応用システム工学
第六回　回帰分析

平成23年05月27日

X^2分布

- 各々標準正規分布N(0,1)の独立な確率変数 $Z_1, Z_2, \cdots, Z_n$ に対する次式で定義される確率変数X^2

$$\chi^2 = Z_1{}^2 + Z_2{}^2 + \cdots + Z_n{}^2 = \sum_{i=1}^{n} Z_i{}^2$$

の確率分布→自由度nのX^2分布　$\chi^2(n)$

$$\sum_{i=1}^{n}\left(\frac{Z_i-\mu}{\sigma}\right)^2 = \frac{(n-1)s^2}{\sigma^2} + \left(\frac{\bar{Z}-\mu}{\sqrt{\sigma^2/n}}\right)^2$$

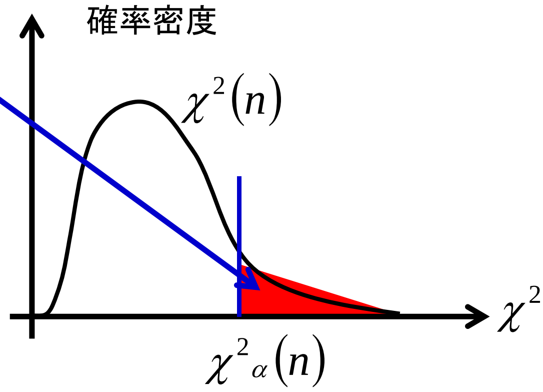
自由度nのx^2分布　自由度n-1のx^2分布　自由度1のx^2分布

母分散の区間推定

- x^2分布する確率変数x^2がある範囲をとる確率
  - x^2分布表を用いる
  - 自由度nのx^2分布の上側確率

$$P\left(\chi^2 > \chi^2{}_{\alpha}(n)\right) = \alpha$$

における100α%点$x^2{}_{\alpha}(n)$



X^2分布

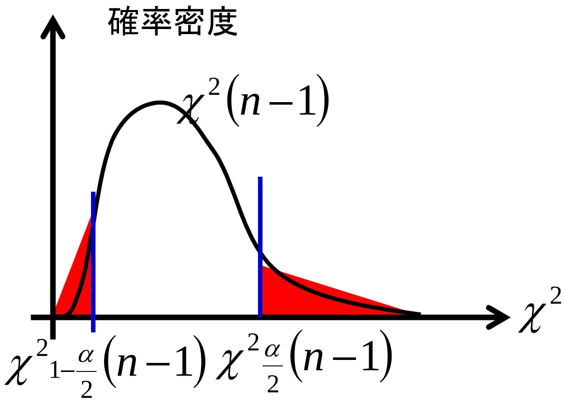
母分散の区間推定

- 正規母集団（母分散$σ^2$）
  - 標本サイズn，不偏分散s^2
  - $\dfrac{(n-1)s^2}{\sigma^2}$ は自由度n-1のx^2分布となる
    - 上側確率

$$P\left(\chi^2 > \chi^2{}_{\frac{\alpha}{2}}(n-1)\right) = \frac{\alpha}{2}$$

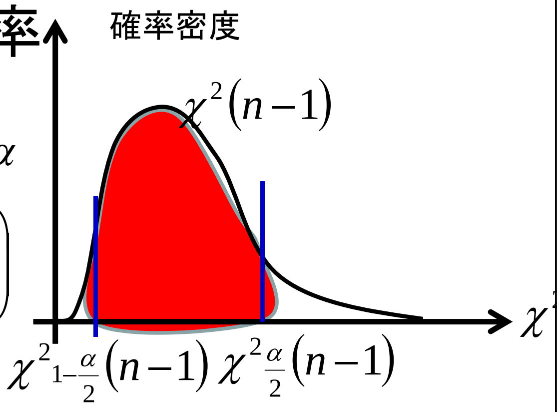
    - 下側確率

$$P\left(\chi^2 > \chi^2{}_{1-\frac{\alpha}{2}}(n-1)\right) = \frac{\alpha}{2}$$



母分散の区間推定

- 上側と下側の間に入る確率



$$P\left(\chi^2{}_{1-\frac{\alpha}{2}}(n-1) \le \chi^2 \le \chi^2{}_{\frac{\alpha}{2}}(n-1)\right) = 1-\alpha$$

$$P\left(\chi^2{}_{1-\frac{\alpha}{2}}(n-1) \le (n-1)\frac{s^2}{\sigma^2} \le \chi^2{}_{\frac{\alpha}{2}}(n-1)\right)$$
$$= 1-\alpha$$

$$P\left(\frac{(n-1)s^2}{\chi^2{}_{1-\frac{\alpha}{2}}(n-1)} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2{}_{\frac{\alpha}{2}}(n-1)}\right) = 1-\alpha$$

