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**関数または**COVARIANCE.P**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)
```

COVARIANCE.P （共分散の値を返す）

- 書式：COVARIANCE.P(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R^2（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。

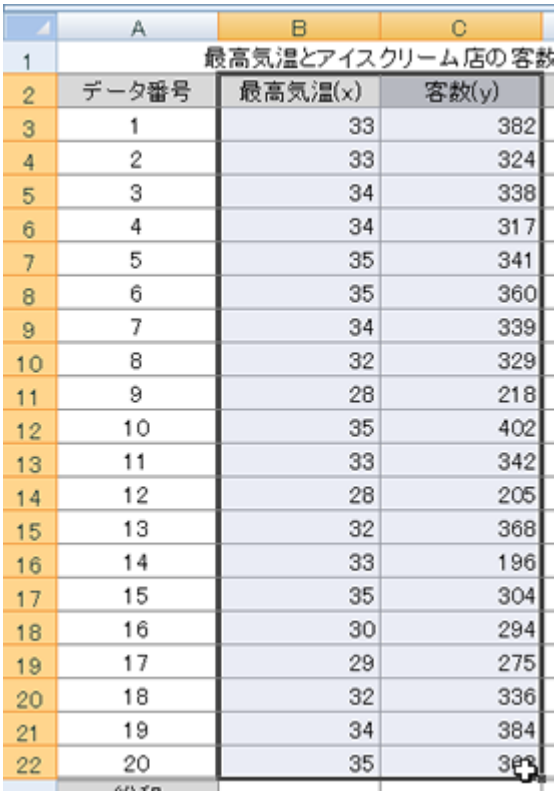
Excelで散布図の作成

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。

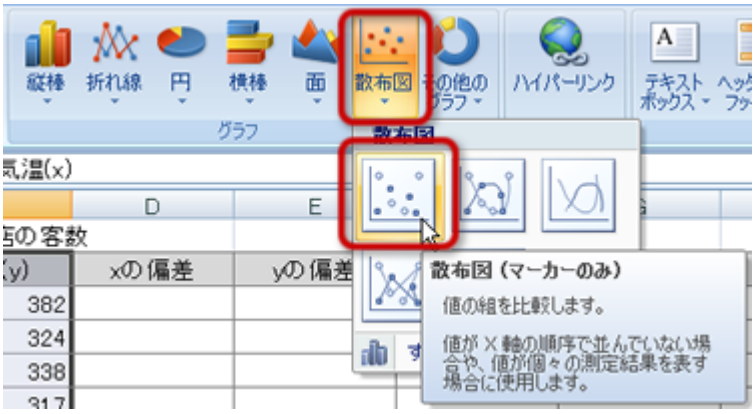
次のようにして、散布図を作成してみましょう。

