2017/12/29

藪友良

第 12 章の答え

12.1 (1)それぞれの t 統計量を計算すると、-0.1/0.05=-2、0.15/0.04=3.75、0.2/0.08=2.5 となります。サンプルサイズ 500 もあれば、t 分布は標準正規分布とほぼ同じですから、有意性は t 値の絶対値が 1.96 より大きいかで判断されます。ここで、-2、3.75、2.5 は絶対値で 1.96 より大きいですから、係数が 0 という帰無仮説が棄却されます。

(2)兄弟の数が増えると、自分以外の兄弟への教育費などがかさみ、家計に財政的余裕がなくなると考えられます。これが兄弟の数が増えると、教育年数が減ってしまう原因かもしれません。

(3)父母の教育年数が高いと、彼らの所得も高い傾向にあるでしょうから、家計に財政的な余裕がうまれ、教育により多くの出費を厭わなくなる可能性があります。また、教育年数が高い親ほど、子供の教育により熱心な傾向があるのかもしれません。

(4)太郎君の教育年数は 10-0.1×2+0.15×16+0.2×12=14.6 年と予測されます。

12.2. 睡眠時間を除いて回帰分析をしてみましょう。

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

このとき、1=(X_1+ X_2+ X_3+ X_4)/24 ですから、

$$Y = \alpha\left(\frac{1}{24}X_1 + \frac{1}{24}X_2 + \frac{1}{24}X_3 + \frac{1}{24}X_4\right) + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

となります。上式を展開すると、

$$Y = \left(\beta_1 + \frac{\alpha}{24}\right)X_1 + \left(\beta_2 + \frac{\alpha}{24}\right)X_2 + \left(\beta_3 + \frac{\alpha}{24}\right)X_3 + \frac{\alpha}{24}X_4 + u$$

となります[1]。したがって、睡眠時間 X_4 の効果は $\alpha/24$ として測られます[2]。

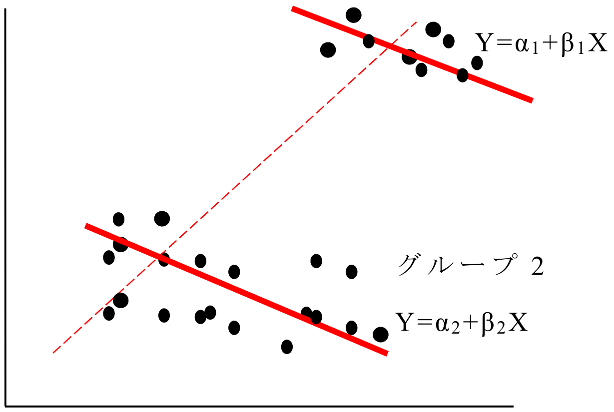
[1] そもそもモデルを定数項なしの式で推定したら、結果の解釈は容易かもしれません。つまり、モデルを $Y=\beta_1^*X_1+\beta_2^*X_2+\beta_3^*X_3+\beta_4^*X_4$ とします（$\beta_1^*=\beta_1+\alpha/24$、$\beta_2^*=\beta_2+\alpha/24$、$\beta_3^*=\beta_3+\alpha/24$、$\beta_4^*=\alpha/24$）。係数 β_i^*は、X_i が 1 単位変化したとき Y が何単位変化するかを表します。

[2] この結果から、β_i は、X_i と X_4 との効果の差を測っていることが分かります。たとえば、X_1 の Y

12.3.双子は能力や家庭環境などにほとんど差がないと考えられますから、双子の変数の差を考えると､能力や家庭環境の影響などを取り除くことができます。したがって、Y を X で回帰すれば、バイアスなく X から Y への影響を測ることができます。これが双子のデータを用いる利点です。

しかしながら、双子のデータを用いれば何の問題もなく、X から Y への影響を正確に推定できるわけではありません。現実問題として、双子であれば教育年数にほとんど違いはないでしょう。つまり、双子の教育年数差 X はほぼ 0 の値をとり､(説明変数の変動が非常に小さいため) 推定結果は不安定となってしまいます。これが双子のデータを用いる欠点です。

4. 様々な原因が考えられますが、1 つの例を紹介します。下図をみると、各グループとも、係数 β はマイナスとなっています。しかし、グル—プを区別しないで一緒にすると、推定された β（図の点線の傾き）はプラスとなります。



5. 以下のような 11 個のダミー変数を定義します。

$$M1_t=\begin{cases}1 & 1\text{月}\\0 & \text{その他}\end{cases}、\quad M2_t=\begin{cases}1 & 2\text{月}\\0 & \text{その他}\end{cases}、\ \cdots、\quad M11_t=\begin{cases}1 & 11\text{月}\\0 & \text{その他}\end{cases}$$

