

正規性の評価

1. 目的

統計解析の種々のアルゴリズムを実行する条件の一つは、データが正規分布に従っていることです。

母集団が正規分布に従うことの確認は難しいですが、与えられたデータが明らかに正規分布から外れているなら、解析アルゴリズムを適用しても意味がありません。つまり、所与のデータについて前もって正規性の評価を行ない、正規分布から外れるものを計算対象から外すことには 意味があります。

検定のアルゴリズムは 2 種類あります。

(1) シャピロー-ウィルク検定

データ数が 4 ~ 20 の場合に利用します。

- ・ 帰無仮説：データは正規分布に従う
- ・ 対立仮説：データは正規分布に従わない

として、所与のデータから有意確率 p 値を計算し、

- ・ p 値 \geq 有意水準なら 帰無仮説 を採用
- ・ p 値 $<$ 有意水準なら 対立仮説 を採用

とするものです。

(2) コルモゴロフ-スミルノフ検定

データ数が 20 超の場合に利用します。

- ・ 帰無仮説：データは正規分布に従う
- ・ 対立仮説：データは正規分布に従わない

として、所与のデータから有意確率 p 値を計算し、

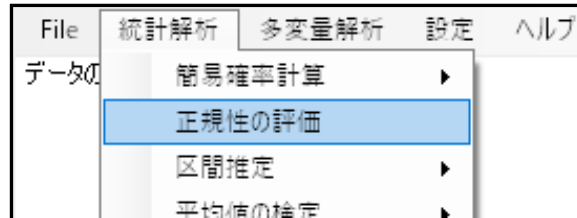
- ・ p 値 \geq 有意水準なら 帰無仮説 を採用
- ・ p 値 $<$ 有意水準なら 対立仮説 を採用

