

エクセルでの統計処理

五石敬路

(1) 散らばりについて (分散、標準偏差、変動係数)

| | |
|------|------------|
| 分散 | =var.p() |
| 標準偏差 | =stdev.p() |
| 変動係数 | =標準偏差/平均 |

(2) 基本統計量 (最大値、最小値、算術平均、標準偏差)

| | |
|------|------------|
| 最大値 | =max() |
| 最小値 | =min() |
| 算術平均 | =average() |
| 標準偏差 | =stdev.p() |

(3) 相関係数

| | |
|------|-----------|
| 相関係数 | =correl() |
|------|-----------|

参考：ジニ係数について

平等度を表す指数。0 から 1 の値をとる。0 に近いほどより平等で、1 に近いほどより不平等。所得では、北欧諸国が 0.2~0.3、アメリカや日本が 0.3~0.5、中南米やサハラ以南アフリカは 0.6~0.7。0.7 は社会が不安定化/崩壊する (ベネズエラ、シエラレオ)。ただ、ジニ係数の計算は難しく、統計ソフトでも用意されていないことが多いので、変動係数で代替できる。

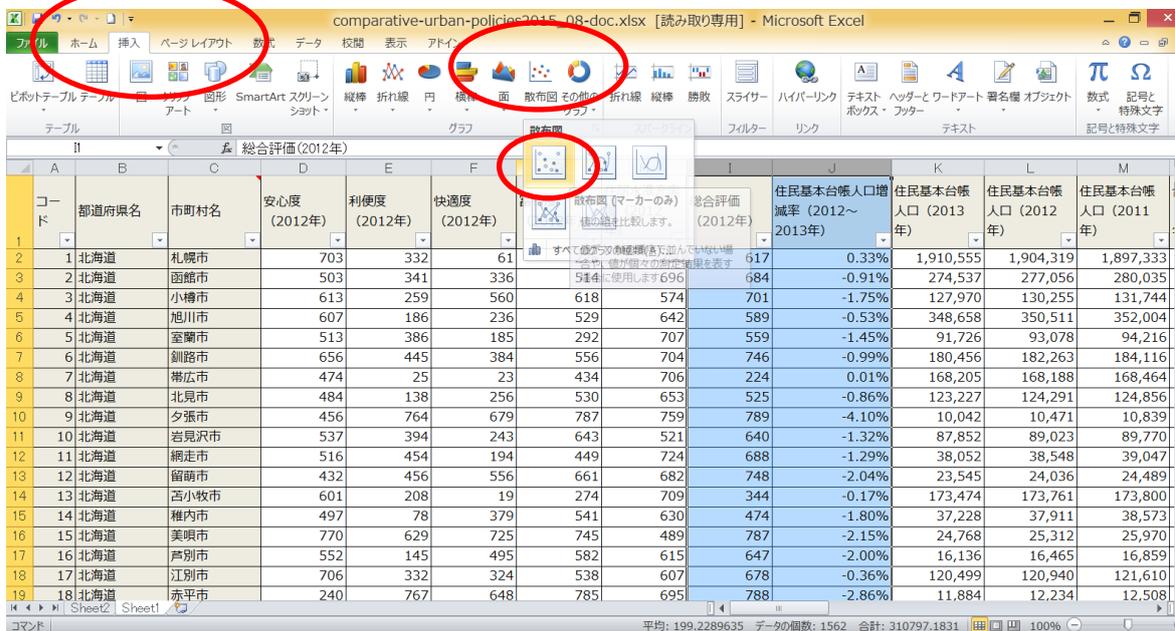
- (4) エクセルでのプロットの描き方、図のワードへの貼り付け (p.2~p.4)
- (5) 基本統計量 (p.5~6)
- (6) 分析ツールの設定 (p.7~8)
- (7) 回帰分析 (p.9~p.12)
- (8) 各自治体における人口増減の要因を回帰分析により調べてみる

