

第2回 平均と分散・データの要約

A. データの要約

1. 100個以上のデータを要約する 5品種のイネの1穂穎花数について

ぱっとデータを見て、何かいえるか

大きさの順に並べてみる 何かいえるか

分布を書いてみる

度数分布とヒストグラムの書き方 いくつに分級すべきか

区間の幅をどう決めるか

書き方にはいくつかの方法がある．ホームページを参照のこと

代表値を考える

() , () , ()

() を考える

() , () , () , ()

()

2. さまざまな統計量の定義

サンプル数を n , データを $x_1, x_2, x_3, x_4, \dots, x_n$ とする .

代表値 中心を表す値

()

$$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4 + \dots + x_{n-1} + x_n}{n} = \sum_{k=1}^n x_k \div n = \frac{\sum_{k=1}^n x_k}{n}$$

()

()

ばらつきを表す値

偏差

()

$$V = \frac{S}{n-1}$$

平方和

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

()

$$S.D. = \sqrt{V}$$

()

()

()

別のデータについて、もう一度データの要約をやってみよう。

針に糸を通す時間										
	8	26	11	16	21	15	13	39	18	10
	8	25	5	14	35	5	6	43	22	21
	49	12	4	40	19	10	15	35	7	33
	9	6	20	6	21	1	14	15	7	16
	19	7	6	32	15	9	27	13	23	18
	19	17	12	6	30	5	6	15	23	9
	13	8	10	31	24	8	16	11	15	2
	13	47	13	28	8	6	8	7	40	13

上のデータを入力する。

大きさの順に並び替える

度数分布とヒストグラムを書く

統計量を求める。

平均

分散

標準偏差

メジアン

最大値

最小値

レンジ

第1四分位量

第3四分位量

変動係数

3. 実際に計算してみよう.

次のデータから平均, 分散, 標準偏差, メジアン, レンジ, 変動係数 (%) を計算せよ.

例題 3, 5, 6, 8, 11

11, 13, 15, 17, 22, 27, 36, 41, 55

1, 2, 5, 6, 10, 13, 15, 18, 22

1, 3, 5, 6, 100

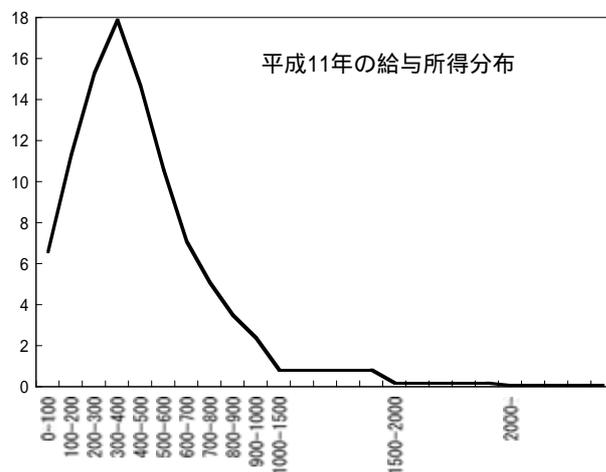
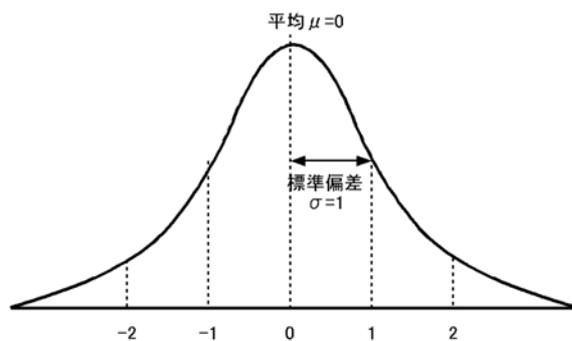
4. データの要約について

- 1) データを何らかの傾向を表す表にまとめる 大きさの順に並べる, 度数分布
- 2) ヒストグラムなどの図を書く
- 3) 平均など中心的傾向を示すような値を求める
- 4) ばらつきを評価する値を求める

B. 要約されたデータから何を読みとるか?

1. それぞれの統計量は何を意味するのか? どういう利用価値があるのか?

