

表計算を使っのアンケート集計

専修大学商学部 高萩栄一郎

1 はじめに

この章は、Excel を使って、アンケートの集計を行う方法を実習します。実習は、紙 (図 1) または、Web (<http://163.136.122.41/enquete/enquete.htm>) でアンケートを行い、その結果を集計します。ただし、調査項目の 9 番は、AHP などと利用されている一対比較表から、固有値ベクトルや固有値をもとめ、評価項目の重要度や回答の信頼性を求めます。アンケートがどのように収集されたのかは、§5 をご覧ください。

この章では、すでにアンケート調査が行われており、その結果が入力されており、表計算ソフトウェアで利用可能になっている状態から説明を始めます。なお、アンケートのデータファイルは、

<http://www.senshu-u.ac.jp/~thc0456/text/>

からリンクを張ってあります。そこから、「練習用のデータ」をダウンロードしてください。

次のような項目についてアンケートが取られています。それぞれの項目について説明していきます。

アンケート番号 機械的につけられた回答票の番号です。

年度 1995 年から毎年行っていますので、年度を入力してあります。ただし、練習用のデータは 1995 年のものだけです。全データ 1995 です。

大学 専修大学と中央大学で行っています。専修大学は「S」、中央大学は、「C」と表示しています。

学年 学年は、授業の履修条件により偏っています。1995 年のデータでは、専修大学は 2 年生のみ、中央大学では 1 ~ 4 年生までです。

性別 男性は 1、女性は 2 が入力されています。

住居 自宅の場合 1、下宿の場合 2 です。

通学時間・身長・体重 記入された数値をそのまま入力してあります。

入学の種別 選択した番号を記入しています。1. 試験 (現役) 2. 試験 (浪人) 3. 推薦 (附属) 4. 推薦 (附属以外) となっています。

授業数・出席・遅刻・休講 記入された数値をそのまま入力してあります。

表計算ソフトウェアの理解度 1. セルにデータを入力 2. 足し算、引き算の計算式、3. 合計や平均の計算式、4. グラフの作成 (棒グラフ)、5. グラフの作成 (散布図、XY グラフ)、6. 絶対番地・相対番地の 6 つ項目について、「まったくできなかった」を 1、「あまりできなかった」を 2、「どちらともいえない」を 3、「理解した」を「4」、「完全に理解した」を「5」で記入してあります。

授業の評価基準 9 番のアンケート項目は、科目名の重要度、担当教員の重要度、成績評価の重要度、時間割の重要度の各項目の重要度を対にして比較しています (一対比較)。これを AHP という理論に当てはめて、それぞれの重要度を計算しています。4 つの重要度の合計は 1 になるようになっています。

重要度の回答の整合性 評価項目間を比較していると矛盾する回答をしている場合があります。例えば、科目名より担当教員、担当教員より成績評価を重要視し、成績評価より担当科目を重要視するといったものです (これは、ジャンケンの三すくみと同じ状態)。AHP では、整合性を計算できます。低ければ低いほど整合性が高いとされています。0.15 より高ければ、整合性が悪く利用しないほうがよいとされています。

情報処理演習アンケート

高萩栄一郎

以下のアンケートは、コンピュータ基礎演習（専修大学）、ゼミナール（専修大学）、情報処理演習（中央大学）での実習でアンケート処理の演習で使うデータを集めることを目的としています。アンケートは、無記名です。

1. 学年 () 年 性別 (男・女)
2. 住居 (自宅・下宿) 通学時間おおよそ () 分
3. 身長 () cm 体重 () kg (この部分は、後期、回帰分析で使います。)
4. 入学の種別 1.試験（現役） 2.試験（浪人） 3.推薦（附属） 4.推薦（附属以外）
5. 履修している授業数（前期・1週間あたり） () 時限
6. 先週、1週間で出席した授業数 () 時限 内遅刻 () 時限
7. 先週、1週間で休講になった授業数 () 時限
8. 表計算ソフトウェアをどれくらい理解したかたずねます。

