医学統計勉強会

第2回　推定　信頼区間　仮説検定

帝京大学臨床研究センター（TARC）・帝京大学大学大学院公衆衛生学研究科　共催

帝京大学大学大学院公衆衛生学研究科

宮田　敏

パラメターの推定

データ解析では，様々なパラメターが推定される.

- 期待値　$\mu = E(X)$
- 割合（成功確率）$p = P(X=1),\ X = \begin{cases} 1 : \text{event occurrence} \\ 0 : \text{no event} \end{cases}$
- 回帰係数　$Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \epsilon_i$
- オッズ比：ロジスティック回帰モデル
  e^{β_j} : $\log(p_i/(1-p_i)) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}$
- ハザード比：Cox比例ハザードモデル
  e^{β_j} : $h(t|x_i) = h_0(t)\exp\{\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}\}$

定義:
点推定：未知母数の推定問題における，標本から計算される推定量の値による推定

区間推定：真の母数を含む確率がある値以上になるように，区間を求めることによる推定

点推定

Parameter (Estimand)	推定量(Estimator)
μ : 母集団平均	$\overline{X}$：標本平均
σ^2 : 母集団分散	S^2: 標本分散
p : 成功確率	X/n: X=成功回数

パラメターは，サンプルから計算した推定量（estimator）を用いて推定される.

- 期待値 ⇒ 標本平均 $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
- 割合 ⇒ $\hat{p} = \frac{X}{n}$ X：イベント数，n：試行回数

その他のモデルの推定量も，統計ソフトで計算できる.
推定量の観測値 = 推定値（estimate）をもって，パラメターの値を推定する.

推定量はサンプルから計算されるが，サンプルには観測誤差が含まれる.

例：ある疾患の罹患率を調べるとする．$n=10$ 人のサンプルを調べたところ，疾患に罹患していたのは $x=0$ 人であった．罹患率の推定値は 0 である．

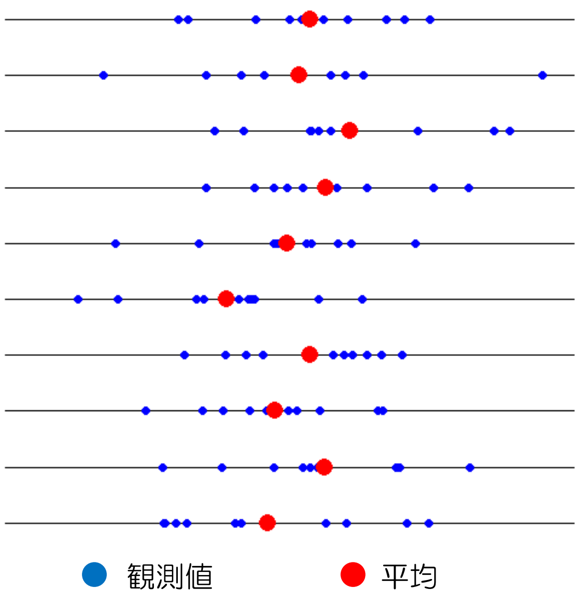
$$\hat{p} = 0/10 = 0$$

このとき，その疾患の罹患率は 0% である，と結論づけてよいのであろうか？

例えば本当の罹患率は 3% であるのに，たまたま $n=10$ 人の中に罹患した人がいなかっただけではないのか？

もう一度別の10人を調べたら，今度は一人罹患者がいて $\hat{p} = 1/10 = 0.1$ となることも有るのでは無いか？

同じ母集団からサンプリングを繰り返した時の，観測値と平均



推定量はサンプルから計算されるが，サンプルには確率的な変動や観測誤差が含まれる．

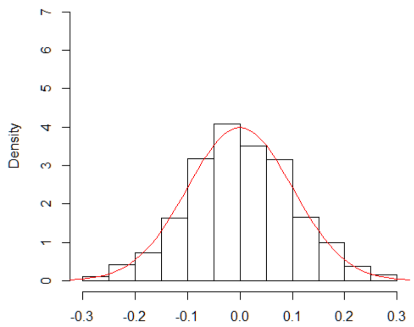
⇒　推定量（この場合は平均）にも，観測誤差=分散がある．

⇒　サンプリングのたびに，推定値にもばらつきがある．

⇒　しかし，推定量の分散はサンプルの分散より小さい．
　サンプル数 n が増えれば，分散が小さくなる．

　推定量の観測値=推定値からは，推定の精確さは評価できない．
　サンプル数が大きくなると，推定量の分散あるいは標準誤差(standard error) は小さくなる．

