2014年6月

# データのモーメント

今月は計量経済学で利用する「モーメント」という言葉の定義を確認することから始めます。

## モーメントとは

JAMES D.HAMILATONの著書 *Time Series Analysis*,Princeton"によれば、標本のモーメントは次のように定義されています。まず、データ(標本)が $x_1, x_2, \ldots, x_T$ とある場合、1次のモーメントとして標本平均は、

$$\bar{x} \equiv (1/T) \cdot (x_1 + x_2 + \cdots + x_T)$$

次に二次モーメントである標本分散は、

$$s^2 \equiv (1/T) \cdot \sum_{t=1}^{T} (x_t - \bar{x})^2$$

一般的にr次の標本モーメントは次のようになります。

$$(1/T) \cdot (x_1^r + x_2^r + \cdots + x_T^r)$$

ここで、2014年2月号で紹介した歪度の計算式を確認します。

$$S = \frac{1}{N} \sum_{i=1}^{N} \left( \frac{y_i - \bar{y}}{\hat{\sigma}} \right)^3$$

この計算式はr=3とした3次のモーメントで､yを標準化した値をxとしていることが分かります。

次に､歪度の計算式で分母にはいっているシグマハットの計算式とEViewsのコマンドについて確認します。

確率変数Xの標準偏差を求める場合、次に示す標本標準偏差を求めるのであれば@stdev(x)、

$$\sqrt{\frac{(x_i - \bar{x})^2}{N-1}}$$

これに対し母集団の母標準偏差を求める場合は@stdevp(x)となります。

$$\sqrt{\frac{(x_i - \bar{x})^2}{N}}$$

また、@stdev(x)と同じ計算を行うコマンドとして@stdevs(x)というコマンドもあります。まとめますと、標本(sample)標準偏差を求めるコマンドは@stdev(x)と@stdevs(x)、母集団(population)の標準偏差を求めるコマンドは@stdevp(x)となります。よって、歪度を求める場合のシグマハットは@stdevp(x)で推定することになります。

以上の基礎知識を元に、歪度の計算を行うプログラムを書いてみましょう。具体的にはN個の

データを正規分布の乱数によって作成し、その歪度を計算します。一方、EViews のヒストグラムを作成する機能を実行して、プログラムの結果と一致することを確認します。
メインメニューで File/New/Program として、次に示すコマンドを入力します。

```
'歪度の計算

!obs=100
'データの個数を設定

wfcreate(wf=moment) u !obs
'ワークファイルの新規作成。Unstructuredワークファイル。データの
'個数は!obs個

rndseed 777
'乱数発生のシード(種)の設定。任意の整数。

series y=10*nrnd
'N(0,1)の正規分布から乱数作成し、10倍した値をyとする。

!sd=@stdev(y)
'yの標本標準偏差。(N-1)で割る

!sdp=@stdevp(y)
'yの母標準偏差。Nで割る

show !sd !sdp
'!sd>!sdpの関係を確認する

'歪度(3次のモーメント)の計算---------------
!mu=@mean(y)
'平均

series z=((y-!mu)/!sdp)^3
'3乗項の計算

!sum=@sum(z)
show (1/!obs)*!sum
```

```
'計算した歪度の表示

show y.hist
'EViews のヒストグラムによる歪度と一致することを確認---終了
```

Run ボタンをクリックしてプログラムを実行すると、次に示すヒストグラムを表示します。

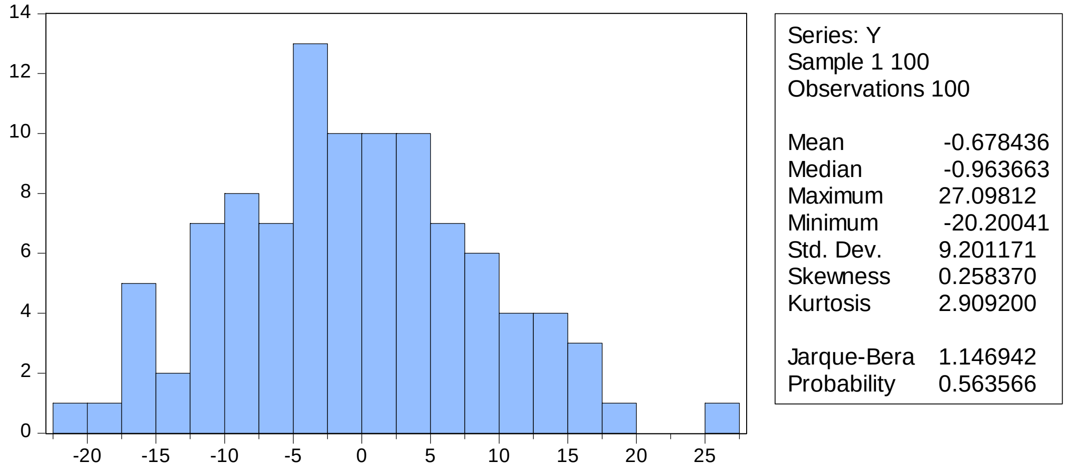



図 1.乱数 Y のヒストグラム

プログラムによって計算した歪度は 0.258370 に一致することを確認してください。また、出力結果として@stdevp と@stdev の違いも示しますので、母標準偏差の方が、標本標準偏差よりも小さくなっていることを確認します。

