第 8 章

仮説と検定

ある命題が確からしいかどうかを統計学的に判定することを**統計的検定 (statistical test)** または**仮説検定 (hypothetical test)** という．また，判定される命題を**仮説 (hypothesis)** という．統計的検定とは，ある仮説が妥当であるかそうでないかについて，一定の確率的根拠に基づいて判断するために行われる作業である．

8.1 ひょうたん島での仮説検定

どのような状況で仮説検定が行われるのか，その例を往年の名作「ひょっこりひょうたん島」の舞台に託して示すことにしよう．断るまでもなく，ひょっこりひょうたん島は故井上ひさし氏の手になる名作テレビドラマであるが，ここで紹介しているのはまったくのパロディであって，オリジナルとは何の関係もない[*1]．

8.1.1 大統領の陰謀の巻

【ここまでのあらすじ】ひょうたん島の住民たちの間で，ある日次のような事件が発生した．身勝手で尻が軽くて口の達者なひょうたん島大統領ドンガバチョは，トラヒゲデパートの社長トラヒゲに商品の在庫処分を頼まれた．なにせ人口が少ない島のことだから，一度売れた商品は捌(さば)けなくなってしまうのである．そこはガバチョ，「ほいほい，ワタクシの舌先三寸で不良在庫の山なんか軽く一掃してあげますよ．その代わりといっては何ですが，売り上げの半分はワタクシめにくださるでしょうな？」と，トラヒゲの渋い顔を他処に，商品のサンプルの包装紙を別のものに取り替えて，ひょうたん島小学校の前で叩き売りを始めたのである．もちろんトラヒゲだって売り上げをガバチョに渡す気なんか毛頭ない．お互いに相手を利用して一儲けしようというコンビなのだ．

はじまり

[*1] なお，小波秀雄が井上ひさしに弟子入りして文章道を教わったことがあるというのは事実である．

ドンガバチョ：　みなさあーーーん，耳寄りの話でございますよ！わが社の最新式の扇風機，ソヨリン X は，従来の製品に比べて 1 時間当たりの電気代がぐぐぐぐっと少ないのでありまーす．ここで買わなきゃ損をすること間違いなし．さあ，さあ，お集まりのほどを！

ハカセ：（「へんだなあ」と独白）ガバチョさん，だったら持ってきた製品を 10 個出してみて（ガバチョしぶしぶ出す）．さあ，みんな，試供品のソヨリン X をとりあえず 1 時間動かしてみて，エネルギー消費を測ってみてくれないかな．

トラヒゲ：　やめろやめろーい（ガバチョ「おやめなっしゃれー！」）．そんなことしたら俺のデパートの倉庫から持ってきたってことがばれちゃうじゃないかよ（と，言いかけてあわてて）っとっとっと，つまりだな，ようするに人間は正直が一番てことよ．

サンデー先生：　そうよ，トラヒゲさん．いいこと言うわねえ（トラヒゲ，でへへへと照れながら頭を掻く）．

サンデー先生と子どもたち：　♪「正直はすてきなこと」の歌．その間に時間がすぎる．

ハカセ：（1 時間後）さて，みんなのレポートをまとめてみようか．ええとダンディさんのデータだと，エネルギー消費は 1250 ガンバだったんだね．サンデー先生のは 1305 ガンバと，・・・・ふむふむ，10 人のデータを平均すると 1273 ガンバ，標準偏差は 29.5 ガンバってことだな．・・・・これくらいは暗算でこのとおりスイスイさ，ボクにとってはネ．

ドンガバチョ：　おっほっほっほ，さっすが天才ですなあ，ハカセさん．もう結論が出ましたですね．わがニューソヨリン X はやはりすんばらしい性能でがしょう？

ハカセ：　まだ何にもきまってないじゃないよ，ガバチョさん．えっと，ボクの頭の中のデータベースによると，もともとトラヒゲデパートにあったソヨリンオリジナルだと時間当たりエネルギー消費は 1286 ガンバで，その標準偏差は 35.4 だったはずだ．

トラヒゲ：　そーれ見ろよ．やっぱり少なくなってるじゃねえかよ，ハカセ．おいらの言ってることはまちがいないだろう？そうだよな．

サンデー先生：　ハカセさん，ガバチョさんだって正直なときは正直なのよ．人を疑うのはよくないわ．1286 ガンバ が 1273 ガンバに減ったってことは，やっぱりソヨリン X って少しはいいんじゃないの？

ハカセ：　いやいや，ものごとは疑ってかかることだって必要だと思うな．えっと，つまり，ガバチョさんが持ってきた 10 個の試供品はトラヒゲデパートにあったソヨリンオリジナルの包装だけ変えたんじゃないかってことは十分考えられる．それでも成績はオリジナルよりもよかったってことは，たまたままぐれで成績がよかったのかも知れない．その可能性がありうるかどうかを考えてみないとね．

ドンガバチョ，トラヒゲ：　ギョギョギョギョッ！

来週に続く

8.1.2　ハカセは仮説を検定する

■ふたつの仮説

さて，以上のシナリオからみて，一体なにを検証できればガバチョとトラヒゲの悪だくみを看破することができるのだろうか．双方の言い分をもう一度みてみよう．

ドンガバチョが主張しているのは，「ソヨリン X はオリジナルから抜き取られたものではない」ということである．つまり，ソヨリン X のデータはオリジナルよりも優れているのであるから別物だというのである．つまり次の主張だ．

