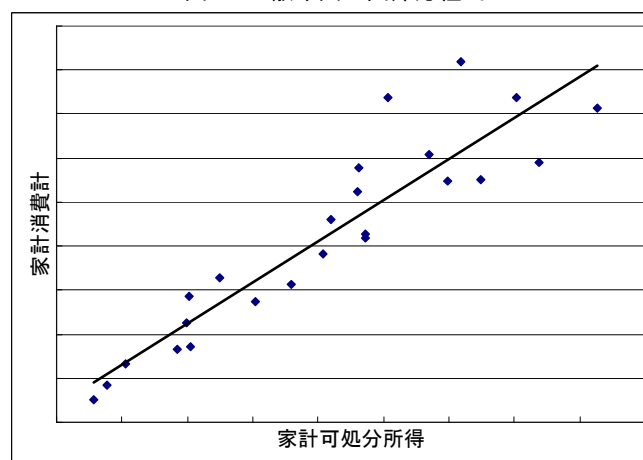
注) 詳しい使い方は、Excel 画面の右上にある「質問を入力して下さい」に調べたい関数や求めたい指標を入力して検索してみよう。

#### 3-1 回帰分析とは

前回、コンビニ M&M の浦安地域マネージャーのあなたは、浦安市に新規出店して十分に売上げがあるのか（消費者がどのくらい購買意欲 (= 支出) があるのか）を考えるため、家計可処分所得と家計消費支出の散布図を描き検討しました。また、その相関係数は 0.9951 と非常に大きく、可処分所得が増加すると家計消費支出が増加する傾向が強いことが分かりました。そこで、あなたは将来の売上げを考えるうえで参考となる将来の家計消費支出を回帰分析で予測することにしました。

**回帰分析**とは、2変数 X, Z のデータがあるとき、Z を X で定量的に説明する**回帰式(あるいは、回帰方程式)**とよばれる式を求めることを目的とします。前回の作業で描いた散布図は、家計可処分所得と家計消費支出の2種類のデータが平面上に点となって散らばっていました。簡単にいうと、**回帰方程式は、その散らばっている無数の点を最もよく表現できる直線のこと**です。

図 3-1 散布図と回帰方程式



クラス担当教員名※ <sup>1</sup>	学部・学科名	学籍番号※ <sup>2</sup>	氏名※ <sup>2</sup>	提出日

※1:履修登録したクラスの担当教員名を書く ※2:学籍番号及び氏名が未記入のもの、また授業終了後に提出されたものは採点しないので、注意すること。

前々回、検討したように、消費支出は、「消費支出  $C$  は所得  $Y$  に依存するが、**限界消費性向**は1未満の正の数である」という仮説をたてて、次式のように線形関数に特定化しました。消費支出  $C$  のように「説明される変数」は**被説明変数**（あるいは従属変数や内生変数）とよばれ、所得  $Y$  のように「説明する変数」は**説明変数**（あるいは独立変数や外生変数）とよべれます。また、 $a$  は切片、 $b$  は傾きといい、所得  $Y$  が1単位の増加に対する消費支出  $C$  の増分を示す限界消費性向を表しています。

$$\text{回帰式: } C = a + bY, a \geq 0 \text{ and } 0 < b < 1 \quad \dots(1)$$

しかし、(1)式の消費関数には、定量的に把握することができないデータ（夏のすごしやすさや花粉症のひどさのこと）の影響が考慮されていません。そこで、 $i$  番目の観測データ  $(C_i, Y_i)$  では把握できないバラツキ（誤差）である  $\varepsilon_i$  を考慮したモデルが、次式の確率モデルです。この方程式は母回帰式（あるいは、母回帰方程式）とよばれ、 $a$  や  $b$  は母回帰係数とよべれます。

$$\text{母回帰式: } C_i = a + bY_i + \varepsilon_i, a \geq 0 \text{ and } 0 < b < 1 \quad \dots(2)$$

説明変数  $Y_i$  や誤差項  $\varepsilon_i$  が、以下の4つの条件、

- 条件①:  $Y_i$  は確率変数ではなく、確定した値をとる。
- 条件②:  $\varepsilon_i$  は確率変数で期待値  $0$ 。  $E(\varepsilon_i) = 0, i = 1, 2, \dots, n$ 。
- 条件③: 異なった誤差項は無相関。すなわち、 $i \neq j$  であれば、 $\varepsilon_i$  と  $\varepsilon_j$  の共分散は  $0$ 。
- 条件④: 誤差項の分散が一定で  $\sigma^2$ 。

