Single-equation regression models

Elementary statistics

（単一方程式モデルで必要な初歩的な統計学）

(1) Desirable properties of estimators（推定量の望ましい属性）
(2) Probability distribution（確率分布）
(3) Hypothesis testing and confidence intervals（仮説検定と信頼区間）
(4) Glossary
(5) References

(1.0) Desirable properties of estimators（推定量の望ましい属性）

以後、例えば、回帰式$y=\alpha+\beta x$の係数βを推定するといったことを考える。そこで回帰分析の結果、求められたβの推定量が望ましいものとなるために必要な性質は以下の４つである。

1. Lack of bias（不偏性）
2. Efficiency（有効性）
3. Minimum Mean square error（ＭＳＥの最小）
4. Consistency（一致性）

(1.1) Lack of bias（不偏性）

＜仮定＞

（実際のβの値）＝β、（推定されたβ）＝$\hat{\beta}$

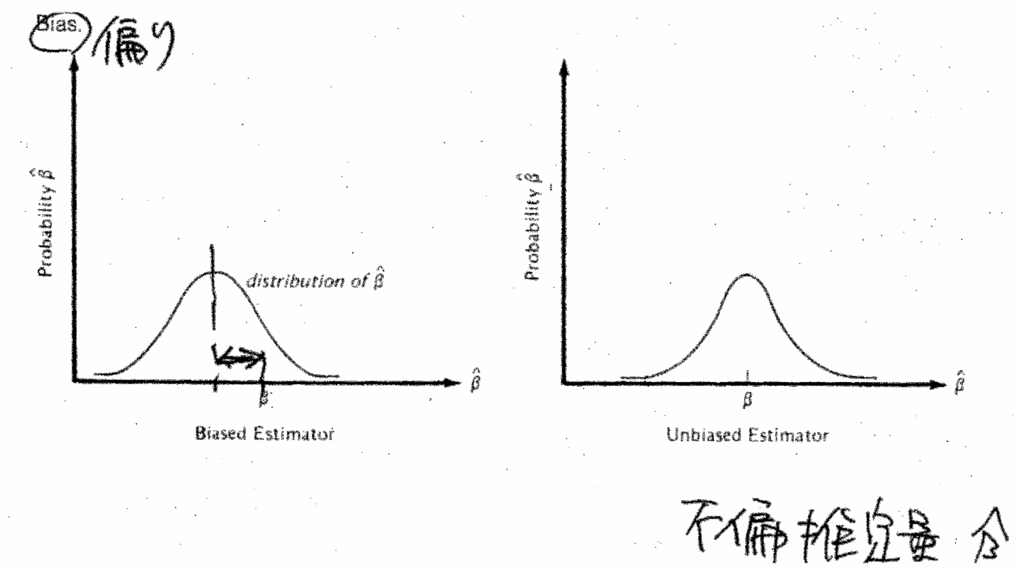
このとき、$E(\hat{\beta})=\beta$ となるような $\hat{\beta}$ を unbiased estimator（不偏推定量）

