統計解析論特殊講義　模擬試験（１）のヒント

１．$n=10,\quad \sum_{i=1}^{n} x_i^2 = 10,\quad \sum_{i=1}^{n} x_i y_i = 10,\quad \sum_{i=1}^{n} y_i^2 = 40,\quad \bar{x}=0,\quad \bar{y}=1$であった．

（ア）回帰式$y_i = \alpha + \beta x_i + \varepsilon_i$の最小二乗推定値$\hat{\alpha}, \hat{\beta}$，残差二乗和$\sum_{i=1}^{n} \hat{u}_i^2$，ＴＳＳ，ＥＳＳ，$R^2$，$\bar{R}^2$を求めよ．

＜ヒ＞

$\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}$，$\sum_{i=1}^{n}(x_i-\bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$，$\sum_{i=1}^{n}(y_i-\bar{y})^2 = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2$を用いる．$ESS = \hat{\beta}^2 \sum_{i=1}^{n}(x_i-\bar{x})^2$を使って，ＲＳＳを$RSS = TSS - ESS$から求める．

（イ）$H_0:\beta=0,\quad H_1:\beta\neq 0$に対するｔ統計量値を求め有意水準５％で検定せよ．

＜ヒ＞

ｔ統計量値の計算は，$\hat{\beta} \Big/ \left(S \Big/ \sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2} \right)$．両側検定になる．境界値は自由度８から２．３１．これは，両側に棄却域が取られ，その両側の棄却域の合計確率が５％であるところから求まる．したがって，絶対値がこれより大きい場合棄却．

（ウ）$H_0:\beta=1,\quad H_1:\beta<1$に対するｔ統計量値を求め有意水準５％で検定せよ．

＜ヒ＞

ｔ統計量値の計算は，帰無仮説が$\beta=1$なので，$(\hat{\beta}-1) \Big/ \left(S \Big/ \sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2} \right)$．片側検定．対立仮説が正しい場合，ｔ統計量値は負になる．従って，対立仮説を採択する＝帰無仮説を棄却する範囲，つまり，棄却域はｔ統計量値が負の側に取られる．境界値は自由度８から－１．８６．

（エ）$H_0:\alpha=\beta=0$に対するＦ統計量値を求め有意水準５％で検定せよ．

＜ヒ＞

帰無仮説での回帰モデルは$y_i=\varepsilon_i$であるので，この回帰モデルのRSS_0は$\sum_{i=1}^{n} y_i^2$で

ある．したがって，$F=\dfrac{(RSS_0-RSS_a)/2}{RSS_a/(n-2)}$となる．分子の自由度が２なのは，帰無仮説の式が２つで構成されているからである．RSS_aはすでに計算済みなのでそれを利用する．境界値は４．４６．これより大きければ棄却.

（オ）βの９５%信頼区間を求めよ.

＜ヒ＞

ｔ統計量の自由度が８なので，その有意水準５%の境界値を求めておく．これを，Aとおくと，$\hat{\beta}-A\dfrac{S}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}},\hat{\beta}+A\dfrac{S}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}}$

（カ）$x=2$の場合，ｙの予測値はいくらか．また，予測値の９５%信頼区間を求めよ.（予測誤差の分散を使い，後はパラメータ値の区間推定と同様に行う.）

＜ヒ＞

スライド４の４０枚目から，$V[y-\hat{y}]=\sigma^2\left(1/n+(x-\bar{x})^2/\sum_{i=1}^{n}(x_i-\bar{x})^2+1\right)$であることがわかる．これを利用する．前のヒントでは$\dfrac{S}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}}$が$\sqrt{V(\hat{\beta})}$の推定値であった． 今度のは予測誤差の標準偏差の推定値として$S\left(1/n+(x-\bar{x})^2/\sum_{i=1}^{n}(x_i-\bar{x})^2+1\right)$を用いると考えればよい．従って，前のヒントの$\dfrac{S}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}}$が$S\left(1/n+(x-\bar{x})^2/\sum_{i=1}^{n}(x_i-\bar{x})^2+1\right)$に変わると考えればよい.

２． $n=7,\ \sum_{i=1}^{n}x_i^2=4,\quad \sum_{i=1}^{n}x_iy_i=8,\quad \sum_{i=1}^{n}y_i^2=43,\quad \bar{x}=0$， $\bar{y}=1$とする．回帰式$y_i=\alpha+\beta x_i+\varepsilon_i$において$H_0:\alpha=\beta,\quad H_1:\alpha<\beta$を有意水準５%で検定せよ.

