1.6　度数分布表の平均値と分散

度数分布表の平均値と分散とは？

まずは復習から。

n 個の**量的変数**の集合であるデータ

x_1，x_2，x_3，$\cdots$，x_n

に対して，次の計算で得られる**代表値**を，**平均値**といいます。

［平均値］　$$\mu = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{k=1}^{n} x_k$$

データの各要素と平均値との差を，**偏差**といい，

「偏差の２乗」の総和をデータ数 n で割ったものを**分散**といいます。

また，分散の平方根をとった値を，**標準偏差**といいます

［分散］　$$\sigma^2 = \frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + \cdots + (x_n - \mu)^2}{n} = \frac{1}{n}\sum_{k=1}^{n} (x_k - \mu)^2$$

次に，前回の課題のデータを利用しながら，説明します。

20 個のデータ：　43　47　52　53　54　61　66　67　68　69

70　71　72　73　76　78　82　83　84　91

40 以上 50 未満を階級の一つとした，階級の幅が 10 の度数分布表

階　級	階級値(x)	度数(f)
４０～５０	４５	2
５０～６０	５５	3
６０～７０	６５	5
７０～８０	７５	6
８０～９０	８５	3
９０～100	９５	1
総度数	―	２０

※度数分布表に整理された 20 個のデータは，次の様に，階級値を用いたデータに書き直されたものとして取り扱われます。

20 個のデータ：　45　45　55　55　55　65　65　65　65　65

75　75　75　75　75　75　85　85　85　95

このため，今回の度数分布表の**中央値**は　70　**最頻値**は　75　です。

このとき度数分布表の**平均値**は，次の様な計算になります。

$$\mu = \frac{45\times 2 + 55\times 3 + 65\times 5 + 75\times 6 + 85\times 2 + 95\times 1}{20}$$

$$= \frac{90 + 165 + 325 + 450 + 170 + 95}{20} = \frac{1380}{20} = 69$$

一般の場合について，平均値と分散の計算の方法をまとめておきます。

度数分布表に整理されたデータ

階級値(x)	x_1, x_2, x_3, $\cdots$, x_n	合計
度　数(f)	f_1, f_2, f_3, $\cdots$, f_n	N

（但し， $N = f_1 + f_2 + f_3 + \cdots + f_n$ ）

に対して，平均値と分散は，次式によって求められる。

［平均値］　$\mu = \dfrac{x_1 f_1 + x_2 f_2 + \cdots + x_n f_n}{N} = \dfrac{1}{N}\sum_{k=1}^{n} x_k f_k$

［分　散］　$\sigma^2 = \dfrac{(x_1-\mu)^2 f_1 + (x_2-\mu)^2 f_2 + \cdots + (x_n-\mu)^2 f_n}{N}$

$$= \frac{1}{N}\sum_{k=1}^{n}(x_k-\mu)^2 f_k$$

例題　次のように度数分布表に整理されたデータの平均値と分散を求めよ。

階級値(x)	度数(f)
４５	2
５５	3
６５	5
７５	6
８５	3
９５	1
合計	２０

［解法］度数分布表の右側に，表を付け加えて解きます。

［解答］

x	f	xf	$x-\overline{x}$	$(x-\overline{x})^2$	$(x-\overline{x})^2 f$
45	2	90	-24	576	1152
55	3	165	-14	196	588
65	5	325	-4	16	80
75	6	450	6	36	216
85	3	255	16	256	768
95	1	95	26	676	676
合計	$\boxed{20}$	$\boxed{1380}$	—	—	$\boxed{3480}$

度数分布表の　平均値　$\mu = \dfrac{1380}{20} = 69$　　分散　$\sigma^2 = \dfrac{3480}{20} = 174$

（※課題(直接計算)の結果：平均値　$\mu = 68$　　分散　$\sigma^2 = 161.1$）

1.7　２乗平均値

２乗平均値とは？

以下では，「度数分布表の」は省略します。
また，平均値を E，分散を V で表記します。

通常の平均値は

［平　均　値］　$\overline{x} = E(x) = \dfrac{x_1 f_1 + x_2 f_2 + \cdots + x_n f_n}{N} = \dfrac{1}{N}\sum_{k=1}^{n} x_k f_k$

