

データの**関係**を調査

相関係数を計算してみよう

得られたデータの関係を調査する方法

- データが数値の場合
 1. 基本統計量（平均，標準偏差，最大，最小）
 2. 度数分布表 → ヒストグラム
 3. 散布図
 4. 相関係数
- データが数値ではない場合
 1. クロス集計表（ピボットテーブル）
 2. 積み上げ棒グラフ
 3. 数値へ置き換え
 4. 相関係数

データが数値 場合

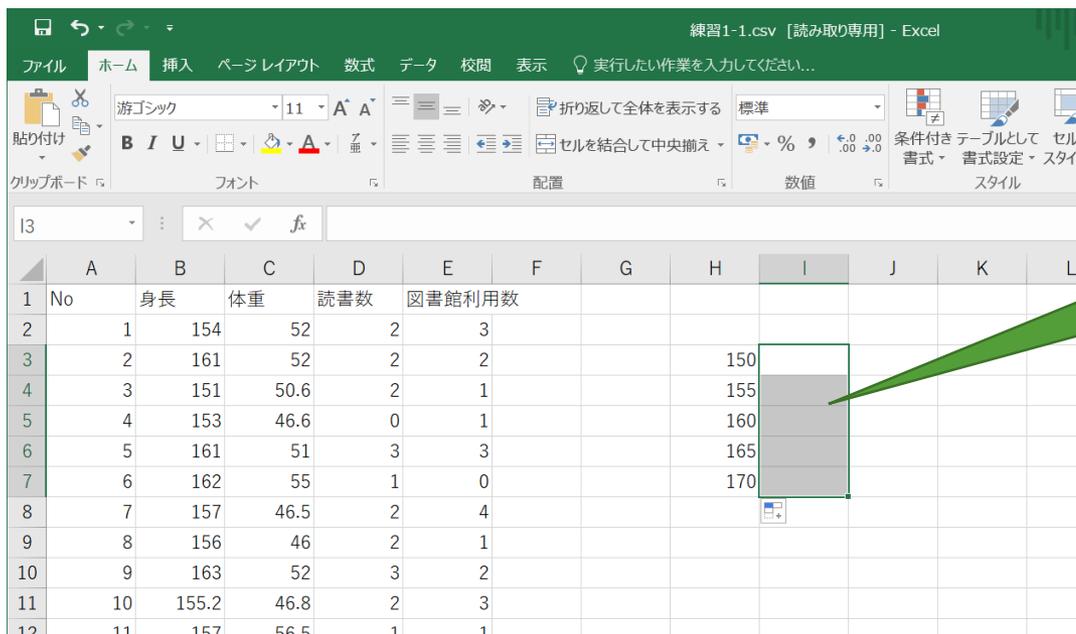
基本統計量の関数

- 平均 =AVERAGE(範囲)
- 標準偏差 =STDEV.P(範囲)
- 最大 =MAX(範囲)
- 最小 =MIN(範囲)

- 度数分布表 =FREQUENCY(範囲, 区間)

度数分布表(FREQUENCY その1)

1. 最小から最大までの区間をあらかじめ入力する
2. 区間の右側を複数選択する (計算結果)



練習1-1.csv [読み取り専用] - Excel

ファイル ホーム 挿入 ページレイアウト 数式 データ 校閲 表示 実行したい作業を入力してください...

貼り付け フォント 配置 数値 スタイル

| | A | B | C | D | E | F | G | H | I | J | K | L |
|----|----|-------|------|-----|--------|---|---|-----|---|---|---|---|
| 1 | No | 身長 | 体重 | 読書数 | 図書館利用数 | | | | | | | |
| 2 | 1 | 154 | 52 | 2 | 3 | | | | | | | |
| 3 | 2 | 161 | 52 | 2 | 2 | | | 150 | | | | |
| 4 | 3 | 151 | 50.6 | 2 | 1 | | | 155 | | | | |
| 5 | 4 | 153 | 46.6 | 0 | 1 | | | 160 | | | | |
| 6 | 5 | 161 | 51 | 3 | 3 | | | 165 | | | | |
| 7 | 6 | 162 | 55 | 1 | 0 | | | 170 | | | | |
| 8 | 7 | 157 | 46.5 | 2 | 4 | | | | | | | |
| 9 | 8 | 156 | 46 | 2 | 1 | | | | | | | |
| 10 | 9 | 163 | 52 | 3 | 2 | | | | | | | |
| 11 | 10 | 155.2 | 46.8 | 2 | 3 | | | | | | | |
| 12 | 11 | 157 | 56.5 | 1 | 1 | | | | | | | |