を満たすとき、消費支出  $C_i$  の期待値は、次式のようになります。

$$E[C_i] = a + bY_i \quad \dots(3)$$

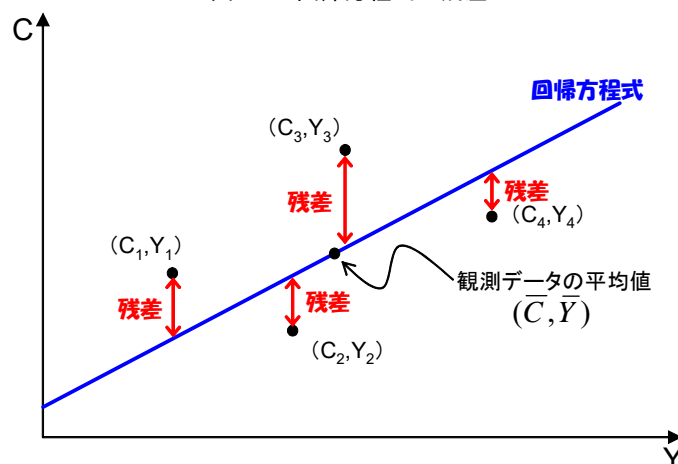
すなわち、説明変数や誤差項が①～④の条件を満たすと考えられるとき、それぞれの観測データで把握することができないバラツキ（誤差）をならした平均的な直線は、回帰式にほかならないということです。

### 3-2 最小二乗法

誤差項  $\varepsilon_i$  は、(2)式を変形した次式から、消費支出の**観測値**  $C_i$  と回帰式  $(a + bY_i)$  から計算される**理論値(あてはまり値)**を差し引いた数値であることが分かります。誤差項は、散布図では下図の縦矢印部分で表され、**残差**とよべれます。

$$\text{誤差項: } \varepsilon_i = C_i - (a + bY_i) \quad \dots(4)$$

図 3-2 回帰方程式と残差



回帰式は、この残差が最も小さくなるような母回帰係数  $a, b$  として求められます。ただし、(4)式で表される残差は、プラスやマイナスの両方があり相殺されることがあるため、回帰式は残差を**二乗**したものの合計が**最も小さく**なるように求められます。この方法を**最小二乗法**といいます。

さて、この最小二乗法をつかって、母回帰係数をどうやって求めればいいのか？ みなさんが苦手な微分を使います^^; 講義時間が足りませんので、ここでは具体的な求め方のステップだけ紹介するにとどめておきます。詳しく知りたい方は、計量経済学や多変量解析のテキストをじっくり読んで下さい。

STEP1: 残差平方和を求める母回帰係数  $a, b$  で偏微分したものを  $0$  とする（一階条件）。

クラス担当教員名 <sup>※1</sup>	学部・学科名	学籍番号 <sup>※2</sup>	氏名 <sup>※2</sup>	提出日

※1:履修登録したクラスの担当教員名を書く ※2:学籍番号及び氏名が未記入のもの、また授業終了後に提出されたものは採点しないので、注意すること。

STEP2: 正規方程式 (a と b に関する連立方程式) から a, b を算出する。

以上の STEP から母回帰係数の最小二乗推定値が次式のように計算されます。

$$\text{母回帰係数の推定値:} \quad \dots(5)$$

ここで、母回帰係数 (推定値)  $\hat{b}$  の分母にある偏差平方和は前回学んだ方法で計算でき、分子にある偏差積和は次式で計算できます。なお、 $\bar{C}, \bar{Y}$  は「C バー」「Y バー」と読み、 $C, Y$  の平均値を表します。(5)式から分かるように、回帰式は必ず観測データの平均値 ( $\bar{C}, \bar{Y}$ ) を通ります。

$$\text{偏差積和:} \quad \dots(6)$$

作業 1. econome4.xlsx を DL して、\*\*\*\*data04.xlsx と名前を変えて各自の UF の計量分析フォルダの中にセーブしてください。sheet name “家計消費支出 (その 3)” で、ステップ①～⑧の順番で作業を行って、母回帰係数(推定値)  $\hat{a}, \hat{b}$  を求めなさい。\*\*\*\*は「学籍番号+氏名」。(2 点×8)

作業 2. \*\*\*\*data04.xlsx の sheet name “家計消費支出 (その 3)” において、作業 1 で推定した母回帰係数を使い、家計消費の理論値 (あてはまり値) を求めなさい (⑨に記入)。また、家計消費支出と家計可処分所得の散布図に観測値と理論値をプロットしなさい。(2 点, 図: 5 点)

