

回帰分析・Excelで統計

樋口さぶろお

龍谷大学理工学部数理情報学科

確率統計☆演習 I L05(2014-10-25 Fri)

今日の目標

- 回帰係数, 回帰直線を, データから手で求められる.
- データが与えられたとき, Excel を使って, 平均値, ..., 回帰直線を求められる. この科目や他の科目のレポート作成に利用できる.



<http://hig3.net>

L04-S2 Quiz 解答:共分散 $\bar{x} = 4, s_x^2 = 4, s_x = 2.$

$\bar{y} = 13, s_x^2 = 122/5 = 24.4, s_y = \sqrt{122/5} = 4.94.$

共分散 $C = \frac{1}{5}[(1-4)(5-13) + (3-4)(15-13) + (4-4)(14-13) + (5-4)(11-13) + (7-4)(20-13)] = 41/5 = 8.2.$

相関係数 $r = \frac{41/5}{2 \cdot \sqrt{122/5}} = 0.83.$

ここまで来たよ

1 略解:2 変量データと相関係数

2 回帰分析

- 回帰分析

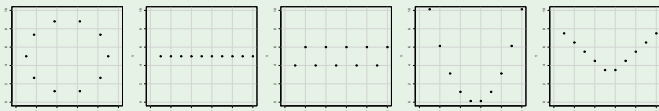
3 Excel で統計

- Excel で統計

相関係数=0 にだまされるな:にせの無関係

Quiz(相関係数)

次のうち、相関係数 r がもっとも大きいものはどれ?



相関係数 $r = 0 \Leftrightarrow x$ と y の間に '関係' がない?

- 相関係数 $r = 0 \Leftrightarrow x$ が増えた

ら

言えない

- 相関係数 $r = 0$ だから x, y は無関係な量, というわけではない

相関係数 $\neq 0$ にだまされるな:にせの因果関係

被シュートと失点は正の相関

- 原因:被シュートが多い, 結果: 失点が多い?
- 原因:失点が多い, 結果: 被シュートが多い?

フリーキックと被シュートは負の相関

- 原因:フリーキックが多い, 結果:被シュートが少ない?
- 原因:被シュートが少ない, 結果:フリーキックが多い?
- 原因:???, 結果:被シュートが少ない, かつ, フリーキックが多い?

- 相関が強くても

- 因果関係があっても

回帰分析

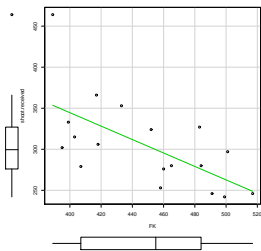
回帰 (regression), 単回帰分析=1 変数回帰分析

2 変量データ (x, y) が

相関係数 $r = \pm 1$ に近い \Leftrightarrow 散布図上のデータ点 (x, y) がほぼ直線に乗っている

その直線 () の式 $y = ax + b$ を知りたい!

つまり a, b を決めたい.



y : **目的変数** (従属変数)

x : **説明変数** (独立変数)

何でそんなことしたいの?

- 法則を見つけない
- x から y を予測したい

回帰直線の決め方

- ① 定規をあてて '真ん中' を通るように
- ② 最小 2 乗法で.

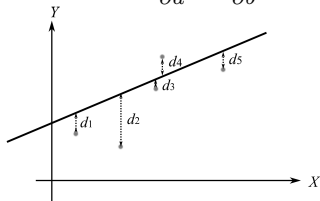
最小 2 乗法

直線からのずれの 2 乗 d^2 の合計

$$f(a, b) = \sum_{i=1}^n d_i^2 = \sum_{i=1}^n (y_i - (ax_i + b))^2$$

の最小条件 $\frac{\partial f}{\partial a} = \frac{\partial f}{\partial b} = 0$ で a, b を決める.

微積分・演習 I



物理実験

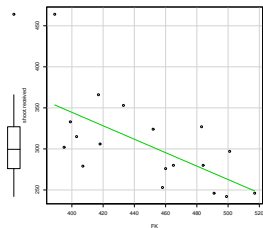
最小 2 乗法の公式

回帰直線

x_i, y_i ($i = 1, \dots, n$) の平均値を \bar{x}, \bar{y} , 標準偏差を s_x, s_y , 相関係数を r とする. このとき回帰直線は,

$$y = \frac{r \times s_y}{s_x} \times (x - \bar{x}) + \bar{y}.$$

傾き $a = \frac{r \times s_y}{s_x}$, 切片 $b =$ 点 (\bar{x}, \bar{y}) を通るような b



回帰直線の傾きのおぼえ方

(x, y) が (m, kg) だとする.