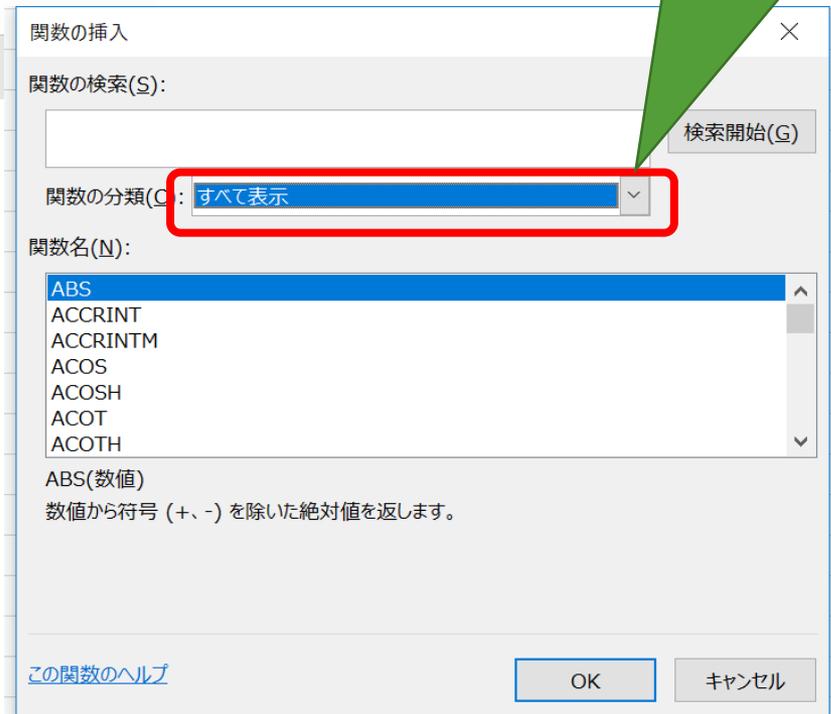
ここを複数選択

度数分布表(FREQUENCY その2)



関数の挿入

「すべて表示」に変更



度数分布表(FREQUENCY その3)

関数の挿入

関数の検索(S):

関数の分類(C): すべて表示

関数名(N):

- FORECAST.ETS.STAT
- FORECAST.LINEAR
- FORMULATEXT
- FREQUENCY**
- ETEST
- FV
- FVSCHEDULE

FREQUENCY(データ配列,区間配列)

範囲内でのデータの度数分布を、垂直配列で返します。返された配列要素の個数の範囲は、区間配列の個数より 1 だけ多くなります。

[この関数のヘルプ](#)

FREQUENCYを選択

関数の引数

FREQUENCY

データ配列 = 参照

区間配列 = 参照

=

範囲内でのデータの度数分布を、垂直配列で返します。返された配列要素の個数は、区間配列の個数より 1 だけ多くなります。

データ配列 には度数分布を求めたい値の配列、または参照を指定します。空白セルおよび文字列は無視されます。

数式の結果 =

[この関数のヘルプ\(H\)](#)

度数分布表(FREQUENCY その4)

| A | B | C | D | E | F | G | H | I | J | K | L | M |
|----|-------|------|----|---|---|---|-----|--------|---|---|---|---|
| No | 身長 | 体重 | 読書 | | | | | | | | | |
| 1 | 154 | 52 | | | | | | | | | | |
| 2 | 161 | 52 | 2 | 2 | | | 150 | H3:H7) | | | | |
| 3 | 151 | 50.6 | 2 | 1 | | | 155 | | | | | |
| 4 | 153 | 46.6 | 0 | 1 | | | 160 | | | | | |
| 5 | 161 | 51 | 3 | 3 | | | 165 | | | | | |
| 6 | 162 | 55 | 1 | 0 | | | 170 | | | | | |
| 7 | 157 | 46.5 | 2 | 4 | | | | | | | | |
| 8 | 156 | 46.5 | 4 | 5 | | | | | | | | |
| 9 | 163 | 56 | 5 | 6 | | | | | | | | |
| 10 | 155.2 | 46.5 | 48 | 5 | | | | | | | | |
| 11 | 157 | 56 | 5 | 6 | | | | | | | | |
| 12 | 150 | 48 | 5 | 6 | | | | | | | | |
| 13 | 155 | 49 | 4 | 5 | | | | | | | | |
| 14 | 163 | 49 | 4 | 5 | | | | | | | | |
| 15 | 164 | 49 | 4 | 5 | | | | | | | | |
| 16 | 165 | 49 | 4 | 5 | | | | | | | | |
| 17 | 156 | 49 | 4 | 5 | | | | | | | | |
| 18 | 168 | 49 | 4 | 5 | | | | | | | | |
| 19 | 159 | 49 | 4 | 5 | | | | | | | | |
| 20 | 158 | 49 | 4 | 5 | | | | | | | | |

データ (Red box around B2:B31)

区間 (Blue box around H3:H7)

関数の引数 dialog box:

FREQUENCY

データ配列: B2:B31 = {154;161;151;153;161;162;157;156}

区間配列: H3:H7 = {150;155;160;165;170}

= {2;5;11;10;2;0}

範囲内でのデータの度数分布を、垂直配列で返します。返された配列要素の個数は、区間配列の個数より 1 つだけ多くなります。

区間配列 にはデータ配列で指定したデータをグループ化するため、値の間隔を配列または参照として指定します。

数式の結果 = 2

[この関数のヘルプ\(H\)](#) OK キャンセル

度数分布表(FREQUENCY その5)

関数の引数

FREQUENCY

データ配列 B2:B31 = {154;161;151;153;161;162;157;1...}

区間配列 H3:H7 = {150;155;160;165;170}

= {2;5;11;10;2;0}

範囲内のデータの度数分布を、垂直配列で返します。返された配列要素の個数は、区間配列の個数より 1 つだけ多くなります。

区間配列 にはデータ配列で指定したデータをグループ化するため、値の間隔を配列または参照として指定します。

数式の結果 = 2

[この関数のヘルプ\(H\)](#)