＜ヒ＞

$y_i=\alpha+\beta x_i+\varepsilon_i$を変形する． $y_i=\alpha-\beta+\beta+\beta x_i+\varepsilon_i$となるので，$y_i=(\alpha-\beta)+\beta(x_i+1)+\varepsilon_i$という回帰式で考える．この１番目の係数に関するｔ検定を利用する．$\hat{\alpha}$と$\hat{\beta}$を$y_i=\alpha+\beta x_i+\varepsilon_i$の推定量値とすると，係数推定値は

$\hat{\alpha}-\hat{\beta}$ である．この分散は $\dfrac{\sigma^2}{\sum_{i=1}^{n}(1を x_i+1 に回帰した残差)^2}$ である．これは，多変数回帰の $V\left(\hat{\beta}_j\right)=\sigma^2 \Big/ \sum_{i=1}^{n}\left(x_{j,i}^{*}\right)^2$ に対応している．

$\sum_{i=1}^{n}(1を x_i+1 に回帰した残差)^2=\sum_{i=1}^{n}1^2-(回帰係数推定値)^2\sum_{i=1}^{n}(x_i+1)^2$ となる．

$1を x_i+1 に回帰したときの回帰係数推定値=\dfrac{\sum_{i=1}^{n}(x_i+1)}{\sum_{i=1}^{n}(x_i+1)^2}$．

$\sum_{i=1}^{n}(x_i+1)=\sum_{i=1}^{n}x_i+n=n(\bar{x}+1)$，

$\sum_{i=1}^{n}(x_i+1)^2=\sum_{i-1}^{n}x_i^2+2\sum_{i=1}^{n}x_i+n=\sum_{i-1}^{n}x_i^2+2n\bar{x}+n$ を利用して，

$\sum_{i=1}^{n}(1を x_i+1 に回帰した残差)^2=n-\dfrac{n^2(\bar{x}+1)^2}{\sum_{i=1}^{n}x_i^2+2n\bar{x}+n}=\dfrac{n\sum_{i=1}^{n}x_i^2-n^2\bar{x}^2}{\sum_{i=1}^{n}x_i^2+2n\bar{x}+n}$．先ほどの回帰式の変形は誤差項を変化させないから，$y_i=\alpha+\beta x_i+\varepsilon_i$ の残差も，$y_i=(\alpha-\beta)+\beta(x_i+1)+\varepsilon_i$ の残差も同じ．ＲＳＳは，$y_i=\alpha+\beta x_i+\varepsilon_i$ を用いて，

$\sum_{i=1}^{n}(y_i-\bar{y})^2-\dfrac{\left\{\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x})\right\}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$ で計算できる．ここから，S^2 も解るので，

$V\left(\hat{\alpha}-\hat{\beta}\right)$ も推定できる．従って，$y_i=(\alpha-\beta)+\beta(x_i+1)+\varepsilon_i$ の $\alpha-\beta$ に関するｔ統計量値も計算できる．対立仮説が，$\alpha-\beta<0$ なので片側検定で，棄却域は負の範囲に取る．自由度５のｔ分布なので，－２．０２が検定の境界値である．

３．$y_i=\alpha+\beta x_i+\varepsilon_i$ における $H_0:\beta=0$ に対する検定のｔ値をｔと置くとき，$t^2=\dfrac{TSS-RSS}{RSS/(n-2)}$ の関係がある．これを元に決定係数 R^2 を t^2 で表し，決定係数 R^2 の値を使って $H_0:\beta=0,\ \ H_1:\beta\neq 0$ の検定を有意水準５％で行うための境界値をｎ＝３～１０について計算せよ．

＜ヒ＞

$R^2 = 1 - \frac{RSS}{TSS}$ なので，$t^2 = (n-2)\frac{1 - RSS/TSS}{RSS/TSS} = (n-2)\frac{R^2}{1-R^2}$ となる．したがって，$R^2 = \frac{t^2}{t^2+n-2} = 1 - \frac{n-2}{t^2+n-2}$ となる．これは t^2 に関する単調増加関数であるから，$\beta = 0$ の検定の境界値をｔに代入すればよい．答えは，{0.993844,0.9025,0.77148,0.658372,0.569259,0.499474,0.444067,0.399294,0.362487,0.331756,0.305746,0.283463,0.264173,0.247316,0.232465,0.219284,0.207508,0.196926}である．

