

第2回 統計学の基礎

★ 教材「生物統計学_代表値の計算 2013」を予習しながら空所を埋めておくこと

A. データの要約と代表値

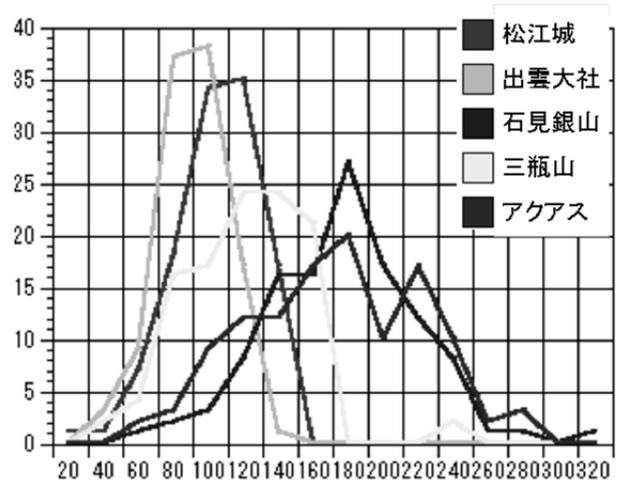
1. 100個以上のデータを要約する

① 中心を表す代表値を考える

()

()

()



② () を考える

()

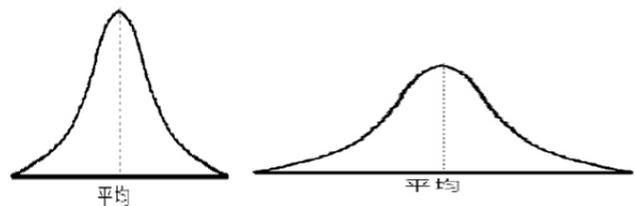
()

()

()

()

	平均	メジアン	モード
松江城 (赤)			
石見銀山 (青)			



2. 中心を表す代表値

① ()

例題

5羽のにわとりはそれぞれ1, 2, 0, 2, 0個の卵を産んだ。平均でいくつの卵を産んだか？

*補足 算術平均 (相加平均) と幾何平均 (相乗平均)

ふつう、平均というと今回紹介したすべての観測値を足し合わせて、観測数で割るものをさす。しかし、年平均成長率のように指数関数的成長をもとにした平均は幾何平均 (相乗平均) といってかけ算したものを n 乗根する平均を計算する。成長率の場合は注意が必要である。

- ② () データを大きさの順に並べて
データが奇数個なら中央に位置するデータの値,
データが偶数個なら中央に位置する2つのデータの平均である

例1 11, 14, 15, 17, 20 例2 2, 4, 8, 10, 12, 16

- ③ () 集めたデータの中で、最も多く現れた値あるいは階級のこと。
名目データではモードを代表値とする

例題

5羽のにわとりはそれぞれ1, 2, 0, 2, 0個の卵を産んだ。モードはいくらですか？

モードの例

★ 名目データの場合

アサガオを20株調べた。10株は青い花、6株は紫の花、4株は白い花だった。花の色のモードは？

★ ミニトマト5品種の着果数のデータの場合、以下のように度数分布をまとめたら、モードは松江城では101~120個、出雲大社では81~100個、石見銀山では161~180個、三瓶山では101~120個と121~140個、アクアスでは161~180個がモードである。この場合、三瓶山ではモードが2つある。

	松江城	出雲大社	石見銀山	三瓶山	アクアス
20	1	0	0	0	0
40	1	3	0	2	0
60	7	9	1	4	2
80	18	37	2	16	3
100	34	38	3	17	9
120	35	17	8	24	12
140	17	1	16	24	12
160	0	0	16	21	17
180	0	0	27	0	20
200	0	0	17	0	10
220	0	0	12	0	17
240	0	0	8	2	10
260	0	0	1	0	2
280	0	0	1	0	3
300	0	0	0	0	0
320	0	0	1	0	0
合計	113	105	113	110	117

3. データのばらつきを表す代表値

データがばらつくとは？

偏差

平方和 平均からの偏差の二乗の和

() V

単位は元の値の2乗となる.

()

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

()

最大値と最小値の差

もっとも両極端な値を使っているので信頼は低い

両方とも異常値であることもある.

四分位範囲

第3四分位点と第1四分位点の差である.

第1四分位点はデータを大きさの順に並べて、最小値から数えて、全体の4分の1番目に当たるデータの値である. 第3四分位点は同じく全体の4分の3番目に当たるデータの値である. メジアンは第2四分位点と一致する.

