

今回・シ

「あぶない」時

やめま

Danger!!

(危1) 時系列デー

(危1) 時系列デー

Do NOT apply GLM to time



Danger! time-ser

(危2) 時系列

「見せかけの回帰

時系列データの結

```

spurious_regression.R *
Source on Save Run
1 x <- cumsum(rnorm(100))
2 y <- cumsum(rnorm(100))
3 plot(ts(x), col = "blue", ylim = range(x, y))
4 lines(ts(y), col = "red")
5 print(summary(glm(y ~ x))$coefficients)

5:40 (Top Level)
Console ~1
> plot(ts(x), col = "blue", ylim = range(x, y))
> lines(ts(y), col = "red")
> print(summary(glm(y ~ x))$coefficients)
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.67120    0.90288  -1.8510 6.7186e-02
x              0.64551    0.10803   5.9753 3.7127e-08

```

・ 安易に「回帰」し

・ ランダムウォーク

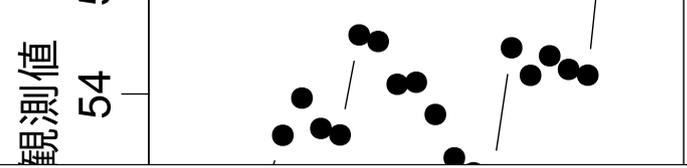
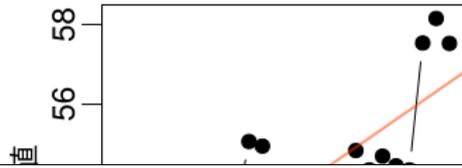
・ 統計モデルが生

パターンを意識す

・ 階層ベイズモデ

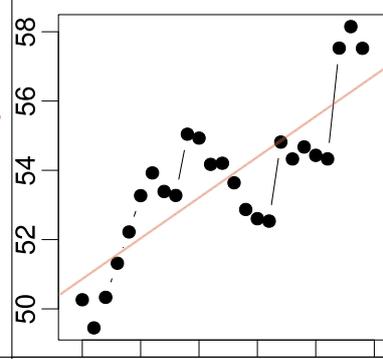
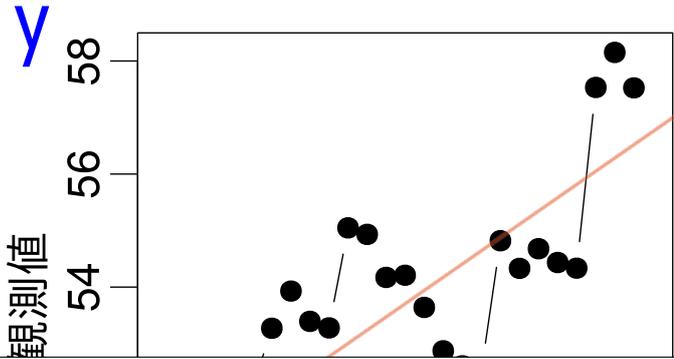
Use state-space

No! Time_series



時系列データの観測値

「やったーゆーし」



```
> summary(glm(form
Deviance Residuals
  Min      1Q
-2.1295 -1.0583
```