標本平均の分布 (n=100)

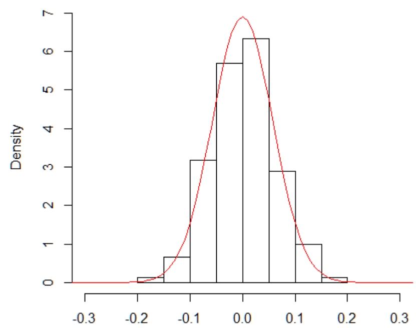


サンプル数の増加

標本平均の分散

$$\frac{\sigma^2}{n}$$

標本平均の分布 (n=300)



2. 信頼区間

点推定量は未知パラメター θ に対して数値的な情報を与えるが，推定の正確さ，信頼度については何の情報も与えない．
　推定の信頼度を示すような，「もっともらしい」推定値の区間（集合）が有用である．

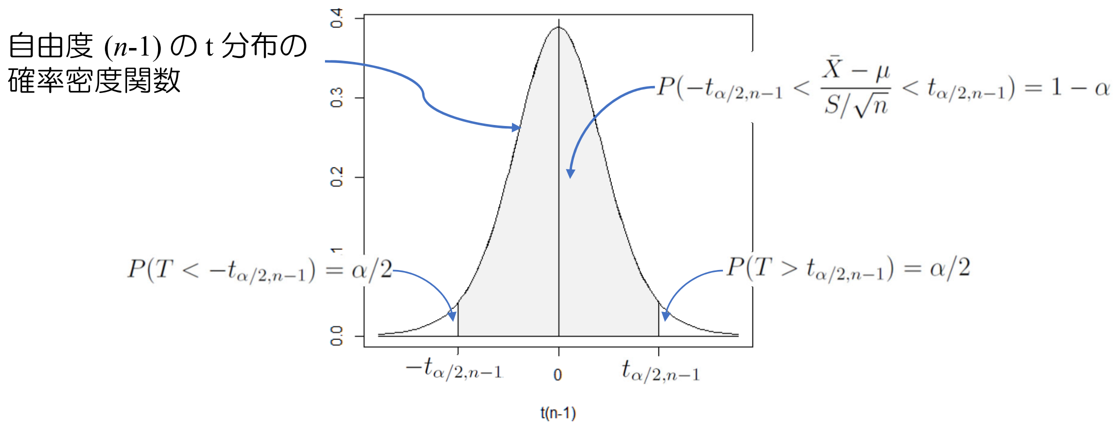
Definition: Parameter θに対して，以下の条件を満たす確率的な区間(二つの確率変数の組) $(l(X_1,\ldots,X_n), u(X_1,\ldots,X_n))$を区間推定量という．

$$P(l(X_1,...,X_n) \leq \theta \leq u(X_1,...,X_n)) = 1-\alpha.$$

信頼水準 (confidence level) $(1 - \alpha)$ の信頼区間 (confidence intervall, CI)とは，区間推定量の $(X_1,\ldots,X_n)$ に観測値 $(x_1,\ldots,x_n)$ を代入した区間である．

例： $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$ 独立かつ同一の期待値 μ 分散 σ^2 の正規分布に従うとする.
この時，以下の T は自由度 (n-1) の t 分布に従うことが知られている.