該当する場所に○印を付けてください	まったくできない	あまりできない	どちらともいえない	理解した	完全に理解した
1.セルにデータを入力					
2.足し算、引き算の計算式					
3.合計や平均の計算式					
4.グラフの作成（棒グラフ）					
5.グラフの作成（散布図、XYグラフ）					
6.絶対番地・相対番地					

9. 授業の選択基準についておたずねします。選択する基準とその意味は、以下に示すものです。それぞれの選択基準について、どちらをどれくらい重要視しているか教えてください。
 1. 科目名 :科目の名称が自分の履修計画や自分の興味
 2. 担当教員 :担当教員の授業の進め方、授業内容（除：成績評価）
 3. 成績評価 :単位をとりやすい、いい成績がとりやすいこと
 4. 時間割 :自分の都合のよい曜日や時限に開講されていること

それぞれの評価基準について、比較して、どちらがどれくらい重要か○を付けてください。重要な程度がその中間のときは、間のマス目に丸を付けてください。	左の基準が絶対的に重要									右の基準が絶対的に重要								
	9	8	7	6	5	4	3	2	1	1/2	1/3	1/4	1/5	1/6	1/7	1/8	1/9	
科目名																		担当教員
科目名																		成績評価
科目名																		時間割
担当教員																		成績評価
担当教員																		時間割
成績評価																		時間割

図 1: アンケート用紙

2 計算式の設定

2.1 階層の設定

通学時間は分単位で、アンケートがとられています。しかし、場合によっては、階層を設定して、そちらを利用した場合がよいことがあります。そこで、表 1 のような対応にします。

1. 「通学時間」の右隣に「通学時間階層」の列を作成します。しかし、すでに「身長」の列が使用しているので新たに列を挿入して作成します。

(a) H 列をクリック

(b) メニューの 挿入 → 列

(c) H1 を「通学時間階層」に設定

表 1: 通学時間階層の対応表

通学時間	通学時間階層
0 ~ 14	1
15 ~ 29	2
30 ~ 59	3
60 ~ 89	4
90 ~ 119	5
120 ~ 149	6
150 ~	7
空白	空白

表 2: VLOOKUP 用通学時間階層の対応表

通学時間	通学時間階層
0	1
15	2
30	3
60	4
90	5
120	6
150	7

2. 表 2 のような VLOOKUP 用の通学時間と通学時間階層の対応表を作成します。ここでは左上のセル(「通学時間」セル)を AE1 にしました。
3. VLOOKUP 関数を使い、通学時間階層を設定
 - (a) - (b) 上の設定だと、通学時間が記述されていないとき(空白のセルのとき)、不都合が起こります。通学時間が空白のときは空白を表示するように設定します。
 - (c) - (d) 計算式を複写します。 H2 H3:H162

2.2 欠席数・出席率などの計算

ここでは、「欠席数」、「出席率」、「欠席率」、「遅刻率」を計算します。
 まず、O 列の右に 4 列挿入し、それぞれに名前をつけます。

1. P 列をクリック
2. メニューの →
3. 2. を 4 回繰り返す。

欠席数(授業数 - 出席数 - 休講数)の計算を行います。

$$\text{P2:} = \text{L2} - \text{M2} - \text{O2}$$

P2 P3:P162

負の値などおかしなデータがありますが、おかしなデータは §3 で対処します。

2.2.1 練習問題

1. Q 列に出席率: $\frac{\text{出席数}}{\text{授業数} - \text{休講数}}$ を計算しましょう。
2. R 列に欠席率: $\frac{\text{欠席数}}{\text{授業数} - \text{休講数}}$ を計算しましょう。
3. S 列に遅刻率: $\frac{\text{遅刻数}}{\text{出席数}}$ を計算しましょう。ただし、出席数が 0 のときは、空白にするようにします (IF 関数を使い、出席数が 0 より大きいときのみ計算するようにします)。

3 データのチェック

表を見ていると、欠席数が負になっていたり、出席率などに 0 で割り算をした印「#DIV/0!」が表示されていたりします。このようなデータは、授業数から遅刻率までのデータは、使わないこととします。そこで、おかしなデータ部分とそれに対応する部分のセルをクリアして使わないようにします。

