第 5 章

検定

5.1　仮説検定の考え方

ある製薬メーカーが新しい睡眠薬を開発.
「従来の製品による平均睡眠増加時間は 1 時間だったが，この新製品はそれよりも効果が大きい」と言っている.
この睡眠薬を購入しようと考えている卸売業者は，メーカーの主張の真偽を調べるために，無作為に抽出した 10 人にこの睡眠薬を服用してもらい，睡眠増加時間を調べた.
その結果，睡眠増加時間の平均は 1.8 時間であった．睡眠増加時間は 分散 1 の正規分布に従うことがわかっているとき，この結果から卸売業者はメーカーの主張を正しいと判断することができるか？

統計学的には，２つの可能性が考えられます.

可能性１　平均は同じであるが，たまたま高めの標本のみ抽出された（図 5.1 左).
可能性２　平均が異なるので，高めの標本が揃った（図 5.1 右).

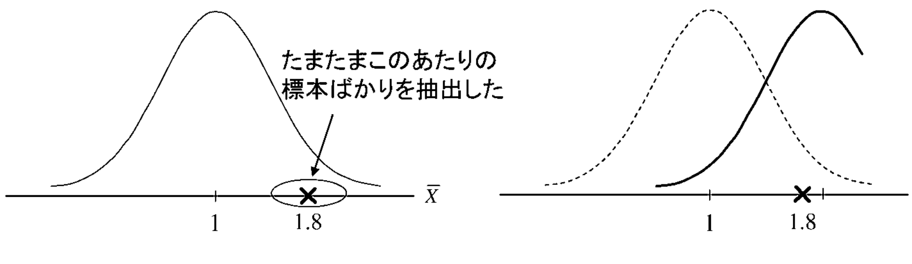

図 5.1

得られた標本平均が，平均が 1 であるとした場合に**非常に低い確率でしか得られない**ような値であるならば，「そもそも平均が異なるのでこのような標本が得られた」と考えようとするでしょう．反対に，平均が 1 である場合にも**十分高い確率で得られる**ようなもので

あるならば，「平均が同じであるからこそこのような標本が得られたのでは？」と疑うでしょう．

「非常に低い確率でしか得られない」とか「十分高い確率で得られる」と言うとき，**基準**を明確にする必要があります．たとえば，平均 1 がであるとき 95% の確率で得られる範囲内の値を「高い確率で得られる値」と考え，残りの 5% の確率でしか得られない範囲の値を「めったに起こらない値」と考えることができるでしょう．
ところで，「95% の確率で得られる値の範囲」と「5% の確率でしか得られない範囲」の決め方は，大きく次の二通りが考えられます（図 5.2)．

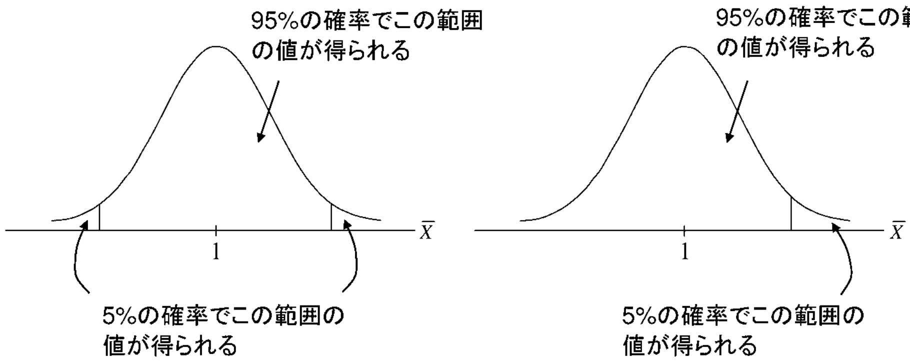

図 5.2

このケースでは，平均が 1 より高い可能性を疑っているわけですので，**1 より大きな**値が出る可能性を疑っていることになります．したがって，5.2 右図のような範囲を考えるのが妥当でしょう．

