

相関図と相関係数

p.17~26

身長とリーチの関係

身長	リーチ	身長	リーチ
173	174	174	171
171	167	163	161
162	159	157	152
161	161	168	169
176	175	168	166
172	173	160	161
173	170	165	161
174	176	154	156

身長とリーチ

- 身長とリーチはほぼY=Xの直線上に分布
 - 身長≒リーチ
- ボクサーのリーチ
 - ジョン・ルイス(ヘビー級チャンピオン) 身長 188・リーチ 203
 - オスカ・ラリオス(スーパーバンタム級) 身長 169・リーチ 180
- 同級の仲里繁 身長 163・リーチ 170
- 辰吉丈一郎 身長 164・リーチ 175

変数変換による直線化

国民1人当たりのGDP(100ドル)と平均寿命 (2000年)

国名	GDP	logGDP	平均寿命	国名	GDP	logGDP	平均寿命
ネパール	2.4	0.38	59	スリランカ	8.4	0.92	73
カンボジア	2.4	0.38	54	中国	8.5	0.93	70
ラオス	3.3	0.52	54	フィリピン	9.8	0.99	69
ベトナム	3.4	0.53	69	タイ	19.6	1.29	69
ベトナム	4.0	0.60	69	マレーシア	38.5	1.59	73
モンゴル	4.1	0.61	67	韓国	96.7	1.99	73
パキスタン	4.5	0.65	63	シンガポール	229.6	2.36	78
インド	4.6	0.66	63	日本	381.6	2.58	81
インドネシア	7.3	0.86	66				

log10=1, log100=2,
log10^m=m

GDPが大きいと平均寿命が長い 全体として対数関数的な傾向

宇宙飛行船チャレンジャー号爆発

1986年1月28日の早朝
原因は固体燃料室を高熱ガスから密封する「O-ring」(直径3フィートの円板)の腐食

相関図→直線化→相関係数

- 対のデータの関連の度合を示す
- $(x_1, y_1), \dots, (x_n, y_n)$ の相関係数 r は

$$r = \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y})}{\sqrt{(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2} \cdot \sqrt{(y_1 - \bar{y})^2 + \dots + (y_n - \bar{y})^2}}$$

- 相関係数は測定単位に依存しない
- 相関係数は直線関係の指標
非線形なら変数変換により直線化

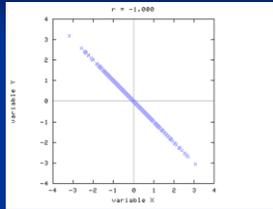
相関係数の計算式

$$r = \frac{x \text{ と } y \text{ の 共分散}}{x \text{ の 標準偏差} \times y \text{ の 標準偏差}}$$

$$= \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}}$$

$$= \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y})}{\sqrt{(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2} \cdot \sqrt{(y_1 - \bar{y})^2 + \dots + (y_n - \bar{y})^2}}$$

相関係数のアニメーション



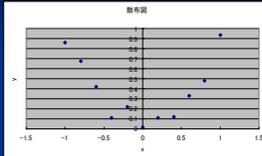
Excel関数

- 相関係数 CORREL(範囲1, 範囲2)
- 共分散 COVAR(範囲1, 範囲2)
- 対数(底10) LOG10(値)
- 平方根 SQRT(値)

問4

$$\begin{aligned}
 & (x_1 - \bar{x})(y_1 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y}) \\
 &= \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n (x_i y_i - x_i \bar{y} - \bar{x} y_i + \bar{x} \bar{y}) \\
 &= \sum_{i=1}^n x_i y_i - \bar{y} \sum_{i=1}^n x_i - \bar{x} \sum_{i=1}^n y_i + n \bar{x} \bar{y} \\
 &= \sum_{i=1}^n x_i y_i - \bar{y} n \bar{x} - \bar{x} n \bar{y} + n \bar{x} \bar{y} = \sum_{i=1}^n x_i y_i - n \bar{x} \bar{y} \\
 &= x_1 y_1 + \dots + x_n y_n - n \bar{x} \bar{y}
 \end{aligned}$$

相関係数の適用が不適切な例



- 相関係数 = -0.03
- 実際には2次の関数関係