これから、範囲 A1:AD162 は、しばしば利用するので、「DATABASE」という名前をつけておきます。

- A1:AD162 を範囲指定
- メニューの **挿入** → **名前** → **定義**
- 「DATABASE」と入力し、**追加** をクリックし、**OK** をクリック

3.1 欠席数などのチェック

欠席数は、授業数 - 出席数 - 休講数で計算しました。そこで、欠席数が負の値になっていたらおかしなデータです。そこで、欠席数が負のデータをクリアします。

1. ソート（並べ替え）を使ったり、セルの値をクリアしたりします。念のため、一度保存をしておきましょう。
2. 欠席数でソートします。
 - (a) 名前ボックスを「DATABASE」にします。
 - (b) メニューの **データ** → **並べ替え**
 - (c) 範囲の先頭行：タイトル行
最優先されるキー：欠席数, 昇順
 - (d) **OK**
3. 欠席数が負のサンプルが上の方に集まったと思います。2 行目から 14 行目のサンプルが負になっています。
4. 2 行目から 14 行目までで、授業数関係のセルをクリアします。
 - (a) おかしなデータの部分 (L2:S14) を範囲指定します。
 - (b) **Delete** キーを押す。

同じように、今度は、授業数について空白のサンプルをクリアします。

1. 名前ボックスを「DATABASE」にし、「授業数」を最優先するキー（昇順）としてソートする。
2. 146 行目から 162 行目までが授業数が空白などでおかしなデータだということがわかります（先ほどの欠席数でクリアしたデータも含まれます）
3. L146:S162 を範囲指定して、**Delete** キーを押す。

3.1.1 練習問題

出席数、遅刻数、休講数についても、授業数と同様におかしなデータ（空白のデータ）がないかどうか、ソートを使って調べ、あったらクリアしましょう。

3.2 重要度の整合性

重要度の整合性は、0.15 より大きいと使うことはできません。§3.1 と同様に、対応する部分をクリアします。

1. 名前ボックスを「DATABASE」にし、「重要度の回答の整合性」を最優先するキー（昇順）としてソートする。
2. 「重要度の回答の整合性」の値が0.15 より大きな行の各評価項目の重要度は、利用できないのでその部分をクリアする。

(a) Z64:AD164 を範囲指定して、**Delete** キーを押す。

最後にアンケート番号順にソートをして元の順番に戻しておきましょう。

4 集計

4.1 ピボットテーブル

集計（1次元集計）やクロス集計（2次元集計）を行うには、ピボットテーブルを使うと便利です。例として、大学別（Sは専修大学、Cは中央大学）、住居別（1:自宅、2:下宿）での通学時間の平均値を図5のような表を求め、図7のようなグラフを作成します。ピボットテーブルを用いて計算します。

まず、図5の表を作成します。

1. 名前ボックスを「DATABASE」にします。
2. メニューの **データ** → **ピボットテーブルとピボットグラフレポート**
3. 出力先は、図2のように「新規ワークシート」にします。
- 4.