()

標準偏差を平均で割ったものでばらつきの大きさを相対的に評価する時に使う.

おおまかに 5%以下だと小さいばらつき, 6-14%はふつうのばらつき, 15%以上は大きいばらつきと考える

B. 代表値の計算の実際

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 山を調査して発見したツキノワグマの頭数 3, 5, 6, 8, 11 頭

1. エクセルの関数を使う計算方法

データ				=AVERAGE(C4:C8)
3	平均	6.6		=VAR(C4:C8)
5	分散	9.3		=STDEV(C4:C8)
6	標準偏差	3.04959		=MEDIAN(C4:C8)
8	メジアン	6		=MAX(C4:C8)-MIN(C4:C8)
11	レンジ	8		=STDEV(C4:C8)/AVERAGE(C4:C8)*100
	変動係数	46.20591		

2. エクセルの分析ツールを使う計算方法

① データ→データ分析 (エクセル 2003 ではツール→分析ツール)



② 基本統計量を選ぶ

③ 入力範囲にデータの範囲を入れる. データの最初にラベル (データ名) があるときはチェックを入れる. 出力先を指定する.



④ 指定した出力先に計算結果が出る.



データ	
平均	13
標準誤差	1.47196
中央値 (メジアン)	13
最頻値 (モード)	#N/A
標準偏差	7.359801
分散	54.16667
尖度	-1.2
歪度	0
範囲	24
最小	1
最大	25
合計	325
標本数	25

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.

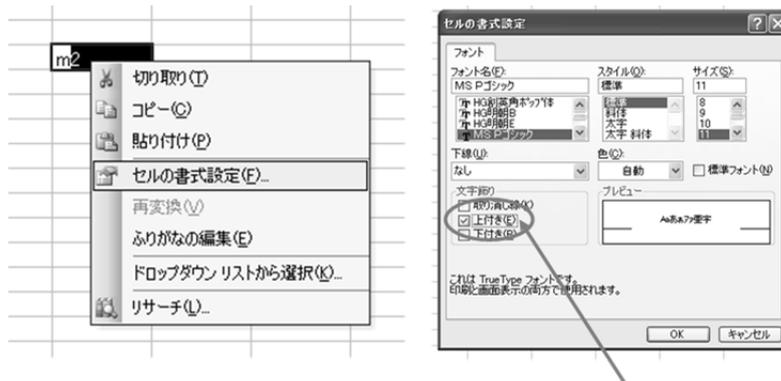
① 農園で収穫したミカンの数 11, 13, 15, 17, 22, 27, 36, 41, 55 個

- 平均
- 分散
- メジアン
- レンジ
- 標準偏差
- 変動係数 (%)

② 根の長さ 1.4, 2.7, 5.9, 6.3, 10.5, 13.1, 15.0, 18.2, 22.7cm

- 平均
- 分散
- メジアン
- レンジ
- 標準偏差
- 変動係数 (%)

★ 分散の単位は cm^2 のように 2 乗がつくことが多い。エクセルで 2 乗を示すのは以下のよう
すればよい。



予習問題

問2.	番号	距離 (cm)
右のデータはカエルのジャンプした距離を100回測定したものである。	1	64
① 平均, メジアン, 分散, 標準偏差, レンジ, 変動係数を計算せよ。	2	58
	3	51
② 問1-②で作成した3つのヒストグラムそれぞれについて, モードを求めよ。	4	74
	5	45
	6	28
	7	31
	8	47
	9	25
	10	45
	11	44
	12	65
	13	33

平均

メジアン

分散

標準偏差

レンジ

変動係数 (%)

モード

階級範囲その1	
0~21cm	10
22~35cm	31
36~49cm	37
50~63cm	18
64~cm	3

階級範囲その2	
0~19cm	9
20~26cm	12
27~33cm	12
34~40cm	18
41~47cm	22
48~54cm	15
55~61cm	6
62~68cm	4
69~75cm	1
76~cm	1

階級範囲その3	
0~18cm	5
19~23cm	8
24~28cm	12
29~33cm	8
34~38cm	15
39~43cm	12
44~48cm	14
49~53cm	12
54~58cm	7
59~63cm	3
64~68cm	2
69~73cm	0
74~78cm	1
79~83cm	1
84~cm	0

4. データの要約について

- 1) データの全体的な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) () などの図を書く
- 3) 平均など () 化の傾向を示すような値を求める
- 4) () を評価する値を求める

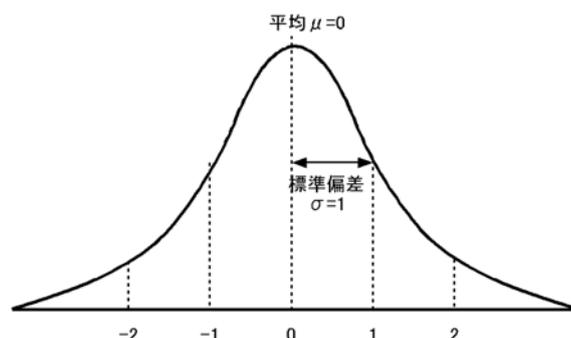
C. 要約されたデータから何を読みとるか?

