統計資料の見方（目次）

出題傾向（統計資料の見方）

１．統計資料の見方は、なぜ出題されるか。

「都職員として身につけるべき基礎的知識を検証するための科目とする。」

（主任選考の見直しについて　平成 17 年 12 月 22 日人事委員会）

２．出題内容について

○ 出題構成見直しの概要

（平成 29 年度主任級職選考の出題構成等の見直しについて　平成 28 年 9 月 12 日人事委員会）

・ 採用試験との重複を解消するため、「統計資料の見方」の出題数を削減

・ 資料解釈を廃止し、統計データの分析（※）から出題

※統計データの分析：職員ハンドブック 2023（第Ⅲ編第９章第３節）

○ 出題数

ＡⅠ類（事務）　２／５５問　（２時間４５分）

ＡⅠ類（４大技術）　２／４５問　（２時間１５分）

ＡⅡ類　２／３０問　（１時間３０分）

※ 出題数を時間で除すると 1 問を 3 分で解答することになります。

○ 出題内訳

出題内容	28 年度まで	**29 年度以降**	解法
資料解釈	4	**0**	表やグラフから読み取って計算する問題
統計知識	1	**2**	公式を用いて計算する問題
計	5	**2**	

３．統計知識について

○ 出題範囲

これまでも職員ハンドブック第Ⅲ編第９章の「統計」の第３節「統計データの分析」からの出題が主となっている。→下記の４パターンに分類できる。

① 代表値…算術平均、中央値、最頻値

② 分散度…分散、標準偏差、偏差値

③ 統計比…構成比、変化率、寄与度・寄与率、指数

④ 回帰分析…回帰方程式、最小二乗法、相関係数

（参考）過年度の出題内容　※平成１４年度までは２題、１５年度以降は１題出題。グレーの網掛けは第３節以外からの出題。

年度	出題分野	出題内容
9	統計データのまとめ方（第２節）、②分散度	統計用語（知識問題）、標準偏差（正規分布・ポワソン分布の性質）（知識問題）
10	①代表値、④回帰分析	算術平均・中央値・最頻値（計算問題）、回帰方程式（知識問題）
11	③統計比、④回帰分析	寄与率・寄与度（計算問題）、散布図（相関係数）（知識問題）
12	統計データのまとめ方（第２節）、②分散度	統計データの分析結果（知識問題）、偏差値（計算問題）
13	統計の役割（第１節）、①代表値・③統計比	統計と行政（知識問題）、用語問題（知識問題）
14	②分散度、④回帰分析	分散（計算問題）、回帰方程式（知識問題）
15	②分散度	標準偏差（正規分布における性質）（知識問題）
16	④回帰分析	散布図（相関係数）（知識問題）
17	①代表値	中央値（計算問題）
18	③統計比	寄与度（計算問題）
19	②分散度	標準偏差（正規分布における性質）（知識問題）
20	②分散度	分散（計算問題）
21	①代表値・②分散度	算術平均・中央値・最頻値・分散・標準偏差（計算問題）
22	統計データのまとめ方（第２節）	グラフの種類と用途（知識問題）
23	②分散度	標準偏差（正規分布における性質）（知識問題）
24	①代表値	中央値（計算問題）
25	②分散度	分散の性質（計算問題）
26	③統計比	寄与率（計算問題）
27	②分散度	偏差値（計算問題）
28	③統計比	指数（ラスパイレス指数）（計算問題）
29	②分散度、③統計比	分散（計算問題）、寄与率（計算問題）
30	統計に関する知識問題、③統計比	統計用語等（知識問題）、寄与度（計算問題）
R1	④回帰分析、①代表値	相関係数（知識問題）、中央値（計算問題）
R2	統計に関する知識問題、②分散度	統計用語等（知識問題）、偏差値（計算問題）
R3	統計データのまとめ方（第２節）、③統計比	グラフの種類と用途（知識問題）、指数（ラスパイレス指数）
R４	統計に関する知識、②分散度	用語（知識問題）、分散（計算問題）

