データ解析　課題　第２回

概要：　期待値と分散，回帰分析

提出方法：結果をレポート（紙）にまとめて西８ W のレポートボックス「データ解析」へ提出．

締め切り：５月２３日（金）の 15:00

2.1　最小分散ポートフォリオの導出（マーコビッツ・モデル）

$\boldsymbol{X}$ は m 次元確率ベクトル，$\boldsymbol{w} \in \mathbb{R}^m$ は定数ベクトルとする．とくに X_i は i 番目の資産の収益率とすれば，$Y = \boldsymbol{w}'\boldsymbol{X}$ は重み $\boldsymbol{w}$ で構築したポートフォリオの収益率となる．期待収益率 $E(Y)$ があらかじめ定めた値 μ に等しく，重みの和が 1 という条件

$$E(Y) = \mu, \quad \sum_{i=1}^{m} w_i = 1$$

のときに収益率の分散 $V(Y)$ を最小にする重みベクトル（ただし負の成分を許す）が以下で与えられることを示せ．

$$\boldsymbol{w} = \boldsymbol{\Sigma}^{-1}\boldsymbol{A}\left(\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\right)^{-1}\begin{bmatrix}\mu \\ 1\end{bmatrix}$$

ただし $\boldsymbol{\Sigma} = V(\boldsymbol{X})$ は正定値行列，$\mathbf{1}_m = (1, \ldots, 1)'$ は成分がすべて 1 で長さ m のベクトル，$\boldsymbol{A} = (E(\boldsymbol{X}), \mathbf{1}_m)$ はランク 2 の $m \times 2$ 行列である．

2.2 ポートフォリオの数値例

$m = 5$ **個の資産の期待収益率** $E(X_i)$**，標準偏差** $\sqrt{V(X_i)}$ **を次のように与える．**

$$E(X_1) = 0.2, E(X_2) = 0.15, E(X_3) = 0.1, E(X_4) = 0.05, E(X_5) = 0.08$$

$$\sqrt{V(X_1)} = 0.2, \sqrt{V(X_2)} = 0.2, \sqrt{V(X_3)} = 0.1, \sqrt{V(X_4)} = 0.05, \sqrt{V(X_5)} = 0.1$$

X_i **と** X_j **の相関係数** ρ_{ij} **は**

$$\rho_{ij} = 0.3, \quad 1 \leq i \neq j \leq 4$$

$$\rho_{i5} = \rho_{5j} = -0.5, \quad 1 \leq i, j \leq 4$$

このとき，下記の計算をおこなう R のスクリプトを作成して，実行せよ．スクリプトの定義と実行結果がわかるように「コンソール出力」示すこと．

- **ポートフォリオの期待収益率が** $E(Y) = 0.15$ **のときに** $V(Y)$ **を最小にする重みと，そのときの** $\sqrt{V(Y)}$ **を求めよ．**
- **ポートフォリオの「平均-標準偏差ダイアグラム」を図示せよ．**

- X_5 がない場合 $(w_5 = 0)$ とくらべて，X_5 を導入することによってどのような効果があったのかを議論せよ．

2.3 回帰係数の推定誤差と最適予測

$\boldsymbol{x}_t = (x_{t1}, x_{t2}, \ldots, x_{tm})$, $t = 1, \ldots, n$ が m 次元ベクトルとする．重回帰モデル (multiple regression model) では，x と y に次の関係があると考える．

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} \cdots + \beta_m x_{tm} + \epsilon_t, \quad t = 1, \ldots, n \tag{2.1}$$

誤差 $\epsilon_1, \ldots, \epsilon_n$ が互いに独立で，その分散が $V(\epsilon_t) = \sigma^2$ とする．

- 最小２乗法で推定した回帰係数 $\hat{\beta}$ の期待値ベクトルと分散共分散行列が次式で与えられることを示せ．

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}, \quad V(\hat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1} \tag{2.2}$$

  ただし，利用した重回帰モデル (2.1) が正しいと仮定し，また説明変数 x_{ti} はすべて定数として扱う．

- 任意の $m+1$ 次元ベクトル $\boldsymbol{d} = (d_0, d_1, \ldots, d_m)'$ をひとつ決める．$\gamma = \boldsymbol{d}'\beta$ の不偏推定量が $\hat{\gamma} = \boldsymbol{d}'\hat{\beta}$ で与えられること，および分散が $V(\hat{\gamma}) = \sigma^2\boldsymbol{d}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{d}$ であることを示せ．

- γ の不偏推定量として $y_1, \ldots, y_n$ の重み付き和を考える．重みベクトルを $f \in \mathbb{R}^n$ とすれば，推定量は $f'y$ と書ける．このような線形不偏推定量のなかで分散を最小にするものが，$\hat{\gamma}$ になることを示せ．

2.4　回帰分析の数値例

ある地域の年平均気温は

西暦 (年)	1960	1970	1980	1990	2000
気温（°C）	16.4	16.2	16.4	18.0	17.9

このデータから西暦 2100 年の年平均気温を予測せよ．線形不偏推定量のなかで分散を最小にする予測値と，その標準誤差を計算する R スクリプト，および，その実行結果を示せ．

ヒント： ここで「標準誤差」とは，線形不偏推定量 $\hat{\gamma}$ の標準偏差（分散の平方根）の意味である．ただし，σ^2 の真値は未知だから，これをデータで推定したもので置き換える必要がある．通常は，残差 e_t をつかって「不偏分散」

$$\hat{\sigma}^2 = \frac{1}{n-(m+1)} \sum_{t=1}^{n} e_t^2$$

を計算する．この例では $m = 1$．

2.5 主成分の導出

$\boldsymbol{x}_t = (x_{t1}, x_{t2}, \ldots, x_{tm})$, $t = 1, \ldots, n$ が m 次元ベクトルとする．これらを n 個の点とみなし，m 次元の単位ベクトル $\boldsymbol{v} = (v_1, \ldots, v_m)$ 方向への射影を考える．次の合成変量

$$y_t = v_1 x_{t1} + \cdots + v_m x_{tm}, \quad t = 1, \ldots, n$$

を定義すれば，射影した点は $y_t \boldsymbol{v}$, $t = 1, \ldots, n$ と書ける．この誤差の２乗和

$$\sum_{t=1}^{n} \|\boldsymbol{x}_t - y_t \boldsymbol{v}\|^2$$

を最小にする v をもとめよ．

2.6 主成分分析の数値例

`X2000data.txt` を読み込み，変数を１０個程度選ぶ．主成分分析を行うプログラムを実装して，バイプロットを図示せよ．

- 関数定義を含め，データの読み込みから結果の出力までをおこなう「スクリプト」を作成する．図示のために `biplot` 関数を利用するのはかまわないが，主成分の計算は行列演算を用いて自ら実装すること．

- **データをあらかじめ「標準化」してから分析すること．**
- **スクリプトを実行したときの，コンソール出力を示せ．スクリプトの定義，および，実行結果が含まれるようにしておくこと．バイプロットを印刷してレポートに含めること．**
- **選んだ変数には，**E09504 **のようなコード番号ではなく，**Gakureki **のようなわかりやすい名前をつけること．さらに，選んだ変数の説明を** `X2000name.txt` **からコピーして示すこと．**
- **友達とレポートの相談をするのはかまわないが，全く同じ変数の組み合わせにならないように配慮すること．** `gakureki-rikon-12.txt` **にしめした１２変数からは変数を選ばないこと．**
- **分析結果にたいして，なんらかの解釈を与えること．**