

<回帰分析>

1. 準備

今回の実習では、あらかじめ河田が作成した**消費関数**のファイルを用いる。

【課題 12】 統計学の講義用 HP から、消費関数のファイルをダウンロードしてみよう。

📖 手順

- ① 徳山大学のトップページ (<http://www.tokuyama-u.ac.jp>) で、「在学生の皆様へ」のメニューから、「教員紹介」を選び、河田の教員紹介を開く。
- ② そのページの下の方にある、「個人サイト」のところをクリックすると、「河田研究室」のページにジャンプする。(ここまでの手順は、<http://www2.tokuyama-u.ac.jp/kawada> とアドレスを直接入力してもよい)
- ③ 「統計学」をクリックし、「第28回 7月21日(木)」の配布資料にある、「消費関数」をクリックし、自分の使いやすい場所に保存する。

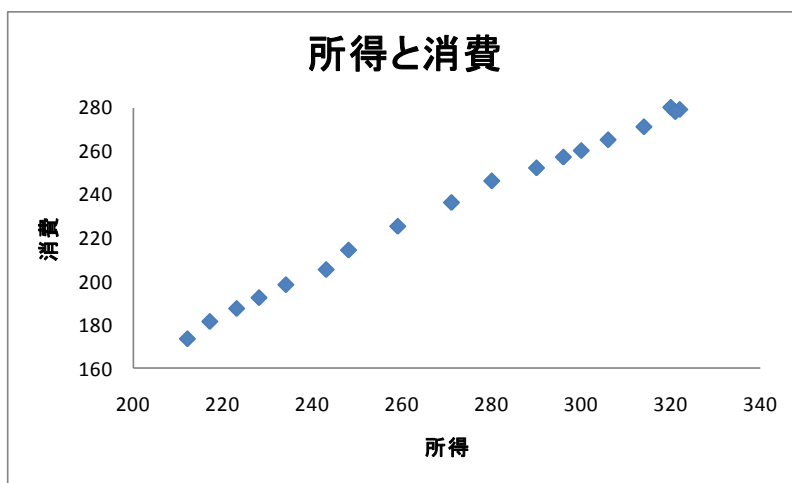
2. 散布図

【課題 13】 消費関数のファイルには、日本の実質家計可処分所得と実質家計最終消費支出のデータ(単位: 兆円)がある。このデータについて、可処分所得を横軸に、消費支出を縦軸にとった散布図を描いてみよう。

📖 散布図は次のような手順で描けばよい。

- ① B2:C19 を範囲指定し、リボン内にグラフのグループにある、散布図 のボタンをクリックする。(挿入タブをクリックすることで表示される。)
- ② 散布図グラフのフォーマット (型式) メニューにおいて、「散布図(マーカーのみ)」(左上) をクリックし、リボンの中のレイアウト 1 (左端) をクリックし、タイトルや軸ラベルなどが書き込めるようにする。
- ③ 下の図のようにタイトル、軸ラベルを入力し、目盛線、凡例を非表示にする。
- ④ グラフの作成を終えた後で、それぞれの軸の書式設定をおこない、

縦軸 最小値 : 160
 最大値 : 280
 目盛間隔 : 20
 横軸 最小値 : 200
 最大値 : 340
 目盛間隔 : 20
 とする。



3. 相関係数の導出

【課題 14】 所得と消費のデータについて相関係数を求めてみよう。

相関係数は次のような式で求められる。

$$R = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}}$$

したがって相関係数を導出するためには、

- 1) x と y の偏差をそれぞれ求め、
- 2) その 2 乗和($\sum(x - \bar{x})^2$, $\sum(y - \bar{y})^2$)、交差積の和($\sum(x - \bar{x})(y - \bar{y})$)を求めたうえで、相関係数を計算する。

📖 手順

- ① B21 セル、C21 セルにそれぞれ平均を求める。B21 なら **=AVERAGE(B2:B19)** とすればよい。
- ② D 列に x の偏差を、E 列に y の偏差を求める。D2 セルに **=B2-B\$21** と入力し、これを D3:D19 セルにコピーする。E 列も同様である。
- ③ F 列に x 偏差 2 乗を、G 列に y 偏差 2 乗を求める。2 乗を表す演算子は ‘^’ であり、F2 セルに **=D2^2** と入力し、これをコピーする。G 列も同様である。
- ④ H 列に偏差交差積を求める。H2 セルに **=D2*E2** と入力し、これをコピーする。
- ⑤ F22:H22 セルに合計を求める。F22 なら **=SUM(F2:F19)** とすればよい。
- ⑥ C24 セルに **=H22/SQRT(F22*G22)** と入力する。相関係数の式と見比べてみよ。

