

「データ」のいろいろ

学芸大学 気象 森

1. 地図

中心極限定理

平均
標準偏差・分散

「誤差」の計算

推定・検定

2. データの現実

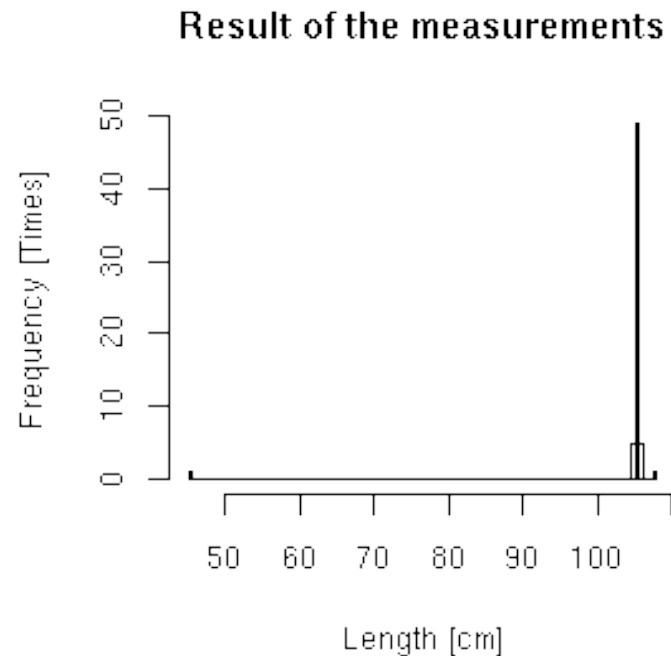
- 長さの計測結果
 - ヒストグラム(頻度分布)
 - データの吟味
 - 平均・分散の計算

105.37, 105.47, 105.14, 105.32, 105.17, 105.33, 104.82, 105.19, 105.79, 105.20
105.73, 105.30, 105.42, 105.12, 105.35, 105.43, 105.24, 105.11, 105.22, 105.36
105.41, 105.48, 107.53, 105.49, 105.18, 105.28, 105.87, 105.27, 105.30, 104.82
105.33, 45.44, 105.26, 105.21, 105.88, 105.15, 105.21, 105.44, 105.13, 105.42
105.04, 105.26, 105.43, 105.36, 105.25, 105.39, 105.71, 105.14, 105.46, 104.77
105.26, 104.96, 105.20, 105.35, 105.37, 105.32, 105.12, 105.05, 105.29, 105.12
104.90

後で課題を出します！

2. データの現実

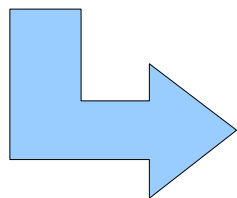
- 長さの計測結果
 - ヒストグラム(頻度分布)
 - データの吟味
 - ふたつのデータの除外



105.37, 105.47, 105.14, 105.32, 105.17, 105.33, 104.82, 105.19, 105.79, 105.20
105.73, 105.30, 105.42, 105.12, 105.35, 105.43, 105.24, 105.11, 105.22, 105.36
105.41, 105.48, 107.53, 105.49, 105.18, 105.28, 105.87, 105.27, 105.30, 104.82
105.33, 45.44, 105.26, 105.21, 105.88, 105.15, 105.21, 105.44, 105.13, 105.42
105.04, 105.26, 105.43, 105.36, 105.25, 105.39, 105.71, 105.14, 105.46, 104.77
105.26, 104.96, 105.20, 105.35, 105.37, 105.32, 105.12, 105.05, 105.29, 105.12
104.90

2. データの現実

- 今回の場合
 - 物差しの精度の違い
 - 紙の湿度の違い
 - 温度の違い
 - 読み取りのクセの違い
 - その他のミス
 - ...

