推定と検定と検定廃止論

寺子屋・統計庵　其之一

杉本解析差有比数

杉本典夫

推定と検定と検定廃止論

- 中心極限定理
- 区間推定の手順
- 信頼区間の解釈
- 統計的仮説検定の手順
- 有意確率p値の意味
- 検定と推定と科学的判断の関係−検定廃止論

推定と検定

統計学的推論

推定
estimate

定量試験

母数を推測
↓
pH計に相当

検定
test

定性試験

母数が基準値と
等しいかどうかを
○×式で推測
↓
リトマス試験紙に相当

検定よりも推定の方が重要

ところが研究現場や厚労省では検定が偏重されている
↑
○×式の方が採点が楽！

推定と検定の基本原理−中心極限定理

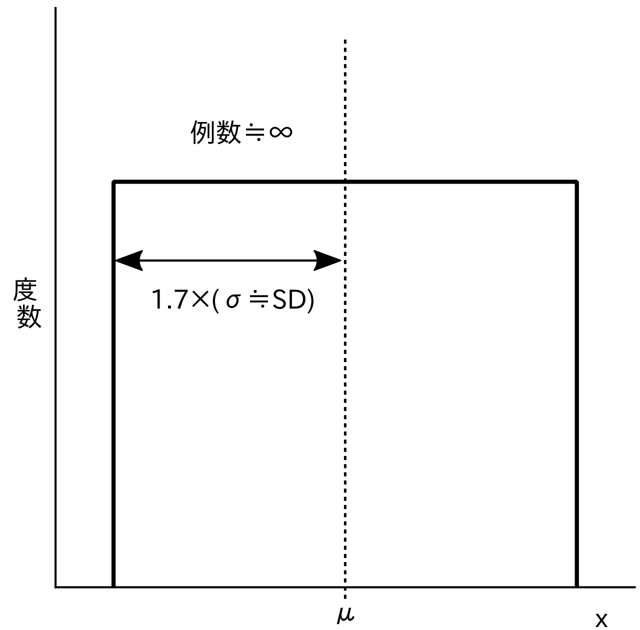


図1.3.3 母集団のデータ分布

n例を
無作為抽出して
標本平均値mを
無限回求める

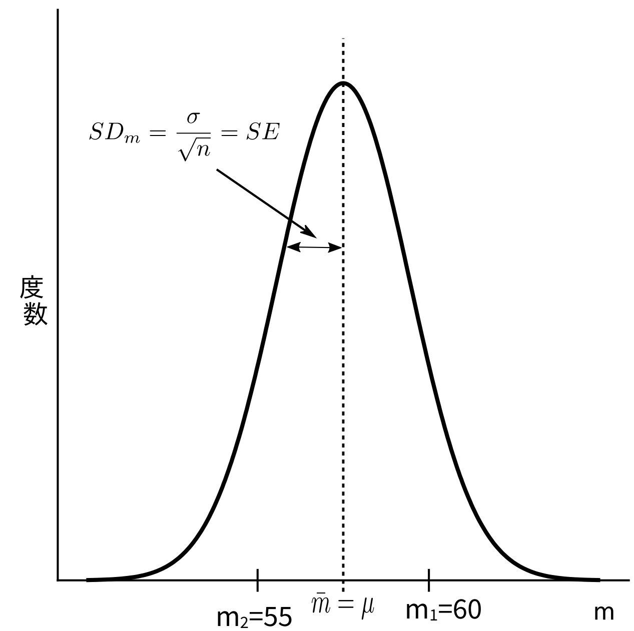


図1.3.4 標本平均値の分布

標本平均値の分布の特徴

1. 母集団がどんな分布をしていてもnが多いほど正規分布に近似する
   →中心極限定理…推測統計学の基本定理
2. 標本平均値mの平均値$\bar{m}$は母平均値μと一致する
3. 標本平均値の標準偏差SD_mは次のような値になる→標準誤差SE

$$SD_m = \frac{\sigma}{\sqrt{n}} = SE \fallingdotseq \frac{SD}{\sqrt{n}}$$

σ：母標準偏差　　n：標本集団の例数
SD：標本集団から求めたσの推測値

中心極限定理のシミュレーション−平均値

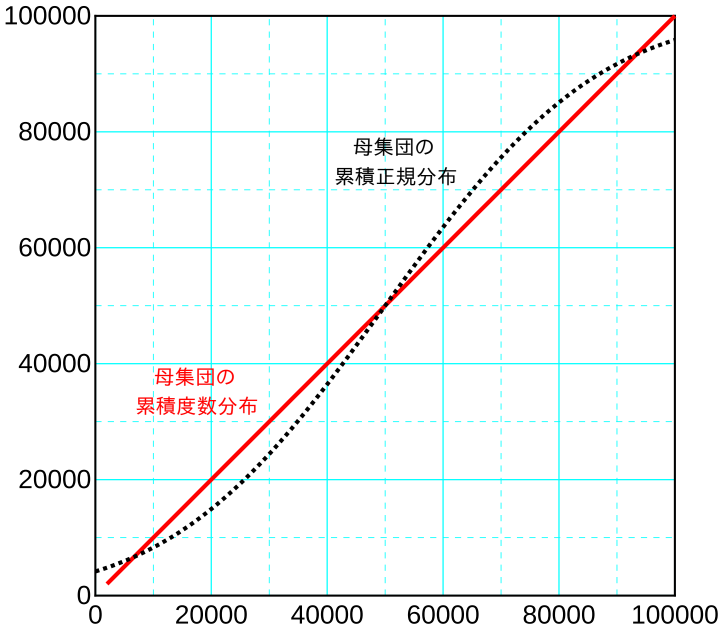


図 付録2.1 母集団:整数

n例を
無作為抽出して
標本平均値を
1万回求める

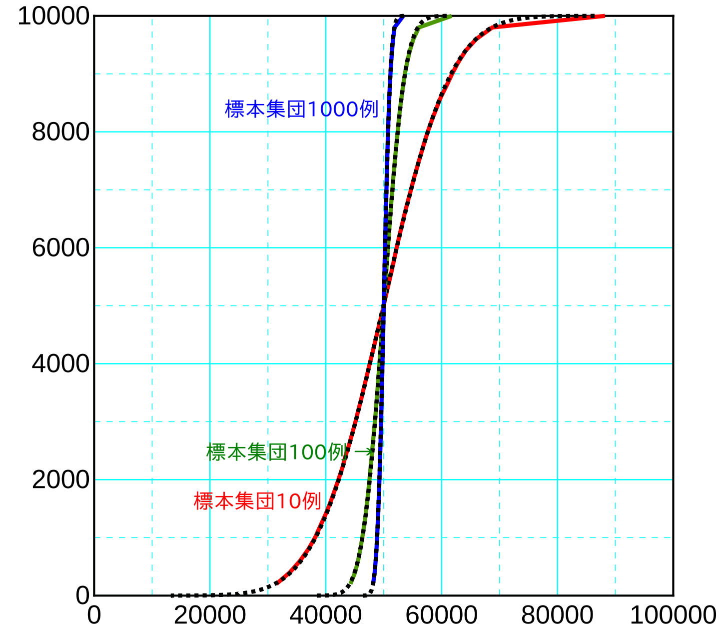


図 付録2.2 標本平均値(母集団:整数)

- 母集団：1から10万までの整数10万個…離散型一様分布

  赤色の実線：実際のデータの累積度数分布(直線)

  黒色の点線：平均値と標準偏差が同じ時の累積正規分布(シグモイド曲線)

- 標本平均値の分布：n例の標本集団を1万回無作為抽出した時の標本平均値1万個

  赤色の実線：10例の標本集団における標本平均値の累積度数分布

  緑色の実線：100例の標本集団における標本平均値の累積度数分布

  青色の実線：1000例の標本集団における標本平均値の累積度数分布

  黒色の点線：平均値と標準誤差が同じ時の累積正規分布

中心極限定理のシミュレーション−平均値

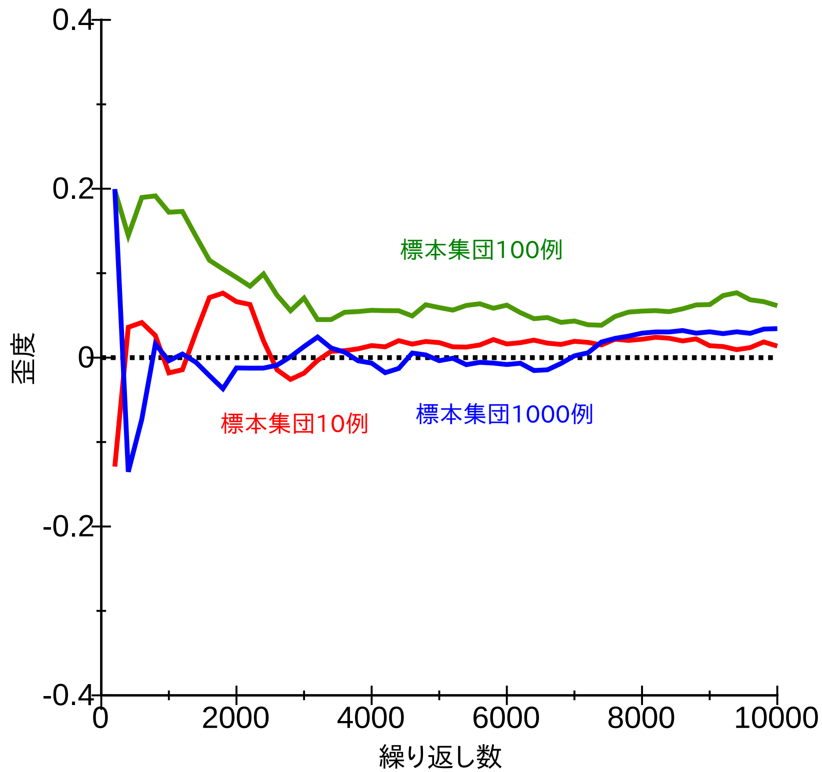


図 付録2.5 歪度の推移(母集団:整数)

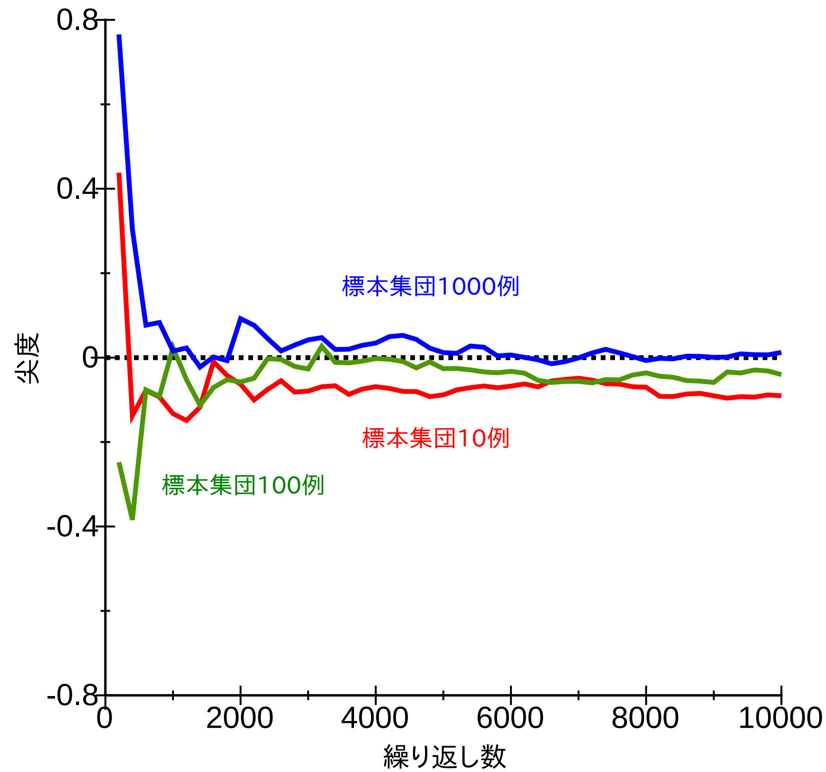


図 付録2.6 尖度の推移(母集団:整数)

-2<歪度<2ならほぼ左右対称

-4<尖度<4ならほぼ正規

標本平均値の分布の歪度も尖度も繰り返し数が5000回以上になると値が安定

→昔は手作業のため1000回程度が限界だったので、昔の統計学解説書には

「標本集団が20〜30例未満の時は中心極限定理が成り立たない」と書かれているものがある

中心極限定理のシミュレーション−平均値

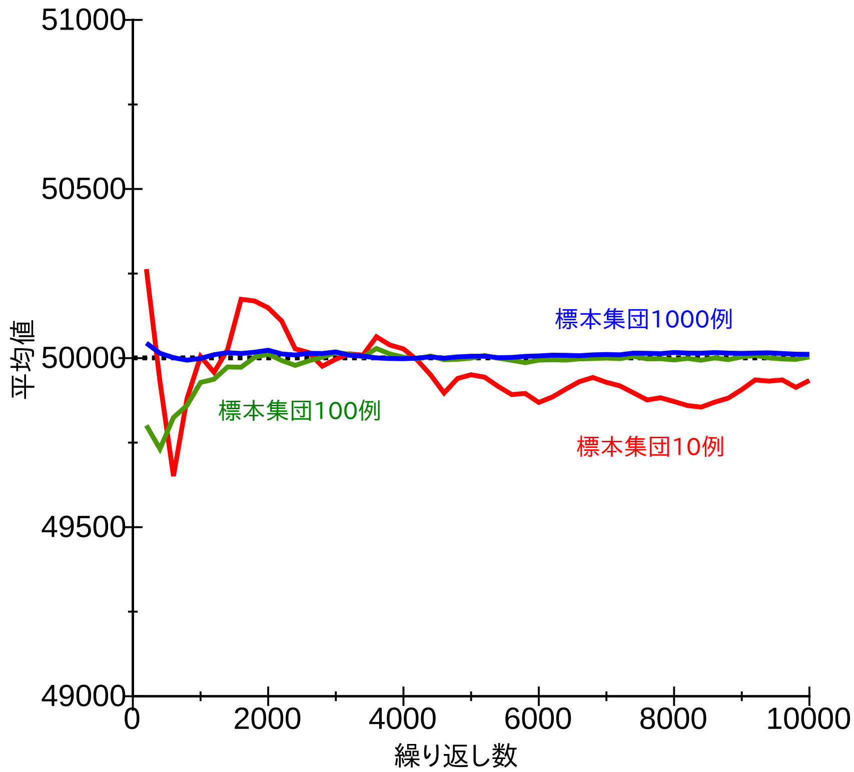


図 付録2.3 平均値の推移(母集団:整数)