という。実際には、$\hat{\beta}$ は、β に対し bias（偏り）が存在し、

$\boxed{Bias = E(\hat{\beta}) - \beta}$　で表すことができる。

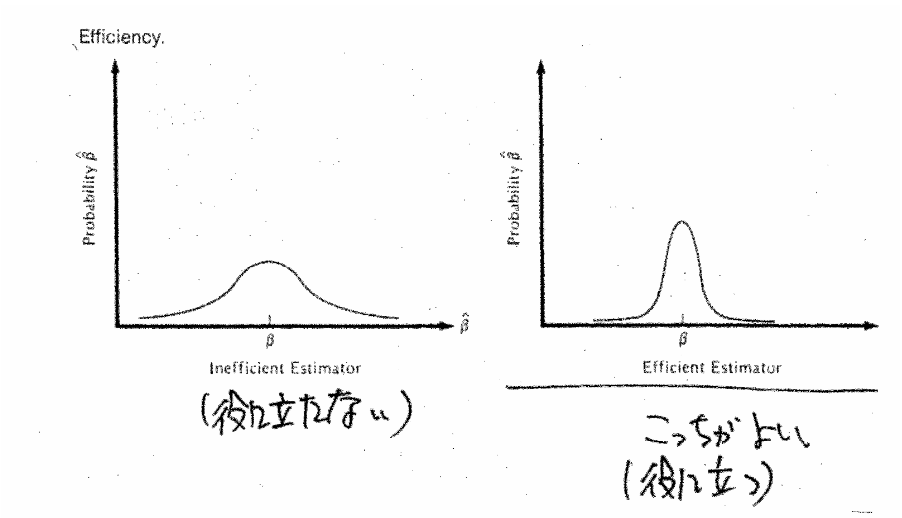
推定において Bias は、当然、存在しない方がよく（*Bias = 0*）、そのとき、

$\hat{\beta}$ は true parameter（実際の値）であるといえる。



(1.2) Efficiency（有効性）

$\hat{\beta}$ の分散ができるだけ小さいほうが、推定量として有効である。



(1.3) Minimum Mean square error（ＭＳＥの最小）

分散と Bias の大きさを示す指標にＭＳＥ（Mean square errore，平均平方誤差）というものがあり、以下の式で定義される。

$$MSE = E(\{\hat{\beta} - \beta\}^2)$$

これを展開すると、

$$MSE = E\left(\left\{\left(\hat{\beta} - \overline{\hat{\beta}}\right) + \left(\overline{\hat{\beta}} - \beta\right)\right\}^2\right) = E\left(\left\{\left(\hat{\beta} - \overline{\hat{\beta}}\right)^2 + \left(\overline{\hat{\beta}} - \beta\right)^2 + 2\left(\hat{\beta} - \overline{\hat{\beta}}\right)\left(\overline{\hat{\beta}} - \beta\right)\right\}\right)$$

$$= E\left(\left\{\hat{\beta} - \overline{\hat{\beta}}\right\}^2\right) + E\left\{\left(\overline{\hat{\beta}} - \beta\right)^2\right\} + E\left(2\left\{\hat{\beta} - \overline{\hat{\beta}}\right\}\left\{\overline{\hat{\beta}} - \beta\right\}\right) \quad \cdots（*）$$

$E\left(2\left\{\hat{\beta} - \overline{\hat{\beta}}\right\}\left\{\overline{\hat{\beta}} - \beta\right\}\right)$だけ取り出して考えると、

$2\left\{\overline{\hat{\beta}} - \beta\right\}$は定数なので、

$$E\left(2\left\{\hat{\beta} - \overline{\hat{\beta}}\right\}\left\{\overline{\hat{\beta}} - \beta\right\}\right) = 2\left(\overline{\hat{\beta}} - \beta\right)E\left(\left\{\hat{\beta} - \overline{\hat{\beta}}\right\}\right) = 2\left(\overline{\hat{\beta}} - \beta\right)\left\{E(\hat{\beta}) - E(\overline{\hat{\beta}})\right\}$$

$$= 2\left(\overline{\hat{\beta}} - \beta\right)\left(\overline{\hat{\beta}} - \overline{\hat{\beta}}\right) = 0$$

$E\left(2\left\{\hat{\beta} - \overline{\hat{\beta}}\right\}\left\{\overline{\hat{\beta}} - \beta\right\}\right) = 0$　これを（*）に代入すると、

$$MSE = E\left(\left\{\hat{\beta} - \overline{\hat{\beta}}\right\}^2\right) + E\left\{\left(\overline{\hat{\beta}} - \beta\right)^2\right\}$$ [1]

(1.2)より、$Bias = E(\hat{\beta}) - \beta = E(\hat{\beta}) - E(\beta) = E(\hat{\beta} - \beta)$　となるので、

