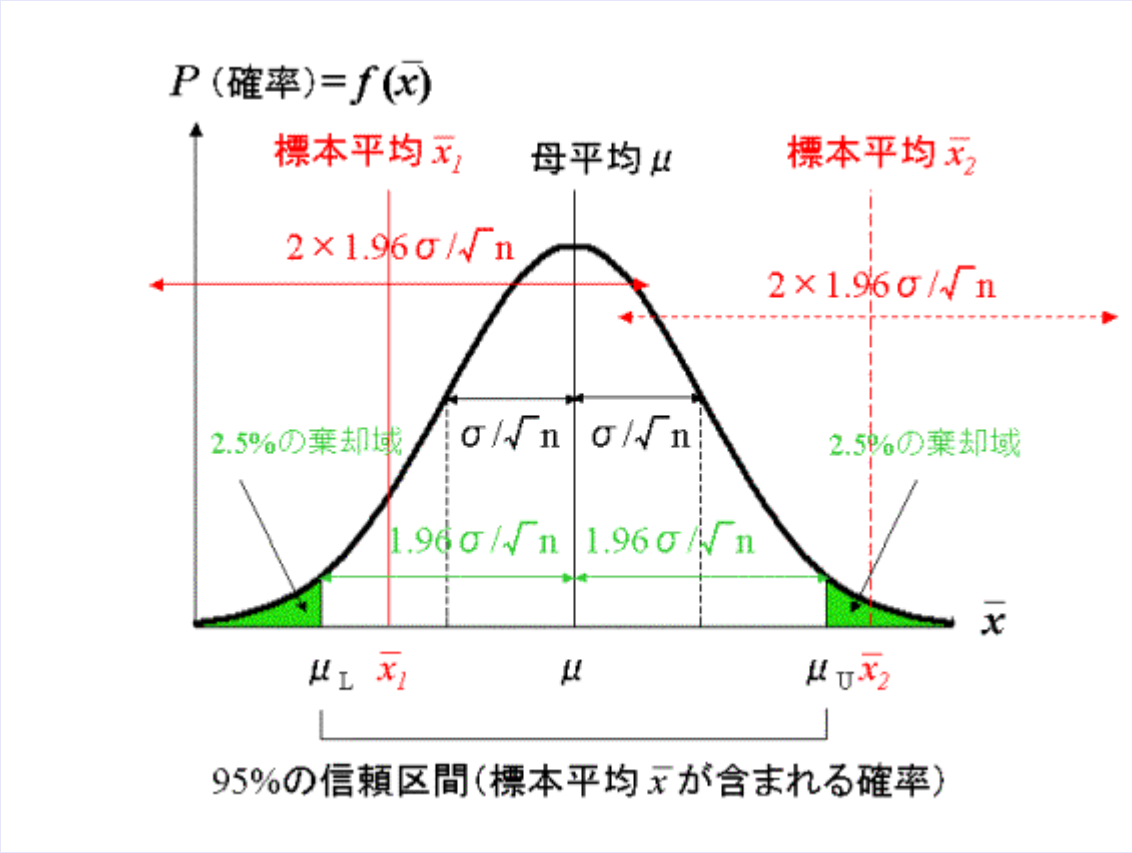


統計的仮説検定（信頼区間、信頼率、危険率、検出力、第一種および第二種の過誤）

母平均が μ で母分散が σ^2 の母集団から n の大きさの任意標本を抽出したとき、母集団がどんな分布であろうとも n が大きければ（ $n \geq 10$ ）中心極限定理により、その標本平均 \bar{x} は、下図のように、平均が μ で分散が σ^2 / \sqrt{n} の正規分布 $N(\mu, \sigma^2 / \sqrt{n})$ に近似できることが知られています。



上図で見られるように、母集団から無作為抽出をした大きさ n の標本の平均 \bar{x} は母平均 μ を中心として $\pm\infty$ まで分布し得ることがわかります。しかし、標本平均 \bar{x} の値が母平均 μ から離れるほどその実現確率 P は小さくなり（ $\bar{x}-\mu$ ）が $\pm\infty$ の極限では $P \rightarrow 0$ となり、その場合は決して実現しそうな標本平均 \bar{x} の値だと言えます。したがって、実際問題として標本平均 \bar{x} がどの程度なら実現可能な値であろうかという推定が必要になってきますが、その問題を解決する前に統計的判断の基本的考え方に触れておくことにします。