> 仮説 **I**：ソヨリン **X** という商品のエネルギー消費率の平均は，「本当に」オリジナルの平均よりも小さい．

「本当に」というのは，たまたまそうなったのではなく，ソヨリン X がオリジナルとは異なる母集団に属していて，その母平均はオリジナルの母平均よりも小さいということを強調するために入れている．

なお，ここで**仮説 (hypothesis)** と言っているのは，上の主張がこれから検証されるべき命題であって，その真偽はまだニュートラルな状態にあることを意味している．

しかし，通常の統計的検定では次のように上と反対の仮説を設けることが多い．これはハカセの取っている立場である．

> 仮説 **II**：ソヨリン **X** という商品のエネルギー消費率の平均は，オリジナルの平均に等しい．

ここで注意しておかなければならないのは，ソヨリン X の 10 個のサンプルから計算される 標本平均 が，母平均，つまりオリジナルの平均と一致しているかどうかを問題にしているのではないということである（1273 ガンバと 1286 ガンバだから，実際一致していない）．標本平均というのは抽出の度に変化する確率変数なのだから，たまたま一致することはまずありえないのだ．

そうではなく，**ソヨリン X はオリジナルから抽出されたものであって，本来の平均はオリジナルのものと一致するといっても確率的にはかまわない**ということを，「等しい」と表現しているのである．

■帰無仮説と対立仮説

上の仮説のうち，仮説 II のようなものは，それが**棄却** (= 否定) されたときに意味をもつことになる．つまり，棄却されれば「等しくない」ということになるのだが，それは実際に「ちがう」という積極的な意味を持っている．逆に仮説 II が採択されたら，ソヨリン

X がソヨリンオリジナルと同じ性能を持つからといって，これらが同じものであるとは言えないのだから，せいぜい「ちがうわけではない」という程度の意味しか持たせられない．

このように，棄却されたときに意味を持つような仮説を**帰無仮説 (null hypothesis)** という[*2]．**帰無仮説は棄却されるときに意味をもつ．**

一方，仮説 I のように帰無仮説と対立して，「差がある」という言明を含む仮説のことは，**対立仮説 (alternative hypothesis)** と呼ばれる．統計的検定では，帰無仮説のことを $\boldsymbol{H_0}$, 対立仮説のことを $\boldsymbol{H_1}$ と表記することがしばしばある[*3]．

■仮説を検定する

それでは統計的処理をどういうふうに使ったら，これらの仮説の当否を**検定 (test)** できるのだろうか．

この問題では，母集団のソヨリンオリジナルの母平均 μ および母標準偏差 σ は，すでにハカセの頭の中のデータベースに入っていて既知である．もし，そこから $n = 10$ 個の標本を無作為抽出したとしたら，得られる標本平均 $\bar{X}$ の期待値は μ と等しくなり，また $\bar{X}$ の分散は σ^2/n に等しくなるはずである（84 ページの式 (6.3)(6.4) 参照）．

したがって，オリジナルから抽出された大きさ 10 の標本によって得られる標本平均は正規分布 $N[\mu, \sigma^2/n]$，具体的には $N[1286, 125.3]$ に従うことになるはずである[*4]．つまり，1286 ガンバを中心として，標準偏差が 11.2 $(= \sqrt{125.3})$ ガンバの正規分布が，オリジナルからの抽出で得られる大きさ 10 の標本の標本平均のとる分布である．**その分布の中で，今回の調査で得られた 1273 ガンバという値は，明白に小さい側に外れているのだろうか．**これが問題の中心なのである．外れていればドンガバチョの勝ち，そうでなければハカセの疑いには妥当性があるし，もちろんソヨリン X を買う意味はない．

なお，このように対象が分布の左右どちらかに外れているかどうかを見るような検定を**片側検定**あるいはもっと詳しく**左側検定**，**右側検定**と言う．それとちがって分布の中心から遠いかだけを判断するような検定は**両側検定**である．

■分布の中のどの位置にあるかで検定する

さて，ここまで来れば問題の解決は容易である．正規分布 $N[1286, 11.2^2]$ における 1273 という値を標準化変換すればよい．

$$\frac{1273 - 1286}{11.2} = -1.16$$

[*2] 本によっては，比較対象の差がゼロになるような仮説を帰無仮説と説明しているものがある．これらは微妙に意味が異なるが，実際上は重なる概念である．

[*3] この辺の話はややこしくて，学習者にとっては頭が混乱するところである（著者もよく混乱する）．しかし，後に見るように，分布の中での位置関係というイメージで考えれば，それほどややこしい概念操作は必要ではないことがわかる．

[*4] 125.3 というのは $35.4^2/10$ を計算すれば出てくる．

一方，標準正規分布では，-1.28 以下の部分に入る面積は 10% ($\alpha = 0.1$) になるので，結局 1273 ガンバという値は，全体の 10% という狭い領域には入っていない．なお，図の影を付けた領域のことを**棄却域**といい，棄却域に入ることを「棄却域に落ちる」と表現する[*5]．ここではもちろん棄却域に落ちてはいない．

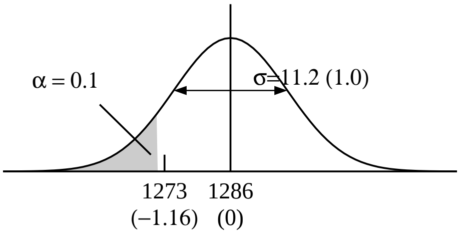


図 8.1 $N[1286, 11.2^2]$ の中での 1273 の位置：カッコ内は標準化された値．

この図から，ソヨリン X の標本平均はソヨリンオリジナルからの抽出標本のものとは**それほど外れていない**ということがわかる．10 回に一度程度おきる現象はさほど確率的にめずらしいものではないが，ドンガバチョが持ち込んだサンプルはその程度の**有意差**を見せなかったのである．このことを次のように表現する．

