

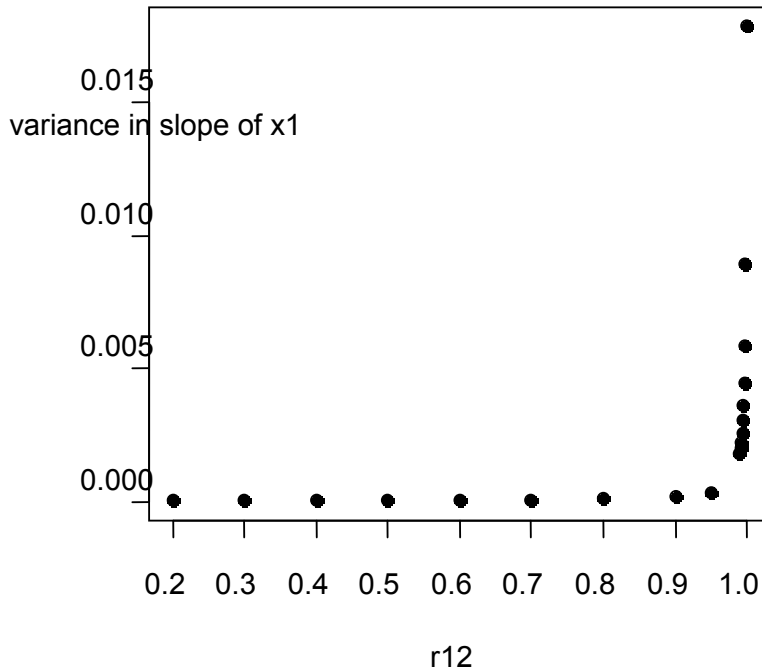
## 多重共線性—計算してみる

重回帰では、説明変数間に一次式で表される関係やそれに非常に近い関係があると、係数(偏回帰係数)の推定が不安定になる。多重共線性(multicollinearity)である。

重回帰で最も単純な、2説明変数の場合について計算して、多重共線性の起こる様子を見てみた。以下の例では、説明変数が  $x_1$  と  $x_2$  の2つ、サンプルサイズが  $n=1000$  のときに  $x_1$  と  $x_2$  の間の相関によってそれぞれの説明変数の係数(偏回帰係数、単回帰なら傾きにあたる)の分散がどう変わるかを見たものである。

データは母平均が  $x_1$  は 1.0 で  $x_2$  は 1.4、母分散が  $x_1$  と  $x_2$  が両方とも 4.0、 $x_1$  と  $x_2$  との間の母相関係数を  $r_{12}$  とし(相関係数としては、いわゆる普通の相関係数  $r$  [つまり細かく言うと積率相関係数を使っている])、2変量正規乱数でまず  $x_1$  と  $x_2$  とを生成した。次に平均が  $1.7x_1 + 3.3x_2 + 12.8$  で標準偏差が 1.2 の正規乱数で目的変数を生成した。これを  $n=1000$  だけ生成したものが1つのサンプルである。1つの  $r_{12}$  の値について一万回計算して、説明変数の係数の分散を求めた。計算は R で行なっている

以下の図は、説明変数  $x_1$  の係数の分散を  $r_{12}$  に対してプロットしたものである。説明変数間の相関係数  $r_{12}$  の値が 0.95 付近よりも大きくなるあたりで、大きな変化が起こり、説明変数間の相関係数が非常に大きいと、説明変数のばらつきが急激に大きくなることがわかる。このばらつきの急激な増大が、説明変数間に非常に強い関係があるときに係数の推定が不安定になること、すなわち多重共線性である。



以下の図は、説明変数  $x_2$  の係数の分散を  $r_{12}$  に対してプロットしたものである。上と同様に、説明変数間の相関係数  $r_{12}$  の値がおよそ 0.95 よりも大きくなると、大きな変化が起こっている。説明変数間の相関がとても強いと、説明変数のばらつきが急激に大きくなることがわかる。