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$$



信頼区間の構成

1. 以下の条件を満たすピボット確率変数 (pivotal random variable) $h(X_1,\ldots,X_n:\theta)$ を見つける:
   - $h(X_1,\ldots,X_n:\theta)$ は標本 $X_1,\ldots,X_n$ と未知母数 θ に依存する.
   - $h(X_1,\ldots,X_n:\theta)$ の確率分布は θ に依存しない.
2. 以下の条件を満たす $h(X_1,\ldots,X_n:\theta)$ の percentile, a, b を求める:

$$P(a < h(X_1,\ldots,X_n) < b) = 1-\alpha. \quad (1)$$

3. (1) を変形して信頼区間を求める:

$$P(l(X_1,\ldots,X_n) \le \theta \le u(X_1,\ldots,X_n)) = 1-\alpha.$$

平均の信頼区間の導出

$$P\left(-t_{\alpha/2,n-1} < \frac{\bar{X}-\mu}{S/\sqrt{n}} < t_{\alpha/2,n-1}\right) = 1-\alpha$$

$$-t_{\alpha/2,n-1} < \frac{\bar{X}-\mu}{S/\sqrt{n}} < t_{\alpha/2,n-1}$$ 左辺括弧内の不等式

$$\iff -t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} < \bar{X}-\mu < t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}$$ 各項に $S/\sqrt{n}$ を掛ける

$$\iff -\bar{X}-t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} < -\mu < -\bar{X}+t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}$$ 各項から $\bar{X}$ を引く

$$\iff \bar{X}+t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} > \mu > \bar{X}-t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}$$ 各項に-1 を掛ける

$$P\left(\bar{X}-t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} < \mu < \bar{X}+t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right) = 1-\alpha$$

$$\mathrm{CI}:(\bar{x}-t_{\alpha/2,n-1}\times s/\sqrt{n}, \bar{x}+t_{\alpha/2,n-1}\times s/\sqrt{n})$$

2. 信頼区間

推定量を標準誤差（あるいはその推定量）で割ったものは，標準正規分布（あるいは t 分布）などに従う．

期待値の信頼区間 $\left(\bar{X}-t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}, \bar{X}+t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right)$

サンプル数 n が大きくなると，信頼区間の幅は小さくなる．

推定量と信頼区間の組み合わせで，観測誤差も含めたパラメターの量的な推測が可能になる．

3. 仮説検定

帰無仮説 (Null hypothesis)： H_0 母集団に関する命題で，検定の当初，真であると仮定されるもの.

対立仮説 (Alternative hypothesis) : H_1 帰無仮説 H_0 と対立する仮説.

Type I error : H_0 が真であるとき H_0 を棄却する過誤

Type II error : H_0 が偽であるとき H_0 を棄却しない過誤

有意水準(Significance level) α : 事前に決められた定数で，検定は P(type I error) = αとなるように定められる

	H_0 is true	H_1 is true
Accept H_0	○	Type II error
Reject H_0	Type I error	○

Type I errorとType II errorの確率を同時に小さくすることは不可能．仮説検定はP(type I error) = α となる範囲の中で，type II errorの確率が最も小さくなるようにデザインされている.

検定統計量 (Test statistic) : 標本から計算される統計量で，それを用いて検定の決定を行うもの.

棄却域 (Rejection region) : 検定統計量がとりうる値の集合のうち，H_0 が棄却される領域.

例：　$X_1,\ldots,X_n \overset{\text{iid}}{\sim} N(\mu,\sigma^2)$

$$H_0: \mu = \mu_0 \text{ vs. } H_1: \mu = \mu_1 (> \mu_0)$$

検定統計量

$$T = \frac{\bar{X}-\mu_0}{S/\sqrt{n}} = \underbrace{\frac{\bar{X}-\mu}{S/\sqrt{n}}}_{\sim t(n-1)} + \frac{\mu-\mu_0}{S/\sqrt{n}} \begin{cases} = 0: \text{ under } H_0 \\ > 0: \text{ under } H_1 \end{cases}$$

信頼区間を導出したときの
$T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$
と似ている.