- **観測値は、実際に観測された値**のことである。
- **理論値は、理論上の推計値**である。回帰方程式を用いて計算される。
- $Y_i, C_i$  の小文字  $i$  は同一のデータセットを表す。1980 年から 2003 年までの各年の所得  $Y$  と消費  $C$  が観測されたとする。 $(Y_i, C_i)$  は、 $(Y_{1980}, C_{1980}), (Y_{1981}, C_{1981}), \dots, (Y_{2003}, C_{2003})$  を表す。この例では、 $Y_i$  が  $i$  年の所得の観測値を、 $C_i$  が  $i$  年の消費の観測値を表す。
- $\hat{Y}, \hat{C}$  は「Y ハット」「C ハット」と読み、 $Y, C$  の**理論値**を表す。
- $\bar{Y}, \bar{C}$  は「Y バー」「C バー」と読み、 $Y, C$  の**平均値**を表す。

クラス担当教員名※ <sup>1</sup>	学部・学科名	学籍番号※ <sup>2</sup>	氏名※ <sup>2</sup>	提出日

※1:履修登録したクラスの担当教員名を書く ※2:学籍番号及び氏名が未記入のもの、また授業終了後に提出されたものは採点しないので、注意すること。

### 3-3 回帰分析の意味（重要です）

それでは、作業1~2で推定した回帰式（下式）を使って、コンビニ M&M の浦安地域マネージャーの業務に役立ててみよう。前回の作業では、「家計消費計」と「家計可処分所得」の関係を調べるのが目的でした。マネージャーとしては、新規出店を検討している浦安地域の消費動向が将来どのように変化するか気になるところです。ここでは、説明変数の「家計可処分所得」で、被説明変数の「家計消費計」を説明していました。すなわち、将来の「家計可処分所得」が分かれば、将来の「家計消費計」を予測することができるわけです。

$$\text{推定した回帰式} : C_i = -23804 + 0.9711Y_i$$

回帰式からは重要な情報が読み取れます。回帰式の **0.9711 という値は、「家計可処分所得(Y)が1単位増加したとき、「家計消費計(C)が0.9711だけ増加する」ということを表しています。家計可処分所得と家計消費計はともに10億円単位の値でしたので、家計可処分所得が1単位(すなわち10億円)増加するとき、家計消費計は0.9711単位(すなわち9.711億円)増加すると読み取れます。このように、回帰式(回帰係数)を推定すれば、説明変数(家計可処分所得)の値を代入することで、被説明変数(家計消費計)を予測することができます。これって、かなりすごいことですよ？ いっちょ、株とか土地の価格予測モデルを考えて…なんてこともできるかも^^;**

作業3. rep04.docx を DL して、\*\*\*\*rep04.docx と名前を変えて各自の UF の計量分析の中にセーブしてください。

\*\*\*\*は「学籍番号+氏名」。作業1で推定した回帰係数を使って教えてください。(5点×3)

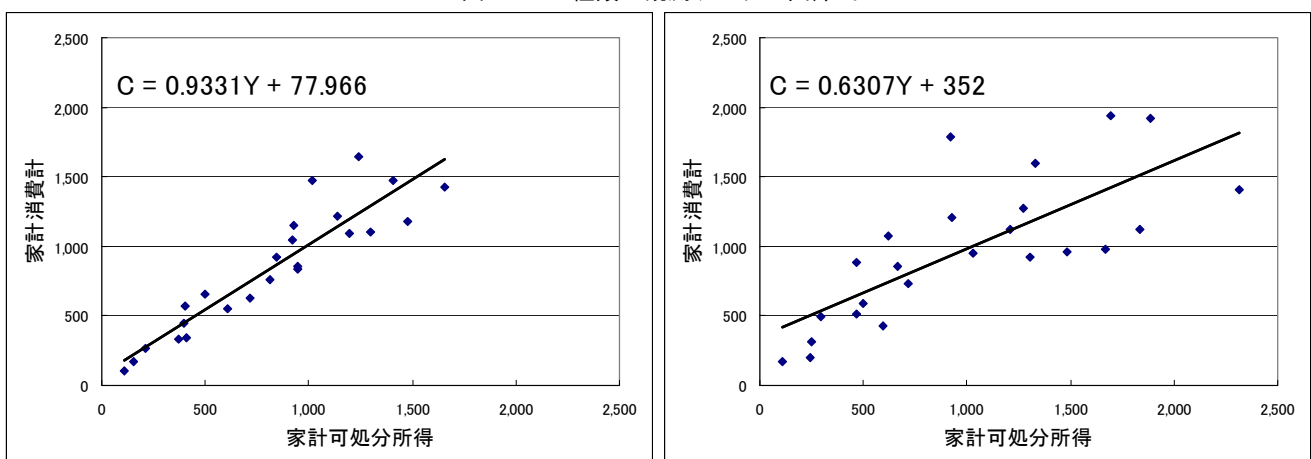
- (1)家計可処分所得が3,500,000億円(単位に注意)のとき、家計消費計はいくらになりますか？計算の仕方も説明してください。
- (2)家計可処分所得が1,000億円(単位に注意)増加するとき、家計消費計はいくら増減しますか？計算の仕方も説明してください。
- (3)家計可処分所得が5,000億円(単位に注意)減少するとき、家計消費計はいくら増減しますか？計算の仕方も説明してください。