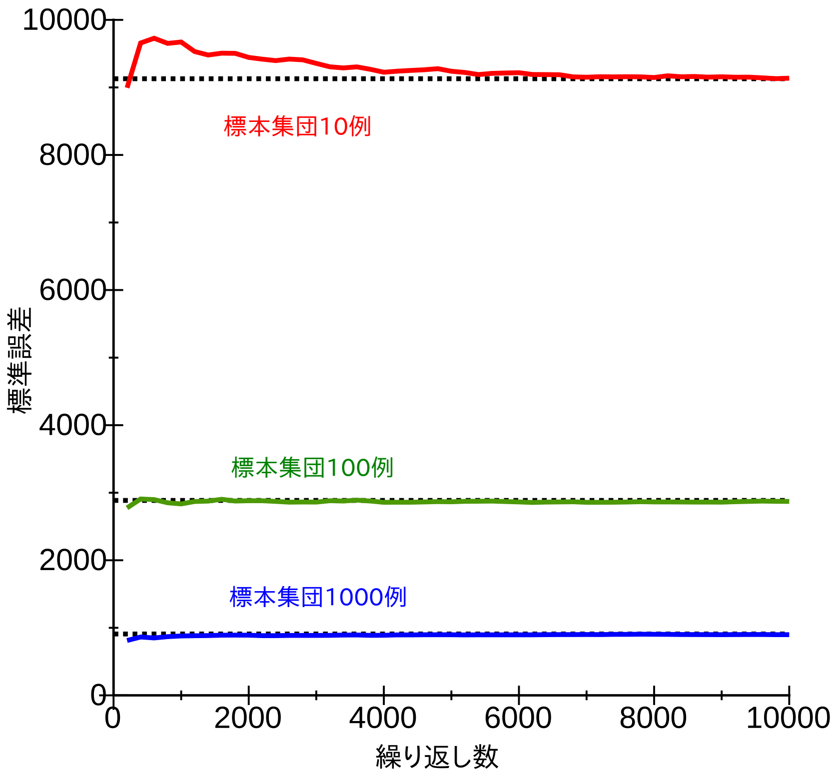


図 付録2.4 標準誤差の推移(母集団:整数)

標本平均値の平均値は母平均値に近似し、標準偏差は標準誤差に近似する

→標本平均値と不偏分散の不偏性(推定値の平均値が母数に一致する性質)による

中心極限定理のシミュレーション−平均値

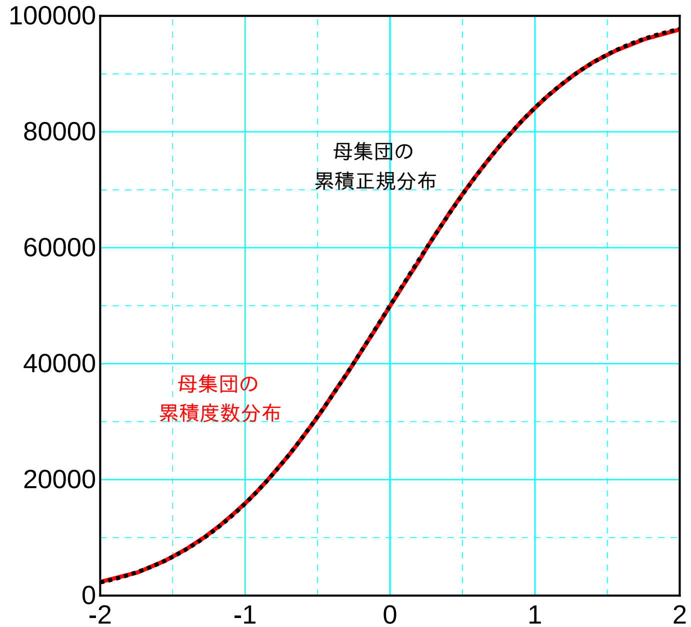


図 付録2.7 母集団:正規乱数

n例を
無作為抽出して
標本平均値を
1万回求める

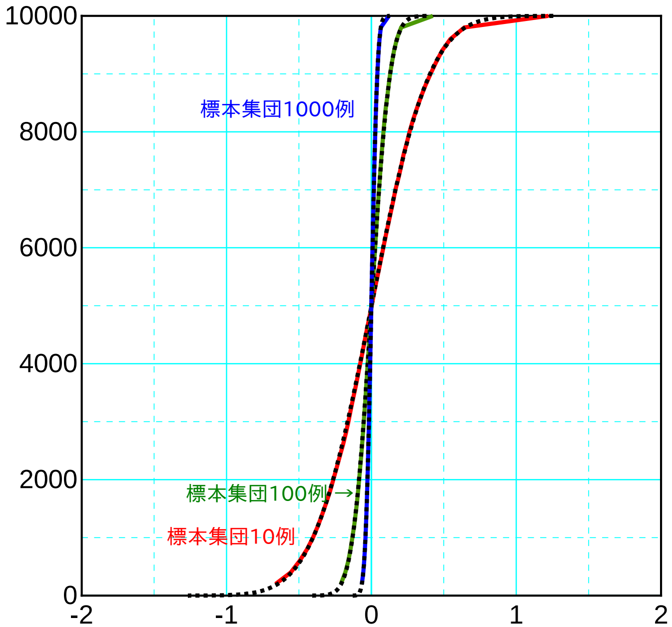


図 付録2.8 標本平均値(母集団:正規乱数)

- 母集団：母平均値が0、母標準偏差が1の正規乱数10万個…正規分布
  赤色の実線：実際のデータの累積度数分布
  黒色の点線：平均値と標準偏差が同じ時の累積正規分布

- 標本平均値の分布：n例の標本集団を1万回無作為抽出した時の標本平均値1万個
  赤色の実線：10例の標本集団における標本平均値の累積度数分布
  緑色の実線：100例の標本集団における標本平均値の累積度数分布
  青色の実線：1000例の標本集団における標本平均値の累積度数分布
  黒色の点線：平均値と標準誤差が同じ時の累積正規分布

中心極限定理のシミュレーション−平均値

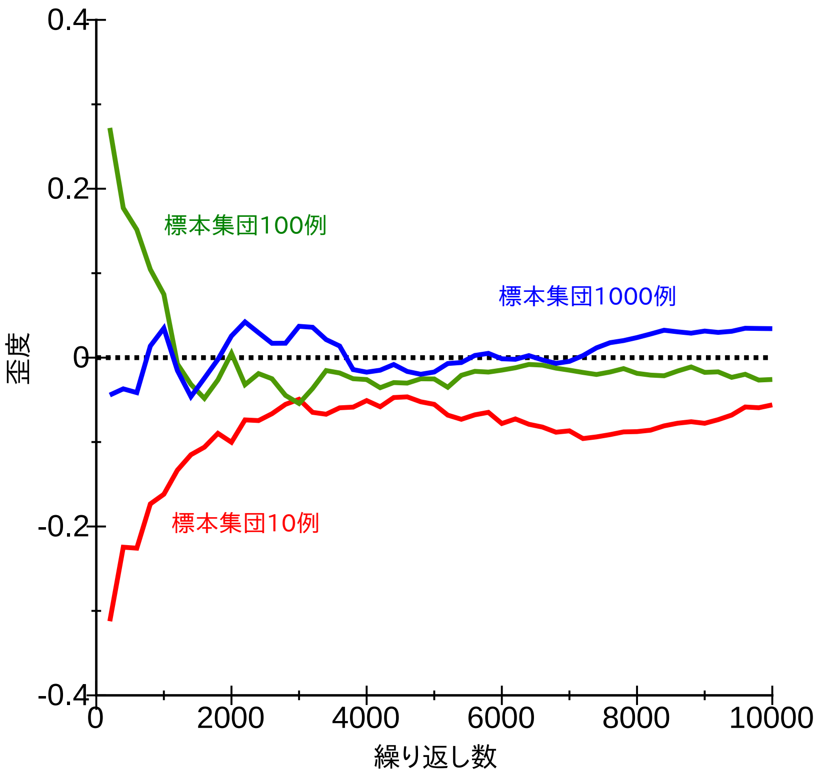


図 付録2.11 歪度の推移(母集団：正規乱数)

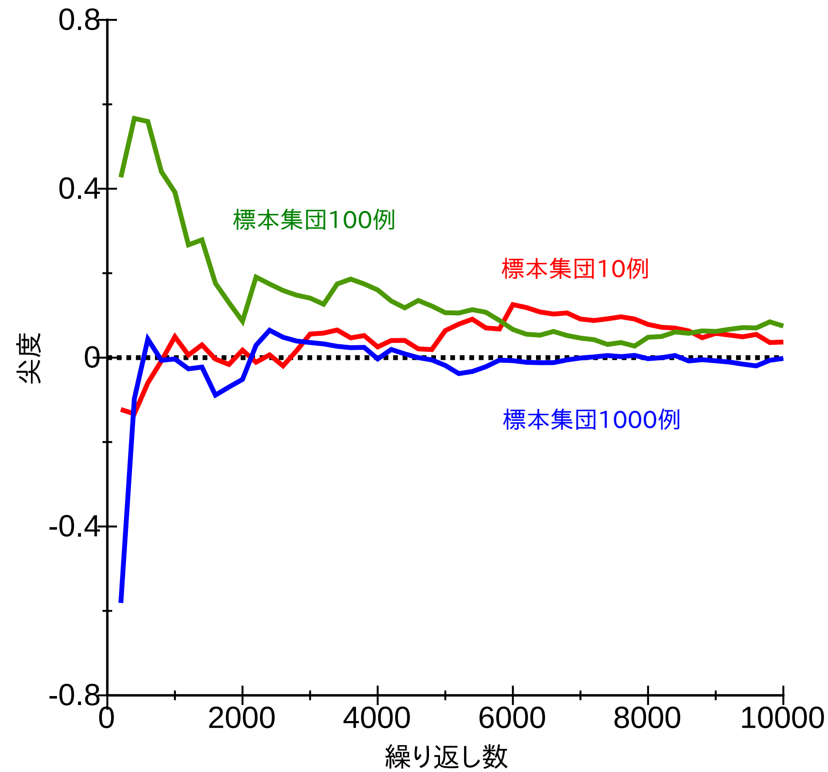


図 付録2.12 尖度の推移(母集団：正規乱数)

-2<歪度<2ならほぼ左右対称

-4<尖度<4ならほぼ正規

母集団が正規分布でも離散型一様分布でも標本平均値の分布はほぼ同じ

→標本集団が10例以上あれば標本平均値の分布は近似的に正規分布になる！

中心極限定理のシミュレーション−平均値

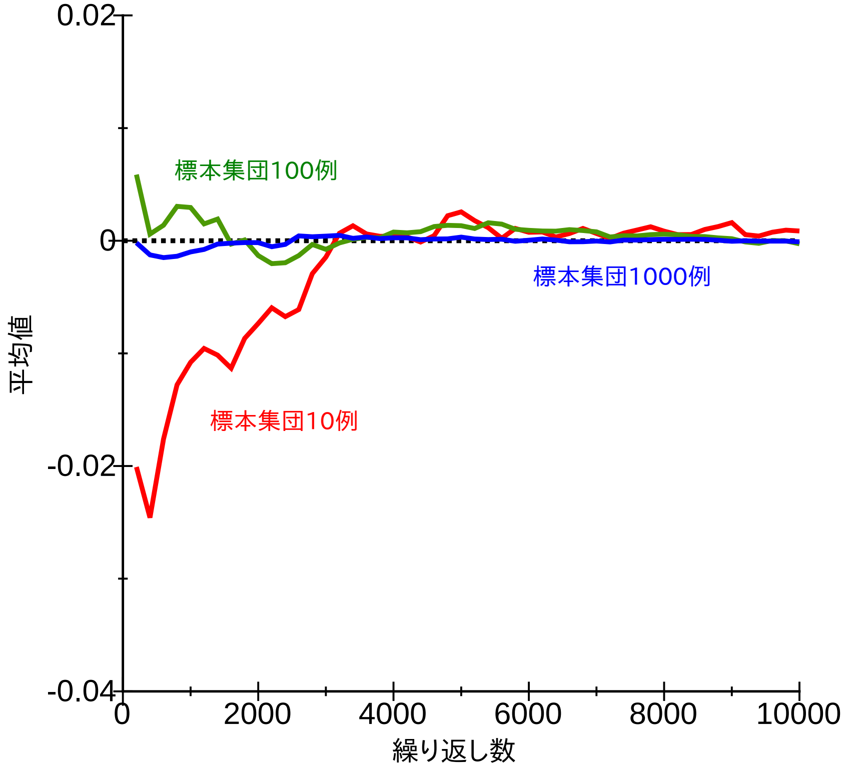


図 付録2.9 平均値の推移(母集団：正規乱数)