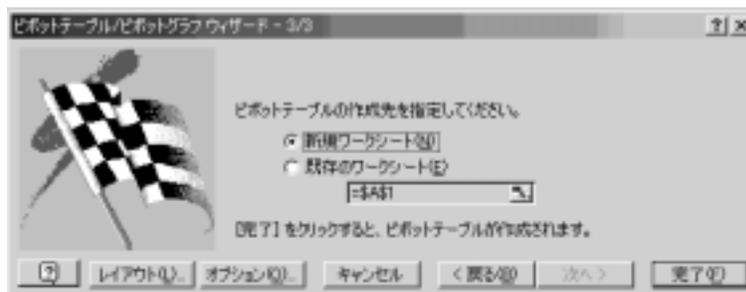


図 2: 出力位置の設定

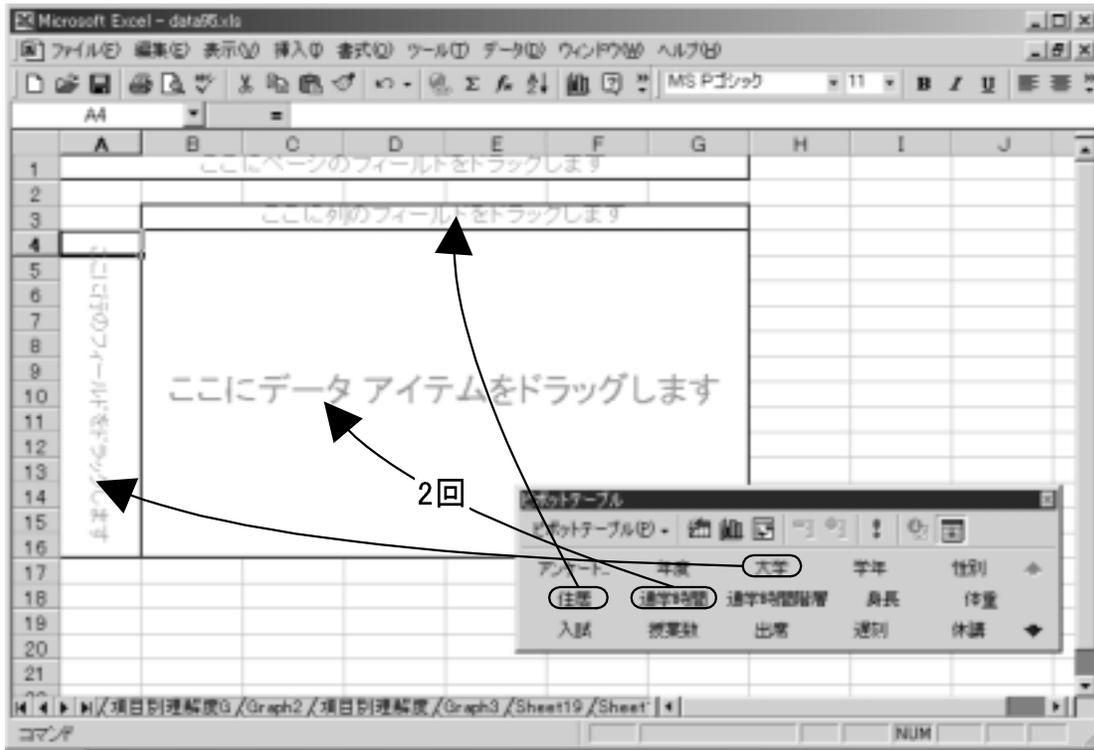


図 3: ピボットテーブルの設定

図 3 の設定は、次のようにします。

1. **大学** のボタンをドラッグして、行の領域にドロップする。
2. **住居** のボタンをドラッグして、列の領域にドロップする。
3. **通学時間** のボタンをドラッグして、データの領域にドロップする（2回：個数と平均値の2つ求めるため）
4. 図 4 のような画面が表示されたと思います。
5. 図 4 のデータの領域の 1 つめの通学時間の部分を右クリックし、**フィールドの設定** を選択します。集計の方法を「数値の個数」にして、**OK** をクリックします。
6. 図 4 のデータの領域の 2 つめの通学時間の部分を右クリックし、**フィールドの設定** を選択します。集計の方法を「平均」にして、**OK** をクリックします。

		住居		
大学	データ	1	2 (空白)	総計
C	データの個数: 通学時間	41	19	60
	平均: 通学時間2	93.90243902	36.21052632	75.63333333
S	データの個数: 通学時間	56	42	1
	平均: 通学時間2	86.51785714	31.42857143	15
全体のデータの個数: 通学時間		97	61	1
全体の平均: 通学時間2		89.63917526	32.91803279	15

- この部分を右クリック
- 「フィールドの設定」をクリック
- 「数値の個数」を選択

- この部分を右クリック
- 「フィールドの設定」をクリック
- 「平均」を選択

図 4: ピボットテーブルの設定 2

図 5 のような集計表が表示されたと思います。

		住居		
大学	データ	1	2 (空白)	総計
C	数値の	41	19	60
	平均: 通学時間	93.90243902	36.21052632	75.63333333
S	数値の	56	42	99
	平均: 通学時間	86.51785714	31.42857143	62.42424242
全体の 数		97	61	159
全体の 平均: 通学時間		89.63917526	32.91803279	67.40880503