（※「階級値(x)」と「度数(f)」の積の総和）

で計算します。次の計算式で求められる値を、**２乗平均値**と呼びます。

［２乗平均値］　$E(x^2) = \dfrac{x_1^2 f_1 + x_2^1 f_2 + \cdots + x_n^2 f_n}{N} = \dfrac{1}{N}\sum_{k=1}^{n} x_k^2 f_k$

（※「階級値(x)」の $\boxed{\text{２乗}}$ と「度数(f)」の積の総和）

分散は，次の計算式でも求めることができます。

［分散］　$\sigma^2 = V(x) = E(x^2) - \{E(x)\}^2$

（※「分散」は「２乗平均値」と「平均値の２乗」との差）

証明）$V(x) = \dfrac{1}{N}\displaystyle\sum_{k=1}^{n}(x_k - \mu)^2 f_k$

$$= \frac{1}{N}\sum_{k=1}^{n}(x_k^2 - 2\mu x_k + \mu^2) f_k$$

$$= \frac{1}{N}\sum_{k=1}^{n}x_k^2 f_k \; - 2\mu \times \frac{1}{N}\sum_{k=1}^{n}x_k f_k \; + \mu^2 \times \frac{1}{N}\sum_{k=1}^{n} f_k$$

※線形性により，定数(添え字　k　が付いていない文字)は
総和の記号の前に出すことができます。

$$= E(x^2) - 2\mu \times \mu + \mu^2 \times \frac{N}{N}$$

２乗平均値　$E(x^2) = \dfrac{1}{N}\displaystyle\sum_{k=1}^{n}x_k^2 f_k$　　平均値　$E(x) = \dfrac{1}{N}\displaystyle\sum_{k=1}^{n}x_k f_k$

総度数(度数の和)　$N = \displaystyle\sum_{k=1}^{n} f_k = f_1 + f_2 + \cdots + f_n$

$$= E(x^2) - 2\mu^2 + \mu^2 = E(x^2) - \mu^2 = E(x^2) - \{E(x)\}^2$$

例題　次のように度数分布表に整理されたデータの平均値と分散を求めよ。

階級値(x)	度数(f)
４５	2
５５	3
６５	5
７５	6
８５	3
９５	1
合計	２０

［解法］度数分布表の右側に，表を付け加えて解きます。

［解答］

x	f	xf	$x^2 f$
45	2	90	$45^2 \times 2 = 4050$
55	3	165	$55^2 \times 2 = 9075$
65	5	325	$65^2 \times 2 = 21125$
75	6	450	$75^2 \times 2 = 33750$
85	3	255	$85^2 \times 2 = 21675$
95	1	95	$95^2 \times 2 = 9025$
合計	20	1380	98700

平均値　$\mu = E(x) = \dfrac{1380}{20} = 69$　　２乗平均値　$E(x^2) = \dfrac{98700}{20} = 4935$

分　散　$\sigma^2 = V(x) = E(x^2) - \{E(x)\}^2 = 4935 - 69^2 = 4935 - 4761 = 174$

問 3. 7　次の表は，ある試験の得点を，階級値の１つを１点，階級の幅を２点として整理した度数分布表である。次の問いに答えよ。

階級値 x	1	3	5	7	9	合計
人　数 f	1	4	16	12	7	40

(1) 中央値と最頻値を求めよ。

(2) 下の表を完成して，平均値と分散を求めよ

x	f	xf	$x^2 f$
1	1		
3	4		
5	16		
7	12		
9	7		
合計	40		

1.8　変数変換

変数変換とは？

○変数 x を，変換式 $y=\varphi(x)$ で新しい変数を作ることを**変数変換**という。
○変数 x の平均を $E(x)$，分散を $V(x)$ とし，$y=ax+b$ で変換したとき
　変数 y の平均値 $E(y)$，分散 $V(y)$ の間には次の関係式が成り立つ。