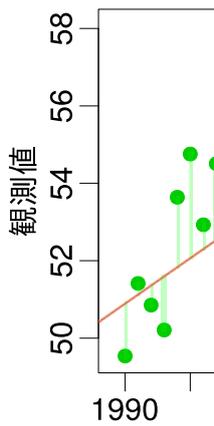
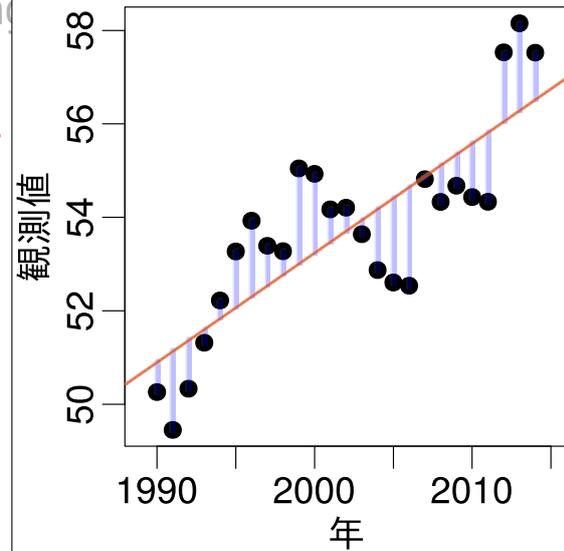
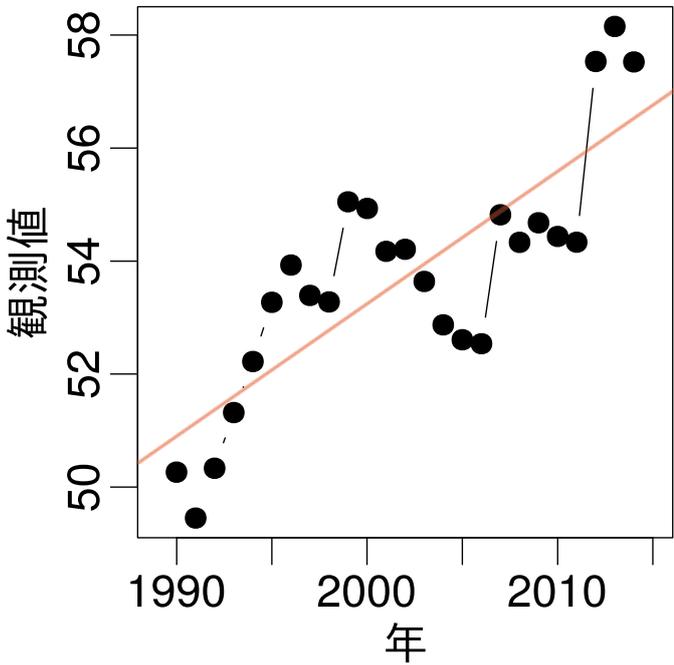
時系列の各点