12 個のダミー変数を定義して、説明変数として用いてしまうと、多重共線性が生じてしまうことに注意してください。

への効果は $\beta_1+\alpha/24$、X_4 の Y への効果は $\alpha/24$ ですから、両者の差は β_1 となるわけです。

6. 1)恒等式は、変数間の関係が分かっているため、そもそも式を推定する必要はありません。恒等式の推定は、何の意味もありません。2) GDP の定義から、

GDP = 消費 + 投資 + 政府支出 + 輸出 － 輸入

となります（これはまさに恒等式)。ここで被説明変数を GDP、説明変数を消費、投資、政府支出、輸出、輸入としましょう。このとき、消費、投資、政府支出、輸出の係数は 1、輸入の係数は－1 と推定されます。加えて、決定係数は 1 となります（恒等式ですから、当てはまりは 100%です)。しかし、GDP は定義から、上記の結果が得られることは自明であり、そもそも推定する必要などはないでしょう。

7. モデルは、$Y_{i,t}=\alpha+\beta X_{i,t}+\gamma_1 Z_i+\gamma_2 S_t+u_{i,t}$（$i$=1,⋯,$N$、$t$=1,⋯,$T$）となります。$Z_i$ は時間を通じて一定な i 固有の要因（固定効果)、S_t は時間を通じて変化し全体に同じ影響を与える要因（時間効果）です。たとえば、S_t は景気変動を捉えるマクロ変数かもしれません。このモデルは、$\alpha_i=\alpha+\gamma_1 Z_i$、$\omega_t=\gamma_2 S_t$ と定義すると、

$$Y_{i,t}=\beta X_{i,t}+\alpha_i+\omega_t+u_{i,t}$$

と表現できます。α_i は固定効果、ω_t は時間効果を捉えています。ここで ω_t は、時間を通じて変化する要因をコントロールしています。

時間効果をコントロールするため、以下のダミー変数を定義しましょう。

$$T2_t=\begin{cases}1 & t=2\\0 & \text{その他}\end{cases}、\quad T3_t=\begin{cases}1 & t=3\\0 & \text{その他}\end{cases},\ldots,\quad TT_t=\begin{cases}1 & t=T\\0 & \text{その他}\end{cases}$$

上記のダミー変数は、時間 t だけに依存するため（i に依存しない)、i の添字を除いています。これらの変数だけでなく、固定効果をコントロールするため、以下のダミー変数を定義します(12 章参照)。

$$D2_i=\begin{cases}1 & i=2\\0 & \text{その他}\end{cases}、\quad D3_i=\begin{cases}1 & i=3\\0 & \text{その他}\end{cases},\ldots,\quad Dn_i=\begin{cases}1 & i=n\\0 & \text{その他}\end{cases}$$

このとき、先の式は

$$Y_{i,t}=\beta X_{i,t}+c+\theta_2 D2_i+\ldots+\theta_n Dn_i+\eta_2 T2_t+\ldots+\eta_T TT_t+u_{i,t}$$

と表現できます。ただし、パラメータは $c=\alpha_1+\omega_1$、$\theta_2=\alpha_2-\alpha_1$、…、$\theta_n=\alpha_n-\alpha_1$、$\eta_2=\omega_2-\omega_1$、 …、$\eta_T=\omega_T-\omega_1$ と定義します。ここで i=1、t=1 なら全てのダミー変数が 0 となりますから、上式は

$$Y_{i,t}=\beta X_{i,t}+c+u_{i,t}$$

$$=\beta X_{i,t}+\alpha_1+\omega_1+u_{i,t}$$

となります。また、i=1、t=2 なら $T2$ 以外のダミー変数が 0 となりますから

$$Y_{i,t}=\beta X_{i,t}+c+\eta_2+u$$

$$=\beta X_{i,t}+(\alpha_1+\omega_1)+(\omega_2-\omega_1)+u_i$$

$$=\beta X_{i,t}+\alpha_1+\omega_2+u_{i,t}$$

となります。同じ i であっても、1 期と 2 期で ω の値が異なりますから、時間効果が考慮されていることが分かります。

8. 1）最小 2 乗法は、残差 2 乗和である

$$\sum_{i=1}^{n}\tilde{u}_i^2=\sum_{i=1}^{n}(Y_i-\tilde{\beta}X_i)^2$$

を最小化するように $\tilde{\beta}$ を選びます。上式を $\tilde{\beta}$ で微分して 0 と置くと、

$$-2\sum_{i=1}^{n}(Y_i-\hat{\beta}X_i)X_i=0$$

となります（この式を満たす $\tilde{\beta}$ は最小 2 乗推定量なので $\hat{\beta}$ と表記しています）。さらに両辺を－2 で割ると、