4．回帰式 $y_i = \alpha + \beta x_i + \gamma z_i + \delta v_i + \varepsilon_i$ において，$H_0 : \begin{cases} \alpha + \beta = \delta + 2 \\ 2\alpha + 4\beta = 2\gamma + 2 \end{cases}$ を検定したい．帰無モデル回帰式はどうなるか？また，このときのＦ統計量を対立モデル回帰式のＲＳＳを RSS_a，帰無モデル回帰式のＲＳＳを RSS_0，標本数ｎを用いて表せ．

（ヒ）

帰無モデルは，$\begin{cases} \alpha = -\gamma + 2\delta + 3 \\ \beta = \gamma - \delta - 1 \end{cases}$ なので，これを回帰式に代入する．$y_i = -\gamma + 2\delta + 3 + (\gamma - \delta - 1)x_i + \gamma z_i + \delta v_i + \varepsilon_i$ となるが，未知係数のある部分は右辺，それ以外は左辺に移項し，$y_i + x_i - 3 = -\gamma + 2\delta + (\gamma - \delta)x_i + \gamma z_i + \delta v_i + \varepsilon_i$ とする．さらに，未知係数でくくる．$y_i + x_i - 3 = \gamma(z_i + x_i - 1) + \delta(v_i - x_i + 2) + \varepsilon_i$．これが，一つの帰無モデルである．それ以外にも帰無モデルは考えられるが，これが一番考えやすいであろう．

5． $\underline{n = 10}$，$\sum_{i=1}^n x_i^2 = 10, \sum_{i=1}^n x_i y_i = 10, \sum_{i=1}^n y_i^2 = 40, \sum_{i=1}^n z_i y_i = 18, \sum_{i=1}^n z_i^2 = 18, \ \bar{x} = 0, \ \ \bar{y} = 1,$ $\bar{z} = 0$ とし，さらに，$\sum_{i=1}^n z_i x_i = 0$ とする．このとき，回帰式 $y_i = \alpha + \beta x_i + \gamma z_i + \varepsilon_i$ において，最小二乗推定値 $\hat{\alpha}, \hat{\beta}, \hat{\gamma}$，残差二乗和 $\sum_{i=1}^n \hat{u}_i^2$，ＴＳＳ，ＥＳＳ，R^2，$\bar{R}^2$ を求めよ．（ｘとｚはどのような関係か？それを利用できないか？）

＜ヒ＞

$\sum_{i=1}^n (x_i - \bar{x})(z_i - \bar{z}) = \sum_{i=1}^n x_i z_i - n\bar{x}\,\bar{z} = 0$ より，β と γ を別々に推定しても良いことがわかる．（スライド３の２２枚目）したがって，$\hat{\beta} = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$，

$\hat{\gamma} = \dfrac{\sum_{i=1}^{n}(y_i - \bar{y})(z_i - \bar{z})}{\sum_{i=1}^{n}(z_i - \bar{z})^2}$ である．これを $\sum_{i=1}^{n} x_i^2$ などから計算する方法は，１を参照

のこと．$ESS = \sum_{i=1}^{n}\left\{\hat{\beta}(x_i - \bar{x}) + \hat{\gamma}(z_i - \bar{z})\right\}^2 = \hat{\beta}^2 \sum_{i=1}^{n}(x_i - \bar{x})^2 + \hat{\gamma}^2 \sum_{i=1}^{n}(z_i - \bar{z})^2$ である

から，ここから，$RSS = TSS - ESS$ で RSS も計算できる．$V(\hat{\beta}) = \dfrac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$ な

ど，１とほぼ同様である．

６．回帰式 $y_i = \alpha + \beta x_i + \varepsilon_i$ において $\hat{\beta}$ の分散について，以下のそれぞれの値が変わること

によって受ける影響について答えよ．X の標本分散が大きくなると（ア，大きくなる，小さくなる，同じ）．また，標本数ｎが大きくなると，（イ，大きくなる，小さくなる，同じ）．誤差項の分散が大きくなると，（ウ，大きくなる，小さくなる，同じ）．真の値βが大きくなると，（エ，大きくなる，小さくなる，同じ）．Ｘの単位を変換して１０００分の１の数値にした場合, 傾きの推定値の分散は,（オ，大きくなる，小さくなる，同じ）．Ｙの単位を変換して１０００分の１の数値にした場合, 傾きの推定値の分散は,（カ，大きくなる，小さくなる，同じ）．Ｙの単位を変換して１０００分の１の数値にした場合，傾きの推定値の分散は，（キ，大きくなる，小さくなる，同じ）．Ｘの単位を変換して１０００分の１の数値にし，かつ，Ｙの単位も変換して１０００分の１の数値にした場合，傾きの推定値の分散は，（ク，大きくなる，小さくなる，同じ）．