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?

1. 中心化の傾向を表す代表値として

① 平均

平均は分布に () のないとき () にもっとも有用な統計量である. とくに母集団が () を示すときに非常に有用である.



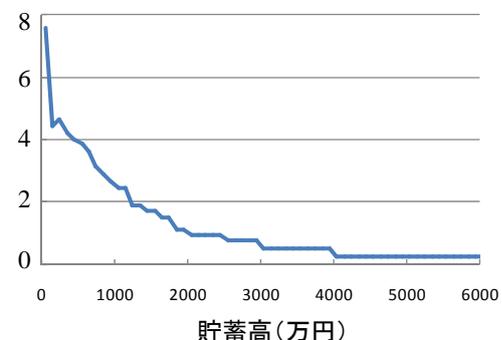
分布がゆがんでいるとき, 平均の利用には注意が必要である

② メジアン

右の給与分布の場合, 平均値とメジアンにはずれがある. 一般的な年収の人を考えると, 年収が平均の人とメジアンの人を選べば, それぞれ別のタイプの人になる.

メジアンは正規性のないデータあるいは標本全体を同時に調査できないようなデータ (性成熟日数, 生存日数, 電球の寿命のような打ち切り標本, 以上・以下を含む標本など) などの形質に用いる. データに異常値があるときのように, 極端に離れた値を含む標本では平均よりメジアンを用いる方が無難である.

(%) 日本の世帯別貯蓄高の分布(2007年)



③ モード

色などの名目 (質的) データならモードを使うしかない.

右の表はある交差点を通過した車の色を調べた結果である. この場合, モードは白となる.

白	47
黒	10
黄	8
赤	7
緑	2
青	1
その他	6

量的データでは度数分布の階級のとりかたによってモードが変わる.

宿題のカエルのジャンプのデータでは3つの度数分布 (ヒストグラム) それぞれについてモードを求めることができる.

このように量的データではデータのまとめ方次第でモードが変わるので、モードの使用には注意が必要となる。

2. ばらつきを評価する指標として

分散・標準偏差

たいていの場合、平均と分散（標準偏差）の2つをばらつきの指標としたらよい

レンジ

計算が簡単なので、標準偏差の代用とすることもある。異常値に左右されやすいので、信頼性は低い

四分位範囲

あまり使わないが、メジアンといっしょに分布のゆがんだデータで利用される。

★ 演習問題

A君はある昆虫の体重を100匹測定した。その結果はエクセルファイルにある。

- ① 平均、メジアン、分散、標準偏差、レンジ、変動係数を計算せよ。
- ② 度数分布とヒストグラムをかけ。
- ③ ②をみて、この昆虫の集団の平均的な姿を考えよ。

