第 1 講　記述統計学①　------------------------------------

●統計学とは何か?

1 記述統計学

2 推測統計学

●データの整理

ex テストの点数(12 人)

41 59 61 64 18 89

38 27 64 75 81 31

☆データの分類

(i)階級

(ii)階級値

(iii)度数

(iv)累積度数

(v)相対度数

(ii)累積相対度数

階級	階級値	度数	累積度数	相対度数	累積相対度数

（発展）

Q 階級数はどうきめるべきか

A

[1]平方根選択

[2]スタージェスの公式

導出

演習問題 1

次のデータに対し、度数分布表を完成させ、ヒストグラムを作れ

37 40 32 26 30

41 32 35 37 34

階級	階級値	度数	累積度数	相対度数	累積相対度数

演習問題 2

データ数が 64 であるときの階級数を(i)平方根選択、(ii)スタージェスの公式で見積もれ

第 2 講　記述統計学②

☆データの特徴づけ

ex　テストの点数(12 人)

41 59 61 64 18 89

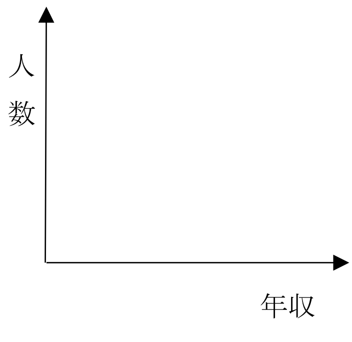
38 27 64 75 81 31

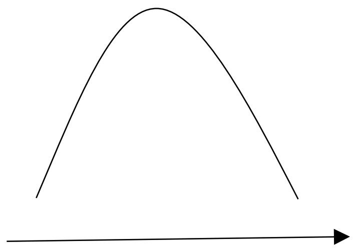
(i)平均値μ

(ii)中央値Me

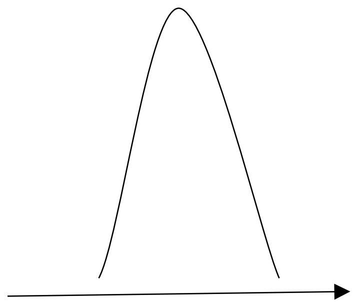
(iii)最頻値Mo

ex　年収





→ データのバラツキを定量化したい



(iv)分散σ^2

ex

演習問題 1

$\sigma^2 = \overline{x^2} - (\bar{x})^2$を示せ

分散はデータの単位の 2 乗になっている

(v)標準偏差σ

(vi)共分散σ_{xy}

(vii)相関係数r

演習問題 2

$\sigma_{xy} = \overline{xy} - \bar{x} \cdot \bar{y}$を示せ

演習問題 3

次のデータに対し、平均、分散、標準偏差を求めよ

5、7、3、2、6、1

●平均と分散の性質

度量xから作った新しい度量

$$y = ax + b$$

に対し、

(i)平均値は

となる

証明

(ii)分散は

となる

証明

意味

(iii)標準偏差は

となる

オマケ　偏差値とは何か?

テスト A	たくみ君	70 点	平均 50 点
テスト B	まなか君	60 点	平均 50 点

Q どっちが優秀?

第 3 講　確率論の基本①

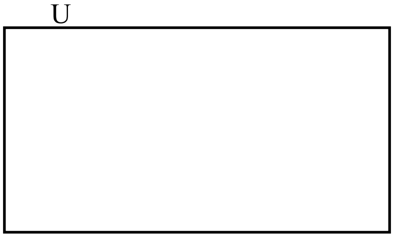
●確率とは何か

●確率変数

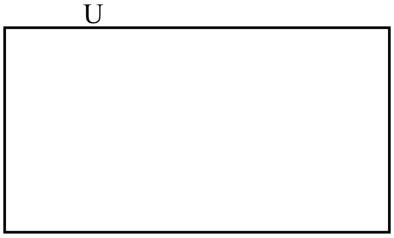
ex

●加法定理

U



U



基本事項

互いに排反な事象A_1, A_2, A_3に対して、

$$P(A_1 \cup A_2 \cup A_3 \cup \cdots)$$
$$= P(A_1) + P(A_2) + P(A_3) + \cdots$$

が成り立つ

●独立試行

基本事項

2 つの独立な試行で得られる事象をA, Bとするとき
事象A, Bが同時に起こる確率は

で与えられる

●反復試行（独立試行の応用）

基本事項

事象Aが起こる確率をPとする。
この試行を独立にn回くりかえしたときに事象Aがr回起こる確率は

と表される

ex

演習問題 1

A, Bの 2 人があるゲームを行い、先に 3 勝したほうが優勝とする。
1 回の試合でAが勝つ確率が$^1/_3$であるとき、Aが優勝する確率を求めよ

(解答)