［変数変換］　$y=ax+b$　のとき

　平均値　$E(y)=a\ E(x)+b$

　分　散　$V(y)=a^2\ V(x)$　（※b は寄与しない）

証明）
$$E(y)=\frac{1}{N}\sum_{k=1}^{n}y_k f_k=\frac{1}{N}\sum_{k=1}^{n}(ax_k+b)f_k$$
$$=a\times\boxed{\frac{1}{N}\sum_{k=1}^{n}x_k f_k}+b\times\frac{1}{N}\boxed{\sum_{k=1}^{n}f_k}$$
$$=a\times\boxed{E(x)}+b\times\frac{\boxed{N}}{N}=aE(x)+b$$

$$E(y^2)=\frac{1}{N}\sum_{k=1}^{n}y_k^2 f_k=\frac{1}{N}\sum_{k=1}^{n}(ax_k+b)^2 f_k=\frac{1}{N}\sum_{k=1}^{n}(a^2x_k^2+2abx_k+b^2)f_k$$
$$=a^2\times\frac{1}{N}\sum_{k=1}^{n}x_k^2 f_k+2ab\times\frac{1}{N}\sum_{k=1}^{n}x_k f_k+b^2\times\frac{1}{N}\sum_{k=1}^{n}f_k$$
$$=a^2E(x^2)+2abE(x)+b^2$$
$$V(y)=E(y^2)-\{E(y)\}^2$$
$$=a^2E(x^2)+2abE(x)+b^2-\{\ aE(x)+b\ \}^2$$
$$=a^2E(x^2)+2abE(x)+b^2-a^2\{E(x)\}^2-2abE(x)-b^2$$
$$=a^2[E(x^2)-\{E(x)\}^2]=a^2\ V(x)$$

例題　次のように度数分布表に整理されたデータの平均値と分散を，変数変換

$$y=\frac{x-65}{10}\left(=\frac{1}{10}x-\frac{13}{2}\right)$$

を用いて求めよ。

階級値(x)	度数(f)
４５	2
５５	3
６５	5
７５	6
８５	3
９５	1
合計	２０

［解答］

x	y	f	yf	y^2f
45	-2	2	-4	8
55	-1	3	-3	3
65	0	5	0	0
75	1	6	6	6
85	2	3	6	12
95	3	1	3	9
合計		20	8	38

変数 y については

$$E(y)=\frac{8}{20}=\frac{2}{5}\qquad E(y^2)=\frac{38}{20}=\frac{19}{10}\qquad V(y)=\frac{19}{10}-\left(\frac{2}{5}\right)^2=\frac{19}{10}-\frac{4}{25}=\frac{174}{100}$$

このとき，変数 x については

$$\frac{1}{10}E(x)-\frac{13}{2}=E(y)=\frac{2}{5}\ \text{より}\ \frac{1}{10}E(x)=\frac{2}{5}+\frac{13}{2}=\frac{69}{10}\qquad \therefore E(x)=69$$

$$\left(\frac{1}{10}\right)^2V(x)=V(y)=\frac{174}{100}\ \text{より}\ V(x)=\frac{174}{100}\times 100=174$$

問 3.8　次のように度数分布表に整理されたデータの平均値と分散を

変数変換　$y=\dfrac{x-25}{10}$　を用いて求めよ。

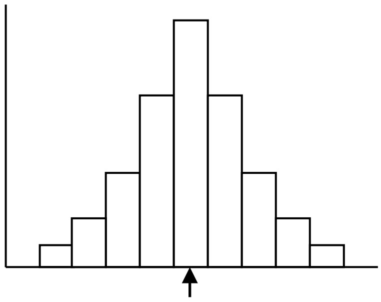
x	5	15	25	35	45	計
f	1	8	20	12	9	50

□□□□【参考】□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

度数分布表の代表値についてお話します。

平均値，中央値，最頻値ですね。

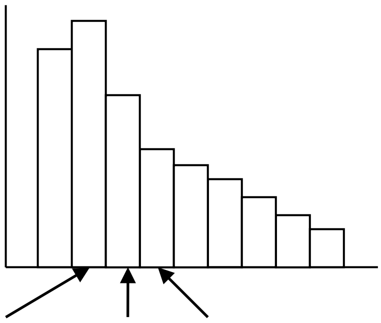
右図のように，度数分布表の形が
左右対称で，中央に盛り上がった
山のような形状に近い場合は
平均値≒中央値≒最頻値
のような関係になります。



平均値≒中央値≒最頻値

しかし，中央が左[右]によって，
右[左]側のすそ野が広がるときは
最頻値＜中央値＜平均値
[平均値＜中央値＜最頻値]
のような関係になります。

※通常よく起きる(最頻値)から，
平均的(平均値)とは限らない！



最頻値＜中央値＜平均値

$\boxed{\text{重要}}$ 度数分布表の形状は様々です。データの代表値として，平均値，中央値，最頻値のどれがふさわしいのかは，データや代表値が持っている特色をよく考えて判断します。

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□