

大学生のための 統計学入門

—テーマ4. 2変量の関連—

京都大学大学院医学研究科
臨床統計学/臨床統計家育成コース 田中司朗



テーマ4. 2変量の関連

- 地球の大きさを測る
 - 相関
 - 回帰分析
 - 最小二乗法
- 相関と因果
- ランダム化臨床試験と医療の進歩
 - ランダム化による因果関係の証明

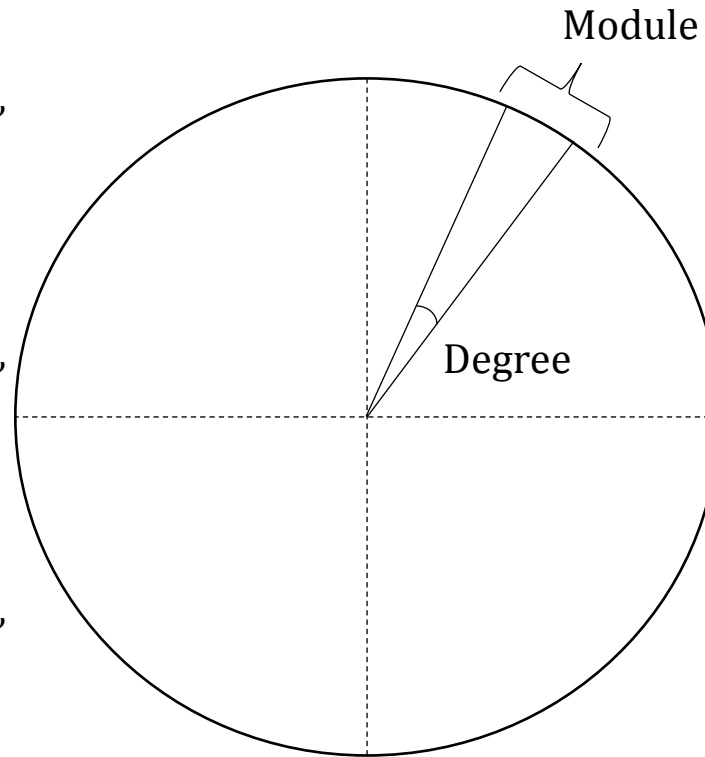
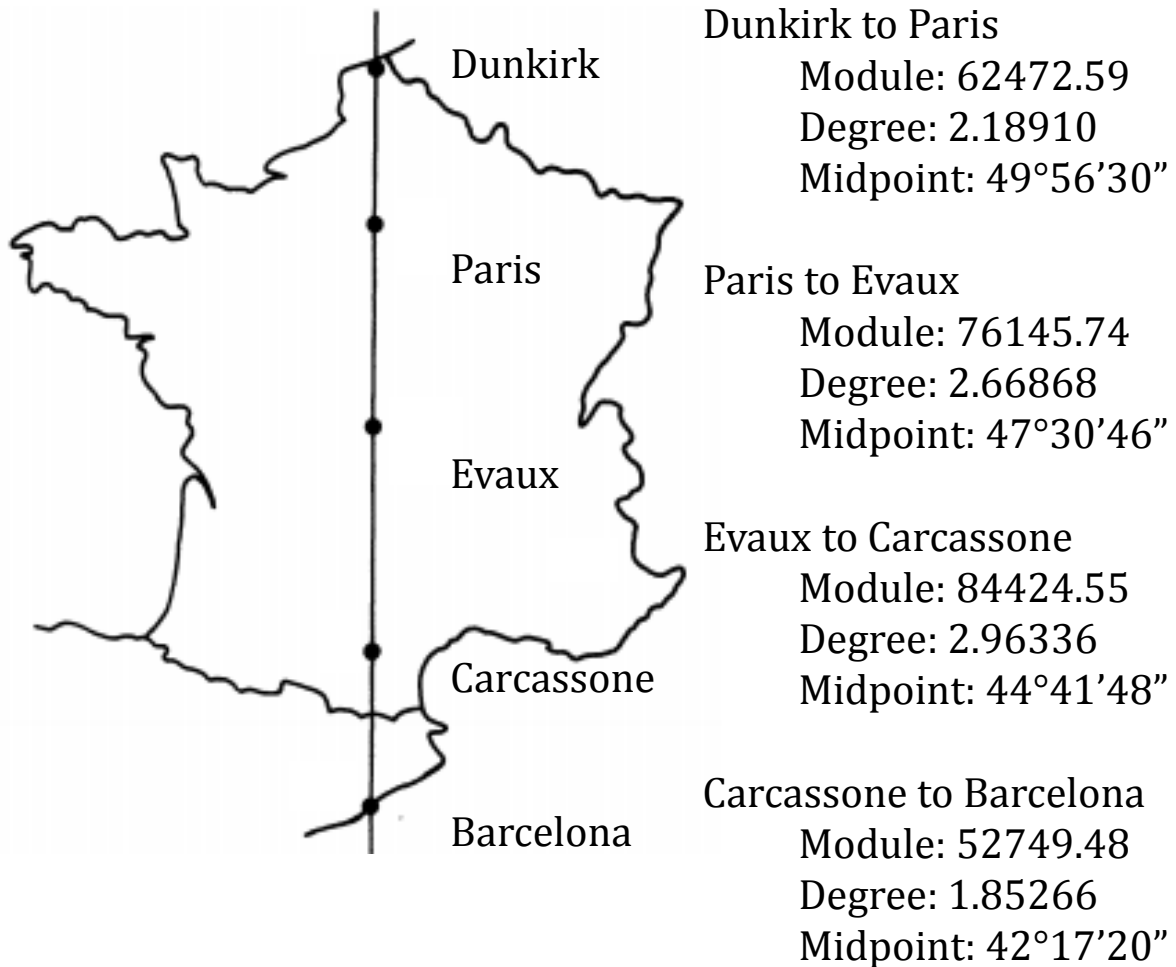