$$\sum_{i=1}^{n}(Y_i-\hat{\beta}X_i)X_i=0$$

となります。さらに整理すると、

$$\sum_{i=1}^{n}X_iY_i-\hat{\beta}\sum_{i=1}^{n}X_i^2=0$$

となり、この式を $\hat{\beta}$ について解くと

$$\hat{\beta}=\frac{\sum_{i=1}^{n}X_iY_i}{\sum_{i=1}^{n}X_i^2}$$

2)真のモデルは $Y_i=\alpha+\beta X_i+u_i$ ですから、

$$E[\hat{\beta}]=\frac{\sum_{i=1}^{n}X_i(\alpha+\beta X_i+u_i)}{\sum_{i=1}^{n}X_i^2}$$

$$=\frac{\alpha\sum_{i=1}^{n}X_i+\beta\sum_{i=1}^{n}X_i^2+E[\sum_{i=1}^{n}X_iu_i]}{\sum_{i=1}^{n}X_i^2}$$

$$=\frac{\alpha\sum_{i=1}^{n}X_i+\beta\sum_{i=1}^{n}X_i^2+\sum_{i=1}^{n}X_iE[u_i]}{\sum_{i=1}^{n}X_i^2}$$

$$=\frac{\alpha\sum_{i=1}^{n}X_i+\beta\sum_{i=1}^{n}X_i^2+0}{\sum_{i=1}^{n}X_i^2}$$

$$=\beta+\frac{\alpha\sum_{i=1}^{n}X_i}{\sum_{i=1}^{n}X_i^2}$$

となります。真のモデルに定数項がなければ（α=0)、不偏性が満たされます。しかし、真のモデルに定数項があれば（$\alpha\neq 0$)、定式化の誤りによって、不偏性が満たされません。

3)定数項がないとき、理論値と残差は

$$\hat{Y}_i=\hat{\beta}X_i、\ \hat{u}_i=Y_i-\hat{\beta}X_i$$

となります。また、残差が満たすべき性質は、

$$\sum_{i=1}^{n}X_i(Y_i-\hat{\beta}X_i)=\sum_{i=1}^{n}X_i\hat{u}_i=0$$

だけであり、この性質から、理論値と残差の積和は 0 となります。

$$\sum_{i=1}^{n}\hat{Y}_i\hat{u}_i=\sum_{i=1}^{n}(\hat{\beta}X_i)\hat{u}_i=\hat{\beta}\sum_{i=1}^{n}X_i\hat{u}_i=0$$

しかし、残差の和が 0 という性質は満たされません。

以上から、Y の全変動は、

$$\sum_{i=1}^{n}(Y_i-\overline{Y})^2=\sum_{i=1}^{n}(\hat{Y}_i+\hat{u}_i-\overline{Y})^2=\sum_{i=1}^{n}((\hat{Y}_i-\overline{Y})+\hat{u}_i)^2$$

$$=\sum_{i=1}^{n}(\hat{Y}_i-\overline{Y})^2+\sum_{i=1}^{n}\hat{u}_i^2+2\sum_{i=1}^{n}(\hat{Y}_i-\overline{Y})\hat{u}_i$$

$$=\sum_{i=1}^{n}(\hat{Y}_i-\overline{Y})^2+\sum_{i=1}^{n}\hat{u}_i^2-2\overline{Y}\sum_{i=1}^{n}\hat{u}_i$$

となります（式展開では$\sum_{i=1}^{n}\hat{Y}_i\hat{u}_i=0$を用いました）。つまり、

$$\frac{\sum_{i=1}^{n}(\hat{Y}_i-\overline{Y})^2}{\sum_{i=1}^{n}(Y_i-\overline{Y})^2}=1-\frac{\sum_{i=1}^{n}\hat{u}_i^2}{\sum_{i=1}^{n}(Y_i-\overline{Y})^2}+\frac{2\overline{Y}\sum_{i=1}^{n}\hat{u}_i}{\sum_{i=1}^{n}(Y_i-\overline{Y})^2}$$

となりますから、

$$\frac{\sum_{i=1}^{n}(\hat{Y}_i-\overline{Y})^2}{\sum_{i=1}^{n}(Y_i-\overline{Y})^2}\neq 1-\frac{\sum_{i=1}^{n}\hat{u}_i^2}{\sum_{i=1}^{n}(Y_i-\overline{Y})^2}$$

を意味します。よって、決定係数として、左辺か右辺のどちらを用いるべきか分かりません。どちらが正しい当てはまりの尺度とも言えませんから、定数項がないモデルでは、決定係数の解釈には注意が必要です。

9. 最小 2 乗法は、残差 2 乗和である

$$\sum_{i=1}^{n} \tilde{u}_i^2 = \sum_{i=1}^{n} (Y_i - \tilde{\alpha} - \tilde{\beta}_1 X_{1i} - \tilde{\beta}_2 X_{2i})^2$$