時系列の「ずれ」

time autocorrelation among

auto-correlation

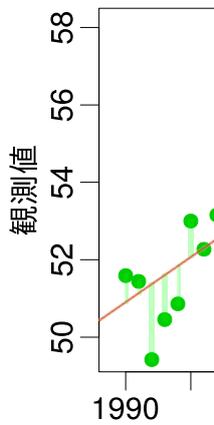
no cor

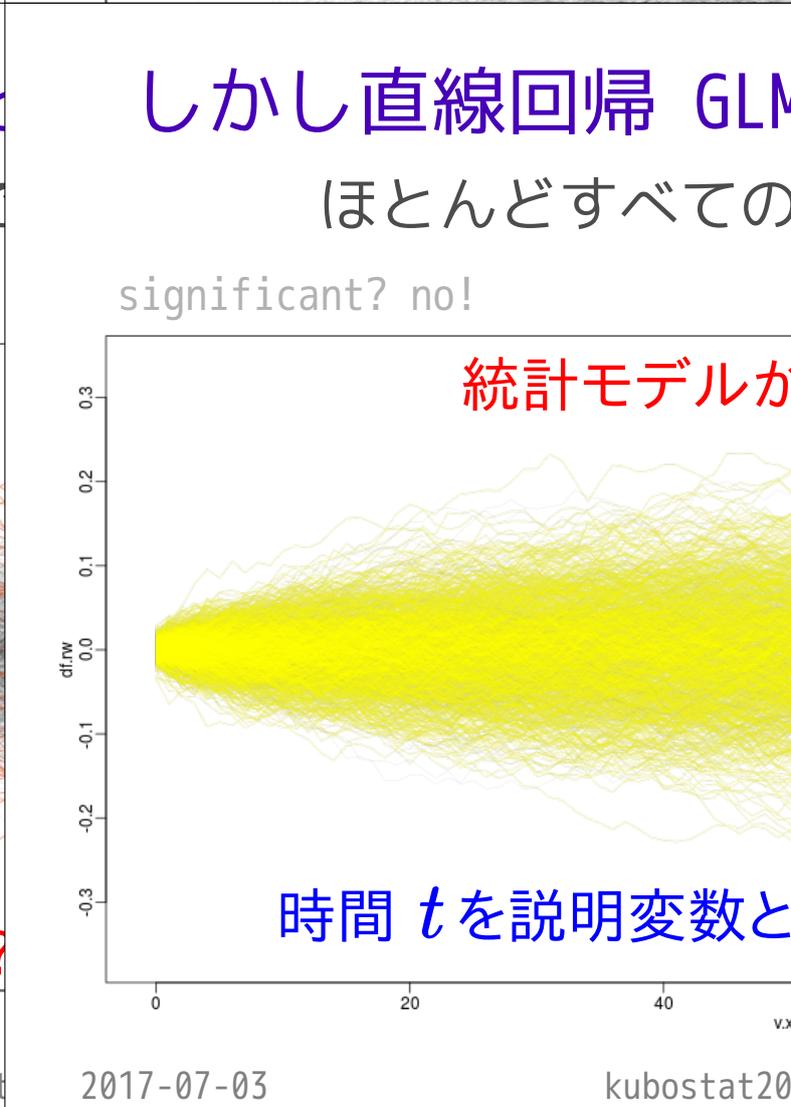
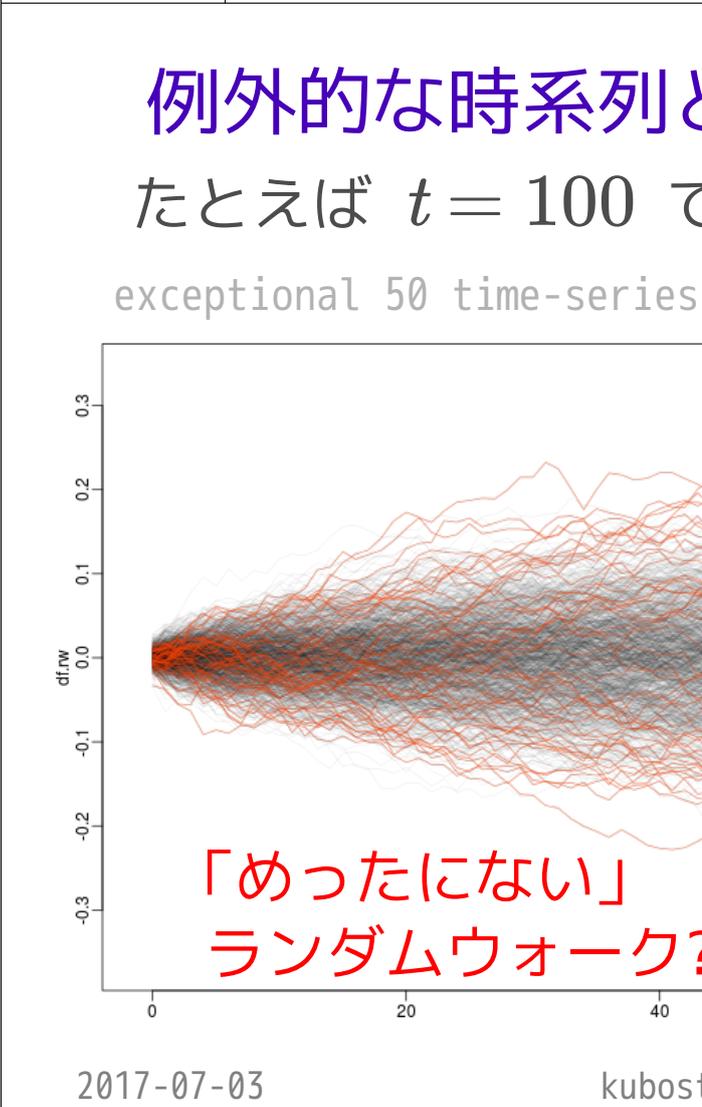
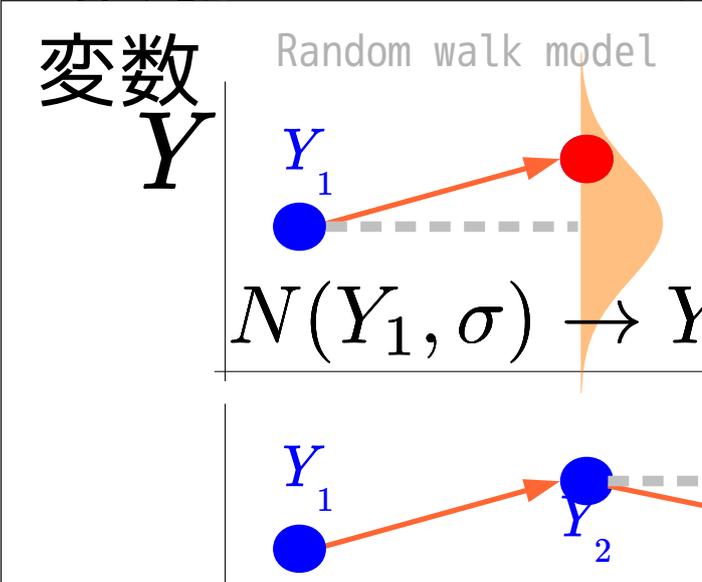
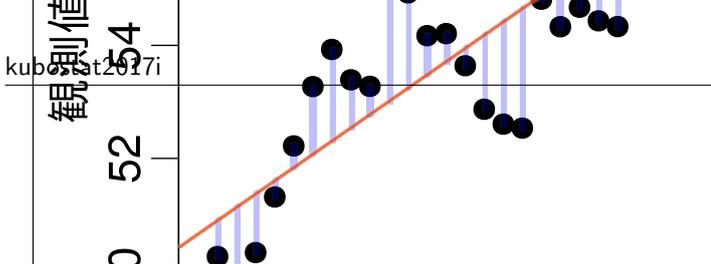