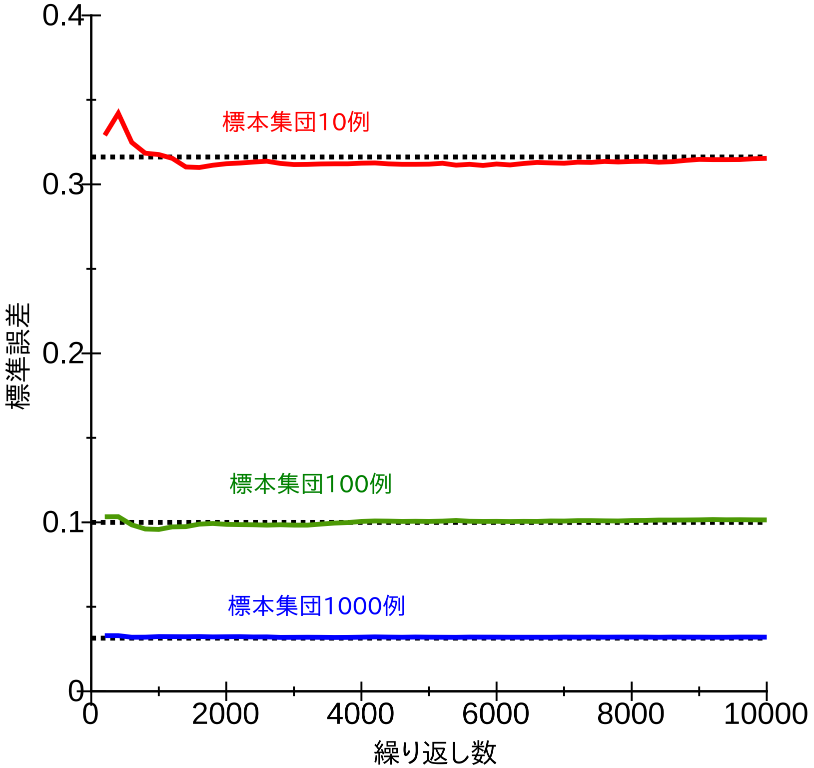


図 付録2.10 標準誤差の推移(母集団：正規乱数)

母集団が正規分布でも離散型一様分布でも標本平均値の分布はほぼ同じ

→標本集団が10例以上あれば標本平均値の平均値は母平均値に近似し、標準偏差は標準誤差に近似する

推定と検定と検定廃止論

- 中心極限定理
- 区間推定の手順
- 信頼区間の解釈
- 統計的仮説検定の手順
- 有意確率p値の意味
- 検定と推定と科学的判断の関係−検定廃止論

推定の原理

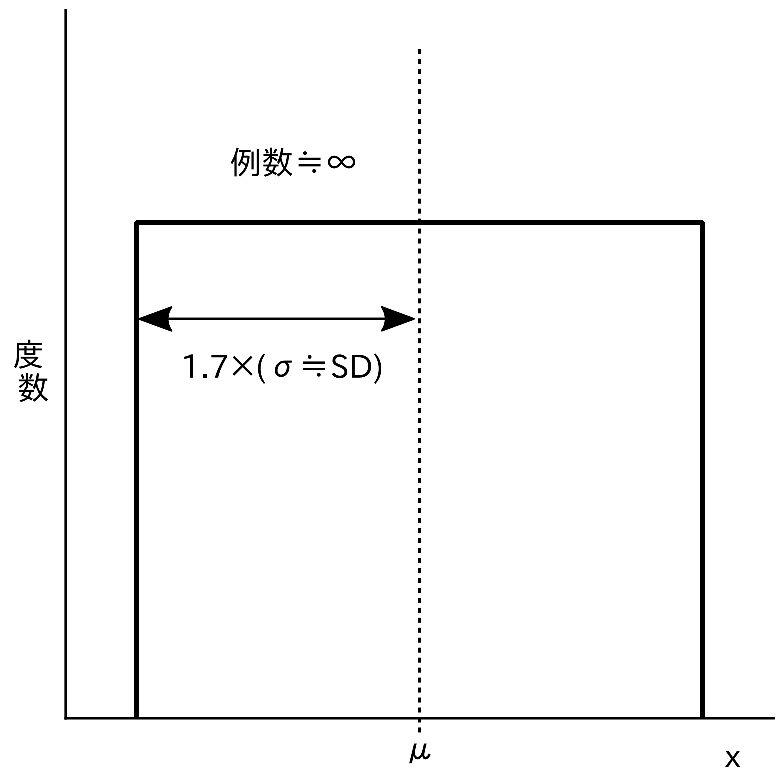


図1.3.3 母集団のデータ分布

n例を
無作為抽出して
標本平均値mを
無限回求める

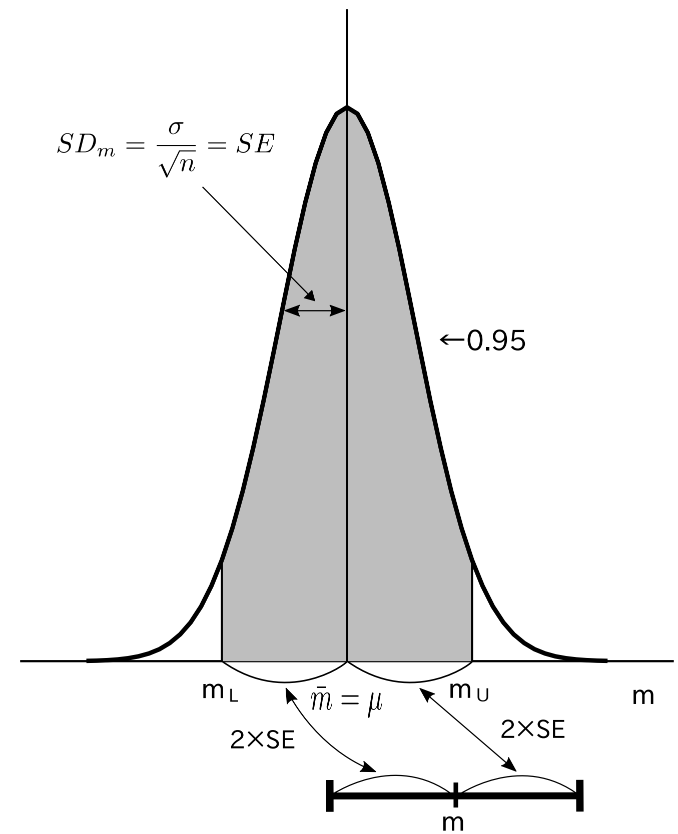


図1.4.3 標本平均値の分布と信頼区間

点推定(point estimation)：母数(母平均値)をピンポイント(標本平均値)で推測

区間推定(interval estimation)：母数をある程度の幅(信頼区間)を持たせて推測

平均値の区間推定法

標本平均値mの分布は近似的に正規分布になる

mの平均値は母平均値μになり、mの標準偏差は標準誤差SEになる
μ±2×SEの範囲に約95%のmが含まれる

ある標本平均値mがμ±2×SEの範囲に含まれる確率は約95%→母集団から求めた95%信頼区間

逆にm±2×SEの範囲にμが含まれる確率も約95%→標本集団から求めた95%信頼区間

95%信頼区間：$\mu \fallingdotseq m \pm 2 \times SE \rightarrow \mu_L = m - 2 \times SE \quad \mu_U = m + 2 \times SE$

95%信頼区間(95%CI)：母平均値が95%の確率で含まれる区間、95%信頼限界(95%CL)

μ_L：信頼区間下限　　μ_U：信頼区間上限　　95%：信頼係数

区間推定の模式図

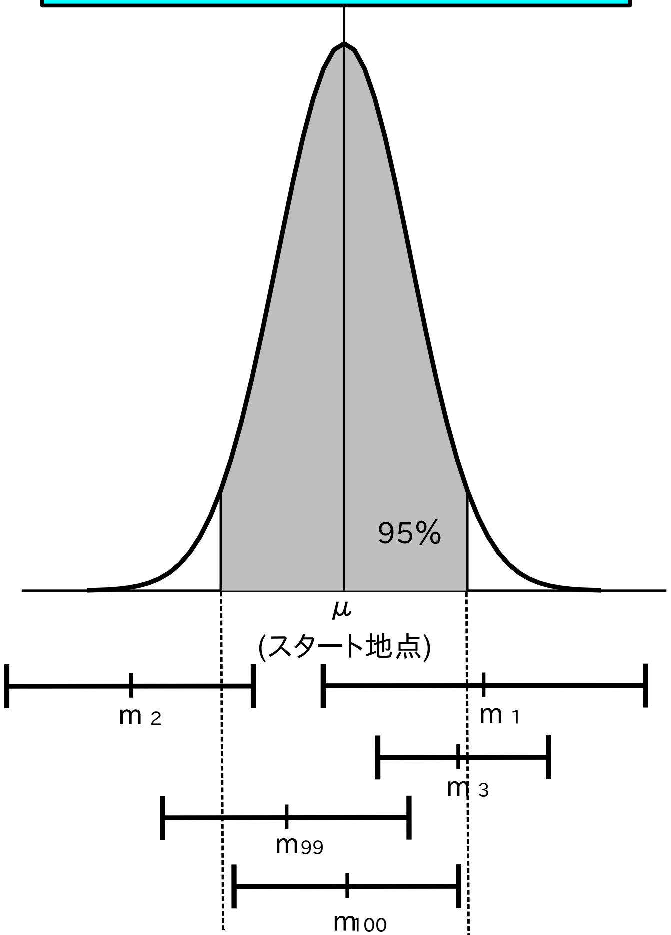


図1.4.4 Neyman-Pearson釣り大会

Neyman-Pearson会主催の釣り大会に100名の釣師(fisher!)が参加した。

彼等はm_1〜m_{100}のゼッケンを付けてスタート地点μから沖に出て、釣り船をランダムに動かしながら釣りをした。

そして終了時間になってスタート地点に戻ろうとしたが、釣りに没頭していたのでスタート地点μがわからない！

区間推定の模式図

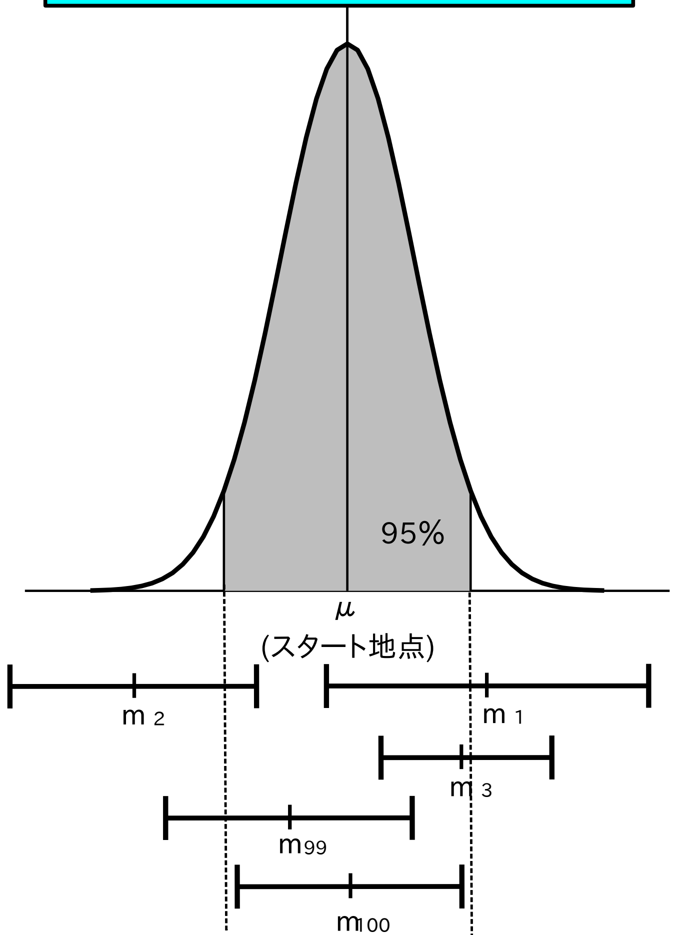


図1.4.4 Neyman-Pearson釣り大会

Neyman-Pearson会は全ての釣り船のランダムな動き(母標準偏差)に基づいてスタート地点μを中心にした釣師達の理論的な分布を描き95名の釣師が含まれると考えられる範囲を求めて釣師達を探す→母集団から求めた95%信頼区間

釣師達は今いる位置から真っ直ぐに海岸に戻ってそこを仮のスタート地点とし、自分の釣り船のランダムな動き(標本標準偏差)に基づいて釣師達の理論的な分布を描き95名の釣師達が含まれると考えられる範囲を求めてスタート地点を探す→標本集団から求めた95%信頼区間