上の式は、$\boxed{MSE = Var(\hat{\beta}) + (Bias)^2}$　と結論付けることができる。

もし、$\hat{\beta}$が unbiased estimator（不偏推定量）なら、*Bias = 0* なので、

---

[1] $E\left(\left\{\hat{\beta} - \overline{\hat{\beta}}\right\}^2\right)$は、$\hat{\beta}$と$\hat{\beta}$の平均との差の２乗の平均、つまり$\hat{\beta}$の分散を表す。よって、$E\left(\left\{\hat{\beta} - \overline{\hat{\beta}}\right\}^2\right) = \mathrm{var}(\hat{\beta})$となる。ちなみに、var とは variance の略。

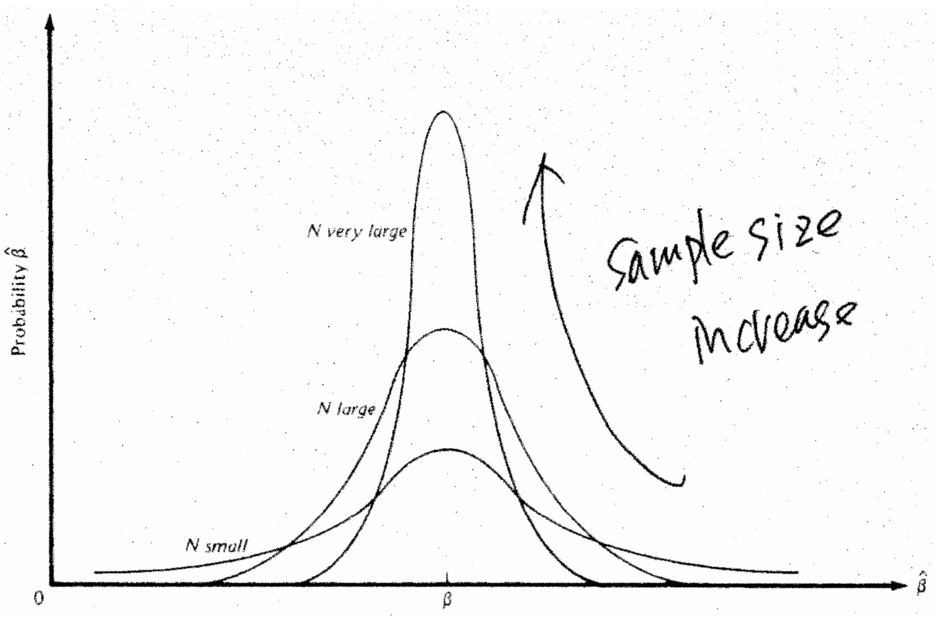
$\underline{MSE = Var(\hat{\beta})}$　となる。

つまり、推定においては、ＭＳＥが分散にできるだけ近いほうが好ましいといえる。

(1.4) Consistency（一致性）

次の場合、$\hat{\beta}$は consistency（一致性）をもち、consistent estimator（一致推定量）であるという。

$\boxed{\lim_{N\to\infty} P(|\beta - \hat{\beta}| < \delta) = 0}$　（Nは標本の大きさ、δは任意の正の実数）[2]



つまり、$\hat{\beta}$が unbiased estimator なら、標本の大きさを大きくしていったとき、$\hat{\beta}$はβに収束するということを示す。統計学的には、Lack of bias（普遍性）よりも、こちらの consistency がより重要視されている。

