

5.3 F分布

様々な確率分布が次々に出てきて、閉口してしまった方もいらっしゃるかもしれません。このF分布で終わりです（あと第9章で q 分布というのが出でますが、限界値を読み取る専用です）。もちろん確率分布は他にもいろいろあるのですが、統計学の初心者ならば、ここまでに学んだ離散型の二項分布とポアソン分布、連續型の正規分布、 z 分布、 t 分布、 χ^2 分布、そして本節で学ぶF分布を理解すれば十分でしょう。もう一頑張りしてください。

F分布 (*F-distribution*) も“F”という統計量が従う連續型確率分布の一種です。F分布という名称は、G・W・スネデカーというアメリカの統計学者が、この統計量を考案したR・フィッシャー (Fisher, トピックス⑥) の頭文字から取りました（名付け親の名前からスネデカー分布とも呼ばれます）。

このF分布は、これまでの確率分布とは決定的に違う性質を持っています。それは、これまでの確率分布が、1つの母集団から何度も無作為抽出した標本の統計量が従う分布であったのに対し、F分布は2つの母集団からそれぞれ無作為抽出した標本から計算した統計量（これがFです）が従う分布であるという点です。その性質を利用して、抽出元である2つの母集団の分散が同じかどうかを判定するとき（**等分散の検定**）などに用いられます。例えば、中学校Aと中学校Bの数学の成績のバラツキに差があるかどうかを調べたいときなど、いろいろな応用場面が考えられるでしょう。等分散の検定は、第7章の最後に

扱いますので、とりあえずここでは「F分布とは何ぞや」ということについて解説しておきます。

F分布に従う統計量であるFの具体的な内容ですが、それはすばり「2つの χ^2 の比」です(図5.4)。ただし、 χ^2 の値は自由度によって大きく異なるので、2つ(比になる分母と分子の) χ^2 をそれぞれの自由度で割ることで、分母と分子の値の大きさを自由度あたりに修正して対等の状態にします。このように、Fには自由度が2つになる点に注意してください(もちろん同じ自由度になることもあります)。

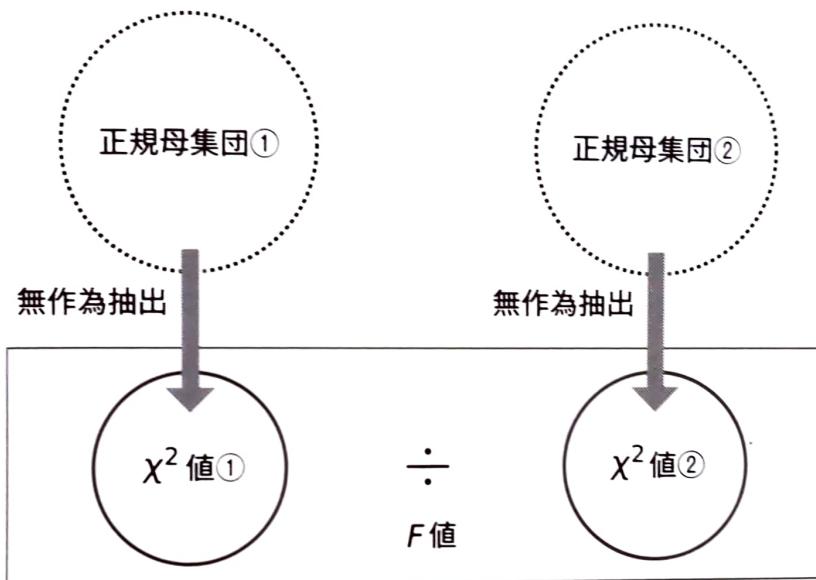


図 5.4 Fの意味

Fを式で表すと次のようにになります。 ν_1 は分子の χ^2 の自由度(第1自由度と呼びます)、 ν_2 は分母の χ^2 の自由度(第2自由度)です。

$$F \text{ 統計量 } F_{(\nu_1, \nu_2)} = \frac{\chi_{(\nu_1)}^2 / \nu_1}{\chi_{(\nu_2)}^2 / \nu_2}$$

なお、このFが従うF分布の確率密度関数は、次のような、とても複雑な式です(入門書では普通は掲載しません……)。

$$F \text{ 分布の確率密度関数 } f(F) = \frac{\Gamma(\frac{\nu_1 + \nu_2}{2}) F^{\frac{\nu_1 - 2}{2}}}{\Gamma(\frac{\nu_1}{2}) \Gamma(\frac{\nu_2}{2}) (1 + \frac{\nu_1}{\nu_2} F)^{\frac{\nu_1 + \nu_2}{2}}} \left(\frac{\nu_1}{\nu_2} \right)^{\frac{\nu_1}{2}}$$

みなさんに、この複雑な式で確認して欲しいのは、 χ^2 分布同様、分布の形を決める母数は自由度のみ（ただし2つ）ということです（それ以外は一般化した階乗を意味するΓ関数がたくさんあるだけ）。 χ^2 の比をとった統計量が従う分布なので、当たり前といえば当たり前なのですが……。

F 分布の平均（期待値）と分散は次のようになります。こちらも χ^2 分布同様、証明は複雑なので省略しますが、平均は第2自由度 ν_2 が大きくなると1に近づくことが簡単にわかりますね。

$$F \text{ 分布の期待値 (平均)} \quad E(F) = \frac{\nu_2}{\nu_2 - 2} \quad (\text{ただし } \nu_2 > 2)$$

$$F \text{ 分布の分散} \quad V(F) = \frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)} \quad (\text{ただし } \nu_2 > 4)$$

F 分布の形状は図5.5のようになります。横軸である確率変数の F は（ χ^2 の）“比”ですから負の値は取らず、右に歪んで左右非対称の形をしています。

自由度が2つあるので様々な組合せがあり、その度に異なる形状を取るのですが、 χ^2 分布のように自由度によって劇的に形が変わるわけではありません。また、いくら自由度が大きくなっても正規分布に近づくこともありません。