そこで、日本古来から伝わる丁半賭博を例にして統計的判断を考えてみることにします。サイコロを振って丁か半が出る確率はイカサマをしなくても何回振ってもその確率は $P = 1/2$ であることは経験的にも理論的にも知られています。しかし、サイコロを2回続けて振ったときにたまたま丁丁あるいは半半と続けて同じ目が出る場合も偶然にはあり得ることも考えられ、その確率は $P = 1/2 \times 1/2 = (1/2)^2 = 1/4$ です。同様に3回続けて振ったときに偶然に丁丁丁あるいは半半半と出る確率は $P = 1/2 \times 1/2 \times 1/2 = (1/2)^3 = 1/8$ となり、一般に n 回続けてサイコロを振ったときに偶然丁ばかりあるいは半ばかりの目が出る確率は $P = (1/2)^n$ となります。

いま、読者の二人（甲と乙）が運悪く丁半賭博をするはめに陥り、二人ともなけなしのお金を勝負の度にいつも半に賭けたものと仮定します。この賭博場の謳い文句はクリーンなイメージの「イカサマはない」（仮説）というものでした。が、しかし、現実には、最初の5回までの勝負では謳い文句通りイカサマは行われなかったのですが、たまたまその日は丁ばかりの目が続き、6回目以降はサイコロに細工をしていつも丁が出るようなイカサマをしていたとします。このとき、甲と乙のとった行動が統計的判断の考え方の基本となります。甲は生来アワテモノのせいか丁半勝負を始めてまだ2回しか丁が続かないのに、この賭博はイカサマであると判断し席を蹴って退場しました。これはイカサマをしていないという事実（仮説）が正しいのにイカサマをしているといった誤った判断をアワテテしてしまった過誤（error）であり、これを「第一種の過誤」（アワテモノの誤り、Type I errorあるいはerror of the first kind）と言います。それに対し、乙は生来ボンヤリモノのせいか、丁半勝負を始めてから10回も丁の目が続いて出ているのに、博打だからまあそんなこともあるだろうとちっともイカサマに気付かない様子で、お金がすってんてんになるまで何度も半を賭け続けていました。もちろん乙の考え方は確立統計の理論からすれば10回続けて丁

の目が出る確率は $P = (1/2)^{10}$ だけはある得るわけですから、謳い文句（仮説）通りイカサマが全く行われていなかったなら、この乙の判断には間違いがなかったと言えますが、現実には6回目以降の丁半勝負ではイカサマが行われており事実は仮説とは違うのに、その仮説をいつまでも正しいと判断したボンヤリモノの乙の判断にはやはり重大な過誤がります。このような事実は間違っている（仮説は正しくない）のに仮説を正しいと判断する過誤のことを「第二の過誤」（ボンヤリモノの誤り、**Type II error**あるいは **error of the second kind**）と言います。そして、甲乙いずれもその犯した過誤の時点での確率 P のことを「危険率」（critical rate）と統計学では呼んでいます。すなわち、甲の危険率（甲が第一種の過誤を犯す危険率）は $P = (1/2)^2 = 1/4 = 0.25 = 25\%$ であり、乙の危険率（乙が第二種の過誤を犯す確率）は $P = (1/2)^{10} \cong 0.001 = 0.1\%$ ということです。

この事実と判断の関係を表にすると下表のようになります。いつも判断が事実に等しい方がいいのですが、統計的判断ではある危険率で第一種の過誤や第二種の過誤を犯す可能性があります。つまり、第一種の過誤も第二種の過誤も同時に起こさないような判断は難しいということです。すなわち、本来どちらの過誤もできるだけ小さくなるような判断ができればいいのですが、現実にはある一定のデータ数では第一の過誤を小さくするとそれに応じて第二種の過誤は大きくなります。そこで、統計学で行う仮説検定は、第一種の過誤を犯す確率（危険率）を決めて判断し、第二種の過誤については評価を行わないのが普通ですが、その場合、データのバラツキ（分散）を減らすこととデータの数を増やすことによって、すなわち、 σ/\sqrt{n} を小さくすることによって、第二種の過誤を少なくすることができます。

| | | |
|--------------|-------------------|--------------------------------|
| 事実 仮説は正しい | 仮説は正しいと判断する ○ | 仮説は誤りと判断する 第一種の過誤（アワテモノの誤り） |
| 仮説は誤り | 第二種の過誤（ボンヤリモノの誤り） | ○ |