○　出題パターン

・　実際に値を算出する問題もしくは用語の性質等を問う問題が出題される。
資料解釈と異なり、**解き方を知らないと正答できないため、暗記を要する。**

・　①代表値：中央値を算出する計算問題 or 算術平均・中央値・最頻値を算出するミックスの計算問題の出題

・　②分散度：分散を算出する計算問題、偏差値を算出する計算問題、正規分布における標準偏差の性質に関する知識問題の出題

・　③統計比：寄与率・寄与度を算出する問題が主。指数（ラスパイレス指数）は平成28・令和３年度出題。（同分野内の構成比、変化率はこれまで資料解釈で問われていた内容のため、統計知識での取り扱いについては不明だが、平成 29 年度は寄与率、平成 30 年度は寄与度の算出が出題された。）

・　④回帰分析：回帰方程式、相関係数（散布図）に係る知識問題の出題。計算問題はここ 20 年は出題なし。

４．資料解釈について（参考）　<u>※29 年度からは出題なし</u>

○ 主な出題範囲

・ 表またはグラフによって示された統計資料から、正しい情報を導き出す。

・ 表やグラフは、以下の３パターンに絞られる。

①「増加率」で表示されるもの
②「構成比」で表示されるもの
③「指　数」で表示されるもの

○ 出題パターン

選択肢の問いは、表やグラフ内に直接示されていない「実数」「増加率」「構成比」に関するものや、「実数」「増加率」「構成比」を複合させたものについて大小比較をおこなって、正しい選択肢を選ぶというのが出題パターンとなっている。

５．解法のポイント

（１）最後に解く。

先に知識問題を解いて、貯金した時間で統計資料の見方（統計データの分析）を解く。

	試験時間（分）	知識問題			統計資料		
		問題数（問）	1問あたり（分）	合計（分）	問題数（問）	1問あたり（分）	合計（分）
AⅠ類（事務）	165	53	2	106	2	29.5	59
AⅠ類（四技）	135	48	2	96	2	19.5	39
AⅡ類	90	33	2	66	2	12	24

（２）正解肢が見つかったら、残りの選択肢を解くのをとりあえずはやめる。

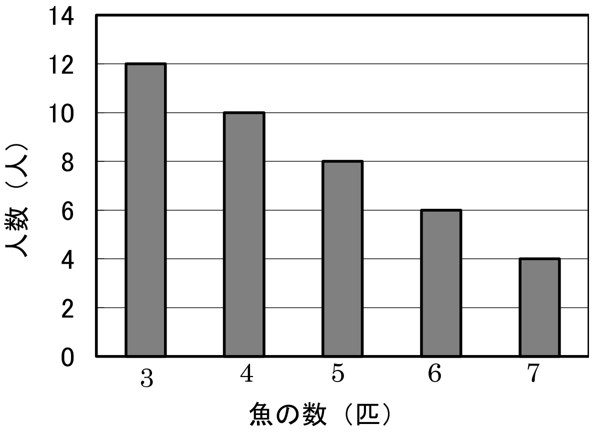
すべて解答してから、残りの選択肢の検算を行う。

６．参考図書

・ 職員ハンドブック 2023　〔第９章統計（Ｐ617～Ｐ641)〕
・ 本研修資料
・ 東京都主任試験「統計データ分析」択一問題集（公人の友社）
・ 統計について学ぼう（厚生労働省ホームページ）
https://www.mhlw.go.jp/toukei/learning/index.html
※中学生、小学生向けの統計学習サイトがまとめられている。

【代表値】

資料１．釣り大会成績表



度数

級

資料 2. 収入別分布表

収入額（万円）	人数 f（人）	中央値 x（万円）	fx（万円）
10 未満	2	5	10
10～20 〃	15	15	225
20～30 〃	60	25	1500
30～40 〃	90	35	3150
40～50 〃	55	45	2475
50～60 〃	11	55	605
60～70 〃	1	65	65
合計	234	−	8030