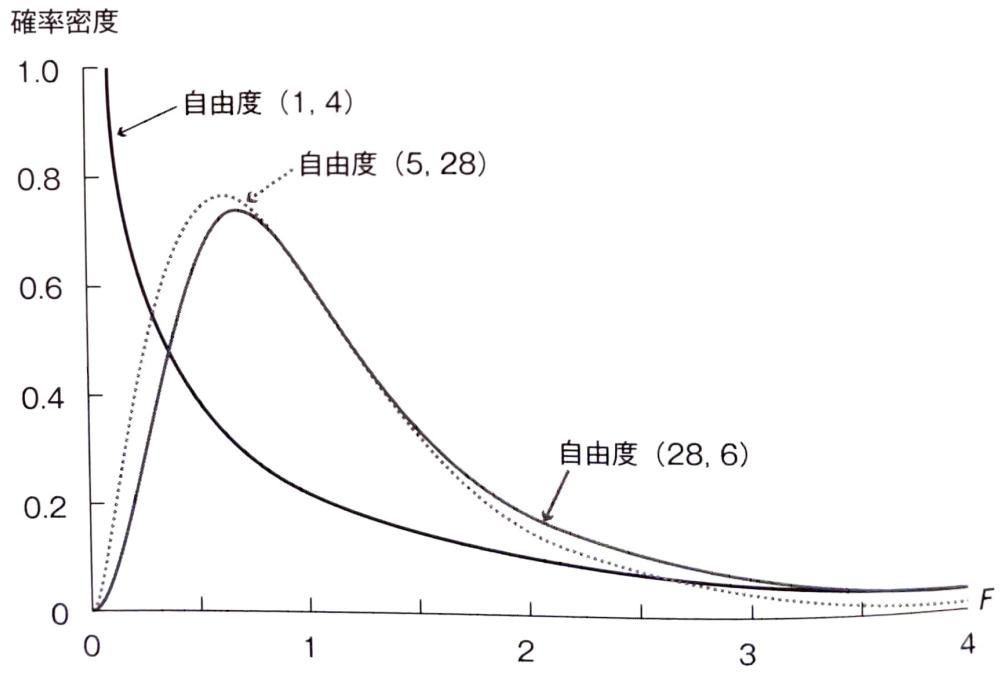


図 5.5 2つの自由度によって形の変わる F 分布

5.4 特別な F 値

● F 値が2つの不偏分散の比になる場合

本節では、ある条件の下で、 F が2つの不偏分散の比になることや、 t と密接な関係があることを導いておきます。とくに不偏分散の比となる F は、等分散の検定（第7章）や分散分析（第8章）で実施される F 検定の統計量となりますので、ここで一旦、理解しておいてください。

まず、節5.2で導いた自由度が $n - 1$ の χ^2 の式を思い出してください。この式の $n - 1$ を ν にして表現し直すと、

$$\chi_{(n-1)}^2 = \frac{(n-1) \times \hat{\sigma}^2}{\sigma^2} \rightarrow \chi_{(\nu)}^2 = \frac{\nu \hat{\sigma}^2}{\sigma^2}$$

となります。ということは、前節で示した F の式（2つの χ^2/ν の比）は、不偏分散 $\hat{\sigma}^2$ を使って、次のように整理できます。

$$F_{(\nu_1, \nu_2)} = \frac{\chi_{(\nu_1)}^2 / \nu_1}{\chi_{(\nu_2)}^2 / \nu_2} = \frac{\frac{\nu_1 \hat{\sigma}_1^2}{\sigma_1^2} / \nu_1}{\frac{\nu_2 \hat{\sigma}_2^2}{\sigma_2^2} / \nu_2} = \frac{\hat{\sigma}_1^2 / \sigma_1^2}{\hat{\sigma}_2^2 / \sigma_2^2}$$

ここで、もし分子の不偏分散 $\hat{\sigma}_1^2$ の真の値である母分散 σ_1^2 と、分母の不偏分散 $\hat{\sigma}_2^2$ の真の値である母分散 σ_2^2 とが同じ σ^2 だったら（検定ではそのような仮説の下で統計量 F を計算します），次のような不偏分散 $\hat{\sigma}_1^2$ と $\hat{\sigma}_2^2$ の単純な比になります。

$$F_{(\nu_1, \nu_2)} = \frac{\hat{\sigma}_1^2 / \sigma^2}{\hat{\sigma}_2^2 / \sigma^2} = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}$$

当然ながら、2つの標本が同じ分散の母集団から抽出されたならば、そこから推定した2つの不偏分散の値も近くなるため、その比である統計量 F は“1”前後の値を取りやすいでしょう。その性質を利用して、等分散の検定では母集団の分散が等しいか否かを、分散分析では処理の効果があるかないかを検証するのです。

● F と t の関係

最後に、 F は t とも関係が深いことを導いておきましょう。

まず、標本平均 \bar{x} の準標準化変量 $t_{\bar{x}}$ の式（節4.4）を、次のように $z_{\bar{x}}$ を使った式に変形します。

$$t_{\bar{x}} = \frac{\bar{x} - \mu}{\hat{\sigma}_{\bar{x}}} = \frac{\bar{x} - \mu}{\hat{\sigma}/\sqrt{n}} = \frac{\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}}{\frac{\hat{\sigma}/\sqrt{n}}{\sigma/\sqrt{n}}} = \frac{z_{\bar{x}}}{\hat{\sigma}/\sigma}$$

この式に、先ほどの

$$\chi^2_{(\nu)} = \frac{\nu \hat{\sigma}^2}{\sigma^2}$$

という式を使って、次のように右辺分母を直します。

$$t_{\bar{x}} = \frac{z_{\bar{x}}}{\sqrt{\frac{\chi^2_{(\nu)}}{\nu}}}$$

この式を2乗すると、次のように、 t の2乗と、第1自由度が1の F (自由度1の χ^2 は z^2 でしたね) とが同じになることがわかります。

$$t_{(\nu)}^2 = \frac{z_{\bar{x}}^2}{\chi^2_{(\nu)}/\nu} = \frac{\chi^2_{(1)}/1}{\chi^2_{(\nu)}/\nu} = F_{(1,\nu)}$$

例えば、 F を使う分散分析（第8章）で、2群の場合は第1自由度が1になります。つまり、2群の分散分析を実施するのと2群の平均の差の t 検定（第7章）を実施することは同じことなのです。みなさんはまだ検定を学んでいないので、とりあえずここでは t 分布と F 分布が密接な関係になるんだなと感じていただければ結構です。

そして、この2つ以外にも、第2章以降学んできた確率分布は少なからず関連しているのです。48ページの図2.13の拡張版として、ここまでに学んだ全ての確率分布の関係を、図5.6に整理しておきます。