とするものです。


2. 使用法

(1) メニューの選択

メニューの「統計解析→正規性の評価」を選択します。

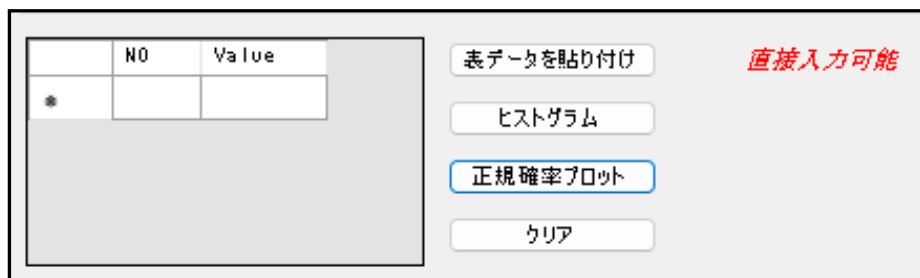


(2) パネルが表示されます。



(3) データの入力

パネルの（下の部分の）グリッドにデータを入力します。



データの入力方法は 2 種あります。

- ・表データをコピーペーストする方法
この場合、データに ID は不要。
- ・直接 Value にデータ記述する方法

正規乱数により発生させたデータを用いて例示します。
このデータは 正規分布に かなりの程度で従っていると考えられます。

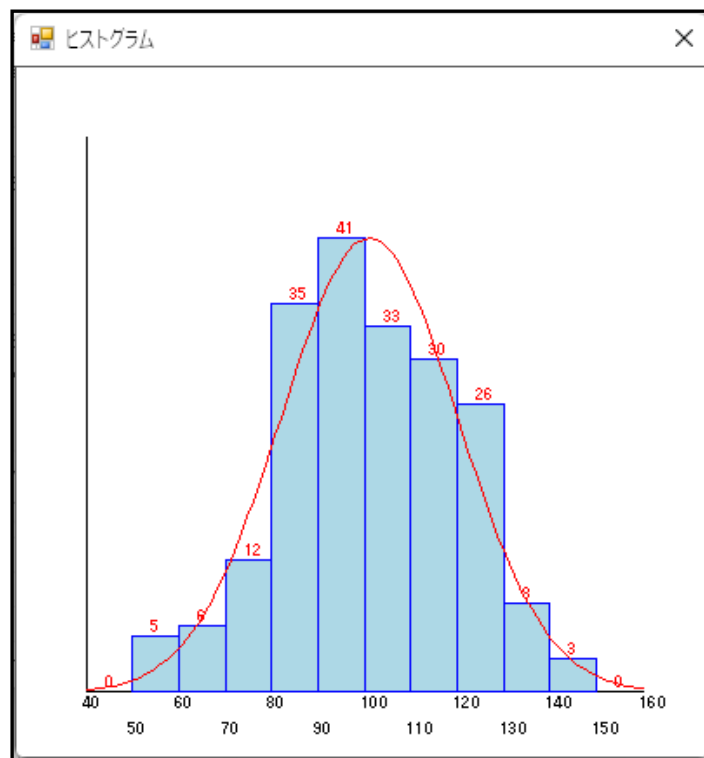
| |
|--------|
| 53.43 |
| 57.55 |
| 58.27 |
| 59.04 |
| |
| 139.31 |
| 140.39 |
| 145.13 |
| 145.44 |