<作成見本>

	A	B	C	D	E	F	G	H
1	年	所得(x)	消費(y)	x偏差	y偏差	x偏差2乗	y偏差2乗	偏差交差積
2	1981	212	173	212	173	44944	29929	36676
3	1982	217	181	217	181	47089	32761	39277
4	1983	223	187	223	187	49729	34969	41701
5	1984	228	192	228	192	51984	36864	43776
6	1985	234	198	234	198	54756	39204	46332
7	1986	243	205	243	205	59049	42025	49815
8	1987	248	214	248	214	61504	45796	53072
9	1988	259	225	259	225	67081	50625	58275
10	1989	271	236	271	236	73441	55696	63956
11	1990	280	246	280	246	78400	60516	68880
12	1991	290	252	290	252	84100	63504	73080
13	1992	296	257	296	257	87616	66049	76072
14	1993	300	260	300	260	90000	67600	78000
15	1994	306	265	306	265	93636	70225	81090
16	1995	314	271	314	271	98596	73441	85094
17	1996	322	279	322	279	103684	77841	89838
18	1997	320	280	320	280	102400	78400	89600
19	1998	321	278	321	278	103041	77284	89238
20								
21	平均	271.3333	233.2778					
22	合計					1351050	1002729	1163772
23								
24		相関係数	0.999863					

1) 単回帰モデル(その 1)

1. 回帰直線の導出

【課題 15】 所得と消費のデータについて $Y = a + bX$ という 1 次式をあてはめ、回帰係数 a, b の推定値を求めよ。(C26 セルに b の推定値、C25 セルに a の推定値を求める。)

回帰係数の推定値を求める式は次のようなものである。

$$\hat{b} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2}$$
$$\hat{a} = \bar{y} - \hat{b}\bar{x}$$

よって、次のようにすればよい。

📖 手順

- ① C26 に b の推定値を求める。=**H22/F22** とすれば求まる。
- ② C25 に a の推定値を求める。=**C21-C26*B21** とすれば求まる。

2. 予測値と残差の計算

【課題 16】 所得と消費のデータについて各年の X のデータに対する予測値 \hat{Y} と残差を求めよ。

📖 手順

- ① I 列に予測値を求める。予測値 \hat{Y} は各 X_i について $a + b X_i$ を計算すればよいので、I2 セルに 1981 年の X (B2 セル)に対応する予測値を求めるなら=**\$C\$25+\$C\$26*B2** とし、これをコピーすればよい。ここでは、コピーの際に絶対参照をするので、'\$'がついている。
- ② J 列に残差を求める。残差は Y から予測値 \hat{Y} を引いたものなので、J2 セルに =**C2-I2** とし、これをコピーすればよい。

3. 回帰直線のグラフへの書き入れ

散布図に回帰直線を書き入れる場合、Excelでは各 X に対応する予測値をグラフに書き入れ、それを直線でつなぐという手順をとる。

【課題 17】 所得と消費のデータについて散布図に回帰直線を書き入れよ。

📖 手順

- ① **グラフをアクティブ**にした状態で、リボンの中の「データの選択」ボタン(出ていない場合には上部の「グラフツール」をクリックする)をクリックし、凡例項目(系列)の「追加」ボタンをクリックする。そして、「系列Xの値」を**B2:B19**、「系列Yの値」を**I2:I19**とする。
- ② ①の操作で散布図上に赤色のマーカーが現れたはずである。これを直線で結ぶ。赤色のマーカーのひとつを右クリックして、「データ系列の書式設定」を選ぶ。そこで「マーカーのオプション」をクリックし、マーカーの種類として「なし」をチェックし、「線の色」をクリックし、線の色として「**線 (単色)**」をチェックする。グラフエリアの外をクリックすると回帰直線が引けたことがわかるはずである。

