第 IX 部　統計的検定

1　標準偏差の意味

- 平均からの偏差
- 標準偏差
- 標準化
- z 値の意味

1.1　はじめに

- 極めて感覚的な話なのですが、平均的なことが起こっても驚かないと思います。
- それは普通だからです。
- 何をもって普通とするかは議論があるのは承知しています。
- 普通の範囲に一定の基準を持つことを考察します。

命題 1

a, b は共に正値とする。

$$a^2 < b^2 \quad \Longrightarrow \quad a < b \tag{1}$$

が成り立つ。

証明

$$a^2 < b^2 \tag{2}$$

$$0 < b^2 - a^2 \tag{3}$$

$$0 < (b - a)(b + a) \tag{4}$$

仮定により、$a > 0, b > 0$ なので、右辺第二因数は

$$a + b > 0 \tag{5}$$

である。

従って、

$$0 < b - a \tag{6}$$

$$a < b \tag{7}$$

(終)

1.2　相対的な位置の指標

- 表 1 はある小学校の生徒の身長と体重の記録である。身長を x_i であらわし、体重を y_i であらわす。
- ここで、それぞれの値が、全体に対してどのような位置にあるのかを考える。
- 序数データとして見れば、x_1 は 1 番目に小さく、x_2 は 2 番目に小さい。
- ここで考えたいのはどの程度小さいのかを把握することである。

表 1　小学生の身長と体重

i	x_i	y_i
1	116.5	21.3
2	122.6	24.2
3	128.6	27.7
4	133.5	31.1
5	138.6	34.4
6	145.0	39.2

1.2.1　平均を基準とする指標

ここで、それぞれの変数の平均を求めと、x の平均は

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{116.5 + 122.6 + 128.6 + 133.5 + 138.6 + 145.0}{n} = 130.8 \tag{8}$$

であって、y_i の平均は

$$\bar{y} = 29.65 \tag{9}$$

である。

このそれぞれの平均を基準として数値の位置を表現する。

1.2.2　平均からの偏差

それぞれの値と平均との差

$$x_i - \bar{x} \tag{10}$$

を『平均からの偏差』という。

- 『平均からの偏差』は、それぞれの値が平均より小さければ負の値をとり、平均より大きければ正の値をとる。
- 符号により平均を基準とする大小は容易に判定できる。

表 2　平均からの偏差

i	x	y
1	−14.3	−8.35
2	−8.2	−5.45
3	−2.2	−1.95
4	2.7	1.45
5	7.8	4.75
6	14.2	9.55

1.2.3　分散

符号の問題[1]により『平均からの偏差』の総和は常に 0 となる。

符号の問題を解決するために『平均からの偏差』の二乗を考える。『平均からの偏差』の二乗の平均を分散といい V_x であらわす。

$$V_x = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{11}$$

1）　正負の値が打ち消しあい総和が過少に評価されること。

表 3　平均からの偏差の二乗

i	$(x-\bar{x})^2$	$(y-\bar{y})^2$
1	204.49	69.7225
2	67.24	29.7025
3	4.84	3.8025
4	7.29	2.1025
5	60.84	22.5625
6	201.64	91.2025
Total	546.34	219.0950

1.2.4 標準偏差

x と y の分散はそれぞれ

$$V_x = 91.05666666\cdots \tag{12}$$

$$V_y = 36.51583333\cdots \tag{13}$$

である。

『平均からの偏差』の二乗の平均が分散なので、それぞれの『平均からの偏差』が平均から隔たっていれば、隔たっているほど分散は大きな値を示す。この特性を利用して、データのばらつきの指標として利用される。

分散の正の平方根を『標準偏差』といい、x の標準偏差を σ_x という記号であらわす。

x と y の標準偏差はそれぞれ

$$\sigma_x = \sqrt{V_x} = 9.542361692\cdots \tag{14}$$

$$\sigma_y = \sqrt{V_y} = 6.042833220\cdots \tag{15}$$

である。標準偏差は共に正なので命題 1 により

$$V_x < V_y \quad \Longrightarrow \quad \sigma_x < \sigma_y \tag{16}$$

である。

1.2.5 標準偏差で測った『平均からの偏差』

定義

『平均からの偏差』を『標準偏差』で割った値を z_i であらわし z 値 という。

$$z_i = \frac{x_i - \bar{x}}{\sigma_x} \tag{17}$$

x_i から z_i を求めることを『標準化』という。

標準化することにより、それぞれの x_i が平均から標準偏差の何倍離れているかを知ることができる。

表 4　標準化された x と y

i	x の z 値	y の z 値
1	−1.499	−1.382
2	−0.859	−0.902
3	−0.231	−0.323
4	0.283	0.240
5	0.817	0.786
6	1.488	1.580