さて、いよいよ本題に戻ることにしましょう。上図は母平均が μ で母分散が σ^2 の母集団から n の大きさの任意標本を抽出したときの標本平均 \bar{x} の分布（正規分布 $N(\mu, \sigma^2/\sqrt{n})$ ）を表していますが、標本平均 \bar{x} の値が母平均 μ から離れれば離れるほど急速にその実現確率 $P = f(\bar{x})$ は小さくなり $\bar{x} = \pm\infty$ では $P \rightarrow 0$ となって実現性は全くあり得ないことになってしまいます。そこで、 \bar{x} の値が標本の平均としてどの程度までが許容できる限界かを定める必要がありますが、わたしたちは経験的に $P = 0.05 = 5\%$ 以下の確率しかない値はめったに起こりそうにもない実現値であろうと考えるのが妥当とみなします。すなわち、上図では両側併せて $P = 0.05 = 5\%$ になるところですから片側ずつにすると $P = 0.025 = 2.5\%$ 以下になる \bar{x} が実現しそうな値だと言えます。この限界点のことを信頼限界と呼び、上図では信頼上限は $\mu_U = \mu + 1.96\sigma/\sqrt{n}$ であり信頼下限は $\mu_L = \mu - 1.96\sigma/\sqrt{n}$ となります。1.96という数字は正規分布のパーセント点の表で0.025に相当する値です。すなわち、標本平均 \bar{x} の95%信頼区間（confidence interval）は

$$\mu - 1.96\sigma/\sqrt{n} \leq \bar{x} \leq \mu + 1.96\sigma/\sqrt{n}$$

あるいは

$$\left| \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \right| \leq 1.96$$

となります。したがって、上図の標本平均 \bar{x}_1 はこの区間内に入っているから採択可能な標本平均であり、標本平均 \bar{x}_2 はこの区間外にあるから棄却すべき標本平均であると言えます。ただし、そのような判断をしたときに第一種の過誤を犯す危険率は5%であるということになります。

このような信頼区間についての関係式を変形すると

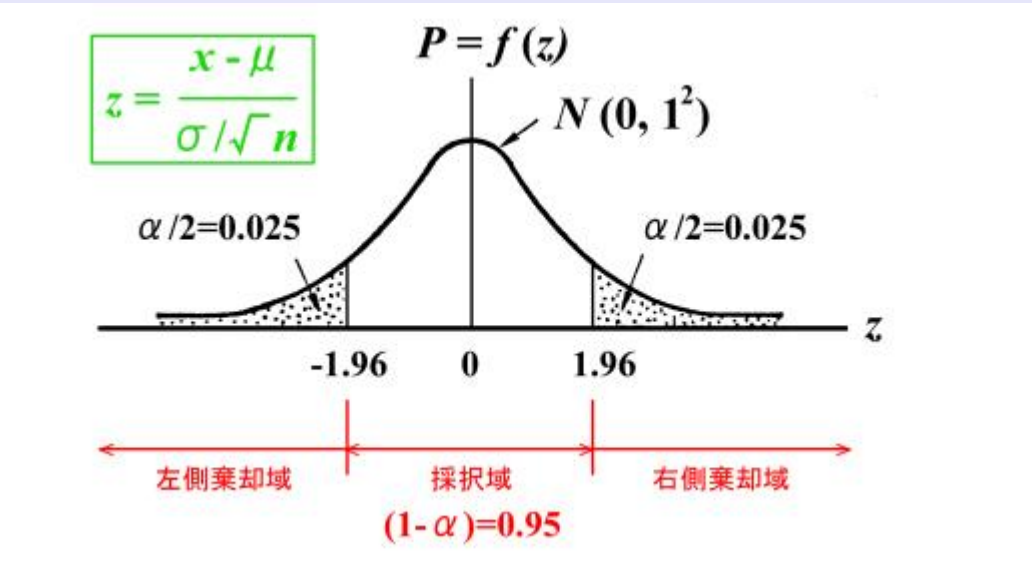
$$\bar{x} - 1.96\sigma/\sqrt{n} \leq \mu \leq \bar{x} + 1.96\sigma/\sqrt{n}$$