検定とかモデル選択と

ずれかたが
ちがってる?

統計モデ





しんはデータ
なのに

を調べたら

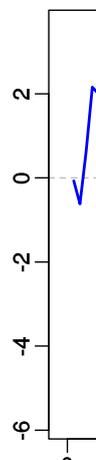
R の ts クラス: `plot(ts(Y))`

`plot(ts(Y))`

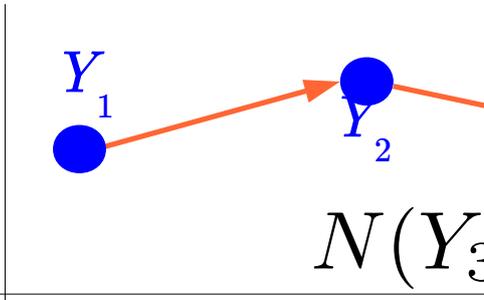
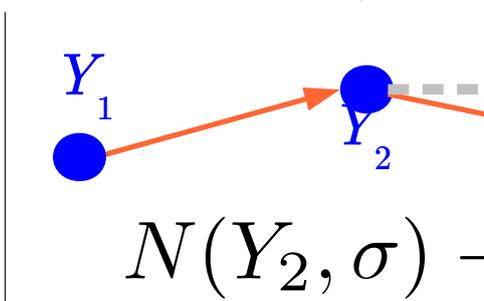
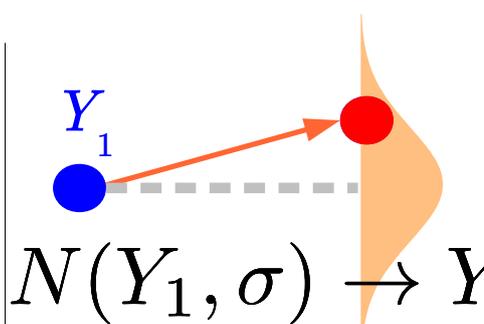
これはたんなる
100 個の正規乱数

自己相関減衰の概

`plot(ts(Y))`



変数
 Y



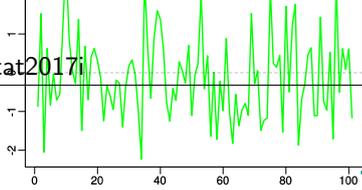
temporal auto-correla

時間的自

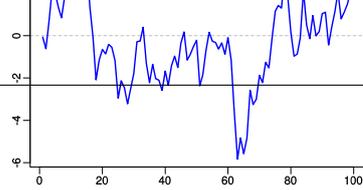
いつも役にたつ

$$\rho_k = \frac{\text{Cov}(y_t, y_{t+k})}{\sqrt{\text{Var}(y_t) \text{Var}(y_{t+k})}}$$