平均 1，分散 1 の正規分布からサイズ 10 の標本抽出が行われるとき，標本平均は平均 1，分散 1/10 の正規分布に従います．

$$\bar{X} \sim N(1, 1/\sqrt{10})$$

これで $\bar{X}$ の確率分布の形が特定されたので，理論上は右側 5% 点を求めることができます．しかし，第 2 章で説明したように，標準正規分布以外の「○○ % 点」を直接求めるのは極めて困難です．したがって，区間推定のときと同様に，$\bar{X}$ を標準化して標準正規分布に「矯正」することで分布を簡単化します．そうすることで，標準正規分布の確率表を使って議論することが可能になります（図 5.3)．

$\bar{X}$ を標準化してできる変数を z と名づけましょう．

$$z = \frac{\bar{X} - 1}{1/\sqrt{10}} \sim N(0, 1)$$

標準正規分布の右側 95% 点は，標準正規分布の確率表から 1.645 であると分かります．標本抽出から得られた標本平均 1.8 は，標準化すれば $(1.8 - 1)/(1/\sqrt{10}) \simeq 2.53$ になりま

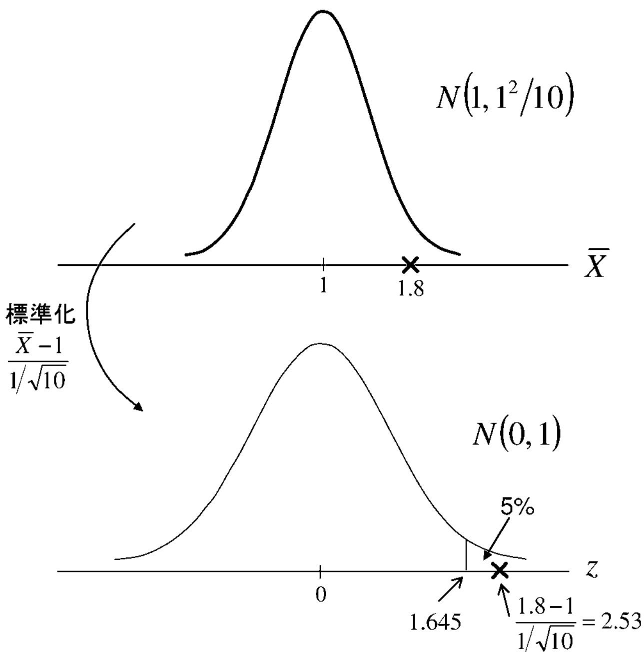

図 5.3

す．これは**棄却域**に入っています．すなわち，今回の標本抽出による標本平均 1.8 は，平均が以前と変わらず 1 のままであるとしたら 5% の確率でしか起こりえないような値ということになります．以上より，右側 5% という判断基準の下では，「新しい睡眠薬の平均睡眠増加時間は 1 と変わらない」という仮説は標本データと矛盾する，と判断されることになるのです．

このとき，「**帰無仮説**$\mu = 1$ は，**対立仮説**$\mu > 1$ に対して**有意水準**5% で棄却された」と言います．

また，得られた標本平均を標準化して新たな変数 $z = (\bar{X} - \mu)/(\sigma/\sqrt{n})$ をつくり，この z が棄却域に入るかどうかで帰無仮説の妥当性を判断しました．この $z = (\bar{X} - \mu)/(\sigma/\sqrt{n})$ を，「帰無仮説の検定に使われる統計量」という意味で**検定統計量**と言います．

仮説検定の手順

1. 帰無仮説（通常 H_0 と表記）と対立仮説（H_1）を決める．
2. 標本抽出を行い，標本平均を計算する．
3. 有意水準を決め（通常は 10%，5%，1% のいずれか），棄却域を決める．
4. 標本平均から検定統計量（後述）を計算し，棄却域に落ちるかどうかを見る．
   (a) 棄却域に落ちる → 帰無仮説の妥当性はデータから指示されない（対立仮説が支持される）
   (b) 棄却域に落ちない → 帰無仮説はデータと矛盾しない（帰無仮説が支持されたわけではない）

5.2　仮説検定の例

5.2.1　宮川 p.261（一部改題）

ある教育学者の「日本の大学生の知能指数はたかだか 110 である」という主張に対して，日本の大学生の I.Q. はもっと高いことを主張したい．そこで 150 人の大学生を無作為に選んで調査したところ，I.Q. の平均値は 111.2 であり，標本標準偏差は 7.2 であった．