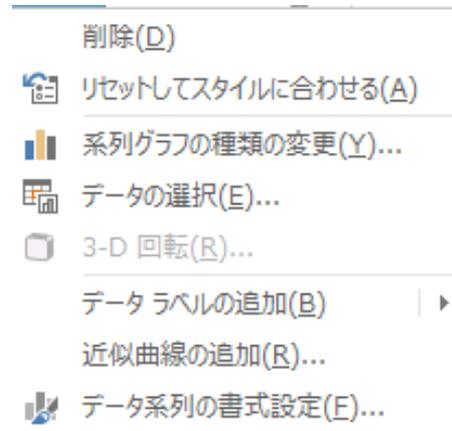
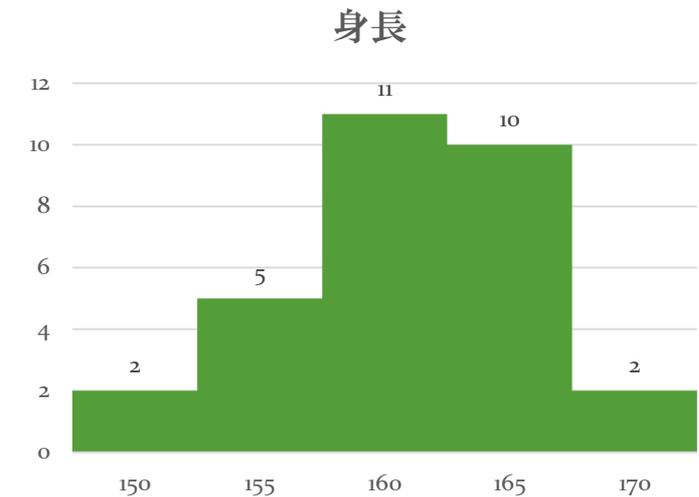
計算結果が同時に表示

[Ctrl] + [Shift]を同時に押しながら
[OK]をクリック

| | G | H | I | J |
|--|---|-----|----|---|
| | | | | |
| | | | | |
| | | 150 | 2 | |
| | | 155 | 5 | |
| | | 160 | 11 | |
| | | 165 | 10 | |
| | | 170 | 2 | |
| | | | | |
| | | | | |

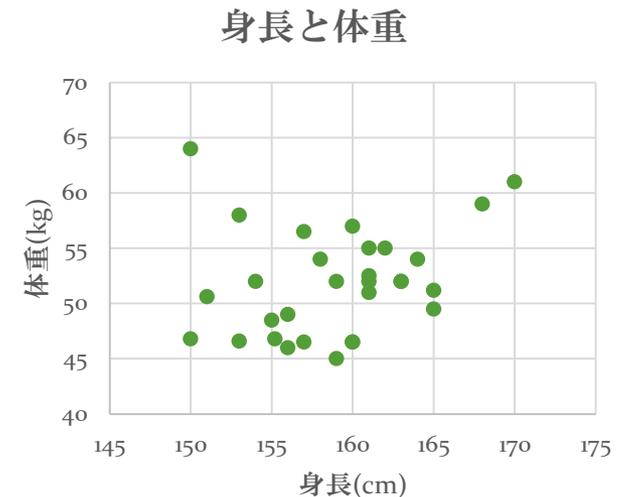
ヒストグラム

- 個々の集計
 - 度数分布表 → ヒストグラム
- 「挿入」→「棒グラフ」
- 棒グラフの間をなくすには、グラフの上で右クリックし、「データ系列の書式設定」→「要素の間隔」を0%



散布図

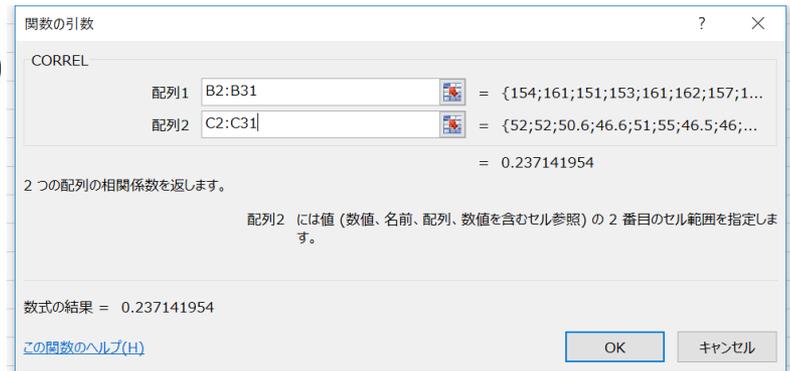
- 2つの関係
 - 2つの並んだデータ → 散布図
 - 2列データを選択して →「挿入」の「散布図」
 - 縦軸横軸の範囲を変更するには、グラフの数値の上で右クリックし、「軸の書式設定」で数値を変更



- 削除(D)
- リセットしてスタイルに合わせる(A)
- A フォント(E)...
- グラフの種類の変更(Y)...
- データの選択(E)...
- 3-D 回転(R)...
- 補助目盛線の追加(N)
- 目盛線の書式設定(M)...
- 軸の書式設定(E)...