1.3 確率

1.3.1 標本空間と事象

- 観測されうる結果を『標本点』とよび ω_i であらわす。
- 標本点の全体の集合を『標本空間』とよび Ω であらわす。
- 標本空間の部分集合を『事象』とよぶ。
- 標本点を一つも含まないことも事象とみなし『空(くう)事象』とよび $\varnothing$ であらわす。
- 標本点 $\omega_1, \omega_2, \cdots, \omega_n$ からなる事象 A_i は、$A_i = \{\omega_1, \omega_2, \cdots, \omega_n\}$ とあらわす。

1.3.2　事象の演算

- 事象 A_i と 事象 A_j 二つの事象において A_i と A_j が共通の標本点を持たない場合 A_i と A_j は『排反(はいはん)事象』であるという。また、A_i と A_j が共通部分を持たないことを A_i と A_j は排反であるという。
- 事象 A_i と事象 A_j 二つの事象のうち少なくとも一つがおこる事象を A_i と A_j の『和(わ)事象』とよび $A_i \cup A_j$ であらわす。
- 事象 A_i と事象 A_j 二つの事象が同時に起こる事象を『積(せき)事象』とよび $A_i \cap A_j$ であらわす。
- 事象 A が起こらない事象を『補(ほ)事象』とよび ${A_i}^C$ であらわす。
- 事象 A_i が起こる確率を $\Pr(A_i)$ とあらわす。$\Pr(A_i) = p_i$

1.3.3 確率の公理

確率 p_i は以下の性質を満たす。

$$0 \leqq p_i \leqq 1 \tag{18}$$

$$\sum_{i=1}^{n} p_i = 1 \tag{19}$$

互いに排反な事象 $A_1, A_2, A_3, \cdots$ に対して

$$\Pr\left(A_1 \cup A_2 \cup A_3 \cup \cdots\right) = p_1 + p_2 + p_3 + \cdots + \tag{20}$$

1.4　神の作ったコイン

コインを投げたとき裏の出る事象 E の確率を p とし、E の余事象 E^C の確率を $1-p$ とする。

- ここで、$p=0.5$ であるコインを考える。
- このコインを 8 回投げ裏の出る回数を x_i とあらわす。
- コインを投げる試行はそれぞれ独立な試行なので、同時確率は 0.5^8 である。

表 5　X の分布

X	度数	$\Pr(x_i)$
0	1	1/256
1	8	8/256
2	28	28/256
3	56	56/256
4	70	70/256
5	56	56/256
6	28	28/256
7	8	8/256
8	1	1/256

1.4.1 確率変数

取りうる値に対して確率が与えられている変数を『確率変数』という。確率変数は大文字を使ってあらす。

コインを投げたとき、コインの裏の出る回数は確率変数である。

1.4.2 期待値

確率変数の対応する確率による加重平均を『期待値』といい確率変数 X の期待値を $E(X)$ という記号であらわす。

$$E(X) = \sum_{i=1}^{n} x_i p_i \tag{21}$$

期待値は、ある値である。

1.4.3 数値例の期待値

裏の出る確率 (p) が $p = 0.5$ であるコインを 8 回投げたときの裏の出る回数を X とすると、確率変数 X の期待値は、

$$E(X) = \frac{0}{256} + \frac{8}{256} + \frac{56}{256} + \frac{168}{256} + \frac{280}{256} + \frac{280}{256} + \frac{168}{256} + \frac{56}{256} + \frac{8}{256} \tag{22}$$

$$= 4 \tag{23}$$

である。

1.5 期待値の演算

期待値に関して以下の関係が成り立つ。

$$E(c) = c \quad ; c \text{ は定数} \tag{24}$$

$$E(X + c) = E(X) + c \quad ; c \text{ は定数} \tag{25}$$

$$E(aX) = a\ E(X) \quad ; a \text{ は定数} \tag{26}$$

$$E(X + Y) = E(X) + E(Y) \quad ; x_i, y_i \text{の対応する確率はともに } p_i \tag{27}$$

問題 IX–1 – 1

a, c を定数とし、X, Y を確率変数とする。以下の関係が成り立つことを示しなさい。ただし、x_i, y_i に対応する確率は p_i とする。

1. $E(c) = c$
2. $E(X + c) = E(X) + c$
3. $E(aX) = a\ E(X)$
4. $E(X + Y) = E(X) + E(Y)$