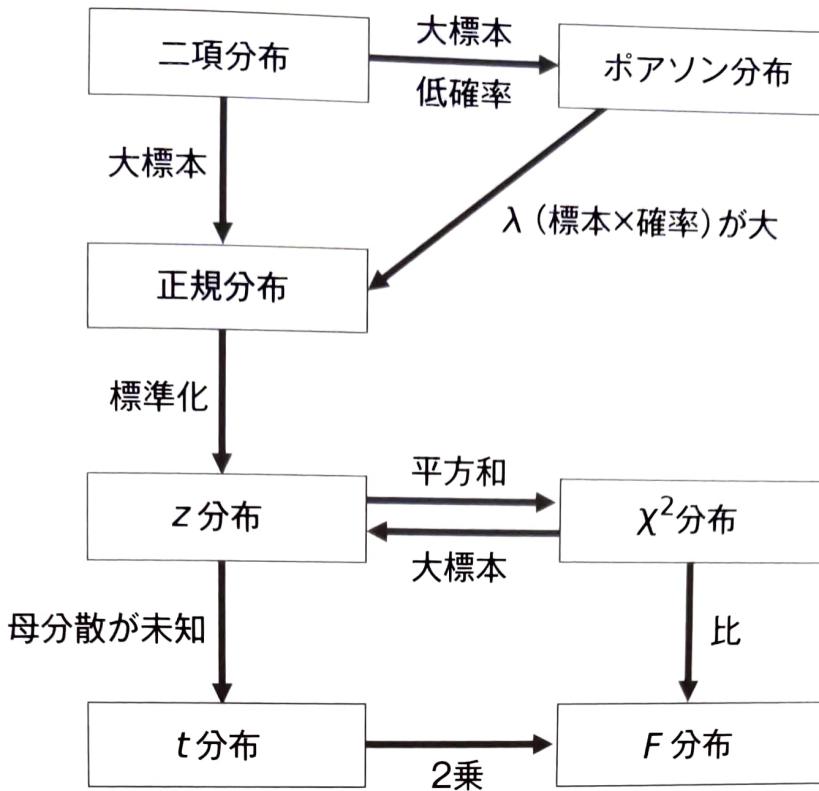


図 5.6 確率分布の関係（図 2.13 の拡張版）

● F 分布表と読み方

F 分布も確率分布ですので、確率密度関数を積分すれば、所与の F の値とそこから上（右）側の確率との対応関係を表にできます。表 5.2 に、その F 分布表の一部を掲載しておきます（全体表は付録 IV と V に掲載しています）。

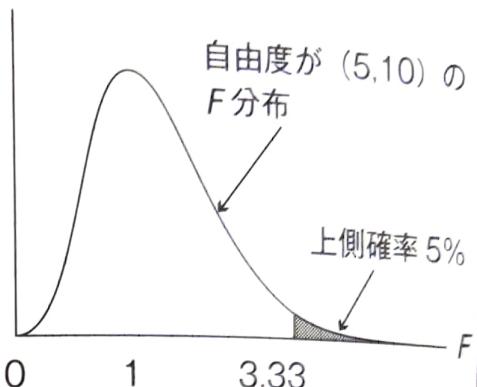
F 値には、分子の自由度（第1自由度 ν_1 ）と、分母の自由度（第2自由度 ν_2 ）の2つがあるので、確率ごとに1つの表が必要になります。ここでは上側確率が 5% ($p = 0.05$) の F 分布表を載せておきます。

さて、読み方ですが、例えば「第1自由度 ν_1 が 5、第2自由度 ν_2 が 10 の F 値（上側確率 5%）」は、表側に 10 と書かれている行と、表頭に 5 と書かれている列とのクロスしている値である“3.33”となります。

Excel ならば $F.INV.RT(0.05, 5, 10)$ という関数で、分子の第1自由度が 5、分母の第2自由度が 10 の F 分布における上側確率 5% の F 値を求めることができます。

表 5.2 F分布表の一部（表頭と表側の自由度に対応するF値、上側確率5%）

		v_1 (分子の自由度)											
		1	2	3	4	5	6	7	8	9	10	15	20
v_2 (分母 の 自由 度)	1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	245.95	248.01
	2	18.51	19.00	19.16	19.25	19.30	19.33						
	3	10.13	9.55	9.28	9.12	9.01	8.94						
	4	7.71	6.94	6.59	6.39	6.26	6.16						
	5	6.61	5.79	5.41	5.19	5.05	4.95						
	6	5.99	5.14	4.76	4.53	4.39	4.28						
	7	5.59	4.74	4.35	4.12	3.97	3.87						
	8	5.32	4.46	4.07	3.84	3.69	3.58						
	9	5.12	4.26	3.86	3.63	3.48	3.37						
	10	4.96	4.10	3.71	3.48	3.33	3.22						



章末問題

問1 χ^2 値とは何かについて、標準化変量 z_i を使って説明しなさい。ただし式を書く必要はありません。

問2 母平均 μ が未知の場合、 χ^2 値の自由度はいくつになるか。標本サイズは n とします。

問3 ある農家の10年間の平均農業所得率の不偏標準偏差は5%でした。この農家の経営上のリスク、つまり農業所得率の母標準偏差が入ると考えられる信頼区間を信頼係数90%で推定しなさい（ヒント：データが10年分ありますから、標本サイズ n は10になります）。

問4 F値とt値の関係について述べなさい。

第6章

仮説検定と検出力

仮説検定：仮説が正しいにしては起きにくいことが起きたと考えられる場合、そもそも仮説は間違っていたと判断する。

検出力：母集団に差があるときに“ある”と正しく判断できる能力。

6.1 検定の概要

本章では、信頼区間の推定とともに推測統計学の二枚看板のもう1つである「仮説検定」の基本と、検定の能力を表す「検出力」を学びます。

仮説検定 (hypothesis testing) は統計的仮説検定、あるいは単に検定 (test)とも呼ばれ、薬を使う・使わないなど、処理の違いで生まれた複数群の母平均などに、差があるかないかを判断する手法です。

仮説検定は図 6.1 のような4つの手順から構成されています。

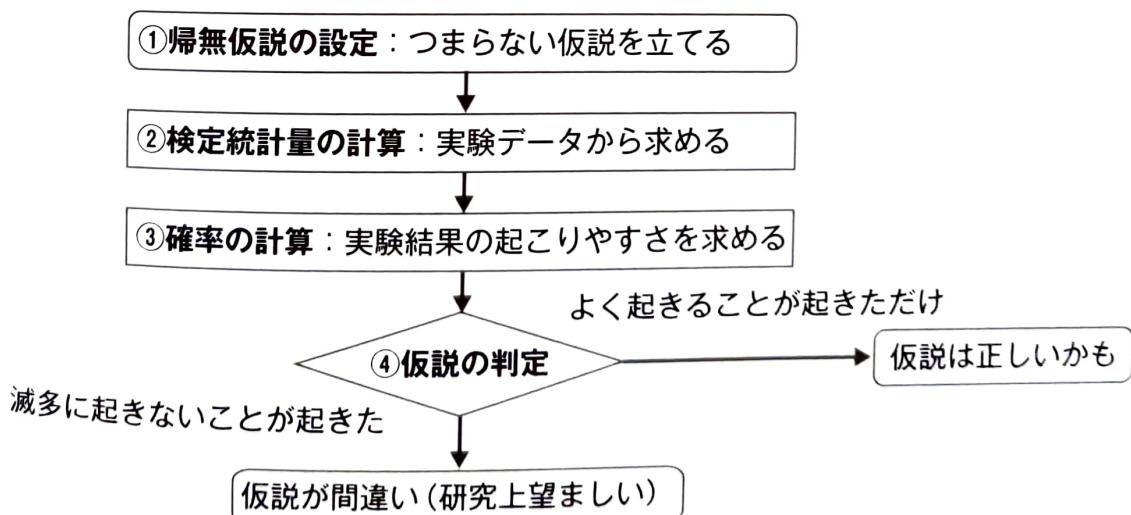


図 6.1 仮説検定の基本手順

手順①：仮説の設定



まず母集団に関する仮説を立てます。仮説とは、未知の事実を説明するため、とりあえず正しいものと仮定しておく道筋のことで、それが正しいか正しくないかを検証します。

手順②：検定統計量の計算



実験で観測されたデータのままで検定できることは滅多にないため、目的に沿った検定統計量を求めます。そのために、ここまで z や t 、 χ^2 、 F などを学んできたのです。