ただし、**被説明変数の予測値(理論値)の精度が低ければ、予測には使えない**ですよ(よく外れる天気予報は当てになりませんよね?)。この点を少し考えてみましょう。

### 3-4 回帰式の精度

回帰式は、観測データさえあれば、最小二乗法の考え方で求めることができます。しかし、求めた**回帰式が優れているかどうかはきちんと確認する必要があります**。下図を見比べてみてください、何か違いがありませんか？

図 3-3 2種類の観測データと回帰式



- 右図と比べると、左図のほうの点と回帰式がよりマッチしている（回帰式の周りに無数の点が集まっている）。

左図のように、**点(観測データ)と回帰式がマッチしているほど、回帰式の精度が高い**といえます。しかし、回帰式の精度をグラフからマッチしているかどうか目で見て判断しては、グラフを見る人の主観が入ってしまいます。そこで、データの散らばり具合を表す客観的な指標として標準偏差や相関係数などを考えたように、回帰式の精度を表す客観的な指標としては**重相関係数**や**決定係数(あるいは**

クラス担当教員名※ <sup>1</sup>	学部・学科名	学籍番号※ <sup>2</sup>	氏名※ <sup>2</sup>	提出日

※1:履修登録したクラスの担当教員名を書く ※2:学籍番号及び氏名が未記入のもの、また授業終了後に提出されたものは採点しないので、注意すること。

は、**奇与率**)があります。

重相関係数は、消費支出の観測値  $C$  と回帰式から計算される理論値  $\hat{C}$  の相関係数のことで、次式で計算できます。(一般的に、理論値やパラメータの推定値にはハット記号が使われます)

重相関係数： …(7)

前回講義で学んだ相関係数の公式とは若干異なっていることに注意して下さい。(共分散が偏差積和へ、標準偏差が偏差平方和へ変わっています)

- 重相関係数は、観測値と理論値の相関係数である。このため、回帰式の精度が高いほど 1 に近づき、精度が悪くなるほど 0 に近づく。

回帰式の精度を表す 2 つ目の指標である決定係数は、重相関係数の二乗に等しい。決定係数は消費支出の観測値  $C_i$  の平均値まわりでの変動和 (**総変動**という) と、理論値  $\hat{C}$  のその平均まわりでの変動和 (**回帰変動**という) の比率に等しく、回帰式が観測値をどのくらい説明しているかを表している。総変動のうち、回帰変動で説明できないものは残差による変動になる。したがって、次式のように式変形できる。

決定係数： …(8)

総変動、残差変動： …(9)

- 決定係数は、観測値を理論値がどれだけ説明しているかを表し、回帰式の精度が高いほど 1 に近づき、精度が悪くなるほど 0 に近づく。
- 決定係数が 1 のとき、観測値は回帰式によって 100% 説明されている。

作業 4. \*\*\*\*data04.xlsx の sheet name “家計消費支出 (その 4)” で、作業 1 で推定した母回帰係数  $\hat{a}$ ,  $\hat{b}$  を利用して (作業 1 が完了していればすでに貼り付けてある)、回帰式の決定係数をステップ③~⑧の順番で求めなさい。回帰式の精度について、計算した決定係数を用いてコメント欄に説明しなさい。(2 点×8)

