

健康統計学 第6回

今回は、前回残した順列・組み合わせ（テキスト45～52ページ）について学習します。

また、これまでのまとめとして、表計算ソフト（Excel）を使った、記述統計（ヒストグラム、代表値、散布度、相関、回帰）の処理の実習をします。

テキスト

- 『やさしい保健統計学 改訂第4版』 縣 俊彦著（南江堂）

次回の連絡

- 次回（第7回）は、第2回から今回までの範囲について、中間試験を実施します。
 - 教室は、いつもと同じです。
 - 試験時間は、60分間です。試験時間終了後に、問題の解説をします。
 - 持ち込みは、電卓（関数電卓、携帯電話などの電卓機能など）以外は**すべて可**です。教科書でもメモでも、このページを印刷したものを持ち込んで構いません。

今回の内容

1. [順列と組み合わせ](#)
2. [Excelで度数分布図を作成](#)
3. [Excelで代表値と散布度を計算](#)
4. [Excelで相関係数と回帰直線を計算](#)
5. [Excelで散布図と回帰直線を作成](#)
6. [Excelで順列と組み合わせを計算](#)

順列と組み合わせ

階乗 (factorial)

- ある数 n から1ずつ少ない数を掛けあわせることを「**階乗**」という

$$n! = n \times (n-1) \times (n-2) \times \dots \times 2 \times 1$$

- なお、0(ゼロ)の階乗は、便宜上、1である。

順列 (Permutation)

- 異なる n 個のものから r 個を選んだ「**並べ方**」を、 n 個から r 個をとる「**順列**」という

$$\begin{aligned} {}_n P_r &= n \times (n-1) \times (n-2) \times \dots \times (n-r+1) \\ &= \frac{n!}{(n-r)!} \end{aligned}$$

- n 個から n 個をとる順列

$${}_n P_n = n \times (n-1) \times (n-2) \times \dots \times 2 \times 1 = n!$$

- 7 個から 3 個をとる順列

$${}_7 P_3 = 7 \times 6 \times 5 = 210$$

- また次のようなことがいえる (n 個から 0 個をとる順列)

$${}_n P_0 = 1$$

組み合わせ (Combination)

- 異なる n 個のものから r 個を選ぶときの組み合わせを、 n 個から r 個をとる「**組み合わせ**」という

$$\begin{aligned} {}_n C_r &= \frac{{}_n P_r}{r!} \\ &= \frac{n \times (n-1) \times \dots \times (n-r+1)}{r!} \\ &= \frac{n!}{r!(n-r)!} \end{aligned}$$

- 4 個から 3 個をとる組み合わせ

$${}_4 C_3 = \frac{4 \times 3 \times 2}{1 \times 2 \times 3} = 4$$

- なぜ ${}_n C_r = {}_n P_r / r!$ となるか?

◦ $\{a, b, c, d, e\}$ の5つから $\{a, b, c\}$ の3つを選ぶ場合を考えると...

- 順列は次の $3! = 6$ 通りとなる
 $\{a, b, c\}, \{a, c, b\}, \{b, a, c\}, \{b, c, a\}, \{c, a, b\}, \{c, b, a\}$
- 組み合わせでは順序を考えないので、順列の結果を $3! = 6$ で割ってやればよい

- また次のようなことがいえる

$$\begin{aligned} {}_n C_0 &= 1 \\ {}_n C_1 &= n \\ {}_n C_n &= 1 \\ {}_n C_r &= {}_n C_{n-r} \end{aligned}$$

度数分布図の作成

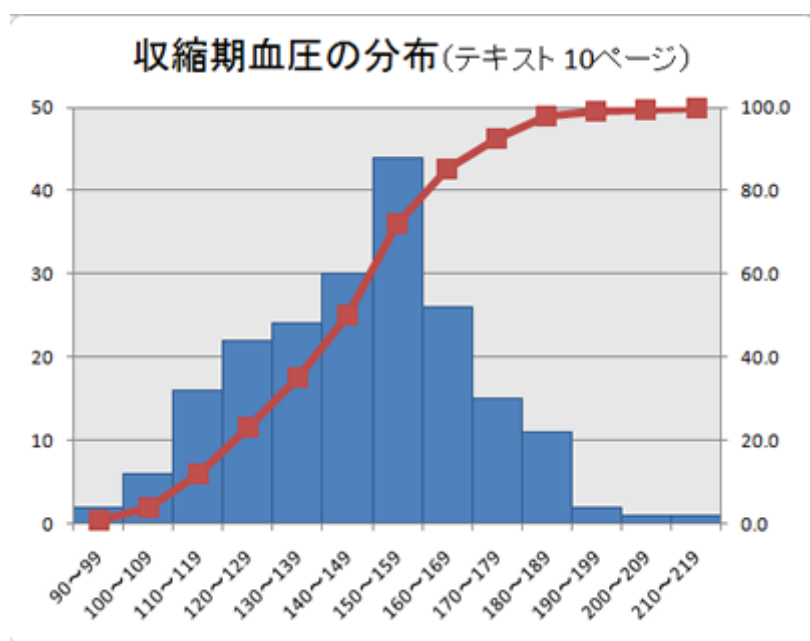
度数分布表は、データを表にまとめて全体的な傾向を値から読み取ることができます。その一方で、値をしっかりと読み取らなければ全体像を想像することはできません。