> **「仮説 II：ソヨリン X という商品のエネルギー消費率の平均は，オリジナルの平均に等しい」は，危険率を 0.1 と取ったときには棄却できない．**

ここで使われている**危険率 (risk)** という言葉の意味は，むしろ仮説が棄却されたときのほうがわかりやすい．つまり，「仮説 A は危険率 0.01 で棄却される」は，「仮説 A を棄てても，外れる危険 (リスク) は 0.01 である」ということなのである．逆に棄却できないときには，それ以上の確率で外れるかも知れないから，何も言わないでおこうということになる．

8.1.3 $\boldsymbol{p}$ 値

上の操作でソヨリン X の標本平均を標準化変換して得られた値 $z = -1.16$ は，どの程度の「めずらしさ」をもっているのだろうか．正規分布曲線でみると，この値は端から 0.12 の面積のところに位置することがわかる．このように，検定されるべき値が，分布の端からどれだけの面積のところに位置するかを $\boldsymbol{p}$ **値**という名前で表すことにする（図 8.2）．

[*5] 対立仮説の側から考えると，棄却域に落ちるのは，対立仮説が棄却されないことを意味していて，考える上では厄介である．

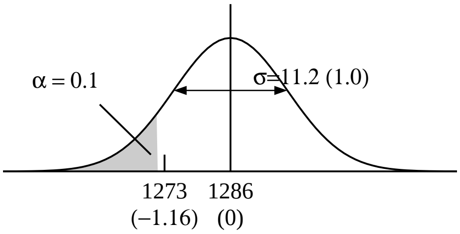


図 8.2　$z = -1.16$ のデータの p 値は 0.12 であり，それほど「めずらしい」値ではない．

仮説検定においては，p 値だけを用いて，「この観測値の p 値は 0.015 だから有意に異なると判断できる」という言い方もする．p 値は，いわばデータの「めずらしさの度合い」であるから，一般に値が小さいほど有意であるという判断が下されることになる．

8.1.4　別のシナリオ—贋物を検定する

さて，ひょうたん島でのトラヒゲとドンガバチョの悪だくみは，オリジナルそのものを別の新しい製品であるかのように見せかけようということであった．その意図はハカセによる統計的検定のおかげで見事に見抜かれて，不良在庫を売りつけられる事態は免れたのである．

ところが，世の中いくらでも悪いやつはいるものであるし，悪いやつらには必ず悪事のチャンスが到来するものである．わがひょうたん島を根城にしている海賊たち，ガラクータ，ヤッホー，トウヘンボク，それにキッド坊やの 4 人は，ある日孤島の洞窟からニセ金貨の詰まった箱を発見した．これがなんと！ひょうたん島で発行されている 1000 ガバス金貨とそっくりだったのだ．しかもニセ金貨は全部で 25 枚，25000 ガバスといえばトラヒゲデパートの年間売り上げを上回る大金である．

これを海賊たちが見逃すはずがない．さっそくニセ金貨を両替するためにトラヒゲ銀行ひょうたん島本店に持ち込んだ．

しかし，さすがはかつて自分も海賊だった頭取のトラヒゲ — その昔に海賊大学校を落第したことでガラクータたちにひそかにコンプレックスを抱いているのだが — は，海賊たちが突然大金を持ち込んだことに不審をいだいて，持ち込まれた金貨を鑑定してもらうためにハカセを呼んだ．

もちろんハカセの頭脳のデータベースには，ひょうたん島の 1000 ガバス金貨の重量の平均値と標準偏差の数値なんかしっかり入っている．二つ返事で承知すると，持ち込まれた 25 枚の金貨を秤で精密に量ることにした．

ここで再び物語のシナリオが再開される．今度の主な登場人物は銀行頭取のトラヒゲ，海賊の首領ガラクータと子分のトウヘンボクとヤッホー，キャプテンキッドの息子キッド

坊や[*6]，そしていうまでもなく天才の名をほしいままにしてきたハカセである．

またはじまり

ガラクータ　（丁寧で上品な女言葉ふうの，いかにもインテリヤクザといった口の利き方がこの海賊の首領の特徴である）ハカセさん，私たちが持ってきたこの金貨はニセモノじゃないかって，疑っておいでのようネ？

ハカセ　もちろん疑ってるヨ（海賊たち「ナニッ！」と気色ばむ）．おっとそんなに焦らないで，海賊さんたち．どんなことでもうたぐってかかるというのが科学者というものなんだからね．ボクとしては，データに基づいて客観的に判定するだけだから，海賊さんたちが本当にホントのことを言っているんだったら，何も心配することはないと思うよ (海賊たち，不安そうに目を見合わせる．)．

トラヒゲ　やいハカセ，オレ様はだな，銀行頭取としてだな，ハカセにこの金貨がニセモノかどうかを鑑定してもらいたいんだよ．トラヒゲ銀行が仮にも損をしないようにやってくれなくちゃ困るわけだ．だから頼むからサァ，海賊の野郎どもに遠慮なんかしないでやってくれヨォ！