クラス担当教員名※ <sup>1</sup>	学部・学科名	学籍番号※ <sup>2</sup>	氏名※ <sup>2</sup>	提出日

※1:履修登録したクラスの担当教員名を書く ※2:学籍番号及び氏名が未記入のもの、また授業終了後に提出されたものは採点しないので、注意すること。

### 3-5 説明変数の採択判断は？

すでに学んだ相関係数から、家計消費支出と家計可処分所得の間には強い正の相関関係があることがわかりました。そこで、家計可処分所得を説明変数とする線形関数で家計消費支出を説明する回帰分析を行った結果、その回帰式は決定係数が高く、説明力が高いことがわかりました。

決定係数が高いということは、モデルの説明力が高いということを意味しています。しかし、決定係数は、(今回は説明変数が1つの単回帰分析でしたが)モデルに含まれている説明変数が本当に必要なのか、それとも必要でないのかについての情報は与えてくれません。回帰分析では、回帰係数を推定するだけでなく、その説明変数が被説明変数に影響を及ぼしているのか(すなわち、回帰係数が0ではない値をとること)についての検定を行わなくてはなりません。特に、この検定は複数の説明変数を採用した重回帰分析(次回から学びます)で重要になります。

### 3-6 t 検定

回帰係数は、説明変数が被説明変数をどのように説明しているのか(値の正負や大きさ)を表す重要なパラメータです。ですから、最小二乗法で推定された回帰係数が信頼できるのか、もしかすると真の値が0ではないのか(説明変数が被説明変数に全く影響を及ぼさないのではないのか)を検定することは非常に重要です。

一般的に、仮説の検定では、望ましくない、検定で否定されることを期待される仮説内容を**帰無仮説(きむかせつ)**とし、その**仮説を棄却して無に帰すようにして検証**します。また、**対立仮説**としては望ましい仮説内容をたてます。

回帰分析では、説明変数が被説明変数を説明することができるのか、すなわち、回帰係数が0ではないことを検証することになります。したがって、帰無仮説は、説明変数が被説明変数を全く説明していない(影響を及ぼさない)という意味の「回帰係数が0に等しい」を仮説とし、対立仮説としては、説明変数が被説明変数を正負やその大きさは分からないが説明しているという意味の「回帰係数が0と等しくない」を仮説とします。なお、回帰係数の符号条件が分かっている場合には、下表②や③のように対立仮説をたてることもできます。下表の帰無仮説や対立仮説は回帰係数が $\theta$ に一致するかどうかを検証するように、より一般的に書いてありますが、 $\theta = 0$ とすれば通常の回帰係数の検定になります。

表 3-2 帰無仮説と対立仮説

帰無仮説 $H_0$	回帰係数は $\theta$ に等しい。 $b = \theta$
対立仮説 $H_1$	① (両側検定) 回帰係数は $\theta$ と等しくない。 $b \neq \theta$
	② (右片側検定) 回帰係数は $\theta$ より大きい。 $b > \theta$
	③ (左片側検定) 回帰係数は $\theta$ より小さい。 $b < \theta$

※  $\theta = 0$  のとき、通常の回帰係数の検定。

仮説の検証には、次式から計算される**t統計検定量 (t値)**が使われます。(詳しい計算方法は7ページ以降を読んでください)

$$t\text{値} : \dots(19)$$

帰無仮説  $H_0$  (回帰係数は0に等しい;  $\theta = 0$ ) が正しいとき、すなわち、説明変数が被説明変数を系統的に説明できないということが正しいならば、回帰係数の推定値  $\hat{b}$  は0近辺の値をとるため、t値も0近辺の値をとります。

一方、対立仮説  $H_1$  (回帰係数は0と等しくない) が正しいならば、回帰係数の推定値  $\hat{b}$  は0ではない値をとるため、t値も0ではない値をとることになります。また、回帰係数の推定値  $\hat{b}$  がその標準誤差  $\hat{\sigma}$  に比べて大きな値をとればほど、t値は絶対値で大きな値となっていきます。t値が0から離れれば離れるほど、回帰係数の推定値  $\hat{b}$  は0近辺の値をとるという帰無仮説  $H_0$  にとって不利な証拠となり、帰無仮説  $H_0$  を棄却する根拠となります。

難しく説明しましたが^^;、一般に、**t値の絶対値が(概ね)2以上であれば回帰係数は有意である(その回帰係数は0ではない、すなわち、回帰式にその説明変数が必要である)と判断**されます。ただし、回帰式を推計するときにつかうデータセットの個数が20前後の少ないときにはt分布表を使って判断する必要があります。なお、**t値が大きければ、その回帰係数が0ではない可能性が高いことを示しているに過ぎません。推定された回帰係数が真の値である可能性が高いことを表していることではないことに注意して下さい。**