区間推定の模式図

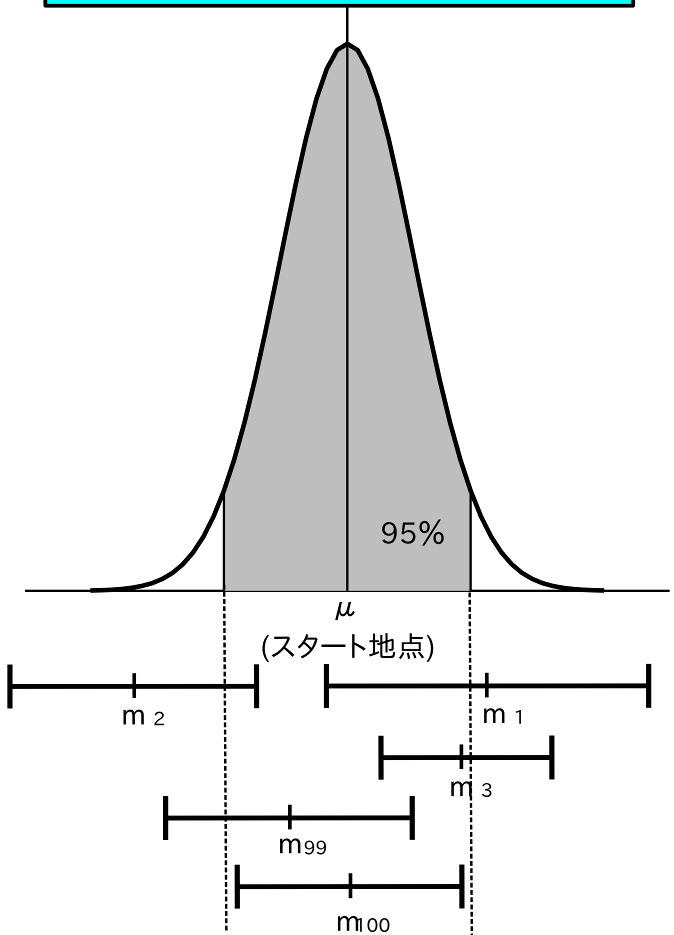


図1.4.4 Neyman-Pearson釣り大会

釣師によって仮のスタート地点(標本平均値)と95%信頼区間の幅が微妙に異なる

→釣師m_1は自分ではスタート地点μを探し出せるのに主催者はm_1を探し出せない

→釣師m_3は自分ではスタート地点μを探し出せないのに主催者はm_3を探し出せる

中心極限定理のシミュレーション−平均値

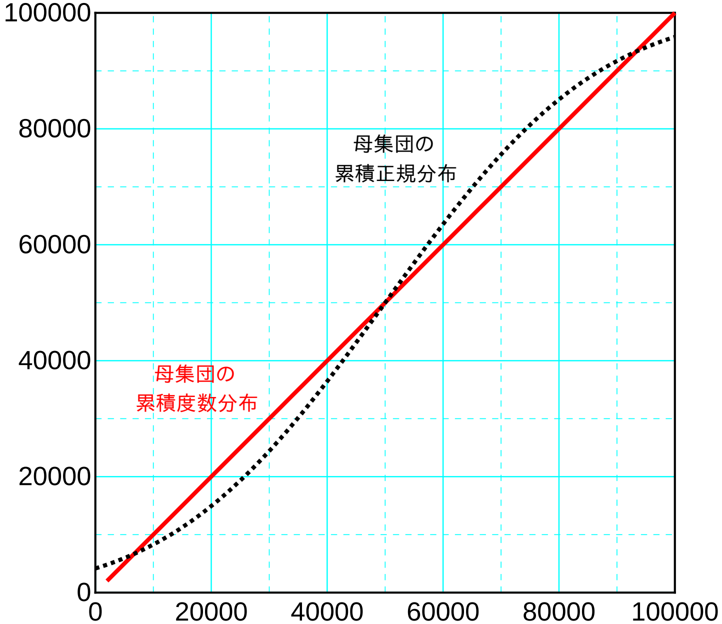
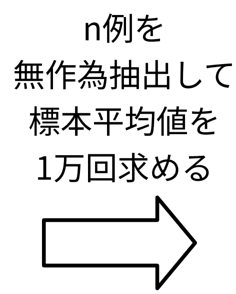


図 付録2.1 母集団:整数



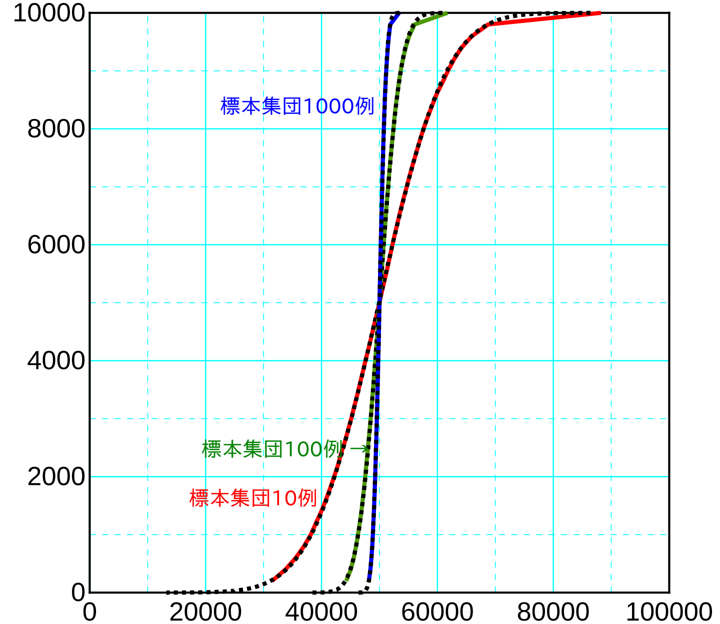


図 付録2.2 標本平均値(母集団:整数)

- 母集団：1から10万までの整数10万個…離散型一様分布

  赤色の実線：実際のデータの累積度数分布(直線)

  黒色の点線：平均値と標準偏差が同じ時の累積正規分布(シグモイド曲線)

- 標本平均値の分布：n例の標本集団を1万回無作為抽出した時の標本平均値1万個

  赤色の実線：10例の標本集団における標本平均値の累積度数分布

  緑色の実線：100例の標本集団における標本平均値の累積度数分布

  青色の実線：1000例の標本集団における標本平均値の累積度数分布

  黒色の点線：平均値と標準誤差が同じ時の累積正規分布

クロス集計表−母集団：整数10万個　標本集団：10例

母集団＼標本集団	信頼区間(%)	信頼区間外(%)	計(%)
信頼区間	9326(93.26)←(e)	179(1.79)	9505(95.05)←(a)
信頼区間外	135(1.35)	360(3.60)←(e)	495(4.95)←(b)
計	9461(94.61)←(c)	539(5.39)←(d)	10000(100)

(a) 母平均値と母標準偏差から求めた95%信頼区間(正規分布利用：μ_L=32109.4〜μ_U=67891.6)に入った
標本平均値の個数=9505(95.05%：本来の信頼係数)

(b) 母平均値と母標準偏差から求めた95%信頼区間に入らなかった標本平均値の個数=495(4.95%)

(c) 標本平均値と標本標準偏差から求めた95%信頼区間(自由度9のt分布利用：標本集団によって異なる)に
母平均値=50000.5が入った回数=9461(94.61%：被覆確率=実質的な信頼係数)

(d) 標本平均値と標本標準偏差から求めた95%信頼区間に母平均値=50000.5が入らなかった回数=539(5.39%)

(e) 母集団と標本集団の判定が一致している標本平均値の個数=9326+360=9686(96.86%：一致率)

クロス集計表−母集団：整数10万個　標本集団：100例

母集団＼標本集団	信頼区間(%)	信頼区間外(%)	計(%)
信頼区間	9464(94.64)←(e)	43(0.43)	9507(95.07)←(a)
信頼区間外	45(0.45)	448(4.48)←(e)	493(4.93)←(b)
計	9509(95.09)←(c)	491(4.91)←(d)	10000(100)

(a) 母平均値と母標準偏差から求めた95%信頼区間(正規分布利用：μ_L=44345.4～μ_U=55655.6)に入った

標本平均値の個数=9507(95.07%：本来の信頼係数)

(b) 母平均値と母標準偏差から求めた95%信頼区間に入らなかった標本平均値の個数=493(4.93%)

(c) 標本平均値と標本標準偏差から求めた95%信頼区間(自由度99のt分布利用：標本集団によって異なる)に

母平均値=50003.4が入った回数=9509(95.09%：被覆確率=実質的な信頼係数)

(d) 標本平均値と標本標準偏差から求めた95%信頼区間に母平均値=50003.4が入らなかった回数=491(4.91%)

(e) 母集団と標本集団の判定が一致している標本平均値の個数=9464+448=9912(99.12%：一致率)

クロス集計表−母集団：整数10万個　標本集団：1000例

母集団＼標本集団	信頼区間(%)	信頼区間外(%)	計(%)
信頼区間	9510(95.10)←(e)	11(0.11)	9521(95.21)←(a)
信頼区間外	20(0.20)	459(4.59)←(e)	479(4.79)←(b)
計	9530(95.30)←(c)	470(4.70)←(d)	10000(100)

(a) 母平均値と母標準偏差から求めた95%信頼区間(正規分布利用：μ_L=48220.3～μ_U=51780.7)に入った
標本平均値の個数=9521(95.21%：本来の信頼係数)

(b) 母平均値と母標準偏差から求めた95%信頼区間に入らなかった標本平均値の個数=479(4.79%)

(c) 標本平均値と標本標準偏差から求めた95%信頼区間(自由度999のt分布利用：標本集団によって異なる)に
母平均値=50011.1が入った回数=9530(95.30%：被覆確率=実質的な信頼係数)

(d) 標本平均値と標本標準偏差から求めた95%信頼区間に母平均値=50011.1が入らなかった回数=470(4.70%)

(e) 母集団と標本集団の判定が一致している標本平均値の個数=9510+459=9969(99.69%：一致率)