表データを貼り付け

を押して、グリッドにデータを定義します。

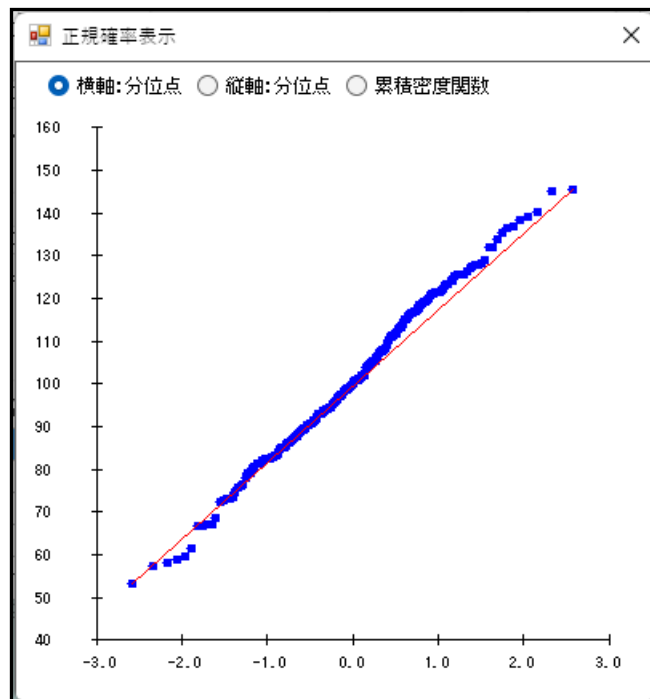
ヒストグラム

で、ヒストグラムが表示されます。



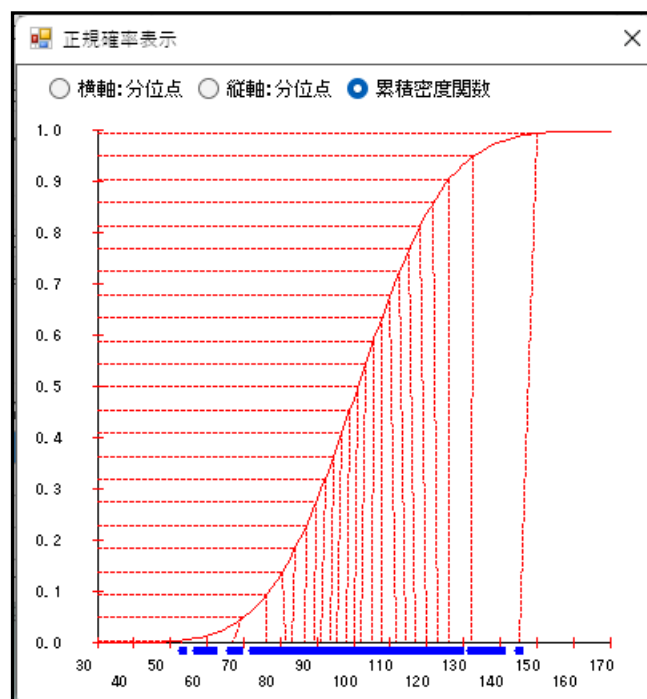
正規確率プロット

で、正規確率プロットが表示されます。



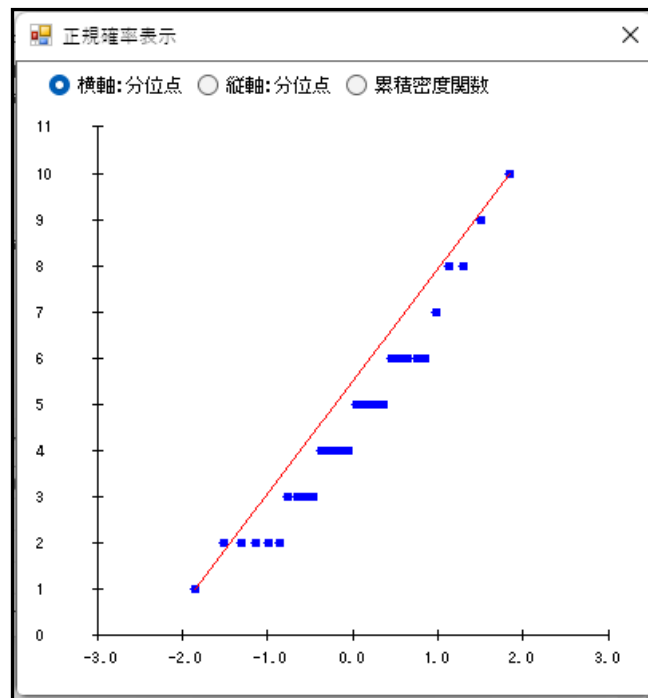
正規確率プロットでは、正規分布に近いデータは 上図のように直線から 大きく離れない位置に並びます。

また ☒ 累積密度関数 を選ぶと、

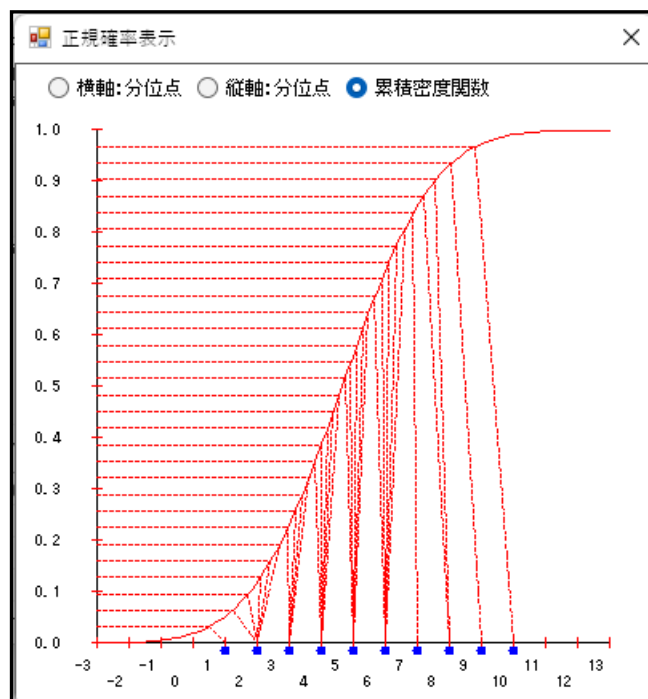


のように、正規分布に近いものほど、きれいにデータが並びます。

実際に用いるデータ、つまり正規分布からやや外れたサンプルを使用すると、下図のようになります。



☒ 累積密度関数 を選ぶと、



のように、ゆがんだ形状となります。

(4) 正規性の評価

データ数が4～20の場合には、「シャピローウィルク検定」を実行します。

例えば、データとして

| |
|-----|
| 71 |
| 86 |
| 92 |
| 95 |
| 100 |
| 102 |
| 105 |
| 108 |
| 118 |
| 123 |