標本数 n	150
標本平均 $\bar{X}$	111.2
$\hat{\sigma}$	7.2
母集団の分布	？ (正規分布と考える)

帰無仮説	$\mu = 110$
対立仮説	$\mu > 110$

標本データは，$\mu = 110$ であっても十分高い確率で発生し得るものでしょうか？それとも，$\mu = 110$ の下ではとうてい発生し得ないもの，したがって $\mu = 110$ を棄却するのに十分なものなのでしょうか？後者であれば，帰無仮説を棄却して対立仮説 $\mu > 110$ の妥当性を主張することができます．標本データが帰無仮説と整合的かどうかをチェックしてみましょう．

まず，棄却域を設定します．ここでは，5% としましょう．
帰無仮説が正しければ，$\bar{X}$ は平均 110，分散 $\sigma^2/150$ の正規分布に従います．したがって，$z = (\bar{X} - 110)/(\sigma/\sqrt{150})$ は標準正規分布に従います．しかし，ここでは標準偏差 σ が未知であるので，代わりに標本標準偏差 $\hat{\sigma}$ を利用します．標本数が大きいため，$(\bar{X} - 110)/(\hat{\sigma}/\sqrt{150})$ も標準正規分布に従うと考えて差し支えないでしょう．
帰無仮説より高いI.Q. が得られる可能性を疑っていますので，右側 5% 点を棄却域に設定します．標準正規分布の右側 5% 点は 1.645 です．得られた標本平均から計算した検定統計量 $(\bar{X} - 110)/(\hat{\sigma}/\sqrt{150}) = (111.2 - 110)/(7.2/\sqrt{150}) = 2.04$ は棄却域に落ちますので，帰無仮説は有意水準 5% で棄却されます．すなわち，「日本の大学生の I.Q. はせいぜい 110 である」という仮説は標本データとは矛盾するという結論になります．

5.2.2　宮川 p.259（一部改題）：両側検定

ある工場で直径 1 インチの軸棒を標準偏差 0.03 インチの管理水準で製造している．ある日の製造品の中から 10 本をとって直径を測定したところ，平均値が 0.978 インチであった．品質管理上問題なしと考えてよいであろうか．

平均 1 インチの母集団からの標本抽出で 0.978 インチという標本平均が得られたとき，それはこれらの標本抽出が異なる母集団からのものである（＝製造ラインに何らかの変化が起こった）証拠なのでしょうか．それとも，同じ母集団からの標本抽出であっても，このくらいの値が出現する可能性は十分あるのでしょうか．統計理論の観点から検討してみましょう．

標本数 n	10
標本平均 $\bar{X}$	0.978
σ	0.03
母集団の分布	？ (正規分布と考える)

帰無仮説	$\mu = 1$ （製造ラインに変化は起こっていない）
対立仮説	$\mu \neq 1$ （製造ラインに変化が起こった）

ここでは対立仮説を $\mu < 1$ ではなく，$\mu \neq 1$（すなわち $\mu < 1$ あるいは $1 < \mu$）としました．これは，製造ラインに変化が起こった場合，軸棒の直径が 1 インチより小さくなるのみでなく，大きくなる可能性もあるからです．
この対立仮説の設定の仕方に伴って，棄却域の設定も図 5.2 の左側のようになります．すなわち，真の平均が 1 より大きくなっていればより右側の値が得られるでしょうし，1 より小さくなっていればより左側の値が得られるでしょう．したがって，左右両端の値が出るほど，帰無仮説の妥当性が疑わしくなるわけです．
このように，分布の両側に棄却域を設定する検定を**両側検定**と言います．これに対して，前例のように片側のみに棄却域を設定する検定を**片側検定**と言います．

さて，帰無仮説が正しければ，$z = (\bar{X} - \mu)/(\sigma/\sqrt{n})$ は標準正規分布します．したがって，-1.96 以下および 1.96 以上の値は 5% の確率でしか発生し得ません．得られた標本平均 0.978 から z を計算してみると，

$$z = \frac{0.978 - 1}{0.03/\sqrt{10}} = -2.319$$

と，棄却域に落ちます．したがって，製造ラインに異常はないという帰無仮説は，有意水準 5% で棄却されることになります．

5.2.3 比率の検定：蓑谷 p.178

あるメーカーは，まったく同一仕様の商品を，色のみ変えて赤と黄の２種類売り出そうとしており，消費者がどちらの色にも同じ選好を示すかどうか調べたい．100 人の消費者を無作為に抽出して調査したところ，64 人が黄色のほうを好み，残り 36 人は赤色のほうを好んだ．この調査結果は，２種類の色の間で選好に差があることを表しているか．有意水