図 5: ピボットテーブルの計算結果

上段にそれぞれのカテゴリーに当てはまるデータの個数が下段にその平均が表示されています。例えば、中央大学で下宿の学生で、通学時間を回答したのは、19 件で平均約 36 分であることを示しています。また、空白の列は住居に関して回答しなかった（空白）データの集計値です。この場合、住居未回答、専修大学で、通学時間を回答したのは 1 件で、その平均は 15 分であることを示しています。

4.2 棒グラフの作成

図 5 のままだと縦横の項目数がそれぞれ 2 つ少ないので、図 7 のようなグラフを作成するために、図 6 のような表を作成し、棒グラフを作成します。ただし、93.90... のところは、「=C4」のような式で入力してください。

中央大学・自宅	93.90243902
中央大学・下宿	36.21052632
専修大学・自宅	86.51785714
専修大学・下宿	31.42857143

図 6: 棒グラフを作成するための準備

棒グラフを作成するには、次のようにします。

1. 図 6 の範囲を範囲指定します。
2. **グラフウィザード** のボタンをクリック
3. グラフの種類は、「横棒」を選択し、形状は、左上のものを選択
4. **次へ** を 2 回クリック
5. グラフタイトルや凡例を設定
6. **次へ** をクリック
7. グラフの場所は「新しいシート」にする。
8. **完了** をクリック

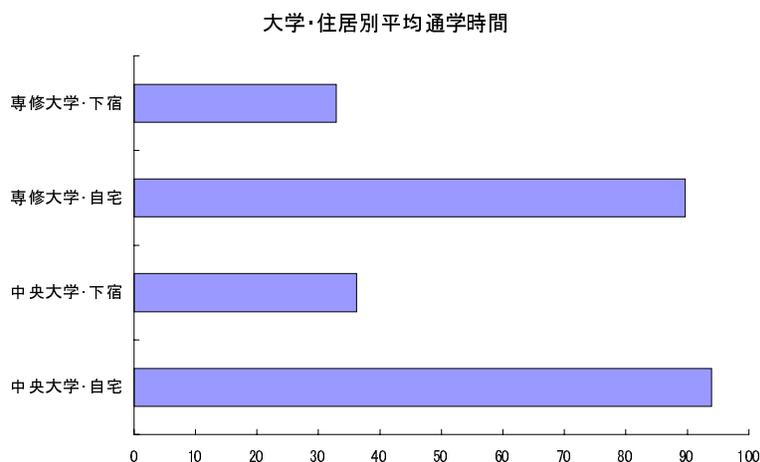


図 7: 大学住居別の平均通学時間のグラフ

シートが増えてきたので、それぞれに名前をつけましょう。例えば、Sheet1 は「データ」、Sheet2 は「平均通学時間(表)」、Graph1 は「平均通学時間(グラフ)」としてみます。

4.3 男女・通学時間階層別の遅刻率

4.1 と同様にピボットテーブルを使って、集計表を作りましょう。重要な点を繰り返すと次のようになります。

範囲指定 名前ボックスを「DATABASE」にする。

また、「同じ元のデータから...」というメッセージが表示。この場合、「はい」を選んでください(ただし、元のデータを変更した場合は、「いいえ」を選んでください)。

「使用するデータを含むピボットテーブルを選択してください」と表示。この場合「[ENQ95.XLS] 平均時間表!ピボットテーブル」を選んでください。

行にドロップするボタン

列にドロップするボタン (通学時間のボタンは2つのように見えますがマウスをあてると「通学時間階層」と表示されるほうです)

データにドロップするボタン

集計の方法を「数値の個数」に変更

集計の方法を「平均」に変更

		通学時間階層							
性別	データ	1	2	3	4	5	6	7	総計
1	数値の個数: 遅刻率	7	14	10	21	12	11	1	76
	平均: 遅刻率	0.230612245	0.041865079	0.148311688	0.081246531	0.131746957	0.034848485	0.090909091	0.097959168
2	数値の個数: 遅刻率	1	3	8	7	8	6	2	37
	平均: 遅刻率	0.142857143	0.03030303	0.106894841	0.110047096	0.104166667	0.173789174	0.038461538	0.103033678
■		8	17	18	28	20	17	3	113
全体の平均: 遅刻率		0.219642857	0.039824718	0.129904201	0.088446672	0.120714841	0.083686375	0.055944056	0.099620733