D. 代表的な離散分布

★ 教材「生物統計学__二項分布 2013」を予習しながら空所を埋めておくこと

1. 二項分布

① 二項分布

大きさ n の標本で、事象 E の起こる確率を p とするとき、そのうち x 個に E が起こる確率 $P(x)$ は二項分布に従う。

例 さいころを 10 回振ったときに 1 の出る回数 x の確率分布は二項分布に従う。

この場合、 $n = 10$, $p = \frac{1}{6}$ の二項分布になる

さいころを 10 回振ったときに 1 が 0 回出る ($x = 0$) 確率は、 $P(0) = 0.162$ である。

さいころを 10 回振ったときに 1 が 3 回出る ($x = 3$) 確率は、 $P(3) = 0.155$ である。

★ 二項分布では母平均 $\mu = np$, 母分散 $\sigma^2 = np(1-p)$ となる。

エクセルでの計算

例 さいころを 10 回振ったときに 1 の出る回数 x の確率分布は二項分布に従う。

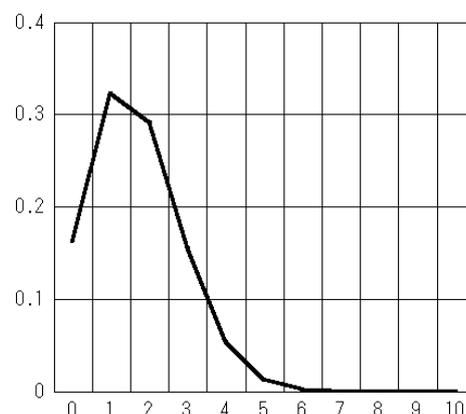
$n = 10$, $p = \frac{1}{6}$ の二項分布になる

エクセルでは $P(x) = \text{BINOMDIST}(x, n, p, \text{false})$

例をエクセルで計算すると以下の通りになる。

さいころの場合		
1 の出る回	確率	エクセルの二項分布関数
0	0.161506	=BINOMDIST(0,10,1/6,FALSE)
1	0.323011	=BINOMDIST(1,10,1/6,FALSE)
2	0.29071	=BINOMDIST(2,10,1/6,FALSE)
3	0.155045	=BINOMDIST(3,10,1/6,FALSE)
4	0.054266	=BINOMDIST(4,10,1/6,FALSE)
5	0.013024	=BINOMDIST(5,10,1/6,FALSE)
6	0.002171	=BINOMDIST(6,10,1/6,FALSE)
7	0.000248	=BINOMDIST(7,10,1/6,FALSE)
8	1.86E-05	=BINOMDIST(8,10,1/6,FALSE)
9	8.27E-07	=BINOMDIST(9,10,1/6,FALSE)
10	1.65E-08	=BINOMDIST(10,10,1/6,FALSE)

上の結果をグラフで示すと右のようになる。



予習での練習 A社のチョコレートにはくじが入っていて、当たる確率は 0.15 である. 10 個買って1つも当たりが入っていない確率, 2つだけ当たりの入っている確率を求めよ. さらに下の表を完成させよ.

$n =$, $p =$

当たりの数	エクセルでの計算式	確率
0		
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		

② 二項分布の利用

二項分布すると考えられる場合, それを利用していくつかの予測や推論を立てることができる.

(1) 確率の計算

例: シロクマチョコレートを買うと, 20 個に 1 個の割合でシロクマのぬいぐるみが当たる. どうしてもシロクマのぬいぐるみがほしい A さんはシロクマのぬいぐるみが手に入る確率を 95% 以上にするにはいくつシロクマチョコレートを買えばよいか? 99% 以上にするにはいくつ買えばよいか?

95% 以上の確率にするにはいくつ買えばよいか?

求める確率 = $1 -$ (ひとつもシロクマのぬいぐるみが当たらない確率) である.

99% 以上の確率にするにはいくつ買えばよいか?

(2) 信頼区間の計算

例：シロクマチョコレートを買うと、20個に1個の割合でシロクマのぬいぐるみが当たる。シロクマチョコレートを20個買うと、シロクマのぬいぐるみが当たる個数はいくつか？

当たる個数	確率
0	0.358
1	0.377
2	0.189
3	0.060
4	0.013
5	0.002
6	0.000
7	0.000
8	0.000
9	0.000
10	0.000
11	0.000
12	0.000
13	0.000
14	0.000
15	0.000
16	0.000
17	0.000
18	0.000
19	0.000
20	0.000

平均すれば1個当たるけれども、実際には1つも当たらない場合もあれば、2つ当たることもある。このような場合、決められた確率で何個から何個まで当たると表現する。これを**区間推定**といい、このような範囲を**信頼区間**、決められた（宣言した）確率を**信頼率**という（詳細は第4回の授業で学ぶ）。

20個のシロクマチョコレートを買うと1個ぬいぐるみが当たる確率が一番高いといっても0.377の確率であり、それほど確実に起こることではない。そこで92%の確率で0個から2個当たるといように確率を宣言して、範囲で示す。

(3) 真偽を推測する：

例：シロクマチョコレートを買うと、20個に1個の割合でシロクマのぬいぐるみが当たる。どうしてもシロクマのぬいぐるみがほしいAさんはシロクマチョコレートを50個買ったのに1つも当たらなかった。Aさんはシロクマチョコレート株式会社はうそつきだと断定した。しかし、その推論は正しいか？

50個シロクマチョコレートを買っても1つもシロクマのぬいぐるみが当たらない確率は二項分を利用すると（ ）である。

この確率はあまりにも小さいから不自然なのでシロクマチョコレート株式会社は嘘をついたと判断できるだろうか？あるいはたまには起きそうなことなので嘘をついたとは断言できないだろうか？