信頼係数1-αの信頼区間 $\left[\frac{(n-1)s^2}{\chi^2{}_{1-\frac{\alpha}{2}}(n-1)}, \frac{(n-1)s^2}{\chi^2{}_{\frac{\alpha}{2}}(n-1)}\right]$

t分布

- 母集団 → 母平均μ, 母分散σ^{2}の正規分布N(μ, σ^2)
  - 母平均μを標本サイズn個の標本$X_1, X_2, \cdots, X_n$で推定・検定する
    - 標本平均 $\bar{X}$ は正規分布 $N\left(\mu, \frac{\sigma^2}{n}\right)$ となる
      - 変形して $X' = \frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}}$ は標準正規分布となる
      - 母分散σ^{2}は不明 → 不変分散s^{2}で代用 t統計量 $t = \frac{\bar{X}-\mu}{\sqrt{\frac{s^2}{n}}}$

t分布

- t統計量の分布

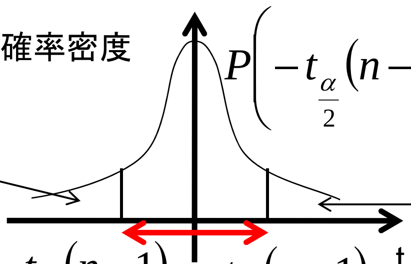
$$t=\frac{\overline{X}-\mu}{\sqrt{\frac{s^2}{n}}}=\frac{\frac{\overline{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}}}{\frac{\sqrt{\frac{s^2}{n}}}{\sqrt{\frac{\sigma^2}{n}}}}=\frac{X'}{\sqrt{\frac{s^2}{\sigma^2}}}=\frac{X'}{\sqrt{\frac{(n-1)\frac{s^2}{\sigma^2}}{n-1}}}$$

標本平均 $\overline{X}$ と不偏分散s^2は独立 → X'と $(n-1)\frac{s^2}{\sigma^2}$ も独立

$(n-1)\frac{s^2}{\sigma^2}$ は自由度n-1のX^2分布X^2(n-1)となる。
→ 統計量tを自由度n-1のt分布と呼ぶ

t分布を用いた母平均の区間推定

- 正規分布する母集団に対し，標本サイズnの標本平均 $\overline{X}$ ，不偏分散s^2から，母平均μの100(1-α)%信頼区間を求める
  - 自由度n-1のt分布t(n-1)
    - T統計量がその値以上の確率が100α/2%の値 $t_{\frac{\alpha}{2}}(n-1)$
    - T統計量がその値以下の確率が100α/2%の値 $-t_{\frac{\alpha}{2}}(n-1)$



$$P\left(-t_{\frac{\alpha}{2}}(n-1)\le t\le t_{\frac{\alpha}{2}}(n-1)\right)=1-\alpha$$

$$P\left(t<-t_{\frac{\alpha}{2}}(n-1)\right)=\frac{\alpha}{2}$$

$$P\left(t>t_{\frac{\alpha}{2}}(n-1)\right)=\frac{\alpha}{2}$$

$-t_{\frac{\alpha}{2}}(n-1)$ $t_{\frac{\alpha}{2}}(n-1)$

100α%信頼区間

t分布を用いた母平均の区間推定

$$P\left(-t_{\frac{\alpha}{2}}(n-1) \le \frac{\overline{X}-\mu}{\sqrt{\frac{s^2}{n}}} \le t_{\frac{\alpha}{2}}(n-1) \right) = 1-\alpha$$

$$P\left(\overline{X} - t_{\frac{\alpha}{2}}(n-1)\sqrt{\frac{s^2}{n}} \le \mu \le \overline{X} + t_{\frac{\alpha}{2}}(n-1)\sqrt{\frac{s^2}{n}} \right) = 1-\alpha$$

100α%信頼区間 $\left[\overline{X} - t_{\frac{\alpha}{2}}(n-1)\sqrt{\frac{s^2}{n}}, \overline{X} + t_{\frac{\alpha}{2}}(n-1)\sqrt{\frac{s^2}{n}} \right]$