表 3-3 t 検定

t 値	仮説は
t 値の絶対値が 2 以上	回帰係数は 0 ではない (帰無仮説は棄却される)
t 値の絶対値が 2 未満	回帰係数が 0 の可能性がある (帰無仮説を棄却できない)

クラス担当教員名※ <sup>1</sup>	学部・学科名	学籍番号※ <sup>2</sup>	氏名※ <sup>2</sup>	提出日

※1:履修登録したクラスの担当教員名を書く ※2:学籍番号及び氏名が未記入のもの、また授業終了後に提出されたものは採点しないので、注意すること。

### Tips

前回講義では、母回帰係数や重相関係数・決定係数を、Excel で数回のステップに分けて求めました。考え方を学ぶといううえでは必要なことですが、毎回、同じように計算をしていては時間ももったいないですね。そこで、Excel にある便利な関数を紹介しますので、このさい覚えちゃいましょう～。

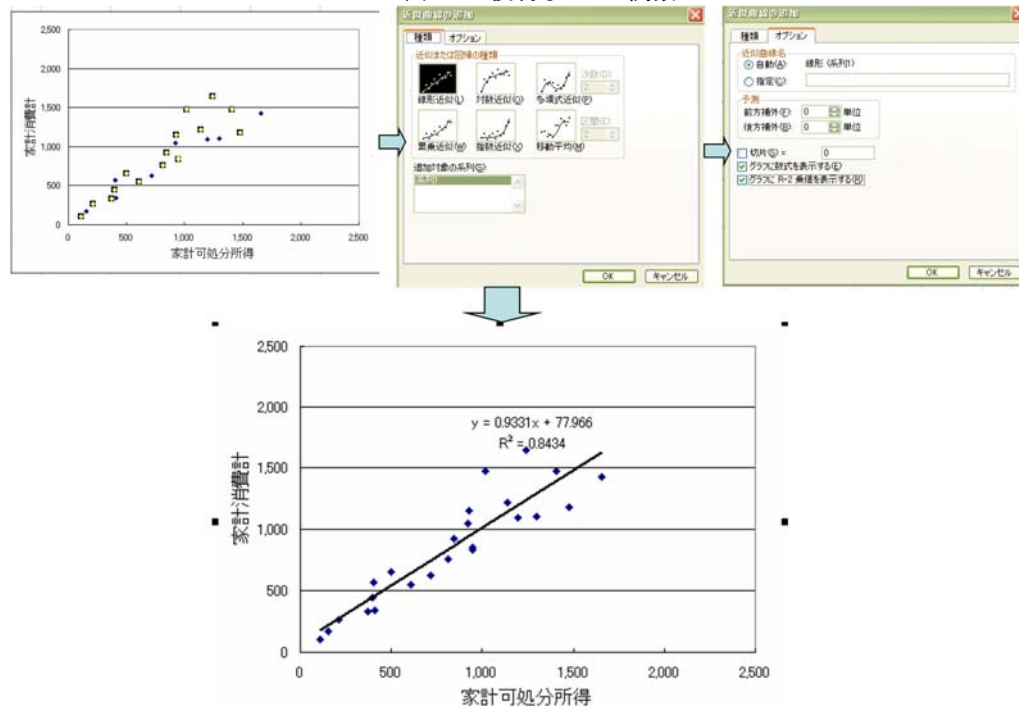
表 3-4 便利な Excel 関数

	関数	簡単な使い方
母回帰係数の切片	INTERCEPT	“=INTERCEPT(D5:D28,C5:C28)” と入力すると、D5 から D28 のセルにあるデータを被説明変数とし、C5 から C28 のセルにあるデータを説明変数とした線形回帰の母回帰係数 (切片) が計算される。
母回帰係数の傾き	LINEST	“=LINEST(D5:D28,C5:C28)” と入力すると、D5 から D28 のセルにあるデータを被説明変数とし、C5 から C28 のセルにあるデータを説明変数とした線形回帰の母回帰係数 (傾き) が計算される。
重相関係数	CORREL	“=CORREL(C5:C28,D5:D28)” と入力すると、C5 から C28 のセルにあるデータと D5 から D28 のセルにあるデータの相関係数が計算される。
決定係数	RSQ	“=RSQ(D5:D28,C5:C28)” と入力すると、D5 から D28 のセルにあるデータを被説明変数とし、C5 から C28 のセルにあるデータを説明変数とした線形回帰の決定係数が計算される。