という関係式が得られますが、これは標本平均 \bar{x} から95%の信頼率で母平均 μ の信頼区間を推定することができるということと同じになります。

ここで、変数 \bar{x} を規準変数 z で変換すると

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

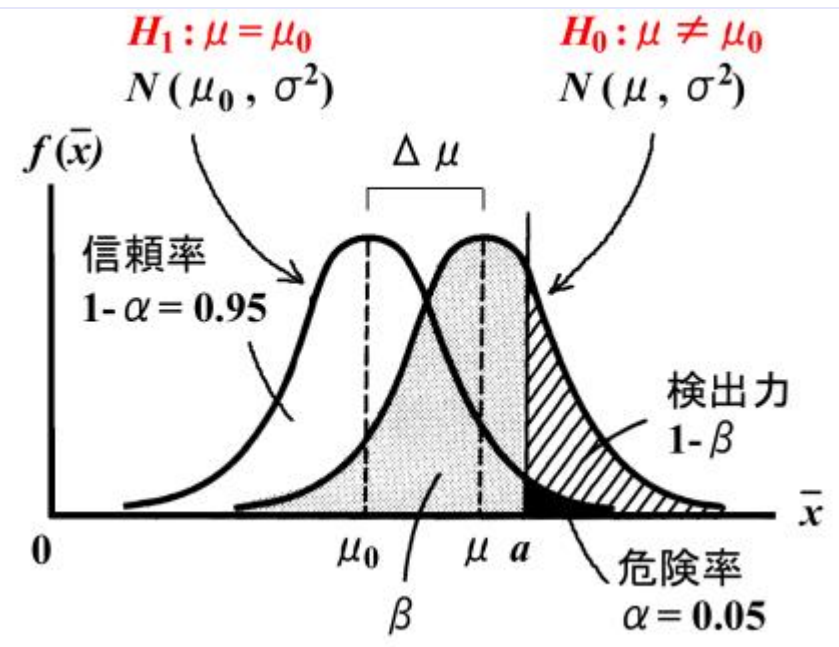
のようになるが、この変数 z もまた下図のような標準正規分布 $N(0,1^2)$ をするので、この図で信頼区間を決める方がわかりやすいかも知れません。



ちなみに、棄却域 (rejection region) に相当する確率 α のことを「危険率」 (critical rate) と言うが、判断の基準となる確率という意味で「有意水準」 (significance levelあるいはlevel of significance) と呼ぶこともあります。また、これとは逆に採択域 (acceptance region) は $(1 - \alpha)$ に相当する確率であり、この確率のことを $(1 - \alpha) \times 100\%$ 信頼率あるいは信頼係数 (confidence coefficient) と呼んでいます。それでは、有意水準が5%というのはどういうことを意味しているかということ、原則として棄却することを予測して立てた仮説 (このような仮説のことを帰無仮説 (null hypothesis) と言い、通常 H_0 という記号を使います) を棄却したとしても、そのことによる誤りは5%以下であるということです。言い換えれば、本当は帰無仮説が正しかったにもかかわらず、それを棄却することによって誤り (第一種の過誤) を犯すかも知れない危険を100回のうち5回までは認めてもいいということと同じですので、有意水準 α のことを危険率と言う場合が多いのです。帰無仮説 (たとえば、 $H_0 : \mu = \mu_0$) に対し、 $\mu \neq \mu_0$ となるような仮説のことを対立仮説 (alternative hypothesis) と呼び $H_1 : \mu \neq \mu_0$ という記号を用います。当然、帰無仮説を棄却する場合は対立仮説を採用することになります。

このように、統計的仮説検定においては、帰無仮説 H_0 が正しいにもかかわらずそれを棄却してしまう誤りが第一種の過誤であり、一般にこの誤りの確率を α という記号で表し、この α のことを危険率と呼んだり有意水準と呼んだりします。また、その確率 α に相当する規準変数 z 上の点 z_α のことを有意水準点と呼んでいます。したがって、危険率や有意水準というのは第一種の過誤を犯す確率であると簡潔に言うこともできます。これに対し、対立仮説が正しい (帰無仮説は誤り) のにそれを検出できないのが第二種の過誤で、その確率を記号 β を使って表すのが一般的です。また、 $(1 - \beta)$ に相当する確率のことを「検出力」と呼んでいます。

統計的仮説検定における危険率 (有意水準) α 、信頼率 $(1 - \alpha)$ 、第二種の過誤を犯す確率 β 、および検出力 $(1 - \beta)$ の関係は下図に示してあります。