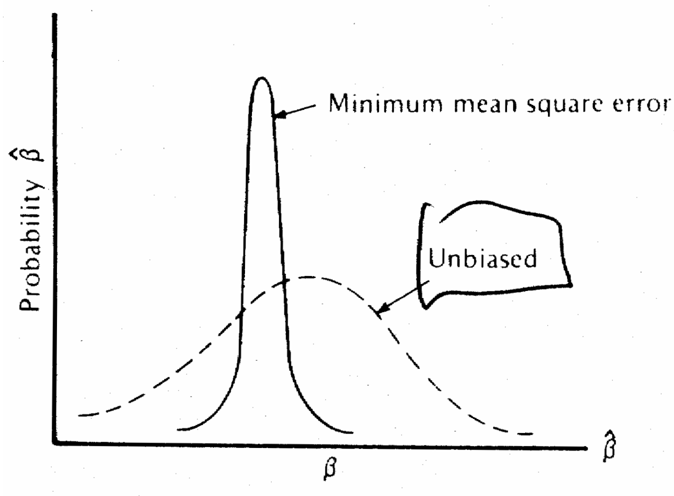
---

[2] $\lim_{N\to\infty} P(|\beta - \hat{\beta}| \geq \delta) = 1$と同じ。また、$\text{plim}\hat{\beta} = \beta$と置き換えることもできる。この法則を大数の法則という。

また、consistency で注意しなければならないのは、unbiased asymptotically である。ＭＳＥが０に近づくということは、Variance と Bias がどちらも０に近づく必要がある。対して、$\hat{\beta}$ が consistent estimator（標本の大きさが無限大のとき $\hat{\beta} = \beta$）だからといって、Variance が低いとは限らない。つまり、

$$MSE \approx 0 \quad \Rightarrow \quad \hat{\beta}\text{が consistent estimator}$$ [3]

と書くことができる。この逆は成立しない。



(2) Probability distribution（確率分布）

本項では以下の４つの確率分布[4]を取り上げる。

1. Normal distribution（正規分布）
2. Chi-square distribution（カイ２乗分布）
3. t distribution（ｔ分布）
4. F distribution（Ｆ分布）

---

[3] $MSE \approx 0$ であることは、$\hat{\beta}$ が consistent estimator であるための十分条件である。$\hat{\beta}$ が consistent estimator であることは、$MSE \approx 0$ であるための必要条件である。

[4] （連続）確率分布とは、変数が一定の範囲の値をとるのは、どれぐらいの確率であるかを示したもの。すなわち、確率分布曲線（確率密度関数のグラフ）と横軸との間の面積は、かならず１になる。

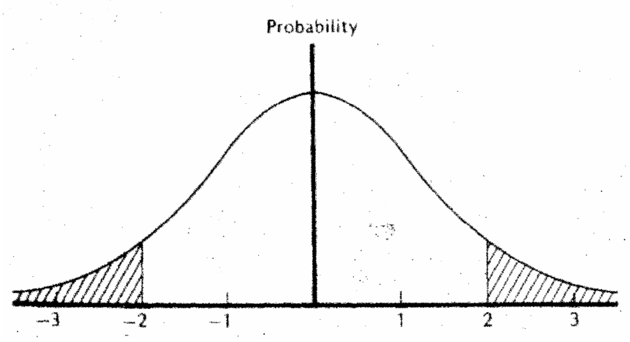
(2.1) Normal distribution（正規分布）

分布とはデータの散らばりのことで、これをグラフに表したものが分布図である。正規分布とは左右対称の釣鐘型で、なだらかな曲線を描く分布であり、全ての統計の基礎となる。

標準分布は以下の式で定義される。

$$\boxed{f(x)=\frac{1}{\sqrt{2\pi}\sigma_x}\exp\left(-\frac{1}{2\sigma_x^2}(x-\mu_x)^2\right)}$$ [5]（σ_x^2 は確率変数 x の分散、μ_x は平均)

これを描くと、次のようなグラフになる。（$\sigma_x^2=1$ ・ $\mu_x=0$）のとき、



確率変数 X が、分散 σ_X^2・平均 μ_X の正規分布に従うとき、$X \sim N(\mu_X,\sigma_X^2)$ と書く。もし、$\mu_X=0,\sigma_X^2=1$ なら、$X \sim N(0,1)$ となり、このような正規分布（上図と同じ）を Standard normal distribution（標準正規分布）という。

正規分布においては以下のことが知られている。

$$P(\mu_X-1.96\sigma_X \le X \le \mu_X+1.96\sigma_X)\approx 0.95$$ [6]

$$P(\mu_X-2.57\sigma_X \le X \le \mu_X+2.57\sigma_X)\approx 0.99$$ [7]