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	3[illegible]

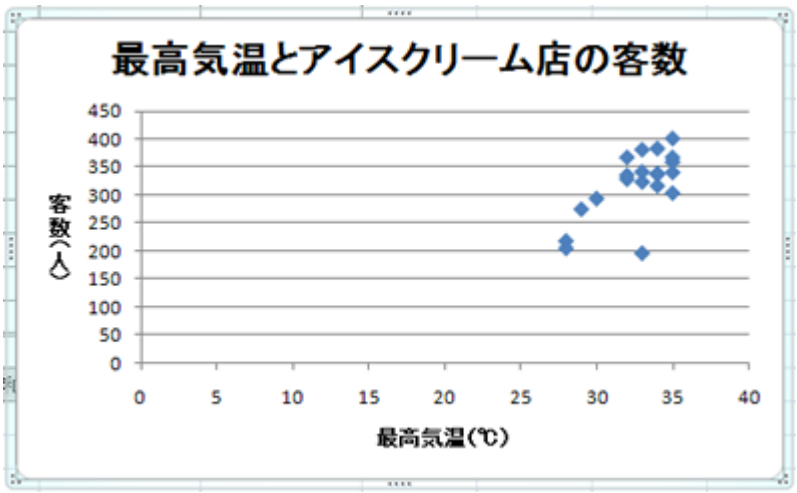
2. 「挿入」タブの「グラフ」グループにある「散布図」ボタンをクリックする
3. メニューから「散布図（マーカーのみ）」を選択する



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**

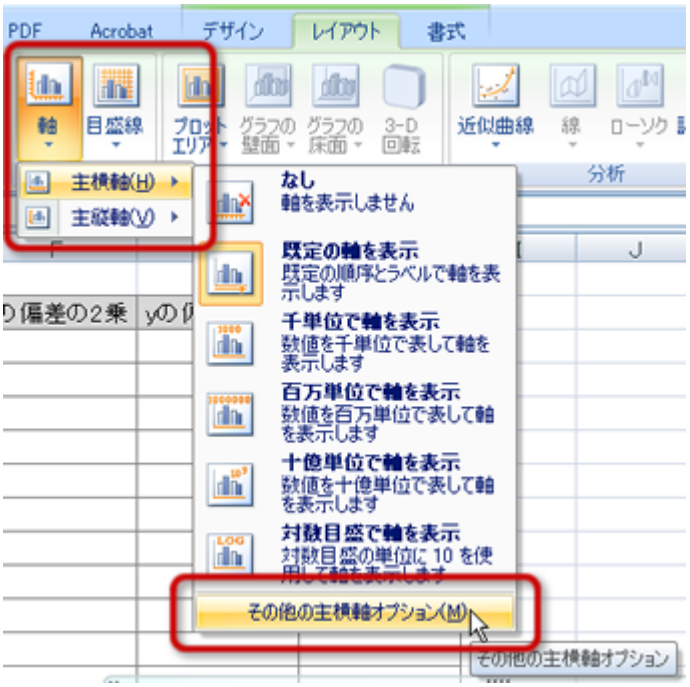


軸の設定

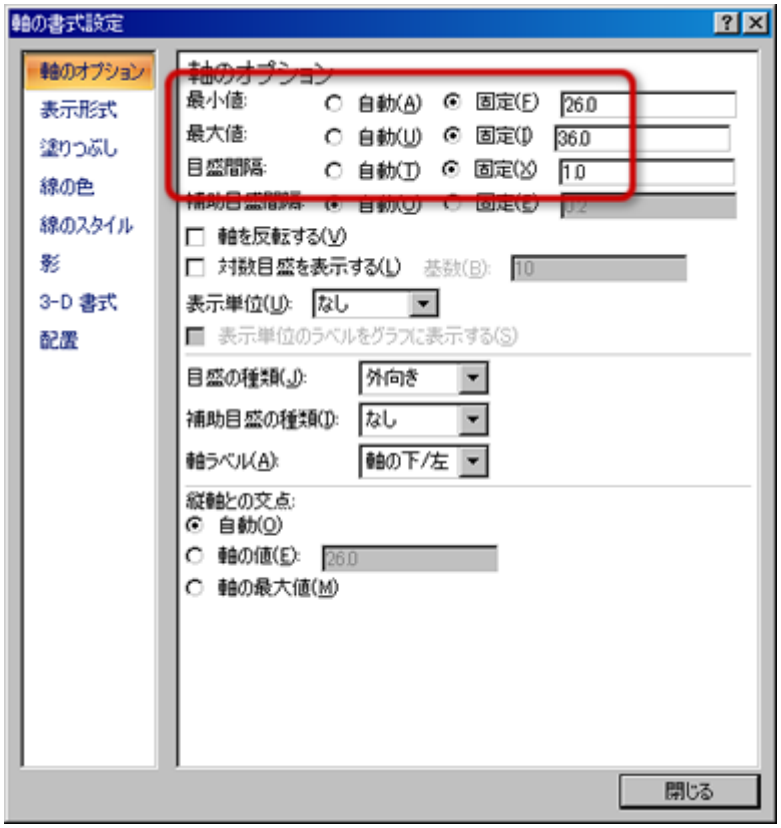
横軸（最高気温）が「0度」から表示されているため、データの分布がグラフの右端に偏ってしまっています。

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。

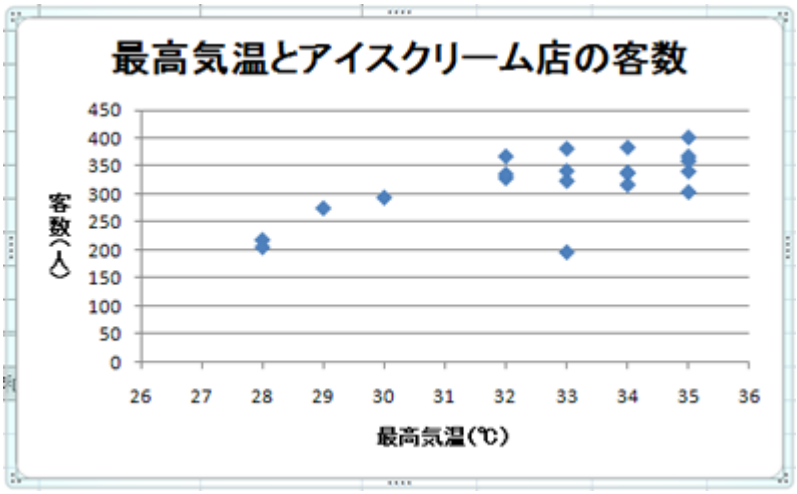