手順③：確率の計算



とりあえず仮説が正しいとした下での実験結果の起こりやすさ（生起確率）を計算し、次の手順の判断材料とします。ただし、確率を計算するのは難しいので、実際には帰無仮説の分布でどれくらい端の方に位置している（極端な値を取っている）のかで判断します。

手順④：仮説の判定

滅多に起きないことが起きたといえるぐらいに確率が低いならば、「そもそも前提していた仮説が間違いだった」と判断します。そうでない場合には判定を保留します（仮説は正しかったとはいわない）。



6.2 仮説の設定

● 帰無仮説の設定

手順①の母集団に関する仮説を立てるにあたって、1つ注意が必要となります。それは、研究目的からすると「つまらない内容（主張したい内容とは逆）」の仮説を立てるということです。

例えば、殺虫剤の効果を検定する場合、殺虫剤処理前と処理後の植物では平均害虫数に「差がある」という仮説を立て、それを証明したくなるのが自然だと思います。しかし、検定では、処理前と処理後とでは平均害虫数に「差はない」という、むしろ棄却したい仮説をまず立てるのです。なぜならば、差（処理の効果）があることを直接証明するためには、どの程度の差なのかをあらかじめ特定し、その大きさの差について検証する必要が出てきてしまうからです。しかし、その差の大きさがわからないから検定するのであって、これでは話が進みません。

そのため、検定では本来の研究目的からすると棄却したい「△△と××とで

は差はない」という仮説を立て、それが結果と矛盾していることを証明するのです。このように、証明したい仮説と相反する仮説を立て、それを棄却することで本来主張したい仮説を証明する論理を、**背理法**と呼びます。そして、検定の対象となる棄却したい仮説を、本来は無に帰すべき仮説という意味で**帰無仮説** (null hypothesis : null は 0 の意味) と呼び、頭文字を使って H_0 で表します。なお、**棄却** (reject) とは仮説が成り立たないこと、つまり正しくないと判断されたために検定結果で採択されなかったことを意味します。

● 対立仮説の設定

一方、帰無仮説が棄却された場合に、代わりに採択される仮説もあらかじめ立てておきます。その内容は帰無仮説を否定するものなので「差（処理効果）がないとはいえない」という二重否定になってわかりにくいのですが、「差がある」と同じ意味と考えてよいでしょう。

このような、本来採択して欲しい（もともと主張したかった）仮説を**対立仮説** (alternative hypothesis) と呼び、 H_1 で表します。「対立仮説など立てなくとも帰無仮説が棄却されれば検定の目的は達成できるじゃないか」と思われるかもしれません、本章後半で扱う検出力や第二種の過誤確率 β が計算できるようになるというメリットが生まれるのです。

なお、帰無仮説が棄却されなかった場合、つまり帰無仮説を受容せざるを得ない場合の解釈には注意が必要です。というのも、「差はない」という帰無仮説が棄却されなかつたからといって、それが直ちに帰無仮説が正しいことを意味しているわけではありません。なぜならば、別の標本や、標本がもっと大きかつたら帰無仮説を棄却できたかもしれないからです。よって、帰無仮説が棄却できなかつた場合でも、あくまで当該標本（今回の実験）からは統計的に差があるとはいえなかつたというだけなので、「差があるともないともいい切れない（今回の結果と帰無仮説は矛盾しない）」程度の消極的な解釈で、判断を保留しておくようにしましょう（図6.2）。

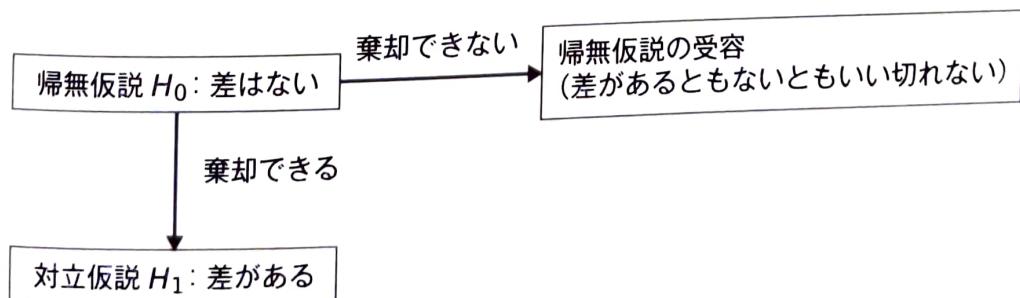


図 6.2 仮説の扱い方

このように、帰無仮説が棄却できなかったときに下される判定が曖昧になる点が、仮説検定の最大の弱点といつてもよいでしょう。とはいえ、結果をはっきり答えてくれない区間推定に比べ、「処理の効果がある」と断言してくれる仮説検定は、誰にとってもわかりやすいため、多くの分野で便利に使われているのです。

6.3 (1標本の) 母平均の検定

● 特定の値と標本平均の差の検定

仮説検定の、手順②（検定統計量の計算）から手順④（仮説の判定）については、早速、正規分布を使った基本的な仮説検定で説明しましょう（ただし、本節では手順②はスキップできます）。

それは（1標本の）**母平均の検定**（one sample test of means）です。この検定は、実験で観測された標本平均が、特定の値（既知の定数）と差があるかどうかを判定します。標本平均の値のままでも検定できるため、検定統計量を別途計算する必要がなく、検定の基本を理解するにはちょうどよいのです。なお、本検定にはいろいろな名称があり、「1サンプルの検定」、「(1変量の分布の) 平均の検定」、「1つの条件の平均値と定数との差の検定」、「平均値の検定」などとも呼ばれています。本書でも改訂前は「定数と平均との差の検定」と呼んでいましたが、あまり一般的ではないようなので「母平均の検定」に変更させていただきました。また、「1標本の」を括弧で付けているのは、次章で扱うしていただきました。「2群の平均の差の検定」を「2標本の母平均の差の検定」と呼ぶ場合があるので、それと区別できるようにしたためです。

具体的な応用例としては、病気が疑われるグループが本当に病気なのか、それとも健康なのかを判定したい場面などが考えられるでしょう（次節の例題で扱います）。病気が疑われるグループに実施した検査の値を、健康なグループの既知の値と比較し、両者間には差が無いという帰無仮説が棄却されたら「このグループはやはり病気だったな」と判定できることになります。