●条件付き確率

基本事項

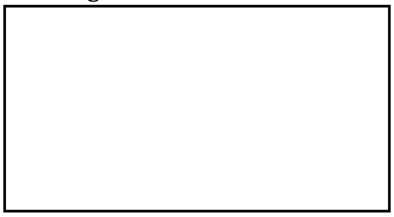
事象A,Bに対して、事象Aが起こったときの事象B起こる条件付き確率は$P_A(B)$と表され

で計算できる

理由

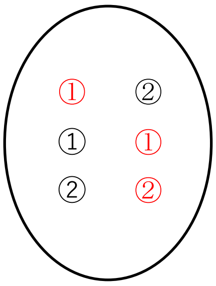
U

U



演習問題 2

1 つの球を取り出したら赤球であった。それにかかれている数字が 1 である確率を求めよ。



演習問題 3

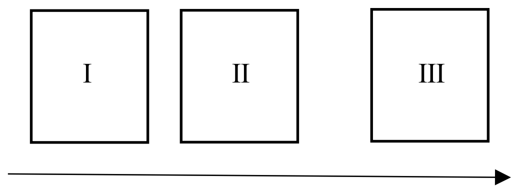
サイコロを 1 回振ったとき、出た目が 3 より大きいことが分かった。

このとき、それが偶数である確率を求めよ

●ベイズの定理

導出

ex



5 回に 1 回ぼうしを忘れる。

IIに忘れた確率は?

演習問題 4

ウソをウソと判定する確率が 99%、ホントをホントと判定する確率が 95%のウソ発見器がある。

この発見器がウソと判定したとき、被疑者がウソをついている確率を求めよ。

ただし、被疑者がウソをつく確率は 0.1%とする。

演習問題 5

体質xをもつ母親の子どもの 80%が体質 x をもち、体質 x をもたない母親の子どもの 50%が体質 x をもつ。

また、体質 x をもつ母親は 60%いる。

いま、検査で子どもが体質 x をもつことがわかった。

この子どもの母親が体質 x をもつ確率を求めよ。

第 4 講　確率論の基本②

●確率分布

ex サイコロ 1 コ

X						
P						

ex サイコロ 2 つの目の和

X
P

●様々な分布

① 離散型一様分布

② 二項分布 $B(n, p)$

$B(15, 0.3)$の場合

③ ポアソン分布 $P_0(\lambda)$

$P_0(1.5)$の場合

●確率密度関数

① 連続一様分布

② 正規分布

●期待値と分散

確率変数

実現値

① 期待値

性質

$$E[c] = c$$

$$E[X + c] = E[X] + c$$

$$E[cX] = cE[X]$$

$$E[X + Y] = E[X] + E[Y]$$

X と Y が独立のとき

$$E[XY] = E[X]E[Y]$$

~記述統計との対応~

演習問題 1

(1) $E[aX+b]=aE[X]+b$

(2) $E[X+Y]=E[X]+E[Y]$

(3) $E[XY]=E[X]E[Y]$ (X,Yは独立)

を証明せよ(離散型の場合でのみ示せばよい)

② 分散

演習問題２

(1) $V[aX+b]=a^2V[X]$

(2) $V[X+Y]=V[X]+V[Y]$ (X, Yが独立)

を証明(離散型の場合のみ示せばよい)

演習問題 3

(1) 連続型一様分布

(2) 二項分布$B(n,p)$

の平均と分散を求めよ(ただし、$V[X]=E[X^2]-E[X]^2$を用いてもよい)

第 5 講　確率論③(正規分布の特徴)

基本事項　正規分布

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

確率変数 X が正規分布に従うとき、

①　期待値(平均値)

$E(x)=$

②　分散

$V(x)=$

☆再生性

基本事項

確率変数X_1とX_2が独立で、それぞれ平均μ_1, μ_2、分散σ_1^2, σ_2^2の正規分布に従うとき、
$X_1 + X_2$は平均$\mu_1 + \mu_2$、分散$\sigma_1^2 + \sigma_2^2$の正規分布に従う。