度数分布図は、データの分布を視覚的（図的）にとらえることができ、ひと目でデータの状況を把握することができます。

度数分布図の種類

ヒストグラム（histogram；柱状図）

- 横軸を階級幅によって等間隔に区切り、各階級の度数にあわせた高さの柱を描いたもの
- 柱の面積が度数に比例するように描く
- データが連続量の場合は、柱と柱は隙間をあけないよう並べて描く



度数多角形

- 階級値を横軸に、度数を縦軸にとった折れ線グラフ
- ヒストグラムの各上辺の中点を直線で結んだもの
- 度数多角形の両端は基線（横軸）につけて描く

度数曲線

- 度数多角形をなめらかな曲線で近似
- 曲線下部の面積が度数多角形の下部の面積あるいはヒストグラムの面積と一致するように描く

幹葉表示

- ヒストグラムでは、分布の形状はわかっていても、データの詳細を把握するのは難しい
- データをすべて提示しつつ、かつ、分布がわかる方法として、ターキー（J. W. Tukey）が開発した手法
 - 度数分布表の階級を幹（stem）、度数のもとである値を葉（leaf）として表現する

幹 葉
9* 0 8
10* 0 3 4 6 9
11* 0 0 2 3 3 3 5 6 6 8
12* 1 1 5 6 7 9 9
13* 1 3 3 6
14* 2 3

Excelでヒストグラムを作成

表計算ソフトの「Microsoft Excel」には、直接ヒストグラムを作成する機能はありません。

そのため、縦棒グラフを利用して、**すべての階級の階級幅が同じ**場合のヒストグラムを作成します。

縦棒グラフの作成

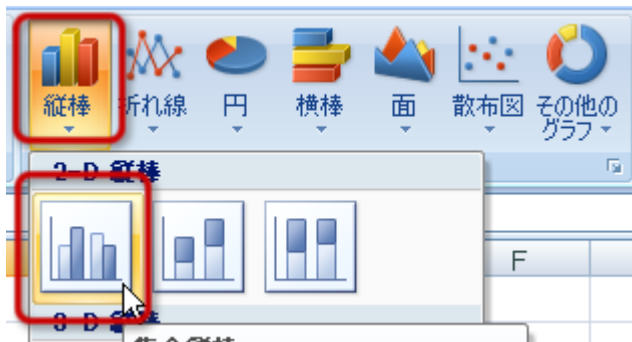
eラーニングの画面からダウンロードできるExcelのファイルを利用してみます。「度数分布図の練習(1)」というシートで作成します。

次のようにして、縦棒グラフを作成してみましょう。

1. 「Ctrl」キーを押しながらマウスをドラッグして、B2～B15セルとD2～D15セルを範囲選択する

	A	B	C	D
1		収縮期血圧(mmHg)の度数分布表(テキスト 9ペー		
2		階級	階級値	度数
3		90~99	94.5	2
4		100~109	104.5	6
5		110~119	114.5	16
6		120~129	124.5	22
7		130~139	134.5	24
8		140~149	144.5	30
9		150~159	154.5	44
10		160~169	164.5	26
11		170~179	174.5	15
12		180~189	184.5	11
13		190~199	194.5	2
14		200~209	204.5	1
15		210~219	214.5	
16		計		200

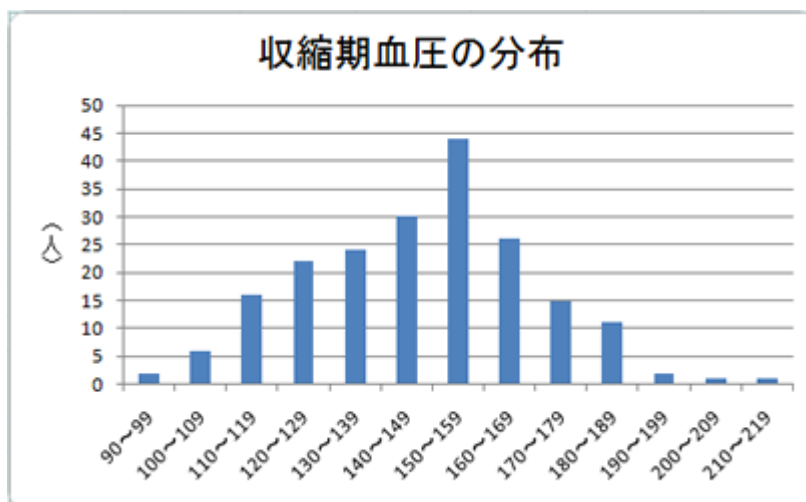
2. 「挿入」タブの「グラフ」グループにある「縦棒」ボタンをクリックする
3. メニューから「2-D縦棒」の「2-D集合縦棒」を選択する