すなわち，検定統計量は帰無仮説の元では t 分布に従い，対立仮説の元では t 分布を正の方向にシフトした分布に従う. $H_1: \mu = \mu_1 (> \mu_0)$ なる対立仮説の元では，検定統計量は「大きい」値をとりがちである.

例: 母集団平均 (σ^2 既知, 有意水準α)

Null hypothesis $H_0: \mu = \mu_0$,

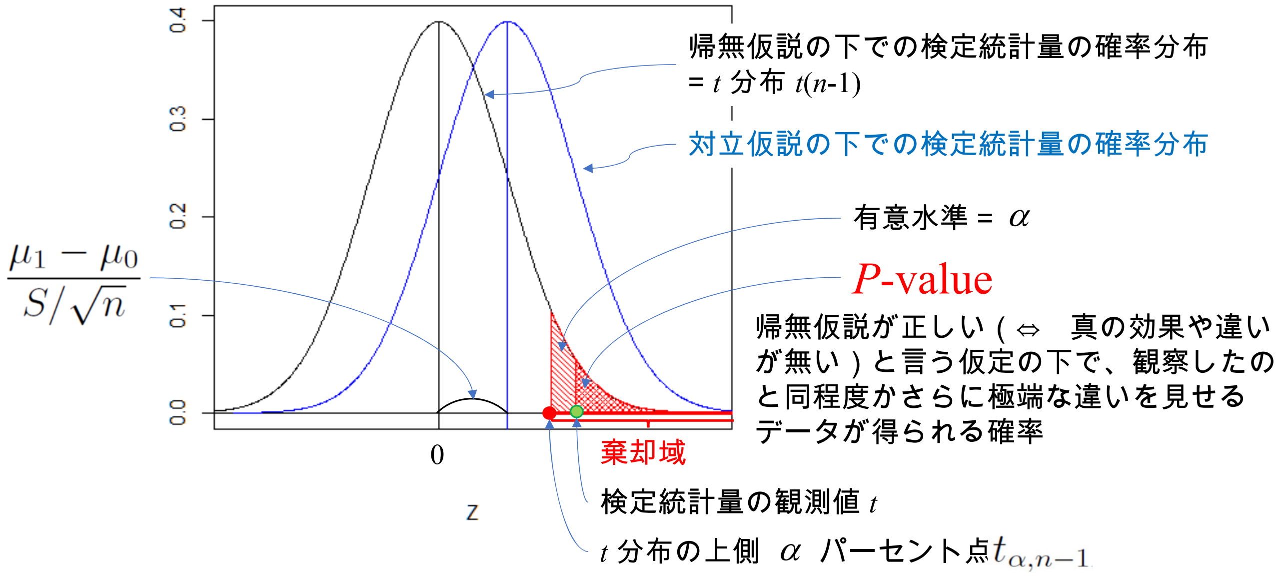
Test statistic $Z = (\bar{X} - \mu_0)/(\sigma/\sqrt{n})$

$Z \sim N(0,1)$ under H_0.

検定統計量は，信頼区間を構成するための pivotal random variable と同じ形をしている. ただし， pivotal random variable における未知パラメターは，H_0で仮定されたものに置き換えられる.

H_0の下で，検定統計量が値を取りやすい領域が採択域に，H_1の下で値を取りやすい領域が棄却域になる.

3.



$$H_0 : \mu = \mu_0 \text{ vs. } H_1 : \mu = \mu_1 (> \mu_0)$$

3. 仮説検定（つづき）

仮説検定は，帰無仮説を棄却するか否かを結論するだけ．仮説検定で，効果 (effect) の大きさを量的に評価することは出来ない．

単純に，サンプル数を増やせば帰無仮説を棄却しやすく，p値が小さくなる．帰無仮説が棄却されたからと言って，効果量 (effect size) が大きいことを意味しない．

帰無仮説と対立仮説の違いが臨床的に意味のある違いであるかは，推定量，信頼区間を用いて別途評価する必要がある．