を入力した場合の結果は以下となります。

| | | | |
|--------------|---------------------------------------|-----------|----------|
| データ数が4～20の場合 | | | |
| シャピローウィルク検定 | | | |
| 統計量 W : | 0.9818131 | P 値 (%) : | 97.41273 |
| 結 果 : | 有意でない: 帰無仮説 (サンプルデータは正規分布に従う) を棄却できない | | |

結果は p 値が 97% で、帰無仮説を棄却できない、つまり正規分布とみなせるという結論です。

データ数が20超の場合には、「コルモゴロフスミルノフ検定」を実行します。

データとして、先に利用した正規乱数からのデータを使用します。

| |
|--------|
| 53.43 |
| 57.55 |
| 58.27 |
| 59.04 |
| |
| 139.31 |
| 140.39 |
| 145.13 |
| 145.44 |

「コルモゴロフスミルノフ検定」を実行した結果は以下となります。

データ数が20を超える場合

コルモゴロフスミルノフ検定 経験分布関数

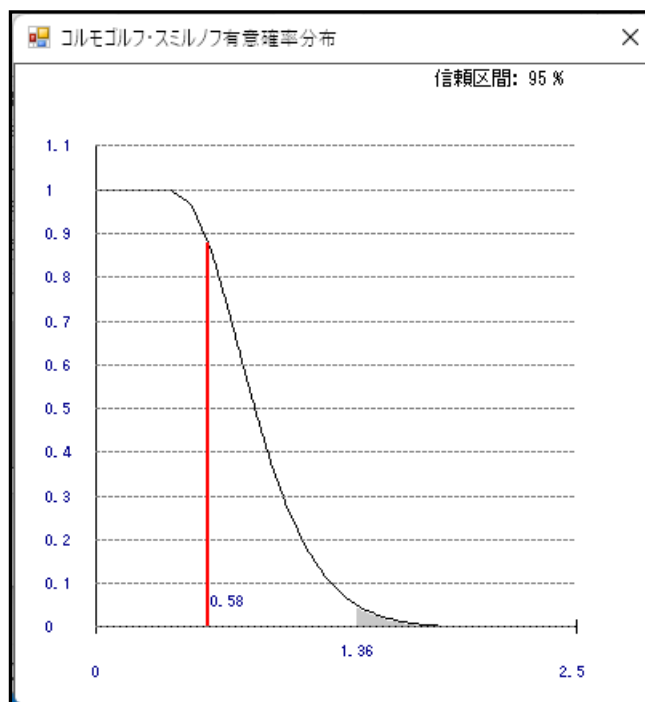
統計量 D_{\max}/\sqrt{n} : 0.5859017 最大離れ D_{\max} : 0.04153347

P値(%) : 87.83 帰無仮説の採択域 : (0 , 1.36367) **有意確率分布**

結 果 : 有意でない: 帰無仮説 (サンプルデータは正規分布に従う) を棄却できない

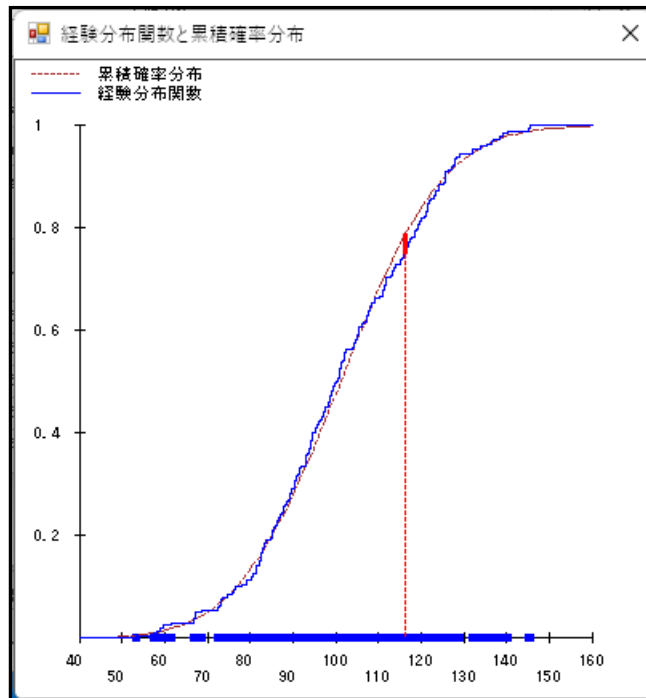
統計量が計算され (0.58)、p 値が 88 % となり、帰無仮説を棄却できない、つまり正規分布とみなせるという結論です。