いま、ある母集団の母平均 μ が μ_0 に等しいという帰無仮説 $H_0: \mu = \mu_0$ を立て、対立仮説として $H_1: \mu > \mu_0$ (すなわち、真の母平均 μ は帰無仮説で仮定した値 μ_0 より $\Delta\mu$ だけ大きい) を信頼率 $(1 - \sigma) = 0.95 = 95\%$ で大きさ n の任意標本について仮説検定してみることにします。この場合、対立仮説が $H_1: \mu > \mu_0$ なので右片側検定をすることになります。したがって、危険率 $\alpha = 0.05 = 5\%$ に相当する有意水準点は正規分布 $N(\mu_0, \sigma^2)$ 曲線上の $a = \mu_0 + 1.645\sigma/\sqrt{n}$ になり、この a 点より右側の黒い部分が帰無仮説 $H_0: \mu = \mu_0$ の棄却域になり、対立仮説 $H_1: \mu > \mu_0$ を採択することになります。また、この a 点より左側の部分はすべて帰無仮説 $H_0: \mu = \mu_0$ が棄却できず、帰無仮説 $H_0: \mu = \mu_0$ を採択することになります。しかしながら、上図で見るように、対立仮説である $\mu > \mu_0$ が真実であったとすると、正規分布 $N(\mu, \sigma^2)$ 曲線上の a 点より左側の部分 (影の部分) の標本平均 \bar{x} は正規分布 $N(\mu_0, \sigma^2)$ 曲線の帰無仮説 $H_0: \mu = \mu_0$ 採択域に入ってしまう、帰無仮説 $H_0: \mu = \mu_0$ が正しくない (対立仮説 $H_1: \mu > \mu_0$ が正しい) のに帰無仮説 $H_0: \mu = \mu_0$ が正しいという間違った判断 (第二種の過誤) をしてしまうことになります。すなわち、正規分布 $N(\mu, \sigma^2)$ 曲線上の影の部分が第二種の過誤を犯す確率 β に相当するわけです。そして、正規分布 $N(\mu, \sigma^2)$ 曲線上の斜線部分が検出力 $(1 - \beta)$ になります。検出力 $(1 - \beta)$ を上げれば相対的に第二種の過誤を犯す確率 β を下げることができるので、次に検出力 $(1 - \beta)$ を高めるにはどうしたらよいかを考えます。

正規分布 $N(\mu, \sigma^2)$ の検出力 $(1 - \beta)$ に対応する正規分布 $N(\mu_0, \sigma^2)$ の有意水準点 a は正規分布 $N(\mu_0, \sigma^2)$ の右片側検定の $\alpha = 0.05 = 5\%$ 点であるから

$$a = \mu_0 + 1.645\sigma/\sqrt{n}$$

となります。一方、この a 点は正規分布 $N(\mu, \sigma^2)$ 曲線上にあるから、この確率 $(1 - \beta)$ の始点に相当する標準変数の値を z_a とすれば

$$a = \mu + z_a\sigma/\sqrt{n}$$

となるので、 $\mu = \mu_0 + \Delta\mu$ の関係を考慮して z_a を求めると

$$z_a = \frac{a - \mu}{\sigma/\sqrt{n}} = \frac{(\mu_0 + 1.645\sigma/\sqrt{n}) - \mu}{\sigma/\sqrt{n}} = 1.645 - \frac{\mu - \mu_0}{\sigma/\sqrt{n}} = 1.645 - \frac{\Delta\mu}{\sigma/\sqrt{n}}$$

となります。

ここで、正規分布 $N(\mu, \sigma^2)$ の標本平均 \bar{x} を規準変数 z に変換して

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

とすれば、検出力 $(1 - \beta)$ は正規分布 $N(\mu, \sigma^2)$ の規準変数 z が正規分布 $N(\mu_0, \sigma^2)$ の棄却域に落ちる確率、すなわち、 $P(z \geq z_a)$ だから