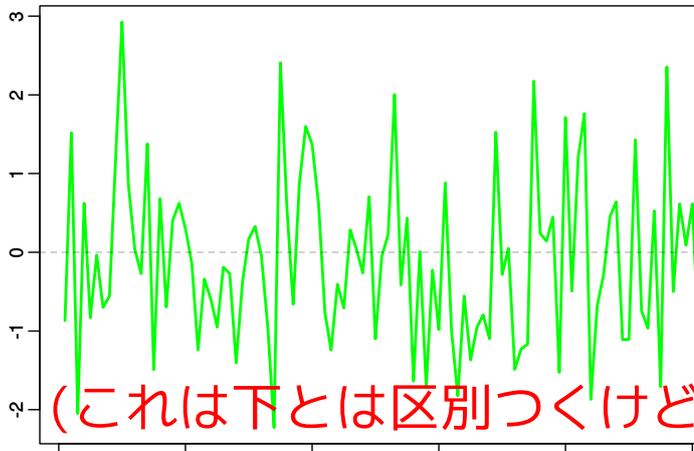


こいつに
ナナメに
したもの
なんだけど..



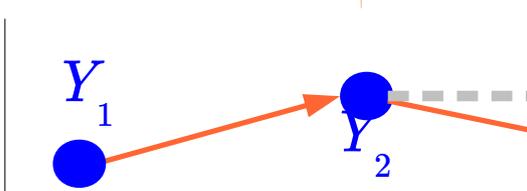
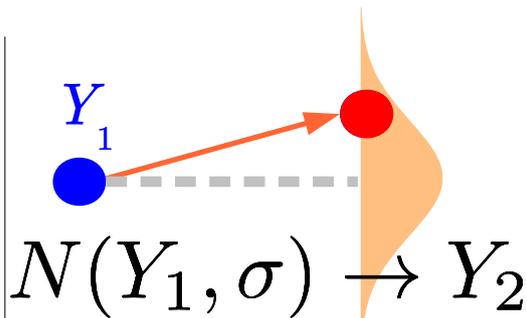
こいつに
ナナメに
したもの

自己相関係数み



(これは下とは区別つくけど)