4. 決定係数の導出

決定係数は回帰における当てはまりの尺度であり、全変動のうち回帰モデルによって説明される変動の割合を示すものである。決定係数は0と1の間の値をとるが、決定係数が0.2や0.3などの小さい値であるということは、あまり関係のないXとYの間に因果関係を想定し、分析を行っているということを意味し、モデルの再検討が必要となる。

決定係数は、

$$R^2 = \frac{\{\sum(x-\bar{x})(y-\bar{y})\}^2}{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}$$

として求められる。

【課題 18】 C27セルに決定係数を求めよ。(手順は自分で考えてみよう)

5. 残差の表示

残差 e_i は従属変数の個々の観測データと回帰直線との間のズレの大きさをあらわすものであった。この残差の状態を調べることで、回帰直線の当てはまり具合など、さまざまな情報を入力することができる。残差を出発点としてモデルの設定やデータ間の関係を検討する分析を、残差分析(residual analysis)という。ここでは、残差を求めてそれをグラフに表示してみよう。

グラフを描くには、残差 e_i を縦軸にとり、横軸には、

① i

② X_i

③ \hat{Y}_i

などを用いる。

【課題 19】 横軸に年をとった残差プロットを描いてみよう。

📖 手順

- ① 年を横軸に、残差を縦軸にとるので、A2:A19を範囲指定し、**Ctrl**キーを押しながらJ2:J19を範囲指定し、リボン内にグラフのグループから散布図のボタン(出ない場合は挿入タブをクリックする。)をクリックする。そして、「散布図(マーカーのみ) (左上)」をクリックする。
- ② そして、最初に描いた散布図と同様に、レイアウトを変更し、グラフタイトル、軸ラベルをそれぞれ記入する。凡例や目盛線も、消去しておこう。

2) 単回帰モデル(その 2)

前章では、相関係数と回帰係数の推定値を偏差を求め、その2乗和と交差積和を計算することによって求めた。

しかしExcelで相関係数や回帰直線などを求めるとき、Excelが備えている関数を用いたり、分析ツールを用いたりすることもできる。ここでは、そのやり方を紹介する。

1. 統計関数による相関係数と回帰直線の導出

(1) 相関係数、決定係数の導出 (関数 CORREL¹ とRSQ)

相関係数を求める関数は**CORREL**を、 $Y=a+bX$ という回帰モデルにおける決定係数を求める関数は**RSQ**を用いる。どちらも引数は2個であり、**CORREL**は引数の順序に関係がないが、**RSQ**は、最初の引数が従属変数(=の左側)の範囲で、2個目の引数が独立変数(=の右側)の範囲をとる。引数の順序に注意しなければならない。

(2) 回帰係数の導出 (関数 SLOPE とINTERCEPT)

$Y=a+bX$ という回帰モデルにおいて、**SLOPE** は回帰直線の傾き **b** を、**INTERCEPT** は切片 **a** を求める関数である。どちらも引数は2個あるが、最初の引数が従属変数(=の左側)の範囲で、2個目の引数が独立変数(=の右側)の範囲をとる。引数の順序に注意しなければならない。

(3) 予測値の導出 (関数 FORECAST²)

予測値 \hat{Y} として、関数**FORECAST** がある。FORECAST(X_i , 回帰のY範囲, 回帰のX範囲) として用いる。なお、この予測値は戻り値は $a + bx_i$ として求められた数値1個である。

【 課題 20 】 統計関数を用いて、相関係数、決定係数、回帰係数の推定値 **a,b** を計算してみよ。

2. 分析ツールの利用

回帰分析も分析ツールを用いておこなうことが可能である。

データタブの中に「データ分析」のボタンをクリックし、下のほうにある、回帰分析を選べばよい。

【 課題 21 】 分析ツールを用いて、消費関数の分析をおこなってみよ。

☒ 本日実習したファイルは、E-mail に添付ファイルとして河田まで提出すること。
(送付先アドレスは kawada@tokuyama-u.ac.jp)

締め切りは 7月 30日(金)17:00 とする。

なお、ファイル名は「消費関数 E40-○○○」のように、**学籍番号をつけること**。

¹PEARSON という関数も CORREL と全く同様の働きをする。

² FORECAST 関数と同様に予測値を求める関数として、**TREND** という関数があるが、ここでは説明は省略する