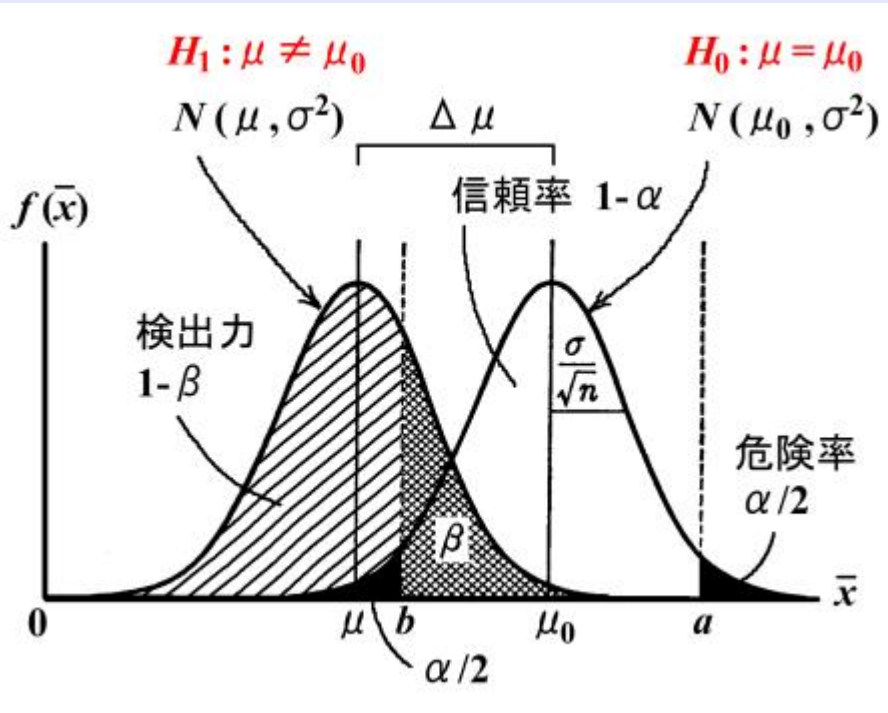
$$1 - \beta = P(z \geq z_a) = P\left(z \geq 1.645 - \frac{\Delta\mu}{\sigma / \sqrt{n}}\right)$$

となります。検出力 $(1 - \beta)$ を高めるためには確率 $P(z \geq z_a)$ を大きくすればよいので、 $\Delta\mu =$ 一定の条件下では

母分散 $\sigma \rightarrow$ 小
標本の大きさ $n \rightarrow$ 大

という条件を満たせば、検出力 $(1 - \beta)$ が高まり、相対的に第二種の過誤を犯す確率 β を小さくできることがわかります。

上述の統計的仮説検定は片側検定の場合を想定して行ったが、両側検定の場合も考慮しておく必要があるでしょう。両側検定の場合は、ある母集団の母平均 μ が μ_0 に等しいという帰無仮説 $H_0: \mu = \mu_0$ を立て、対立仮説として $H_1: \mu \neq \mu_0$ (すなわち、真の母平均 μ は帰無仮説で仮定した値 μ_0 とは違う) を信頼率 $(1 - \sigma) = 0.95 = 95\%$ で大きさ n の任意標本について仮説検定することになります。この場合、対立仮説が $H_1: \mu \neq \mu_0$ なので両側検定をすることになります。したがって、危険率 $\alpha / 2 = 0.025 = 2.5\%$ に相当する有意水準点は、下図の正規分布 $N(\mu_0, \sigma^2)$ 曲線上の $a = \mu_0 + 1.96\sigma / \sqrt{n}$ と $b = \mu_0 - 1.96\sigma / \sqrt{n}$ になり、この a 点より右側の黒い部分および b 点より左側の黒い部分が帰無仮説 $H_0: \mu = \mu_0$ の棄却域になり、対立仮説 $H_1: \mu \neq \mu_0$ を採択することになります。また、この b 点と a 点にはさまれた部分はすべて帰無仮説 $H_0: \mu = \mu_0$ が棄却できず、帰無仮説 $H_0: \mu = \mu_0$ を採択することになります。



しかしながら、上図で見るように、対立仮説である $\mu \neq \mu_0$ （たとえば真の母平均 μ は帰無仮説の値 $\mu = \mu_0$ より $\Delta\mu$ だけずれている）が真実であったとすると、正規分布 $N(\mu, \sigma^2)$ の母平均 μ が明らかに正規分布 $N(\mu_0, \sigma^2)$ の棄却域に入っている（したがって、 $\mu \neq \mu_0$ が正しい）にもかかわらず、正規分布 $N(\mu, \sigma^2)$ 曲線上の b 点より右側の部分（影の部分）の標本平均 \bar{x} は正規分布 $N(\mu_0, \sigma^2)$ 曲線の帰無仮説 $H_0: \mu = \mu_0$ 採択域に入ってしまう、帰無仮説 $H_0: \mu = \mu_0$ が正しくない（対立仮説 $H_1: \mu \neq \mu_0$ が正しい）のに帰無仮説 $H_0: \mu = \mu_0$ が正しいという間違った判断（第二種の過誤）をしてしまうことになります。すなわち、正規分布 $N(\mu, \sigma^2)$ 曲線上の影の部分が第二種の過誤を犯す確率 β に相当するわけです。そして、正規分布 $N(\mu, \sigma^2)$ 曲線上の斜線部分が検出力 $(1 - \beta)$ になります。検出力 $(1 - \beta)$ を上げれば相対的に第二種の過誤を犯す確率 β を下げることができます。