有意確率分布 をクリックすると、



その様子を示しています。

更に **経験分布関数** をクリックすると、下図が表示されます。

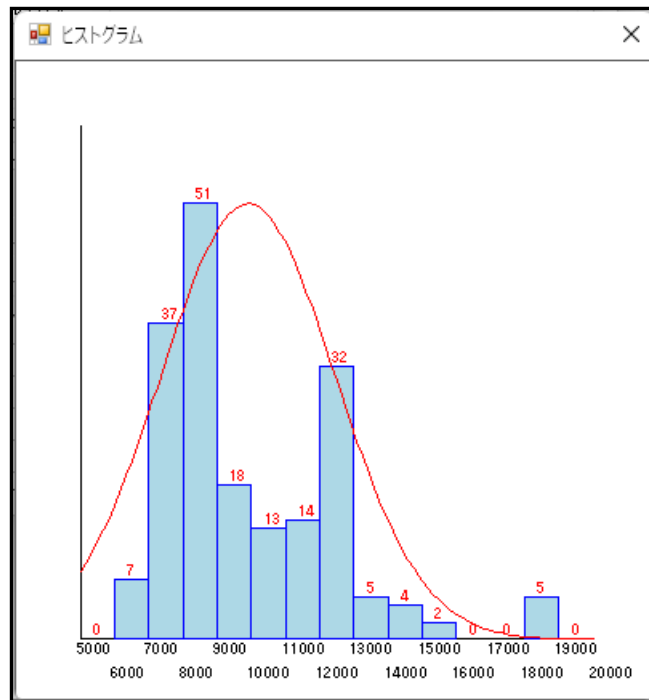


理論上の正規分布の場合の分布（———）と、
入力サンプルデータによる分布（———）との差異が
表現されており、一番離れている位置（最大離れ D_{max} ）が
表示されています。

次に 正規分布からのサンプルとはみなしにくいデータを例に、
正規性の検定を行います。

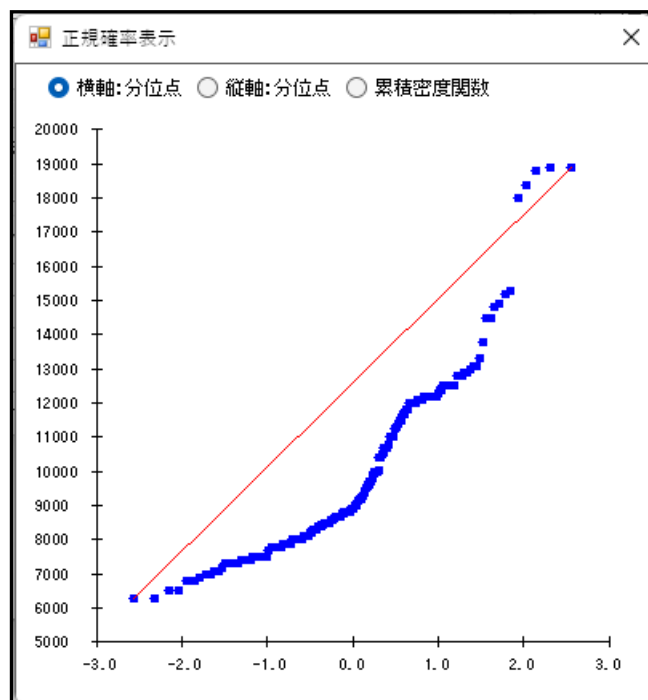
| |
|--------|
| 7,900 |
| 8,500 |
| 10,800 |
| 10,800 |
| |
| 8,100 |
| 8,600 |
| 7,100 |
| 18,400 |