基本事項

平均μ、分散σ^2の正規分布に従う独立したn個の確率変数を$X_1, X_2, \cdots, X_n$とする。
このとき

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

は平均μ、分散$\frac{\sigma^2}{n}$の正規分布に従う。

☆中心極限定理

基本事項

平均μ、分散σ^2の正規分布に従う独立した n 個の確率変数を$X_1, X_2, \cdots, X_n$とする。
n が大きいとき

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

は平均μ、分散$\frac{\sigma^2}{n}$の正規分布に従う。

☆パーセント点

☆確率変数の標準化

基本事項

平均μ、分散σ^2(標準偏差σ)の確率変数 X を標準化した確率変数 Z は

$$Z = \frac{X - \mu}{\sigma}$$

となる

演習問題

2 枚のコインを投げた時の表の回数 X を標準化せよ。

☆確率正規分布(Standard normal distribution)

平均が 0 で分散が 1 である正規分布

演習問題

確率変数$X_1, X_2, \cdots, X_n$が互いに独立で同じ標準正規分布に従うとき、次の確率を求めよ。

(1) $P(-1 \leqq X \leqq 1)$

(2) $P\left(-1 \leqq \frac{X_1+X_2}{2} \leqq 1\right)$

(3) $P\left(-1 \leqq \frac{X_1+X_2+\cdots+X_5}{5} \leqq 1\right)$

第 6 講　推測統計学①(母集団と標本)　---------------------

●母集団と標本

※抽出は独立かつ等確率に行う

●母集団分布 (Population distribution)

母集団分布は多くの場合正規分布が仮定される

●標本分布 (Sample distribution)

基本事項	推定量

母平均 $\mu = \frac{x_1+x_2+\cdots+x_N}{N}$

標本平均 $\bar{X} = \frac{X_1+X_2+\cdots+X_n}{n}$

母分散 $\sigma^2 = \frac{(x_1-\mu)^2+(x_2-\mu)^2+\cdots+(x_N-\mu)^2}{N}$

不偏分散 $U^2 = \frac{(X_1-\bar{X})^2+(X_2-\bar{X})^2+\cdots+(X_n-\bar{X})^2}{n-1}$

Q. 何故、標本分散ではダメなのか?

$$S^2 = \frac{(X_1-\bar{X})^2+\cdots+(X_n-\bar{X})^2}{n}$$

まとめると

$E[S^2] = \frac{n-1}{n}\sigma^2$ …期待値が母数と一致しない。

ここで、$U^2 = \frac{n}{n-1}S^2$であるので、

※何故、標本分散は過小評価してしまうのか?

演習問題

ある雑誌がたくさん並んでいる。標本として 10 冊無作為に抽出してページ数を調べた結果、次の値が得られた。(単位はページ)

84,76,92,85,76
81,67,101,63,73

標本平均、標本分散、不偏分散を求めよ。(分散は少数第 1 位まで)

演習問題

全国の小学 6 年生の男子を無作為に抽出し、それぞれの身長を調べた結果、次の標本が得られた。(単位は cm)

$$148,152,154,146,150$$

標本平均、標本分散、不偏分散を求めよ。

第 7 講　推測統計学②(点推定)　----------------------------

推定 $\begin{cases} \text{・点推定} \\ \text{・区間推定} \end{cases}$

●最尤推定法 (Maximum likelihood estimation)

実際に得られた標本があるとき、それらが得られる確率が最大になるような母数の値をその推定値とする手法

例 表が出る確率がθであるようなコインがある。このコインを 10 回投げたら 6 回表が出た。最尤推定法によりθを推定せよ。

例

あるお菓子の平均内容量(g)を調べるために、標本として 3 つの袋を抽出したところ、

$$60,65,58(\mathrm{g})$$

であった。この標本から最尤推定法により、平均内容量を推定せよ。
ただし、母集団は正規分布に従うとする。

演習問題

平均がμ、分散がσ^2である正規母集団から無作為に n 個抽出したところ、$x_1, x_2, \cdots, x_n$であった。このときμとσ^2を最尤推定法で推定せよ。