推定と検定と検定廃止論

- 中心極限定理
- 区間推定の手順
- 信頼区間の解釈
- 統計的仮説検定の手順
- 有意確率p値の意味
- 検定と推定と科学的判断の関係−検定廃止論

信頼区間の解釈−ネイマン・ピアソン流

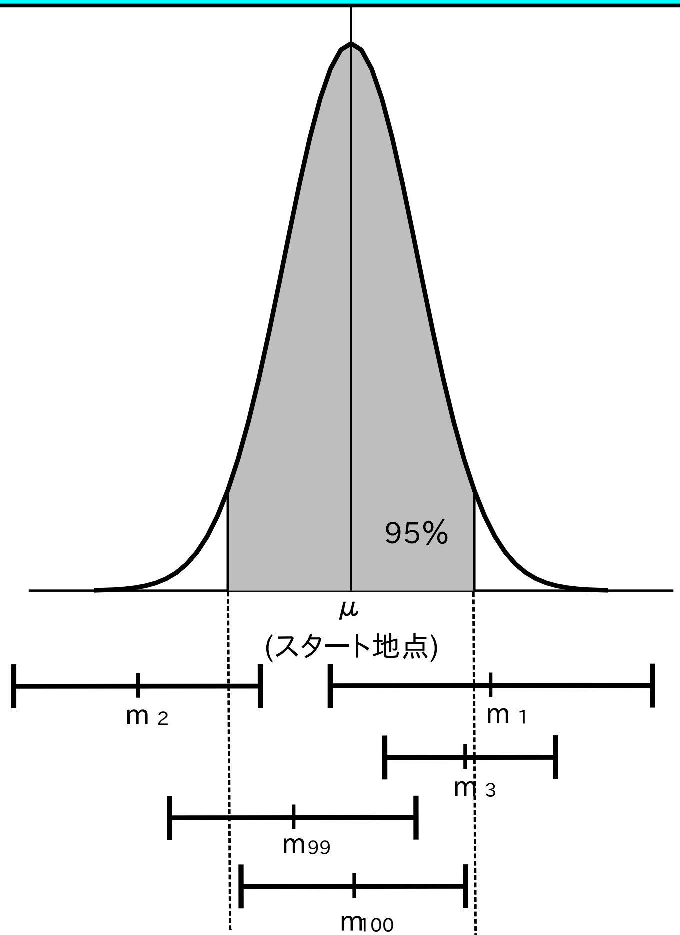


図1.4.4 Neyman-Pearson釣り大会

95%信頼区間：母平均値が標本平均値と同じ値だったら、その母集団から無作為抽出した標本集団の標本平均値の95%のものが含まれる区間

→母平均値が入っている信頼区間を得る確率が95%

信頼区間の解釈−ネイマン・ピアソン流

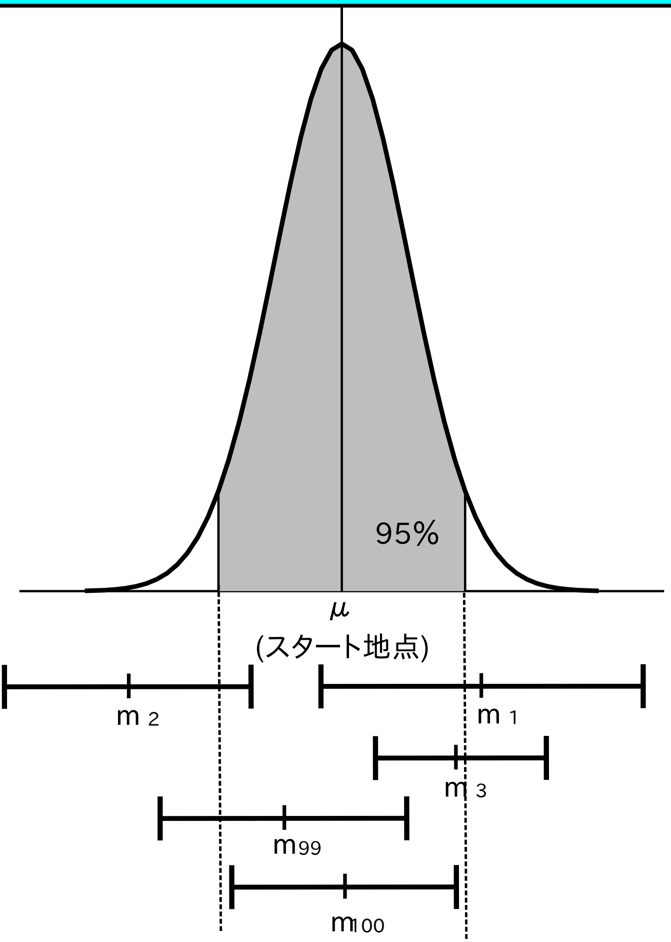


図1.4.4 Neyman-Pearson釣り大会

ネイマン・ピアソン統計学で使う確率は主に一般的な確率＝結果の確率

過去に起きた出来事に基づいてこれから起きる出来事の確率を表す値

ランダム性に基づいて未来に起きる出来事の確からしさを表す頻度的な確率

信頼区間の解釈−ネイマン・ピアソン流

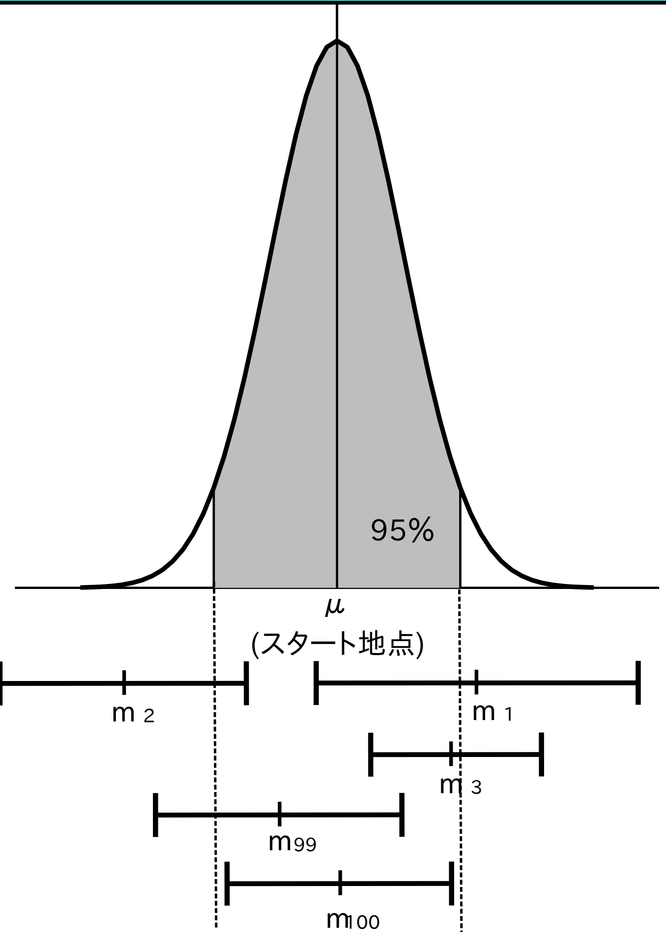


図1.4.4 Neyman-Pearson釣り大会

結果の確率は出来事が起きた後は「0」か「1」に確定する

信頼区間は母集団から標本集団を無作為抽出した後で決まるので母平均値が入っている確率は「0」か「1」

→母平均値が入っている信頼区間を得る結果の確率が95%

信頼区間の解釈−ベイズ流

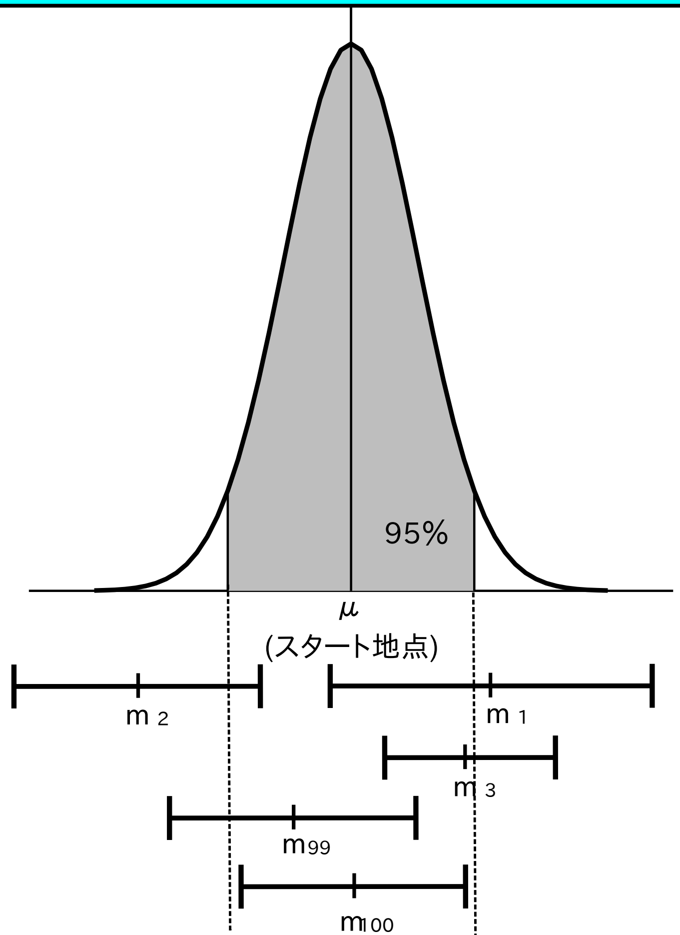


図1.4.4 Neyman-Pearson釣り大会

95%ベイズ信用区間：母平均値が95%の確率で入っている区間

→信用区間に母平均値が入っている確率が95%

信頼区間の解釈–ベイズ流

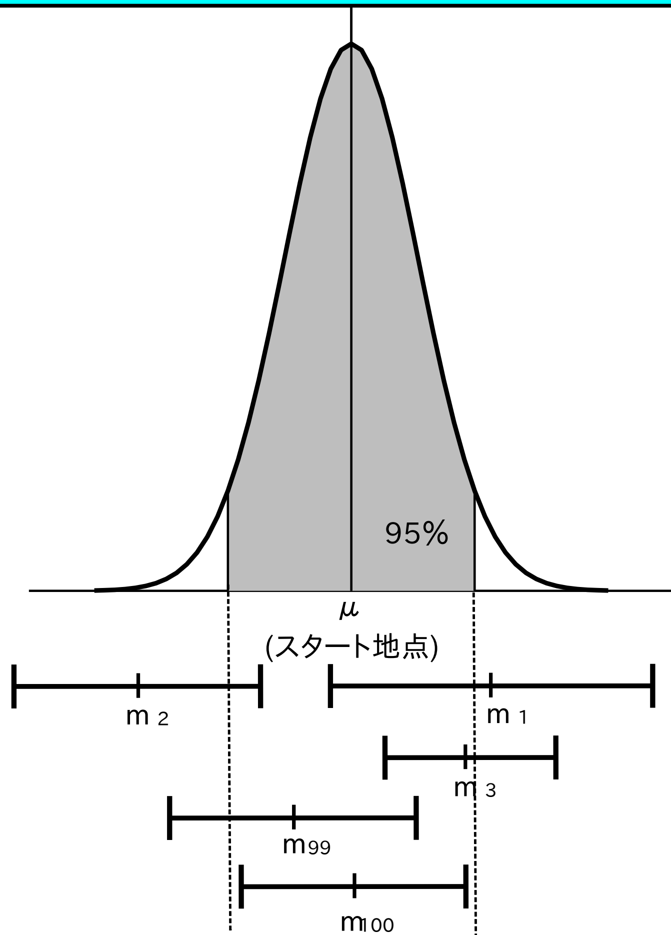


図1.4.4 Neyman-Pearson釣り大会

ベイズ統計学で使う確率は主に逆確率＝原因の確率←ラプラスが定義

現在、起きている出来事に基づいて過去に起きたであろう出来事の確率を表す値

情報不足に基づいて過去に起きた出来事の確からしさを表す主観的な確率

信頼区間の解釈−ベイズ流

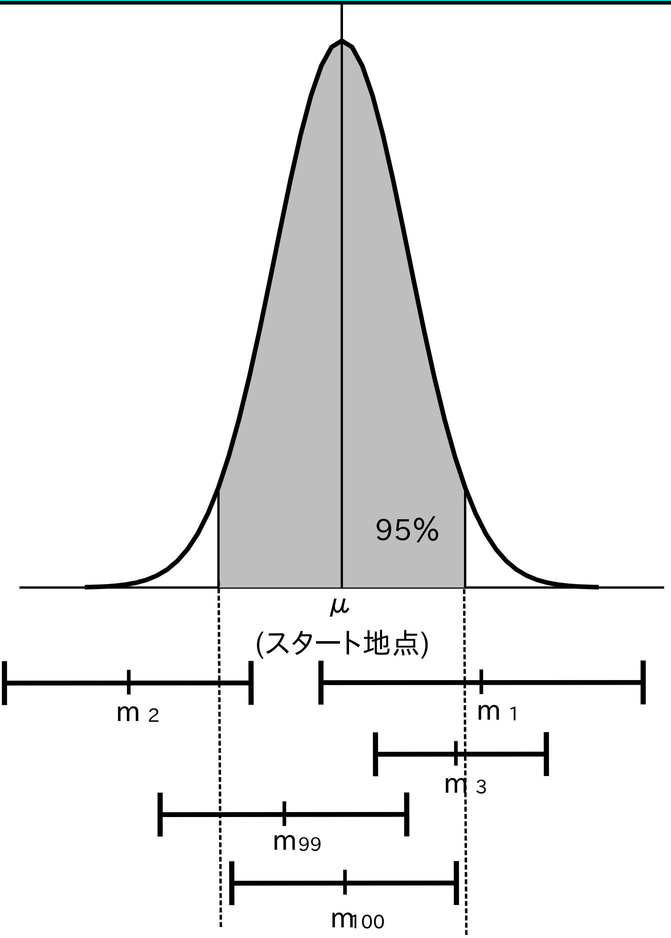


図1.4.4 Neyman-Pearson釣り大会

原因の確率は情報が増えて過去に起きた出来事が特定できれば「0」か「1」に収束する

標本集団の例数をn→∞にすると情報が増えて標準誤差=0になり信用区間が母平均値に収束する

→信用区間に母平均値が入っている原因の確率が95%

信頼区間の解釈−使い分け

発症 → 診断 → 治療

疾患特定
(原因の確率)

治療効果予測
(結果の確率)

感度と特異度は結果の確率で陽性・陰性予測値は原因の確率

治療効果の予測や予後予測は結果の確率

→原因推測と結果予測で信頼区間の解釈を使い分ける

ネイマン・ピアソン統計学とベイズ統計学

ネイマン・ピアソン統計学

事前情報 → 仮説 → 仮説の検証・修正

↑ 科学理論（事前情報 → 仮説）

↑ 観測データ（仮説 → 仮説の検証・修正）

批判的思考法→仮説演繹法による科学法則の検証

パラダイムシフトに対応可→自然科学向き

フィッシャー・ネイマン・ピアソン等が

ベイス統計学を批判して自然科学向きにした近代統計学

ベイズ統計学

事前情報 → 確実性がより高い情報に更新

↑ 観測データ

経験的思考法→経験則の構築

状況の変化に素早く対応可→人文科学向き

パラダイムシフトに対応できないので自然科学向きではない！

推定と検定と検定廃止論

- 中心極限定理
- 区間推定の手順
- 信頼区間の解釈
- 統計的仮説検定の手順
- 有意確率p値の意味
- 検定と推定と科学的判断の関係−検定廃止論

統計的仮説検定の手順-1

問題：日本人の平均体重は50kgか？

帰無仮説H_0:日本人の平均体重は50kgである←問題の答えは○

↕

対立仮説H_1:日本人の平均体重は45kgまたは55kgである←問題の答えは×

1.問題を設定する→科学的意義のある基準値μ_0=50と許容範囲(検出差)$\delta^*=\pm 5$を決める

2.問題の答を○×式で設定する→帰無仮説H_0と具体的な対立仮説H_1を設定する

3.有意水準と検出力を決める→有意水準α=0.05(5%)・検出力(1-β)=0.8(80%)

4.先行研究・探索試験の結果に基づいて結果を予想し、試験の必要例数を求める

→平均値：$\mu \fallingdotseq 55$　$\delta=\mu-\mu_0 \fallingdotseq 5$　標準偏差：$\sigma \fallingdotseq 10$　より　必要例数：n=34