① 算術平均　　H21 出題

（職員ハンドブック 2021　P630「算術平均」参照）

総和を項数で割った、いわゆる平均値のこと。

$$\bar{x} = \frac{\sum fx}{\sum f}$$

【資料１の場合】

$$\frac{(12\times3)+(10\times4)+(8\times5)+(6\times6)+(4\times7)}{40} = 4.5\text{（匹）}$$

【資料２の場合】

級に幅があるので、級の真ん中の値をその級の平均値と推定して計算する。

$$\frac{(2\times5)+(15\times15)+(60\times25)+(90\times35)+(55\times45)+(11\times55)+(1\times65)}{234} = 34.3\text{（万円）}$$

※極端にかけ離れた数値があるとそれに引きずられてしまうので、算術平均を使用する時は、集団のばらつきに注意しなければならない。

② 中央値（メディアン（中位数））　　H17,21,24,R1 出題

（職員ハンドブック 2021　P631「中央値」参照）

級の値を小さい順（または大きい順）に並べた時の中央に位置している値のこと。

人数が奇数であれば、中央に位置する人は１人なので、その値が中央値。

人数が偶数であれば、中央に位置する人は２人いるので、その２人の平均値が中央値。

（例）７人の時⇒４番目の人の値、８人の時⇒４番目と５番目の値の平均

【資料１の場合】
参加者が 40 名なので、20 番目と 21 番目の人が釣った魚の数を見る。
すると、3 匹釣ったのは下位 12 名で、4 匹まで含めると 22 名いる。
したがって、20 番目と 21 番目の人は、共に 4 匹釣ったといえる。
（中央値なので、上位から数えても同じ結果になる）

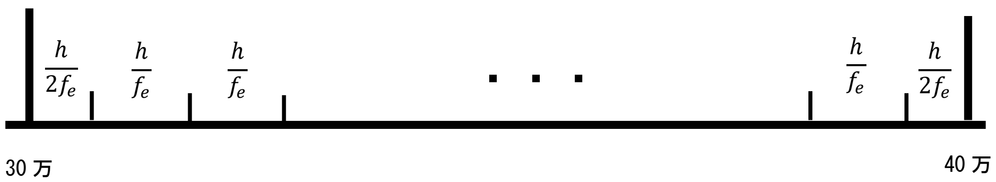
【資料２の場合】

中央値（Me）＝（中央値を含む階級の下限）＋ $\left(\frac{\sum f}{2} - Ne\right) \times \frac{h}{fe}$

$\sum f$：総度数、Ne：中央値より低い階級の度数合計
fe：中央値の入っている階級の度数、h：級間隔

（１）　中央値の属する階級を求める。

（２）　下図のように、中央値を含む階級において、中央値を含む階級の度数が一様に分布していると仮定する。



（３）　収入額を少ない順番に並べたとき、中央値が何番目になるか考えると、（２）の場合は全体で$\frac{\sum f+1}{2}$番目の値を中央値とできることがわかる。

（４）　以上の（１）～（３）から、中央値 Me は以下のように求められる。

$$\mathrm{Me} = (\text{中央値を含む階級の下限}) + \frac{h}{2f_e} + \frac{h}{f_e}\underbrace{\left(\frac{\sum f+1}{2} - Ne - 1\right)}_{\text{中央値を含む階級の下限から数えて、間隔が}\frac{h}{f_e}\text{の区間の数}}$$

（５）　（４）の式を整理すると、冒頭の式

$$\mathrm{Me} = (\text{中央値を含む階級の下限}) + \left(\frac{\sum f}{2} - Ne\right) \times \frac{h}{fe}$$

が得られる。

（６）　この式に、(中央値を含む階級の下限)=30, 総度数$\sum f$=234, 中央値より低い階級の度数合計Ne=77, 中央値を含む階級の度数f_e=90, 級間隔h=10 を代入すると 34.4 を得る。

③ **最頻値（モード）**　　H17,21,30 出題

（職員ハンドブック 2021　P631「最頻値」参照）

一番多くの値が属する変数を表す。（分布の峰）

【資料１の場合】

釣った魚の数の値の一番多くが属するところなので３匹。

【資料２の場合】

最頻値（Mo）＝（最頻値を含む階級の下限）$+\frac{f\beta}{f\alpha+f\beta}\times h$ $f\alpha$：最頻値を含む階級の前の階級の度数、$f\beta$：最頻値を含む階級の後の階級の度数 h：級間隔