ハカセ　誤解がないように言っておきたいんだけど，ボクがこれからやろうとしているのはとっても簡単なことなんだ．つまり，海賊さんたちが持ってきた金貨が本物であるという主張，えっとむつかしくいえば仮説というんだけど，その仮説を否定しても大丈夫かどうかという点を，データを使って確かめようというのさ．

ヤッホー　（自分の頭の上でくるくる手を回しながら）ひゃー，ハカセの言ってるのを聞いていたら，頭の中がぐるぐる回って，そんでもってなんかがパチンとはじけたみたいだなあ，おいトウヘンボク，おいらの頭どうなってるか見てくれよ．

トウヘンボク　どれどれ，わあ，ヤッホーの髪の毛のぐるぐるはちょっとおかしいあるネ．右に巻いたり左に巻いたり，わたし，こんなのは昔，動物園のお猿さんの頭で見たことあるヨ（ヤッホー怒ってトウヘンボクをどつく）．あ，私も分らないあるから勘弁してヨ！

ガラクータ　静かになさい！ハカセさんの言っていることがあなたたちに分るはずなんかないじゃないの．ハカセさんはネ，つまり，あなたたちのお頭（つむ）では死んでもわからないぐらいむつかしいことを言ってるのよ．ワタシにだって分らないんだから（一同ずっこける）・・・・でもハカセさん，あなたのおっしゃることってすごくなんだか，えっと，その，あのォ，とっても回りくどーく聞こえるんだけど？

ハカセ　さすがガラクータさん，いいとこに気がついたね．その通り確かに回りくどいんだ．そもそもボクにできることは限られていて，秤で重さを精密に調べるぐらいの

[*6] 伝説の海賊であるキャプテンキッドの息子が現代に登場するのはおかしいと思うかもしれないが，キャプテンの残した遺産の箱の中から，数百年の眠りを覚まして飛び出してきたのである．

ことしかできない．そのデータから，「『海賊さんたちの持ってきた金貨が本物である』ということを否定できるかどうか」だけを判断できる．否定できれば海賊さんたちの言っていることは多分ウソだろうということになるし，そうでなければ，ウソだとは言えない — 本当だとも言えないけどね．

一同　ふーん．おいら分んない．へえー，そうなのか，なるほどォ．分ったふりなんかするんじゃないよ！（などと口々に騒ぐ）

ハカセ　ちょっと失礼するよ（天秤で 25 枚の金貨の秤量を始める．♪トラヒゲと海賊「お金を量ってどうなるの？」)．さあ，できた．25 個の金貨の重さはこんな結果になった．

32.97, 36.37, 35.24, 36.03, 34.84, 33.63, 37.94, 33.48, 34.09, 33.74, 34.53, 36.86, 31.79, 35.61, 34.14, 34.51, 35.13, 32.83, 34.89, 32.19, 36.67, 36.01, 37.04, 35.1, 33.73 (単位はポンズ)

ハカセ　一方で，ボクのデータベースによると，ひょうたん島の 1000 ガバス金貨の重さは，平均が 35.03 ポンズ，標準偏差が 0.925 ポンズだ．これでデータは出揃ったぞ．

一同　（顔を見合わせて）さあ，どうなるんだろう？

ハカセ　さて，平均値の方をまず検定してみよう．おっと！これは海賊さんたちの言うことは否定できないかも知れないぞ．ちゃんと計算すると・・・データの平均はええと，34.77 ポンズ．なるほど，これだとこの金貨が本物だって話を否定することはできないな．

海賊たち　(ヤッホー，トウヘンボク) わーい，金貨は本物だってよオ，ハカセが言ってるぞーい．（ガラクータ）ハカセさんってやっぱり天才だわねェ．

トラヒゲ　（やや憮然として）なんてこった，金貨は本物かも知れないってエのかよ，ハカセ？おいらにゃどうもうさん臭い気がして仕方ないんだけどヨオ．

ハカセ　さっき言った通りさ．本物であるという海賊さんたちの話を否定するデータにはなってないんだよ．・・・　ふむふむ，もう少し別の角度からデータを検定するとどうなるんだ？

この後の原稿紛失

8.1.5　平均値の一致を検証する

ハカセがニセ金貨の重量の平均値を得たときに検定しようとした仮説は次の通りである．これは海賊たちの立場に立った主張になっていることに注意してほしい．それを棄却できるかどうかを判断するのが，検定の目的なのである．

> 仮説：持ってきた **25** 枚の金貨の重量の「本来の」平均値は本物の金貨の平均値と等しい．

ここで「本来の」平均値としているのは，25 枚の秤量で得られた標本平均のことではない．海賊たちとしては，持ち込んだ金貨 25 枚が本物の金貨を母集団として抽出されたのだと主張しているのである．この構図は，ソヨリン X のときとはちょうど反対であることに注意しよう．

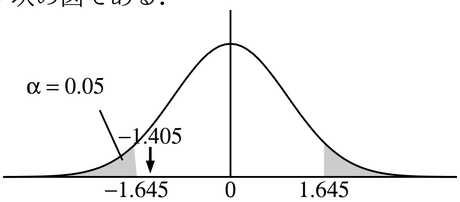
そして，実際に得られた標本平均 $\overline{X}$ はというと，

$$\overline{X} = \frac{32.97 + 36.37 + 35.24 + \cdots + 33.73}{25} = 34.77$$

ソヨリン X の議論の時と同じく，海賊の金貨がひょうたん島の金貨から抽出されたものであるならば，その平均 $\overline{X}$ は分散 $s^2 = \sigma^2/25$(標準誤差 $s = \sigma/\sqrt{n} = 0.185$) をもち，$\overline{X}$ の期待値（「本来」の平均）35.03 にピークを持つ正規分布をなすはずである．これらを使って，得られた平均値 34.77 を標準化変換してみよう．

$$Z = \frac{34.77 - 35.03}{0.185} = -1.405$$

Z の位置を表したのが次の図である．



海賊の金貨の平均値 Z

図を見れば一目瞭然，海賊の持ってきた金貨の質量の平均値は，本物の金貨を 25 個持ってくればあり得たであろう分布の中に入っていて[*7]，これだけの証拠では，**本物ではないということはできない**．このことを次のように表現する．