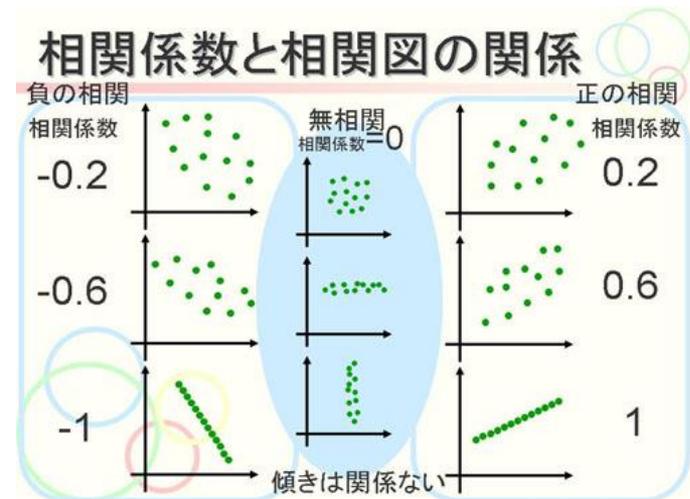
関係を数値で確認

- 相関係数を求める関数（どちらも同じ）
 - = CORREL(データ1, データ2)
 - = PEARSON(データ1, データ2)



相関係数の値 相関係数の強弱

| | |
|-----------|----------|
| 0.7~1 | 強い正の相関あり |
| 0.4~0.7 | 正の相関あり |
| 0.2~0.4 | 弱い正の相関あり |
| -0.2~0.2 | ほぼ関係ない |
| -0.4~-0.2 | 弱い負の相関あり |
| -0.7~-0.4 | 負の相関あり |
| -1~-0.7 | 強い負の相関あり |



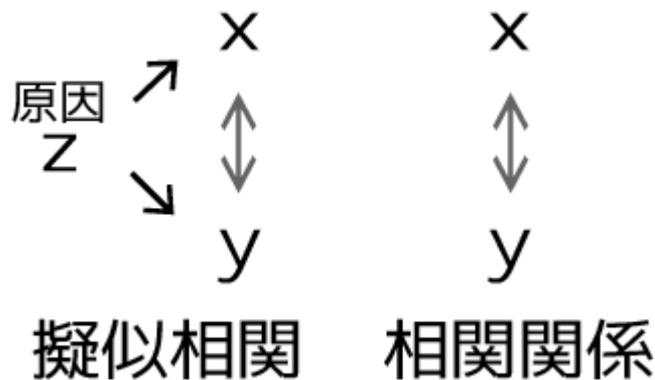
引用元：なるほど統計学園高等部

まとめ

- 得られたデータの特徴を数量で表すために
 1. 基本統計量（平均，標準偏差，最大，最小）
 - AVERAGE, STDEV.P, MAX, MIN関数を利用
 2. 度数分布表（ヒストグラム）
 - FREQUENCY関数(Ctrl+ShiftキーでOK)
 3. 散布図
 - 「挿入」→グラフ
 4. 相関係数
 - CORREL関数
 - 関係を数値で判断

擬似相関の話（見かけの相関）

- 相関係数が高いからといって，因果関係があるわけではない
- 相関があるからといって，本当に関係があるかわからない
- ある事柄AとBに相関が認められたとしても，別の事柄Cによって「見かけの相関」が表れている場合がある



例)
X：算数の成績
Y：身長
Z：年齢

数値データで ない場合

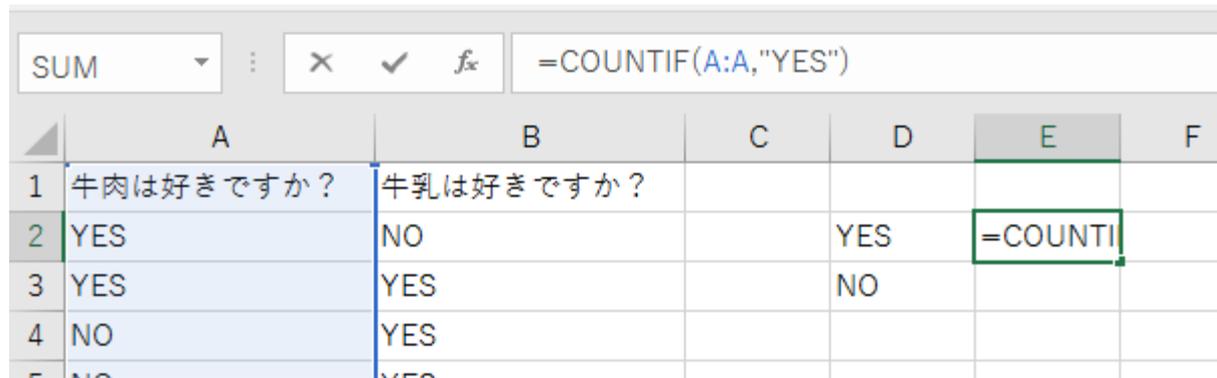
数値データでないもの（記号）

- アンケートなどで得られるデータ
- 順序なし
 - YES, NO
 - 男女
 - 血液型
 - 都道府県
- 順序あり
 - あてはまる, どちらでもない, あてはまらない
 - 好き, どちらでもない, 嫌い
 - 松, 竹, 梅

個数を数えるしか計算方法がない

個数を数える方法

- 関数の利用
 - =COUNTIF(範囲, 検索条件)



The screenshot shows an Excel spreadsheet with the following data:

| | A | B | C | D | E | F |
|---|-----------|-----------|---|-----|---------|---|
| 1 | 牛肉は好きですか？ | 牛乳は好きですか？ | | | | |
| 2 | YES | NO | | YES | =COUNTI | |
| 3 | YES | YES | | NO | | |
| 4 | NO | YES | | | | |
| 5 | NO | YES | | | | |

The formula bar at the top shows the formula: =COUNTIF(A:A,"YES"). The cell E2 is highlighted with a green border, indicating the active cell where the formula is being entered.

