

データ解析基礎レポート課題（第 11 回）

出題 1 月 9 日（金），提出期限 1 月 19 日（月）
提出先：情報システム学研究科棟 2 階事務室脇のポスト

1. 以下のステップを踏んで重回帰分析の数値実験を行なえ。
 - a. 適当な定数 b_0, b_1, b_2 を各自定めよ。
 - b. 二つの変数 x_1, x_2 の組を 10 個各自適当に定めよ。
 - c. b. で定めた 10 セットについて， $y = b_0 + b_1 x_1 + b_2 x_2 + c\varepsilon$ として，それぞれ y を計算せよ。ただし， ε は標準正規分布に従う乱数である（以前やったように，0 から 1 までの一様乱数を 12 個足して 6 を引いて作ればよい）。
 - d. b., c. で得られた 10 個の入出力関係に基づいて線形回帰により b_0, b_1, b_2 を推定し，a. で定めた真の値と比較せよ。
 - e. 平均 2 乗誤差，誤差の標準偏差の不偏推定量，決定係数をそれぞれ求めよ。
 - f. c の大きさをいろいろと変えて，決定係数がどのように変わるかを調べよ（ c が大きすぎても小さすぎても有意義な結果は得られないので，他の項とノイズ項の大きさを考えて実験すること）。なお，同じ大きさの c に対して繰り返し実験することも重要である。
 - g. b. で定めた 10 個の x_1 に対して， $z = b_0 + b_1 x_1 + b_2 x_1^2 + c\varepsilon$ としてそれぞれ z を計算せよ。定数 c の大きさは，ノイズ項の影響が適度に効くように定めよ。
 - h. 以下の四つのモデルについてそれぞれ線形回帰により a_i を推定し，平均 2 乗誤差 (D^2)，決定係数 (R^2)， $AIC (= n \log D^2 + 2p)$ を求め，モデル間で比較せよ。
 - 1) $z = a_0 + a_1 x_1$
 - 2) $z = a_0 + a_1 x_1 + a_2 x_1^2$
 - 3) $z = a_0 + a_1 x_1 + a_2 x_1^2 + a_3 x_1^3$
 - 4) $z = a_0 + a_1 x_1 + a_2 x_1^2 + a_3 x_1^3 + a_4 x_1^4$
 - i. 異なるノイズや x_1 の組みに対して同じことを何度か繰り返し行ない，AIC の最小値を与えるモデルが元のモデルとどの程度の割合で一致するかを調べよ。
2. 以下の三グループの平均に差があるかどうかを有意水準 1% で検定せよ。このデータについて，post hoc test である Tukey の HSD 検定を行なうことは適切か？もし，適切であれば，検定を行なえ（「スチューデント化された範囲 q の分布」の数表はホームページを参照）。

Group 1	8	7	10	9	11
Group 2	6	8	9	10	7
Group 3	7	11	11	12	9

（次ページに続く）

3. ほぼ同じ値段の三つのブランドの電灯の寿命に差があるかどうかを調べるため、それぞれ六つずつサンプルをとって実際に寿命を測ったところ、以下のような結果が得られた。
- 三つのブランドの平均寿命に差があるかどうかを有意水準 5% で検定せよ。
 - このデータについて HSD 検定をすることは適切か？もしそうならば HSD 検定を行なえ。

Brand 1	23	20	27	25	23	25
Brand 2	33	30	29	28	31	32
Brand 3	21	24	29	27	28	26

4. 第 11 回の講義の内容で理解できなかったこと、また、改善してほしいことなどについて、自由に記述せよ（なければ、特に記載しなくてよい）。