（ヒ）

$V(\hat{\beta}) = \dfrac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \dfrac{\sigma^2}{n\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} = \dfrac{\sigma^2}{n\left\{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right\}}$ を用いる．その他の答

えは，練習問題１の解答やスライド１の説明変数，被説明変数の単位変換を参照のこと．

７．新たに説明変数を付け加えた際の説明力の増加＝決定係数の増加を新たな説明変数と被説明変数の偏相関係数 r とその説明変数を入れない回帰における R^2 を用いて表せ．

（ヒ）

教科書ｐ．９５より $RSS(\text{追加変数あり}) = RSS(\text{追加変数なし})(1 - r^2)$ である

から．$\dfrac{RSS(\text{追加変数あり})}{TSS} = \dfrac{RSS(\text{追加変数なし})}{TSS}(1 - r^2)$ となり，

$$1-\frac{RSS(\text{追加変数あり})}{TSS}=1-\frac{RSS(\text{追加変数なし})}{TSS}\left(1-r^2\right)$$ となり，

$$R^2(\text{追加変数あり})=1-\frac{RSS(\text{追加変数あり})}{TSS}$$ ，

$$R^2(\text{追加変数なし})=1-\frac{RSS(\text{追加変数なし})}{TSS}$$ であるから，

$$\begin{aligned}R^2(\text{追加変数あり})&=1-\frac{RSS(\text{追加変数なし})}{TSS}+\frac{RSS(\text{追加変数なし})}{TSS}r^2\\&=1-\frac{RSS(\text{追加変数なし})}{TSS}-\left\{1-\frac{RSS(\text{追加変数なし})}{TSS}\right\}r^2+r^2\\&=R^2(\text{追加変数なし})-R^2(\text{追加変数なし})r^2+r^2\\&=R^2(\text{追加変数なし})\left(1-r^2\right)+r^2\end{aligned}$$

となる.

８．最小二乗推定量はＢＬＵＥ（ブルーと読む）＝最小分散線形不偏推定量である．この事実を用いて，母集団の平均の推定量としては標本平均が最小分散線形不偏推定量であることを示せ.

＜ヒ＞

母集団の平均の推定量を考えるために， $y_i=\mu+\varepsilon_i$ という回帰モデルを考える． $E[y_i]=\mu$ であるので， μ の推定は $E[y_i]$，つまり，母平均を推定していることになる．この回帰モデル $y_i=\mu+\varepsilon_i$ の μ の推定量の内，観測値 y_i の線形和で表される線形推定量で不偏なものの中で最小分散なのは最小二乗推定量である．

$y_i=\mu+\varepsilon_i$ の μ の最小二乗推定量 $\hat{\mu}$ は $\hat{\mu}=\dfrac{\sum_{i=1}^{n}(1\times y_i)}{\sum_{i=1}^{n}1^2}=\dfrac{\sum_{i=1}^{n}y_i}{n}=\bar{y}$ であるので，標本平均が最小分散線形不偏推定量である.

９．母集団の平均の推定量として標本平均が一致推定量であることを示せ.

＜ヒ＞

上の問題と同じ．母集団の平均の推定量を考えるために， $y_i=\mu+\varepsilon_i$ という回帰モデルを考える． $E[y_i]=\mu$ であるので， μ の推定は $E[y_i]$，つまり，母平均を推定していることになる． μ の最小二乗推定量 $\hat{\mu}$ は $\hat{\mu}=\dfrac{\sum_{i=1}^{n}(1\times y_i)}{\sum_{i=1}^{n}1^2}=\dfrac{\sum_{i=1}^{n}y_i}{n}=\bar{y}$ であ

るが，最小二乗推定量は一致推定量なので，標本平均は一致推定量である．

１０． $y_i = \mu + \varepsilon_i$ ， $E[\varepsilon_i] = 0$ ， $V[\varepsilon_i] = \sigma^2$ ， ε_i は i に関して独立とする．

$S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$ において．$E[S^2] = \sigma^2$ であることを示せ.

＜ヒ＞

$y_i = \mu + \varepsilon_i$ で考えると，$S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}\hat{u}_i^2$ であるが，$\hat{u}_i = y_i - \hat{y}_i = y_i - \bar{y}$ であるので，$S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$ である．$V[y_i] = V[\mu] + V[\varepsilon_i]$ であるが，$V[\mu] = 0$ であるから，$V[y_i] = V[\varepsilon_i] = \sigma^2$．ここで，線形回帰の性質から，$E[S^2] = \sigma^2 = V[y_i]$ である．