

Python勉強会@HACHINONE

## 第16章

うそ、真っ赤なうそ、

そして統計

# お知らせ

Python勉強会@HACHINOHEでは、ジョン・V・グッターグ『Python言語によるプログラミングイントロダクション』近代科学社、2014年をみんなで勉強しています。

この本は自分で読んで考えて調べると力が付くように書かれています。

自分で読んで考えて調べる前に、このスライドを見るのは、いわば**ネタバレ**を聞かされるようなものでもったいないです。

是非、本を読んでからご覧ください。

# 統計の諸課題

## Python勉強会@HACHINOHE

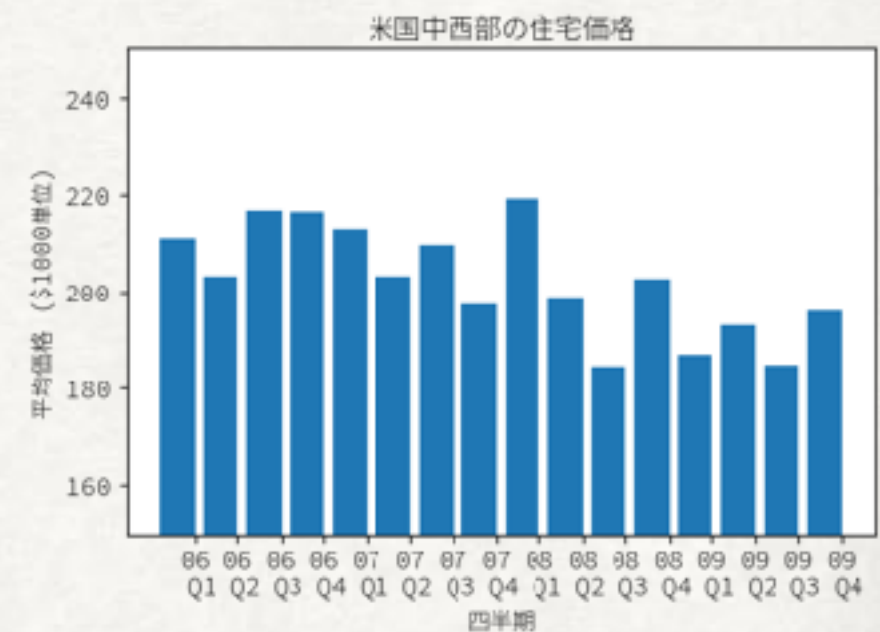