中央値の場合と同様、求めたい値（最頻値）の属する階級における分布について仮定して求める必要がある。このような場合において、最頻値を含む階級をその両端の階級の度数で比例分割する方法を用いる（職員ハンドブック 2021　P631）。

$$\mathrm{Mo}=30+\frac{55}{60+55}\times 10=34.8\text{ 万円}$$

○ 以上から、**算術平均**は最も一般的であるが、極端な変数の影響を避けたいときは**中央値**を、典型的な値を求めたいときは**最頻値**を選ぶ。

【分散度】

①　分散と標準偏差　　分散 H20, 25, 29, 30, R4 出題、標準偏差 H15, 19, 23 出題

（職員ハンドブック 2021　P632「分散度」参照）

> データが平均値を中心にどの程度散らばっているのかを示す値
>
> $$分散(\sigma^2) = \frac{\sum f\left(変数^{*} - 算術平均\right)^2}{総度数} = \frac{\sum f(x-\bar{x})^2}{\sum f}【①】 = \frac{\sum fx^2}{\sum f} - \bar{x}^2【②】$$
>
> σ^2：分散、x：変数*、$\bar{x}$：算術平均、f：度数
> ※この場合の変数は、各級の中央値
>
> 標準偏差（σ）＝分散の平方根 $= \sqrt{\sigma^2}$

変数

収入額（万円）	度数 f（人）	中央値 x（万円）	fx（万円）	fx^2（万円）
10〜20	4	15	60	900
20〜30	9	25	225	5,625
30〜40	25	35	875	30,625
40〜50	18	45	810	36,450
50〜60	8	55	440	24,200
60〜70	6	65	390	25,350
合計	70	−	2,800	123,150

右表のように、fx^2の値が出ていれば、②の方程式に数値を入れて計算をする。

また、①の方程式を使った解法としては、平均値からずれた位置にどの程度の数が存在するかを計算すればよいので、

・　算術平均（$\bar{x}$）＝ $\frac{2800}{70}$ ＝40

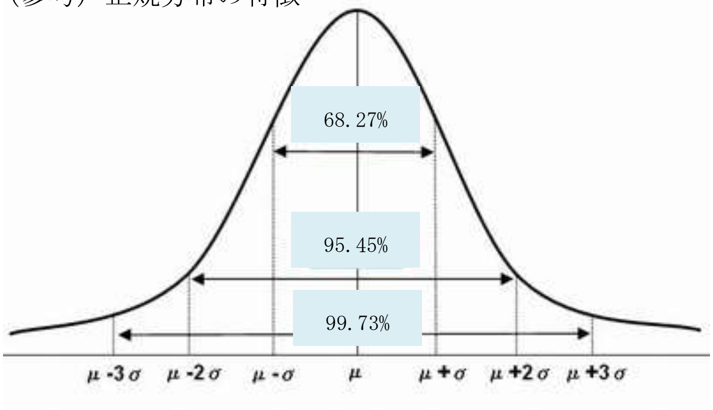
・①による計算

$$\frac{4\times(15-40)^2 + 9\times(25-40)^2 + 25\times(35-40)^2 \cdots + 6\times(65-40)^2}{70}$$

●標準偏差は<u>分散の平方根</u>で表され、<u>平均値</u>からの散らばり具合が分かる。
正規分布の場合、平均値：μ、標準偏差：σ とすると

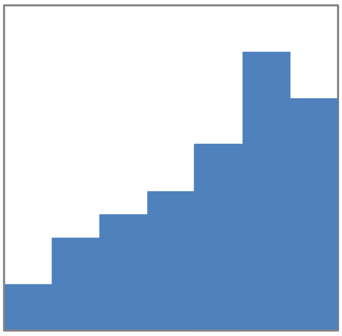
・$\mu \pm \sigma$　の範囲に全体の **68.27%** のデータが含まれる。
・$\mu \pm 2\sigma$ の範囲に全体の **95.45%** のデータが含まれる。
・$\mu \pm 3\sigma$ の範囲に全体の **99.73%** のデータが含まれる。