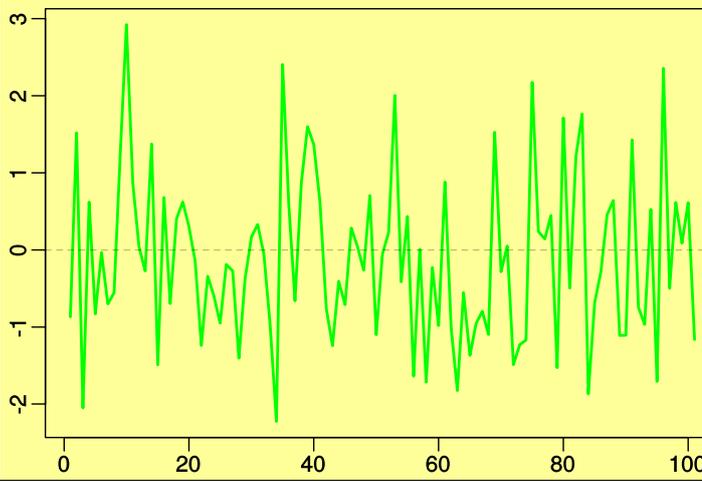
変数 Y



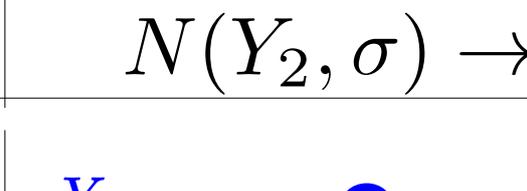
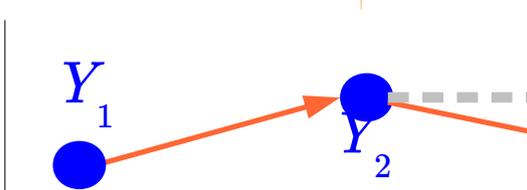
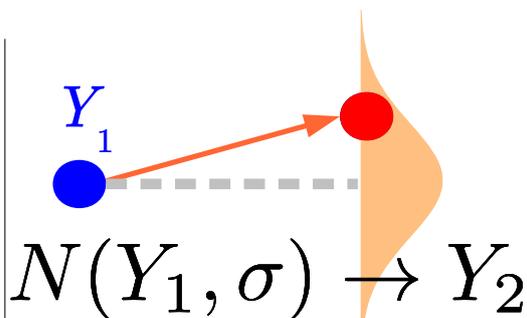
状態空間モデ

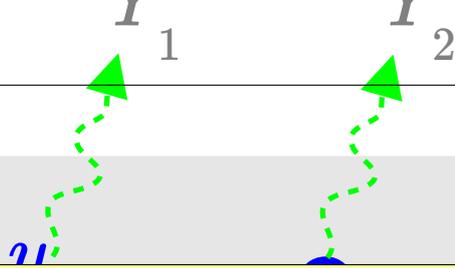
時系列ラ

いろいろな時
統一的にあ

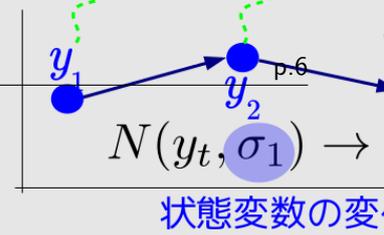


変数 Y





σ_2 大
小



状態空間

state-space

階層ベイズ

どうやってモデ



R の状態
package

library

library

しかしより一般

ついでの理

階層ベイズモ

多数の「似たようなパ
「適切」な制約をか

全デー

こういう問題

BUGS 言語で

階層ベイズモデ



モデルの構造
データと
パラメーター
の初期値
サンプリング
の詳細
Input

BUGS言語

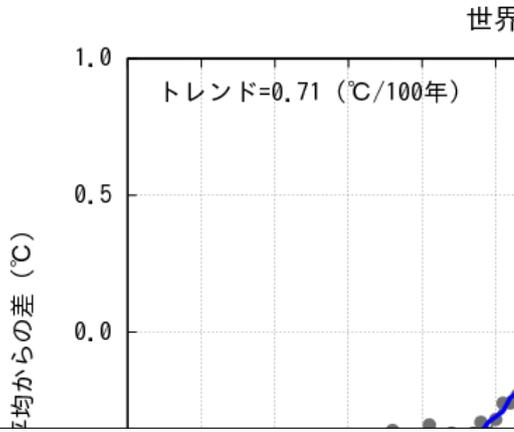
JAGS

```
for (t in 2:N.Y)
```

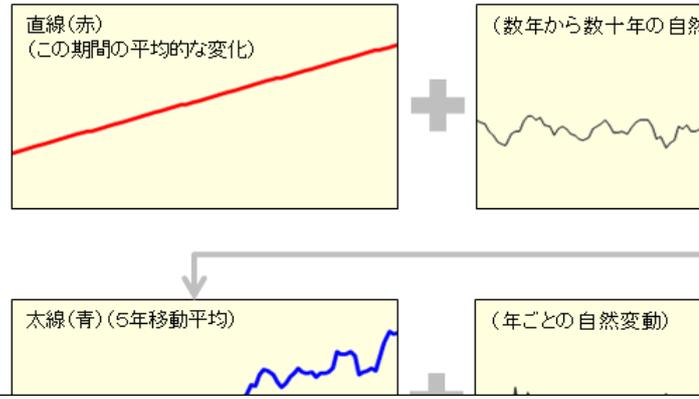
```
Y[t] ~ dnorm(
y[t] ~ dnorm(
m[t] <- delta
```