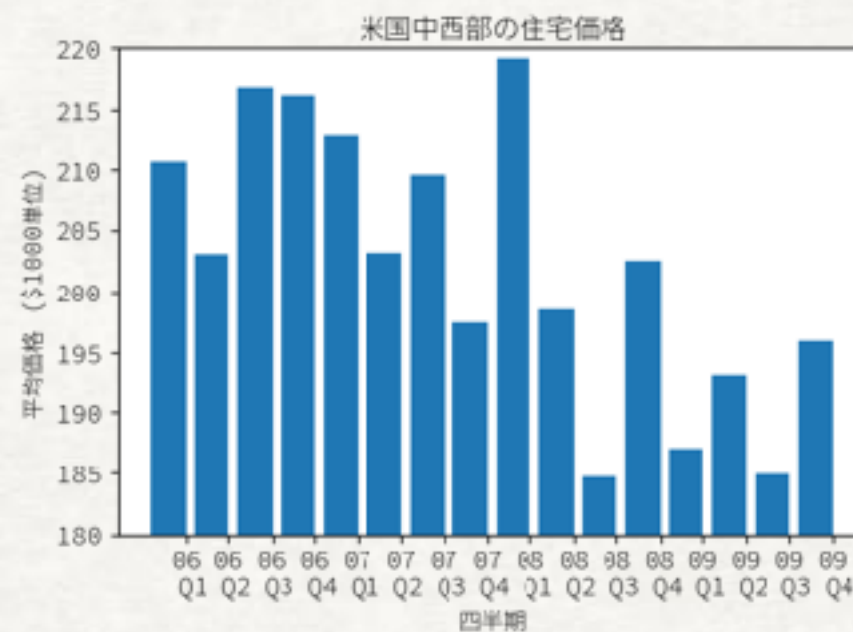
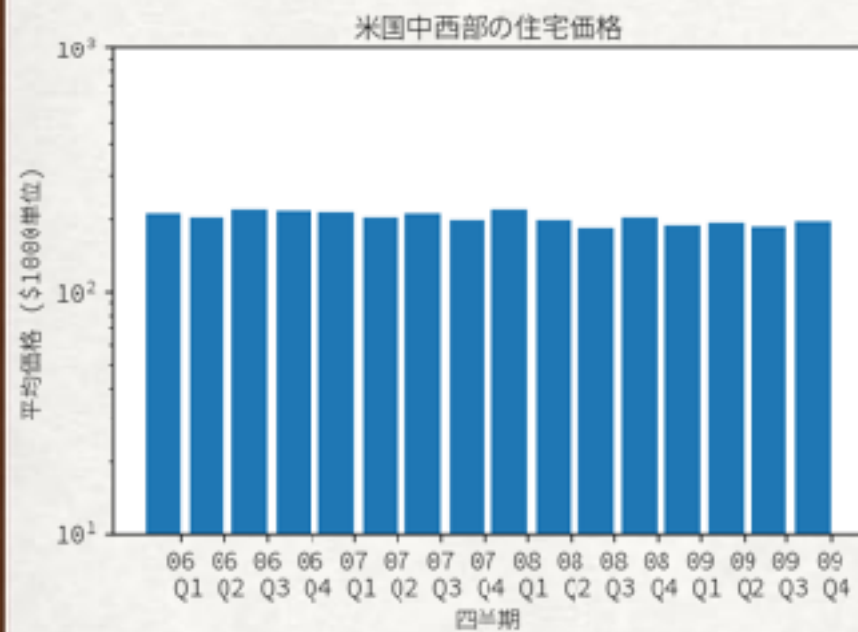
- 元データがダメだと、結論もダメ: 奴隷制度の話
- グラフの縦軸の目盛り: リーマンショックの話
- 相関と因果の錯誤: 通学とインフルエンザの流行、輸入レモンと高速道路の傷み、ホルモン入れ替え療法と心臓血管疾患
- まとめとしての統計量の限界: 同じ統計量で分布が異なるデータ
- 生存バイアス: 被弾と撃墜、同性愛と寿命
- 文脈の理解: 豚インフルエンザ、銃犯罪
- 外挿の限界: インターネット利用率の予測
- テキサスの名射撃手の誤り: 誕生月と拒食症
- 割合の母数: 株価の上下変動の説明