ここから検定の原理を解説しますが、まずは正規分布の性質を思い出してください（図6.3）。正規分布は、植物の長さや質量、生物の知能など、自然界で起こる現象の多くが当てはまる釣り鐘型の確率分布のことです。推定時の誤差を評価できるように標本平均 \bar{x} の分布を考えた場合でも、左右対称で真の値である母平均 μ が真ん中にある正規分布となります。標本平均 \bar{x} が取る横軸が連続型確率変数、縦軸が確率密度（その標本平均 \bar{x} の起きやすさ）で、何度も

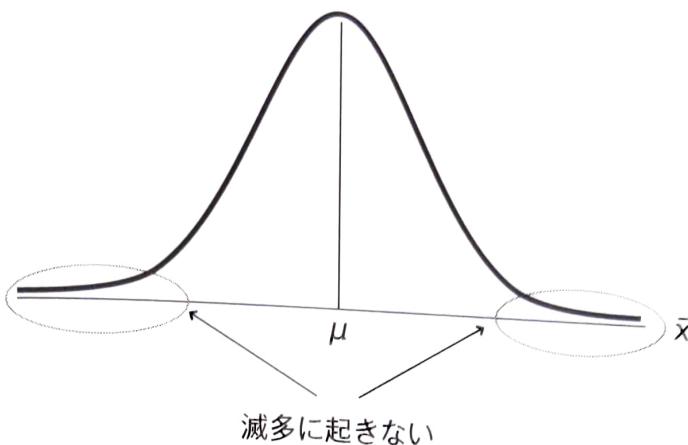


図 6.3 母平均から標本平均が大きく離れるることは滅多にない

験を独立に実施したとすると、真ん中の母平均の周辺には多くの標本平均が集まりやすく（値を取る確率が高く）、逆にそこから離れた両端の方には行きにくい（極端な値を取る確率は低い）という性質がありました。

母平均の検定では、この性質を利用して、観測された標本平均 \bar{x} が、特定の値 μ_0 を中心とした帰無仮説の分布のどのあたりに位置しているのかをみると、それらが同じ母集団から抽出されたものかどうかを判定します。なお、比較対象である特定の値 μ_0 は、後ほど示す帰無仮説 H_0 が正しい場合に共通の母平均になるので、0の添字をつけて標本抽出元の母平均 μ と区別します。

検定の流れを見ていきましょう。先ほど紹介した手順（図6.1）よりも詳細に説明したいのでステップという言葉を使わせていただきます（同じ4段階に分割していますが、内容は一致していません）。

ステップ①：2つの仮説を考える

母平均が μ の母集団から無作為抽出された標本があります（実験や調査で観測したデータのことです）。その標本から計算した標本平均が \bar{x} だとします。

今回、この \bar{x} の真の値である μ と特定の値 μ_0 との間で、差があるか／ないかを検定することを考えます。特定の値 μ_0 は既知で、具体的な値が入っていますが、同時にその母集団の母平均（真の値）でもあります。

実際には μ は未知ですので、比較する（検定対象となる）のは、その標本平均 \bar{x} と特定の値 μ_0 になります。ただし、仮説は母集団について立てるものなので、帰無仮説は「標本平均 \bar{x} の抽出元の母平均 μ と、特定の値の抽出元の母平均 μ_0 との間には差がない ($H_0 : \mu = \mu_0$)」となります。差があるとかないとかいうと意味がわかりにくくなってしまいますが、 \bar{x} の母平均 μ と μ_0 との間に差がないということは、図6.4左のように、 \bar{x} は比較対象である μ_0 と同じ母に差がないということは、図6.4左のように、 \bar{x} は比較対象である μ_0 と同じ母

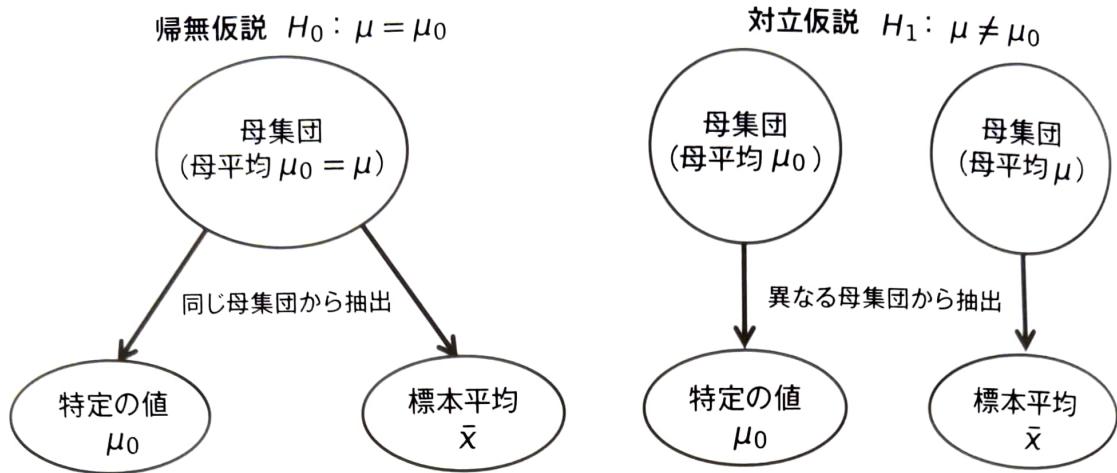


図 6.4 「(1 標本の) 母平均の検定」の仮説の考え方

集団から抽出されたということ、つまり μ_0 は \bar{x} の真の値でもあるということを意味します。逆に μ と μ_0 とで差があるということ（対立仮説 $H_1: \mu \neq \mu_0$ ）は、図 6.4 右のように、 \bar{x} は別の母集団から抽出されたということ、つまり μ_0 は \bar{x} の真の値ではないということを意味します。

なお、 μ_0 は μ より小さく（大きく）ならないと仮定できる場合や、 μ_0 が μ より小さく（大きく）なったことに興味が無い場合には、対立仮説 H_1 を $\mu < \mu_0$ ($\mu > \mu_0$) とすることもあります。このような場合には、分布の片側だけで判定すればよいので**片側検定** (one-sided test) と呼ばれます。ただし、そのような場面はそれほど多くはないでしょうから、本書では $H_1: \mu \neq \mu_0$ とした**両側検定** (two-sided test) で説明を進めます（ただし、第8章で学ぶ分散分析は片側検定が基本です）。

ステップ②：仮説が正しい下での標本の位置を考える

帰無仮説と対立仮説、それぞれ正しい場合の、標本分布（母平均の検定では標本平均の確率分布）の位置関係を図で整理しておきましょう。

図 6.5 が、帰無仮説 $H_0: \mu = \mu_0$ が正しいと仮定した下での標本分布（略して「帰無仮説の分布」）です。帰無仮説が正しいということは、 μ_0 を真の値（母平均）とした破線の標本分布（←比較対象）と、 μ を真の値（母平均）とした実線の標本分布（←実験結果）が重なっている状態です。ただし、実際の母平均の検定では、前者は具体的な値が与えられているので、破線のような分布を考える必要はありません（あくまで理解のために描きました）。この（帰無仮説が正しい）場合、標本平均 \bar{x} は μ_0 に近い値を取る（観測される）確率が高くなるはずです。