- ピボットテーブル
 - 単純集計
 - クロス集計

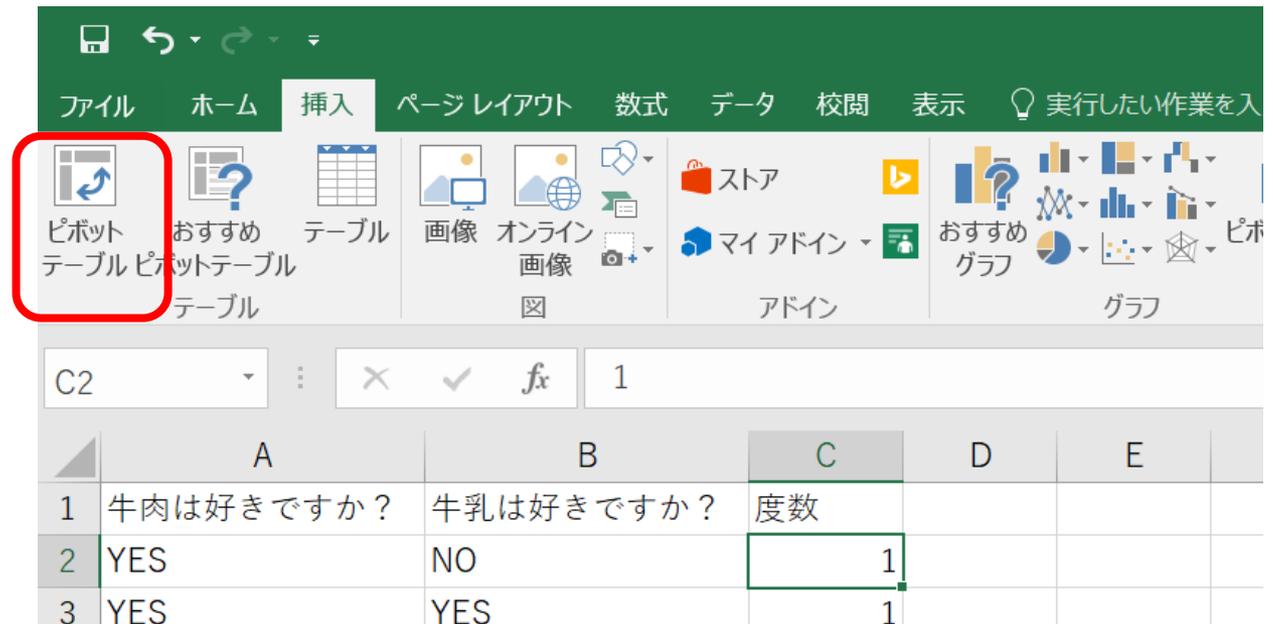
ピボットテーブル（その1）

- データの用意
- 数えたいデータ
- 度数の列を作り，数値 1 をコピー
- 注意
 - 並んでいること（列が空いていないこと）が必須

| | A | B | C | D | E |
|----|-----------|-----------|----|---|---|
| 1 | 牛肉は好きですか？ | 牛乳は好きですか？ | 度数 | | |
| 2 | YES | NO | 1 | | |
| 3 | YES | YES | 1 | | |
| 4 | NO | YES | 1 | | |
| 5 | NO | YES | 1 | | |
| 6 | YES | YES | 1 | | |
| 7 | YES | YES | 1 | | |
| 8 | YES | NO | 1 | | |
| 9 | YES | YES | 1 | | |
| 10 | YES | YES | 1 | | |
| 11 | YES | YES | 1 | | |
| 12 | NO | NO | 1 | | |
| 13 | YES | YES | 1 | | |
| 14 | NO | YES | 1 | | |
| 15 | YES | NO | 1 | | |
| 16 | YES | YES | 1 | | |
| 17 | NO | YES | 1 | | |
| 18 | YES | NO | 1 | | |
| 19 | YES | YES | 1 | | |
| 20 | YES | NO | 1 | | |
| 21 | YES | YES | 1 | | |