(4) エクセルでのプロットの描き方

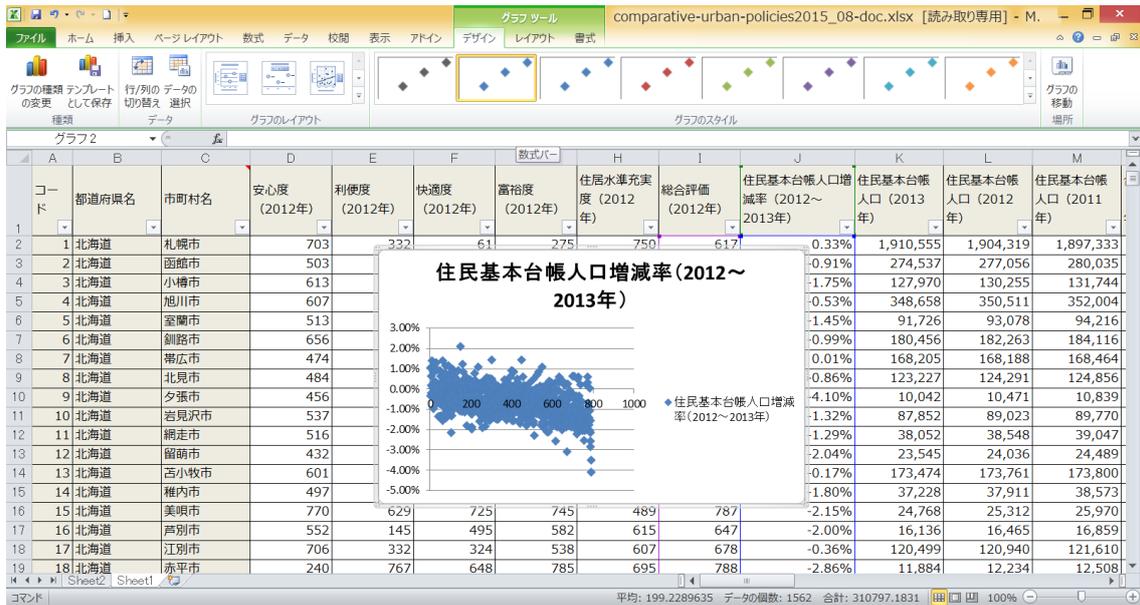
| コード | 都道府県名 | 市町村名 | 安心度 (2012年) | 利便度 (2012年) | 快適度 (2012年) | 富裕度 (2012年) | 住居水準充実度 (2012年) | 総合評価 (2012年) | 住民基本台帳人口増減率 (2012~2013年) | 住民基本台帳人口 (2013年) | 住民基本台帳人口 (2012年) | 住民基本台帳人口 (2011年) |
|-----|-------|------|-------------|-------------|-------------|-------------|-----------------|--------------|--------------------------|------------------|------------------|------------------|
| 1 | 北海道 | 札幌市 | 703 | 332 | 61 | 275 | 750 | 617 | 0.33% | 1,910,555 | 1,904,319 | 1,897,333 |
| 2 | 北海道 | 函館市 | 503 | 341 | 336 | 514 | 696 | 684 | -0.91% | 274,537 | 277,056 | 280,035 |
| 3 | 北海道 | 小樽市 | 613 | 259 | 560 | 618 | 574 | 701 | -1.75% | 127,970 | 130,255 | 131,744 |
| 4 | 北海道 | 旭川市 | 607 | 186 | 236 | 529 | 642 | 589 | -0.53% | 348,658 | 350,511 | 352,004 |
| 5 | 北海道 | 室蘭市 | 513 | 386 | 185 | 292 | 707 | 559 | -1.45% | 91,726 | 93,078 | 94,216 |
| 6 | 北海道 | 釧路市 | 656 | 445 | 384 | 556 | 704 | 746 | -0.99% | 180,456 | 182,263 | 184,116 |
| 7 | 北海道 | 帯広市 | 474 | 25 | 23 | 434 | 706 | 224 | 0.01% | 168,205 | 168,188 | 168,464 |
| 8 | 北海道 | 北見市 | 484 | 138 | 256 | 530 | 653 | 525 | -0.86% | 123,227 | 124,291 | 124,856 |
| 9 | 北海道 | 夕張市 | 456 | 764 | 679 | 787 | 759 | 789 | -4.10% | 10,042 | 10,471 | 10,839 |
| 10 | 北海道 | 岩見沢市 | 537 | 394 | 243 | 643 | 521 | 640 | -1.32% | 87,852 | 89,023 | 89,770 |
| 11 | 北海道 | 網走市 | 516 | 454 | 194 | 449 | 724 | 688 | -1.29% | 38,052 | 38,548 | 39,047 |
| 12 | 北海道 | 留萌市 | 432 | 456 | 556 | 661 | 682 | 748 | -2.04% | 23,545 | 24,036 | 24,489 |
| 13 | 北海道 | 苫小牧市 | 601 | 208 | 19 | 274 | 709 | 344 | -0.17% | 173,474 | 173,761 | 173,800 |
| 14 | 北海道 | 稚内市 | 497 | 78 | 379 | 541 | 630 | 474 | -1.80% | 37,228 | 37,911 | 38,573 |
| 15 | 北海道 | 美瑛市 | 770 | 629 | 725 | 745 | 489 | 787 | -2.15% | 24,768 | 25,312 | 25,970 |
| 16 | 北海道 | 芦別市 | 552 | 145 | 495 | 582 | 615 | 647 | -2.00% | 16,136 | 16,465 | 16,859 |
| 17 | 北海道 | 江別市 | 706 | 332 | 324 | 538 | 607 | 678 | -0.36% | 120,499 | 120,940 | 121,610 |
| 18 | 北海道 | 赤平市 | 240 | 767 | 648 | 785 | 695 | 788 | -2.86% | 11,884 | 12,234 | 12,508 |