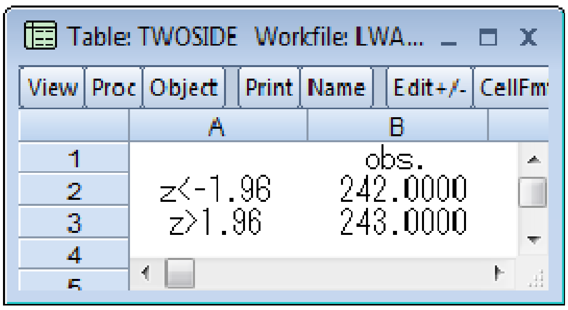
ここまで分布ということについて、いくつかの角度から見てきました。我々は、分布がわかることによって「珍しい」値の存在を知ることができます。ここで言う「珍しい」とは、例えば、100 人の人に聞いたら、体験者が 5 人よりも少ないような事象(値)を意味します。

2014 年 4 月号で利用した対数正規分布による仮想賃金データのときのコマンドを再び利用して、賃金データが対数正規分布に従うと仮定した場合、外れ値(=珍しい値)を取る人の人数を調べてみましょう。File/New/Programs として、プログラム画面に次のコマンドを入力します。

```
!obs=100
'データの個数を決めます。100から初めて徐々に増やして、結果を比較してください

wfcreate(wf=lwage) u !obs
rndseed 123
```

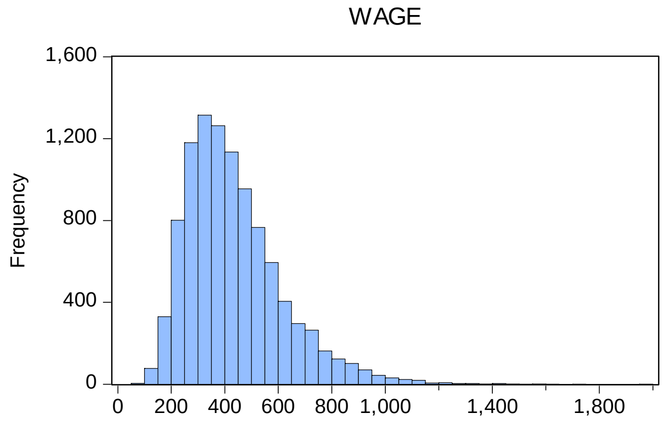
```
series wage=@rlognorm(6,0.4)
’平均6、標準偏差0.4の対数正規分布に従う乱数によるwageの作成

series lwage=log(wage)
'wageの対数を取ったlwageの作成。これが正規分布する

series z=(lwage-@mean(lwage))/@stdevs(lwage)
'lwageを正規化(標準化)してzを作成する

group group01 wage lwage z
show group01.distplot(m)

table twoside
'tableオブジェクトtwosideの作成
twoside(1,2)="obs."
twoside(2,1)="z<-1.96"
twoside(3,1)="z>1.96"

smpl @all if z<@qnorm(0.025)
twoside(2,2)=@obssmpl
’下側2.5%の範囲に入るzの個数を表に格納

smpl @all if z>@qnorm(0.975)
twoside(3,2)=@obssmpl

show twoside
smpl @all
```

このプログラムで!obs の値を 100, 1,000, 10,000…と増やしていきます。上下 2.5%の領域に含まれるデータの個数は、徐々に全体の 5%に近づいていきます。参考までに、!obs=10000 とした時の結果を示します。正規分布の両側を対して 5%の範囲には 485 個のデータが含まれています。1 万個の 5%は 500 個ですから、485 はかなり近い値であることが



分かります。

ヒストグラムを確認しましょう。
Wage は対数正規分布になっており、その対数を取った log(wage)は左右対称な正規分布になっています。
さらに、log(wage)を標準化した z は N(0,1)にしたがっている様子は分かります。




ここまでのお話を少し整理しましょう。確率変数がある分布関数に従っているとすると、そのパラメータ(平均や分散など)が分かれば、「珍しい値」とは具体的にどのようなデータなのか、知ることができる、というのが統計学という「道具」の一つの使い方です。
ここで大事なのは、「普通」のデータがどんな値か、ということを示すことが目的ではなく、「珍しい値」であることを主張することです。
では、分布関数として、どのようなものがあるのか、次号から見てゆくことにします。次号では t 分布について考えます。

■