---

*7　図の中で，この平均値はかなりきわどいところにあって，それを暗算で判定したハカセの頭脳はさすがに大したものなのである．

上の仮説は，危険率 0.1 で棄却されない．

なお，この場合の棄却域，つまり図の斜線をほどこした部分は，両側にとられていることに注意してほしい．これは検定の目的が本物と異なっているかどうかを問題にしているからである．このように，両側に棄却域をとる検定を**両側検定**と呼ばれる．

一方，ソヨリン X の検定のように，サンプルの値が比較対象よりも優れて（劣って）いるかどうかを検定する場合には片側に棄却域がとられるので，**片側検定**という．

8.1.6　分散の一致を検証する——χ^2 検定

さてハカセがつぶやいた「もう少し別の角度」とは何だろうか．失われた原稿に書かれてあったシナリオを考えてみよう．

ハカセは 25 個のデータを使って平均値を計算し，それと本物の平均値および分散を利用して検定を行った．しかし，25 個のデータから得られる量は平均だけではない．分散もそうである．あるいはもっと別の量だって計算できる．

さて，分散とその類似の量がどのような分布をなすかということは，すでに第 6 章 6.5 節で詳しく取り扱った．そこではまず，次のことが示されている (p.93)．

【定理 I】母集団は平均値 μ，分散 σ^2 をもつ正規分布 $N[\mu, \sigma^2]$ をしているとする．そこから n 個の無作為抽出を行ったとき，

$$Z = \frac{1}{\sigma^2}\left((X_1 - \mu)^2 + (X_2 - \mu)^2 + \cdots + (X_n - \mu)^2\right) \tag{8.1}$$

なる Z は，自由度 n の χ^2 分布 に従う．

また，標本平均 $\overline{X}$，標本分散 s^2 を使って次のような定理も述べられている (p.93)．

【定理 II】母集団は平均値 μ，分散 σ^2 をもつ正規分布 $N[\mu, \sigma^2]$ をしているとする．そこから n 個の無作為抽出を行ったとき，

$$Z = \frac{1}{\sigma^2}\left((X_1 - \overline{X})^2 + (X_2 - \overline{X})^2 + \cdots + (X_n - \overline{X})^2\right) = \frac{ns^2}{\sigma^2} \tag{8.2}$$

なる Z は，自由度 $n-1$ の χ^2 分布 に従う．ただし，ここで s^2 は式 (6.2) で表される標本分散である．

■母平均と母分散を使って検定する

最初に式 (8.1) を使った検定を試みることにしよう．それにはこの式に現れる Z を，データを使って求める必要がある．ハカセのデータベースによれば，母集団の平均 μ は 35.03 ポンズ，標準偏差 σ は 0.925 ポンズである．なお，母集団といっても，海賊がもっ

てきた 25 枚の金貨がそこから抽出されたわけではないが，仮説としては「25 枚の金貨は本物である」として検定するのだから，ここでは仮にニセ金貨の母集団を本物の 1000 ガバス金貨としているのである．また金貨の質量の数値は，それぞれ 32.97, 36.37, ... ，33.73 となっているから（つまり $X_1 = 32.97, X_2 = 36.37, ...$），$Z$ は次のようにして求めることができる．

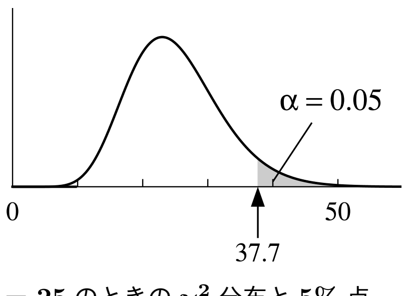
$$Z = \frac{1}{0.925^2}\left((32.97 - 35.03)^2 + (36.37 - 35.03)^2 + \cdots + (33.73 - 35.03)^2\right) = 70.49$$

この種の計算は面倒なので，表計算ソフトをうまく使うとよい．

定理 I によれば，Z は自由度が 25 の χ^2 分布に従う．つまり，ひょうたん島の本物の金貨から 25 個の抽出を何度も行って，上の手順で Z を求めると，自由度 25 の χ^2 分布に従ってあちこちに値が散らばるわけだ．さて問題は，**海賊の持ってきた金貨のデータから得られた 70.49 という値は，その分布の中に現れる可能性がどの程度あるのか**ということである．それを確かめるには χ^2 分布表を見ればよい．

χ^2 分布表からの抜粋

α	0.995	0.975	0.95	0.9	0.5	0.05	0.025	0.01	0.005
$\nu = 23$	9.260	11.688	13.090	14.848	22.337	35.172	38.076	41.638	44.181
$\nu = 24$	9.886	12.401	13.848	15.659	23.337	36.415	39.364	42.980	45.558
$\nu = 25$	10.520	13.120	14.611	16.473	24.336	37.652	40.646	44.314	46.928
$\nu = 26$	11.160	13.844	15.379	17.292	25.336	38.885	41.923	45.642	48.290