このように、仮説検定では、帰無仮説の分布（図 6.5 のように重なっている

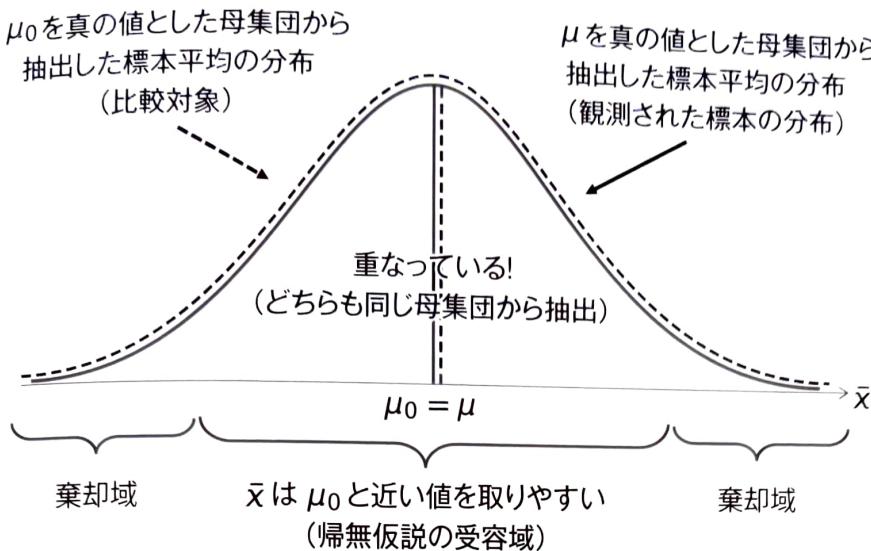


図 6.5 帰無仮説が正しい場合の分布

状態)において、統計量（標本平均 \bar{x} ）がどのあたりに位置しているか（ \bar{x} と μ_0 との距離）が重要となります。

というわけで、母平均の検定では、観測された標本平均 \bar{x} が、帰無仮説の分布の中央にある μ_0 に近い値を取ったときに、「(帰無仮説が正しいとすると)起こるべきことが起きただけなので、帰無仮説が間違いというのには無理がある」と判断します。ですから帰無仮説の母平均 μ_0 付近を、帰無仮説の受容域(採択域: acceptance area)と呼びます。

一方、標本平均 \bar{x} が帰無仮説の分布の両端に位置したとき、つまり μ_0 から離れた値を取ったときには「偶然でもなかなか起きないことが起きたので、そもそも帰無仮説は間違いであったと考える方が合理的である」と判断します。よって、帰無仮説の母平均 μ_0 から遠い、分布の両裾の端を（帰無仮説の）棄却域(rejection area)と呼びます。

帰無仮説が間違っていて対立仮説が正しい場合の標本分布である「対立仮説の分布」は図6.6のようになります。 μ が真の値である母集団から抽出された標本分布（←実験結果；右・実線）が、 μ_0 が真の値である母集団から抽出した標本分布（←比較対象；左・破線）とは違う場所に位置しています。なお、図6.6では対立仮説の分布を帰無仮説の分布の右側に描いていますが、ステップ①の最後に触れたように、 μ が必ず μ_0 よりも大きくなるという根拠がない限り、左側に位置することも想定します。

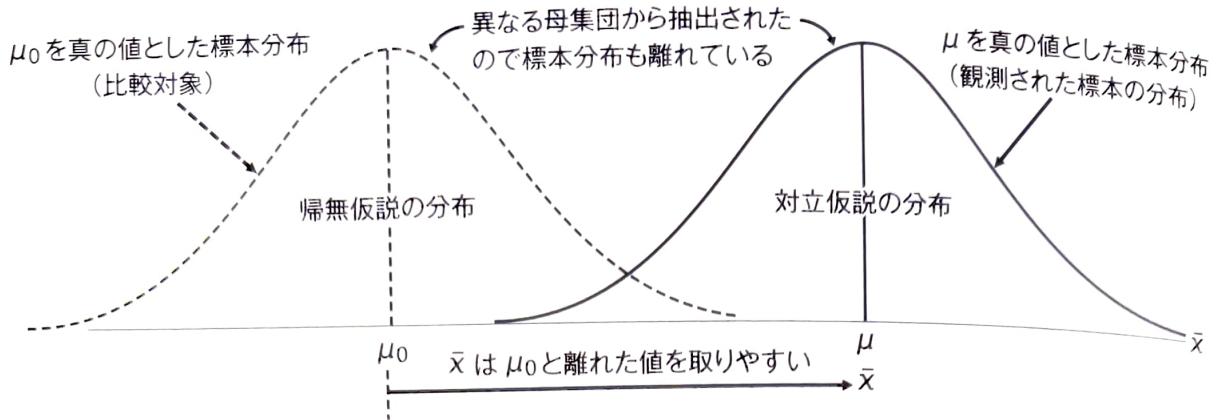


図 6.6 帰無仮説が間違っている（対立仮説が正しい）場合の分布

ステップ③：帰無仮説をどこから棄却するかを考える

それでは、どの程度、標本平均 \bar{x} が母平均 μ_0 から離れていれば、帰無仮説が間違ったと判断できるでしょうか。それが人によって変わるように、とても客観的な検定とはいえません。こうした帰無仮説を棄却するための、いわば足切り水準を**限界値**（**臨界値**、**境界値**；critical value）と呼びます。

図 6.7 に限界値の決め方を描きました。帰無仮説の下で検定統計量（標本平均 \bar{x} ）が従う分布において、限界値よりも外側の確率（グレー部分）をあらかじめ具体的に設定し、その確率に対応する値（= 限界値）を求めるのです。限界値を決めるこの両裾の確率の基準を**有意水準**（significance level）と呼び、確率の大きさを $\alpha = ○○$ と表しますが、それをいくつに設定するのかに決まりはありません。

大きく設定すれば限界値が μ_0 に近い値となるため、検定統計量が棄却域に入りやすくなりますが、簡単に棄却されるような甘い検定では実施する意味がなくなってしまいます。ですから、一般的には、数字の切りがよくて適度に厳しい $\alpha = 0.05$ (5 %) が用いられます。

