

データの取り扱い MS Excelを利用した統計処理 (分散・平均の検定)

2019 年 10 月 10 日
合成生物学研究分野
W5-728室 濱田 浩幸

九州大学農学部生物資源環境学科応用生物科学コース応用生命化学分野

教材の配布

1. 以下のURLにアクセスする。

<http://www.brs.kyushu-u.ac.jp/~hamada/B2ET.htm>

2. 圧縮ファイル (**Stat-Word.zip**) を取得する。

C. MS Excelを利用した統計処理 (分散・平均の検定)

手 続 き	内 容
ダウンロード	圧縮ファイル (Stat-Word.zip) の取得
保 存 先	デスクトップ
内 容 物	MS Word file および PDF

ここをクリック

3. 取得したファイルを解凍する。

教材の確認

解凍後、以下2点のファイルがあることを確認する。

- 1) C. MS Excelを利用した統計処理（分散・平均の検定）
提出ファイル（配布資料と同じもの）
- 2) 分布表

3

講義および演習の内容

0. はじめに
1. 相加平均・相乗平均・調和平均・対数平均
2. 正規分布・対数正規分布
3. 分散・標準偏差
4. 中央値・最頻値
5. 検定とは

演習課題：カタログ値との比較

分散の検定・平均値の検定

演習課題：二群の比較

分散の検定・平均値の検定

4

はじめに

実験データを図表にまとめることは、**実験データの誤りの発見、傾向の理解に有用**である。また、実験データに**統計処理**を加えることにより、実験データの**本質を捉えた考察**が可能となる。

統計処理

平均値 (mean value) ・ 標準偏差 (SD)

回帰分析 (regression analysis)

平均値の差の検定 (test)

平均値の推定 (estimation)

などなど

5

1. 平均

実験データを表現する代表値の1つ

- 1) 算術平均 (相加平均) ・ 加重平均
- 2) 自乗平均
- 3) 幾何平均 (相乗平均)
- 4) 調和平均
- 5) 対数平均

調和平均 <= 相乗平均 <= 対数平均 <= 相加平均

6

1. 算術平均（相加平均）・加重平均

重み $n_1, n_2, \dots, n_{k-1}, n_k$
 測定値 $x_1, x_2, \dots, x_{k-1}, x_k$

算術平均 $\bar{x} = \frac{x_1 + x_2 + \dots + x_{k-1} + x_k}{k} = \sum_{i=1}^k x_i / k$

加重平均 $\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_{k-1} x_{k-1} + n_k x_k}{n_1 + n_2 + \dots + n_{k-1} + n_k}$

x の重みが等しい場合 ($n_1 = n_2 = \dots = n_{k-1} = n_k$) : **算術平均**

母集団が非常に大きい場合、母平均を得ることは困難。
 母集団から無作為に抽出した測定値（複数）を用いて、
 平均を求める。なお、**測定値の変化が小さいほど、
 その算術平均は母平均に等しくなる。**

7

1. 自乗平均・幾何平均（相乗平均）

k 個の正または負の測定値 $x_1, x_2, \dots, x_{k-1}, x_k$ の

自乗和を測定値の個数で除算した値 : **自乗平均**

$$\bar{x} = \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_{k-1}^2 + x_k^2}{k}}$$

x が平均値—測定値で表されるような場合などに用いられる（誤差の評価）

k 個の正の測定値 $x_1, x_2, \dots, x_{k-1}, x_k$ の積の k 乗根 : **幾何平均**

$$\bar{x} = (x_1 \cdot x_2 \cdots x_{k-1} \cdot x_k)^{1/k}$$

年ごとに測定値が幾何級数的に増加している場合などに用いられる
 （例：情報通信技術の分野）

8

1. 調和平均・対数平均

k 個の正の測定値 $x_1, x_2, \dots, x_{k-1}, x_k$ の逆数の

算術平均の逆数：**調和平均**

$$\bar{x} = \frac{k}{(1/x_1) + (1/x_2) + \dots + (1/x_{k-1}) + (1/x_k)}$$

測定値の逆数を取り扱う場合に用いられる

x_1 と x_2 の自然対数 (\log_n) の差に対する x_1 と x_2 の差：**対数平均**

$$\bar{x} = \frac{x_1 - x_2}{\ln(x_1) - \ln(x_2)} = \frac{x_1 - x_2}{\ln(x_1/x_2)}$$

近年、1細胞計測技術が進み、生命現象の多くにおいて、
測定値が対数正規分布に従うことが知られるようになった。

9

1. 平均 練習問題

1. 案外きつかった。

先日、合成生物学研究分野の総員は久住山登山に出かけた。
11:30に標高1333mの牧ノ戸峠を出発し、14:00に標高1787mの
山頂へ到着した。記念撮影後、14:30に下山を開始し、16:00に
牧ノ戸峠に帰着した。牧ノ戸峠から山頂までの距離は4.6km
であった。**往復の平均移動速度を求めよ**（山頂の休憩時間は
除く）。