4. グラフが作成される

作成できたら、グラフに次の設定をしてください。

- 凡例は表示を「なし」にする
- グラフのタイトルを設定する
- 縦軸の軸ラベルを「(人)」と設定する

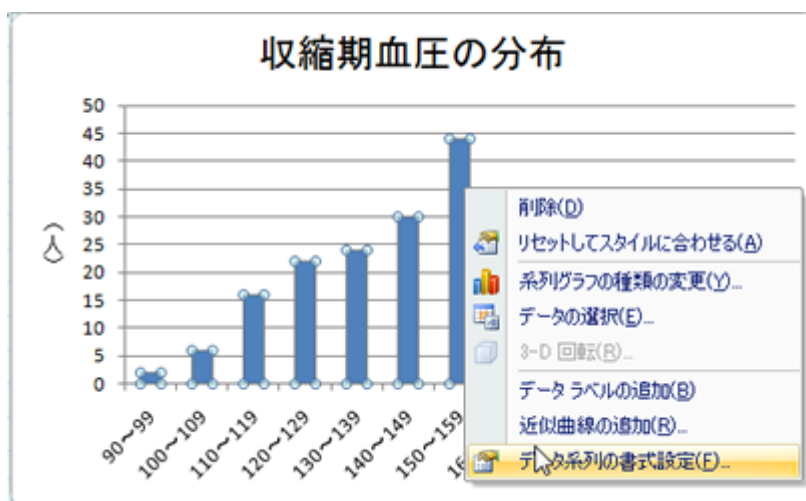


棒グラフからヒストグラムへ

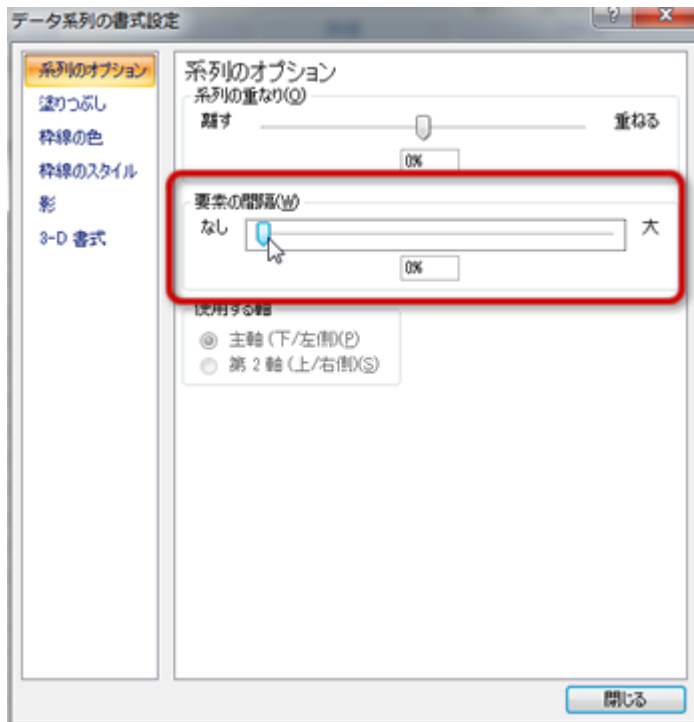
縦棒グラフのままでは、縦棒と縦棒の間に隙間があいてしまいます。

次のような設定をして、隙間をなくして、ヒストグラムの形にします。

1. グラフの縦棒部分をマウスで右クリックして、メニューから「データ系列の書式設定」を選択する



2. 「データ系列の書式設定」の書式設定が表示されたら、「系列オプション」の「要素の間隔」を「なし」(0%)に設定する



3. 「閉じる」ボタンをクリックすると、縦棒の隙間がなくなる

Excelで代表値と散布度を計算

数式の入力

Excelでは、セルに「**数式**」を入力することで、計算ができます。数式を入力するときの基本的なルールは、次のとおりです。

- 最初は「=」ではじめる
- カッコ「()」を使って計算する順番を指定できる
- 四則演算が使える（半角で入力）

演算	数学での記号	Excelでの記号	計算式の例	表示される結果
足し算	+	+	=1+2	3
引き算	-	-	=2-3	-1
掛け算	×	*	=4*5	20
割り算	÷	/	=1/2	0.5
べき乗	^	^	=2^3	8

数式の入力例

たとえば、身長と体重のデータから人の肥満度をはかる指標である、BMI（ボディマス指数）を計算する場合を考えてみましょう。

BMI = 体重 (kg) ÷ 身長 (m) の2乗

身長がB2～B11セルに、体重のデータがC2～C11セルに入力されており、それらから求めたBMIをD2～D11セルに表示させるには、次のように操作します。

1. D2セルに次の計算式を入力する

=C2/((B2/100)^2) （「/100」としているのは、身長がcm単位のため）

2. 「Enter」キーを押すと、計算結果が表示される
3. D2セルの計算結果を、D3～D11セルへコピーする

平方根、n乗根の計算

- 正の平方根()を計算するには、**SQRT**関数を利用します。

SQRT(平方根を計算する)

- 書式：SQRT(数値)
- 引数：平方根を求める数値
- 例：A12セルの数値の平方根を計算する

=SQRT(A12)