long-term change of

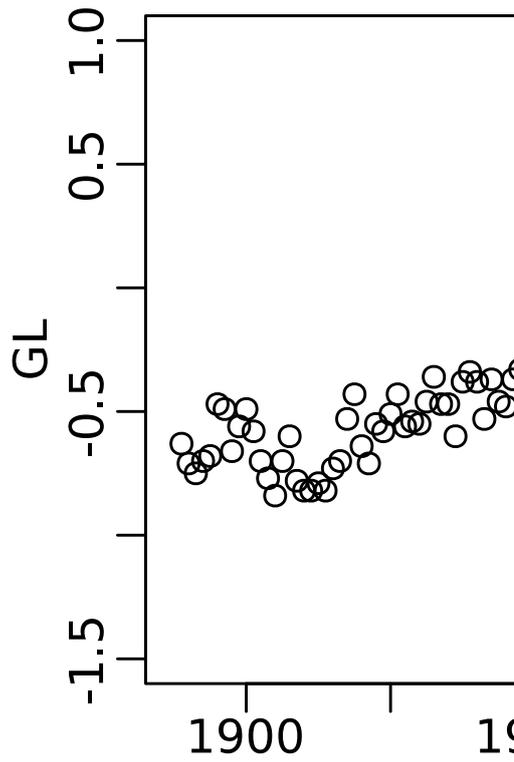
気象庁の長期変化傾向



気象庁の長期変化傾向



公開データを



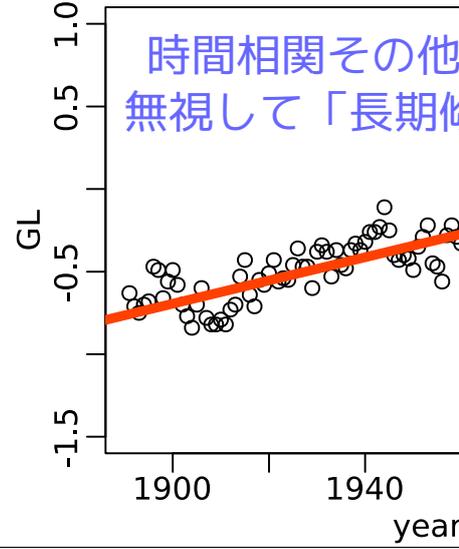
「とりあえず、直線

```
> summary(glm(GL ~ year, d
```

Coefficients:

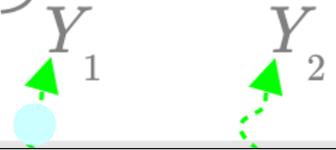
	Estimate	Std.
(Intercept)	-1.41e+01	6.
year	7.03e-03	3.

100年
あたり
0.70°C



	Estimate	S
(Intercept)	-1.41e+01	
year	7.03e-03	

観測データ



状態空間モデル：

ランダムウォーク



状態空間モデル：す

```

Y[1] ~ dnorm(y[1],
y[1] ~ dnorm(0.0,
for (t in 2:N.Y) {
  Y[t] ~ dnorm(y
  y[t] ~ dnorm(m
  m[t] <- delta
}
delta ~ dnorm(0, T
for (t in 1:2) {

```

GLM under-estimate

状態空間モデルが

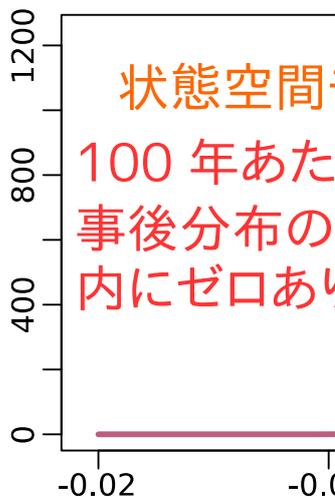
観測値間に相関あり→サン

> summary(glm(GL ~ year

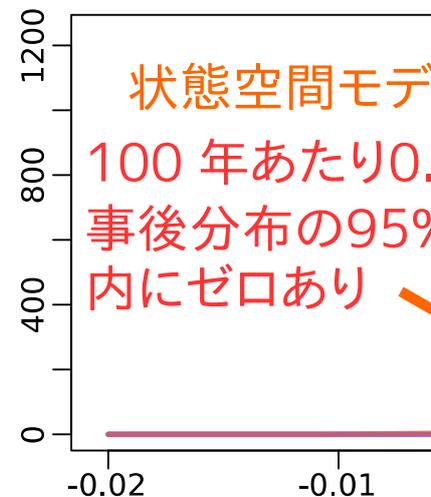
Coefficients:

	Estimate	S
(Intercept)	-1.41e+01	
year	7.03e-03	

100年あたり 0.70°C



100年あたり 0.70°C



・ GLM: $Y(t) \sim t$ と

・ 統計的解析・検定

時系列

「見せかけの回帰

ノイズの大きな時系列

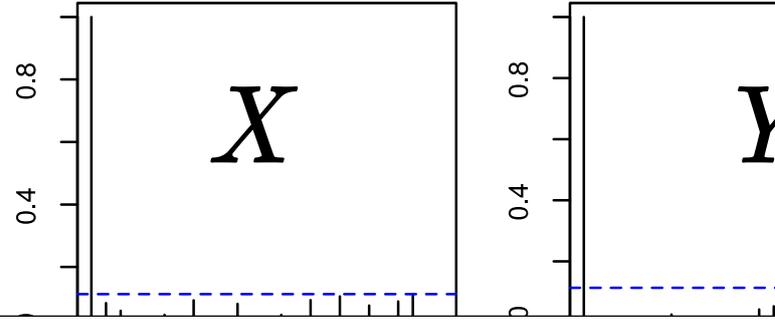
時間的自己相関のない時

```

spurious_regression.R x
Source on Save Run
1 x <- cumsum(rnorm(100))
2 y <- cumsum(rnorm(100))
3 plot(ts(x), col = "blue", ylim = range(x, y))
4 lines(ts(y), col = "red")
5 print(summary(glm(y ~ x))$coefficients)

5:40 (Top Level)
Console ~|
> plot(ts(x), col = "blue", ylim = range(x, y))

```

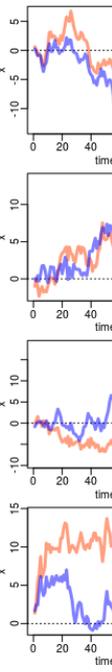
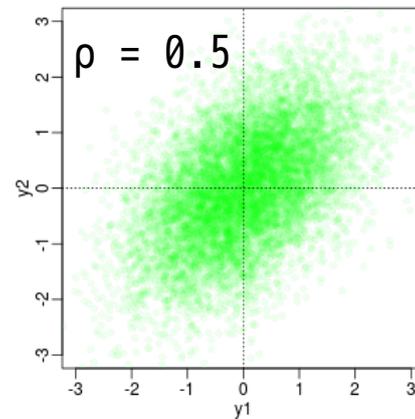
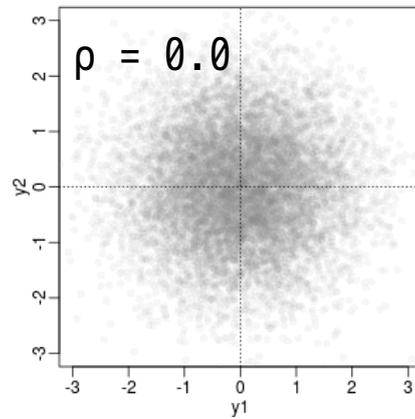


二変量正規分布とラ

疑わし spurious

状態空間モ

あつか



```

}
mu[1] ~ dunif(-1.0E+4, 1.0E+4)
mu[2] ~ dunif(-1.0E+4, 1.0E+4)
Omega[1:2, 1:2] <- diag(1, 2)
VarCov[1, 1] <- sigma^2

```

```

3 chains, each with 5200 iterations (total 15600)
n.sims = 15000 iterations saved
      mean      sd   2.5%   25%
mu[1]  -0.122  0.110 -0.342 -0.195
mu[2]  -0.157  0.100 -0.355 -0.224

```

時間的な相関

情報量を減

おれ

時系列の「ずれ」



時系列データの統計モデリング

- ・ 安易に「回帰」してはいけない
- ・ ランダムウォークモデルが基本
- ・ 統計モデルが生成する時系列パターンを意識する
- ・ 階層ベイズモデルで推定

状態空間モデル