① プロットしたい変数を2つ選択する。ここでは総合評価と住民基本台帳人口増減率。

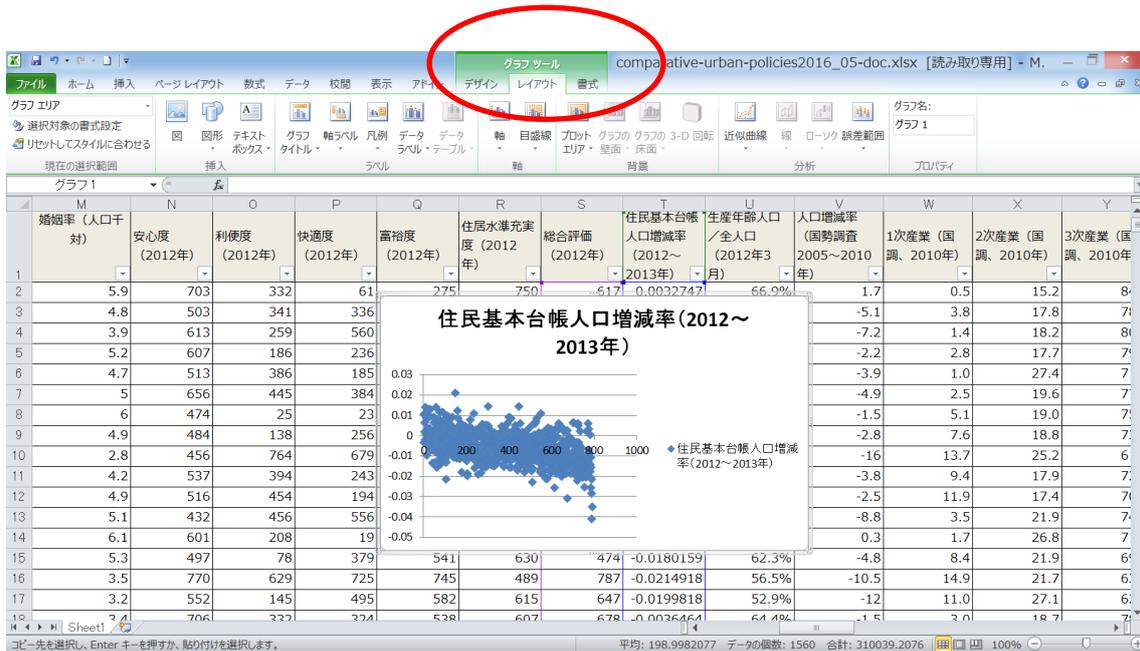
(注) 左側の変数がX軸、右軸の変数がY軸に表示されます。



② 挿入→散布図をクリック。散布図のうち、左上を選択。

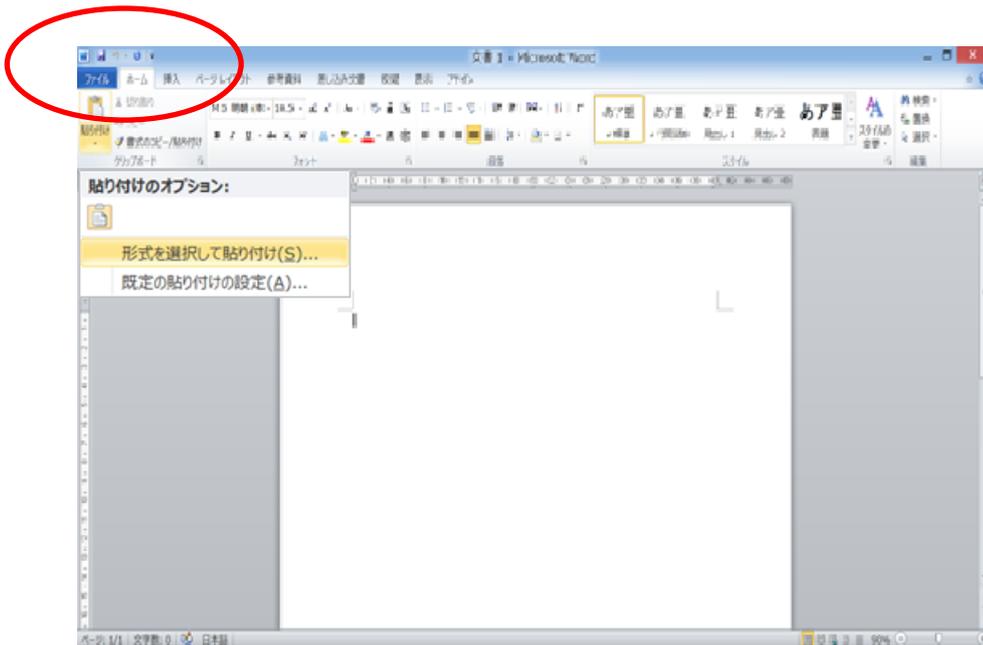


③ 散布図 (プロット) が表示されます。

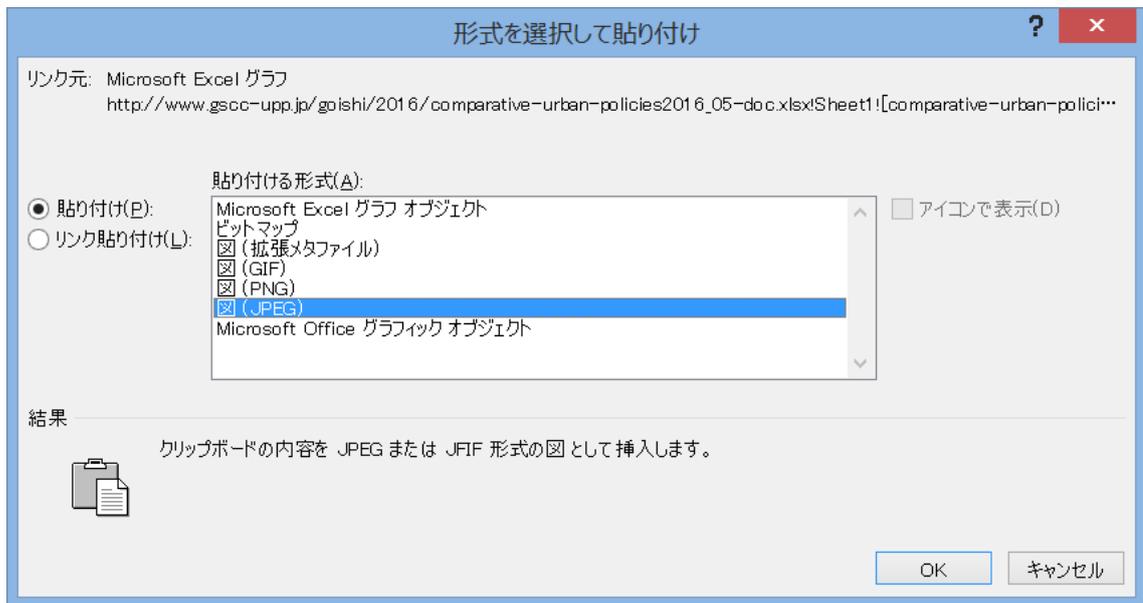


④ 図のレイアウトを整える場合、レイアウトをクリックする。グラフタイトル、軸ラベル、凡例等のメニューがある。

⑤ ワード等にコピーする場合、まず、図を選択し Ctrl-C。

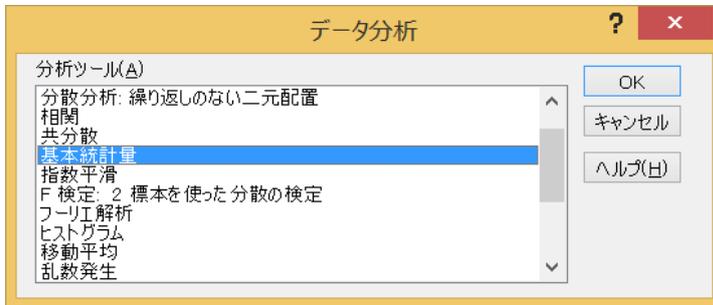


⑥ ワードを開き、ホームー貼り付けー「形式を選択して貼り付け」をクリック



⑦ ワード上で扱いやすいのは「図 (JPEG)」。「Microsoft Excel グラフ オブジェクト」は、エクセルのデータも合わせてコピーされるので、データの加工はしやすくなるが、全体的に扱いにくい。