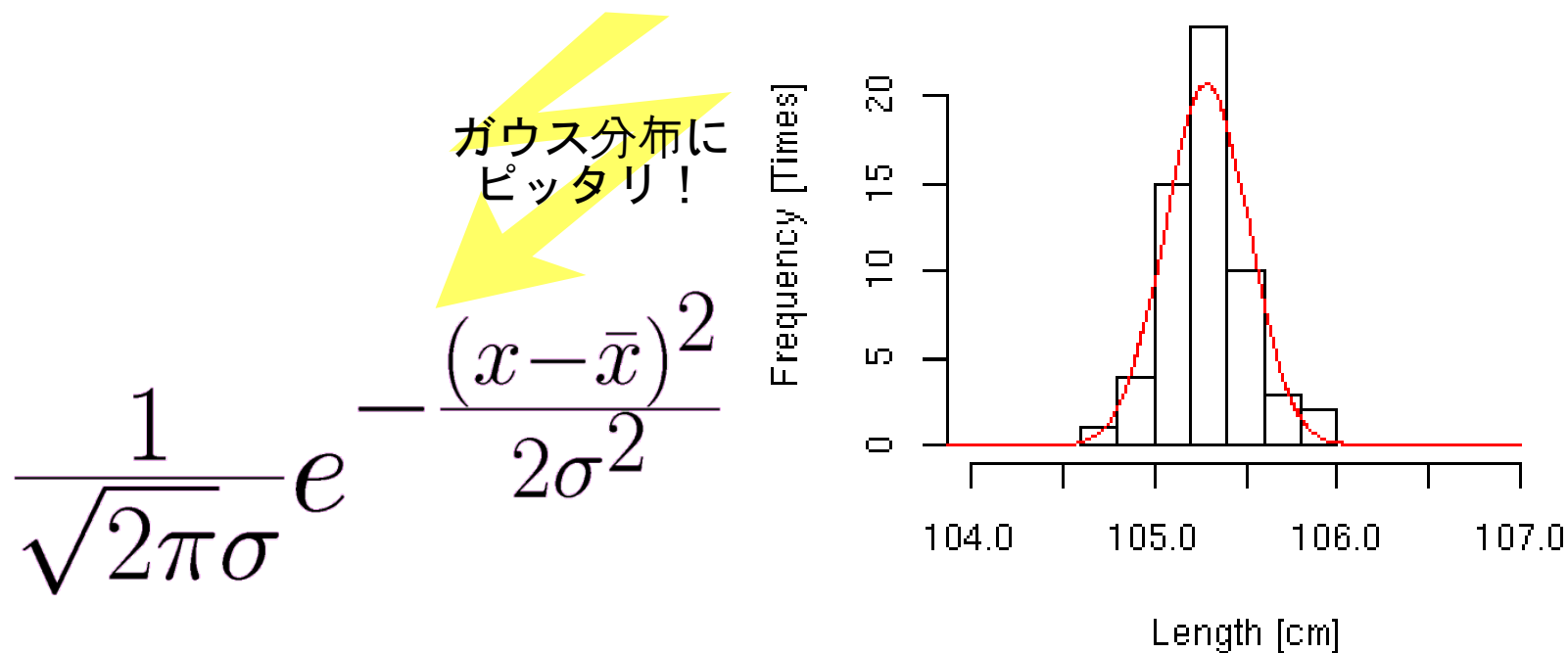


じゃあ実際にどうするの？

2. データの現実

- 長さの計測結果
 - ヒストグラム(頻度分布)
 - データの吟味

Result of the measurements (Revised)