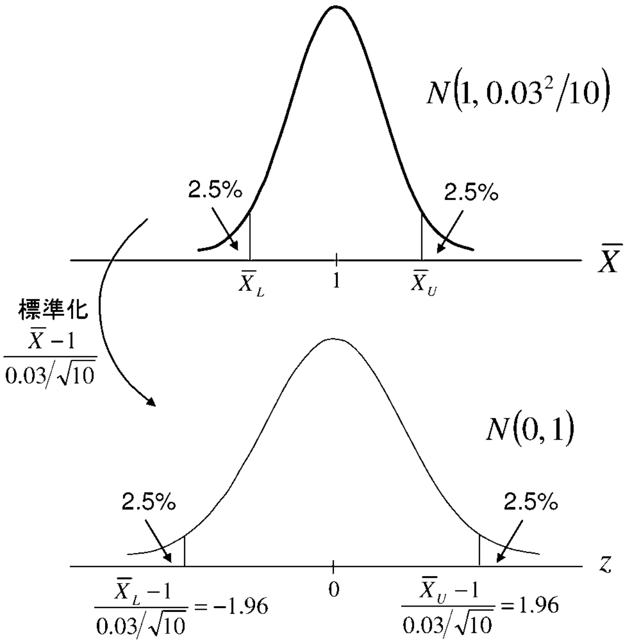

図 5.4

準 5% で検定せよ.

これは，「黄色のほうを好む人の比率が 0.5 かどうか」という，比率を問題としています．したがって，まずは２項変数を定義しましょう．

$$X = \begin{cases} 1 & (\text{黄色を好む}) \\ 0 & (\text{赤を好む}) \end{cases}$$

文字通り「全ての人」（被験者 100 人という意味ではありません）の X を調べ，その平均（$\mu = (1/N)\sum_{i=1}^{N} X_i$）をとれば，それはまさに黄色を好む人の「比率」になっています．しかし，私達は全ての人について X を調べるわけにはいきませんので，無作為に選んだ 100 人についての情報を頼りに μ についての推論を行っていくわけです．

帰無仮説は「$\mu = 0.5$」とします．ここでは，人々の好みがどちらに偏っていても企業の意思決定にとっては重要ですので，対立仮説は $\mu \neq 0.5$ とし，黄色だけでなく赤に偏っている可能性も残しておきましょう．これに伴って，検定方法は両側検定になります．

２項変数は平均 μ，分散 $\mu(1-\mu)$ の「２項分布」と呼ばれる分布に従うことがわかっています．したがって，この母集団からの標本平均 $\bar{X}$ は，標本サイズが大きいときには平均 μ，分散 $\mu(1-\mu)/n$ の**正規分布**で近似できます．さらに $\bar{X}$ を標準化した $z = (\bar{X}-\mu)/\sqrt{(\mu(1-\mu)/n)}$ は標準正規分布で近似できます．

いま，仮に帰無仮説が正しいとすれば，$z = (\bar{X}-0.5)/\sqrt{(0.5 \times 0.5/100)}$ が標準正規分布に従うことになります．標準正規分布の両側 5% 点は-1.96 および 1.96 ですので，棄却域は $z < -1.96$ および $1.96 < z$ になります．ここで，調査によって得られた標本平均

0.64 を標準化してみると，

$$z = \frac{0.64 - 0.5}{\sqrt{0.5 \times 0.5/100}} = 2.8$$

となり，ゆっくり棄却域に落ちます．したがって，消費者の色に対する選好に差はないという仮説は，有意水準 5% で棄却されることになります．また，標本抽出から得られた値がより右側のものであることより，母集団の平均は 0.5 より**大きい**可能性が疑われます．