ピボットテーブル（その2）

- メニューの「挿入」
- 左はしの「ピボットテーブル」をクリック

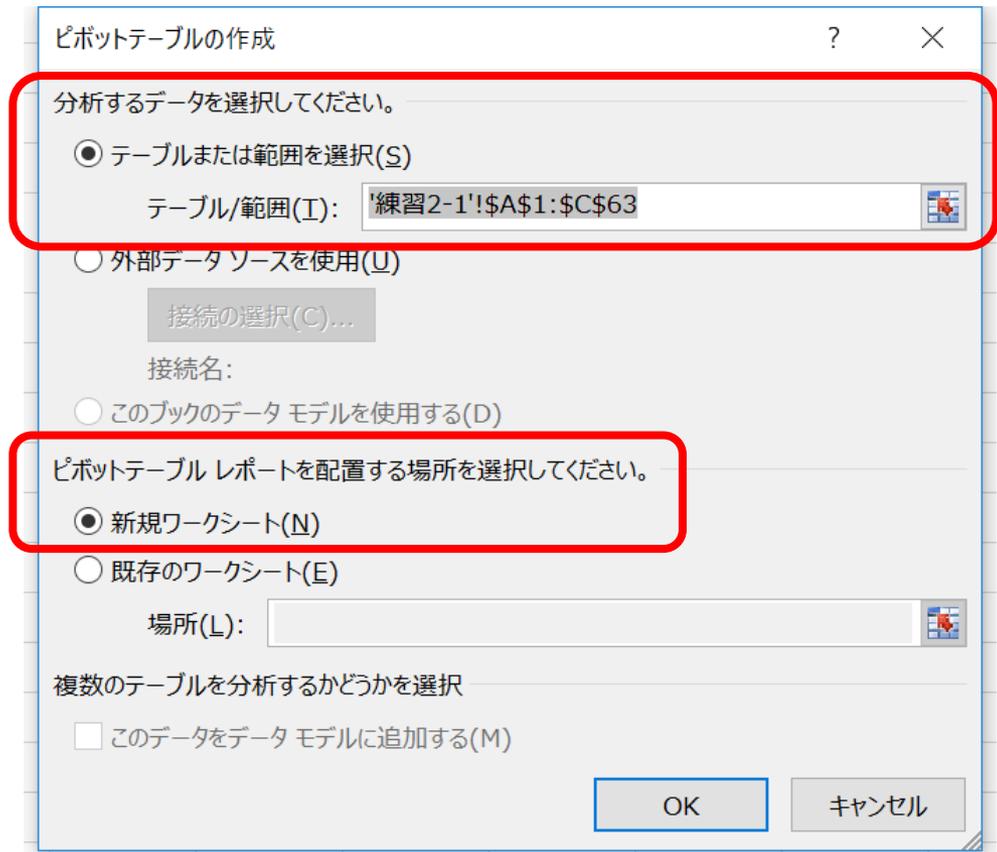


The screenshot shows the Microsoft Excel interface. The 'Insert' ribbon is active, and the 'PivotTable' icon (a grid with a circular arrow) is highlighted with a red box. Below the ribbon, the formula bar shows 'C2' and the value '1'. The spreadsheet data is as follows:

| | A | B | C | D | E |
|---|-----------|-----------|----|---|---|
| 1 | 牛肉は好きですか？ | 牛乳は好きですか？ | 度数 | | |
| 2 | YES | NO | 1 | | |
| 3 | YES | YES | 1 | | |

ピボットテーブル（その2）

- 上部
 - 範囲を選択
- 下部
 - 出力先は必ず「新規ワークシート」
 - 理由：失敗してもワークシートごと削除できるから



ピボットテーブル (その4)

- 単純集計
 - 数をかぞえる

「度数」を「値」へ
ドラック&ドロップ

| 行ラベル | 合計 / 度数 |
|------|---------|
| NO | 13 |
| YES | 49 |
| 総計 | 62 |

ピボットテーブル
レポートに追加するフィールドを指定してください:

検索

牛肉は好きですか?
 牛乳は好きですか?
 度数

その他のテーブル...

次のボックス間でフィールドをドラッグしてください:

| フィルター | 列 |
|-------|---|
| | |

| 行 | Σ 値 |
|----------|---------|
| 牛肉は好き... | 合計 / 度数 |

レイアウトの更新を保留する 更新

「Aのタイトル」を、下の
「行」へドラック&ドロップ

ピボットテーブル (その4)

- クロス集計
- 2つのデータ, それぞれに答えた人の数
 - YES-YES
 - YES-NO
 - NO-YES
 - NO-NO

| | | 牛乳は好きですか？ | | |
|-----------|-----|-----------|-----|----|
| | | NO | YES | 総計 |
| 牛肉は好きですか？ | NO | 3 | 10 | 13 |
| | YES | 13 | 36 | 49 |
| 総計 | | 16 | 46 | 62 |

「Bのタイトル」を「列」へドラック&ドロップ

The screenshot shows the Excel interface with a PivotTable. The PivotTable is structured as follows:

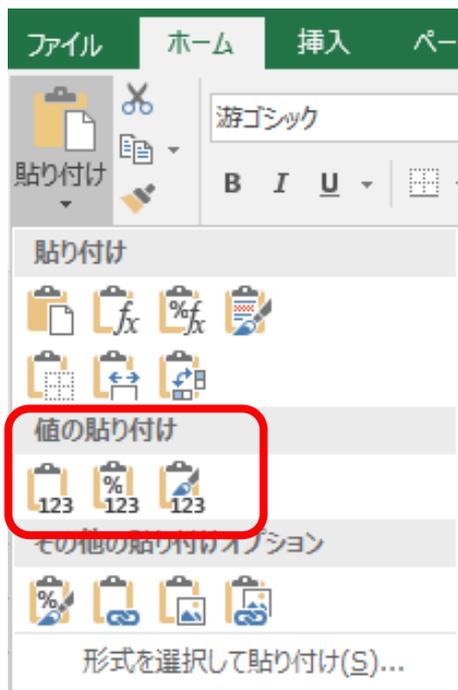
| 合計 / 度数 | 列ラベル | | |
|---------|------|-----|----|
| 行ラベル | NO | YES | 総計 |
| NO | 3 | 10 | 13 |
| YES | 13 | 36 | 49 |
| 総計 | 16 | 46 | 62 |

The PivotTable task pane on the right shows the following settings:

- レポートに追加するフィールドを選択してください:
- 検索:
- 牛肉は好きですか？
- 牛乳は好きですか？
- 度数
- その他のテーブル...
- 次のボックス間でフィールドをドラッグしてください:
- フィルター: [空]
- 列: 牛乳は好き...
- 行: 牛肉は好き...
- Σ 値: 合計 / 度数
- レイアウトの更新を保留する
- 更新

