

第 5 章

表計算を使っのアンケート集計

2025 年 12 月 31 日

学習目標

- (1) アンケートデータを整理する.
- (2) アンケート結果を集計をする.
- (3) 集計結果を適切なグラフで表示し，考察する.

本章は，専修大学商学部の高萩栄一郎の著作である．

1 はじめに

Excel を使って、アンケートの集計を行う方法を実習します。すでに下記のような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。データファイルは、[テキストのページ](#) からダウンロードできます (enq_text2014_2021.xlsx(アンケートファイル, 教科書))。このテキストで説明しているのは、このファイルです。別に、2014 年度～2023 年度までのファイルがあります。

2 アンケートの概要

アンケートは、GoogleDrive のフォーム ([アンケートの画面の PDF ファイル](#)) で実施しました。主に専修大学の「情報入門」「情報基礎」の履修者に行いました。サンプルの件数は 9870 件です。

所属学部 「経済学部」, 「法学部」, 「経営学部」, 「商学部」, 「文学部」, 「ネットワーク情報学部」, 「人間科学部」, 「国際コミュニケーション学部」 から回答 (1 つだけ選択, 必須の回答項目, 質的データ)

性別 「男性」, 「女性」 から回答 (1 つだけ選択, 必須の回答項目, 質的データ)

現在の住居 「一人暮らし (含む: 寮)」, 「家族と一緒に (実家など)」 から回答 (1 つだけ選択, 必須の回答項目, 質的データ)

ペットの飼育経験 犬, 猫, ウサギ, ハムスター について, それぞれ, 「現在飼育中 (含: 実家)」, 「飼育した経験あり」, 「飼育したことはない」 から選択 (それぞれ 1 つだけ選択, 必須の回答項目, 質的データ)

各種類のペットについて「かわいい」と思う程度 犬, 猫, ウサギ, ハムスター について, それぞれ, 「かわいくない」, 「どちらかといえばかわいくない」, 「どちらかといえない」, 「どちらかといえばかわいい」, 「かわいい」 から選択 (それぞれ 1 つだけ選択, 必須の回答項目, 順序データ)。

ペットを購入するとしたら, いくらくらいまで出そうと思いますか? 対象のペット 「犬」, 「猫」, 「ウサギ」, 「ハムスター」,

「この中にはない」, 「ペットは売買の対象とは考えない」から選択（それぞれ 1 つだけ選択, 必須の回答項目, 質的データ）

ペットの想定支出金額 「その金額を回答してください. 前の質問で, 「この中にはない」, 「ペットは売買の対象ではない」と回答したときは, この回答は不要です.」（任意の回答項目, 量的データ）として, 対象ペットの想定支出金額を聞いています.

※当初, 必須の回答, 数値のみの回答などの制約をしないでアンケート実施しました. その後, 制約をかけました. サンプルのデータは, 制約をかけたものと同様になるように調整しました. 各調査年を 1 列目に追加しました. 回答者の ID, タイムスタンプ（回答した日時）も収集していますが, この 2 つの項目は削除したものを提供します.

2023 年度より, 「ペットに関する意見や思いを記述してください.」という設問で自由記述欄を追加しました. この部分の分析は, 第 9 節で行います.

3 記号, 文字列への変換

図 1 は, GoogleDrive のスプレッドシートに格納されたアンケートデータで, アンケートファイルのシート「enqd」に入っています（2014～2022 年度のデータ, 自由記述欄への回答は含まれていません）.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	調査年	所属学部	性別	現在の住居	ペットの飼育経験 [犬]	ペットの飼育経験 [猫]	ペットの飼育経験 [ウサギ]	ペットの飼育経験 [ハムスター]	各種類のペットについて「かわいい」と思う程度を回答してください。 [犬]	各種類のペットについて「かわいい」と思う程度を回答してください。 [猫]	各種類のペットについて「かわいい」と思う程度を回答してください。 [ウサギ]	各種類のペットについて「かわいい」と思う程度を回答してください。 [ハムスター]	ペットを購入すると思ったら、いくらくらいまで出そうと思いますか？	その金額を回答してください。
2	2014	経済学部	男性	一人暮らし(含む:寮)	飼育したことはなし	飼育したことはなし	飼育したことはなし	飼育したことはなし	かわいくない	かわいくない	かわいくない	かわいくない	犬	10000
3	2014	経済学部	男性	家族と一緒に(実家など)	飼育したことはなし	飼育したことはなし	飼育したことはなし	飼育したことはなし	どちらかといえばかわいい	どちらかといえばかわいい	どちらかといえない	どちらかといえない	ペットは元々の対象とは考えない	
4	2014	経済学部	男性	家族と一緒に(実家など)	飼育したことはなし	飼育したことはなし	飼育したことはなし	飼育したことはなし	かわいい	かわいい	かわいい	かわいい	犬	100000
5	2014	文学部	男性	家族と一緒に(実家など)	飼育したことはなし	飼育したことはなし	飼育したことはなし	飼育したことはなし	どちらかといえばかわいい	どちらかといえばかわいい	どちらかといえばかわいい	どちらかといえばかわいい	この中にはない	