- n乗根を計算する関数はないため、べき乗(^)を利用する
 - 「n乗根の計算」は、「1/nのべき乗の計算」と同じ意味になることを利用する

◦ 例：A12セルの数値の4乗根 $\sqrt[4]{A12}$ を計算


```
=A12^(1/4)
```

代表値を計算

平均値

- 算術平均は、**AVERAGE**関数を利用します。

AVERAGE (平均値を計算する)

- 書式 : AVERAGE(数値1, 数値2, ...)
- 引数 : 数値1, 数値2, ... : 平均を計算するセルの範囲
- 例 : F1 ~ F10セルまでのセルの数値の平均値を計算する

```
=AVERAGE(F1:F10)
```

- 幾何平均は、**GEOMEAN**関数を利用します。

GEOMEAN (正の数からなる配列またはセル範囲のデータの幾何平均を計算する)

- 書式 : GEOMEAN(数値1, 数値2, ...)
- 引数 : 数値1, 数値2, ... : 平均を計算するセルの範囲

- 調和平均は、**HARMEAN**関数を利用します。

HARMEAN (1 組の数値の調和平均を計算する)

- 書式 : HARMEAN(数値1, 数値2, ...)
- 引数 : 数値1, 数値2, ... : 平均を計算するセルの範囲

中央値

- 中央値は、**MEDIAN**関数を利用します。

MEDIAN (引数に含まれる数値の中央値を求める)

- 書式 : MEDIAN(数値1, 数値2, ...)
- 引数 : 数値1, 数値2, ... : 中央値を計算するセルの範囲
- 例 : F1 ~ F10セルまでのセルの中央値を求める

```
=MEDIAN(F1:F10)
```

四分位数

- 四分位数は、**QUARTILE**関数を利用します。

QUARTILE (配列に含まれるデータから四分位数を抽出する)

- 書式 : QUARTILEE(配列, 戻り値)
- 引数 : 配列 : 対象となるデータを含む配列 (セルの範囲)
- 引数 : 戻り値 : 戻り値として返す四分位数の内容を指定
 - 戻り値: 0: 最小値
 - 戻り値: 1: 第1四分位数 (25%)
 - 戻り値: 2: 第2四分位数 (50%)=中央値
 - 戻り値: 3: 第3四分位数 (75%)

- 戻り値: 4: 最大値

百分位数

- 百分位数は、PERCENTILE関数を利用します。

PERCENTILE (配列に含まれるデータから百分位数 (%) を抽出する)

- 書式 : QUARTILEE(配列, 率)
- 引数 : 配列 : 対象となるデータを含む配列 (セルの範囲)
- 引数 : 率 : 0 ~ 1の値で、目的の百分位の値 (パーセンタイル値) を指定

最頻値

- 最頻値は、MODE関数を利用します。

MODE (引数に含まれるデータのなかで最も頻繁に出現する値を求める)

- 書式 : MODE(数値1, 数値2, ...)
- 引数 : 数値1, 数値2, ... : 最頻値を計算するセルの範囲
- 例 : F1 ~ F10セルまでのセルの最頻値を求める

```
=MODE(F1:F10)
```

散布度を計算

分散

- 分散は、VARP関数を利用します。

VARP (引数を母集団全体と見なし、母集団の分散 (標本分散) を求める)

- 書式 : VARP(数値1, 数値2, ...)
- 引数 : 数値1, 数値2, ... : 母集団に対応するセルの値、セルの範囲

標準偏差

- 標準偏差は、STDEVP関数を利用します。

STDEVP (引数を母集団全体であると見なして、母集団の標準偏差を求める)

- 書式 : STDEVP(数値1, 数値2, ...)
- 引数 : 数値1, 数値2, ... : 母集団に対応するセルの値、セルの範囲

不偏分散

- 不偏分散は、VAR関数を利用します。

VAR (引数を正規母集団の標本と見なし、標本に基づいて母集団の分散の推定値 (不偏分散) を求める)

- 書式 : VAR(数値1, 数値2, ...)
- 引数 : 数値1, 数値2, ... : 母集団の標本に対応するセルの値、セルの範囲

不偏標準偏差

- 不偏標準偏差は、STDEV関数を利用します。

STDEV (引数を標本と見なし、標本に基づいて母集団の標準偏差の推定値を求める)

- 書式 : STDEV(数値1, 数値2, ...)

○引数：数値1, 数値2, ... :母集団に対応するセルの値、セルの範囲

Excelで相関と回帰を計算

相関を計算

相関係数

- 2つの配列データの相関係数は、**CORREL**関数を利用します。

CORREL(相関係数の値を返す)

- 書式 : CORREL(配列1, 配列2, ...)
- 引数 : 配列1 ... : データが入力されたセルの範囲
- 引数 : 配列2 ... : もう一方のデータが入力されたセルの範囲
- 例 : データがA1～A10セルとB1～B10までのセルの数値から、相関関数を計算する

```
=CORREL(A1:A10, B1:B10)
```

- ピアソンの積率相関係数は、**PEARSON**関数を利用します。

PEARSON(ピアソンの積率相関係数 r の値を返す)