$\nu = 25$ のときの χ^2 分布と 5% 点

上の図と表で分るとおり，70.49 という値は 5% 点どころか 0.5% 点をとっても棄却域に落ちてしまい，この分布ではほとんど起こりえない値であることが分る．つまり，χ^2 を使った検定によって，次の結論が得られる．

海賊の持ってきた金貨が本物であるという仮説は危険率 0.5% で棄却される．

要するに，安心してニセモノと決め付けてよいということだ．

■標本分散を使って検定する

さて次に，**定理 II** のほうを使った検定を試みよう．定理 II では，25 枚の金貨の重さから得られる標本分散と標本平均を利用することができ，計算もやや簡単である．

まず標本平均は前に求めたように $\overline{X} = 34.77$ であり，標本分散は 2 乗の平均から平均の 2 乗を引けば得られて

$$s^2 = \frac{32.97^2 + 36.37^2 + \cdots + 33.73^2}{25} - 34.77^2 = 2.347$$

となる．表計算ソフトを使って計算してもよい．これから Z を求めると，

$$Z = \frac{ns^2}{\sigma^2} = \frac{25 \times 2.347}{0.925^2} = 68.57$$

が得られる．この値を今度は自由度が $n-1=24$ の χ^2 分布を使って検定すると，やはり海賊が持ってきた金貨が本物であるという仮説は非常に低い危険率で棄却される．ようするに，海賊たちが持ってきた金貨は，平均の質量こそ本物と見分けがつかないほどに似ていたのだが，作りが悪いために個々の重量がまちまちだったのである．

8.1.7　金貨のできが良過ぎても怪しい—両側検定

ここまでの設定では，海賊たちが持ってきた金貨は質が悪くて，本物に比べて質量のばらつきが大きすぎたところを検定で発見したケースになっている．しかし，悪い奴らがもっとハイテクを駆使できる連中だとしたら，逆のことが起きるかもしれない．次の例題を考えてみよう．

例題 8–1　ひょうたん島銀行に，りゅうとした身なりの紳士が現れて，12 枚の金貨の両替を依頼した．トラヒゲ社長は天性の勘で「こいつは怪しい！」と，紳士を別室で待たせている間に金貨の質量を量って，ハカセに鑑定してもらうことにした．データは次のようになっている（単位 ポンズ）．

$$35.7, 35.03, 35.11, 34.21, 35.08, 34.86, 35.13, 35.09, 34.36, 35.23, 35.24, 35.67$$

ひょうたん銀行発行の本物の金貨の平均質量は 35.03 ポンズ, 標準偏差は 0.925 ポンズである．持ち込まれた金貨は本物であると判定してよいか．

上のデータから定理 I の Z を計算してみよう．

$$Z = \frac{1}{0.925^2}((35.7-35.03)^2 + (35.03-35.03)^2 + \cdots + (35.67-35.03)^2) = 2.451$$

この結果を自由度 12 の χ^2 分布とにらみ合わせて考えることにする．表を見ると，$\alpha = 0.995, 0.005$ の点がそれぞれ 3.074, 28.300 となっている．これらは両端にそれぞれ

0.5% の面積を切り取る Z の値で，データから求めた 2.451 という値はその外にある．ただし前の問題とちがって，大きい方にではなく小さい方に外れている．

つまり，このケースでは，持ち込まれた金貨が本物から抽出されたものであるという仮説は，危険率を 1% にとった場合に棄却されるということになる．つまり贋金だと判定できる．もっとも，分布の左側に外れたということは，分布が狭い範囲にまとまり過ぎているということを意味する．ニセ金貨を作った連中はあまりにも正確に作ってしまったというわけだ．このように，χ^2-分布で**両側検定**を行うことで，意味のある検定結果を得ることができる．

8.2　その他の検定

8.2.1　小標本について平均値を検定する —— t-検定

ひょうたん島のソヨリン X 売り込み騒動，ニセ金貨両替詐欺事件では，平均値の検定が使われた（後者では分散の χ^2 検定が絶大な威力を発揮したが）．その場合，母集団の平均値や分散が精密に分っているという条件があって，検定は容易に行われた．

しかし，実は，ソヨリン X のサンプル 10 個を手にしたとき，ハカセの脳のデータベースにはオリジナルのソヨリンのデータは平均値しかなかったとしよう．つまり母分散は知られていないのである．この場合には 8.1.1 節以降のシナリオは変更を余儀なくされるが，7.3.3 節で出てきた t-分布を利用することで検定を行うことができる．

今，正規分布している母集団から抽出された大きさ n の標本があったとして，母平均 μ, 標本平均 $\overline{X}$, 標本標準偏差 s を使って与えられる T という量を次のように求めたとしよう．

$$T = \frac{\sqrt{n-1}(\overline{X}-\mu)}{s}$$

このとき，T は自由度が $n-1$ の t-分布に従う．

さて，ここではすでに論じたように，

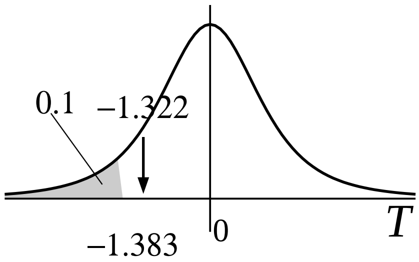
「10 個のソヨリン X の値の平均は，ソヨリンオリジナルのデータの平均と等しい」という帰無仮説が棄却されるかどうかを検定で明らかにしたいのである．そこで，与えられたデータから T を計算してみよう．

$$T = \frac{\sqrt{10-1}(1273-1286)}{29.5} = -1.322$$

例によって，この値が分布のどのへんにあるかを表から確かめよう．

α	0.1	0.05	0.025	0.01	0.005	0.0025
$\nu=8$	1.397	1.860	2.306	2.896	3.355	3.833
$\nu=9$	1.383	1.833	2.262	2.821	3.250	3.690
$\nu=10$	1.372	1.812	2.228	2.764	3.169	3.581