単回帰モデルにおける回帰係数の推定と検定

- 母集団の線形回帰モデル

$$y_i = a_0 + a_1 x_i + e_i, i = 1,2\cdots n$$

  - 誤差e_iは期待値μ=0，分散σ^2の正規分布，相互に無相関とする
  - 母集団からの標本(x_i,y_i),i=1,2，・・・,nを考える
    - 標本に対する回帰係数，定数項と観測値の標本との関係を考える

単回帰モデルにおける回帰係数の推定と検定

- 標本に対する回帰係数

$$\hat{a}_1 = \frac{s_{xy}}{s_{xx}} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{s_{xx}} = \frac{\frac{1}{n}\left\{\sum_{i=1}^{n}(x_i - \bar{x})y_i - \bar{y}\sum_{i=1}^{n}(x_i - \bar{x})\right\}}{s_{xx}}$$

$$= \sum_{i=1}^{n}\frac{(x_i - \bar{x})y_i}{ns_{xx}}$$

- 標本に対する定数項

$$\hat{a}_0 = \bar{y} - \hat{a}_1\bar{x} = \frac{1}{n}\sum_{i=1}^{n} y_i - \frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{ns_{xx}}\bar{x} = \sum_{i=1}^{n}\left\{\frac{1}{n} - \frac{x_i - \bar{x}}{ns_{xx}}\bar{x}\right\}y_i$$

単回帰モデルにおける回帰係数の推定と検定

- 母集団のa_1と標本に対する回帰係数 $\hat{a}_1$ の関係

$$\hat{a}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{ns_{xx}} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(a_0 + a_1 x_i + e_i)}{ns_{xx}}$$

$$\sum_{i=1}^{n}(x_i - \bar{x}) = 0$$

より

$$\sum_{i=1}^{n}(x_i - \bar{x})\bar{x} = 0$$

$$\sum_{i=1}^{n} e_i = 0$$

$$\sum_{i=1}^{n}(x_i - \bar{x})x_i = \sum_{i=1}^{n}(x_i - \bar{x})^2 = ns_{xx}$$

$$\sum_{i=1}^{n}(x_i - \bar{x})e_i = \sum_{i=1}^{n} x_i e_i - \bar{x}\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} x_i e_i$$

$$\hat{a}_1 = \frac{a_1 ns_{xx} + \sum_{i=1}^{n} x_i e_i}{ns_{xx}} = a_1 + \frac{\sum_{i=1}^{n} x_i e_i}{ns_{xx}}$$

単回帰モデルにおける回帰係数の推定と検定

- 母集団のa_0と標本に対する定数項の関係

$$\hat{a}_0 = \sum_{i=1}^{n}\left\{\frac{1}{n} - \frac{x_i - \bar{x}}{ns_{xx}}\bar{x}\right\}y_i = \sum_{i=1}^{n}\left\{\frac{1}{n} - \frac{x_i - \bar{x}}{ns_{xx}}\bar{x}\right\}(a_0 + a_1 x_i + e_i)$$

$$= \frac{1}{n}\sum_{i=1}^{n}(a_0 + a_1 x_i + e_i) - \frac{\bar{x}}{ns_{xx}}\sum_{i=1}^{n}(x_i - \bar{x})(a_0 + a_1 x_i + e_i)$$

$$= a_0 + a_1\bar{x} - \frac{\bar{x}}{ns_{xx}}\left\{a_1\sum_{i=1}^{n}(x_i - \bar{x})^2 + \sum_{i=1}^{n}x_i e_i\right\}$$

$$= a_0 + a_1\bar{x} - \frac{\bar{x}}{ns_{xx}}\left\{a_1 ns_{xx} + \sum_{i=1}^{n}x_i e_i\right\} = a_0 - \frac{\bar{x}}{ns_{xx}}\sum_{i=1}^{n}x_i e_i$$