- 書式 : PEARSON(配列1, 配列2)
- 引数 : 配列1 ... : 独立変数に対応するセルの範囲
- 引数 : 配列2 ... : 従属変数に対応するセルの範囲
- 例 : 独立変数がA1～A10セル、従属変数がB1～B10までのセルの数値から、積率相関関数を計算する

```
=PEARSON(A1:A10, B1:B10)
```

- なお、Excel2004以降は、CORREL関数の結果とPEARSON関数の結果は同じになります。

共分散

- 共分散(2種類のデータ間での標準偏差の積の平均値)は、**COVAR**関数を利用します。

COVAR(共分散の値を返す)

- 書式 : COVAR(配列1, 配列2)
- 引数 : 配列1 ... : データが入力されたセルの範囲
- 引数 : 配列2 ... : もう一方のデータが入力されたセルの範囲
- 例 : データがA1～A10セルとB1～B10までのセルの数値から、共分散を計算する

```
=COVAR(A1:A10, B1:B10)
```

偏差平方和

- 偏差平方和(標本の平均値に対する各データの偏差の平方和)は、**DEVSQ**関数を利用します。

DEVSQ(偏差平方和の値を返す)

- 書式 : DEVSQ(数値1, 数値2, ...)
- 引数 : 数値1, 数値2 ... : データが入力されたセルの範囲
- 例 : データがA1～A10セルのセルの数値から、偏差平方和を計算する

```
=DEVSQ(A1:A10)
```

回帰を計算

回帰直線の傾き

- 既知の y と既知の x のデータから回帰直線の傾きには、**SLOPE**関数を利用します。

SLOPE(回帰直線の傾きを返す)

- 書式 : SLOPE(配列1, 配列2)
- 引数 : 配列1 ... : 既知の y(従属変数)に対応するセルの範囲
- 引数 : 配列2 ... : 既知の x(独立変数)に対応するセルの範囲
- 例 : 既知の y(従属変数)がA1 ~ A10セル、既知の x(独立変数)がB1 ~ B10までのセルの数値から、回帰直線の傾きを計算する

```
=SLOPE(A1:A10, B1:B10)
```

回帰直線のy切片

- 既知の y と既知の x のデータから(線形)回帰直線のy切片には、**INTERCEPT**関数を利用します。

INTERCEPT(回帰直線の切片を返す)

- 書式 : INTERCEPT(配列1, 配列2)
- 引数 : 配列1 ... : 既知の y(従属変数)に対応するセルの範囲
- 引数 : 配列2 ... : 既知の x(独立変数)に対応するセルの範囲
- 例 : 既知の y(従属変数)がA1 ~ A10セル、既知の x(独立変数)がB1 ~ B10までのセルの数値から、回帰直線のy切片を計算する

```
=INTERCEPT(A1:A10, B1:B10)
```

決定係数

- 既知の y と既知の x のデータからR²(決定係数)を求めるには、**RSQ**関数を利用します。

RSQ(r²の値を返す)

- 書式 : RSQ(配列1, 配列2)
- 引数 : 配列1 ... : 既知の y(従属変数)に対応するセルの範囲
- 引数 : 配列2 ... : 既知の x(独立変数)に対応するセルの範囲
- 例 : 既知の y(従属変数)がA1 ~ A10セル、既知の x(独立変数)がB1 ~ B10までのセルの数値から、決定係数 R²を計算する

```
=RSQ(A1:A10, B1:B10)
```

Excelで散布図と回帰直線を作成

2組のデータの関係性を視覚的に把握するには、「**散布図**」を用います。

また、表計算ソフトのExcelでは、作成した散布図を利用して、「**回帰直線**」（単回帰直線）を描いたり、回帰式を表示することができます。

Excelで散布図の作成

散布図の作成

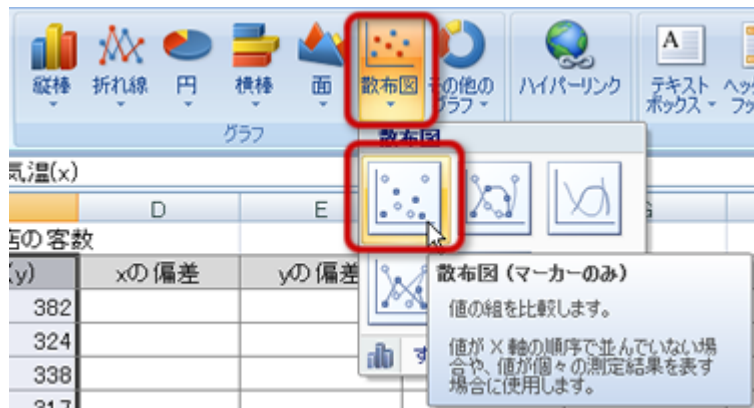
eラーニングの画面からダウンロードできるExcelのファイルを利用してみます。「**【練習】** 相関と回帰」というシートで作成します。

