第３回 課題

1．　一般逆行列 A^+ を計算する関数 `geninv` を作れ．ただし，最大特異値との比が `tol`(デフォルト値 10^{-7}) 以下の特異値はゼロとみなす．

```
geninv <- function(A,tol=1e-7) {
ここでAの一般逆行列を計算
}
```

2．次の二つの行列について一般逆行列を計算し，数値計算の誤差を除いて $AA^+A - A = 0$ となることを確かめよ．

```
A1 <- matrix(1:15,5)
A2 <- matrix(rnorm(15),5)
```

3．上記の二つの行列について A^+A を計算し，もし単位行列でない場合はその理由を述べよ．

第４回 課題

１．直線当てはめ（単回帰）の関数kaiki1を作れ．

```
kaiki1 <- function(x,y) {
# x,yは同じ長さの実数ベクトル
# y = coef[1] + coef[2]*x + residの形の単回帰分析を行う
# 以下のcoef, pred, residを計算する
#   coef(係数)は２次元ベクトル
#   resid(残差)はyと同じ長さのベクトル
#   pred(予測値) = coef[1] + coef[2]*xはyと同じ長さのベクトル
# 次の行は結果をリストとして返す.
list(coef,pred,resid)
}
```

２．重回帰分析の関数kaiki2を作れ．

```
kaiki2 <- function(x,y) {
# xはn * p次元の行列
# yは長さnのベクトル
```

```
# y = coef[1] + coef[2]*x[,1] + ... + coef[p+1]*x[,p] + resid
#   の形の重回帰分析を行う
# 以下のcoef, pred, residを計算する
#    coef(係数)はp+1次元ベクトル
#    resid(残差)はyと同じ長さのベクトル
#    pred(予測値)はyと同じ長さのベクトル
# 次の行は結果をリストとして返す.
list(coef,pred,resid)
}
```

3. kaiki2の返す値から重相関係数の二乗R^2を計算する関数jyusokansqを作れ.

```
jyusokansq <- function(kout) {
# kout$predは予測値, kout$residは残差
# これらから重相関係数の二乗を計算しrsqに代入
rsq
}
```

4. kaiki2, jyusoukansqを使い，X2000$xから適当な項目を選んで重回帰分析する．係数$\beta$と重相関係数$R$を計算する．myfunc20020919.Rにあるmylsfitをつかって同じ分析をして，結果が同じになるかどうか確認する．

5. 上で得られた結果について，predをx軸，yをy軸とするプロットをする．x軸=y軸となる直線を描く(abline(0,1)をつかう)．さらに県名を使ったプロットをする．myfunc20020919.Rのmyplot関数を参考にせよ．プロットはmyfunc20020919.Rにあるpsinit関数などを使いepsファイルとして出力し，それをプリンタで印刷する．

```
psinit("ファイル名") # これ以後のプロットの結果をファイルにeps形式で書き出す
ここでプロットをおこなう...
dev.off() # ファイルをクローズする
```

第５回 課題

１．平均ベクトルと共分散行列が次のように与えられる２次元正規分布に従う確率変数xを10000個生成し，サイズ2×10000の行列に代入せよ．

$$\mu = \begin{bmatrix} 10 \\ 20 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 2 \end{bmatrix}$$

ヒント： コレスキー分解$\Sigma = R'R$と$z \sim N_2(\mathbf{0}, \boldsymbol{I}_2)$から$x = R'z + \mu$と書ける．

２．　上記の確率変数の成分の和$y = x_1 + x_2$のヒストグラムを`hist(y,prob=T)`などを使って書き，そこにyの確率密度関数の理論曲線を重ね合わせて表示せよ．

３．適当な自由度のカイ二乗分布に従う確率変数を２種類以上生成し，それらの和が再びカイ二乗分布になることをヒストグラムと確率密度関数の理論曲線を使って例示せよ．

4．まず次のようにデータ$(x_i, y_i), i = 1, \ldots, n$を生成し，これに３次の多項式回帰モデルを当てはめ推定した回帰係数を示せ．データの散布図に推定した曲線を重ねて表示せよ．

```
n <- 30  # データ数 n = 30
x <- runif(n,min=-3,max=3)  # x ~ U(-3,3)をn個生成
y <- 3 + 2*x + x^2 # 理論式を y = 3 + 2x + x^2  とする
y <- y + rnorm(n,mean=0,sd=1) # 誤差をN(0,1)とする.
```

5． 上記のデータについて決定係数R^2を計算せよ.「モデル(0)の検定」を行い，そのF統計量と確率値を示せ．それを`f$fsummary`と比較せよ．

6． 上記のデータについてすべての回帰係数の有意性を検定せよ．t統計量と確率値を示せ．それを`f$tsummary`と比較せよ．

第6回 課題

1．信頼領域の計算で用いた

$$r_1(p) = \sqrt{(p+1)F^{\alpha}_{p+1,n-p-1}}, \qquad r_2(p) = \sqrt{F^{\alpha}_{1,n-p-1}}$$

について $n = 30, \alpha = 0.05$ とおき，$r_1(p)$ と $r_2(p)$ を $p = 0, 1, \ldots, 10$ の範囲で計算せよ．$r_1(p)$ と $r_2(p)$ の比較をして違いを述べよ．それは回帰直線(曲面)の信頼領域についてどのような結果をもたらすか?

2．雪日数(B02304)をｘ，最高気温(B02102)をｙとする多項式回帰分析を次数 $p = 1, 2, 3$ について行え．それぞれの次数についてｘ,ｙの散布図上に推定した回帰直線（曲線）とその95％同時信頼区間および95％信頼区間を重ねて描け．