単回帰モデルにおける回帰係数の推定と検定

- 標本に対する回帰係数の期待値

$$E[\hat{a}_1] = E\left[\sum_{i=1}^{n}\left\{\frac{1}{n} - \frac{x_i - \bar{x}}{ns_{xx}}\bar{x}\right\}y_i\right] = E\left[\sum_{i=1}^{n}\frac{(x_i - \bar{x})(a_0 + a_1 x_i + e_i)}{ns_{xx}}\right]$$

$$= \sum_{i=1}^{n}\frac{(x_i - \bar{x})a_1 x_i}{ns_{xx}} = \sum_{i=1}^{n}\frac{(x_i - \bar{x})a_1 x_i}{ns_{xx}} - \sum_{i=1}^{n}\frac{(x_i - \bar{x})a_1\bar{x}}{ns_{xx}}$$

$$= \sum_{i=1}^{n}\frac{(x_i - \bar{x})^2 a_1}{ns_{xx}} = a_1$$

$$E[\hat{a}_0] = E[\bar{y} - \hat{a}_1\bar{x}] = E[\bar{y}] - E[\hat{a}_1]E[\bar{x}]$$

$$= a_0 + a_1\bar{x} - a_1\bar{x} = a_0$$

標本に基づく回帰係数・定数項　　　は，母集団のa_0,a_1に対して偏りのない推定値となっている→不偏推定値

単回帰モデルにおける回帰係数の推定と検定

- 標本に対する回帰係数の期待値（別解）

$$E[\hat{a}_1] = E\left[a_1 + \frac{\sum_{i=1}^{n} x_i e_i}{ns_{xx}}\right] = E[a_1] + E\left[\frac{\sum_{i=1}^{n} x_i e_i}{ns_{xx}}\right]$$

$$= a_1 + \frac{1}{ns_{xx}} E\left[\sum_{i=1}^{n} x_i e_i\right]$$

$$= a_1 + \frac{1}{ns_{xx}} \sum_{i=1}^{n} x_i E[e_i]$$

$$= a_1$$

単回帰モデルにおける回帰係数の推定と検定

- 標本に対する定数項の期待値(別解)

$$E[\hat{a}_0] = E\left[a_0 - \frac{\bar{x}}{ns_{xx}} \sum_{i=1}^{n} x_i e_i\right] = E[a_0] - E\left[\frac{\bar{x}}{ns_{xx}} \sum_{i=1}^{n} x_i e_i\right]$$

$$= a_0 - \frac{\bar{x}}{ns_{xx}} E\left[\sum_{i=1}^{n} x_i e_i\right]$$

$$= a_0 - \frac{\bar{x}}{ns_{xx}} \sum_{i=1}^{n} x_i E[e_i]$$

$$= a_0$$

標本に基づく回帰係数・定数項 $\hat{a}_1, \hat{a}_0$ は，母集団のa_o,a_1に対して偏りのない推定値となっている→不偏推定値

単回帰モデルにおける回帰係数の推定と検定

- 標本に対する回帰係数の分散

$$V[\hat{a}_1]=V\left[\sum_{i=1}^{n}\frac{(x_i-\overline{x})y_i}{ns_{xx}}\right]=\sum_{i=1}^{n}\left(\frac{x_i-\overline{x}}{ns_{xx}}\right)^2 V[y_i]$$

$$=\left(\frac{1}{ns_{xx}}\right)^2\sum_{i=1}^{n}(x_i-\overline{x})^2\sigma^2=\frac{\sigma^2}{ns_{xx}}$$

$$V[\hat{a}_0]=V\left[\sum_{i=1}^{n}\left\{\frac{1}{n}-\frac{x_i-\overline{x}}{ns_{xx}}\overline{x}\right\}y_i\right]=\sum_{i=1}^{n}\left(\frac{1}{n}-\frac{x_i-\overline{x}}{ns_{xx}}\overline{x}\right)^2 V[y_i]$$

$$=\left[\sum_{i=1}^{n}\left(\frac{1}{n}\right)^2-2\sum_{i=1}^{n}\left(\frac{1}{n}\frac{x_i-\overline{x}}{ns_{xx}}\overline{x}\right)+\sum_{i=1}^{n}\left(\frac{x_i-\overline{x}}{ns_{xx}}\overline{x}\right)^2\right]V[y_i]$$

