



Topic 5.

相関と回帰

スコットアトム

大学サッカー界のデータ解析を牽引
するためのスキルを身につける。

それぞれの指標の意味

平均値

→ 観測値の総和を観測値の個数で割ったもの

中央値

→ データを順に並べたとき中央に位置する値

最頻値

→ データ群や確率分布で最も頻繁に出現する値

標準偏差

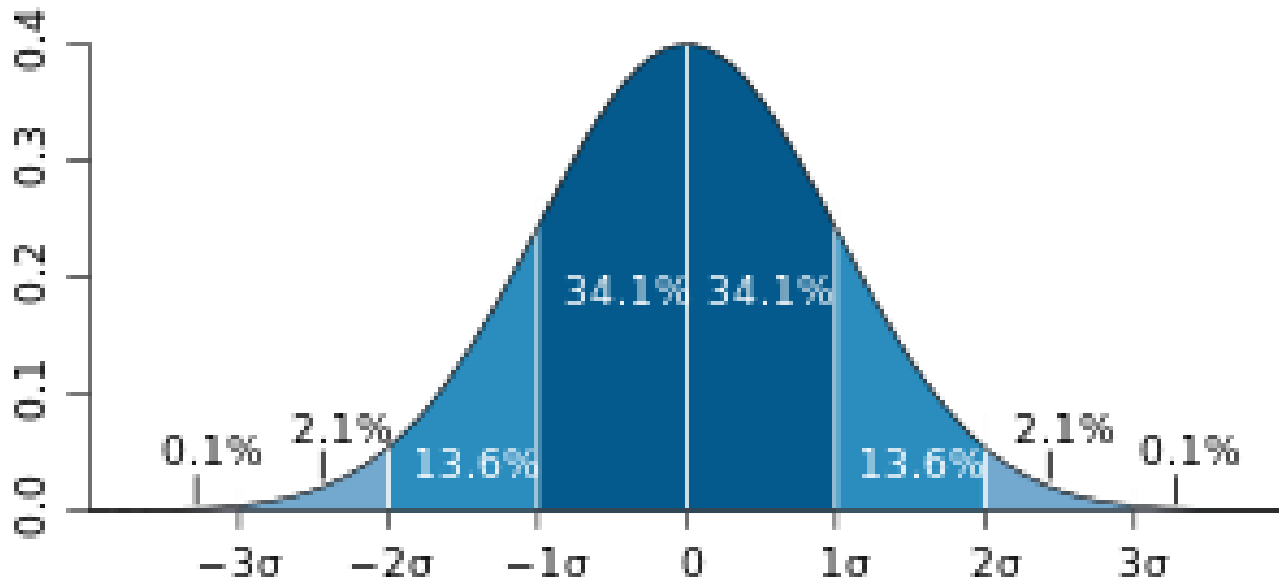
→ データや確率変数の散らばり具合を表す数値

分散

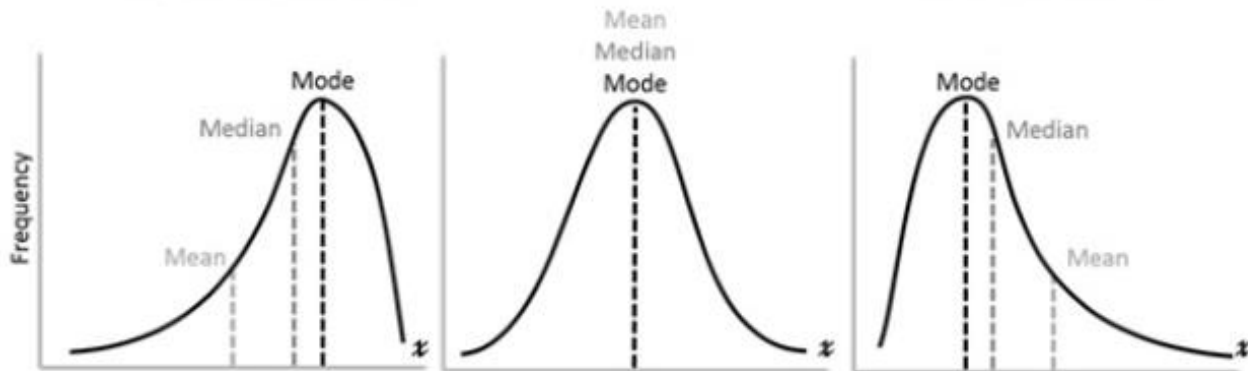
→ データや確率変数の散らばり具合を表す数値

分散 = 標準偏差の2乗

記述統計補足② : Additional Info



記述統計補足③ : Additional Info



負の非対称分布

対称分布

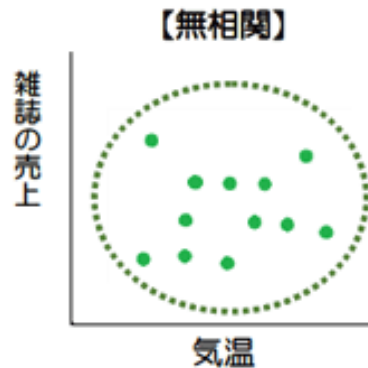
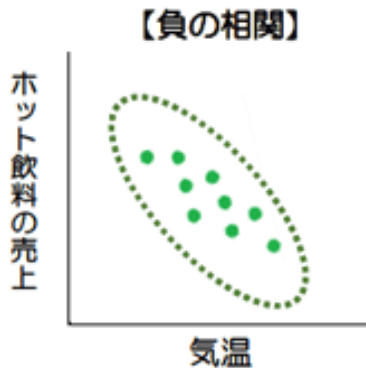
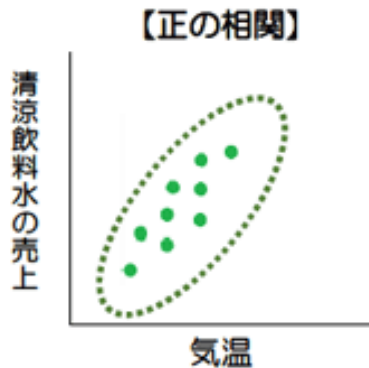
正の非対称分布

対象分布なら平均・中央値・最頻値は一致する.

確認してみよう エクセルファイルへ

相関とは

二つのものが密接にかかわり合っていること。



相関 ②: Correlation