統計的仮説検定の手順-2

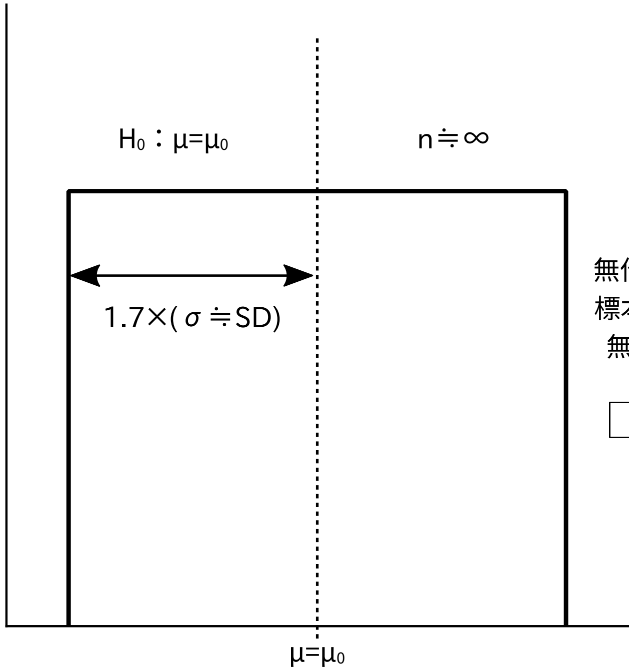


図1.5.2 母集団のデータ分布

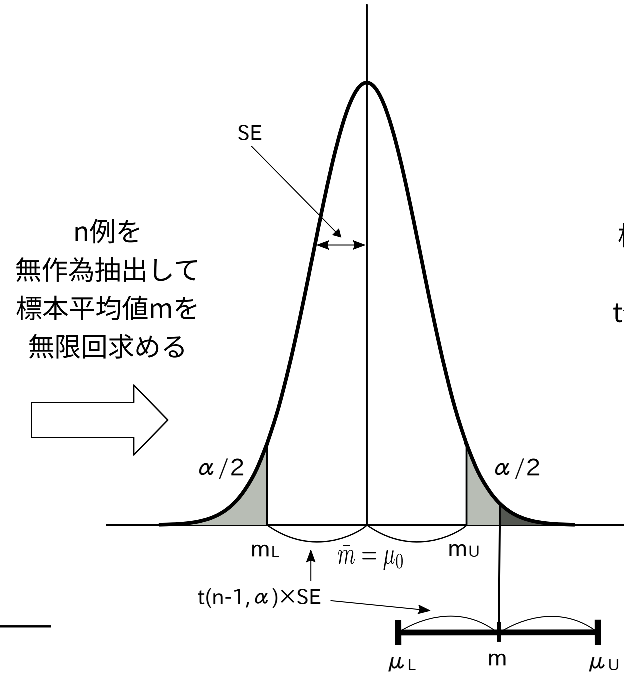


図1.5 帰無仮説の棄却域と信頼区間

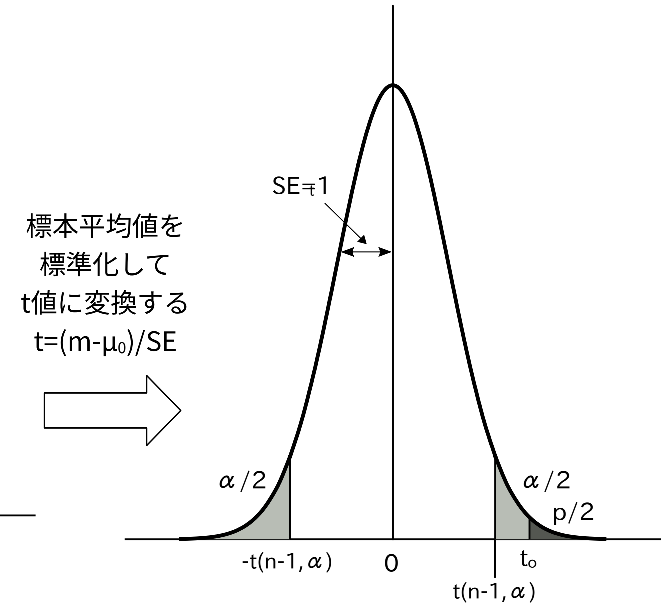


図1.5.4 t分布

5.母集団からn例の標本集団を無作為抽出して標本平均値と標本標準偏差を求める

→n=100　m=60　SD=10

6.帰無仮説が正しいと仮定した時の標本平均値の分布を描く→図1.5

7.標本平均値を$t=(m-\mu_0)/SE$と標準化した時のt値の分布を描く→図1.5.4

8.標本平均値の分布とt値の分布における帰無仮説の棄却域(信頼区間外)を求める

→図1.5と図1.5.4の薄い灰色部分…標本平均値がここに入ったら帰無仮説を棄却

中心極限定理のシミュレーション−平均値

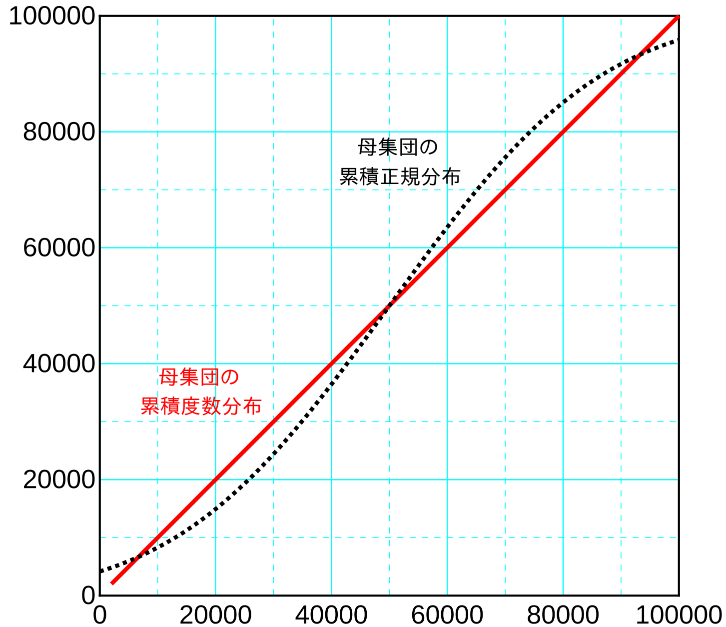


図 付録2.1 母集団:整数

n例を
無作為抽出して
標本平均値を
1万回求める

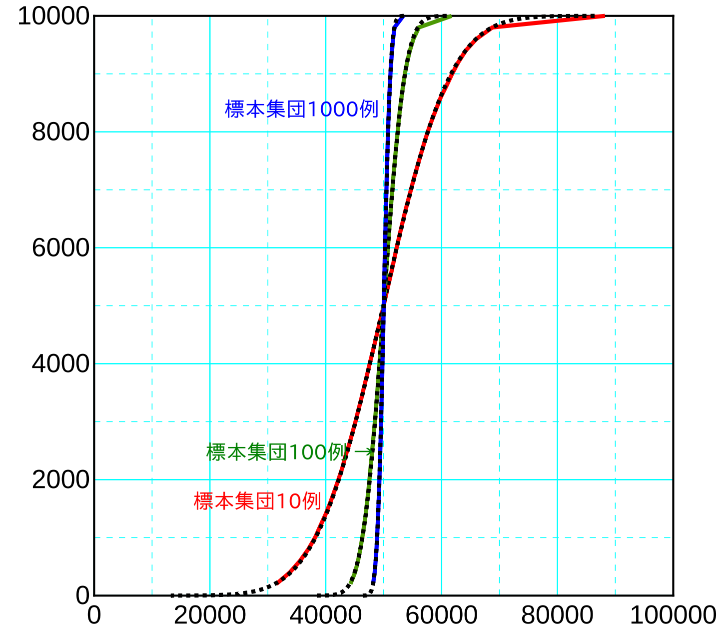


図 付録2.2 標本平均値(母集団:整数)

- 母集団：1から10万までの整数10万個…離散型一様分布
  赤色の実線：実際のデータの累積度数分布(直線)
  黒色の点線：平均値と標準偏差が同じ時の累積正規分布(シグモイド曲線)

- 標本平均値の分布：n例の標本集団を1万回無作為抽出した時の標本平均値1万個
  赤色の実線：10例の標本集団における標本平均値の累積度数分布
  緑色の実線：100例の標本集団における標本平均値の累積度数分布
  青色の実線：1000例の標本集団における標本平均値の累積度数分布
  黒色の点線：平均値と標準誤差が同じ時の累積正規分布

中心極限定理のシミュレーション–平均値

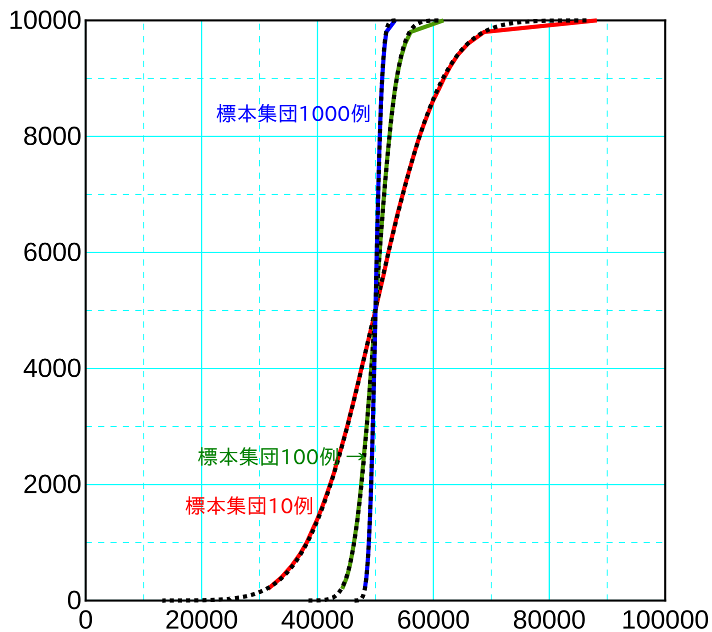


図 付録2.2 標本平均値(母集団:整数)

標本平均値を
標準化して
t値に変換する
t=(m-μ_0)/SE

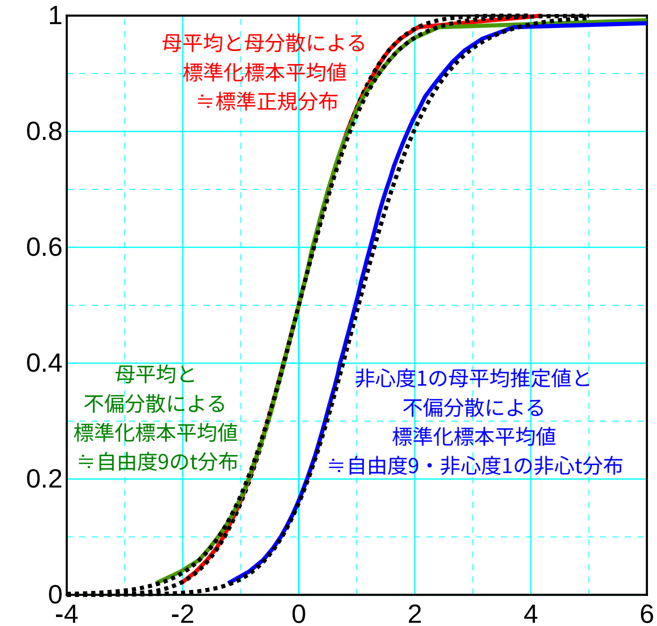


図 付録2.23 標準化標本平均値(母集団：整数)

- t値の分布：10例の標本集団における標準化標本平均値1万個

赤色の実線：母平均値と母標準偏差を用いた標準誤差によって標準化したt値の累積度数分布

黒色の点線：累積標準正規分布

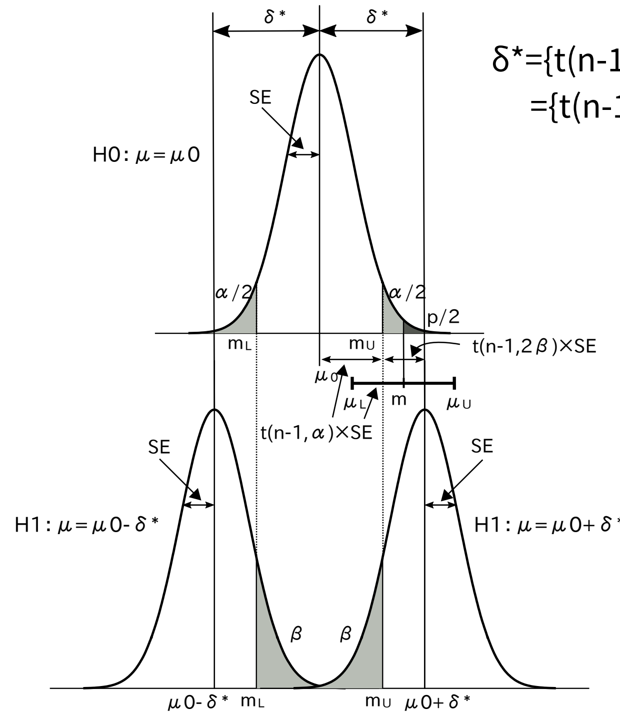
緑色の実線：母平均値と標本標準偏差を用いた標準誤差によって標準化した
t値の累積度数分布←帰無仮説が正しい時のt値の累積分布

黒色の点線：自由度9の累積t分布

青色の実線：(標本平均値＋標準誤差)を母平均値と標本標準偏差を用いた標準誤差によって
標準化したt値の累積度数分布←対立仮説が正しい時のt値の累積分布

黒色の点線：自由度9・非心度1の累積非心t分布

統計的仮説検定の手順-3



$\delta^* = \{t(n-1,\alpha)+t(n-1,2\beta)\}SE$
$= \{t(n-1,\alpha)+t(n-1,2\beta)\}(SD/\sqrt{n})$

図1.14 統計的仮説検定の模式図

9.対立仮説が正しいと仮定した時の標本平均値の分布を描く→図1.14の下2つの分布

10.標本平均値の分布における対立仮説の棄却域を求める

→図1.14の下2つの分布の薄い灰色部分＝帰無仮説の信頼区間(棄却域外)

11.実際のデータを用いて検出力分析を行い実際の検出力が

事前に設定した検出力以上であることを確認する

→n=100　SD=10　δ*=5　より　検出力：(1-β’)=0.999 ＞ (1-β)=0.8

統計的仮説検定の手順-3’