●推定量に望む性質（θ: 母数、$\hat{\theta}$: 推定量）

基本事項

標本平均　$\bar{X}=\frac{X_1+X_2+\cdots+X_n}{n}$

不偏分散　$U^2=\frac{(X_1-\bar{X})^2+(X_2-\bar{X})^2+\cdots+(X_n-\bar{X})^2}{n-1}$

☆大数の法則(law of large numbers)（標本平均が母平均に限りなく近付く）

公　式　**マルコフの不等式**

Xを非負の値をとる確率変数とする。このとき、任意の$c>0$に対して

$$P[X\geq c]\leq\frac{E[X]}{c}$$

が成り立つ。

公　式　**チェビシェフの不等式**

$E[Y]=\mu, V[Y]=\sigma^2$とするとき、任意の$a>0$に対して

$$P[|Y-\mu|\geq a]\leq\frac{\sigma^2}{a^2}$$

が成り立つ。

演習問題

平均μ、分散 4 の正規母集団から抽出した標本$X_1 + X_2 + \cdots + X_9$について、
$P(|\bar{X} - \mu| \geqq 1)$の正確な確率をチェビシェフの不等式による評価と比較をせよ。
ただし、標準正規分布に従う確率度数 Z について$P(Z \geqq 1.5) = 0.0668$となることをもちいいてよい。

第 8 講　推測統計学③(区間推定Ⅰ)　-----------------------

推定
- ・点推定
- ・区間推定

<u>考え方</u>

母

ex　日本の成人男性の平均身長を調べるために、一人だけ無作為に抽出して調べたところ 165cm であった。平均身長を信頼度 95%で区間推定せよ。ただし、母集団の身長 X は分散6^2の正規分布に従うとする。

Q 標本の大きさを大きくしたらどうなるか?

ex 日本の成人男性の平均身長を調べるために、標本として次の 9 人の身長のデータを得た。

$$165\ 170\ 163\ 171\ 161$$

$$162\ 180\ 158\ 164$$

これらをもとに平均身長を信頼度 95%で区間推定せよ。ただし、母集団の身長 X は分散6^2の正規分布に従うとする。

比較

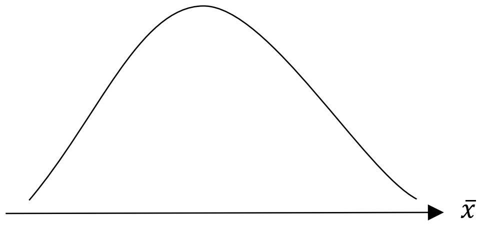
<u>信頼区間の意味</u>

まとめ

正規母集団に対する信頼度 α の信頼区間は

$$\bar{X} - k\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + k\frac{\sigma}{\sqrt{n}}$$

となる。ここで、k は標準正規分布の両側$100(1-\alpha)\%$点である



演習問題 1

神奈川県の 20 代男子の 100 人を無作為に抽出したところ、体重の平均は 65kg であった。この体重が分散10^2の正規分布に従うとするとき、平均体重μを信頼度 95%で区間推定せよ

演習問題 2

前問において、信頼度 99%で区間推定せよ。ただし、標準正規分布の両側 1%点は2.58である

演習問題 3

分散 9 の正規母集団から標本を抽出し、母平均μの区間推定を信頼度 95%で行う、信頼区間の幅を 4 以下にするには標本の大きさ n はいくら以上にすればよいか

$$\bar{X} - k\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + k\frac{\sigma}{\sqrt{n}}$$

を別の方法で導出してみる

第 9 講　推測統計学④(区間推定Ⅱ)　------------------------

〈母平均の推定〉

・母分散が既知

・母分散が未知

・母集団分布が未知

Q 母分散が分からない場合は?

A

公　式	t 分布
自由度$\nu(=n-1)$の t分布の確率密度関数	

例

ある場所を通る車の台数を調べた。その結果、各々の日で通った車の台数は

94 99 86 101 (台)

であった。正規母集団を仮定するとき、その平均台数μを信頼度 95%で区間推定せよ

まとめ

正規母集団に対する信頼度 α の信頼区間は

となる。ここで、kは自由度$n-1$の t 分布の両側$100(1-\alpha)\%$点である

Q 母集団分布が分からないときはどうする?

A

定　理　**中心極限定理**

平均μ、分散σ^2の母集団から抽出した標本の大きさが十分大きいとき、

標本平均$\bar{X}$は近似的に平均μ、分散$\frac{\sigma^2}{n}$の正規分布に従う

標本の大きさnが十分に大きいとき$\bar{X}$は正規分布に従う

→

例

中学生がもらうお年玉の平均値を調べるため、大きさ 100 の標本を取り出したところ、
$\bar{x} = 25000, u^2 = 5000^2$であった。平均金額$\mu$を信頼度 95%で推定せよ

演習問題 1

ある雑誌のページ数を 9 冊調べたところ、その標本の平均ページ数は 168
、不偏分散は6^2であった。この雑誌の平均ページ数μを信頼度 95%で推定せよ。
なお、自由度 8 の t 分布の両側 5%点は 2.30 である(正規母集団を仮定)

演習問題 2

中身の見えない箱に入ったボールの数の平均値を知るために大きさ 100 の標本を取り出したところ、$\bar{x} = 40, u^2 = 5^2$であった。箱に入ったボールの平均数μを信頼度 95%で推定せよ

●母比率の推定

Q この意見に賛成か?