前者を $X \sim N(0,1)$ で考えれば、$P(-1.96 \le X \le 1.96)\approx 0.95$ となるが、

---

5 $\exp(x)\equiv e^x$　ちなみに、$e=\lim_{h\to\infty}\left(1+\frac{1}{h}\right)^h=\lim_{t\to 0}(1+t)^{\frac{1}{t}}$

6 $\int_{\mu_X-1.96\sigma_X}^{\mu_X+1.96\sigma_X} f(x)\,dx=\int_{\mu_X-1.96\sigma_X}^{\mu_X+1.96\sigma_X}\frac{1}{\sqrt{2\pi}\sigma_x}\exp\left(-\frac{1}{2\sigma_x^2}(x-\mu_x)^2\right)dx\approx 0.95$

7 $P(a\le X\le b)=\int_a^b f(x)\,dx$

これは、X が $-1.96 \le X \le 1.96$ の値をとる確率が、約 0.95 であることを示している。

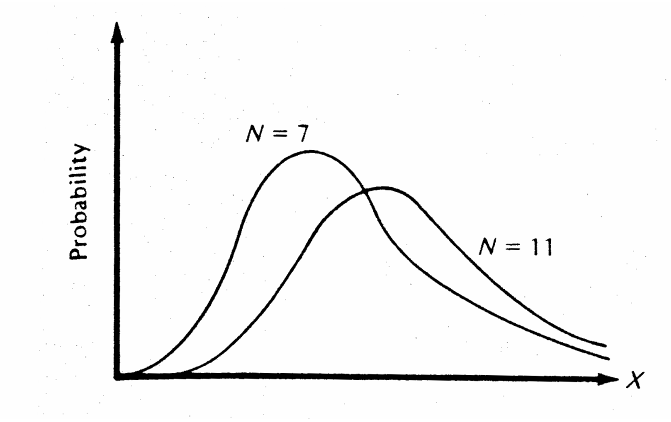
　　ちなみに、$X \sim N(\mu_X, \sigma_X^2)$ において、$\mu_X - n\sigma_X \le X \le \mu_X + n\sigma_X$ となる範囲のことを、n シグマ範囲という。X を標準化[8]した確率変数を Z とすると、$Z \sim N(0,1)$ に対し、n シグマ範囲は $-n \le Z \le n$ となる。つまり、前ページの標準正規分布は２シグマ範囲より外側の領域を示したものとなる。
$P(\mu_X - 1.96\sigma_X \le X \le \mu_X + 1.96\sigma_X) \approx 0.95$ より、２シグマ範囲では確率の９５％が網羅されていることになる。