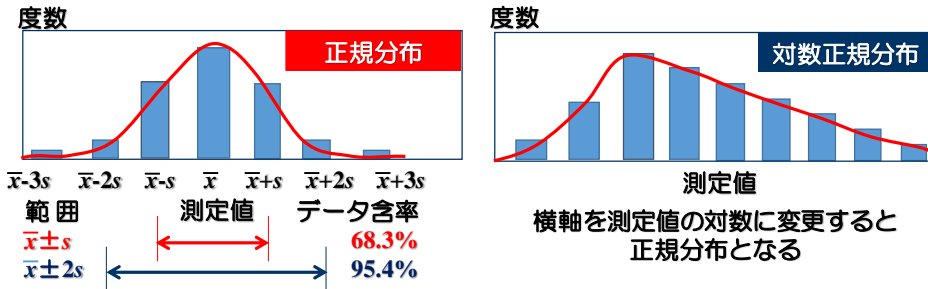
2. 1限の講義は大変だね。

バスで伊都地区から馬出地区へ移動した。
朝の渋滞に巻き込まれ、往路の速度は20km/hr,
復路の速度は30km/hrであった。**往復の平均移動速度を
求めよ。また、この平均の名称を答えよ。**

10

2. 正規分布（ガウス分布）・対数正規分布

自然界で観測される事象は平均の近くで起こることが多く、平均から離れるに従って起こりにくくなる。



正規分布の確率密度関数：左右対称の釣鐘状の分布曲線

$$f(x) = \frac{1}{\sqrt{2\pi}s} \exp\left(-\frac{(x-\bar{x})^2}{2s^2}\right)$$

s ：標準偏差 s^2 ：分散
 x ：測定値 \bar{x} ：測定値の平均

測定値の分布が正規分布に従うことを前提に統計量を求めている

11

3. 分散・標準偏差

測定値が平均近傍に集まっているか、あるいは平均から離れた点にも存在しているかどうかを評価する（群間比較に適用）。

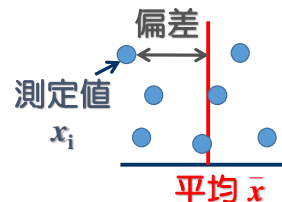
分散 (s^2) variance

測定値 x_i と平均 \bar{x} との差の自乗和を測定値の数 n で除算

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

標準偏差 (s) SD = 分散の平方根

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



分散と標準偏差を求めるとき、**分母に自由度 ($n-1$) を適用**することが多い（MS Excel）：不偏分散・不偏標準偏差

12

4. 中央値 (median) ・ 最頻値 (mode)

中央値 (正規性が観られない状況で用いられる)

測定値を大きさの順に並べた場合、その中央となる測定値。

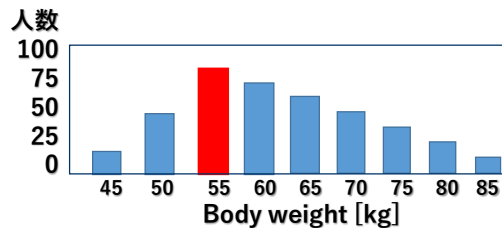
測定データの個数が奇数ならば、その中央にある値。

一方、偶数ならば、中央にある2値の算術平均。

測定値の最小値や最大値が集団から大きく離れている場合、平均よりも中央値を使用した方が良い。(年収の統計量)

最頻値

度数分布表 (ヒストグラム) において、頻度が最高となる測定値。



13

では

本題にはいります

14

5. 検定とは（１）

C. MS Excelを利用した統計処理（分散・平均の検定）

コイントスにてコインの表裏を当てるゲームを考えよう。あなたは表裏を確実に当てることができる特別な能力を有しているかな？—さて、この手のゲームにおいて、よく見られる１シーンに次の様な応酬がある。

「君にそんな能力はないよ!」、「5回連当（3.125%の確率）したら能力があると認めるよ!」。前者は「君の能力は特別ではない」、後者は「まぐれにしては出来過ぎだから認めるが、一方でその確率で誤りを犯す」といった見解である。

15

5. 検定とは（２）

C. MS Excelを利用した統計処理（分散・平均の検定）

統計学では、この一連のやり取りを**検定**といい、前者（特別能力はないと設定する）を**帰無仮説**、後者（判断基準になる確率）を**有意水準または危険率**という。ここでは、配布される教材を用いて以下の演習１～４を実践し、**分散および平均の検定手法を学ぶ**。

帰無仮説は「**差が無い**」と置くものである。

有意水準（危険率）は**解析者が決める**。

16

演習要領

配布資料「C. MS Excelを利用した統計処理（分散・平均の検定）」
の内容に従い、演習1～演習4を解き、**その回答を提出ファイル**
「C. MS Excelを利用した統計処理（分散・平均の検定）」に記入する。