単回帰モデルにおける回帰係数の推定と検定

- 標本に対する回帰係数の分散つづき

$$V[\hat{a}_0]=\left[\frac{1}{n}-\frac{2\overline{x}}{n^2s_{xx}}\sum_{i=1}^{n}(x_i-\overline{x})+\left(\frac{\overline{x}}{ns_{xx}}\right)^2\sum_{i=1}^{n}(x_i-\overline{x})^2\right]V[y_i]$$

$$=\left[\frac{1}{n}-\frac{2\overline{x}}{n^2s_{xx}}0+\left(\frac{\overline{x}}{ns_{xx}}\right)^2 ns_{xx}\right]\sigma^2$$

$$=\left[\frac{1}{n}+\frac{\overline{x}^2}{ns_{xx}}\right]\sigma^2$$

単回帰モデルにおける回帰係数の推定と検定

- 標本に対する回帰係数の共分散

$$Cov[\hat{a}_1, \hat{a}_0] = Cov\left[\sum_{i=1}^{n} \frac{(x_i - \bar{x})y_i}{ns_{xx}}, \sum_{i=1}^{n} \left\{\frac{1}{n} - \frac{x_i - \bar{x}}{ns_{xx}}\bar{x}\right\} y_i\right]$$

$$= \sum_{i=1}^{n} \frac{(x_i - \bar{x})}{ns_{xx}} \left\{\frac{1}{n} - \frac{x_i - \bar{x}}{ns_{xx}}\bar{x}\right\} V[y_i]$$

$$= \left\{\frac{1}{n^2 s_{xx}} \sum_{i=1}^{n} (x_i - \bar{x}) - \frac{\bar{x}}{(ns_{xx})^2} \sum_{i=1}^{n} (x_i - \bar{x})^2\right\} V[y_i]$$

$$= \left\{\frac{1}{n^2 s_{xx}} 0 - \frac{\bar{x}}{(ns_{xx})^2} ns_{xx}\right\} \sigma^2 = -\frac{\bar{x}}{ns_{xx}} \sigma^2$$

単回帰モデルにおける回帰係数の推定と検定

- 標本に対する回帰係数の分散(別解)

$$V[\hat{a}_1] = V\left[a_1 + \frac{\sum_{i=1}^{n} x_i e_i}{ns_{xx}}\right] = V\left[\frac{\sum_{i=1}^{n} x_i e_i}{ns_{xx}}\right] = \left(\frac{1}{ns_{xx}}\right)^2 V\left[\sum_{i=1}^{n} x_i e_i\right]$$

$$= \left(\frac{1}{ns_{xx}}\right)^2 \sum_{i=1}^{n} x_i^2 V[e_i] = \left(\frac{1}{ns_{xx}}\right)^2 \sigma^2 \sum_{i=1}^{n} x_i^2$$

$$= \left(\frac{1}{ns_{xx}}\right)^2 \sigma^2 \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{\sigma^2}{ns_{xx}}$$

単回帰モデルにおける回帰係数の推定と検定

- 標本に対する定数項の分散(別解)

$$V[\hat{a}_0] = V\left[a_0 - \frac{\bar{x}}{ns_{xx}}\sum_{i=1}^{n} x_i e_i\right] = V\left[\frac{\bar{x}}{ns_{xx}}\sum_{i=1}^{n} x_i e_i\right] = \left(\frac{\bar{x}}{ns_{xx}}\right)^2 V\left[\sum_{i=1}^{n} x_i e_i\right]$$

$$= \left(\frac{\bar{x}}{ns_{xx}}\right)^2 \sum_{i=1}^{n} x_i{}^2 V[e_i] = \left(\frac{\bar{x}}{ns_{xx}}\right)^2 \sigma^2 \sum_{i=1}^{n} x_i{}^2$$

$$= \left(\frac{\bar{x}}{ns_{xx}}\right)^2 \sigma^2 \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{\bar{x}\sigma^2}{ns_{xx}}$$

単回帰モデルにおける回帰係数の推定と検定

- 標本分散

$$s^2 = \frac{\sum_{i=1}^{n}\{y_i - (\hat{a}_0 + \hat{a}_1 x_i)\}^2}{n} = \frac{\sum_{i=1}^{n}\{y_i - (\bar{y} - \hat{a}_1\bar{x} + \hat{a}_1 x_i)\}^2}{n}$$

$$= \frac{1}{n}\left\{\sum_{i=1}^{n}(y_i - \bar{y})^2 + \hat{a}_1{}^2\sum_{i=1}^{n}(x_i - \bar{x})^2 - 2\hat{a}_1\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})\right\}$$

$$= \frac{1}{n}\left\{ns_{yy} + \hat{a}_1{}^2 ns_{xx} - 2\hat{a}_1 ns_{xy}\right\} = s_{yy} + \hat{a}_1{}^2 s_{xx} - 2\hat{a}_1 s_{xy}$$

$$= s_{yy} + \left(\frac{s_{xy}}{s_{xx}}\right)^2 s_{xx} - 2\frac{s_{xy}}{s_{xx}} s_{xy} = s_{yy} - \frac{s_{xy}{}^2}{s_{xx}}$$