片側検定の場合と同じように、両側検定の場合についても検出力 $(1 - \beta)$ を求めてみることにします。正規分布 $N(\mu, \sigma^2)$ の検出力 $(1 - \beta)$ に対応する正規分布 $N(\mu_0, \sigma^2)$ の有意水準点 a と b はそれぞれ正規分布 $N(\mu_0, \sigma^2)$ の両側検定の $\alpha/2 = 0.025 = 2.5\%$ 点であるから

$$a = \mu_0 + 1.96\sigma/\sqrt{n} \qquad b = \mu_0 - 1.96\sigma/\sqrt{n}$$

となります。一方、この a 点と b 点はいずれも正規分布 $N(\mu, \sigma^2)$ 曲線上にあるから、この確率 $(1 - \beta)$ の始点に相当する規準変数の値を z_a および z_b とすれば

$$a = \mu + z_a\sigma/\sqrt{n} \qquad b = \mu - z_b\sigma/\sqrt{n}$$

となるので、 $\mu = \mu_0 + \Delta\mu$ の関係を考慮して z_a と z_b を求めると

$$\begin{aligned} z_a &= \frac{a - \mu}{\sigma/\sqrt{n}} = \frac{(\mu_0 + 1.96\sigma/\sqrt{n}) - \mu}{\sigma/\sqrt{n}} = 1.96 - \frac{\mu - \mu_0}{\sigma/\sqrt{n}} = 1.96 - \frac{\Delta\mu}{\sigma/\sqrt{n}} \\ z_b &= \frac{b - \mu}{\sigma/\sqrt{n}} = \frac{(\mu_0 - 1.96\sigma/\sqrt{n}) - \mu}{\sigma/\sqrt{n}} = -1.96 - \frac{\mu - \mu_0}{\sigma/\sqrt{n}} = -1.96 - \frac{\Delta\mu}{\sigma/\sqrt{n}} \end{aligned}$$

となります。

ここで、正規分布 $N(\mu, \sigma^2)$ の標本平均 \bar{x} を規準変数 z に変換して

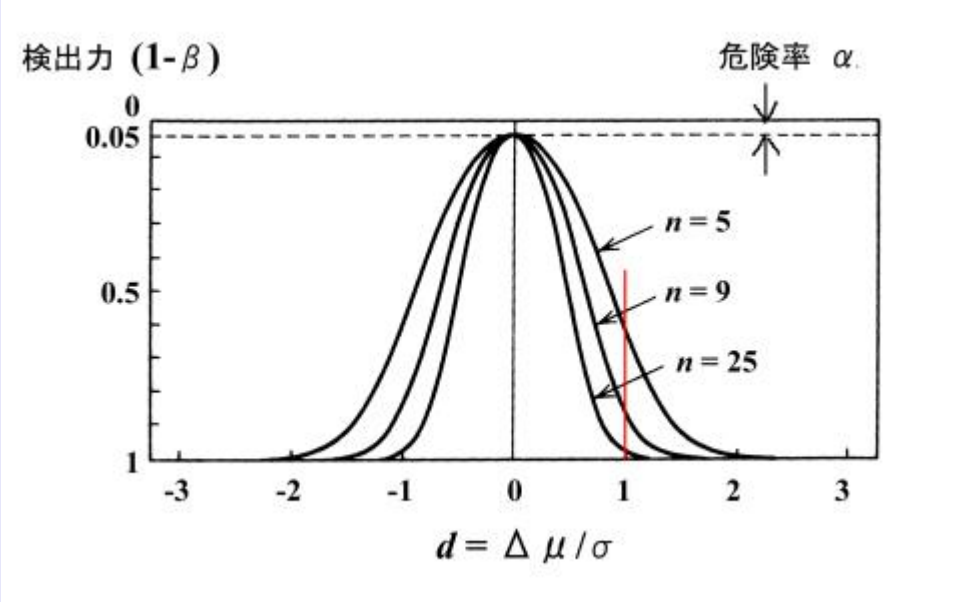
$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