演習問題 3

ある芸能人の認知度を調べるために、20 代の男性 400 人に「この人を知っていますか?」と聞いたところ、320 人が「知っている」と答えた。20 代男性への認知度pを信頼度 95%で推定せよ

第 10 講　推測統計学⑤(区間推定Ⅲ、検定)　---------------

●母分散の推定

定　理

分散σ^2の同一の正規分布に従う独立な n 個の確率変数を$X_1, X_2, \cdots, X_n$について、
このとき

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

として、次の和を考える

$$W = \left(\frac{X_1 - \bar{X}}{\sigma}\right)^2 + \left(\frac{X_2 - \bar{X}}{\sigma}\right)^2 + \cdots + \left(\frac{X_n - \bar{X}}{\sigma}\right)^2$$

この確率変数 W は自由度$n-1$のχ^2分布に従う

公　式　χ^2分布

自由度$\nu(=n-1)$の χ^2 分布の確率密度関数

ex

「1 日のうちゲームに費やす時間」の分散σ^2を推定するためにおおきさ 10 の標本を抽出したところ

126 34 224 11 38 14

178 199 25 89 (分)

という結果が出た。正規母集団を仮定し、分散σ^2を信頼度 95%で推定せよ。

まとめ

正規母集団に対する信頼度 α の信頼区間は

となる。ここで、k_1, k_2は自由度$n-1$の χ^2 分布の

と

である

演習問題 1

分散についての信頼区間を構成するために正規母集団から 20 個のデータをとったところ$u^2 = 30.0$が得られた。これをもとに母分散を信頼度 95%で区間推定せよ。ただし自由度 19 の χ^2 分布の下側 2.5%点は 4.40、上側 2.5%点は 23.3 である

演習問題 2

ある町の住民の体重の分散σ^2を推定するために、大きさ 10 の標本を抽出して調べたところ$u^2 = 32.6$であった。この町の体重の分散σ^2を信頼度 95%で推定せよ。ただし、自由度 9 の χ^2 分布の下側、及び上側 2.5%点はそれぞれ 2.70、19.0 である

●検定

あるコインを 20 回投げたとき、表が 15 回出た。このコインは裏より表の方が出やすいと言えるか?

ex

コインを 60 回投げる

① 「表と裏の出る確率は等しい」

「裏の出る確率は表が出る確率より大きい」

② 帰無仮説を認めたとき、裏の出る回数Xの分布を求める

③ 5%以下のことが起こったら、それは偶然とは言えないと約束しておく

④ 60 回中 38 回裏だった

流れのまとめ

①

②

③

④

※注　対立仮説を変えると結果は変わり得る

例

演習問題 2

ある機械で、ある食品を作っている。これまでのこの食品の平均の重さは 95.0g であったが、「重さが変わったのではないか」という指摘を受けた。この仮説を検定するために、10 個のデータを測定したところ次のようになった。

93.8 97.0 97.4 97.3 97.4

95.3 95.1 96.7 97.8 96.1

帰無仮説$H_0: \mu = 95.0$を有意水準 5%で検定せよ

ただし、母集団は母分散$\sigma^2 = 1.4$の正規分布に従うとする

演習問題 3

正規母集団から$n = 20$の標本を無作為に抽出したところ、$\bar{x} = 14.09$であった。これまで母平均は$\mu = 15.0$とされていた。母平均は小さくなったといえるか、有意水準 5%で検定せよ

ただし、母分散は 4 で変わらないものとする(標準正規分布の上側 5%点は 1.64 である)

演習問題 4

ある商品の長さを 11 個無作為に調べたところ

3.4 3.6 3.8 4.1 4.0

3.4 3.2 3.7 3.3 3.6 3.5

であった。母集団が正規分布に従うものとする。このとき、有意水準 1%として、$H_0: \mu = 3.95$を(1)$H_1: \mu \neq 3.35$、(2) $H_1: \mu > 3.35$の 2 つの場合について検定せよ

(自由度 10 の t 分布の両側 1%点は 3.169、上側 5%点は 2.764 である)