

<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**所得と消費**のファイルを用いる。

【課題 19】 統計学基礎の講義用 HP から、所得と消費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 「統計学基礎」をクリックし、「第19回 12月9日(木)」の配布資料にある、「所得と消費」をクリックし、自分の使いやすい場所に保存する。

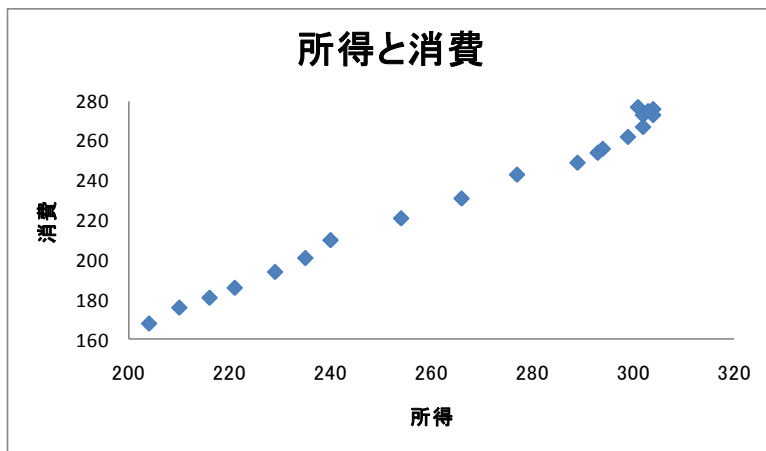
2. 散布図

【課題 20】 所得と消費のファイルには、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(平成 12 年基準、単位: 兆円)¹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:C21 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 : 320
 目盛間隔 : 20
 とする。



¹ 出典: 『平成 21 年版 国民経済計算年報』

3. 相関係数の導出

【課題 21】 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 B23 セル、C23 セルにそれぞれ平均を求める。B23 なら **=AVERAGE(B2:B21)** 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\$23** と入力し、これを D3:D21 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 F24:H24セルに合計を求める。F24 なら **=SUM(F2:F21)** とすればよい。
- ⑥ C26セルに **=H24/SQRT(F24*G24)** 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	204	168	-63.15	-65.65	3987.923	4309.923	4145.798
3	1982	210	176	-57.15	-57.65	3266.123	3323.523	3294.698
4	1983	216	181	-51.15	-52.65	2616.323	2772.023	2693.048
5	1984	221	186	-46.15	-47.65	2129.823	2270.523	2199.048
6	1985	229	194	-38.15	-39.65	1455.423	1572.123	1512.648
7	1986	235	201	-32.15	-32.65	1033.623	1066.023	1049.698
8	1987	240	210	-27.15	-23.65	737.1225	559.3225	642.0975
9	1988	254	221	-13.15	-12.65	172.9225	160.0225	166.3475
10	1989	266	231	-1.15	-2.65	1.3225	7.0225	3.0475
11	1990	277	243	9.85	9.35	97.0225	87.4225	92.0975
12	1991	289	249	21.85	15.35	477.4225	235.6225	335.3975
13	1992	293	254	25.85	20.35	668.2225	414.1225	526.0475
14	1993	294	256	26.85	22.35	720.9225	499.5225	600.0975
15	1994	299	262	31.85	28.35	1014.423	803.7225	902.9475
16	1995	302	267	34.85	33.35	1214.523	1112.223	1162.248
17	1996	302	273	34.85	39.35	1214.523	1548.423	1371.348
18	1997	304	276	36.85	42.35	1357.923	1793.523	1560.598
19	1998	304	273	36.85	39.35	1357.923	1548.423	1450.048
20	1999	303	275	35.85	41.35	1285.223	1709.823	1482.398
21	2000	301	277	33.85	43.35	1145.823	1879.223	1467.398
22								
23	平均	267.15	233.65					
24	合計					25954.55	27672.55	26657.05
25								
26		相関係数	0.994674					
27								

1) 単回帰モデル(その 1)

1. 回帰直線の導出

【課題 22】 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。(C28 セルに b の推定値、C27 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 C28 に b の推定値を求める。=**H24/F24** とすれば求まる。
- ② C27 に a の推定値を求める。=**C23-C28*B23** とすれば求まる。

2. 予測値と残差の計算

【課題 23】 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\$27+\$C\$28*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 24】 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:B21**、「系列Yの値」を**I2:I21**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、