とすれば、検出力 $(1 - \beta)$ は正規分布 $N(\mu, \sigma^2)$ の規準変数 z が正規分布 $N(\mu_0, \sigma^2)$ の棄却域に落ちる確率、すなわち、確率 $P(z \geq z_a)$ および確率 $P(z \leq z_b)$ の総和だから

$$\begin{aligned} 1 - \beta &= P(z \geq z_a) + P(z \leq z_b) \\ &= P\left(z \geq 1.96 - \frac{\Delta\mu}{\sigma/\sqrt{n}}\right) + P\left(z \leq -1.96 - \frac{\Delta\mu}{\sigma/\sqrt{n}}\right) \end{aligned}$$

となります。

標本の大きさ n によって検出力 $(1 - \beta)$ がどのように変わるのかを危険率 $\alpha = 0.05 = 5\%$ の両側検定の条件下で $d = \Delta\mu/\sigma$ を関数にしてプロットしたものを下図に示します。検出力曲線の逆がOP曲線（オペレーション曲線）です。OP曲線は β が統計パラメーターによってどう変わるかをプロットしたもので、抜き取り検査に使用されます。



たとえば、母平均の差が母分散程度 ($d = \Delta\mu/\sigma \cong 1$) である場合は検出力曲線と赤線の交点の読みから、標本の大きさが $n = 5$ のときは検出力 ($1 - \beta$) は約60%、 $n = 9$ のときは約85%、 $n = 25$ のときはほとんど100%近くになることがわかります。また、母平均の差が母分散の2倍以上 ($d = \Delta\mu/\sigma \geq 2$) である場合は、標本の大きさが $n = 5$ であっても検出力 ($1 - \beta$) はほぼ100%になり、第二種の過誤を犯す可能性はほとんどない ($\beta \cong 0 = 0\%$) ことがわかります。

(例題) 母集団は正規分布であり、母分散 σ^2 は既知であるとします。このとき

帰無仮説 $H_0 : \mu = \mu_0$
対立仮説 $H_1 : \mu \neq \mu_0$

という条件の下で、有意水準5%で仮説検定を行い、標本の大きさ $n = 9$ について検出力 ($1 - \beta$) を計算してみましょう。ただし、母平均 $\Delta\mu = \mu - \mu_0$ の差は母分散 σ^2 程度であったと仮定します。

(解答) 仮説 $H_0 : \mu = \mu_0$ および $H_1 : \mu \neq \mu_0$ の有意水準5%の棄却域は、両側検定であるので

$$\frac{|\bar{x} - \mu_0|}{\sigma/\sqrt{n}} \geq 1.96$$

となります。真の母平均が μ であるとき、標本平均 \bar{x} がこの棄却域に落ちる確率すなわち検出力 ($1 - \beta$) は

$$\begin{aligned} 1 - \beta &= P\left(\left|\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right| \geq 1.96\right) = P\left(\left|\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} + \frac{\mu - \mu_0}{\sigma/\sqrt{n}}\right| \geq 1.96\right) \\ &= P\left(\left|\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} + \frac{\Delta\mu}{\sigma/\sqrt{n}}\right| \geq 1.96\right) \end{aligned}$$

となるが、ここで

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

とおけば、この規準変数 z は標準正規分布 $N(0,1^2)$ に従います。よって

$$\begin{aligned} 1-\beta &= P\left(\left|z + \frac{\Delta\mu}{\sigma/\sqrt{n}}\right| \geq 1.96\right) \\ &= P\left(z \geq 1.96 - \frac{\Delta\mu}{\sigma/\sqrt{n}}\right) + P\left(z \leq -1.96 - \frac{\Delta\mu}{\sigma/\sqrt{n}}\right) \end{aligned}$$

となります。題意より

$$\begin{aligned} n &= 9 \\ \Delta\mu &\cong \sigma \end{aligned}$$

だから、求める検出力 $(1-\beta)$ は

$$\begin{aligned} 1-\beta &\cong P(z \geq 1.96 - 3) + P(z \leq -1.96 - 3) \\ &= P(z \geq -1.04) + P(z \leq -4.96) \\ &= [0.5 + \{0.5 - P(z \geq 1.04)\}] + P(z \geq 4.96) \\ &= \{0.5 + (0.5 - 0.1492)\} + 0.0000003525 \\ &= 0.5 + 0.3508 + 0.0000003525 \\ &\cong 0.851 = 85.1\% \end{aligned}$$

となります。