20個に1個の割合でシロクマのぬいぐるみが当たるのだから20個買うと1個は当たりそうにふつうの人は期待する。しかし、20個シロクマチョコレートを買って、ぬいぐるみが少なくとも1個当たる確率は（ ）である。これほど確率が低いのは20個シロクマチョコレートを買って、2個以上当たる幸運な人がいるからで、その確率は（ ）である。

このように確率分布を使うと標本（この場合、数を決めて購入したシロクマチョコレートのくじ）から母集団（シロクマチョコレート株式会社を作るすべてのシロクマチョコレートのくじ）について推論することができる。このような確率分布を利用した推論（推定と検定）については第4回以降の授業で学ぶ。

③ 演習問題 1

A釣り堀には1種類の魚がたくさん泳いでいる。この魚のうち、20%は雌である。魚はランダムに泳ぐので、釣り針にもランダムにかかるでしょう。この釣り堀で5匹の魚を釣ったときに雌が3匹以上いる確率を計算せよ。

④ 演習問題 2

B園芸にある植物を10株納めてほしいという注文が来た。しかし必ず赤い花が咲いた条件で納めることという条件がついていた。この植物は30%の確率で赤い花が咲く。それではいくつ鉢を準備したら、95%以上の確率で赤い花を10株準備できるだろうか？

★ 教材「生物統計学_ポアソン分布 2013」を予習しながら空所を埋めておくこと

2. ポアソン分布

① ポアソン分布

ポアソン分布は一定の長さの時間、一定の大きさの空間においてごくまれに起こる事象を表現するときに用いる。二項分布において p をどんどん小さくする一方で、 n を無限大にすると得られる。非常に大きな集団においてきわめて起こりにくい事象を対象としたときの分布である。

二項分布と違って、分布の大きさ n は必要ない。例えば、交通事故死はきわめてまれなものである。その対象となる n はしかも何人か決めようがない。運転者や歩行者の数は毎日異なるからである。そういうときにポアソン分布は有効である。

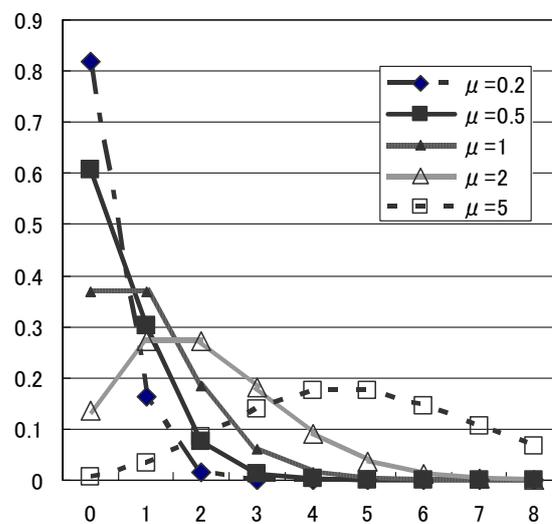
例 ある島では毎年、何千羽ものヒナが生まれる。毎年平均 0.5 羽の出現率で黄金色の羽をもつヒナが生まれるという。 $\mu = 0.5$ であるから、黄金の羽を持つヒナが 0,1,2 羽、出現する確率はそれぞれポアソン分布に従う。(ここでは毎年それぞれ誕生するヒナの数が何千羽であるかを正確にわかっていなくてもよいし、異なっていたとしてもよい。そこが二項分布と違う。)

ポアソン分布の計算には母平均だけが必要である。ここでは母平均 $\mu = 0.5$ である。

エクセルでは、 $P(x) = POISSON(x, \mu, false)$

黄金色の羽をもつヒナの数	確率	エクセルでの式
0	0.606531	=POISSON(0,0.5,FALSE)
1	0.303265	=POISSON(1,0.5,FALSE)
2	0.075816	=POISSON(2,0.5,FALSE)

- ★ ポアソン分布では平均 μ が決まると分布の形が決まる.
- ★ ポアソン分布では, 母平均 μ と母分散 σ^2 は等しい.



予習での練習 A君は1時間, 夜空を見上げると流れ星を平均 0.4 個みるという. ポアソン分布に従うとすれば, 1時間で流れ星を 0, 1, 2 個見る確率を計算せよ.

$\mu =$

当たりの数	エクセルでの計算式	確率
0		
1		
2		

E. 予習と宿題

予習と宿題は <https://moodle.cerd.shimane-u.ac.jp/moodle/> を見てください.