Chapter 1

基礎統計

1.1 統計的検定

1.1.1 用語の説明

まず、この節の理解に必要な基本的用語を説明します。

・母集団

調べたい対象となる人や物の集まり全体の集合のことです。例えば、日本人の身長を調べたいときには、日本人全体が母集団となります。

・母平均

母集団 $\{x_1, x_2, \cdots, x_N\}$ に対して、各 x_i の生起確率を p_i とするとき、次の式で与えられる値のことを母平均 μ といいます。

$$\mu = x_1 p_1 + x_2 p_2 + \cdots + x_N p_N = \sum_{i=1}^{N} x_i p_i$$

1.1.2 統計的検定とは

ある製薬メーカーが、新しい風邪薬の開発に取り組んでいたとします。当然､新しい風邪薬は従来の風邪薬より効果があってほしいものです。このとき、新しい風邪薬と従来の風邪薬を比較し、新しい風邪薬の方が有効であることを客観的に示す方法の一つに統計的検定があります。

大まかに説明すると統計的検定とは以下のような手順で進められます。まず、研究対象としての母集団に対して帰無仮説と対立仮説と呼ばれるものを立てます。風邪薬の例でいえば、患者という母集団に対して、新しい風邪薬と従来の風邪薬の効果は同じであるという帰無仮説を立てます。対立仮説は、新しい風邪薬と従来の風邪薬には違いがある（差がある）と立てます。次に、母集団からサンプルをランダムに抽出しデータを取得します。そして、それ

らの値から検定統計量と呼ばれるものを計算します。検定統計量は、各種の検定ごとに計算式が決まっており、そのそれぞれに対して従う確率分布が決まっている確率変数であるので、その確率分布からそれがどのくらいの確率であるのかを調べます。その確率が極めて低い場合（検定統計量が棄却域に入るとき）、帰無仮説が棄却されて対立仮説が採用されます。

この統計的検定は、仮説の検定とも呼ばれます。

(1) 帰無仮説と対立仮説

母集団についての仮定を帰無仮説といいます。統計においては H_0 という記号で表されることが多いです。この仮定と異なる仮説を対立仮説といい、H_1 という記号で表されます。

例えば、母集団の母平均 μ が μ_0 であると仮定した場合、帰無仮説は、

帰無仮説 H_0：母平均 μ は μ_0 である … $\mu = \mu_0$

となります。この帰無仮説に対して対立仮説は、

対立仮説は、母平均 μ が μ_0 と異なる

となりますから、次の 3 通りの場合が考えられます。

① 対立仮説 H_1：母平均 μ は μ_0 でない … $\mu \neq \mu_0$

② 対立仮説 H_1：母平均 μ は μ_0 より小さい … $\mu < \mu_0$

③ 対立仮説 H_1：母平均 μ は μ_0 より大きい … $\mu > \mu_0$

(2) 両側検定と片側検定

対立仮説の取り方によって、統計的検定は両側検定か片側検定に分けられます。

① 対立仮説 H_1：$\mu \neq \mu_0 \Rightarrow$ 両側検定

② 対立仮説 H_1：$\mu < \mu_0 \Rightarrow$ 片側検定

③ 対立仮説 H_1：$\mu > \mu_0 \Rightarrow$ 片側検定

(3) 有意水準と棄却域

検定を行うときは、初めに定めておく確率 α があります。これは、調べている対象の起こりやすさの基準を表しており、統計では有意水準と呼ばれています。一般的には $\alpha = 0.05$ とします。

この有意水準 α を与える（もとの仮説が誤りと判断する）領域のことを棄却域といいます。

① 両側検定の場合

$$\begin{cases} \text{帰無仮説} & H_0 : \mu = \mu_0 \\ \text{対立仮説} & H_1 : \mu \neq \mu_0 \end{cases}$$

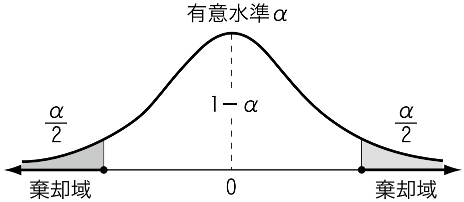

図 1.1.1　両側検定の場合

② 片側検定の場合

$$\begin{cases} \text{帰無仮説} & H_0 : \mu = \mu_0 \\ \text{対立仮説} & H_1 : \mu < \mu_0 \end{cases}$$