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

として求められる。

【課題 25】 C29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、さまざまな情報入手することができる。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 26】 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:A21を範囲指定し、**Ctrl**キーを押しながらJ2:J21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 CORREL² 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 FORECAST³)

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 27】 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

Excel には統計分析を行うためのいくつかの分析ツールが付属している。これらのツールを使えば一度に詳細な分析結果を得ることができる。

分析ツールを最初に使用する場合には、アドイン(有効にすること)しなくてはならない。分析ツールのアドインは次のようにおこなう。

① 左上の **Office ボタン** を押し、下にある「Excel のオプション」のボタンをクリックする。

② 「分析ツール」にチェックをつけ、**OK** ボタンをクリックする。

すると、データタブの中に「データ分析」のボタンが出てくるので、下のほうにある、回帰分析を選べばよい。

アドインを行った後で、再びメニューバーから「ツール」を選ぶと、下のほうに「分析ツール」と表示される。ここで分析ツールを選び、回帰分析を選べばよい。

²PEARSON という関数も CORREL と全く同様の働きをする。

³ 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

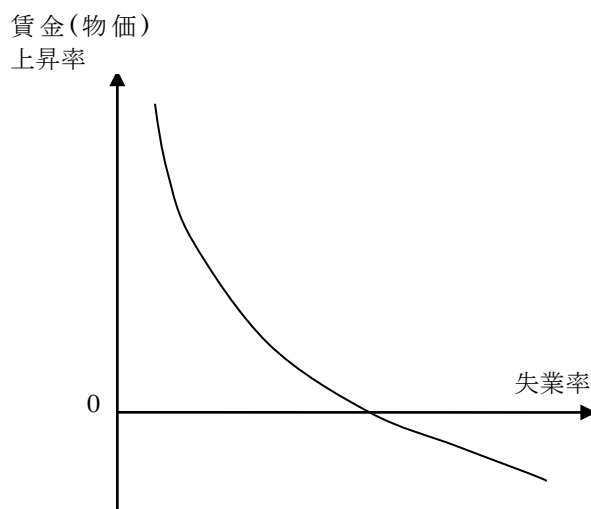
3) 単回帰モデルの他の事例

1. フィリップス曲線

フィリップス(A. W. Phillips) は 1958 年、名目賃金上昇率と失業率の間に右図のような関係があることを、1861 年から 1957 年のイギリスのデータを用いて示した。この曲線は、発見者の名前をとって**フィリップス曲線**と呼ばれる。

その後の研究により、賃金上昇率は物価上昇率との関係が強いことなどから、最近では縦軸に物価上昇率を用いることが多い。

この関係は、インフレ率が高い状況では失業率が低下し、失業率が高い状況ではインフレ率が低下するという、インフレーションと失業とのトレードオフの関係を示している。



2. フィリップス曲線の定式化

フィリップス曲線は、物価上昇率を Y 、失業率を X とするとき、

$$Y = \alpha + \beta \frac{1}{X}$$

という式をあてはめることが考えられる。この式において、 Y と X の関係は直線では表されない(Y と X は**線形関係ではない**という)が、 $X' = \frac{1}{X}$ という変数変換をおこなえば、この式は

$$Y = \alpha + \beta X'$$

という直線(線形式)で表すことができる。

3. フィリップス曲線の実証分析

近年の日本のデータにおいて、フィリップス曲線のような関係が成り立っているかどうかを分析してみよう。

【課題 28】講義用 HP に **Phillips.xls** という名前で 1985 年から 2009 年までの物価上昇率と失業率のデータ⁴が保存されている。このデータを各自のフォルダにダウンロードし、分析をおこなってみよう。

手順は次のようになる。

📖 手順

- ① 失業率を横軸に、物価上昇率を縦軸にとった散布図を描く
- ② D 列に $X' = \frac{1}{X}$ という変換をおこなったデータを作成する。D2 セルに '=1/C2' と入力し、それを D 列全体にコピーすればよい。
- ③ 分析ツールを用いて Y を被説明変数、 X' を説明変数とする回帰分析をおこなう。
- ④ 分析ツールの出力結果にある予測値を、散布図に描き入れ、線でつなぐ。

⁴ 物価上昇率 - 総務省統計局『消費者物価指数』(全国・総合)対前年比
失業率 - 総務省統計局『労働力調査』完全失業率(年平均・男女計)