単回帰モデルにおける回帰係数の推定と検定

$$Cov[\hat{a}_1, e_i] = Cov\left[\sum_{i=1}^{n}\frac{(x_i-\overline{x})y_i}{ns_{xx}}, e_i\right]$$

$$= Cov\left[\sum_{i=1}^{n}\frac{(x_i-\overline{x})e_i}{ns_{xx}}, e_i\right] \qquad \sum_{i=1}^{n} Cov[\hat{a}_1, e_i] = 0$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{(x_i-\overline{x})e_i e_j}{ns_{xx}}$$

$$= \sum_{i=1}^{n}\frac{(x_i-\overline{x})e_i^{\,2}}{ns_{xx}} = \frac{\sigma^2}{ns_{xx}}\sum_{i=1}^{n}(x_i-\overline{x}) = 0$$

単回帰モデルにおける回帰係数の推定と検定

$$Cov[\hat{a}_0, e_i] = Cov\left[\sum_{i=1}^{n}\left\{\frac{1}{n}-\frac{x_i-\overline{x}}{ns_{xx}}\overline{x}\right\}y_i, e_i\right] = Cov\left[\sum_{i=1}^{n}\left\{\frac{1}{n}-\frac{x_i-\overline{x}}{ns_{xx}}\overline{x}\right\}e_i, e_i\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\{\frac{1}{n}-\frac{x_i-\overline{x}}{ns_{xx}}\overline{x}\right\}e_i e_j = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{1}{n}-\frac{x_i-\overline{x}}{ns_{xx}}\overline{x}\right\}e_i^{\,2}$$

$$= \frac{\sigma^2}{n}\sum_{i=1}^{n}\left\{\frac{1}{n}-\frac{x_i-\overline{x}}{ns_{xx}}\overline{x}\right\} = \frac{\sigma^2}{n}\left\{\sum_{i=1}^{n}\frac{1}{n}-\frac{\overline{x}}{ns_{xx}}\sum_{i=1}^{n}(x_i-\overline{x})\right\}$$

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

$$\sum_{i=1}^{n} Cov[\hat{a}_0, e_i] = \sum_{i=1}^{n}\frac{\sigma^2}{n} = \sigma^2$$

単回帰モデルにおける回帰係数の推定と検定

- 不偏推定量 (不偏分散)

$$V_e = \frac{n}{n-2}s^2 = \frac{\sum_{i=1}^{n}\{y_i - (\hat{a}_0 + \hat{a}_1 x_i)\}^2}{n-2}$$

$$E[V_e] = \sigma^2$$

  – 自由度n-2のt分布の統計量

$$t = \frac{\hat{a}_1 - a_1}{\sqrt{\frac{V_e}{ns_{xx}}}} \qquad t = \frac{\hat{a}_0 - a_0}{\sqrt{V_e\left[\frac{1}{n} + \frac{\bar{x}^2}{ns_{xx}}\right]}}$$

単回帰モデルにおける回帰係数の推定と検定

- 100(1-α)%信頼区間
  – 絶対値で表す

$$P\left(\frac{|\hat{a}_1 - a_1|}{\sqrt{\frac{V_e}{ns_{xx}}}} \geq t_\alpha(n-2)\right) = \alpha$$

$$P\left(\frac{|\hat{a}_0 - a_0|}{\sqrt{V_e\left[\frac{1}{n} + \frac{\bar{x}^2}{ns_{xx}}\right]}} \geq t_\alpha(n-2)\right) = \alpha$$

単回帰モデルにおける回帰係数の推定と検定

- 信頼率1-αの信頼区間

$$\hat{a}_1 - t_\alpha(n-2)\sqrt{\frac{V_e}{ns_{xx}}} \le a_1 \le \hat{a}_1 + t_\alpha(n-2)\sqrt{\frac{V_e}{ns_{xx}}}$$

$$\hat{a}_0 - t_\alpha(n-2)\sqrt{V_e\left[\frac{1}{n} + \frac{\bar{x}^2}{ns_{xx}}\right]} \le a_0 \le \hat{a}_0 + t_\alpha(n-2)\sqrt{V_e\left[\frac{1}{n} + \frac{\bar{x}^2}{ns_{xx}}\right]}$$

- レポート締切　7/29
- 提出先　E2-111