5.3　平均の差の検定

今，同じ小学２年生の４月生まれの子と３月生まれの子をそれぞれ 50 人ずつ集め，50 メートル走のタイムを計測したとしましょう．４月生まれの子の平均タイムは 10 秒 76 で，３月生まれの子のそれは 10 秒 85 であったとします．この結果をもって，「４月生まれの子のほうが足が速い」と言えるでしょうか？必ずしもそうとは言えないでしょう．母集団の平均は同じであるにもかかわらず，４月生まれの子についてはたまたま足の速い子ばかりを，３月生まれの子についてはたまたま遅い子ばかりを選んでしまった可能性もぬぐえません（図 5.5 右）．このとき，標本データからふたつの母集団の平均が異なるかど

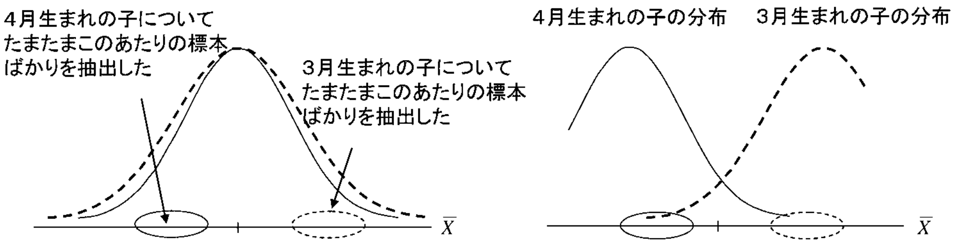

図 5.5

うかを統計的に検定することはできないでしょうか？

標本	４月生まれ	３月生まれ
標本数	n_1	n_2
標本平均	$\bar{X}_1$	$\bar{X}_2$
標本標準偏差	$\hat{\sigma}_1$	$\hat{\sigma}_2$
母集団	４月生まれ	３月生まれ
平均	μ_1	μ_2
標準偏差	σ_1	σ_2
分布	正規分布	正規分布

「平均が等しい」という仮説は，$\mu_1 = \mu_2$ あるいは $\mu_1 - \mu_2 = 0$ と表現できます．一方，「平均が異なる」という仮説は，$\mu_1 \neq \mu_2$ あるいは $\mu_1 - \mu_2 \neq 0$ と表現できます．

ところで，$\mu_1 - \mu_2$ の推定値としてまずは $\bar{X}_1 - \bar{X}_2$ が思い浮かぶでしょう．そして，標本平均の差 $\bar{X}_1 - \bar{X}_2$ の確率分布が扱いやすいものであれば，これまでと同様の検定が可能となることは容易に想像できるでしょう．

■期待値

$$\begin{aligned} E(\bar{X}_1 - \bar{X}_2) &= E(\bar{X}_1) - E(\bar{X}_2) \\ &= \mu_1 - \mu_2 \end{aligned}$$

すなわち，$\underline{\bar{X}_1 - \bar{X}_2 \text{ は } \mu_1 - \mu_2 \text{ を中心に確率分布する}}$ のです．

■分散（標準偏差）

$$\begin{aligned} V(\bar{X}_1 - \bar{X}_2) &= V(\bar{X}_1) + V(\bar{X}_2) \\ &= \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \end{aligned}$$

すなわち，$\bar{X}_1 - \bar{X}_2$ は標本抽出のたびごとに異なった値をとる確率変数ですが，その分散は $\bar{X}_1$ の分散と $\bar{X}_2$ の分散の「和」になります．「差」でないことに注意してください．

■分布　n_1 と n_2 が 30 より大きいとき，$\bar{X}_1 - \bar{X}_2$ は近似的に正規分布することが知られています．

以上をまとめると，次のようになります．

$$\bar{X}_1 - \bar{X}_2 \simeq N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

仮説 $\mu_1 - \mu_2 = 0$ が正しければ，標本抽出によって得られる $\bar{X}_1 - \bar{X}_2$ はゼロ近くの値になる可能性が高いでしょう．一方，仮説 $\mu_1 - \mu_2 = 0$ が真でなければ（すなわち $\mu_1 - \mu_2 \neq 0$ が正しければ)，ゼロから離れた値が得られやすくなるでしょう．

5.4　帰無仮説の設定

本章の最初の例では，「新薬は効果がない」を帰無仮説に，「新薬は効果がある」を対立仮説にしました．なぜ，「効果がある」を帰無仮説にしなかったのでしょうか．ここでは，「何を帰無仮説に設定するか」を考える際の基準について議論していきます．