データのコピー（操作のため）

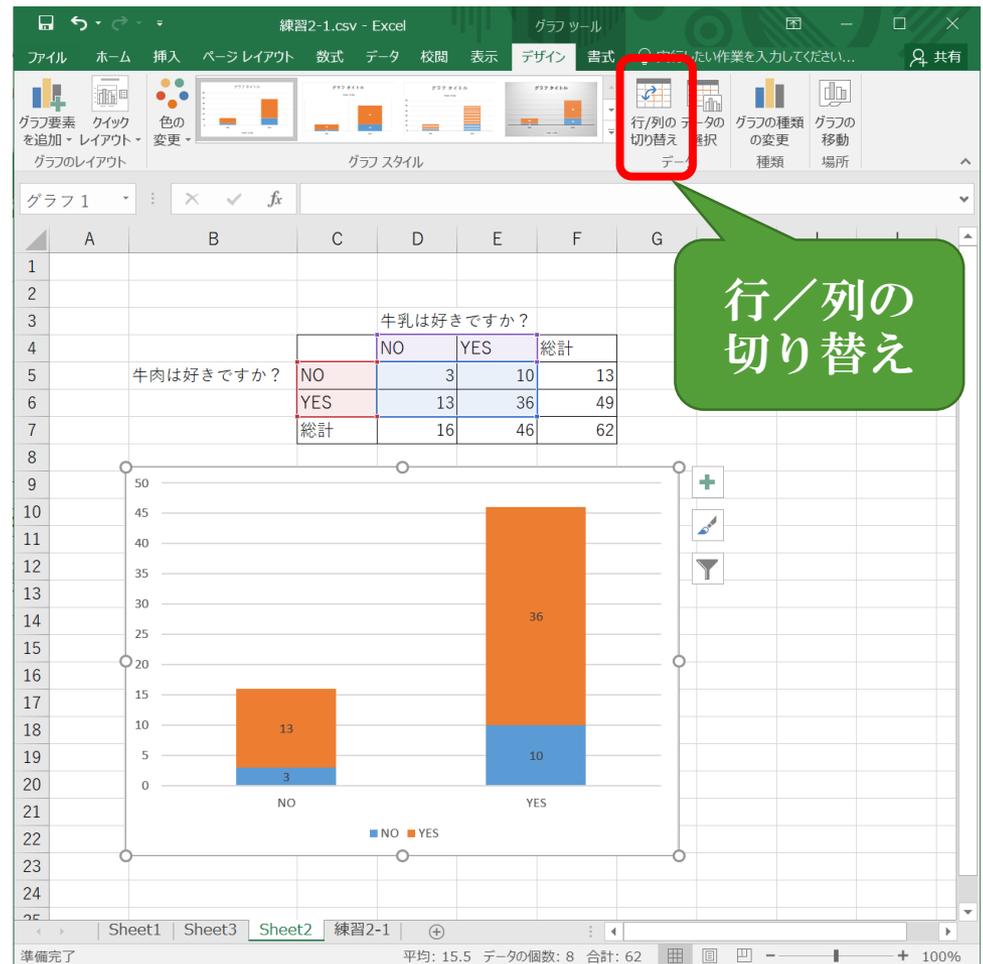
- ピボットテーブルのままでは扱いにくい
- 新しいワークシートへ「**値の貼り付け**」



| | | 牛乳は好きですか？ | | |
|-----------|-----|-----------|-----|----|
| | | NO | YES | 総計 |
| 牛肉は好きですか？ | NO | 3 | 10 | 13 |
| | YES | 13 | 36 | 49 |
| | 総計 | 16 | 46 | 62 |