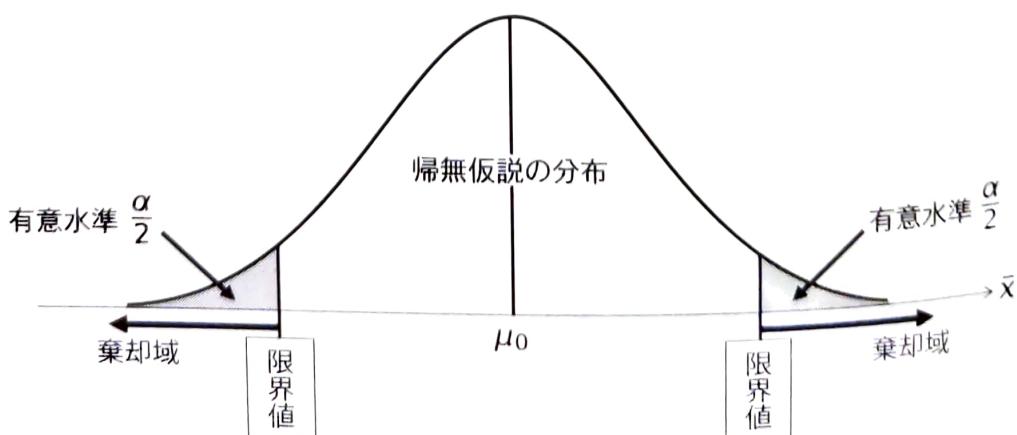


図 6.7 帰無仮説が間違っていると判断する基準（限界値）

ただし、精度の高さが求められる工業分野や自然科学分野では0.01(1%)が、逆に誤差の大きい社会科学分野では0.1(10%)まで許容されることもあります。初学者向けの本書では $\alpha = 0.05$ 、つまり有意水準5%を基本としますが、近年、もっと厳しくすべきだ(例えば $\alpha = 0.005$)という指摘もあることを申し添えておきます。

なお、図6.7は両側検定なので両側の確率を合わせて α (片側 $\alpha/2$ ずつ)としていますが、片側検定の場合には、対立仮説側だけの確率だけで α とします(ですから同じ α でも帰無仮説は片側検定の方が棄却しやすくなります)。

ステップ④：限界値を求めて検定統計量と比較する

最後のステップでは、ステップ③で設定した有意水準(両側5%: $\alpha = 0.05$)に対応する限界値と、検定統計量(観測された標本平均)とを比較して、いよいよ帰無仮説が正しいか否かを判定します。限界値の求め方は、区間推定のときの信頼限界(節4.3)と全く一緒ですが、復習がてら解説しましょう。

標準正規分布において、所与の上側(右裾)確率に対応する確率変数(標本平均)の値は、標準正規(z)分布表から読み取ることができます。付録Iの z 分布表で、“0.0250”となるのは、表側(z 値の小数点第1位まで)は“1.9”的行で、表頭(小数点第2位)は“0.06”的列なので、 z 分布で上側の確率が0.025(2.5%)となる z 値は“1.96”であることがわかります。

ただし、今回は、標本平均のままで標準化されていません(z ではなく \bar{x})ので、 z 分布表から読み取った値(1.96)に母標準誤差 $\sigma_{\bar{x}}$ (標本平均 \bar{x} の母標準偏差)を乗ずる必要があります(z ならば母標準誤差が1なので不要)。よって、両側有意水準5%($\alpha = 0.05$)に対応する限界値は、帰無仮説の分布の中央にある μ_0 の左右に $1.96 \times \sigma_{\bar{x}}$ のところ($\mu_0 \pm 1.96\sigma_{\bar{x}}$)となります。

というわけで、図6.8のように、検定統計量である観測された標本平均 \bar{x} の値が $\mu_0 + 1.96\sigma_{\bar{x}}$ よりも大きいか、あるいは $\mu_0 - 1.96\sigma_{\bar{x}}$ よりも小さい場合に、帰無仮説が正しいとした下では滅多に起こらないことが起きた(偶然としても20回に1回も出ない)ので、むしろ「帰無仮説は正しくなかった(μ_0 と \bar{x} との間の差は大きかった→ μ_0 は \bar{x} の真の値ではなかった)」と判断します。

なお、ここで改めて気がついた方も多いと思いますが、限界値は母標準誤差 $\sigma_{\bar{x}}$ の値に左右されます。つまり、データをたくさん集めて標本が大きくなるほど、限界値は μ_0 と近い値になるため、同じ有意水準であっても棄却域が広がり、帰無仮説は棄却されやすくなるのです(これが帰無仮説を正しいと断言してはいけない理由です)。