（参考）正規分布の特徴



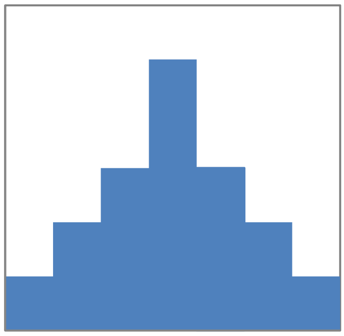
●ヒストグラムと平均値・中央値・最頻値の関係

左に裾が長い分布



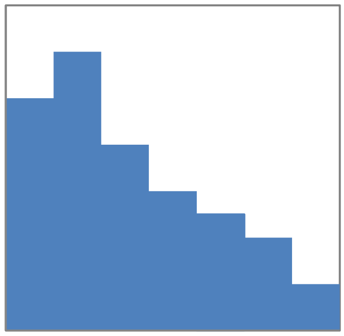
平均値<中央値<最頻値

左右対称の分布



平均値=中央値=最頻値

右に裾が長い分布



最頻値<中央値<平均値

② **偏差値**　　H27，R2 出題

（職員ハンドブック 2021　P634「分散度」参照）

測定値の平均と標準偏差を用いて、平均が 50、標準偏差が 10 になるように変換して求める値

$$偏差値(T) = 50 + 10 \times \frac{x - \bar{x}}{\sigma}$$

x：各人の点数、$\bar{x}$：平均点、σ：標準偏差

（例）あるクラスの英語の試験で A 君の点数が 90 点、クラスの平均点は 75 点、標準偏差は 12 点であったとすれば、A 君の偏差値は、

$$T = 50 + 10 \times \frac{90-75}{12} = 62.5$$

【統計比】

①　寄与度　　H18, 30 出題

（職員ハンドブック 2021　P636「寄与度」参照）

各項目の変化が全体をどの程度の割合で変化させたかを示す。

$$寄与度（\%）=\frac{内訳項目の変化数}{総数の期首の値}\times 100$$

②　寄与率　　H26, 29 出題

（職員ハンドブック 2021　P636「寄与率」参照）

各項目の変化が、全体の変化にどの程度の影響を与えたかを示す。

$$寄与率（\%）=\frac{内訳項目の変化数}{総数の変化数}\times 100$$

※　寄与度と寄与率を混乱しないように注意（寄与率＝変化数の構成比）。
　　各項目の寄与率を合計すると 100 になる。

〇2009 年に対する 2010 年の商品 A の増加寄与度

$$\frac{110-100}{500}\times 100=2\%$$

〇2009 年に対する 2010 年の商品 A の増加寄与率

$$\frac{110-100}{550-500}\times 100=20\%$$

	2009 年	2010 年
商品 A	100	110
商品 B	400	440
合計	500	550

③　指数　　H28, R3 出題

（職員ハンドブック 2021　P636「指数」参照）

同一現象の時間（場所）的な変化を、ある時点（場所）を基準として相対的に示す。

〇指数＝（比較量／基準量）×　100

〇物価指数（総合指数）

ラスパイレス式（%）：$\frac{\sum P_t\ Q_o}{\sum P_o\ Q_o}\times 100$

→基準時の数量を固定

パーシェ式（%）：$\frac{\sum P_t\ Q_t}{\sum P_o\ Q_t}\times 100$

→比較時の数量を固定

P(Price)：価格
Q(quantity)：数量

0：基準時
t：比較時

○ラスパイレス式（基準時固定ウェイト）

$$\frac{\mathbf{300} \times 10 + \mathbf{1200} \times 10}{500 \times 10 + 1000 \times 10} \times 100$$

=100（%）

	2007 年（基準）		2017 年（比較）	
	価格	購入量	価格	購入量
商品 A	500	10	300	20
商品 B	1000	10	1200	5

○パーシェ式（比較時固定ウェイト）

$$\frac{300 \times 20 + 1200 \times 5}{\mathbf{500} \times 20 + \mathbf{1000} \times 5} \times 100$$