採用するのは自由度 $\nu=9$ のところの値である．また検定の意図は「ソヨリン X のエネルギー消費率がソヨリンオリジナルよりも小さいかどうか」の確認にあるのだから，分布で左端の面積 0.1 の領域を棄却域として設定する．



図を見れば，この場合にも上の仮説は棄却されない．つまり，有意な差があるとはいえないという結論になる．

8.2.2　分割表による検定

χ^2 分布を応用して，ある集団がもついくつかの性質が独立であるかどうかを判定することができる．その種の問題としては，たとえば次のようなものがある．

1. ある集団について，生活習慣と健康状態のアンケートを調べて，それらが独立であるか，あるいは関係があるかどうかを判断する．
2. 読書傾向と好きな色に関する調査結果から，それらの間に相関があるといえるかどうかを判断する．

これは単に集計表を見ただけでもある程度の判断はできるものである．しかし，一見して傾向がありそうなデータであっても，実際には傾向がないのに偶然そういうことがあるという程度のものなのか，本当に傾向があるのかという判断を定量的に行うことは難しい．また集計表の数字を定性的に判断しようとすると，先入観や主観的な期待によって判断に偏りが生じることがしばしばある．そこで客観的にこれらの判断を下すためには，以下に説明する**独立性の $\boldsymbol{\chi^2}$ 検定**を行うとよい．

なお，χ^2 分布による検定はきわめて応用が広く，ここで取り上げた以外にもさまざまのタイプの検定がある．詳しくは他の書物を参照してほしい．

■分割表

アンケート調査によって表 8.1 のような表を作成することがしばしばある．

このような表を**分割表 (contingency table)** [*8]という．上の分割表の個別の欄に書き込まれた 39,45,21,83, ... といった数値は，**観測度数 (observed frequency)** という．

たとえばこの分割表を見ると，音楽の好みと色の好みに関連があるかどうかを知ることができるかも知れない．たとえばクラシックの愛好者はどちらかというと青を好み，邦楽

[*8] contingency というのは依存性という意味である．この英語のほうが意味が分かりやすい．

表 8.1　分割表の例

好み	赤	青	緑	計
クラシック	39	45	21	105
邦楽ポップス	83	68	47	198
洋楽ポップス	53	51	65	169
歌謡曲	41	32	55	128
計	216	196	188	600

ポップスの愛好者は赤を好むようにも見える．しかし，このような関連が本当に存在するのか，通常のデータの揺らぎの範囲なのかは，見ただけではなかなか分らない．

このように，2 つの性質によってグループ分けされた分割表から，異なる性質同士（音楽と色の好み）に関連があるかどうかを知りたいときには，独立性の検定という手続きを行う．

■独立性の検定

ある母集団から大きさ N の標本を抽出して，性質 A, B でそれぞれ m 個，n 個ずつに分類される．その結果として次のような分割表が与えられたとしよう．

性質	$B_1,\ B_2,\ B_3,\ \ldots, B_n$	計
A_1	$x_{11}, x_{12}, x_{13}, \cdots,\ x_{1n}$	a_1
A_2	$x_{21}, x_{22}, x_{23}, \cdots,\ x_{2n}$	a_2
$\cdots$	$\cdots, \cdots, \cdots, \cdots, \cdots$	$\cdots$
A_m	$x_{m1}, x_{m2}, x_{m3}, \cdots, x_{mn}$	a_m
計	$b_1,\ b_2,\ b_3,\ \cdots,\ b_n$	N

この時，次の命題が成り立つことが知られている．すなわち，**性質 $\boldsymbol{A, B}$ が独立であるならば，次の $\boldsymbol{X}$ は，自由度が $\boldsymbol{(m-1)(n-1)}$ の $\boldsymbol{\chi^2}$ 分布に従う．**

$$
\begin{aligned}
X = &\\
&\frac{(x_{11}-a_1b_1/N)^2}{a_1b_1/N} + \frac{(x_{21}-a_2b_1/N)^2}{a_2b_1/N} + \cdots + \frac{(x_{m1}-a_mb_1/N)^2}{a_mb_1/N} + \\
&\frac{(x_{12}-a_1b_2/N)^2}{a_1b_2/N} + \frac{(x_{22}-a_2b_2/N)^2}{a_2b_2/N} + \cdots + \frac{(x_{m2}-a_mb_2/N)^2}{a_mb_2/N} + \\
&\cdots + \\
&\frac{(x_{1n}-a_1b_n/N)^2}{a_1b_n/N} + \frac{(x_{2n}-a_2b_n/N)^2}{a_2b_n/N} + \cdots + \frac{(x_{mn}-a_mb_n/N)^2}{a_mb_n/N}
\end{aligned}
\tag{8.3}
$$