図 8: 遅刻率の集計結果

図 8 を見ると、男性の階層 7, 女性の階層 1, 3, 7 の数値の個数が少ないことに注意しなくてはなりません。平均をとっているため、1 つ極端に大きなデータや小さなデータがあるとそのデータに引っ張られて、過度に大きくなったり、小さくなったりします。男性の階層 7 は、1 件だけなので、ある 1 つのデータの値のみで決まってしまう。

§4.2 と同様に、グラフ用の表を作成し、その表から新しいシートにグラフを作成します。

通学時間階層	1(0~14)	2(15~29)	3(30~59)	4(60~89)	5(90~119)	6(120から149)	7(150~)
男性	0.230612245	0.041865079	0.148311688	0.081246531	0.131746957	0.034848485	0.090909091
女性	0.142857143	0.03030303	0.106894841	0.110047096	0.104166667	0.173789174	0.038461538

図 9: 折れ線グラフを作成するための準備

通学階層別の遅刻率

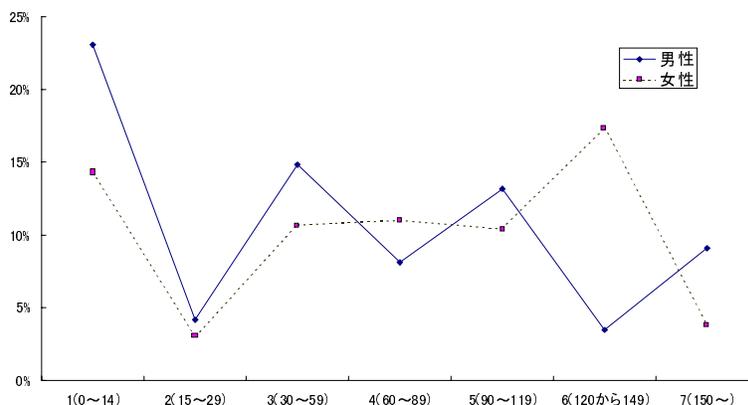


図 10: 男女・通学時間階層別の遅刻率の折れ線グラフ

このグラフから、第1階層（通学時間15分未満）で遅刻率が高く、第2階層（通学時間15分以上30分未満）で遅刻率がやや低いことがわかり、他は、ほぼ同じであることがわかります。ただし、女性の第1,2,7階層、男性の第7階層のデータの個数（サンプル数）は、5未満と少ないので、それらの値は、あまり信頼できません。

4.4 理解度の集計

理解度の集計は、それぞれの番号（1から5まで）につけた件数を数えます。その場合次のように、列に「1.セルにデータを入力」として、データにも「1.セルにデータを入力」（集計方法：データの個数）とします。男女別などでも集計したいときは、行にその項目のボタンをドロップします。

ピボットテーブルの設定で注意する点は以下のとおりです。

範囲指定 名前ボックスを「DATABASE」にする。

また、「同じ元のデータから...」というメッセージが表示 この場合、 はい を選んでください。

「使用するデータを含むピボットテーブルを選択してください」と表示 この場合、2つうちどちらでもよい。

列にドロップするボタン 1.セルにデータを入力

データにドロップするボタン 1.セルにデータを入力 集計の方法は「データの個数」のまま

ピボットテーブルの作成先 新しいシート

「2.足し算、引き算の計算式」も同様に行います。ただし、ピボットテーブルの作成先は、「1.セルにデータを入力」と同じシートの適当な場所（例えば「Sheet6!A5」）にします。

同様に、「3.合計や平均の計算式」、「4.グラフの作成（棒グラフ）」、「5.グラフの作成（散布図、XYグラフ）」、「6.絶対番地・相対番地」を集計します。結果は、図11のようになります。