図 1 収集したアンケートデータ (一部)

データ数が多くセルの移動が大変で、連続しているデータの移動や範囲指定は下記ショートカットを利用すると便利です。

右端まで移動 : (End) (→)

最下行まで移動 : (End) (↓)

最上行まで移動 : (End) (↑)

右端まで範囲指定 : (ctrl) + (Shift) + (→)

最下行まで範囲指定 : (ctrl) + (Shift) + (↓)

一部のノート PC では、(End)キーは、(fn)+(→) ((fn)キーを押しながら(→)キーを入力)や(fn)+(↑)で代用できます。

したがって、この場合、最下行まで移動は、(fn)キーを押しながら(→)キーを入力して、(↓)を入力します。

下方に移動すると、1 行目が見えなくなります。先頭行を固定すると見やすくなります。

メニューの 表示 → ウインドウ → ウインドウ枠の固定 → 先頭行の固定

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	NO	調査年	所属学部	性別	現在の住居	飼育経験 [犬]	飼育経験 [猫]	飼育経験 [ウサギ]	飼育経験 [ハムスター]	かわいさ [犬]	かわいさ [猫]	かわいさ [ウサギ]	かわいさ [ハムスター]	対象ペット	金額
2	1	2014	経済学部	男性	一人暮らし	なし	なし	なし	なし	1	1	1	1	犬	10,000
3	2	2014	経済学部	男性	家族と一緒に	なし	なし	なし	なし	4	4	3	3	対象外	
4	3	2014	経済学部	男性	家族と一緒に	なし	なし	なし	なし	5	5	5	5	犬	100,000
5	4	2014	文学部	男性	家族と一緒に	なし	なし	なし	なし	4	4	4	4	他	
6	5	2014	経済学部	男性	一人暮らし	なし	なし	なし	なし	3	4	1	1	猫	50,000
7	6	2014	人間科学部	女性	一人暮らし	なし	なし	なし	なし	5	5	4	5	他	30,000
8	7	2014	経済学部	男性	一人暮らし	飼育中	なし	なし	なし	5	5	5	5	犬	100,000
9	8	2014	経済学部	男性	一人暮らし	なし	なし	なし	なし	1	4	3	4	対象外	
10	9	2014	経済学部	男性	一人暮らし	なし	なし	なし	経験あり	5	5	4	3	犬	0

図2 分析用シートの完成例（一部分）

3.1 分析用ワークシートの作成と NO, 学部, 性別

図1は、回答は、選んだ選択肢の名称で記述されていて、長い文字列で、集計に手間取ります。そこで、図2のように、短い文字列や記号、番号に変換します。

- (1) 新しいワークシート、たとえば、「alldata」を作成します。
- (2) 1行目を表頭にします。図2のように enqd からコピーしたり、入力したりします。
- (3) A列は回答の番号を振ります。「NO」の列に、オートフィル機能を使って1から9870まで順番に番号を振ります。
(9871行目まで)
 - (a) A2に1, A3に2, A4に3 を入力

(b) A2 から A9871 を範囲指定

(c) メニューの → → →

(d) そのまま をクリック

(4) 「調査年」「所属学部」「性別」はそのまま、シート「enqd」の値を「=」を使った計算式で代入します。

=enqd!A2

=enqd!B2

=enqd!C2

B2:D2 B3:D9871

複写先の上下移動は、ショートカットキーを使うと便利です。

動画 enqNN_cp_wv.mp4:アンケート ワークシートの作成 1(音声付き)

動画 enqNN_noetc_wv.mp4:アンケート ワークシートの作成 2(音声付き)

動画 enqNN_cp.mp4:アンケート ワークシートの作成 1(音声なし)