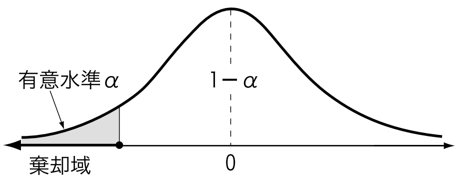

図 1.1.2　片側検定の場合 1

③ 片側検定の場合

$$\begin{cases} \text{帰無仮説} & H_0 : \mu = \mu_0 \\ \text{対立仮説} & H_1 : \mu > \mu_0 \end{cases}$$

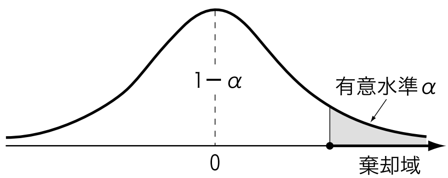

図 1.1.3　片側検定の場合 2

(4) 有意確率と有意水準

「検定統計量が棄却域に入るとき、帰無仮説 H_0 を棄てる」とは、次のこと

と同じ意味です。それは、「有意確率≦有意水準のとき、帰無仮説 H_0 を棄てる」です。図1.1.4、図1.1.5 に検定統計量が棄却域に入るときの例を表しておきます。

① 両側検定の場合

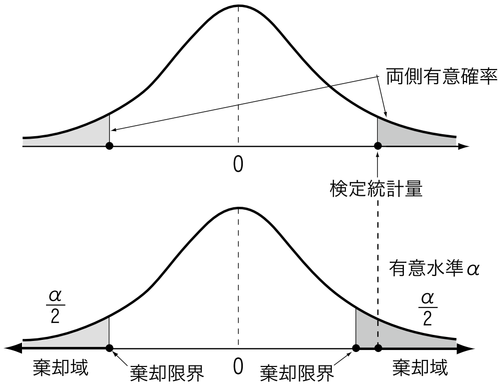

図 1.1.4　両側検定の場合

② 片側検定の場合

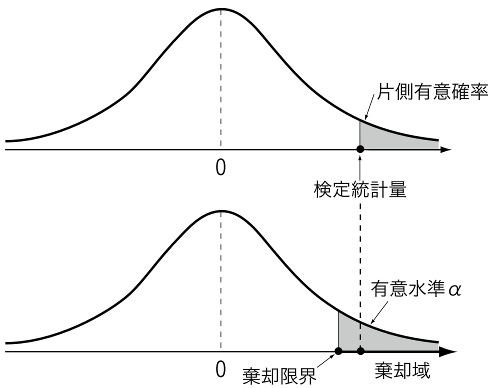

図 1.1.5　片側検定の場合

1.2　正規性の検定

1.2.1　用語の説明

まず、この節の理解に必要な基本的用語を説明します。

・標本

母集団からでたらめに取り出された一部分のことです。例えば、日本人の身長を調べたいときに、日本全国から 1,000 名の身長を測ったとしたら、その 1,000 名の人が標本となります。

・分布

母集団の確率分布やデータの散らばり具合を指します。例えば、日本人全体の身長の分布をグラムで表したりします。確率分としては、正規分布やカイ二乗分布、F 分布などがあります。

・正規分布

母集団が従う分布のうちの一つで、次の式に従うものです。

$$f(x)=\frac{1}{\sqrt{2\pi\sigma}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

ここで、μ：母平均　σ^2：母分散を表します。

・母分散

母集団 $\{x_1, x_2, \cdots, x_N\}$ に対して、各 x_i の生起確率を p_i とするとき、次の式で与えられる値のことを母分散 σ^2 といいます。

$$\begin{aligned}\sigma^2 &= (x_1-\mu)^2p_1+(x_2-\mu)^2p_2+\cdots+(x_N-\mu)^2p_N\\ &=\sum_{i=1}^{N}x_i^2p_i-\mu^2\end{aligned}$$