1. **メニューの「グラフツール」の「レイアウト」にある「軸」→「主横軸」→「主横軸オプション」を選択する**



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

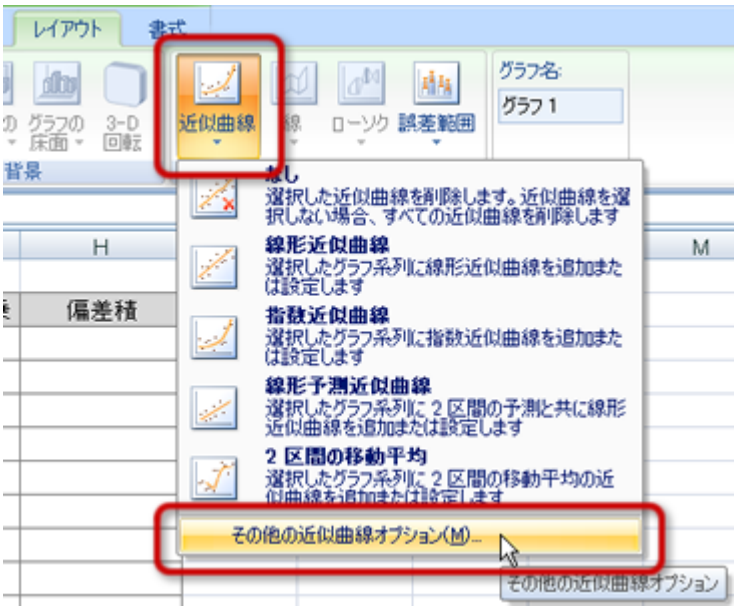


あとは、必要に応じて、縦軸の範囲も設定したり、デザインの変更をすると良いでしょう。

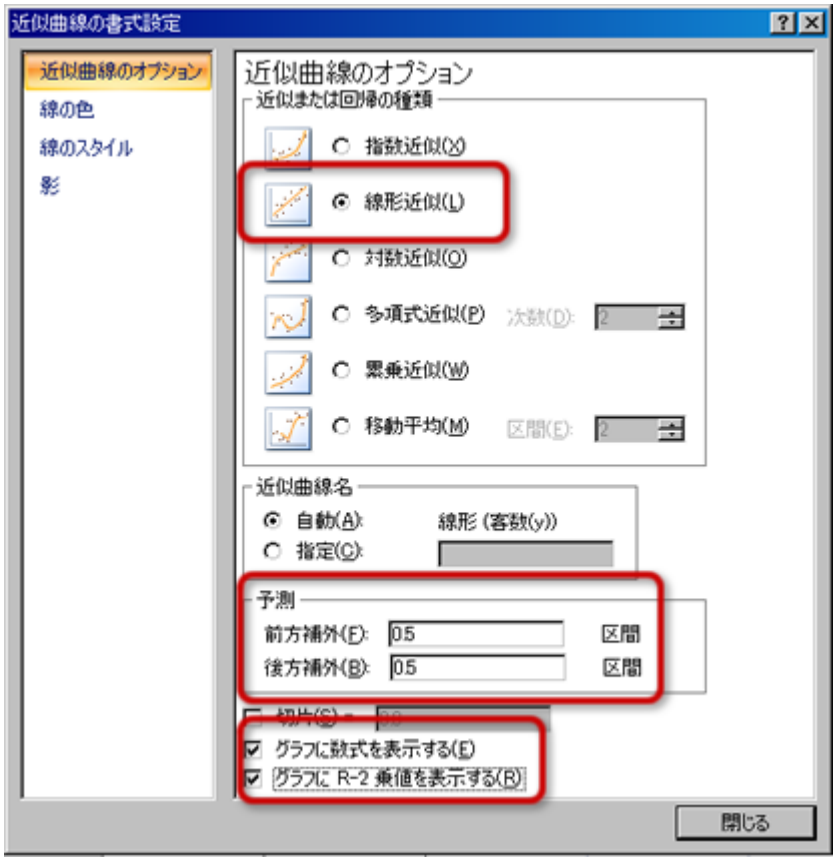
Excelで回帰直線の作成

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。

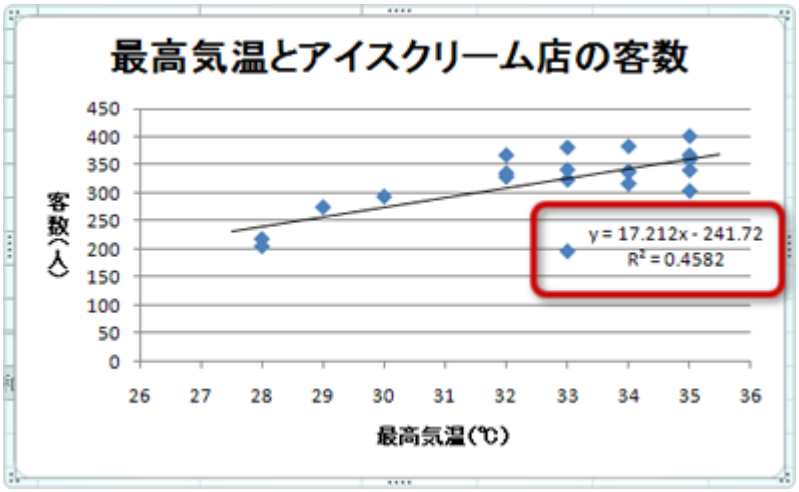
1. メニューの「グラフツール」の「レイアウト」にある「近似曲線」→「その他の近似曲線オプション」を選択する



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。