地球の大きさに関する質問

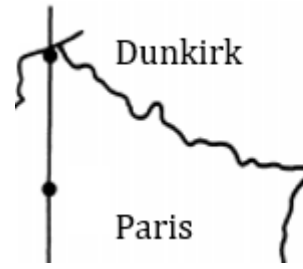
- 質問
 - 地球の円周の長さは40008kmだが端数の8kmはどこから来たか?
- 答え
 - 200年前と現代でメートルの決め方が異なるため

時代	メートルの定義	誤差
1795年	地球の1/4 周の1/10000000	0.5~0.1 mm
1869年	最初のメートル原器（白金製）の長さ	0.05~0.01 mm
1889年	白金イリジウム合金製原器の長さ	0.2~0.1 μm
1960年	クリプトン86原子の発光スペクトル波長の1650763.73倍	0.01~0.005 μm
1983年	真空中で光が1/299792458秒に進む距離	0.1 nm

地球の大きさを測る



地球の円周の 長さを求めるには



Dunkirk to Paris
Module: 62472.59
Degree: 2.18910
Midpoint: 49°56'30"

- 一周は360度だから, 距離を緯度の差で割って360度を掛ければよい
- ダンケルク〜パリ間の測地データ
 - 距離は243.4km
 - 62472.59モジュール
 - 1モジュールは3.895344メートル（当時の長さの単位）
 - 角度は2.18910度
- 地球の円周の長さ
 - $(243.4/2.18910) \times 360 = 40027\text{km}$

もっと正確な 地球の円周の近似式

- 地球は楕円体なので, 緯度によって距離・角度・円周の関係が異なる
- そのため, 以下のような一次式を使った方が正確
- $Y = \alpha + \beta X$
 - $Y = \text{距離/角度}$
 - $X = \sin^2(\text{緯度})$
 - α, β が未知数
 - 地球の円周の長さ $= 360(\alpha + \beta/2)$

測地データから 得られた連立方程式

- $Y = \alpha + \beta X$
 - Y =距離/角度
 - $X = \sin^2(\text{緯度})$
 - α, β が未知数
 - 地球の円周の長さ=360($\alpha + \beta/2$)
- 未知数二つに対して方程式が四つあると解けない