1.2.2　歪度、尖度を用いた検定

区間推定や仮説の検定をするとき、まず問題となるのは、調べたい母集団が正規分布に従っているかということです。そこで、母集団の正規性を調べる方法として 歪度と尖度を利用する方法があります。正規分布の場合、母集団の歪度は 0、尖度は 3 となることがわかっています。

(1) 歪度を用いて対称性からのズレを検定する

【手順 1】データからヒストグラムを描き、その分布の形を調べます。

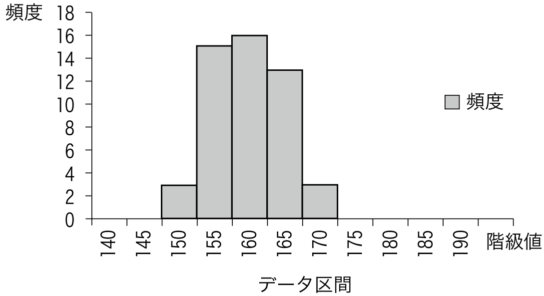

図 1.2.1　ヒストグラム

【手順 2】次の仮説を検定します。

帰無仮説　H_0：母集団の分布は左右対称である（歪度が 0 である）

対立仮説　H_1：母集団の分布のスソは左に長い（または右に長い）

（歪度が 0 ではない）

【手順 3】歪度を計算します。

この計算には、Excel の分析ツールを利用します。

初めに、分析ツールの中の基本統計量を選択します。

図 1.2.2　分析ツール画面

図 1.2.3　基本統計量入力画面

次に、Excel に入力したデータの範囲を指定し、基本統計量を計算します。この出力結果は、表 1.2.1 のようになります。

表 1.2.1　出力結果

身長	
平均	158.1
標準誤差	0.706385
中央値（メジアン）	158
最頻値（モード）	154
標準偏差	4.994895
分散	24.94898
尖度	-0.40203
歪度	0.211566
範囲	21
最小	148
最大	169
合計	7905
標本数	50
信頼区間（95.0 %）	1.419534

【手順 4】上の出力結果の歪度と巻末の〔歪度のパーセント点〕の数表を比較します。

標本数が 50、有意水準 0.05 より 0.211566 ＜ 0.533 なので、
帰無仮説 H_0 は棄てられません。

したがって、母集団の分布は左右対称になっていると仮定できます。

(2) 尖度を用いて、スソの長さを検定する

【手順 1】データからヒストグラムを描き、その分布の形を調べます。

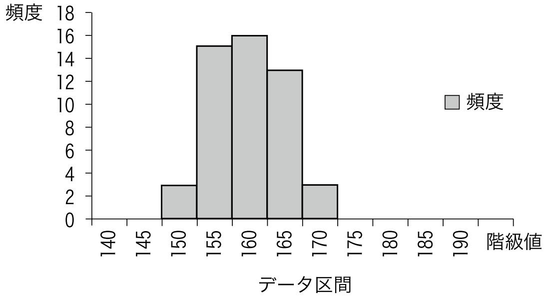

図 1.2.4　ヒストグラム

【手順 2】次の仮説を検定します。

帰無仮説　H_0：母集団の分布は左右対称である

対立仮説　H_1：母集団の分布のスソは長い（または短い）

【手順 3】尖度を計算します。

図 1.2.5　分析ツール画面

Excel の分析ツールを利用します。

始めに、分析ツールの中の、基本統計量を選択します。

図 1.2.6　基本統計量入力画面

次に、Excel に入力したデータの範囲を指定し、基本統計量を計算します。この出力結果は、表 1.2.2 のようになります。

表 1.2.2　出力結果

身長	
平均	158.1
標準誤差	0.706385
中央値（メジアン）	158
最頻値（モード）	154
標準偏差	4.994895
分散	24.94898
尖度	-0.40203
歪度	0.211566
範囲	21
最小	148
最大	169
合計	7905
標本数	50
信頼区間（95.0%）	1.419534

【手順 4】上の出力結果の尖度と巻末の〔尖度のパーセント点〕の数表を比較します。
標本数 50、有意水準 0.05 より $-0.40203 < 3.99$ なので
帰無仮説 H_0 は棄却されません。
したがって、母集団のスソの長さは正規分布と変わらないと仮定できます。