相関係数とは：どれぐらい相関が強いのか

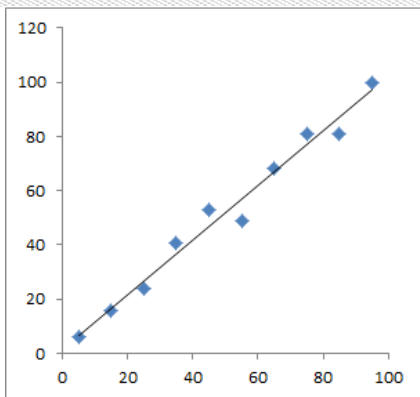
相関係数	相関の強さの目安	相関係数	相関の強さの目安
0～0.3未満	ほぼ無相関	0～-0.3以上	ほぼ無相関
0.3～0.5未満	非常に弱い正の相関	-0.3～-0.5以上	非常に弱い負の相関
0.5～0.7未満	正の相関がある	-0.5～-0.7以上	負の相関がある
0.7～0.9未満	強い正の相関	-0.7～-0.9以上	強い負の相関
0.9以上	非常に強い正の相関	-0.9以下	非常に強い負の相関

相関関係は因果関係を含意しない
(Correlation does not imply causation)

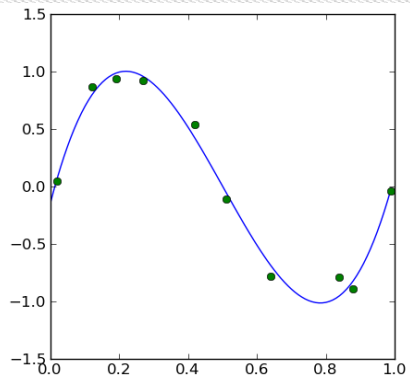
相関関係を考えるときは必ず因果関係を考慮しなければならない

回帰とは

データをモデル(線)をあてはめること



線形回帰

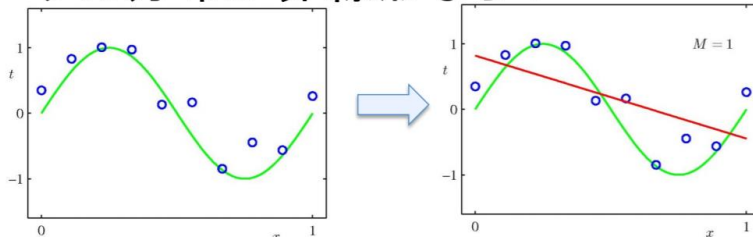


非線形回帰

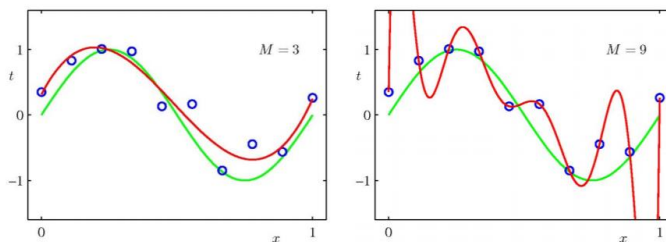
回帰 ①: Regression

回帰は難しい

- データの分布が非線形なら？



- どうなるとうれしい？



今回は線形回帰しか考えない。

確認してみよう エクセルファイルへ