データの個数: 1.セルにデータを入力	1.セルにデータを入力					
計	1	2	3	4	5(空白)	総計
	3	6	21	61	68	159
データの個数: 2.足し算、引き算の計算式	2.足し算、引き算の計算式					
計	1	2	3	4	5(空白)	総計
	1	5	17	71	61	159
データの個数: 3.合計や平均の計算式	3.合計や平均の計算式					
計	1	2	3	4	5(空白)	総計
	1	11	28	73	44	157
データの個数: 4.グラフの作成(棒グラフ)	4.グラフの作成(棒グラフ)					
計	1	2	3	4	5(空白)	総計
	3	5	19	95	36	158
データの個数: 5.グラフの作成(散布図,XYグラフ)	5.グラフの作成(散布図,XYグラフ)					
計	1	2	3	4	5(空白)	総計
	4	23	42	71	18	158
データの個数: 6.絶対番地・相対番地	6.絶対番地・相対番地					
計	1	2	3	4	5(空白)	総計
	10	24	21	14	3	72

図 11: 理解度の集計結果（5つのピボットテーブル）

グラフ化のために、12のような表を作成します。

	まったくできなかつた	あまりできなかつた	どちらともいえない	理解した	完全に理解した
1.セルにデータを入力	3	6	21	61	68
2.足し算、引き算の計算式	1	5	17	71	61
3.合計や平均の計算式	1	11	28	73	44
4.グラフの作成(棒グラフ)	3	5	19	95	36
5.グラフの作成(散布図、XYグラフ)	4	23	42	71	18
6.絶対番地・相対番地	10	24	21	14	3

図 12: 理解度の集計結果 (5 つのピボットテーブルをまとめたもの)