次のようにして、散布図を作成してみましょう。

1. マウスをドラッグして、B2～C22セルを範囲選択する

	A	B	C
1		最高気温とアイスクリーム店の客数	
2	データ番号	最高気温(x)	客数(y)
3	1	33	382
4	2	33	324
5	3	34	338
6	4	34	317
7	5	35	341
8	6	35	360
9	7	34	339
10	8	32	329
11	9	28	218
12	10	35	402
13	11	33	342
14	12	28	205
15	13	32	368
16	14	33	196
17	15	35	304
18	16	30	294
19	17	29	275
20	18	32	336
21	19	34	384
22	20	35	385

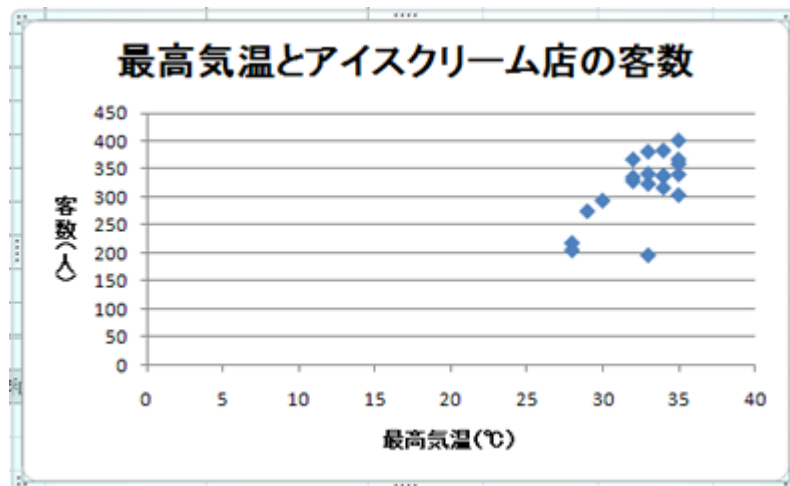
2. 「挿入」タブの「グラフ」グループにある「散布図」ボタンをクリックする
3. メニューから「散布図(マーカーのみ)」を選択する



4. グラフが作成される

作成できたら、グラフに次の設定をしてください。

- 凡例は表示を「なし」にする
- グラフのタイトルを「最高気温とアイスクリーム店の客数」設定する
- 縦軸の軸ラベルを「客数(人)」と設定する
- 横軸の軸ラベルを「最高気温()」と設定する