注) 詳しい使い方は、Excel 画面の右上にある「質問を入力して下さい」に調べたい関数や求めたい指標を入力して検索してみよう。

もう一つ、単相関分析の簡単なやり方も習得しましょう。まず、二変数の散布図を作成します。次に、散布図の一つのポイントを右クリックするとメニューリストが出ますので、近似値曲線の追加をクリックしてください。すると、下図のウィザードが出てきます。種類を線形近似として、オプションから「グラフに数式を表示する」と、「グラフに R-2 乗値を表示する」を指定すると、理論式と単回帰の結果が表示されます。簡単な分析にはこれが使えます。

図 3-4 便利な Excel 関数



クラス担当教員名 <sup>※1</sup>	学部・学科名	学籍番号 <sup>※2</sup>	氏名 <sup>※2</sup>	提出日

※1:履修登録したクラスの担当教員名を書く ※2:学籍番号及び氏名が未記入のもの、また授業終了後に提出されたものは採点しないので、注意すること。

【参考:大学院への進学を考えている人は読んでね】

### t 値の計算方法

t 値は下式から計算することが出来ます。分子にある回帰係数の推定値についての計算方法はすでに学んだ計算と同じです。分母にある“回帰係数の標準誤差”の計算は以下の通りです。

$$t \text{ 値} : t = \frac{(\text{回帰係数の推定値}) - \theta}{(\text{回帰係数の標準誤差})} = \frac{\hat{b} - \theta}{s(\hat{b})}$$

(5)式で推定される母回帰係数は、母集団に含まれている観測された標本（サンプル）を使って最小二乗法で推定されたもの（平均値）なので、ある程度のバラツキを持っています。最小二乗法から計算される母回帰係数を  $a = \bar{C} - b\bar{Y}$  の関係を使って式変形すると、

$$\begin{aligned} \text{母回帰係数} : \hat{b} &= \frac{C \text{ と } Y \text{ の偏差積和}}{Y \text{ の偏差平方和}} = \frac{\sum_i (C_i - \bar{C})(Y_i - \bar{Y})}{\sum_i (Y_i - \bar{Y})^2} = \frac{\sum_i (a + bY_i + \varepsilon_i - \bar{C})(Y_i - \bar{Y})}{\sum_i (Y_i - \bar{Y})^2} \\ &= \frac{\sum_i \{b(Y_i - \bar{Y}) + \varepsilon_i\}(Y_i - \bar{Y})}{\sum_i (Y_i - \bar{Y})^2} = b + \frac{\sum_i (Y_i - \bar{Y})\varepsilon_i}{\sum_i (Y_i - \bar{Y})^2} \quad \dots(10) \end{aligned}$$

が得られます。すなわち、母回帰係数は  $\varepsilon_i$  の確率変数の影響を受けているため、母回帰係数も確率変数としてバラツキが存在します。ここで、最小二乗法的前提条件①～④を思い出すと、条件②確率変数  $\varepsilon_i$  は期待値 0 ( $E(\varepsilon_i) = 0$ ) なので、 $(Y_i - \bar{Y})$  が確率変数でなければ（条件①）、第 2 項の期待値は 0 となります。

$$E(\hat{b}) = b \quad \dots(11)$$

すなわち、母回帰係数  $\hat{b}$  の期待値は  $b$  となります。さらに、次の条件を追加すると、

**条件⑤** : 誤差項  $\varepsilon_i$  は正規分布。

母回帰係数は、確定値  $b$  と正規分布に従う確率変数  $\varepsilon_i$  項の和で表すことができます。したがって、母回帰係数の分布は、

$$N\left(b, \frac{\sigma^2}{\sum_i (Y_i - \bar{Y})^2}\right) \quad \dots(12)$$

の正規分布に従うこととなります（時間がなくて証明は省略。詳しく知りたい方は、計量経済分析や多変量解析のテキストをじっくり読みましょう）。母回帰係数の推定値のバラツキ（標準偏差）は**標準誤差**と呼ばれ、母回帰係数の分散の平方根に等しくなります。母回帰係数の分散に含まれる誤差  $\varepsilon_i$  の分散が分かれば、母回帰係数の分散（あるいは標準偏差）の真の値は計算できます。しかし、観測値が母集団の一部であるため、母集団の誤差の分散（真の値）は知ることができません。このため、母集団の誤差の分散の代わりに推計される誤差、すなわち残差の分散  $\sigma^2$  が利用されます。

$$\text{回帰係数 } \hat{b} \text{ の分散} : \text{var}(\hat{b}) = \frac{\sigma^2}{\sum_i (Y_i - \bar{Y})^2} \quad \dots(13)$$

$$\text{回帰係数 } \hat{b} \text{ の標準誤差} : S(\hat{b}) = \frac{\hat{\sigma}}{\sqrt{\sum_i (Y_i - \bar{Y})^2}} \quad \dots(14)$$