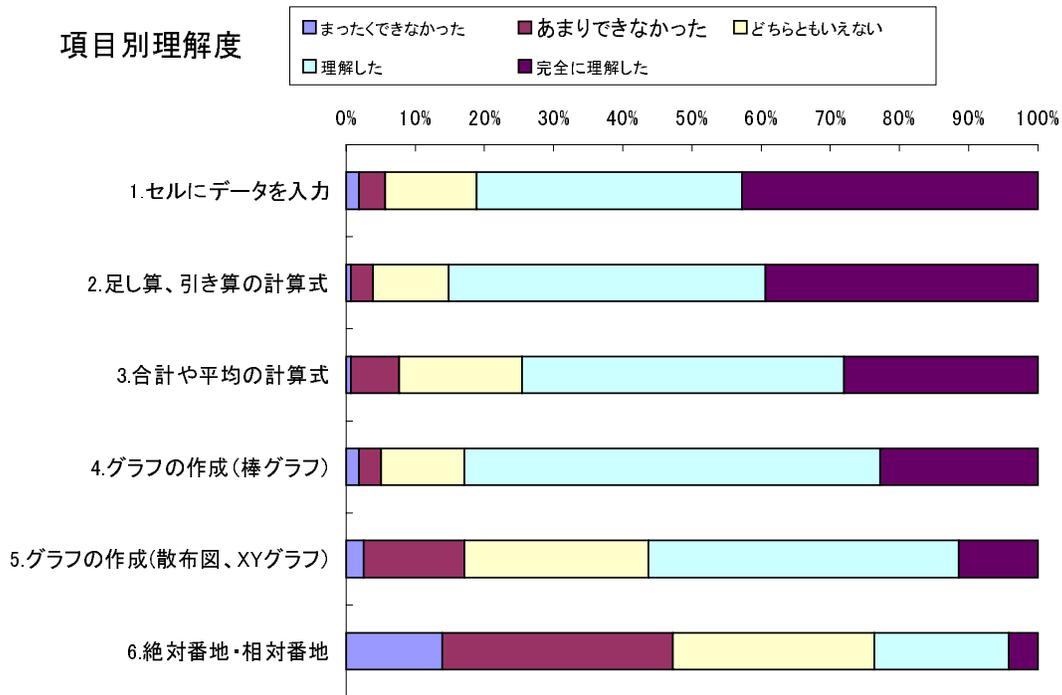


図 13: 項目別の理解度

X 軸の項目の表示順番を逆転させるには、項目軸をダブルクリックし、**目盛**のタグの中の「軸を反転させる」のチェックボックスを ON にします。

また、1つの項目のみ注目するならば、円グラフを使うとよいでしょう。

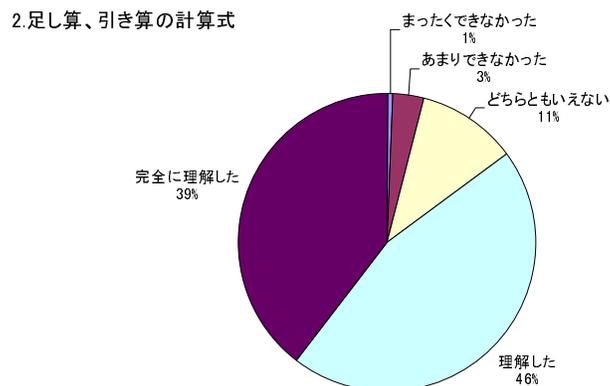


図 14: 合計や平均の計算式の理解度

4.5 重要度の集計

この節では、複数の項目についての平均値を求めます。具体的には、男女別の「科目名の重要度」...「時間割の重要度」を求めます。

範囲指定 名前ボックスを「DATABASE」にする。

また、「同じ元のデータから...」というメッセージが表示 この場合、 はい を選んでください。

「使用するデータを含むピボットテーブルを選択してください」と表示 この場合、どれでもでもよい。

列にドロップするボタン 性別

データにドロップするボタン

科目名の重要度 集計の方法:「数値の個数」

科目名の重要度 集計の方法:「平均」

担当教員の重要度 集計の方法:「平均」

成績評価の重要度 集計の方法:「平均」

時間割の重要度 集計の方法:「平均」

ピボットテーブルの作成先 新しいシート

データ	性別		総計
	1	2	
数値の個数: 科目名の重要度	37	25	62
平均: 科目名の重要度	0.251248995	0.178990072	0.222112333
平均: 担当教員の重要度	0.158737335	0.169956389	0.163261147
平均: 成績評価の重要度	0.397288692	0.440112174	0.414556225
平均: 時間割の重要度	0.192724988	0.210941367	0.200070302

図 15: 各評価基準の重要度

図 16 は、各評価基準の重要度のグラフを作成するための表です。

	男性	女性	合計
科目名の重要度	25.12%	17.90%	22.21%
担当教員の重要度	15.87%	17.00%	16.33%
成績評価の重要度	39.73%	44.01%	41.46%
時間割の重要度	19.27%	21.09%	20.01%

図 16: 各評価基準の重要度 (グラフ作成用)

グラフを作成するとき、「データの系列」(列か行)を変更する必要があるかもしれません。

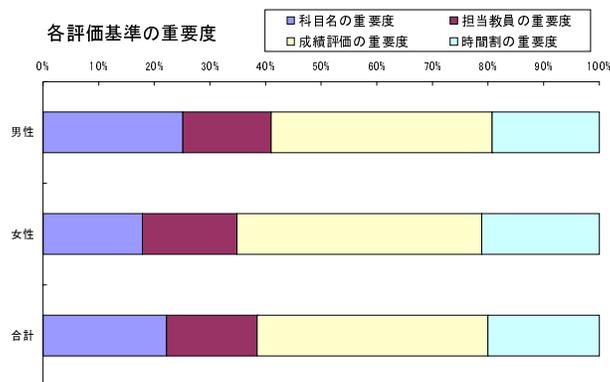


図 17: 各評価基準の重要度のグラフ

5 報告書の作成

アンケート調査がどのように行われたかを記述します．実習用のデータでは以下ようになります．

- アンケートの目的
 - － コンピュータ基礎演習（専修大学），ゼミナール（専修大学），情報処理演習（中央大学）での実習でアンケート処理の演習で使うデータを集めること
 - － 授業での理解度を調べ，今後の授業の進め方の参考にすること
 - － 学生の授業選択の意識構造を調べること
- アンケート対象
 - － 中央大学商学部情報処理演習Ⅰを履修している学生 60 名
 - － 専修大学商学部コンピュータ基礎演習を履修している学生 101 名
 - － 計 161 名
- アンケート方法
 - － 授業時間中配布し，その場で回収

次に，アンケートの大きなまとめりに章建てをし，表やグラフを使って分析をします．