# グラフの目盛り

Python勉強会@HACHINOHE

- 住宅価格は下落したか？
- 異なる縦軸の目盛り



# 因果の錯誤

Python勉強会@HACHINOHE

- インフルエンザの流行
  - 学校に通う時期にインフルエンザが流行する
  - それぞれ寒い時期
- 輸入レモンと高速道路の傷み
  - メキシコから輸入レモンが増えると、高速道路の痛みが減る
  - 最近、レモンの輸入が増え、高速道路の痛みも減った
- ホルモン入れ替え療法と心臓血管疾患
  - ホルモン入れ替え療法を受けている女性は心臓血管疾患リスクが低い
  - ホルモン入れ替え療法を受けている女性は健康に気を使っていた

# 岩崎学@IoT・データ利活用ビジネスフォーラム2018

Python勉強会@HACHINOHE

- 横浜市立大学データサイエンス学部学部長
- ある大学のデータでは、女子学生の割合が高い学科の方が、TOEICの平均点が高いように見えるが？
- その他に多重共線性の話など
  - 強い相関がある3つのデータの重回帰分析の話

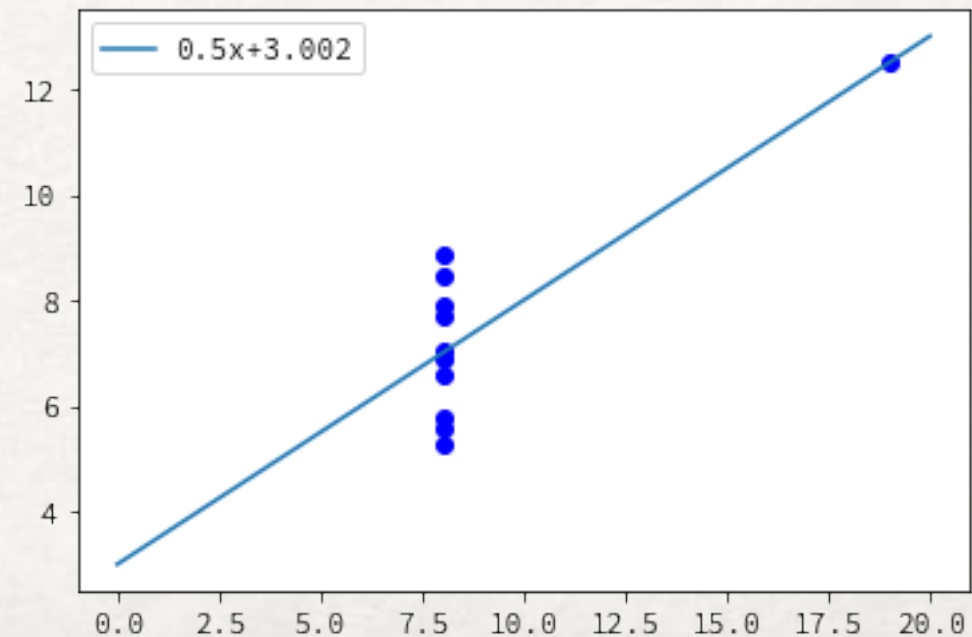
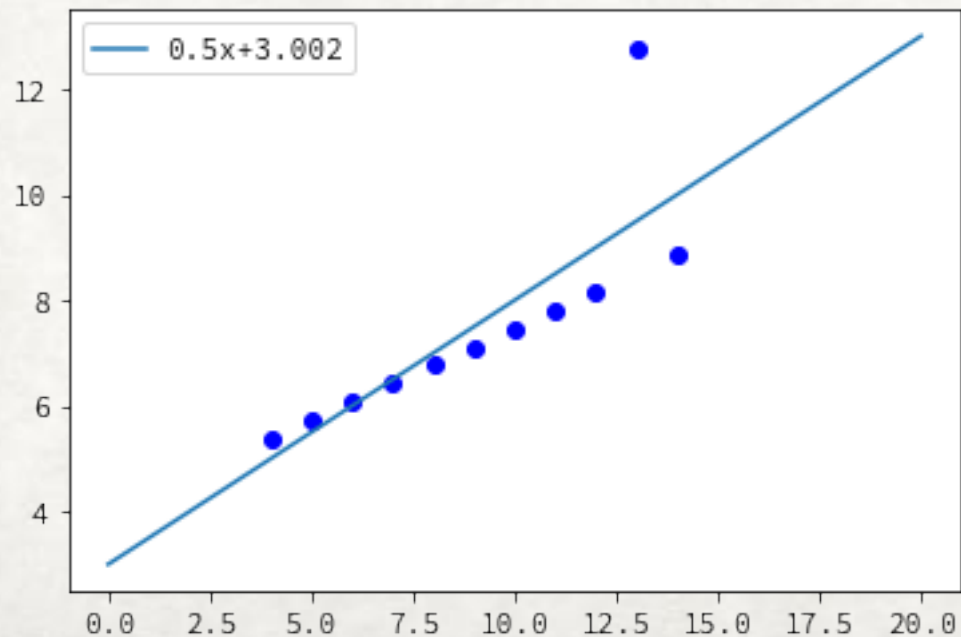
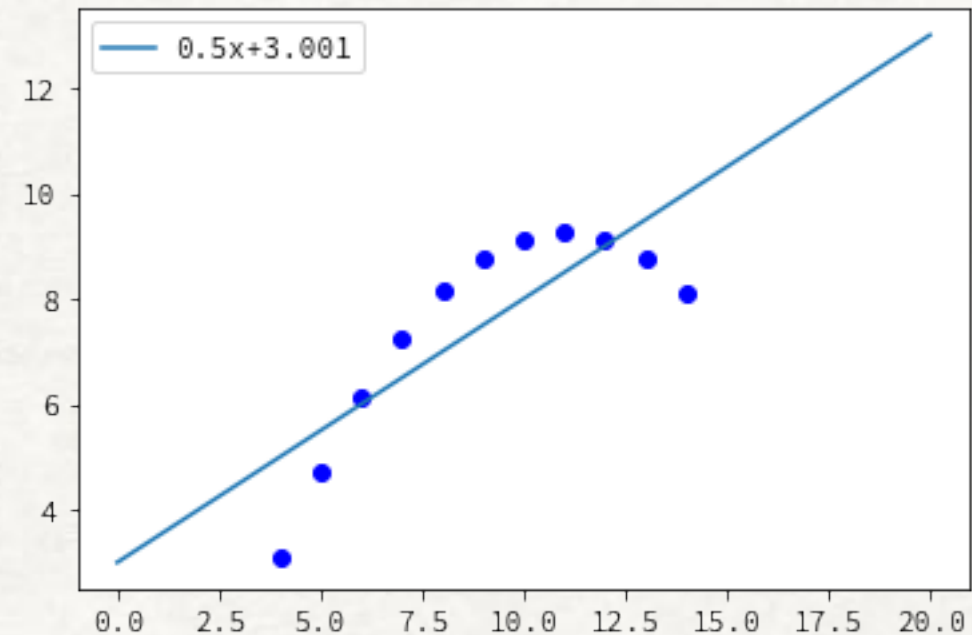
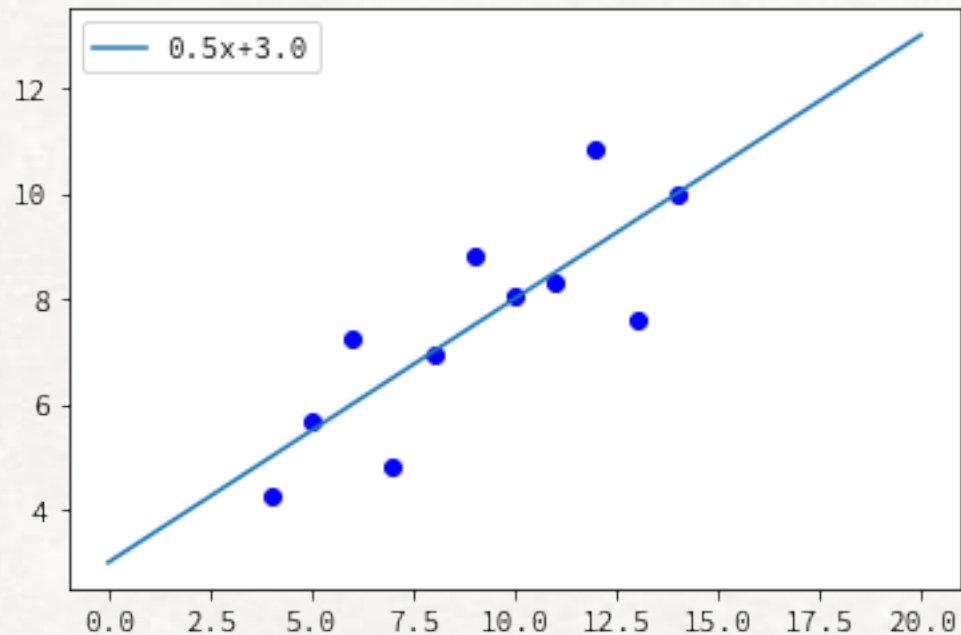