=80（%）

※ラスパイレス式は比較時の価格のみ調べれば済むため、速報性、低コスト等のメリットあり。パーシェ式は比較時の数量及び価格を調べる必要があるが、ラスパイレス式よりも正確性に優れている。

【回帰分析】（職員ハンドブック 2021　P638「回帰分析」参照）

結果となる数値と要因となる数値の関係を調べて、それぞれの関係を明らかにする統計的手法。事象の予測・シミュレーション、検証、要因分析などを行うときに用いられる。

①　回帰方程式

変数間の関係を示す方程式のこと。

○２変数 x と y の間に、変数 x を原因（**独立変数**）、y を結果（**従属変数***）とする関係があれば、y ＝ f（x）で示すことができる。※従属変数*：説明変数とも呼ぶ

○規則的な変化が起こる場合、$\boxed{y=a+bx}$の一次方程式で表された回帰方程式が当てはまる関係が多い。

○係数 a, b を「**回帰母数**」と呼び、特に b を「**回帰係数**」という。

○単回帰分析：独立変数 x が 1 つの回帰分析→　y＝a+bx
　重回帰分析：独立変数 x が２つ以上の回帰分析→　y＝a+b1x+ b2x･･･

②　最小二乗法

適当なモデルから想定される 1 次関数など特定の関数を用いて近似するときに、想定する関数が測定値に対してよい近似となるように、<u>**残差の二乗和を最小とするような係数を決定する方法**</u>、あるいはそのような方法によって近似を行うこと…

→**与えられたデータの関係を表す<u>もっともらしい直線を求める方法</u>**

（回帰線の当てはめ・フィッティング）

（参考）

$d = y - \hat{y}$（残差）の二乗和が最小になるように直線の位置を設定。

$\hat{y} = a + bx$（a：切片、b：傾き）とおく。

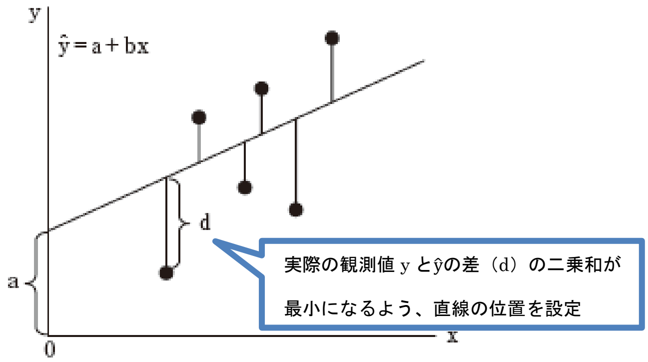
回帰直線からの偏差の二乗和（平方和）D が最小になるように a、b を求めるべく、
式を展開。

$\mathrm{D} = \sum(y - \hat{y})^2 \rightarrow \mathrm{D} = \sum(y - a - bx)^2$

⇩

$$\mathrm{b} = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sum(x - \bar{x})^2}$$