解例 IX−1 − 1

$$
\begin{aligned}
E(c) &= \sum_{i=1}^{n} c\ p_i \\
&= c\sum_{i=1}^{n} p_i \\
&= c
\end{aligned}
$$

$$
\begin{aligned}
E(X+c) &= \sum_{i=1}^{n} (x_i + c)\ \ p_i \\
&= \sum_{i=1}^{n} (x_i p_i + c p_i) \\
&= \sum_{i=1}^{n} (x_i p_i) + \sum_{i=1}^{n} (c\ p_i) \\
&= E(X) + c
\end{aligned}
$$

$$
\begin{aligned}
E(aX) &= \sum_{i=1}^{n} a x_i p_i \\
&= a \sum_{i=1}^{n} x_i p_i \\
&= a\ E(X)
\end{aligned}
$$

$$
\begin{aligned}
E(X+Y) &= \sum_{i=1}^{n} (x_i + y_i)\, p_i \\
&= \sum_{i=1}^{n} (x_i p_i + y_i p_i) \\
&= \sum_{i=1}^{n} x_i p_i + \sum_{i=1}^{n} y_i p_i \\
&= E(X) + E(Y)
\end{aligned}
$$

1.6 期待値からの偏差

確率変数と期待値の差を『期待値からの偏差』という。

$$X - E(X) \tag{28}$$

1.6.1 『期待値からの偏差』の期待値

期待値からの偏差の期待値は、

$$E(X - E(X)) = E(X) - E(X) = 0 \tag{29}$$

1.6.2 確率変数の分散

「『期待値からの偏差』の二乗」の期待値を『分散』といい $V(X)$ であらわす。

$$V(X) = E\left((X - E(X))^2\right) \tag{30}$$

$$= E\left(X^2 - 2XE(X) + (E(X))^2\right) \tag{31}$$

$$= E\left(X^2\right) - 2E(X)E\left(X\right) + (E(X))^2 \tag{32}$$

$$= E\left(X^2\right) - 2E(X)^2 + (E(X))^2 \tag{33}$$

$$= E\left(X^2\right) - (E(X))^2 \tag{34}$$

1.6.3 数値例の分散

裏の出る確率 (p) が $p = 0.5$ であるコインを 8 回投げたときの裏の出る回数を X とすると、確率変数 X の二乗の期待値は

$$\begin{aligned} E(X^2) &= \frac{0}{256} + \frac{8}{256} + \frac{112}{256} \\ &+ \frac{504}{256} + \frac{1120}{256} + \frac{1400}{256} \\ &+ \frac{1008}{256} + \frac{392}{256} + \frac{64}{256} \qquad (35) \\ &= 18 \qquad (36) \end{aligned}$$

である。

表 6 X^2 の分布

X^2	$\Pr({x_i}^2)$
0	1/256
1	8/256
4	28/256
9	56/256
16	70/256
25	56/256
36	28/256
49	8/256
64	1/256

$E(X) = 4$ なので、確率変数 X の分散 $V(X)$ は

$$\begin{aligned} V(X) &= E\left(X^2\right) - (E(X))^2 && (37) \\ &= 18 - 4^2 && (38) \\ &= 2 && (39) \end{aligned}$$

1.6.4 分散の演算

a, c を定数とし、X を確率変数とすると、確率変数 (X) の分散に関して以下の関係が成り立つ。

$$V(c) = 0 \tag{40}$$

$$V(X + c) = V(X) \tag{41}$$

$$V(aX) = a^2 V(X) \tag{42}$$

証明

$$V(c) = E\left((c - E(c))^2\right) \tag{43}$$

$$= E\left((c - c)^2\right) \tag{44}$$

$$= E(0) \tag{45}$$

$$= 0 \tag{46}$$

$$
\begin{aligned}
V(X+c) &= E\left(\left(X+c-E(X+c)\right)^2\right) && (47)\\
&= E\left(\left(X+\cancel{c}-E(X)-\cancel{c}\right)^2\right) && (48)\\
&= E\left(\left(X-E(X)\right)^2\right) && (49)\\
&= V(X) && (50)
\end{aligned}
$$

$$
\begin{aligned}
V(aX) &= E\left(\left(aX-E(aX)\right)^2\right) && (51)\\
&= E\left(\left(aX-aE(X)\right)^2\right) && (52)\\
&= E\left(a^2\left(X-E(X)\right)^2\right) && (53)\\
&= a^2E\left(\left(X-E(X)\right)^2\right) && (54)\\
&= a^2V\left(x\right) && (55)
\end{aligned}
$$