そして、**本日中にe-mailに添付して回答を記入した提出ファイルを提出**
してください。

なお、提出ファイルのファイル名は「C.学籍番号+氏名」とすること。
また、e-mailの本文にも学籍番号と氏名を記載すること。

17

「分析ツール」のアドイン

MS Excelを起動 → 空白のブックを開く

データTabに

「データ分析」ツールがアドインされているか？

Yes の場合 設定の必要なし（終了）

Noの場合 （アドイン無し）

ファイルTab ⇒ オプション ⇒ アドイン
管理：Excelアドイン設定ボタンクリック／
分析ツールに✓ OKボタンクリック／



データTabに「データ分析」ツールが
アドインされていれば、設定終了

18

演習1：カタログ値との比較 - 平均の検定

手続き

1. MS-Excelの新規ファイルをひらく。
2. Sheet1に表7のデータを入力する。
3. 「データ分析」の「基本統計量」ツールを用いて、平均と標準偏差を求める。
4. 式(6)を用いて t_0 値を求める。また、 t 分布表を用いて t 値も求める。
5. 平均の検定結果を得る。
6. 提出ファイルに、平均、標準偏差、 t_0 値、 t 値、検定の結果を記入する。

19

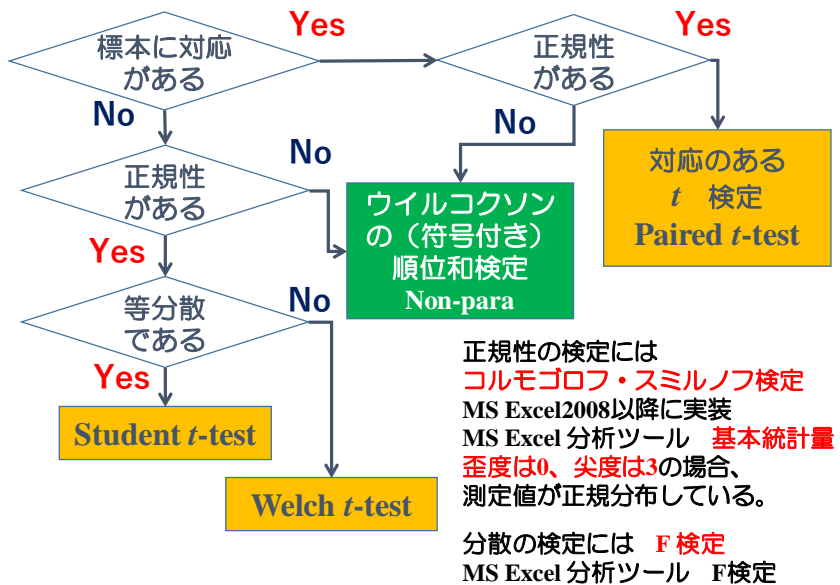
演習2：カタログ値との比較 - 流量の分散の検定

手続き

1. 演習1のMS-ExcelのSheet2に表7のデータを入力する。
2. 「基本統計量」ツールを用いて分散を求める。
3. 式(7)を用いて X^2_0 値を求める。また、 X^2 分布表を用いて X^2 値も求める。
4. 分散の検定結果を得る。
5. 提出ファイルに、分散、 X^2_0 値、 X^2 値、検定の結果を記入する。

20

二群の比較：平均の検定法の選択



21

演習3：装置の比較 - 標本に対応が無い二群の平均の検定

手続き

1. 演習1のMS-ExcelのSheet3に表8のデータを入力する。
2. 「基本統計量」ツールを用いて平均と標準偏差を求める。
3. 「データ分析」・「F検定：2標本を使った分散の検定」ツールを用いて、分散の検定結果を得る。
4. 「データ分析」・「t検定：等分散を仮定した2標本による検定」または「t検定：分散が等しくないと仮定した2標本による検定」を用いて、平均の検定結果を得る。
5. 平均の検定の結果をグラフで表示する。
6. 提出ファイルに、平均、標準偏差、検定の結果、グラフを記入する。

22

演習4：薬効の評価 - 標本に対応が**有る**二群の平均の検定

手続き

1. 演習1のMS-ExcelのSheet4に表9のデータを入力する。
2. 「基本統計量」ツールを用いて平均と標準偏差を求める。
3. 「データ分析」・「t検定：一対の標本による平均の検定」ツールを用いて、平均の検定結果を得る。
4. 平均の検定の結果をグラフで表示する。
5. 提出ファイルに、平均、標準偏差、検定の結果、グラフを記入する。

23

お し ま い

技術習得が目的ですので、操作方法がわからない
課題を放置せず、**全課題を解きましょう。**

レポートを**必ず提出**しましょう。
出席状況の判断材料となります。

24