

統計解析環境「R」

中西渉

watayan@meigaku.ac.jp
名古屋高等学校

2011年2月20日

agenda

- 1 Rとは
- 2 Why R ?
- 3 デモ
- 4 利用場面
- 5 情報入手
- 6 短所
- 7 まとめ

Rとは

- 統計解析に適したプログラミング言語・実行環境
- GPL に基づき配布
- マルチプラットフォーム
- S 言語から派生

他の選択肢

- Excel

- SPSS

他の選択肢

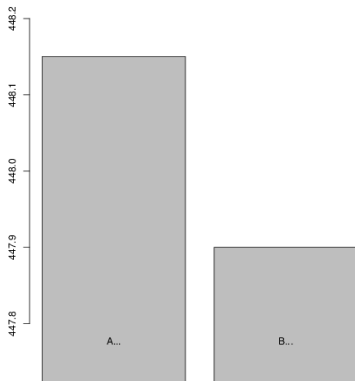
- Excel
 - 統計関数・ヘルプに問題
 - グラフに疑問
 - どうやって箱ひげ図やヒストグラムを?
 - ひどい 3D グラフ（特に円グラフ）
- SPSS

他の選択肢

- Excel
 - 統計関数・ヘルプに問題
 - グラフに疑問
 - どうやって箱ひげ図やヒストグラムを?
 - ひどい 3D グラフ（特に円グラフ）
- SPSS
 - 高価

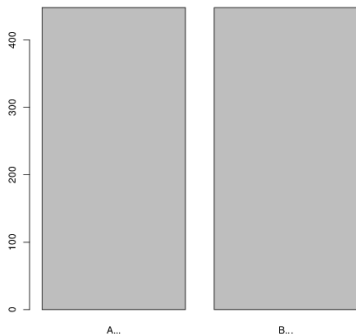
午前中の電池稼働時間の平均

A 社ならこのグラフ?



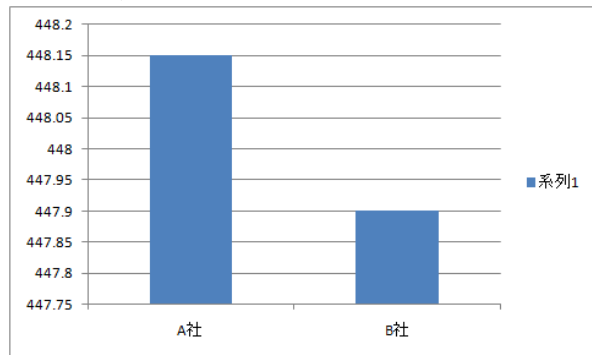
午前中の電池稼働時間の平均

真っ当なグラフ



午前中の電池稼働時間の平均

Excel のデフォルトだと...



R の長所

- 無料
- マルチプラットフォーム (Windows, Mac OS X, Linux)
- グラフ (箱ひげ図, ヒストグラムなど)
- プログラミング言語

デモ

実際のデータで...

数学「データの分析」

Tukey のアレやコレが簡単に...

統計関数

sum	和
length	データの個数
mean	平均
median	中央値
max,min	最大, 最小
var,sd	分散, 標準偏差
table	度数分布
summary	基本統計量
fivenum	5 値要約
cov,cor	共分散, 相関係数
stem	幹葉図

グラフ

barplot 棒グラフ

pie 円グラフ

plot,matplot 散布図

boxplot 箱ひげ図

hist ヒストグラム

シミュレーション

- ベクトル処理 → ループが省ける
- 1 行ずつ試行錯誤

例

あのシミュレーションを...

DNCL で書いてみると

整数 x, y, a, n, i

$n \leftarrow 10000$

$a \leftarrow 0$

i を 1 から n まで 1 ずつ増やしながら,

| $x \leftarrow \text{random}(100)$

| $y \leftarrow \text{random}(100)$

| もし $x * x + y * y < 10000$ ならば

| | $a \leftarrow a + 1$

| を実行する

を繰り返す

$4.0 * a / n$ を印刷する

例

あのシミュレーションを...

R で書いてみると

```
x <- runif(10000)
y <- runif(10000)
mean(x^2 + y^2 < 1) * 4
```

業務

- 我々の仕事に統計は必要
- 使わないものを生徒に教えるか？

Web

総本山 <http://www.r-project.org>

日本語のサイト 参考文献欄を参考に...

書籍

- 最近増えた
- 好みのものを...

勉強会

Nagoya.R

- <http://corpus-study.info/nagoyar/>
- 前半は初心者講習会



短所

- CUI
- 日本語処理
- 資料
高校生向けテキストがほしい
- 質問できる人
- 分散・標準偏差
分母は n でなく $n - 1$

まとめ

- 統計は数値のリテラシ

まとめ

- 統計は数値のリテラシ
- 紙の上の話で終わるのか？

まとめ

- 統計は数値のリテラシ
- 紙の上の話で終わるのか？
- 知ってることを教えるのか
知るべきことを教えるのか