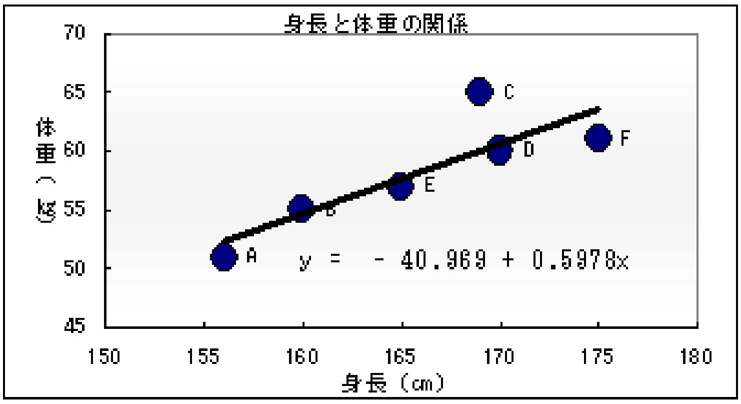
$$\mathrm{a} = \bar{y} - b\bar{x}$$



（例）

次のグラフは、あるクラスの男子（A～F）の身長と体重の関係を表したものです。

一般的に身長と体重は比例しますが、このグラフからも身長と体重の関係について右上がりの直線のイメージが想像されます。



この例で体重と身長を表す式が回帰方程式で、身長を横軸、体重を縦軸とすると y=a+bx（この場合は y＝- 40.969＋0.5978x）とあらわせます。

このように散布図を描いて、回帰方程式を求め、独立変数 x に数値を代入することで、生徒の身長から体重を予測したり、C さんは身長に対して体重がやや重めであると検証したり、さまざまな分析を行うことが可能となります。

③ 相関係数と決定係数　H16, 30, R1 出題

相関係数r：変数間の関係の方向と強さを示す尺度。

-1 から＋1 までの値をとる。（-1≦r≦1）

絶対値が 1 に近いほど相関関係が強く（＋1 に近いときが「正」の相関関係、－1 に近いときが「負」の相関関係。）、0 に近いほど相関関係が弱い（分散的な場合は弱い。）。

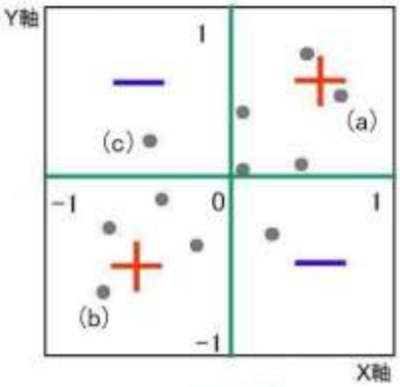
決定係数r^2：求めた回帰直線の実際のデータへの当てはまりの良さを図る尺度。
1 に近いほど当てはまりがよく、従属変数の変動のうち、正相関、負相関にかかわらず、独立変数との回帰によって説明される部分の割合を表す尺度

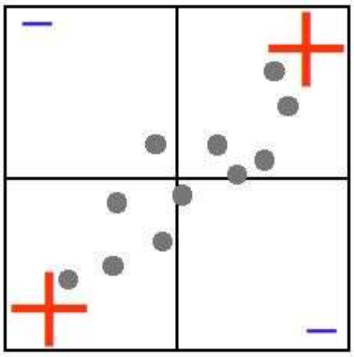
（参考）相関係数・決定係数の計算式

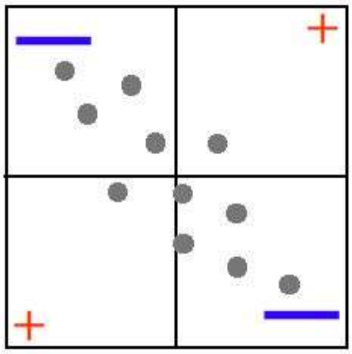
$$\mathrm{r} = \frac{\sum(x-\bar{x})\,(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2\sum(y-\bar{y})^2}}$$

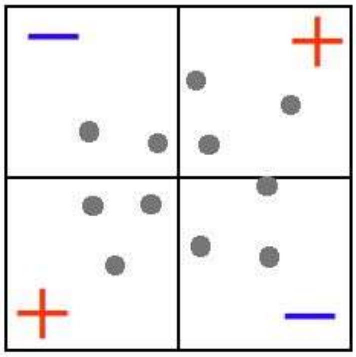
$$r^2 = 1 - \frac{\sum(y-\hat{y})^2}{\sum(y-\bar{y})^2}$$