# アンスコム の例

Anscombe, F. J. (1973). "Graphs in Statistical Analysis". *American Statistician* **27** (1): 17–21

## Python勉強会@HACHINOHE

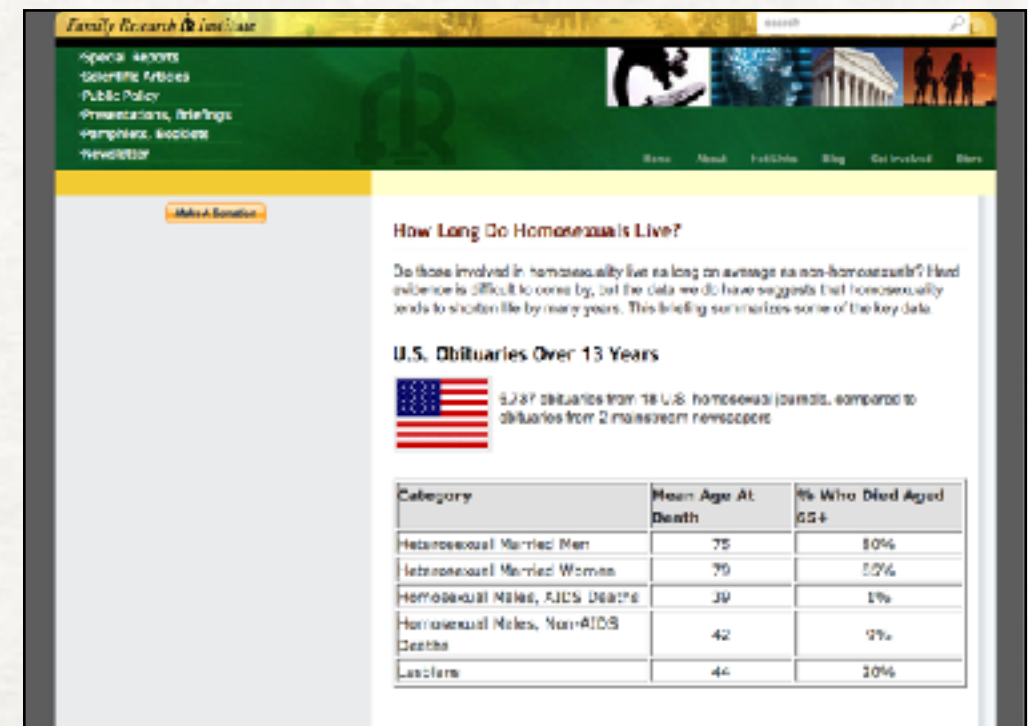
- 平均、分散、相関係数、回帰直線が同じだが...
- 統計量はある意味「まとめ」



# 生存バイアス

Python勉強会@HACHINOHE

- 帰還した航空機の被弾箇所を補強
  - 撃墜された航空機の被弾箇所と異なる
  - ウィリアム・リドウェル他 『要点で学ぶ、デザインの法則150』  
BNN新社、2015年
- 同性愛者の平均死亡年齢
  - 同性愛雑誌に掲載された死亡記事  
と2つの大新聞の死亡記事がソース



<http://www.familyresearchinst.org/2012/01/how-long-do-homosexuals-live/>



# 文脈の理解、外挿

Python勉強会@HACHINOHE

- 文脈
  - 豚インフルエンザで159人が死亡
    - 季節性インフルエンザでは36000人が死亡
  - ほとんどの自動車事故は自宅から10マイル以内
    - 多くの移動は自宅から10マイル以内
  - 銃の99.8%は暴力犯罪に用いられていない
    - 銃は3億丁→60万丁が用いられている
- 外挿
  - 1994年から2000年のインターネット利用率を外挿すると2010年頃に100%を突破

# テキサスの名射撃手の誤り

Python勉強会@HACHINOHE

- 撃ったところに後からの的を描く
- 6月生まれ的女性に拒食症が多い
  - 446人の拒食症の女性のうち、偶然、48人以上が6月生まれになる確率4.5%
  - しかし、偶然、48人以上が生まれた月が存在する確率45%
- 40人学級で、自分と同じ誕生日の人がいる確率は小さいが、誕生日が同じ組み合わせの2人が存在する確率は大い
- 占星術の研究でも
  - 「火星効果」火星の影響を受けると運動能力が高い？

# 割合の母数

Python勉強会@HACHINOHE

- 月平均0.5%株価が上がった → 去年よりも8%下がった??
- 前半は毎月15%減、後半は毎月16%増

月	1	2	3	4	5	6	7	8	9	10	11	12
価値	100	85	72	61	52	44	44	51	60	69	80	93

- 元が大きい値の割合の方が、元が小さい値の割合より大きい
- ある疾患にかかる確率が200%になる
- その疾患にかかる確率が100万人に1人だったら?