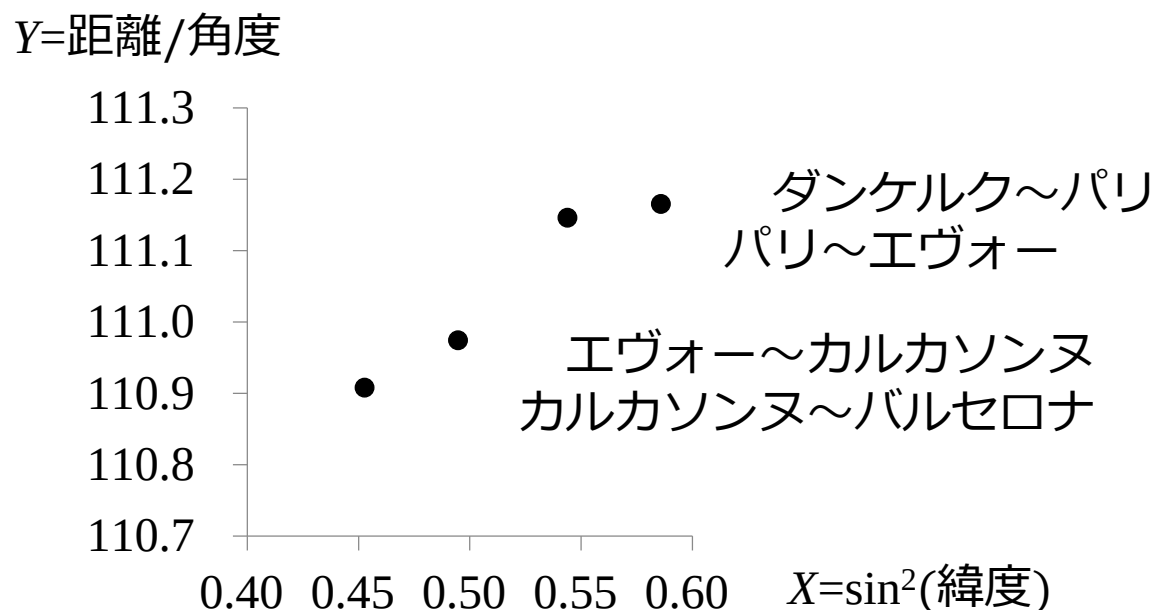
	方程式
ダンケルク～パリ	$111.17 = \alpha + \beta \times 0.586$
パリ～エヴォー	$111.15 = \alpha + \beta \times 0.544$
エヴォー～カルカソンヌ	$110.97 = \alpha + \beta \times 0.495$
カルカソンヌ～バルセロナ	$110.91 = \alpha + \beta \times 0.453$

数学者Gaussはどうか

- 縦軸 Y =距離/角度, 横軸 $X=\sin^2(\text{緯度})$ の散布図を描いた
- 四つの点にいちばん近くなるような直線を引き, その切片を α , 傾きを β の解とした