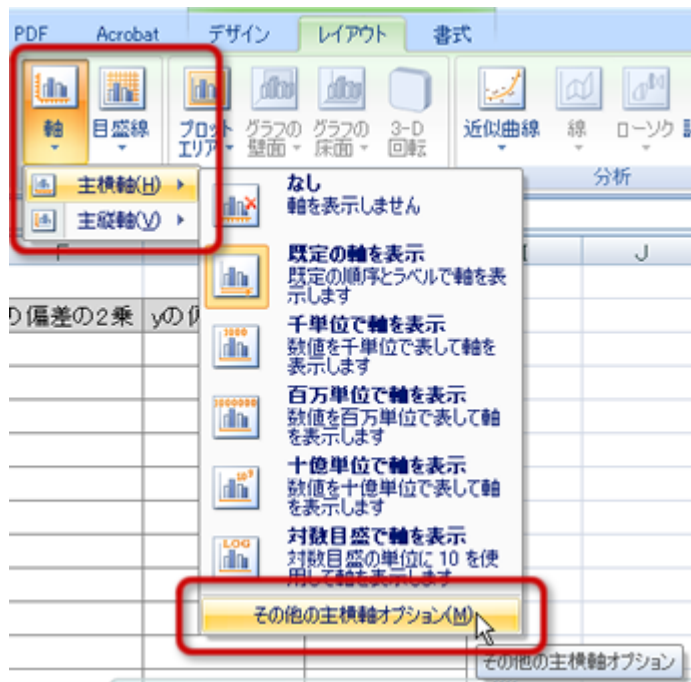


軸の設定

横軸（最高気温）が「0度」から表示されているため、データの分布がグラフの右端に偏ってしまっています。

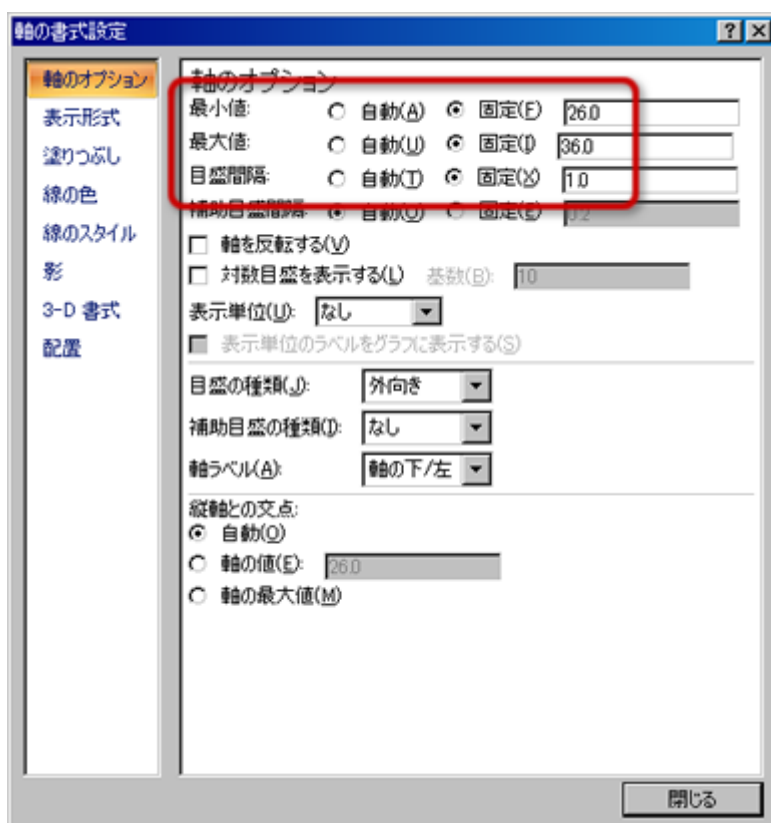
そこで、横軸の範囲を「26度～36度」の範囲に変更してみましょう。

1. メニューの「グラフツール」の「レイアウト」にある「軸」「主横軸」「主横軸オプション」を選択する

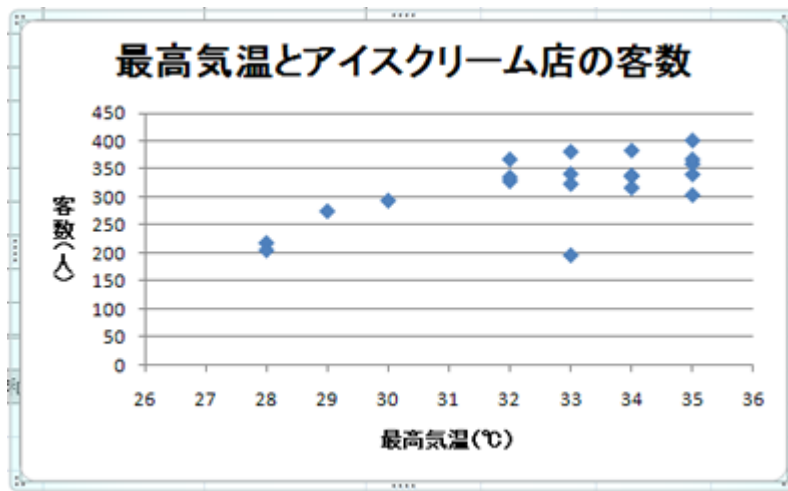


2. 「軸の書式設定」が表示されるので、次のように設定する

- 「最小値」を「固定」に選択して、入力欄に「26.0」と入力
- 「最大値」を「固定」に選択して、入力欄に「36.0」と入力
- 「目盛間隔」を「固定」に選択して、入力欄に「1.0」と入力