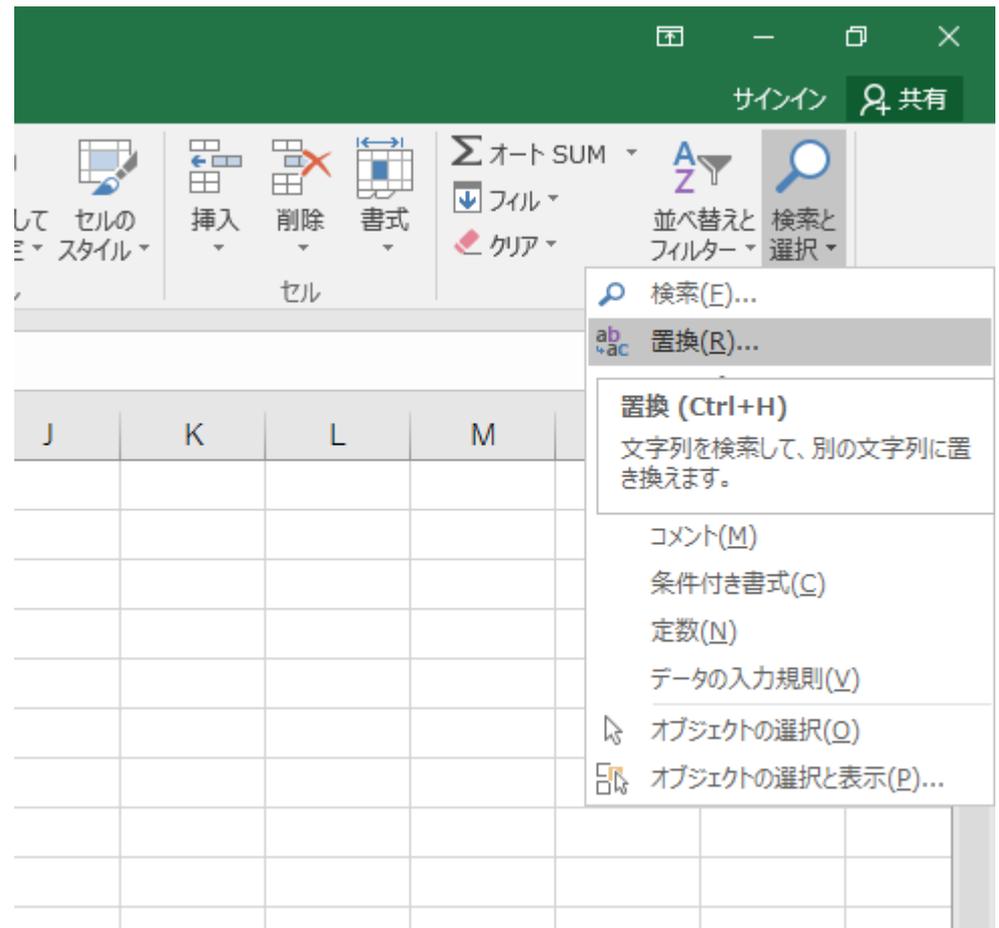
グラフ化（積み上げ棒グラフ）

- 「挿入」→「グラフ」→「積み上げ棒グラフ」
- 「行／列の切り替え」で表のとりのグラフ構成にできる



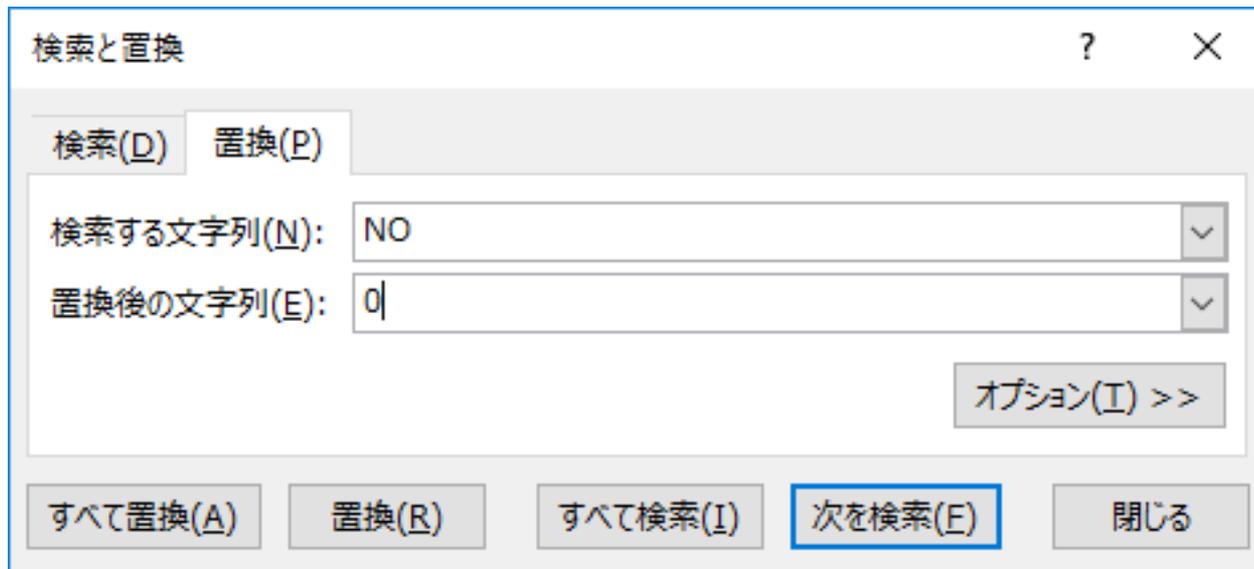
相関係数を算出するために

- YESを1, NOを0として、相関係数を計算
- 置き換え（おきかえ）を利用して一気に置換



置き換え

- NO → 0
- YES → 1
 - それぞれ「すべて置換」をクリック



相関係数

- 相関係数を求める関数（どちらも同じ）
 - = CORREL(データ1, データ2)
 - = PEARSON(データ1, データ2)
- データが0と1の場合の相関係数はファイ係数とも呼ぶ

| | A | B | C | D | E | F |
|---|-----------|-----------|------|---|----------|---|
| 1 | 牛肉は好きですか？ | 牛乳は好きですか？ | | | | |
| 2 | 1 | 0 | | | | |
| 3 | 1 | 1 | 相関係数 | | -0.03213 | |
| 4 | 0 | 1 | | | | |
| 5 | 0 | 1 | | | | |
| 6 | 1 | 1 | | | | |

| 相関係数の値 | 相関係数の強弱 |
|-----------|----------|
| 0.7~1 | 強い正の相関あり |
| 0.4~0.7 | 正の相関あり |
| 0.2~0.4 | 弱い正の相関あり |
| -0.2~0.2 | ほぼ関係ない |
| -0.4~-0.2 | 弱い負の相関あり |
| -0.7~-0.4 | 負の相関あり |
| -1~-0.7 | 強い負の相関あり |

まとめ

- 得られたデータの特徴を表にする
 1. クロス集計表
 - ピボットテーブル機能
 2. 積み上げ棒グラフ
 - 「挿入」→棒グラフ
 3. 数値へ置き換え
 - 「ホーム」→置換
 4. 相関係数
 - CORREL関数
 - 関係を数値で判断

今後の学習について

- 統計の入門書は山のようにあります
 - 気に入ったものを手に取ってみてください
- Webでおすすめ
- アイスクリーム屋さんで学ぶ 楽しい統計学
 - <http://kogolab.chillout.jp/elearn/icecream/index.html>