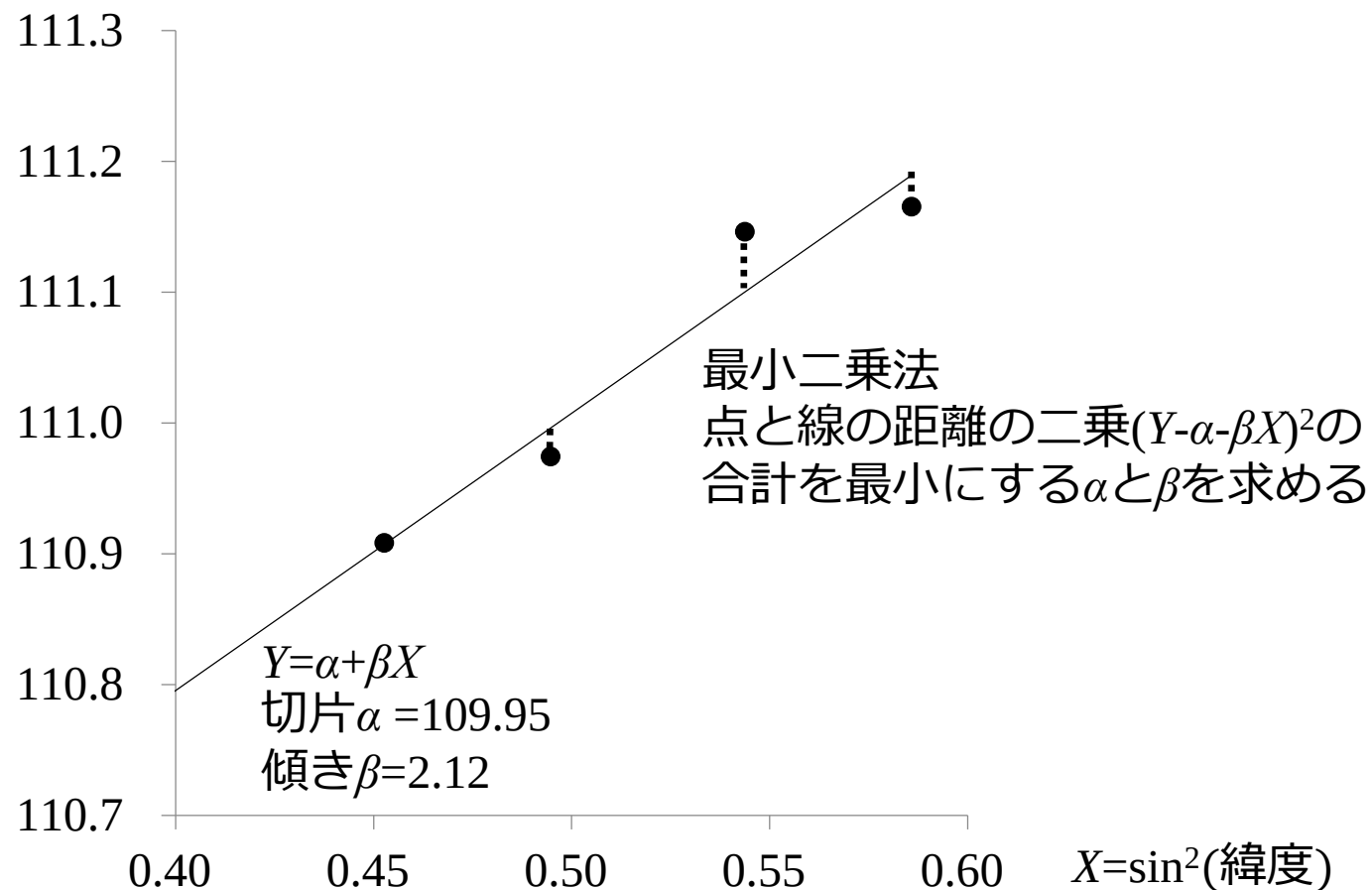


数学者Gauss (1777~1855年)



数学者Gaussはどうか

Y=距離/角度

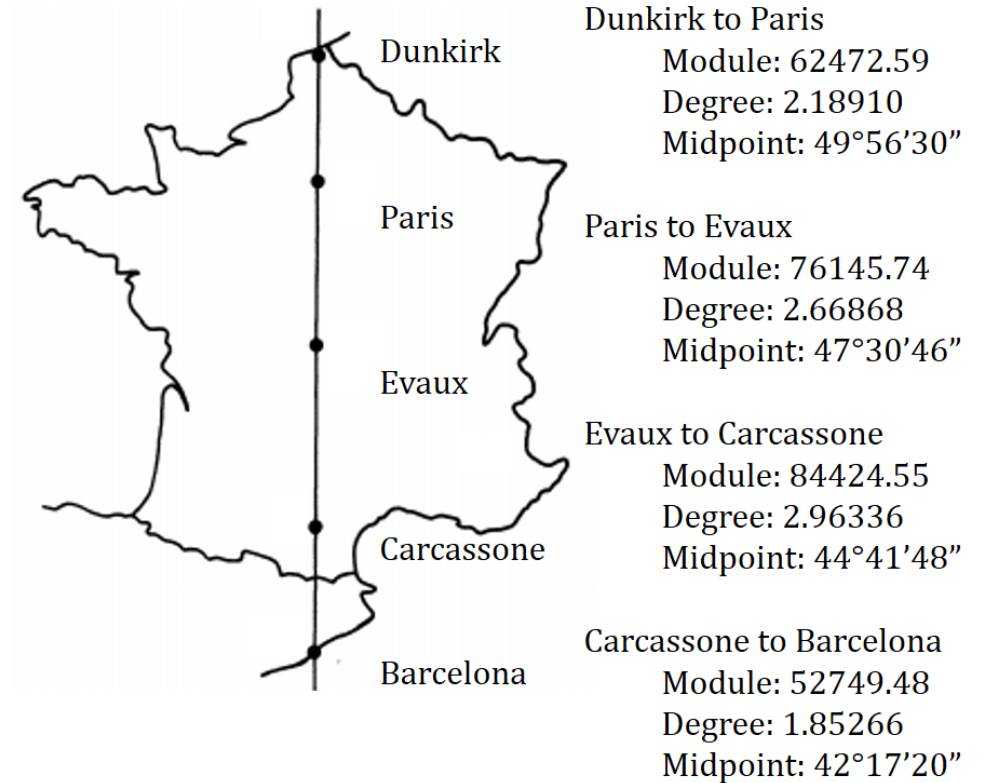


最小二乗法による 地球の円周の推定

- 地球の円周の長さ $=360(\alpha+\beta/2)$
- 先ほどのグラフでは
 - 切片 $\alpha=109.95$
 - 傾き $\beta=2.12$
 - 地球の円周の長さ $=360(109.95+2.12/2)$
 $=40519$
- この計算結果は, 途中で数字を丸めているので精度がよくない

最小二乗法による地球の円周の推定

- 地球の円周の長さ= $360(\alpha+\beta/2)$
- 元データ（モジュール単位）で計算
 - 切片 $\alpha=28224$
 - 傾き $\beta=546.7$
- ここから地球の円周の長さを計算すると39963kmとなり, かなりの精度で40000kmに近くなる



相関と回帰分析

- 相関
 - 二つの確率変数 X と Y が独立でないこと
- 回帰分析
 - 相関関係を調べるための方法
 - 結果変数 Y と説明変数 X が以下の数式に従うと仮定
 - $Y = \alpha + \beta X + \varepsilon$
 - ε は平均0の誤差項
- 最小二乗法
 - Y と $\alpha + \beta X$ の差の二乗を最小にするように α と β を計算