を最小にするように、$\tilde{\alpha}$、$\tilde{\beta}_1$、$\tilde{\beta}_2$を選びます。まず、$\tilde{\alpha}$で微分して 0 と置くと

$$-2\sum_{i=1}^{n} (Y_i - \hat{\alpha} - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i}) = 0$$

となり、両辺を-2 で割ると

$$\sum_{i=1}^{n} (Y_i - \hat{\alpha} - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i}) = 0$$

となります。上式を満たす$\tilde{\alpha}$、$\tilde{\beta}_1$、$\tilde{\beta}_2$は最小 2 乗推定量なのでハットを付けています。この式を展開すると、

$$\sum_{i=1}^{n} Y_i - n\hat{\alpha} - \hat{\beta}_1 \sum_{i=1}^{n} X_{1i} - \hat{\beta}_2 \sum_{i=1}^{n} X_{2i} = 0$$

となりますから、両辺を n で割って、$\hat{\alpha}$について解くと、

$$\hat{\alpha} = \overline{Y} - \hat{\beta}_1 \overline{X}_1 - \hat{\beta}_2 \overline{X}_2$$

となります。

次に、残差 2 乗和を$\tilde{\beta}_1$で微分して 0 と置くと、

$$-2\sum_{i=1}^{n} (Y_i - \hat{\alpha} - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i}) X_{1i} = 0$$

となり、さらに両辺を-2 で割ると、

$$\sum_{i=1}^{n} (Y_i - \hat{\alpha} - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i}) X_{1i} = 0$$

となります。ここで$\hat{\alpha}$の式を上式に代入すると、

$$\begin{aligned}
&\sum_{i=1}^{n} \Big(Y_i - (\overline{Y} - \hat{\beta}_1 \overline{X}_1 - \hat{\beta}_2 \overline{X}_2) - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i}\Big) X_{1i} \\
&= \sum_{i=1}^{n} \Big((Y_i - \overline{Y}) - \hat{\beta}_1 (X_{1i} - \overline{X}_1) - \hat{\beta}_2 (X_{2i} - \overline{X}_2)\Big) X_{1i} \\
&= \sum_{i=1}^{n} (Y_i - \overline{Y}) X_{1i} - \hat{\beta}_1 \sum_{i=1}^{n} (X_{1i} - \overline{X}_1) X_{1i} - \hat{\beta}_2 \sum_{i=1}^{n} (X_{2i} - \overline{X}_2) X_{1i} = 0
\end{aligned}$$

となります。ここで偏差の和は 0 から[3]、上式は

[3]偏差の和は 0 であることから、以下が成立することに注意してください。

$$\sum_{i=1}^{n} (X_{1i} - \overline{X}_1) X_{1i} = \sum_{i=1}^{n} (X_{1i} - \overline{X}_1) X_{1i} - \sum_{i=1}^{n} (X_{1i} - \overline{X}_1) \overline{X}_1 = \sum_{i=1}^{n} (X_{1i} - \overline{X}_1)(X_{1i} - \overline{X}_1)$$

$$\sum_{i=1}^{n} (X_{2i} - \overline{X}_2) X_{1i} = \sum_{i=1}^{n} (X_{2i} - \overline{X}_2) X_{1i} - \sum_{i=1}^{n} (X_{2i} - \overline{X}_2) \overline{X}_1 = \sum_{i=1}^{n} (X_{1i} - \overline{X}_1)(X_{2i} - \overline{X}_2)$$

$$\sum_{i=1}^{n}(Y_i - \overline{Y})(X_{1i} - \overline{X}_1) - \hat{\beta}_1 \sum_{i=1}^{n}(X_{1i} - \overline{X}_1)^2 - \hat{\beta}_2 \sum_{i=1}^{n}(X_{2i} - \overline{X}_2)(X_{1i} - \overline{X}_1) = 0$$

と書き換えられます。ここで左辺の第 3 項は、標本共分散が 0 との仮定から 0 です。よって、上式は

$$\sum_{i=1}^{n}(Y_i - \overline{Y})(X_{1i} - \overline{X}_1) - \hat{\beta}_1 \sum_{i=1}^{n}(X_{1i} - \overline{X}_1)^2 = 0$$

となり、これを $\hat{\beta}_1$ について解けば

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_{1i} - \overline{X}_1)(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_{1i} - \overline{X}_1)^2}$$

が得られます。つまり、X_1 と X_2 の標本共分散が 0 であれば、X_2 を無視して X_1 だけで回帰しても、X_2 を含めて回帰しても、同じ結果が得られるのです。

12 章では、含めるべき変数を考慮しないとき、推定にバイアスが生じるとしました。しかし、この問題から分かることは、説明変数間の相関がなければ、それらの変数を除いて推定しても何の問題もないことが分かります。換言すれば、説明変数間の相関があるときだけ、それらの変数を除いて推定すると問題が生じるのです。この結果は、説明変数が任意の k 個あっても成立します。