(2.2) Chi-square distribution（カイ２乗分布，χ^2 分布）

確率変数 X に対して、標本分散 s^2、標本の大きさ N、母分散 σ^2 とすると、χ^2 分布は以下を表す。

$$\boxed{\frac{(N-1)s^2}{\sigma^2}}$$

これを自由度 $N-1$ の χ^2 分布[9]という。χ^2 分布は、標本分散と母分散の比の分布を表している。この分布を図示したものが以下のものである。



自由度（≒標本の大きさ）が大きくなればなるほど、分布が緩やかになっていることが分かる。

[8] 標準化 X を標準化し Z とするとき $Z = \dfrac{X - \mu_X}{\sigma_X}$ が成立する。

[9] 記号で表すと、$\dfrac{(N-1)s^2}{\sigma^2} \sim \chi^2(N-1)$

＜Appendix＞

そもそもχ^2分布は、互いに独立な標準正規確率変数の２乗和の分布と定義される。

Z_1からZ_kを独立に分布する標準正規確率変数だとすると

$$W = Z_1^2 + Z_2^2 + \cdots + Z_k^2$$

が自由度kのχ^2確率変数である。

$\{X_i\}$が平均μ，分散σ^2の独立な正規確率変数であれば

$$W = \frac{1}{\sigma^2}\sum_{i=1}^{k}(X_i - \mu)^2$$

が自由度kのχ^2確率変数になる。

(2.3) t distribution（t分布）

Xを標準正規確率変数、Zを自由度Nのχ^2確率変数とし、さらにXとZが独立に分布していれば、

$$t = \frac{X}{\sqrt{Z/N}}$$

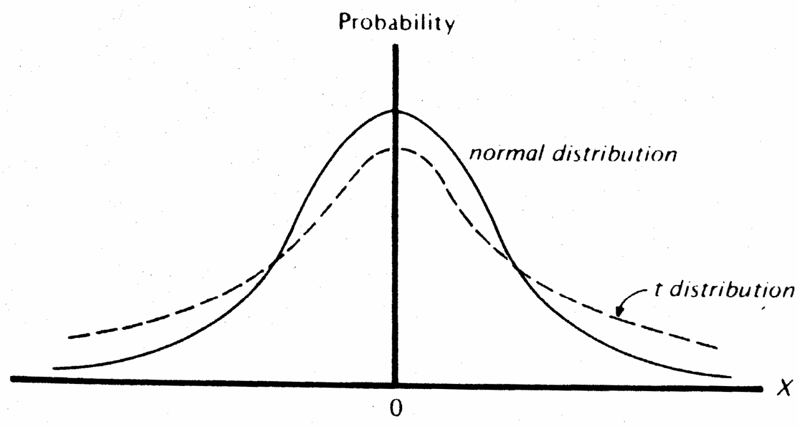
は自由度Nのt確率変数といい、その分布をt分布といえる。

t分布は標準正規分布と同じく縦軸に関して対称な分布である。

また、自由度が１のときは期待値も分散も存在せず、２のときは、期待値は存在するが分散は存在しないので、３以上で初めて両方存在する。

$X = \dfrac{\overline{X} - \mu_X}{\sigma_X/\sqrt{N}}$、$Z = \dfrac{(N-1)s_X^2}{\sigma_X^2}$　をt分布の式$t = \dfrac{X}{\sqrt{Z/N}}$に代入して、

$t = \dfrac{\overline{X} - \mu_X}{s_X/\sqrt{N}}$が導ける。すなわち、$t$分布とは、標本分散と標本平均の比を示している。これを図示したものが以下である。