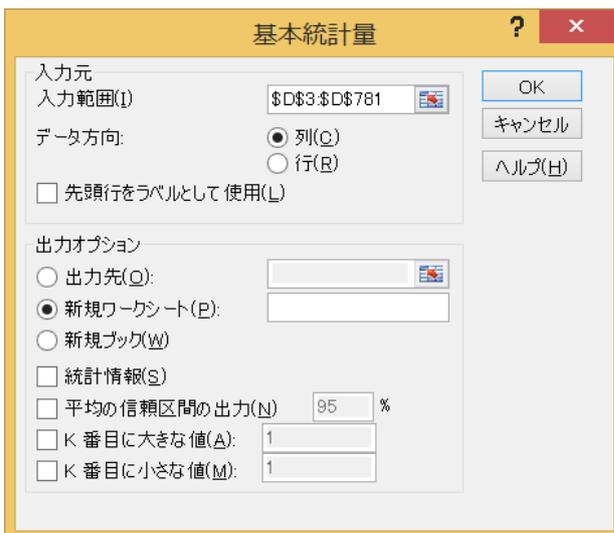
(5) 基本統計量



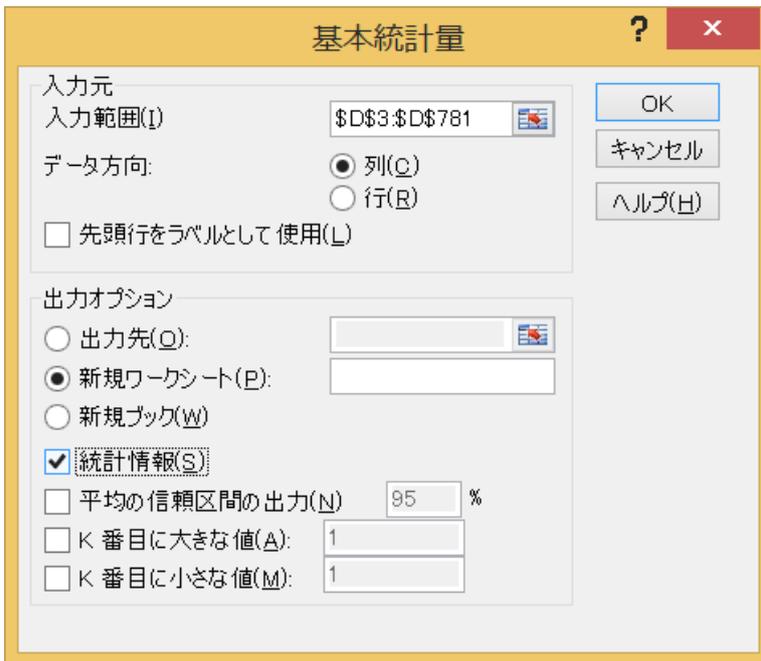
- ① 基本統計量を選択し、OKをクリック



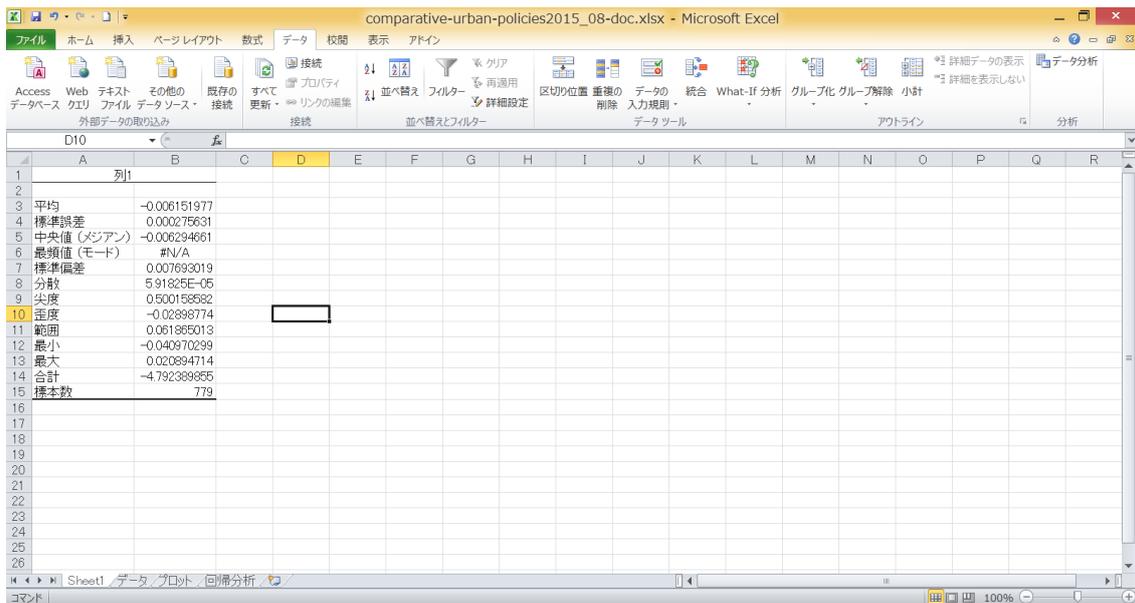
- ② 入力範囲の横にある空欄をクリック（縦棒がチカチカする）



入力範囲の横にある空欄に「住民基本台帳人口増減率」（D欄）のデータを選択する（空欄には、\$D\$3:\$D\$781 と表示される）。



③ 新規ワークシート、統計情報にチェックする。

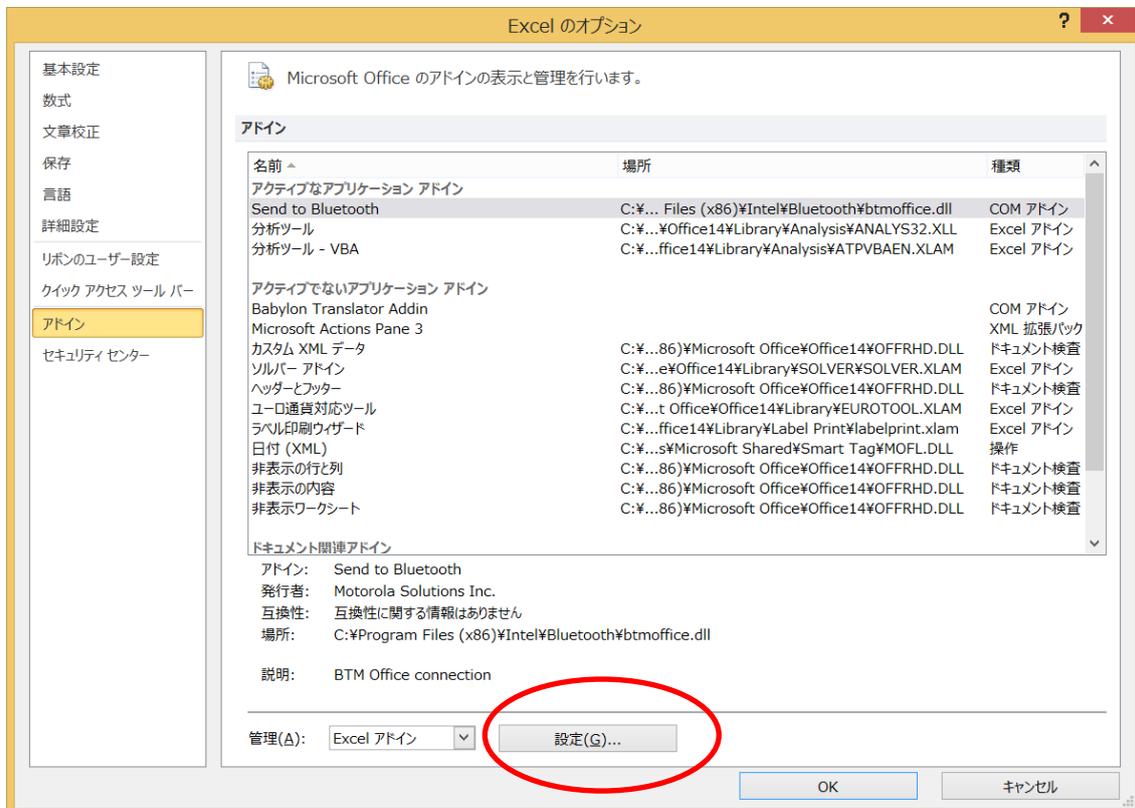