3. 「閉じる」をクリックすると、横軸の表示が変更される

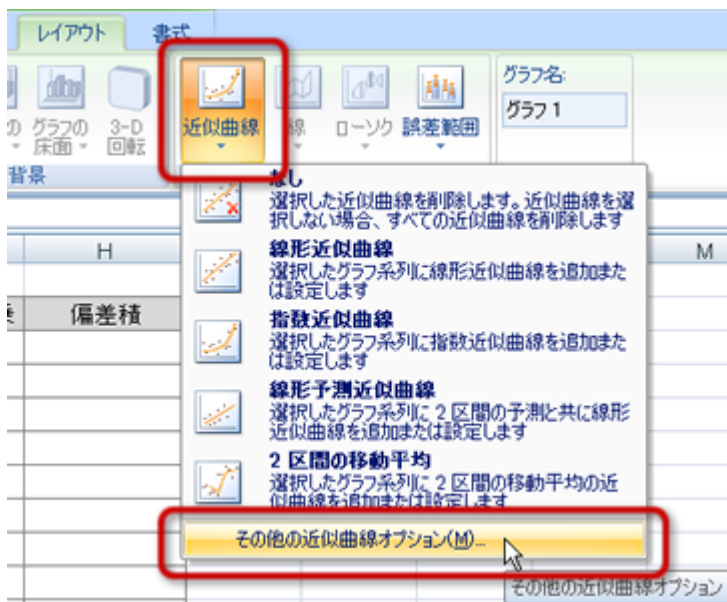


あとは、必要に応じて、縦軸の範囲も設定したり、デザインの変更をすると良いでしょう。

Excelで回帰直線の作成

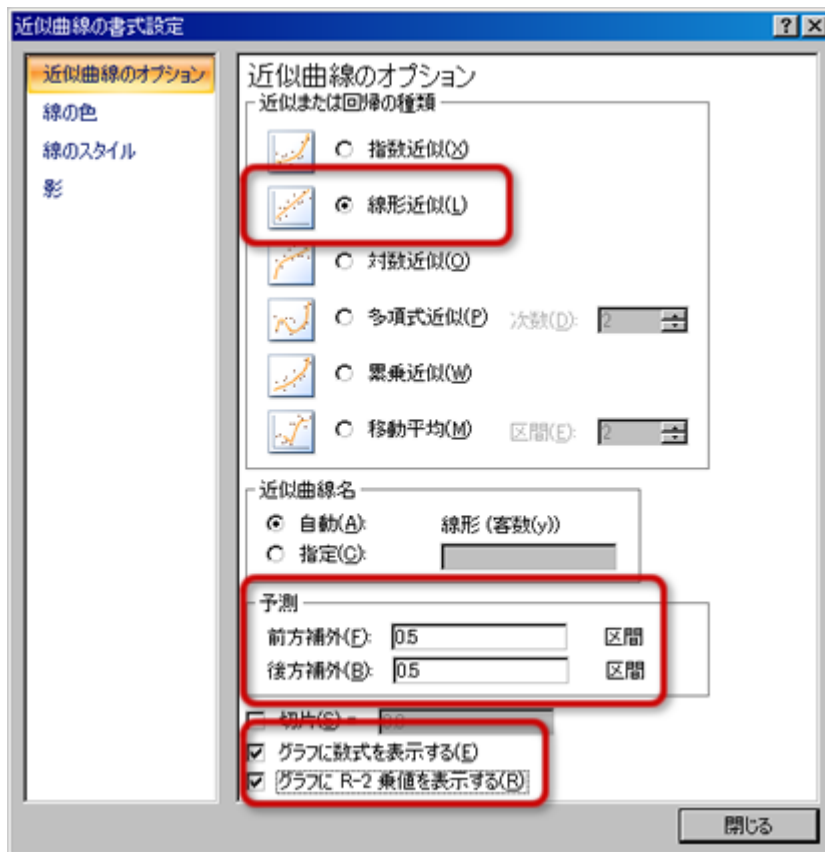
Excelでは、「近似曲線」という機能で、回帰直線を描くことができます。

1. メニューの「グラフツール」の「レイアウト」にある「近似曲線」「その他の近似曲線オプション」を選択する

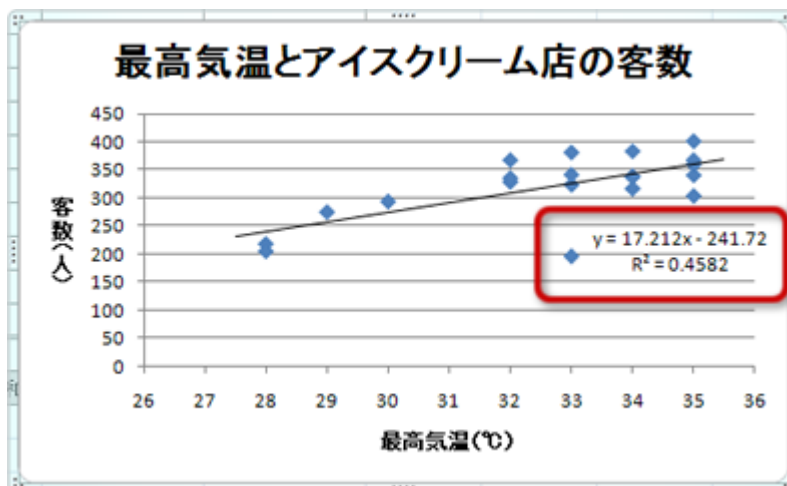


2. 「近似曲線の書式設定」が表示されるので、次のように設定する

- 「近似または回帰の種類」を「線形近似」に選択する
- 「予測」の「前方補外」・「後方補外」の入力欄に、それぞれ「0.5」と入力
- 「グラフに数式を表示する」と「グラフにR-2乗値を表示する」のチェック欄をクリックしてチェックする



3. 「閉じる」をクリックすると、散布図のグラフに回帰直線（と回帰式）が表示される



これで、回帰曲線と同時に、回帰式（傾き、y切片）、決定係数(R^2値）が表示されます。

Excelで順列と組み合わせを計算

階乗

- 数値の階乗を計算するには、**FACT**関数を利用します。

FACT(数値の階乗の値を返す)

- 書式 : FACT(数値)
- 引数 : 数値 ... : 階乗を求める正の整数 (小数点以下がある数値を指定すると小数点以下が切り捨て)
- 例 : 10の階乗を計算する

```
=FACT(10)
```

順列

- 順列 (異なる n 個のものから r 個を選んだ並べ方) を求めるには、**PERMUT**関数を利用します。

PERMUT(順列の値を返す)

- 書式 : PERMUT(総数, 抜き取り数)
- 引数 : 総数 ... : 選びとる対象となる全体の数
- 引数 : 抜き取り数 ... : 選びとる数
- 例 : 8個のものから3個を取る順列の値を計算する

```
=PERMUT(8, 3)
```

組み合わせ

- 組み合わせ (異なる n 個のものから r 個を選ぶときの組み合わせ) を求めるには、**COMBIN**関数を利用します。

COMBIN(組み合わせの値を返す)

- 書式 : COMBIN(総数, 抜き取り数)
- 引数 : 総数 ... : 抜き取る対象となる全体の数
- 引数 : 抜き取り数 ... : 抜き取る組み合わせ 1 組に含まれる項目の数
- 例 : 8個のものから3個を選ぶときの組み合わせの値を計算する

```
=COMBIN(8, 3)
```