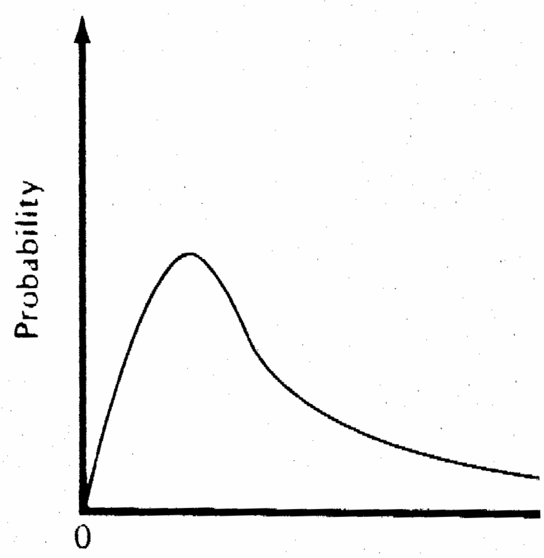
(2.4) F distribution（F分布）

F分布

Vを自由度mのX^2確率変数とし、Wを自由度kのX^2確率変数とする。このときVとWが独立に分布していれば、VとWの比率

$$F=\frac{V/m}{W/k}$$

は自由度がmとkのF分布に従って分布する。
k≧3でなければ期待値は存在しない。

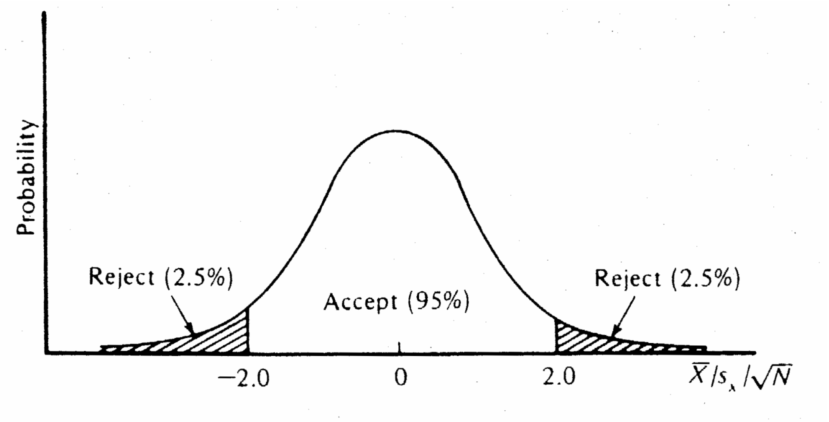


(3) Hypothesis testing and confidence intervals (仮説検定と信頼区間)

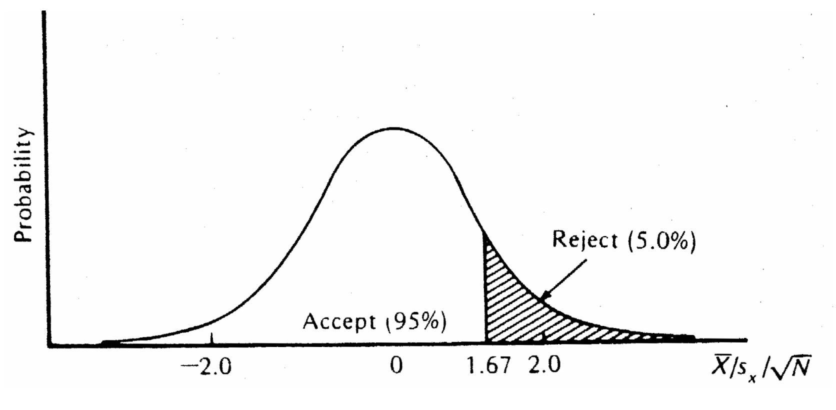
標本平均から両側の“ある範囲”のこと。たとえば、標本分布が正規分布と仮定したとき、測定値から無作為に標本を取り出して平均値を算出したときの標本分布は正規分布となる。このとき、母集団から標本を取り出して求めた平均値が、信頼区間に入らない確率が一定値(例えば 5%)であるような可能性をのこしている。

ただし、信頼区間の上限から下限の間に、母数としての平均値がある確率が 1-0.05 という理解は正確ではない。このような区間推定をして標本を抽出することで、例えばその 95%は正しい推定であろう、と期待することができるということになる。

＜両側検定の様子＞



＜片側検定の様子＞



Ｐ値
検定結果を示すとき，棄権率(有意水準)をどんな値にすれば，検定結果が有意になるかという値を示すことがある。これをＰ値(P-value)という。

(4) Glossary

Property	属性
Regression	回帰分析
Estimator	推定量
True ～	真の（実際の）
Distribution	分布
Bias	偏り
Unbiased estimator	不偏推定量
Lack of bias	不偏性
Dispersion	ばらつき・ちらばり
Efficiency	有効性
Sample size	標本の大きさ
Variance	分散
Numerical value	絶対値
MSE (Mean square error)	平均平方誤差
Precision	正確さ
Consistency	一致性
Asymptotic	非対称
Consistent estimator	一致推定量
Probability distribution	確率分布
Normal distribution	正規分布
Standard normal distribution	標準正規分布
Chi-square distribution	カイ２乗分布
Hypothesis testing	仮説検定
Confidence intervals	信頼区間

(5) References

Robert S. Pindyck and Daniel L. Rubinfeld (1991): Econometric Models and economic forecasts 3rd edition, McGraw-Hill

Mendenhall, Reinmuth and Beaver (1989): Statistics for management and economics 6th edition, Pws-kent publishing company

森棟公夫 (2000): 統計学入門　第２版, 新世社

藤本利躬・西田小百合 (1999): 初級コース　計量経済学, 中央経済社