④ 新しいワークシートに基本統計量が表示される。

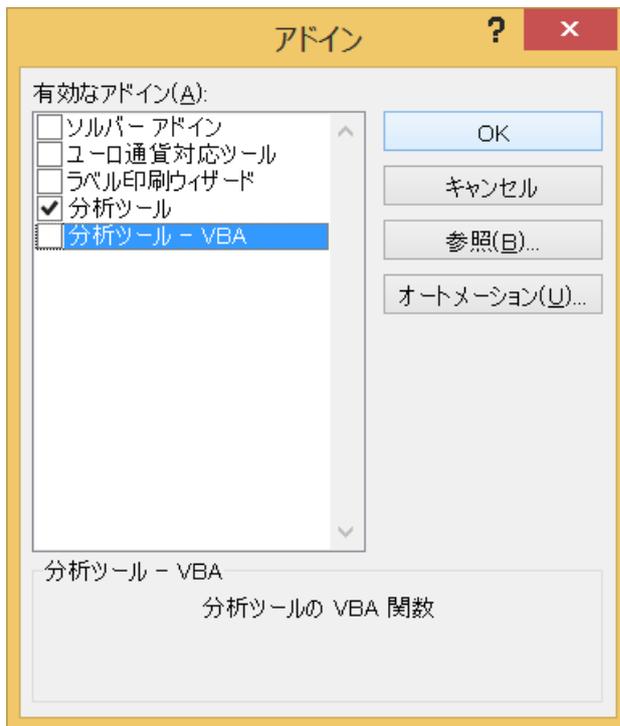
(6) 分析ツールの設定



① ファイルーオプションをクリック



② アドインー設定 (下の方にある設定 (G) のボタン) をクリック

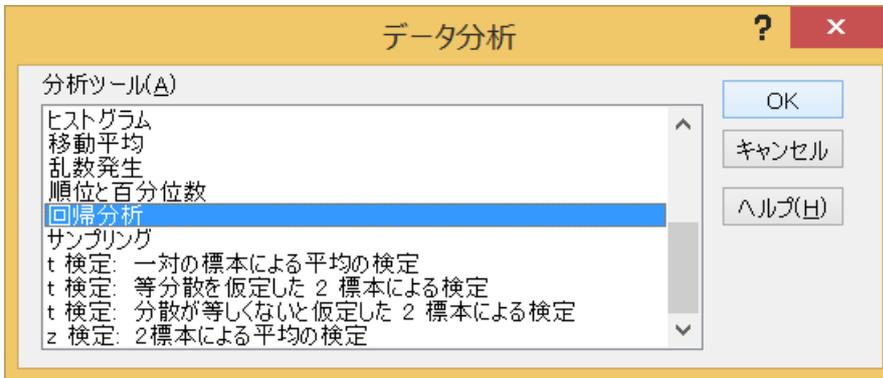


③ 分析ツールをチェックし、OKをクリック

(7) 回帰分析



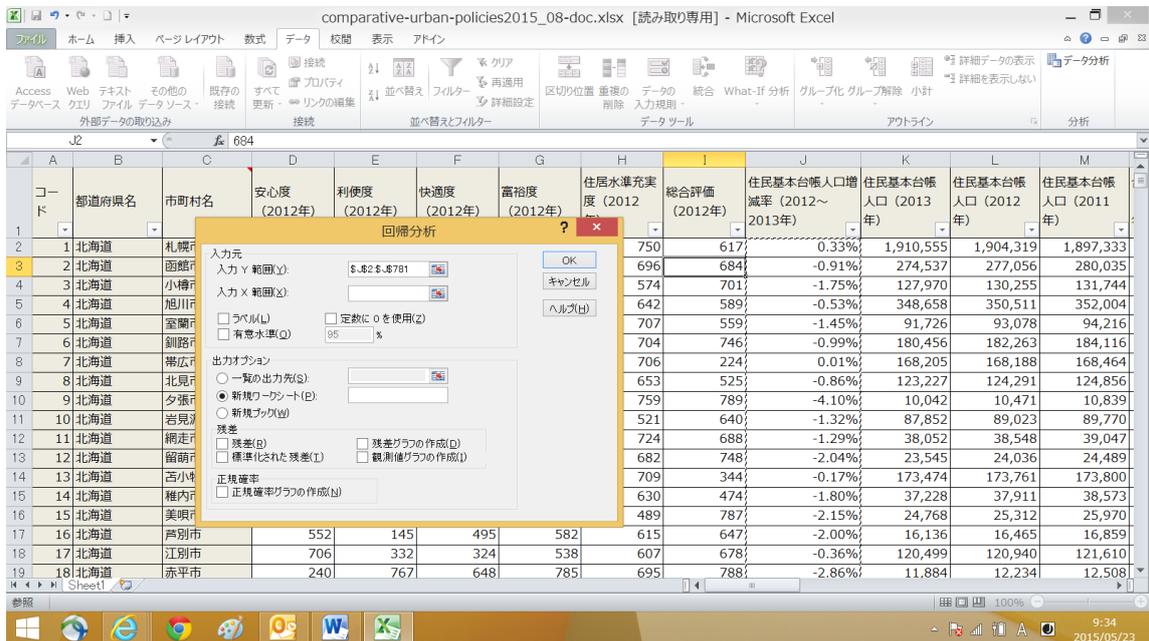
① データーデータ分析をクリック



② 回帰分析を選択し、OKをクリック



③ 入力Y範囲の横にある空欄をクリック（縦棒がチカチカする）



④ 入力Y範囲の横にある空欄に「住民基本台帳人口増減率」（J欄）のデータを選択する（空欄には、\$J\$2:\$J\$781 と表示される）。これが被説明変数（従属変数）になる。

回帰分析 ? X

入力元

入力 Y 範囲(Y): \$J\$2:\$J\$781

入力 X 範囲(X): \$I\$2:\$I\$781

ラベル(L) 定数に 0 を使用(Z)

有意水準(Q) 95 %

出力オプション

一覧の出力先(S):

新規ワークシート(P):

新規ブック(W)

残差

残差(R) 残差グラフの作成(D)

標準化された残差(I) 観測値グラフの作成(I)

正規確率

正規確率グラフの作成(N)