式 8.3 はいかにも複雑に見えるが，実際にはきわめて単純で機械的な計算を意味している．表 8.1 について，X を計算してみよう．この場合には，$N = 600, m = 4, n = 3, a_1 = 105, a_2 = 198, \ldots, b_1 = 216, b_2 = 196, x_{11} = 39, x_{12} = 45, \cdots, x_{43} = 55$ である．従って，

$$X = \frac{(39-105\times 216/600)^2}{105\times 216/600} + \frac{(83-198\times 216/600)^2}{198\times 216/600} + \frac{(53-169\times 216/600)^2}{169\times 216/600} + \frac{(41-128\times 216/600)^2}{128\times 216/600} + \cdots + \frac{(55-128\times 188/600)^2}{128\times 188/600} = 25.9$$

として計算される X は，もし A, B が独立であれば自由度が 6 $(= (4-1)\times(3-1))$ の χ^2 分布に従うから，高い確率で分布曲線の内側に入ってくるはずである．そこで，**もしも X の値が χ^2 分布の曲線の外側の方にはみ出しているならば，A と B が独立であるという仮説は棄却される．**

ただし，「はみ出し」というのは何に対してのはみ出しであるのか，その基準を設けておかないといけない．それが**危険率**である．では，今求められている X の値, 25.9 は，「はみ出し」ているのだろうか．

図 8.3 の χ^2 分布のグラフを見ながら考えてみよう．グラフの陰をつけた部分の面積が α であり，ここでは $\alpha = 0.05$ になるときの X の値は 12.59 であることが示されている (5% 点は 12.59 である)．この値はもちろん χ^2 分布表から拾ったものだ．この値を使うのが，**危険率 5% で検定する**ということである．

この問題の分割表から得られた X の値，25.9 は，あきらかにこの 5% 点の外側に外れている．従って, **音楽の好みと色の好みが独立であるという仮説は危険率 5% で棄却される．**[*9]

表 8.2　χ^2 分布の表

α	0.995	0.975	0.950	0.900	0.500	0.05	0.025	0.01	0.005
$\nu = 6$	0.676	1.24	1.64	2.20	5.35	12.59	14.45	16.81	18.55
$\nu = 9$	1.73	2.70	3.33	4.17	8.34	16.92	19.02	21.67	23.59
$\nu = 10$	2.16	3.25	3.94	4.87	9.34	18.31	20.48	23.21	25.19
$\nu = 11$	2.60	3.82	4.57	5.58	10.34	19.68	21.92	24.73	26.76

このように，分割表から X を求めて，それを χ^2 分布の表と見比べることで，2 つの性質が独立であるかどうかを検定することができる．

[*9] 老婆心ながら誤解を避けるために付け加えると，この結論はでっち上げデータに基づくものであって，実際的な意味はない．

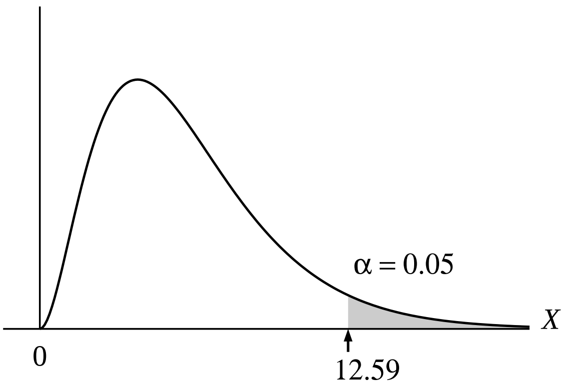


図 8.3　$n = 6$ に対する χ^2 分布関数

【章末問題】

問題 8–1　トウモロコシの中には黄色と白の 2 色の実が混ざってついているものがある．これはバイカラーと呼ばれ，優性の黄色の実の純系品種と劣性の白い実の純系品種の 1 代交配種であるために，その子（純系品種から見たら孫）である種子がもつ 2 個の遺伝子の黄色，白の組み合わせによって色の違いが生じたものである．この場合，最初の品種が純系であれば黄色と白の実の数の比の期待値はメンデルの法則に従って $3:1$ となるが，そうでないと，たとえば花粉が別の品種の花から飛んできたものだったりすると，比率の期待値は変化することになる．

いま，あるトウモロコシの実を調べたところ，144 粒の実のうち黄色の実が 119 個，白が 25 であった．この実が純系種の 1 代交配種から作られたという仮説は棄却できるか．危険率を 5% と 1% にとって検定しなさい．

問題 8–2　遺跡から貝殻が 12 個出土したので，その質量のデータを使って現存種の貝と同種のものであるかどうかを検定したい．データは次のようになっている（単位は g）．

11.78, 12.92, 7.55, 14.52, 12.05, 19.0, 11.29, 11.81, 15.38, 9.62, 14.19, 12.62

これと比較したい現存種の貝の質量の平均値は 12.6 g，標準偏差は 1.9 g である．この母集団は正規分布しているものとする．

1. 平均質量の分布を考えて検定を行ったとき，「この貝殻は現存種のものである」という仮説を危険度 5% で棄却できるか．
2. 式 (8.2) で定義される Z を使って χ^2 検定を行ったとき，「この貝殻は現存種のものである」という仮説を危険度 5% で棄却できるか．

問題 8–3　次の表は，ある疾病にその人がかかっているか，また飲酒の習慣があるかということをたずねたアンケートに基づいて作成した分割表である。この結果から，飲酒とその病気にかかることとは独立であるという仮説を危険率 1% で検定しなさい．

	罹患者	正常	計
習慣あり	37	37	74
習慣なし	93	133	226
計	130	170	300