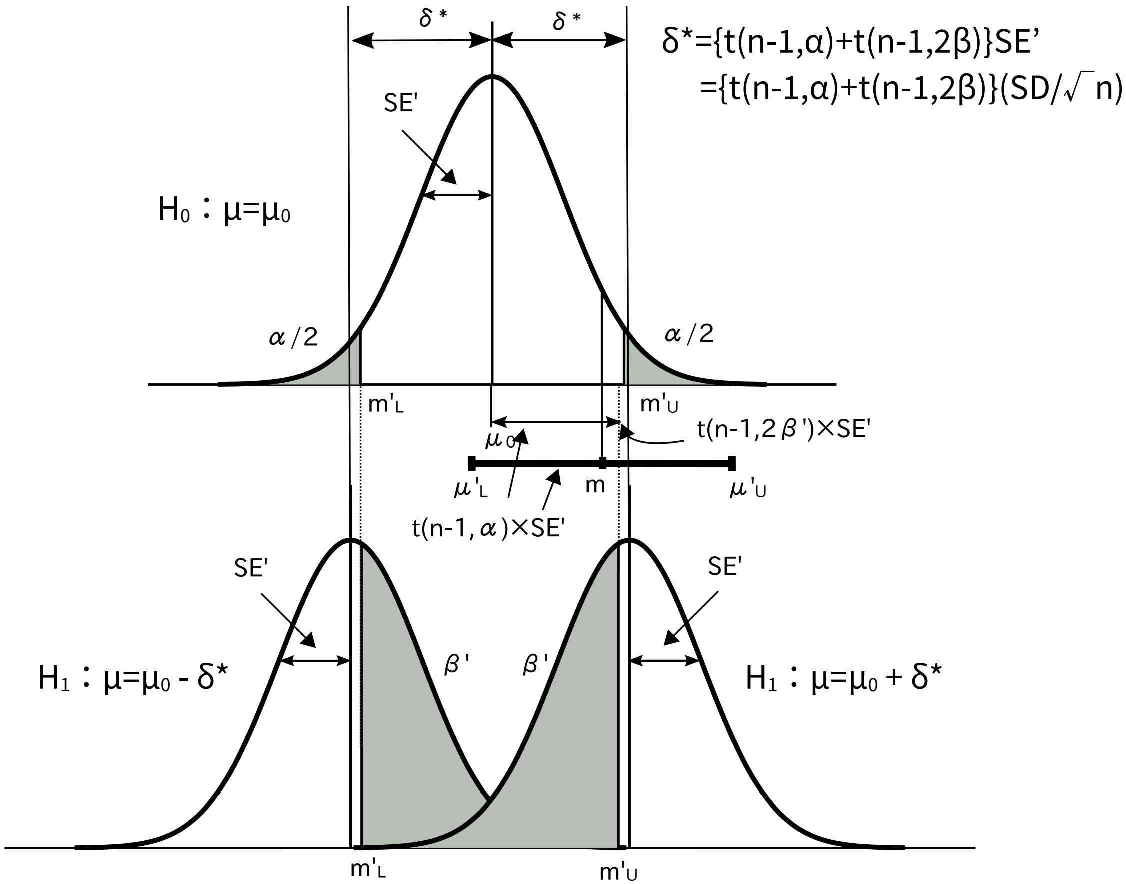


図1.6.8 統計的仮説検定の実際の模式図

実際の検出力(1-β’)が事前に設定した検出力(1-β)未満の時は試験失敗！

→例数を増やして再試験

統計的仮説検定の手順-4

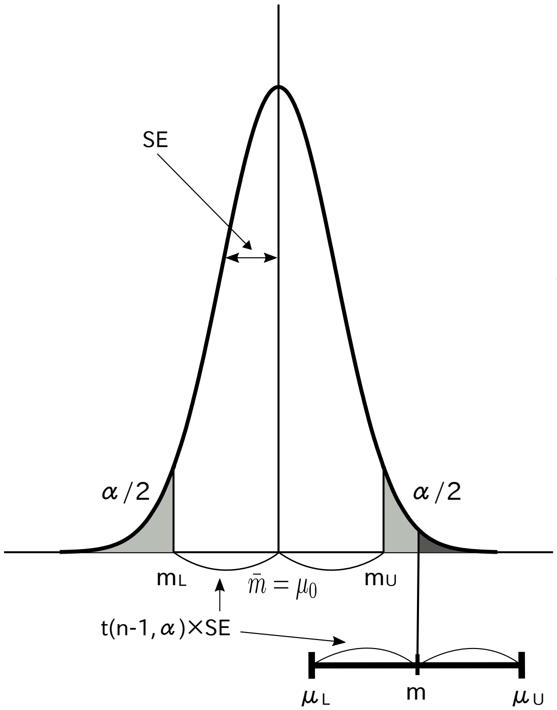


図1.5 帰無仮説の棄却域と信頼区間

標本平均値を
標準化して
t値に変換する
$t=(m-\mu_0)/SE$

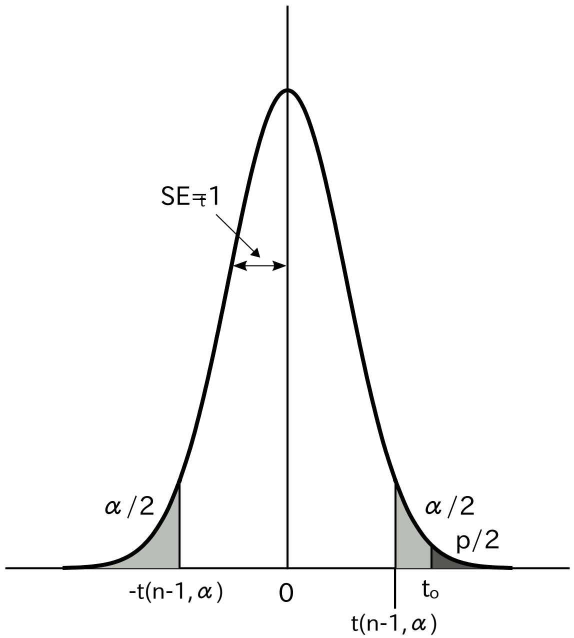


図1.5.4 t分布

12.標本平均値が帰無仮説の棄却域に入っているかどうか調べる

○方法1：標本平均値mと棄却域の境界値m_Lおよびm_Uを比較する

＝信頼区間に基準値μ_0が入っているかどうか調べる

→$m_L \leqq m \leqq m_U$なら棄却域に入っていない＝信頼区間μ_L～μ_Uにμ_0が入っている…有意ではない

→$m < m_L$ または $m_U < m$なら棄却域に入っている＝信頼区間μ_L～μ_Uにμ_0が入っていない……有意

統計的仮説検定の手順-4

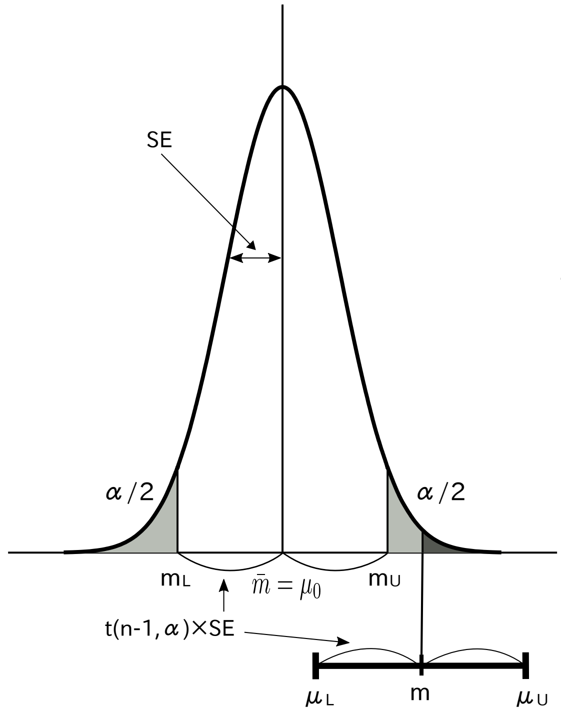
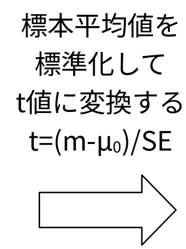


図1.5 帰無仮説の棄却域と信頼区間



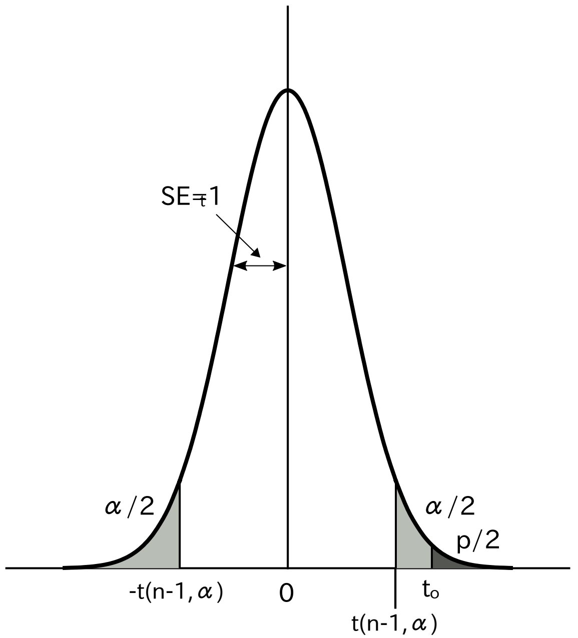


図1.5.4 t分布

12.標本平均値が帰無仮説の棄却域に入っているかどうか調べる

○方法2：標本平均値mを標準化したt_oと棄却域の境界値t_L=-t(n-1,α)およびt_U=t(n-1,α)を比較する

＝|t_o|がt(n-1,α)よりも大きいかどうか調べる…昔の検定法

→|t_o|≦t(n-1,α)なら棄却域に入っていない…有意ではない

→|t_o|＞t(n-1,α)なら棄却域に入っている……有意

統計的仮説検定の手順-4

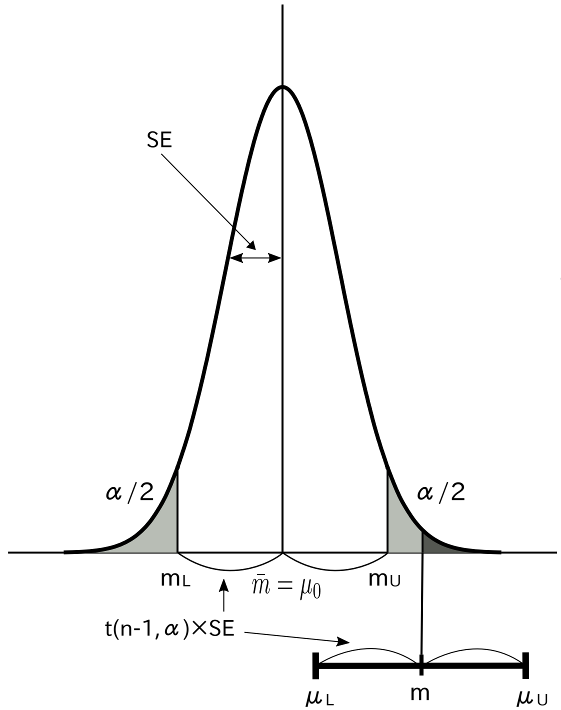


図1.5 帰無仮説の棄却域と信頼区間

標本平均値を
標準化して
t値に変換する
$t=(m-\mu_0)/SE$

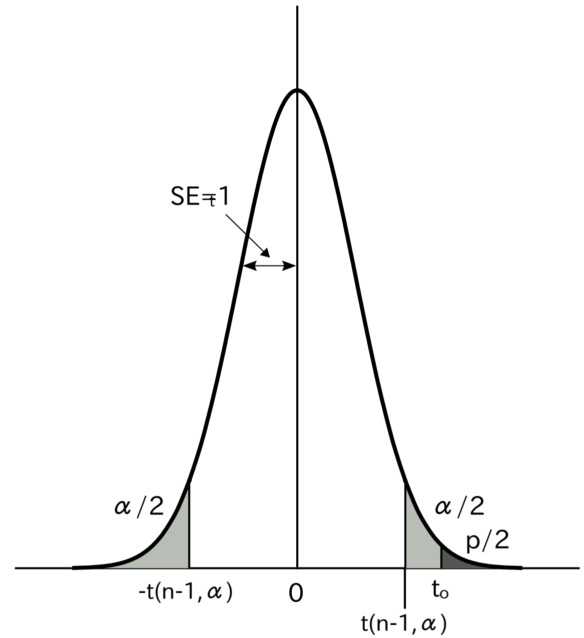


図1.5.4 t分布

12.標本平均値が帰無仮説の棄却域に入っているかどうか調べる

○方法3：t_oから右側の部分の面積p/2(図1.5.4の濃い灰色部分)とα/2を比較する

＝(t分布においてt_o→∞まで積分した値)×2=p値がαよりも小さいかどうか調べる

→p≧αなら棄却域に入っていない…有意ではない

→p＜αなら棄却域に入っている……有意

統計的仮説検定の手順-5

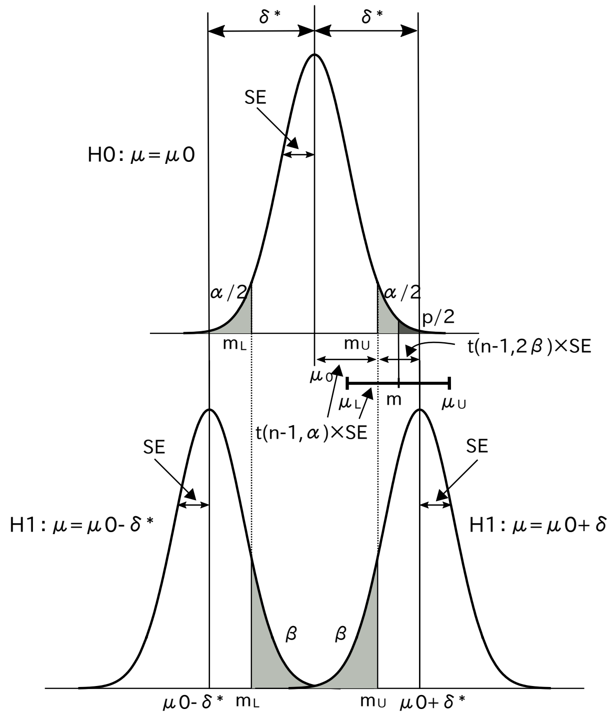


図1.14 統計的仮説検定の模式図

13.検定結果に基づいて統計学的結論を下す

○有意の時：帰無仮説を否定した仮説を統計学的結論として採用する

→$\mu \neq \mu_0$…日本人の平均体重は50kgではない(定性的な結論)←間違っている(原因の)確率はα未満

○有意ではない時：対立仮説を否定した仮説を統計学的結論として採用する

→$\mu_0 - \delta^* < \mu < \mu_0 + \delta^*$…日本人の平均体重は45kgよりも重く55kgよりも軽い(実質時に50kgと同等)

↑間違っている(原因の)確率はβ以下

統計的仮説検定の手順-6

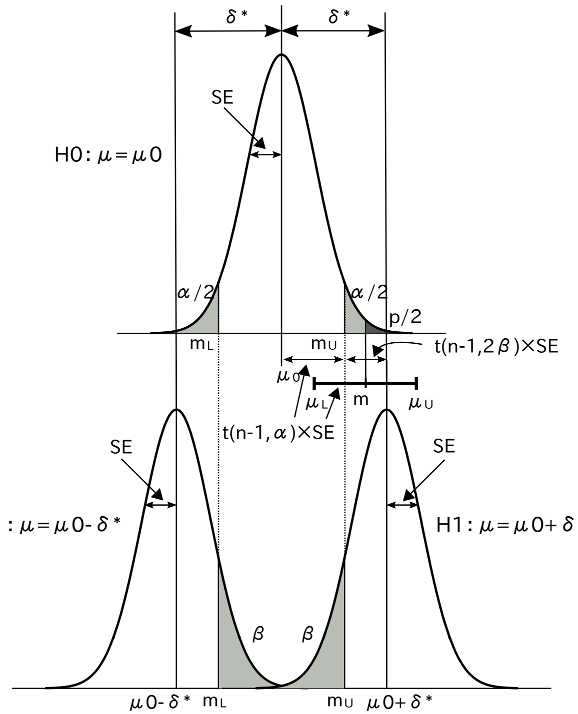


図1.14 統計的仮説検定の模式図

14.区間推定結果と統計学的結論を科学的に評価して科学的結論を下す

○有意の時：日本人の平均体重は60±2=58～62kgである(原因の)確率が高く

　　これは50kgよりも10kgほど重いので日本人は肥満傾向がある

○有意ではない時：日本人の平均体重は45kg～55kgの範囲である(原因の)確率が高く

　　これは医学的には50kgと同等なので日本人は痩せ傾向も肥満傾向もない

推定と検定と検定廃止論

- 中心極限定理
- 区間推定の手順
- 信頼区間の解釈
- 統計的仮説検定の手順
- 有意確率p値の意味
- 検定と推定と科学的判断の関係−検定廃止論

有意確率p値の意味

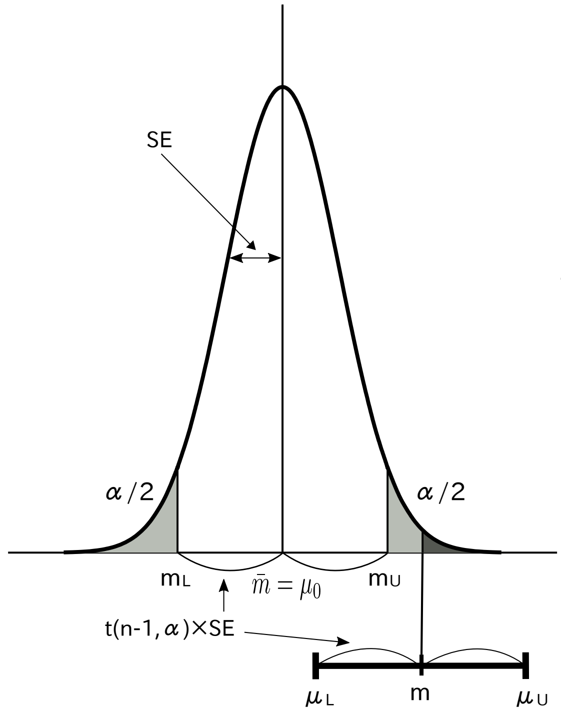
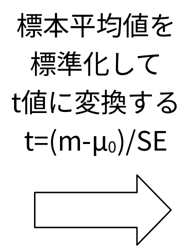