母回帰方程式が次式で表されると、

$$\text{母回帰方程式} : C_i = a + bY_i + \varepsilon_i \quad \dots(15)$$



クラス担当教員名※ <sup>1</sup>	学部・学科名	学籍番号※ <sup>2</sup>	氏名※ <sup>2</sup>	提出日

※1:履修登録したクラスの担当教員名を書く ※2:学籍番号及び氏名が未記入のもの、また授業終了後に提出されたものは採点しないので、注意すること。

### 【参考:大学院への進学を考えている人は読んでね】

すでに学んだ標準偏差の公式から、母集団の誤差の標準偏差は次式で表されます。

$$\text{母集団の誤差の分散: } \sigma^2 = \frac{\sum_i (\text{母集団の真の誤差})^2}{\text{データセットの個数}} = \frac{\sum_i (\varepsilon_i)^2}{n} \quad \dots(16)$$

しかし、母集団の真の回帰係数が分からないのと同じように、母集団の誤差の標準偏差も分かりません。最小二乗法から推計された回帰式を利用すれば、理論値と観測値の差から残差（推定値としての誤差）を計算することができます。

$$\text{残差: } e_i = C_i - \hat{C}_i = C_i - (\hat{a} + \hat{b}Y_i) \quad \dots(17)$$

残差の分散や標準偏差は、残差  $e_i$  を使って計算できます。残差の平均値は0なので、残差の偏差平方和は  $\sum_i (e_i)^2$  となります。母集団の場合にはサンプル数  $n$  で割ることで分散を計算できましたが、残差  $e_i$  を計算する際に、2つの母回帰係数を必要とし、その計算に自由度を2つ使ってしまったので、分母はサンプル数  $n$  から2を差し引いたものとなり、その分だけ計算される誤差の分散は大きくなります。

$$\text{(推定値の) 誤差の分散: } \hat{\sigma}^2 = \frac{\sum_i (\text{観測値 } C_i - \text{理論値 } \hat{C}_i)^2}{\text{データセット個数} - \text{回帰係数の個数}} = \frac{\sum_i (e_i)^2}{n-k} \quad \dots(18)$$

回帰係数の標準偏差は、(18)式から計算される推定値としての誤差（すなわち残差）の分散の平方根を(14)式に代入することで得られます。

作業 5. \*\*\*\*data04.xls の sheet name “家計消費支出（その5）”で、Tips の方法（次ページ）で母回帰係数  $\hat{a}$ ,  $\hat{b}$  を推定し (①②)、推定値としての誤差の分散（残差の分散）を③～⑤のステップで求めなさい。

作業 6. \*\*\*\*data04.xls の sheet name “家計消費支出（その5）”で、母回帰係数の  $t$  値を⑥～⑨の順番で求め、その回帰係数が採用されるかどうかを⑩で答えなさい。（2点×5）

### t 分布と棄却域について

下図には帰無仮説  $H_0$ （回帰係数は0に等しい）が正しいときの  $t$  分布が描かれています。(19)式から分かるように、帰無仮説  $H_0$ （回帰係数は0に等しい）が正しいときは  $t$  値が0になります。したがって、実際に計算される  $t$  値が0近辺の値となる領域  $(1-\alpha)$  ( $t$  分布図の真ん中部分) に含まれなければ、帰無仮説  $H_0$  が棄却されることとなります。この  $\alpha$  は有意水準と呼ばれ、通常5% (0.05) や1% (0.01) が使われます。 $t_{\alpha/2}$  は有意水準  $\alpha$  のとき  $t$  分布図の両側にある領域  $\alpha$  の確率をあたえる  $t$  値の臨界点であり、 $t$  値が絶対値で  $t_{\alpha/2}$  以上の値をとるとき、帰無仮説  $H_0$ （回帰係数は0に等しい）が棄却されます。なお、有意水準  $\alpha$  のときの  $t$  値の臨界点  $t_{\alpha/2}$  は計量経済学や統計学などのテキストに表としてまとめてあります。