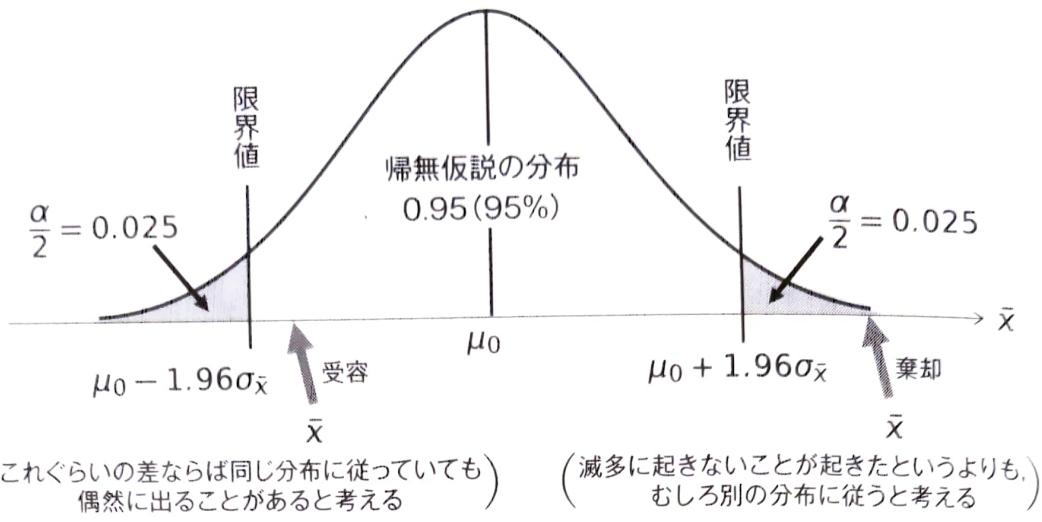


図 6.8 両側 5% 水準で帰無仮説が棄却される場合とされない場合

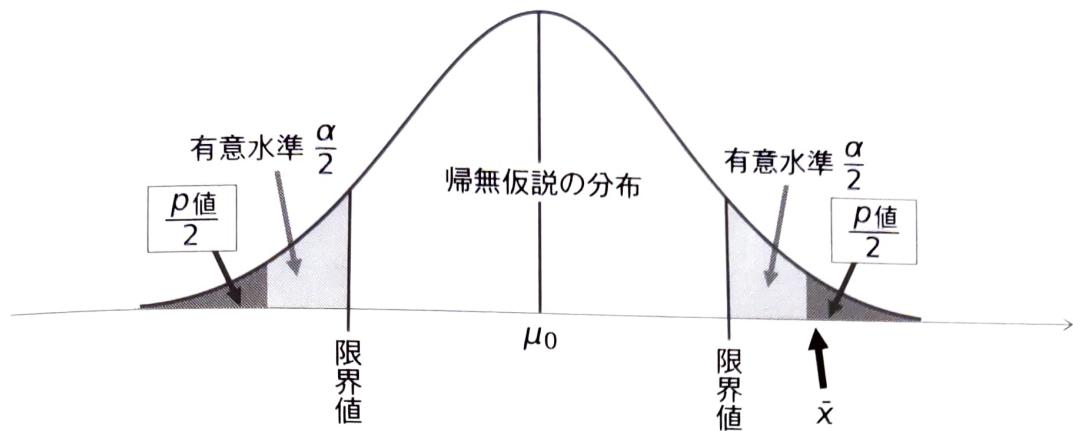
逆に、観測された標本平均 \bar{x} が $\mu_0 + 1.96\sigma_{\bar{x}}$ よりも小さい、または $\mu_0 - 1.96\sigma_{\bar{x}}$ よりも大きい場合には、帰無仮説が正しい下で起こるべきことが起きただけなので、帰無仮説を受容します。いいかえると、同じ母集団から抽出された（同じ標本分布に従っている）としても、その程度の小さい差ならば偶然でもよく出る（20回に1回以上は出る）と考えて、今回は μ_0 と \bar{x} との間の差は大きいとはいえないかった（ μ_0 は \bar{x} の真の値かもしれない）と判断します。

なお、帰無仮説が棄却されたという判定結果を**有意** (significant) と表現することがありますが、この言葉をあまり重く捉えないでください。重く捉えすぎると、有意でない（帰無仮説が棄却されない）場合の実験や研究自体が“無意味”であると感じてしまい、検定で有意にすることだけにこだわってしまう本末転倒な状況になってしまいます（第9章の多重性の発生原因でもあります）。帰無仮説が受容される（有意でない）とか棄却される（有意である）とかは、あくまで実験結果と帰無仮説との乖離の程度が、偶然の範囲内といつてもよいのかどうかの“目安”を示しているだけなのです。

(ステップ④の) 補足：p 値を使った判定

以上が、母平均の検定を事例とした仮説検定の手順ですが、ソフトウェアを使用して検定を実施すると**p 値** (p-value) という指標が計算されます。ちなみに**有意確率**とか**確率値**と訳されることもありますが **p 値**（大文字ではなく小文字）のまま用いましょう。昔はステップ④で解説した方法（限界値と検定統計量との比較）で帰無仮説の是非を判定することが一般的でしたが、近年はこの **p 値** を用了判定方法が用いられることが多くなってきました。

p 値 とは、図 6.9 の両裾の一番濃色の部分のように、帰無仮説の下で、実験

図 6.9 p 値注: p 値も両側の確率を合わせた値とする (両側検定)

結果（検定統計量 \bar{x} ）よりも極端な（分布の外側の）値が観測される確率のことです。限界値を求めるために事前に決めておく確率が有意水準 α であるのに對して、事後に判明する確率を示したのが p 値ということになります。ソフトウェアが一般的でなかった時代は、この確率を求めるることは難しかったため概念のみが存在していました。

さて、この p 値、帰無仮説を棄却することができるもっとも低い（厳しい）有意水準ともいえます。よって、先ほどのステップ④（検定統計量 \bar{x} が限界値よりも小さければ帰無仮説を棄却）の代わりに、 p 値が有意水準 α よりも小さければ帰無仮説を棄却すれば良いのです ($p < 0.05$ などと記載します)。また、有意水準 α の値にこだわらないのであれば、 p 値だけを表記しておけば良いでしょう ($\alpha = 0.05$ では有意ではないが、 $\alpha = 0.06$ ならば有意であるという判定に疑問を感じるようならば、切りの良い有意水準にこだわる必要はないでしょう)。

例題

ある日、A君がキャンパス内で、見たことのないテントウムシを数匹発見しました。しかし、B君に見せたところ、「背の模様が同じなので○○テントウ」という種だ」といわれました。果たして、A君が見つけたテントウムシは○○テントウなのでしょうか？ それとも新しい種を発見したのでしょうか？ なお、○○であることがわかっているとします。

解:

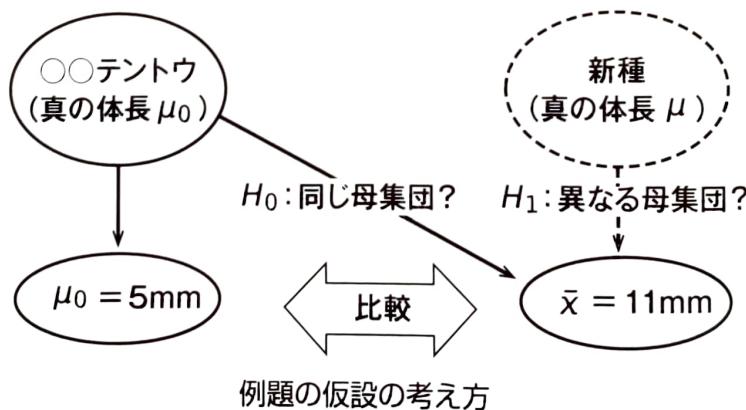
A君が見つけたテントウムシが新種であることを仮説検定で判定するために A君が見つけたテントウムシの平均体長 11 mm と○○テントウの体長 5 mm との差は、採集したテントウムシの平均体長 11 mm

“6 mm”が、統計的に意味を持つ（偶然では起きにくいほど大きい差であると判断できる）必要があります。そこで、本節で学んだ「母平均の差の検定」を使います。有意水準は一般的な両側5%としましょう。

手順①：仮説の設定

まず、主張したいこととは逆の内容の帰無仮説を立てます。この例題では、A君の立場に立てば「自分（A君）が見つけたテントウムシは新種である」ことが主張したい内容（つまり対立仮説）でしょうから、帰無仮説はその逆の「自分が見つけたテントウムシは既知の○○テントウである」という、A君にとってはつまらない内容になります。いいかえれば、この帰無仮説は、「本来、両者の体長に差はないものの、標本でならば今回の6 mm ぐらいの体長差は偶然でも現れることがあるだろう」という内容です。

$$\left\{ \begin{array}{l} \text{帰無仮説 } H_0 : \mu(\bar{x} = 11 \text{ mm}) = \mu_0(5 \text{ mm}) \rightarrow \text{新種ではない (同じ母集団から抽出)} \\ \text{対立仮説 } H_1 : \mu(\bar{x} = 11 \text{ mm}) \neq \mu_0(5 \text{ mm}) \rightarrow \text{新種である (異なる母集団から抽出)} \end{array} \right.$$



手順③（手順②検定統計量計算は不要）：確率の計算

平均体長が11 mmよりも大きい○○テントウが採集される確率（ p 値）を求めたいのですが難しいので、その代わりに帰無仮説の分布で11 mmがどのあたりに位置するのかを考えることにします。

手順④：仮説の判定

あらかじめ決めておいた有意水準（今回は両側で5%）に対応する限界値を求め、それを11 mmと比較することで、帰無仮説の是非を判定します。