OK

キャンセル

ヘルプ(H)

- ⑤ 入力 X 範囲の横にある空欄に、説明変数（独立変数）のデータを選択する。ここでは、総合評価（2012年）を選択した。空欄には、\$I\$2:\$I\$781 と表示される

回帰分析 ? X

入力元

入力 Y 範囲(Y): \$J\$2:\$J\$781

入力 X 範囲(X): \$I\$2:\$I\$781

ラベル(L) 定数に 0 を使用(Z)

有意水準(Q) 95 %

出力オプション

一覧の出力先(S):

新規ワークシート(P):

新規ブック(W)

残差

残差(R) 残差グラフの作成(D)

標準化された残差(I) 観測値グラフの作成(I)

正規確率

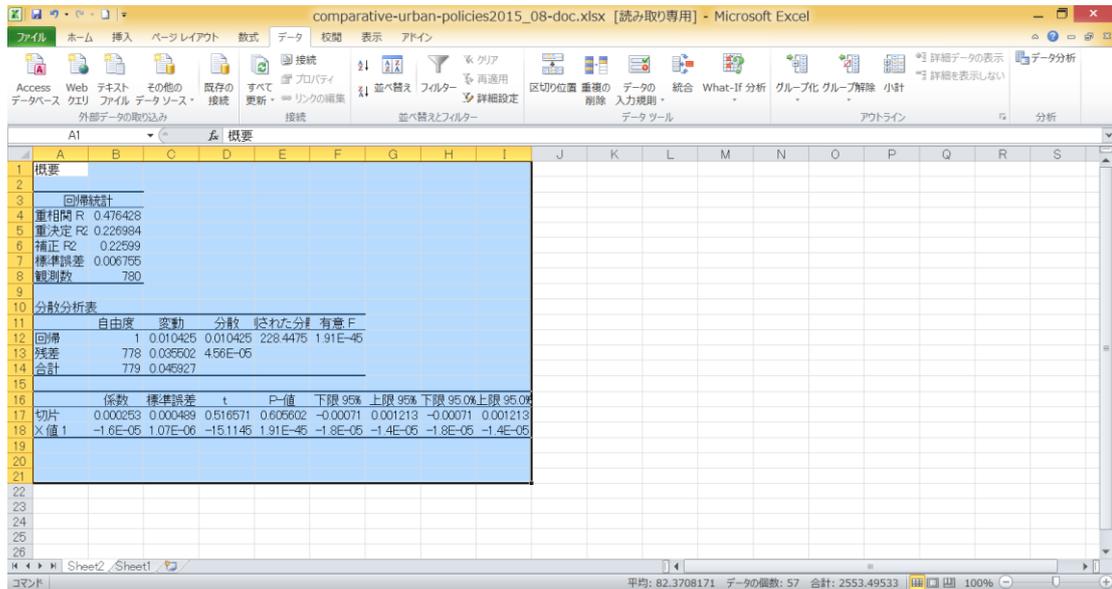
正規確率グラフの作成(N)

OK

キャンセル

ヘルプ(H)

- ⑥ 有意水準をチェック、新規ワークシートをチェックし、OKをクリックする。



- ⑦ 新しいワークシートに、回帰分析の結果が表示されます。
 この場合、「総合評価」が「住民基本台帳人口増減率」と関係しているかどうかは、おおよそですが、「X値1」の「t」が2以上（もしくは-2以下）、「P-値」が、0.05以下であれば、関係がある（統計的に有意）とすることができます。
 この場合、以下の式になります。

$$\text{「住民基本台帳人口増減率」} = 0.000253 + (-0.0000161) \times \text{「総合評価」}$$

ほかに、重要な項目として「決定係数」があります。これは「回帰統計」の欄の「重決定 R2」に示されています。決定係数は被説明変数の散らばりを説明変数によりどの程度説明できているかを表し、0から1の間の値をとり、数値が高いほど当てはまりが良いです。どの程度高ければ良いかの明確な基準はないですが、0.3以下だとあまり当てはまっていないと思われれます。0.7以上であれば高いと判断されるでしょう、

※ 下限 95%、上限 95%は、計算された係数が 95%の確率でとる値の範囲を表す。「統計的に有意」とは、この範囲のなかに0が入っていないということ。

$$-1.96 < \frac{\text{回帰分析の係数} - \text{母集団の係数}}{\text{標準誤差}} < 1.96$$

移項すると、母集団の係数（真の値）の上限 95%、下限 95%が計算できる。

$$1.96 \times \text{標準誤差} + \text{回帰分析の係数 (上限 95\%)} > \text{母集団の係数} > -1.96 \times \text{標準誤差} + \text{回帰分析の係数 (下限 95\%)}$$