傾きは r みたいなもの.

r は無次元の数 () だが

傾き a の単位は .

調整するためには r に $\frac{s_y}{s_x}$ (単位) をかけて傾き a にする.

なんで s_x, s_y とかつくの?

なかったら x, y いれかえても同じ傾きになっちゃうじゃん.

L05-Q1 来週の非参照 Quiz はこんな感じ

Quiz(共分散と相関係数)

下のデータを考える.

x	y
2	4
2	6
4	11
5	9
7	15

- ① 共分散を求めよう.
- ② 相関係数を求めよう.
- ③ 回帰直線の式を求めよう.

ただし, 平均値 $\bar{x} = 4, \bar{y} = 9$, 標準偏差 $s_x = 1.90, s_y = 3.85$ (四捨五入してます) であることを使っていい.

ここまで来たよ

1 略解:2 変量データと相関係数

2 回帰分析

- 回帰分析

3 Excel で統計

- Excel で統計

準備

統計ソフトウェア実習室にインストールされているのは

- R 無料. オープンソース. 解説書が多い.
- SPSS 伝統ある高級品.
- Excel 機能は限られ怪しいところもあるが, 自宅にもインストールされてるでしょ.

今日は Excel を使ってみます.

スタートボタン > Excel 2010

ファイル > オプション > アドイン > Excel のアドイン > 設定 > 分析ツール に
チェックを入れて OK する.

L05-Q2

今日の授業内課題

<http://hig3.net> → RaMMoodle で, データを取得して Excel で指定の
分析 ×2 をして, xlsx ファイルを RaMoodle に提出しよう.

Excel による主な分析

	メニューベース	関数ベース
代表値	データ > 分析 > データ分析 > 基本統計量	平均値 average, 分散 varp, 標準偏 差 stdevp, 最頻値 mode
四分位数	データ > 分析 > データ分析 > 順位と区分位数	中央値 median, 四 分位数 quartile
度数分布表, ヒ ストグラム	データ > 分析 > データ分析 > ヒストグラム	frequency + グ ラフ
散布図	挿入 > グラフ > 散布図	
共分散, 相関係 数	データ > 分析 > データ分析 > 共分散, 相関	covar, correl
回帰分析	データ > 分析 > データ分析 > 回帰分析	covar
クロス集計表	挿入 > テーブル > ピボット テーブル	

レポート課題

締切: 2014-11-07 金 23:55, 提出場所: RaMMoodle

Quiz(多変量分析)

平成 26 年度学校基本調査 http://www.mext.go.jp/b_menu/toukei/chousa01/kihon/1267995.htm の

統計表一覧 <http://www.e-stat.go.jp/SG1/estat/NewList.do?tid=000001011528> から、何か 2 変量データ (x, y) を取り出して、相関係数を求めて回帰分析をした結果の Excel ワークシートを提出しよう。

- 回帰分析が、統計的、教育的、社会的に意味があり、おもしろそうな (x, y) であること。
- シート内に、 x, y がそれぞれ何か、どの表から取得したか (表番号, 表の名前, URL) を明示すること。
- 分析の結果、 x, y の間はどのような関係か (正負の強い弱い相関, x が 1 増えると、 y は…), 短い日本語の文でシート内に書こう。

連絡

- 予習問題ポリシー変更: 点数:最終受験→最高点, 締切:水 9:20 →金 9:00, 正誤表示:締切後→受験後.
- 2014-10-31 金 授業なし. 予習問題の締切は 2014-11-07 金 09:20 です.
- 2014-11-21 金 2 プチテスト. 非参照. 30 ピーナッツ.
- 2014-10-01 水昼 からチューターやっています. 1号館 6階 1-614.
- 2014-10-28 火 4, 29 水 14:00-17:00 数理情報学科特別講義.
- 2014-11-06 木 数学検定団体受検申込締切.
<http://www.math.ryukoku.ac.jp/suken/> で受付中.
- 2014-12-06 土 34 数学検定団体受検.

- Excel 入門コース <https://moodle.media.ryukoku.ac.jp/>
- 情報リテラシー講座 <http://www.seta.media.ryukoku.ac.jp/literacy/literacy.html>
- リメディアル統計学 <https://moodle.media.ryukoku.ac.jp/>