1.6.5 確率変数の標準偏差

『分散』の正の平方根を『標準偏差』といい σ_X または $D(X)$ であらわす。

$$\sigma_X = D(X) = \sqrt{V(X)} = \sqrt{E\left((X - E(X))^2\right)} \tag{56}$$

数値例において

$$V(X) = 2 \tag{57}$$

なので、

$$D(X) = \sqrt{2} \tag{58}$$

1.7　確率変数の標準化

確率変数 X を、期待値 $E(X)$ と標準偏差 σ_X を用いて標準化する。

$$z_i = \frac{x_i - E(X)}{\sigma_X} \tag{59}$$

数値例の確率変数 X の期待値と標準偏差は

$$E(X) = 4 \tag{60}$$

$$D(X) = \sqrt{2} \tag{61}$$

である。

表 7　標準化された確率変数

X	Z	$\Pr(z_i)$
0	−2.8284271	1/256
1	−2.1213203	8/256
2	−1.4142136	28/256
3	−0.7071068	56/256
4	0.0000000	70/256
5	0.7071068	56/256
6	1.4142136	28/256
7	2.1213203	8/256
8	2.8284271	1/256

1.8　標準偏差の $\pm n$ 倍の範囲

数値例において、確率変数 Z が ± 1 の範囲すなわち

$$-1 \leqq z_i \leqq 1 \tag{62}$$

である確率は、

$$\frac{56}{256}+\frac{70}{256}+\frac{56}{256}=\frac{182}{256} \tag{63}$$

$$=0.7109\cdots \tag{64}$$

であって、

± 2 の範囲

$$-2 \leqq z_i \leqq 2 \tag{65}$$

である確率は、

$$\frac{28}{256}+\frac{56}{256}+\frac{70}{256}+\frac{56}{256}+\frac{28}{256} \tag{66}$$

$$=\frac{238}{256}=0.9296\cdots \tag{67}$$

である。

問題 IX–1 – 2

裏の出る確率 (p) が $p = 0.5$ であるコインを 16 回投げたときの裏の出る回数を X とする。

1. 確率変数 X の分布を求め、期待値, 分散, 標準偏差を求めなさい。
2. z_i を求めなさい。
3. $-1 \leqq z_i \leqq 1$ の範囲の確率を求めなさい。
4. $-2 \leqq z_i \leqq 2$ の範囲の確率を求めなさい。

表 8 コインを 16 回投げたときの確率分布

X	Z	$\Pr(z_i)$	X	Z	$\Pr(z_i)$
0	−4.0	1/65,536	9	0.5	11,440/65,536
1	−3.5	16/65,536	10	1.0	8,008/65,536
2	−3.0	120/65,536	11	1.5	4,368/65,536
3	−2.5	560/65,536	12	2.0	1,820/65,536
4	−2.0	1,820/65,536	13	2.5	560/65,536
5	−1.5	4,368/65,536	14	3.0	120/65,536
6	−1.0	8,008/65,536	15	3.5	16/65,536
7	−0.5	11,440/65,536	16	4.0	1/65,536
8	0.0	12,870/65,536			

1.9　まとめ

- 『平均からの偏差』の二乗の平均を『分散』という。
- 『分散』の正の平方根を『標準偏差』という。
- 『平均からの偏差』を『標準偏差』で割ることを『標準化』といい、得られた値を『z 値』という。
- 『標準化』することで、相対的な位置を容易に把握することができる。
- $-1 \leqq z_i \leqq 1$ は、期待値を中心とした標準偏差の ± 1 倍 の範囲であり、$-2 \leqq z_i \leqq 2$ は ± 2 倍 の範囲である。

1.9.1 参考文献

(1) 「行動科学における統計解析法」芝祐順　南風原朝和　著 『東京大学出版会』1990 年 3 月 20 日　初版

(2) 「完全独習　統計学入門」小島　寛之　著 『ダイヤモンド社』2006 年 9 月 28 日　初版