x, y：観察値
$\bar{x}$：xの算術平均
$\bar{y}$：yの算術平均
$\hat{y}$：理論値
r：相関係数
r^2：決定係数




正の相関


負の相関


相関ゼロ=無相関

参　考

【増加率】（職員ハンドブック 2021　P 635「変化率」参照）

期首の値*（期間の始めの値）に対して、どれだけ増えた（減った）のかを示すもの。

$$変化率（\%）=\frac{期間中の増減数}{期首の値^{*}}\times 100$$

【例題１】下表の場合、2002 年の対前年増加率は？

	2001 年	2002 年
商品 A（個）	100	120
商品 B（個）	100	100
商品 C（個）	100	90

〈正解〉

・商品 A：$\frac{120-100}{100}\times 100=20\%$

・商品 B：$\frac{100-100}{100}\times 100=0\%$

・商品 C：$\frac{90-100}{100}\times 100=-10\%$

● 増加率が＋であれば、数値は元の値　より増える。
増加率が－であれば、数値は元の値　より減る。
増加率が０であれば、数値は元の値　と変わらない。

	2001年	増減	2002年	増加率
A	100個	<	120個	20%
B	100個	>	90個	-10%
C	100個	=	100個	0%

● 増加数が同じ場合
元の値の小さい方が増加率は大きい。
元の値の大きい方が増加率は小さい。

	2001年	増加数	2002年	増加率
A	50個	+25個	75個	50%
B	75個	+25個	100個	33%

● 増加率が同じ場合
元の値の大きいほうが増加数は大きい。
元の値の小さいほうが増加数は小さい。

	2001年	増減	2002年	増加率
A	50個	+25個	75個	50%
B	100個	+50個	150個	50%

● 複数にわたる累積の増加率を求める場合、増加率の和で近似できる。
（例）右表において、H16 に対する H19 の増加率の計算は、

⇩

2＋（-1）＋3＝　4（%）

（H16 の“－１%”は H15 に対する増加率なので足さない）

※　なお、増加率が 10%を超えるような大きい時は、増加率の和で近似したこの概算は成り立たないので注意。

	H16	H17	H18	H19
増加率（%）	-1	2	-1	3

【構成比】（職員ハンドブック 2021　Ｐ634「構成比」参照）

全体に占める個々の内訳の割合を示すもの。

$$項目Aの構成比（\%）=\frac{項目Aの値}{全体(合計)の値}\times 100$$

【例題２】下表の商品 A の、2001 年と 2002 年の構成比は？

	2001 年	2002 年
商品 A（個）	150	150
：	：	：
合計	300	500

〈正解〉

2001 年は 50%($=\frac{150}{300}\times 100$)、2002 年は 30% ($=\frac{150}{500}\times 100$)

● 項目の値が同じ場合

合計の値が小さい方が構成比は大きい。

合計の値が大きい方が構成比は小さい。

	A		合計	構成比
2001年	50個	…	100個	50%
2002年	50個	…	150個	33%

● 構成比が同じ場合

合計の値が大きい方が項目の値は大きい。

合計の値が小さい方が項目の値は小さい。

	A		合計	構成比
2001年	50個	…	100個	50%
2002年	75個	…	150個	50%

構成比と増加率の組み合わせ

構成比と増加率を合わせた問題は頻出。

	61 年	62 年	63 年
商品 A（個）	600	800	1200
：	：	：	：
合計	4500	4000	4800

【例題３】

右表から 62 年と 63 年の A の構成比はそれぞれ前年に比べて増加したか？

〈正解〉

<u>61 年と 62 年を比較した場合</u>

合計は減少している中、商品 A は増加しているので、計算するまでもなく、A の構成比は増加している。

<u>62 年と 63 年を比較した場合</u>

構成比を計算すると、

62 年の A の構成比：$\frac{800}{4000}\times 100=20\%$

63 年の A の構成比：$\frac{1200}{4800}\times 100=25\%$

一方で、増加率を比較すると

63 年の A の増加率　：$\frac{400}{800} \times 100 = 50\%$

63 年の合計の増加率：$\frac{800}{4000} \times 100 = 20\%$

- 項目 A 増加率＞全体の増加率⇒構成比は**増加**
  項目 A 増加率＜全体の増加率⇒構成比は**減少**
  項目 A 増加率＝全体の増加率⇒構成比は**同じ**

	62年	増加率	63年	増加率	64年
商品A	800個	+50%	1200個	+20%	1440個
⋮	⋮	−	⋮	−	⋮
合計	4000個	+20%	4800個	+50%	7200個
構成比	20%	＜	25%	＞	20%