

第5章
重回歸分析
練習問題

問 1

<複数回答>決定係数と自由度調整済み決定係数に関する正しい記述をすべて選択しなさい。

1. 説明変数の数 K が増えると、決定係数は 1 に近づく悪い性質がある。
2. 説明変数の数 K が増えても、データとの当てはまりが悪ければ、自由度調整済み決定係数は減少する。
3. 自由度調整済み決定係数はマイナスになることがある。
4. 自由度調整済み決定係数は決定係数よりも小さい値となる。

問2

<複数回答>モデルは $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + u$ とする。被説明変数を Y 、説明変数を X_1 とした単回帰分析に関する正しい記述をすべて選択しなさい。

1. X_1 と X_2 の標本共分散が 0 なら、係数 β_1 の推定に欠落変数バイアスは生じない。
2. $\beta_2=0$ であれば、係数 β_1 の推定に欠落変数バイアスは生じない。
3. X_1 と X_2 の標本共分散が正、 β_2 が正なら、係数 β_1 の推定に負の欠落変数バイアスが生じる。
4. 推定に X_2 が含まれていないため、係数 β_1 の推定に必ず欠落変数バイアスが生じる。

問3

<複数回答>自由度調整済み決定係数に関する記述として正しいものをすべて選択しなさい。

1. 時系列データでも変化率を用いる場合、自由度調整済み決定係数は低くなる傾向がある。
2. 被説明変数が同じなら、自由度調整済み決定係数によって、複数のモデル間の当てはまりの良さを比較できる。
3. 自由度調整済み決定係数を解釈するときは、その値だけを見るのではなく、状況によって判断することが大事である。
4. 横断面データを用いる場合、自由度調整済み決定係数は高くなる傾向がある。

問4

<複数回答>モデルは $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + u$ とする。多重共線性となる場合をすべて選択しなさい。

1.	$1 = X_1 + X_2$
2.	$X_2 = 1 + 10X_1$
3.	$X_2 = \sqrt{1 + X_1}$
4.	$X_2 = X_1^2$

問 5

<複数回答>弱い多重共線性を表す記述として正しいものをすべて選択しなさい。

1. 弱い多重共線性があっても OLS 推定はできるが、推定結果が不安定になる傾向がある。
2. 弱い多重共線性があっても、サンプルサイズが大きければ、OLS 推定の結果は安定する。
3. 弱い多重共線性があると OLS 推定ができない。
4. 弱い多重共線性があったら、それが経済理論に基づいたモデルであっても、説明変数の一部を推定式から除去したほうがよい。

問6

<複数回答>被説明変数を都道府県の転入超過数(=転出者数-転入者数)とし、説明変数を転出者数と転入者数とする。このとき、正しい記述をすべて選択しなさい。

1. 決定係数は1になる。
2. 説明変数の係数はすべて1になる。
3. この式を推定する意味はない。
4. 当てはまりは良く、優れた定式化となる。

問7

<複数回答>中学2年生の男女の点数差が有意に異なるかを知りたい。被説明変数 Y を数学の点数、説明変数の候補として、男性ダミー M と女性ダミー F を考える。男性ダミー M とは、男性なら1となるダミー変数であり、女性ダミー F とは、女性なら1となるダミー変数である。データを推定した結果、次の定式化が得られた。

$$\hat{Y} = 50.21 - 0.422F$$

このとき、正しい記述をすべて選択しなさい。

1. 説明変数を F ではなく M としたら、その係数は 0.422 となる。
2. 男性の平均点は 50.21 点となる。
3. 男性ダミー M が含まれないので、これは無意味な推定である。
4. 女性は男性より点数が 0.422 点高い。

問 8

<複数回答>中学 2 年生の数学の点数を用いて相対年齢効果(実年齢の違いが成績などに影響を与えること)を検証したい。被説明変数は数学の点数 Y とし、説明変数の候補として、生まれた時期を表すダミー変数 Q_1 、 Q_2 、 Q_3 、 Q_4 を考える。ただし、 Q_1 は 4~6 月生まれなら 1、 Q_2 は 7~9 月生まれなら 1、 Q_3 は 10~12 月生まれなら 1、 Q_4 は 1~3 月生まれなら 1 をとるダミー変数となる。データを分析したところ、次の推定結果が得られたとしよう(サンプルサイズは 4508 となる)。このとき、正しい記述をすべて選択しなさい。

$$\hat{Y} = 50.42 + 0.115Q_2 - 0.322Q_3 - 1.603Q_4$$

(0.297) (0.411) (0.421) (0.430)

1. 1~3 月生まれは、4~6 月生まれより点数が 1.603 点低い。
2. 4~6 月生まれの平均点は 50.42 点となる。
3. 両親が生まれ月をコントロールできるなら、生まれた時期を表すダミー変数と両親の属性との間に相関が生じ、欠落変数バイアスが生じる可能性がある。
4. この推定には、ダミー変数 Q_1 が含まれておらず、意味のない推定結果である。

問9

空欄に当てはまる記号を選びなさい。

誤差項 u_i は標準的仮定を満たすとする。説明変数の数が K としたとき、 $\sum_{i=1}^n (\hat{u}_i/\sigma)^2$ は自由度()の χ^2 分布に従う。

1. n
2. $n-2$
3. $n-K$
4. $n-K-1$

問 10

米国 i 州のビール税 X が上がると飲酒運転が減少し、 i 州の交通事故死亡者数 Y が減少するかを知りたい。モデルとして、 $Y = \alpha + \beta_1 X + \beta_2 W + u$ を考える。分析者の関心はビール税 X の係数 β_1 であり、 W はコントロール変数である。この中で、悪いコントロールとなる W はいずれか。

1. i 州のビール消費額
2. i 州の自動車保有台数
3. i 州の地価
4. i 州の道路舗装状況を表す変数