ヒストグラムは 以下のようにです。



正規確率プロット

は 以下のようにです。



「コルモゴロフ・スミルノフ検定」を実行した結果は以下となります。

データ数が20を超える場合

コルモゴロフ・スミルノフ検定

経験分布関数

統計量 $D_{max} \sqrt{n}$: 2.190641

最大離れ D_{max} : 0.1597689

P値(%) : 0.01459

帰無仮説の採択域 : (0 , 1.36367)

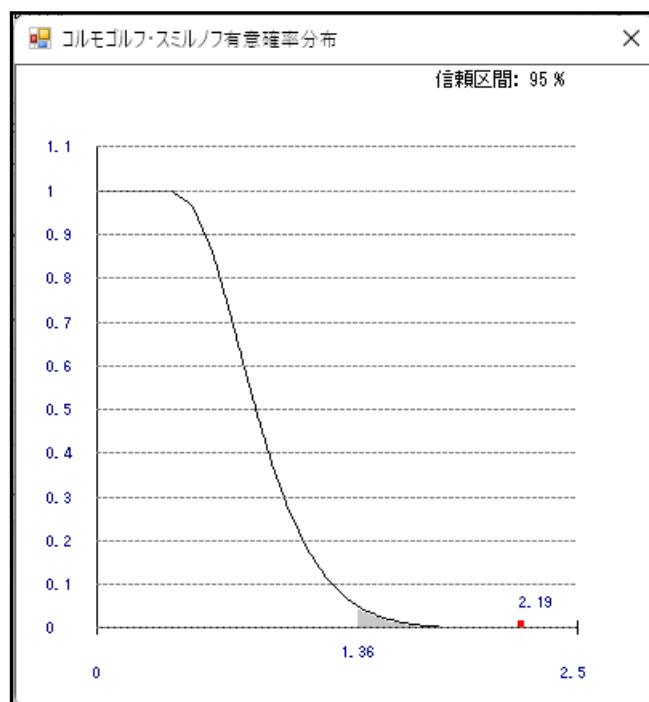
有意確率分布

結 果 : 有意 : 帰無仮説 (サンプルデータは正規分布に従う) を棄却する

統計量が計算され (2.19)、p 値が 0.01% となり、帰無仮説を棄却する、つまり正規分布とみなせないという結論です。

有意確率分布

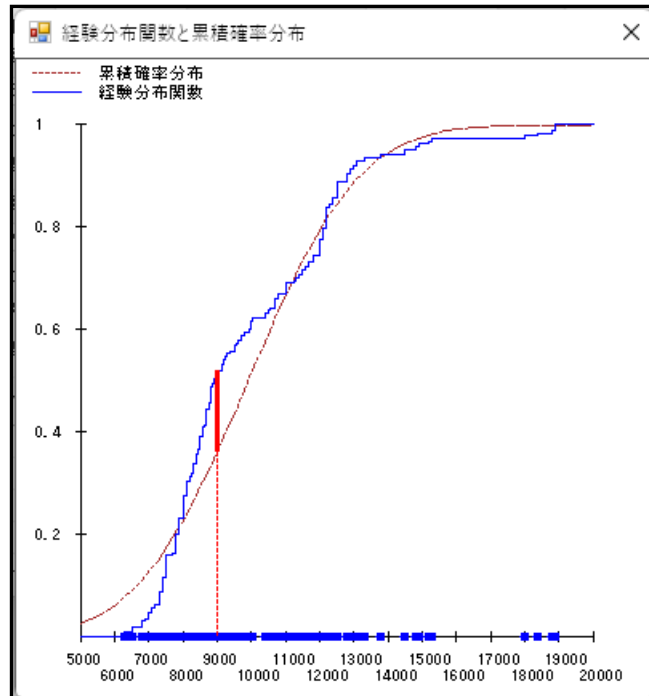
をクリックすると、



その様子が分かります。

経験分布関数

をクリックすると、下図が表示されます。



理論上の正規分布の場合の分布（——）と、
入力サンプルデータによる分布（——）との差異が
表現されており、一番離れている位置（最大離れ D_{max} ）が
0.1597と表示されています。