

# 経済統計学第三回計算機実習要項

2007年12月；担当 星野

- 教科書 95 ページ例題 6.1 のデータを使うことにします。
  - セル A1:A10 に消費 ( $X$ )、B1:B10 に食費 ( $Y_1$ )、C1:C10 に教育費 ( $Y_2$ ) を入力して下さい。
- 単回帰  $Y_1 = a_1 + b_1X$ 、 $Y_2 = a_2 + b_2X$  の係数を最小二乗法で求めます。
  - 消費、食費、教育費の平均を求めます。A12 に “=AVERAGE(A1:A10)” と入力して B12:C12 にコピー & 張り付けして下さい。
  - 次に標準偏差を求めます。A13 に “=STDEVP(A1:A10)” と入力して B13:C13 にコピー & 張り付けして下さい。
  - $X$  と  $Y$  の相関係数を求めます。A15 に “=CORREL(A1:A10,B1:B10)”、C15 に “=CORREL(C1:C10,A1:A10)” と入力します。
  - $b_1$  を求めます。A17 に “=A15\*B13/A13” と入力します。
  - $b_2$  は B17 に “=C15\*C13/A13” と入力します。
  - $a_1$  は A18 に “=-A17\*A12+B12”、 $a_2$  は B18 に “-B17\*A12+C12” と入力します。
- 教科書を見て求めた係数を確認して下さい。(下一桁は四捨五入してあります) 限界性向、平均性向、弾力性の計算は自習にします。以上より  $b = r_{XY}s_Y/s_X$ 、 $a = -b\bar{X} + \bar{Y}$  が確認出来ました。 $n - 1$  ではなく  $n$  で割った標準偏差である事に注意しましょう。
- 次に散布図を書いて直線とデータの関係を見てみましょう。
  - セル D1 に “=A\$17\*\$A1+A\$18” と入力し、D1:E10 までコピー & 張り付けします。これは予測値(「内挿値」という)  $\hat{Y} = a + bX$  を求めています。
  - グラフウィザードを用いて、散布図を二枚書きます。A1:A10 を X 軸の値として用い、1 枚目は B1:B10 と D1:D10、2 枚目は C1:C10 と E1:E10 を比較します。同一散布図で二以上のデータを含めるには、系列を追加します。例えば系列 1 の Y 軸値を B1:B10、系列 2 では D1:D10 で与えましょう。
- 回帰の良さの概念について；全変動 = 残差変動 + 回帰変動という恒等式を確認します。すなわち

$$\sum (y_i - \bar{y})^2 = \sum (y_i - \hat{y}_i)^2 + \sum (\hat{y}_i - \bar{y})^2,$$

但し  $\hat{y}_i$  が  $y_i$  の内挿値、 $\bar{y}$  が  $y_i$  の平均です。以下では、 $Y_1 = a_1 + b_1X$  について確認します。

- 全変動は  $\sum (y_i - \bar{y})^2$ 、例えばセル H1 に “=VARP(B1:B10)\*10” と入力すれば良い。(何故これで良いか考える事)
- 残差変動は  $\sum (y_i - \hat{y}_i)^2$  です。D1:D10 に内挿値が与えられているので、セル F1 に “=(B1-D1)^2” と入力。これを F1:F10 までコピー & 張り付け。F1:F10 までの総和を求めれば良いので、セル H2 に “=SUM(F1:F10)” と入力。
- 回帰変動は  $\sum (\hat{y}_i - \bar{y})^2$ 。G1 に “=(D1-B\$12)^2” と入力。これを G1:G10 までコピー & 張り付け。総和はセル H3 に “=SUM(G1:G10)” と入力。

- セル H4 に “=H2+H3” と入力。これがセル H1 と一致する事を確認。
- 回帰変動 / 全変動を「決定係数」と言います。セル H5 に “=H3/H1” と入力しましょう。
- 決定係数の意味について考えます。内挿値と実現値の相関係数が 1 に近い程、良い予測式だという事は言えるはずですが。セル H6 に “=CORREL(B1:B10,D1:D10)” と入力します。これを「重相関係数」と言います。実は重相関係数の二乗が決定係数です。セル H7 に “=H6^2” と入力、H5 と比較して下さい。

6. ここでは回帰の仕組みを理解してもらう為に面倒な方法で係数を求めましたが、EXCEL にはもっと楽な方法が用意されています。

- メニューバーの「ツール (T)」をクリックしてメニューを出して下さい。
- その中に「データ分析」という項目が無ければ以下の作業をします。有れば飛ばします。
  - メニューの中から「アドインの登録と管理」を選びます。
  - 追加のアドインとして「分析ツール」を選択（四角い箱にチェックを入れる）してください。
  - もう一度「ツール (T)」をクリックしてメニューに「データ分析」が出ることを確認します。
- 「ツール (T)」「データ分析」と選択し、出てきたメニューより「回帰分析」を選択します。
- Y 範囲として “B1:B10”、X 範囲として “A1:A10” を考えます。
- 出力された「切片」と「X 値 1」の係数を A18( $a_1$ )、A17( $b_1$ ) と比較してください。
- 決定係数は出力の「重決定 R2」の値です。セル H5 の値と一致する事を確認。
- 重相関係数は出力の「重相関 R」の値です。セル H6 の値と比較。

7. 同様に教育費と消費の関係についても、回帰分析ツールの出力を確認して下さい。

8. 次に重回帰をしてみます。消費を、教育費と食費で説明する事を考えます。モデルとしては

$$X = a + bY_1 + cY_2 \quad (1)$$

という事です。つまり教育費と食費の支出額を知れば、消費支出の見当が付くと解釈します。

- 回帰ツールの Y 範囲で “A1:A10”、X 範囲で “B1:C10” とします。モデルの変数名と EXCEL の入力で X, Y が食い違う事に混乱無きよう。常に被説明変数の範囲を「Y 範囲」で入力する。説明変数の範囲が「X 範囲」で、重回帰の場合複数コラムを連続的に与える。(おそらく B1:B10, C1:C10 のように、カンマで範囲をつなぐ事は出来ない。)
- 「切片」、「X 値 1」、「X 値 2」の係数がそれぞれ、 $a, b, c$  に対応します。
- 次に単回帰モデル (消費を食費のみで説明):

$$X = a + bY_1 \quad (2)$$

の係数をモデル (1) と比較してみましょう。決定係数の変化は？

- 自由度修正済  $R^2$  は「補正 R2」の項で出力されます。モデル (1),(2) のいずれが良いですか？

9. 変数変換について：EXCEL では  $e$  を底とする対数関数を “LN()” で表します。指数関数は “EXP()” です。GDP( $Y_t$ )、年度 ( $t$ )、について  $Y_t = \exp(a + bt)$  という関係が有るとします。この時  $\log Y_t = a + bt$  です。

- <http://stat.w3.kanazawa-u.ac.jp/econs/list.html> より gde.xls をダウンロードする。
- ワークシート gde.xls で、セル C4 に “=LN(B4)” と入力。これを C5:C25 にコピー & 貼り付け。
- この新たな系列を被説明変数、年度を説明変数とする単回帰を行う。
- 各年度について、GDP の理論値を求めよ (EXP を用いる)。
- 理論値と実現値を、散布図で比較せよ。