中心的傾向を表す代表値として

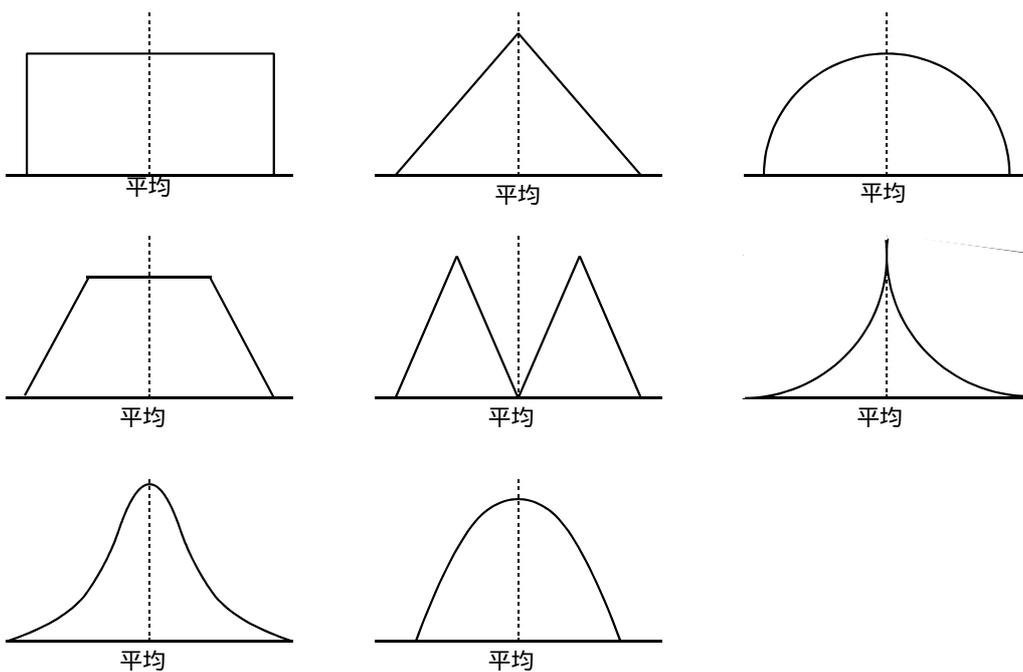


ばらつきを評価する指標として

2. 要約する過程でどのくらいの情報が失われたのか？

3. 種々の統計量から分布を再構築してみよう

何らかのデータについてどのような分布が想定できるか。



C. 宿題

1. 前回の宿題 2. について, 今回学んだデータの要約の仕方を利用して, 平均, メジアン, 分散, 標準偏差, レンジ, 変動係数を求めよ. さらに 3 種類書いたヒストグラムそれぞれについてモードを求めよ.

2. データの分布の形にはいろいろ考えられる. 次回以降の授業では二項分布, ポアソン分布, 正規分布について学ぶ. 予備的に次の調査を行う.

二項分布に従うと考えられる以下の現象のうち, 一つを選んで自分で実際に最低でも 25 回以上実験・調査せよ. その結果を度数分布になおし, ヒストグラム (階級範囲は 1 つずつとし, 0 と 1 をいっしょの階級にしない) を書け. さらに平均, 分散を計算せよ. 集めたデータからわかったことを箇条書きで書け.

さいころを 10 回振って, そのうち 1 が何回出るか?

トランプから 1 枚のカードを抜く. 元に戻して, 再びよく混ぜる. 10 回カードを引いて, ハートが何枚出るか.

麻雀牌から 1 枚の牌を抜く. 元に戻して, よく混ぜる. 10 回牌を引いて, 字牌が何枚出るか.

10 本の鉛筆を用意し, 1 本だけ赤鉛筆を入れる. よく混ぜて 1 本を取り出す. 取り出した後, その鉛筆は元に戻す. これを 10 回行い, 赤鉛筆を何回引いたか.

めったに起こらないことはポアソン分布に従うと考えられる. 以下の現象のうち, 一つを選んで自分で実際に実験・調査せよ. その結果を度数分布になおし, ヒストグラム (階級範囲は 1 つずつとし, 0 と 1 をいっしょの階級にしない) を書け. さらに平均, 分散を計算せよ. 集めたデータからわかったことを箇条書きで書け.

ここ 100 年間日本において火山の噴火した回数 (理科年表でしらべることができる).

ここ 100 年間において, 世界で大地震の起こった回数 (理科年表でしらべることができる).

100 個入りのお菓子里に不良品に入っている数を 100 袋について調べる.

鳥根県住民がここ 100 回の宝くじの特等に当たった人数.

100 日間における鳥根県での交通事故の発生件数

警察庁のホームページ (<http://www.npa.go.jp>) から統計 交通事故発生状況 1. 死者日報 全国で見ると毎日, 交通事故死があるから, 県レベルでみた方がよい.

提出締め切りは 10 月 25 日 (月) 午後 1 時までに生物資源科学部 2 号館 204 室に提出のこと.