2. データの現実

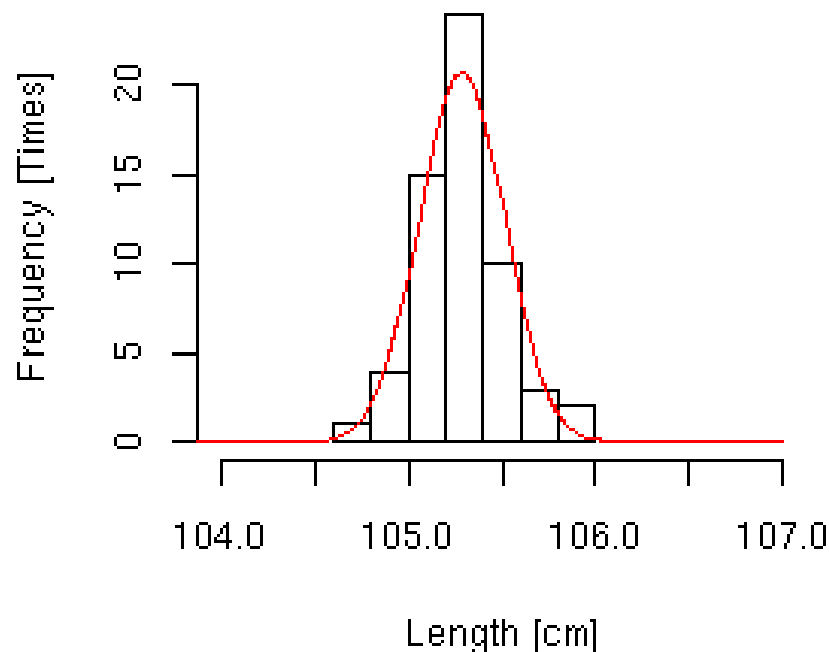
- 長さの計測結果
 - ヒストグラム(頻度分布)
 - データの吟味
 - 平均・分散の計算

cf. 中心極限定理

ガウス分布に
ピッタリ!

$$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}$$

Result of the measurements (Revised)



3. ここまでのまとめ

- 測っても正しい値はわからない。
 - 何度測ってもわからない。
 - 大きく間違えることもある。
 - でも、計測結果はガウス分布に近いみたい。
- じゃあどうする？！
 - 何回も測定してみる
 - 測定誤差を見積もる
 - 0.3mm(目視のずれ)が数回程度 → 1~2mm 程度
 - そんなものと割り切る ⇒ それを前提に
 - 正規分布の性質を使う 扱う方法を考える。
 - 例: データの吟味に使える

4. 実践編

- 準備

- 平均

- 分散

- 平均からのズレの平均

× : ゼロになる。

- 平均からのズレの2乗の平均

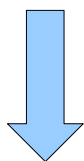
△ : ゼロにはならないが、2乗の値。

- 標準偏差

- 分散の平方根 : 平均からのズレの指標になる

4. 実践編

- 正規分布の予備知識
 - 平均 ± 標準偏差
 - 全データの68%が入っている。
 - 平均 ± 2×標準偏差
 - 全データの95%が入っている。



平均から標準偏差の2倍も離れている
データは異常である。

105.37, 105.47, 105.14, 105.32, 105.17, 105.33, 104.82, 105.19, 105.79, 105.20
105.73, 105.30, 105.42, 105.12, 105.35, 105.43, 105.24, 105.11, 105.22, 105.36
105.41, 105.48, 107.53, 105.49, 105.18, 105.28, 105.87, 105.27, 105.30, 104.82
105.33, 45.44, 105.26, 105.21, 105.88, 105.15, 105.21, 105.44, 105.13, 105.42
105.04, 105.26, 105.43, 105.36, 105.25, 105.39, 105.71, 105.14, 105.46, 104.77
105.26, 104.96, 105.20, 105.35, 105.37, 105.32, 105.12, 105.05, 105.29, 105.12
104.90

4. 実践編

- 誤差を含んだデータの計算

- $a = A \pm \Delta A, b = B \pm \Delta B$
- $x = a \times b$ の値と誤差は？

- 地学実験テキストによると

- $(A - \Delta A) \times (B - \Delta B) < x < (A + \Delta A) \times (B + \Delta B)$
- $|\Delta x/x| = |\Delta A/A| + |\Delta B/B|$

- 物理学実験テキストによると

- $\Delta x = \sqrt{(\Delta A/B)^2 + (\Delta B/A)^2}$
- $|\Delta x/x| = |\Delta A/A| + |\Delta B/B|$