図1.5 帰無仮説の棄却域と信頼区間



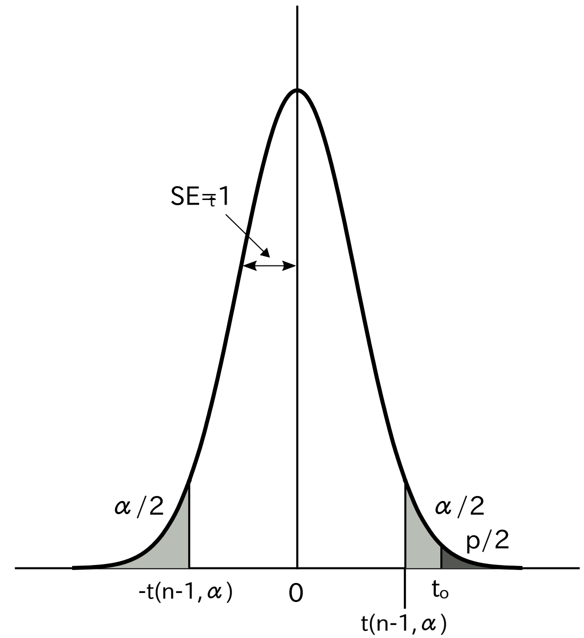


図1.5.4 t分布

p：標本平均値が帰無仮説の棄却域に入っているかどうかを調べるための便宜的な値

p/2：帰無仮説が正しい時に実際の標本平均値以上の標本平均値が得られる確率

は正しいが

p：帰無仮説が正しい時に実際の標本平均値以上に極端な値が得られる確率

は正しいとは限らない！

改善率の2群間比較

群	非改善(%)	改善(%)	計(%)
プラセボ投与群	16(80)	4(20)	20(100)
薬剤A投与群	4(40)	6(60)	10(100)
計	20(67)	10(33)	30(100)

・Fisherの正確検定

片側検定(片側棄却域)：p/2=0.0387 ＞ 0.025 … 片側有意水準2.5%で有意ではない

両側検定(両側棄却域)：p=2×0.0387=0.0774 ＞ 0.05 … 両側有意水準5%で有意ではない

極端な結果が得られる両側有意確率：$p=p_l + p_u=0.0448 < 0.05$ … 両側有意水準5%で有意

↑RやSASやSPSSが採用している両側有意確率計算法←検定結果を誤判定してしまう！

・χ^2検定(連続修正有り)

$\chi_o^2=3.169(p=0.0751) < \chi2(1,0.05)=3.841$ … 有意水準5%で有意ではない

・2群の改善率の差=0.4(40%)　95%信頼区間：下限=-0.03(-3%)　上限=0.83(83%)

改善率の2群間比較

群	非改善(%)	改善(%)	計(%)
プラセボ投与群	16(80)	4(20)	20(100)
薬剤A投与群	4(40)	6(60)	10(100)
計	20(67)	10(33)	30(100)

帰無仮説 H_0：プラセボ投与群の改善率＝薬剤A投与群の改善率=33%

帰無仮説が正しい時に得られる可能性がある結果は次の11種類

群	非改善	改善	計
P	10	10	20
A	10	0	10
計	20	10	30

群	非改善	改善	計
P	11	9	20
A	9	1	10
計	20	10	30

群	非改善	改善	計
P	12	8	20
A	8	2	10
計	20	10	30

群	非改善	改善	計
P	13	7	20
A	7	3	10
計	20	10	30

群	非改善	改善	計
P	14	6	20
A	6	4	10
計	20	10	30

群	非改善	改善	計
P	15	5	20
A	5	5	10
計	20	10	30

群	非改善	改善	計
P	16	4	20
A	4	6	10
計	20	10	30

群	非改善	改善	計
P	17	3	20
A	3	7	10
計	20	10	30

群	非改善	改善	計
P	18	2	20
A	2	8	10
計	20	10	30

群	非改善	改善	計
P	19	1	20
A	1	9	10
計	20	10	30

群	非改善	改善	計
P	20	0	20
A	0	10	10
計	20	10	30

超幾何分布の棄却域と有意確率

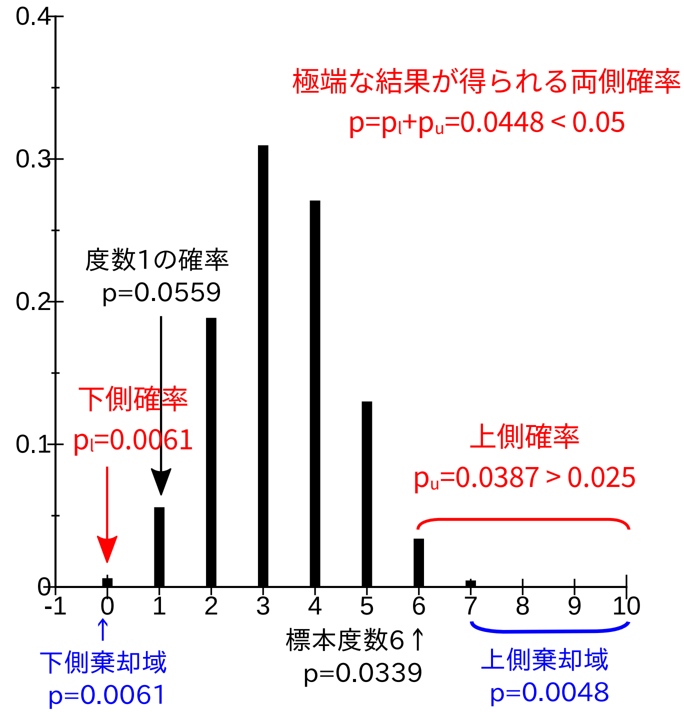


図1.6.6 非対称離散分布の棄却域と両側確率

薬剤A投与群の改善度数の分布：図1.6.6のような超幾何分布(非対称離散分布)

両側棄却域：分布の左右でそれぞれ累積確率が0.025以上にならないところまでの度数

　下側棄却域：度数0のみ(p=0.0061)　上側棄却域：度数7～10(p=0.0048)

標本度数6の上側確率：$p_u=0.0387 > 0.025$ … 有意ではない＝標本度数6は棄却域に入っていない

下側確率：$p_l=0.0061$(累積確率がp_u以上にならない度数まで累積した値)

試験結果以上に極端な結果が得られる両側確率：$p=p_l+p_u=0.0061+0.0387=0.0448 < 0.05$ … 有意

改善率の2群間比較

群	非改善(%)	改善(%)	計(%)
プラセボ投与群	16(80)	4(20)	20(100)
薬剤A投与群	4(40)	6(60)	10(100)
計	20(67)	10(33)	30(100)

・Fisherの正確検定

片側検定(片側棄却域)：p/2=0.0387 ＞ 0.025 … 片側有意水準2.5%で有意ではない

両側検定(両側棄却域)：p=2×0.0387=0.0774 ＞ 0.05 … 両側有意水準5%で有意ではない

極端な結果が得られる両側有意確率：$p=p_l+p_u=0.0448$ ＜ 0.05 … 両側有意水準5%で有意

↑RやSASやSPSSが採用している両側有意確率計算法←検定結果を誤判定してしまう！

・χ^2検定(連続修正有り)

$\chi_o^2=3.169(p=0.0751)$ ＜ χ2(1,0.05)=3.841 … 有意水準5%で有意ではない

・2群の改善率の差=0.4(40%)　95%信頼区間：下限=-0.03(-3%)　上限=0.83(83%)

Fisherの正確検定は片側検定用と割り切り、両側検定にはχ^2検定を用いるのが無難！

→p値は試験結果が帰無仮説の棄却域に入っているかどうかを調べるための便宜的な値

推定と検定と検定廃止論

- 中心極限定理
- 区間推定の手順
- 信頼区間の解釈
- 統計的仮説検定の手順
- 有意確率p値の意味
- 検定と推定と科学的判断の関係–検定廃止論

検定結果だけから科学的な判断をするのは危険

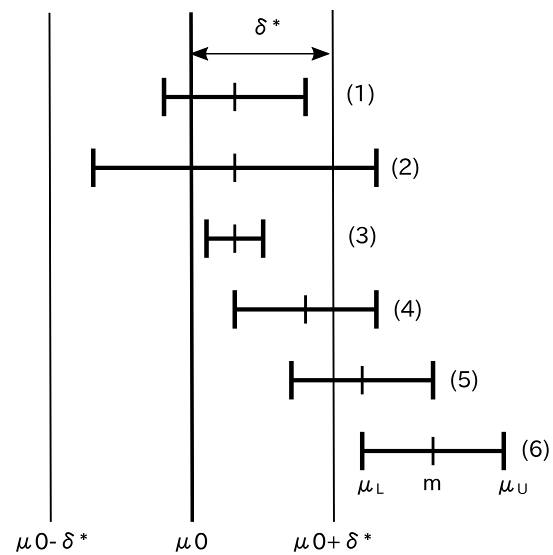


図1.15 検定結果と信頼区間

	検定結果	推定結果	実質科学的な判断
(1)	有意ではない	$\mu \fallingdotseq \mu_0$	母平均値は基準値とほぼ等しい
(2)	有意ではない	$\mu = \mu_0 \sim \mu_0 + \delta^*$	この結果だけでは判断できない 信頼区間をもっと狭くする必要がある(例数を増やす)
(3)	有意	$\mu_0 < \mu < \mu_0 + \delta^*$	母平均値は基準値と実質的に変わらない
(4)	有意	$\mu \fallingdotseq \mu_0 + \delta^*$	母平均値は基準値と実質的に変わらない可能性が高い
(5)	有意	$\mu \fallingdotseq \mu_0 + \delta^*$	母平均値は基準値よりも大きい可能性が高い
(6)	有意	$\mu_0 + \delta^* < \mu$	母平均値は基準値よりも大きい

※生物学的同等性試験では推定結果を重視し、検定結果は参考程度 → 検定廃止論

例数が多いと検定結果は科学的な判断には使えない

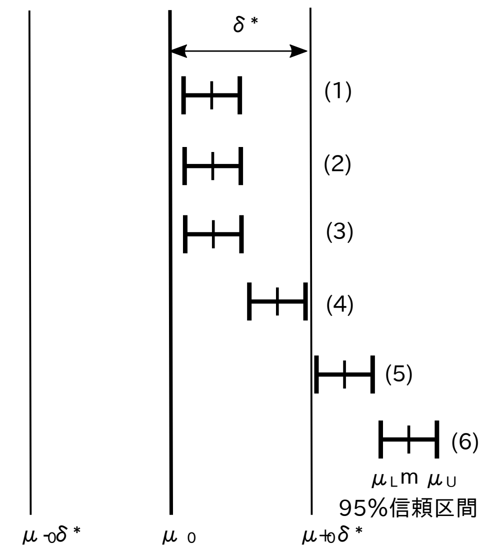


図1.7.6 例数が多い時の検定結果と信頼区間

	検定結果	推定結果	実質科学的な判断
(1)	有意	$\mu_0 < \mu < \mu_0 + \delta^*$	母平均値は基準値と実質的に変わらない
(2)	有意	$\mu_0 < \mu < \mu_0 + \delta^*$	母平均値は基準値と実質的に変わらない
(3)	有意	$\mu_0 < \mu < \mu_0 + \delta^*$	母平均値は基準値と実質的に変わらない
(4)	有意	$\mu_0 < \mu < \mu_0 + \delta^*$	母平均値は基準値と実質的に変わらない
(5)	有意	$\mu_0 + \delta^* < \mu$	母平均値は基準値よりも大きい
(6)	有意	$\mu_0 + \delta^* < \mu$	母平均値は基準値よりも大きい

※生物学的同等性試験では推定結果を重視し、検定結果は参考程度 → 検定廃止論

有意症・有意症症候群は難治性疾患

有意症(significantosis)・有意症症候群(significant syndrome)

「有意差あり＝実質科学的に有意義な差がある」とか

「有意差なし＝実質科学的に有意義な差がない」と誤解する難治性の疾患

医学界や厚生労働省等で大流行中！

有意症・有意症症候群の予防策

- 推定結果を重視して検定結果は参考程度にする
- 検定を廃止して推定を用いる(検定廃止論)

本日の結語

検定結果ではなく

推定結果に基づいて

科学的判断をしましょう！

ご清聴ありがとうございました