２種類の誤り

たとえ新薬の本当の平均が 1 時間のままであったとしても，棄却域より右側の標本平均を得る確率は 5% あります．したがって，帰無仮説が正しいにもかかわらず帰無仮説を棄却してしまうという，**第１種の誤り（type Ⅰ error）**を犯してしまう確率が 5% 存在します．

同様に，たとえ新薬が本当に平均睡眠時間を延ばしているとしても，棄却域より左側の標本平均を得る確率もゼロではありません．たとえば，真の平均が 1.8 であったとすると，それでもこの仮説を採択できない確率が図 5.6 の b だけ存在します．このように，対立仮説が正しいにもかかわらず採択しないという，**第２種の誤り（type Ⅱ error）**を犯してしまう確率もゼロではないのです．

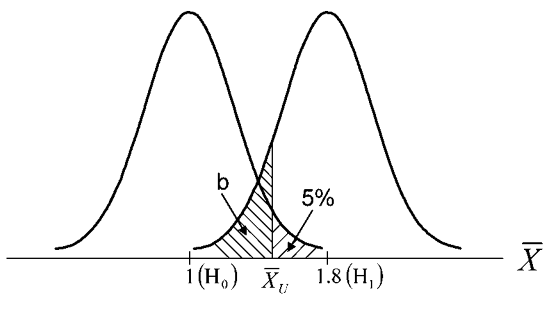

図 5.6

第 1 種の誤りを犯す確率は，有意水準を調整することで事前にコントロールできます．一方，第 2 種の誤りを犯す確率についてはどうでしょうか．図 5.7 を用いて考えてみましょう．真の平均が 1.8 ならば，第 2 種の誤りは確率 b で起こることになります．しかし，真の平均がもっと小さい 1.4 である場合には，確率は c となります．このように，真の平均が未知である以上，第 2 種の誤りを犯す確率については，一般的に事前にコントロールすることができません．

重要なのは，第 1 種の誤りは事前にコントロールできるが，第 2 種の誤りはコントロールできないということです．

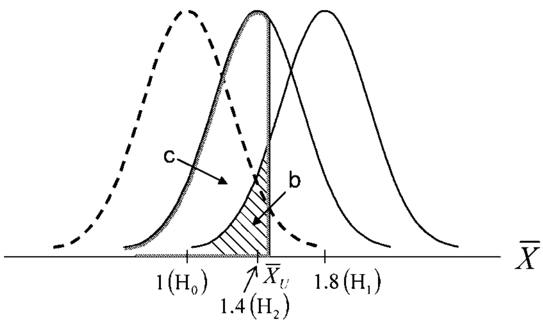

図 5.7

何を帰無仮説にするか

「新薬は効果がない」を帰無仮説にした場合，効果がないにもかかわらず「効果あり」と判定してしまう確率は事前にコントロールされますが，効果があるにもかかわらず「効果なし」と判定してしまう確率のほうはコントロールされないことになります.

一方,「新薬は効果がある」を帰無仮説にすると，効果があるにもかかわらず「効果なし」と判定してしまう確率がコントロールされ，効果がないにもかかわらず「効果あり」と判定される確率がコントロールされないことになります.

表 5.1　帰無仮説と 2 種類の誤り

帰無仮説	第 1 種の誤り（制御可能）	第 2 種の誤り（制御不能）
新薬は効果がない（図）	効果がないのに効果ありと判定	効果があるのに効果なしと判定
新薬は効果がある（図）	効果があるのに効果なしと判定	効果がないのに効果ありと判定

この新薬を小売店に出すかどうか検討している卸売業者にとって，効果がないのに効果ありと判定して顧客の信頼を失う費用と，効果があるのに効果なしと判定して利益獲得機会を逸する費用と，どちらが深刻でしょうか．通常は，前者の費用のほうをはるかに重要視するでしょう．したがって，前者の誤りをコントロールするために,「効果がない」を帰無仮説にするのがこの業者にとって最適な仮説設定ということになります.

ふたつの対立する仮説を設定したとき，それが正しいときに採択できないことが深刻な費用をもたらす仮説を帰無仮説にすることで，その深刻な費用を事前にコントロールしようとするのです.