動画 enqNN_noetc.mp4:アンケート ワークシートの作成 2(音声なし)

3.2 変換表による設定

「一人暮らし（含む：寮）」を「一人暮らし」, 「家族と一緒に（実家など）」を「家族と一緒に」と表示します。Vlookup 関数を利用します。ワークシート「convT」に、変換表を作成してあります。

現在の住居を変換表を用いて変換. 絶対参照を利用して設定します。スピルを使うことも考えられますが、ペットの飼育経験などでの再利用（コピーして式を編集）を想定して、絶対参照を使用します。

E2: =VLOOKUP(enqd!D2,convT!\$B\$2:\$C\$3,2,FALSE)
 複写元: E2 複写先: E3:E9871

動画 enqNN_v1.mp4: アンケート Vlookup 関数による住居の設定

D2 の計算式で、convT!\$B\$2:\$C\$3 の部分は、(F4)キーなどを使って絶対参照(\$マークを付ける)にするのを忘れないようにしましょう。


	A	B	C
1		回答	記号
2		一人暮らし(含む:寮)	一人暮らし
3		家族と一緒に(実家など)	家族と一緒に
4			
5		回答	記号
6		現在飼育中(含:実家)	飼育中
7		飼育した経験あり	経験あり
8		飼育したことはない	なし
9			
10			
11		回答	記号
12		かわいくない	1
13		どちらかといえばかわいくない	2
14		どちらといえない	3
15		どちらかといえばかわいい	4
16		かわいい	5
17			
18			
19		回答	表示
20		犬	犬
21		猫	猫
22		ウサギ	ウサギ
23		ハムスター	ハムスター
24		この中ではない	他
25		ペットは売買の対象とは考えない	対象外
26			

図3 変換表(シート「convT」)

同様に、他のペットの飼育経験、かわいさ、対象ペットについて、設定しましょう（図 2 の F 列から N 列）。E2 の計算式を複製して、必要な部分を修正して作成することもできます。

3.3 金額 (欠損値)

金額は、ワークシートの「enqd」の金額の値を利用します。このアンケート項目は、回答をしないことを許しています。回答しない場合のセルの値は、空値（値が存在しない）となり、欠損値などと呼ばれています。この空値、欠損値は、通常集計対象から外すサンプルなので、0 とは区別して扱います。ある範囲の平均値 (average 関数) を求めたとき、0 は平均値の集計対象に含まれますが、空値は集計対象に含まれません。「3, 9, 0」の平均値は 4 ですが、「3, 9, 空値」の平均値は 6 です。

「=enqd!N2」などの計算式を単純に使うと、空値（未回答）は 0 となってしまいます。そこで、IF 関数を使って、空白 (空値) のときは、そのまま空白 (空値) が表示されるようにします ("" は、キーを 2 回押します)。

O2:	=IF(enqd!N2="", "", enqd!N2)
複製元 O2	複製先 O3:O9871

3.4 金額の補正 (クレンジング処理)

アンケートの金額の列 (N 列) の値を見ていくと、負の値やとてつもなく大きな値が入力されています (例: 1E+15 = 1,000,000,000,000,000 や -1)。一般的なペットの最高価格が 100 万円前後であることを考え、200 万円を超える場合や負の値の場合、空白の場合は、未回答の扱いとして、空白 (空値) にすることにします。

また、金額の入力で、『前の質問で、「この中にはない」、「ペットは売買の対象ではない」と回答したときは、この回答は不要です。』と表示しています。そこで、対象ペットで、「他」または「対象外」の場合も空白 (空値) にします。

そこで、M2 の計算式を次のように変更します。

O2: =IF(OR(enqd!N2="",enqd!N2>2000000,enqd!N2<0,N2="他",N2="対象外"),"",enqd!N2)
 複写元 O2 複写先 O3:O9871

OR 関数は、括弧内の条件のうち、1 つでも成立すれば真となる関数です。また、金額の欄（空値か、0 未満、200 万より大）の判定は、ワークシート enqd を参照していて、対象ペットの判定は、alldata の左隣のセルの値で判定しています。

条件	意味
enqd!N2=""	元の回答が未回答（空値）
enqd!N2>2000000	200 万より大きな値を回答
enqd!N2<0	負の値を回答
N2="他"	変換表による設定により設定した表 (alldata) の N2 が「他」だったら
N2="対象外"	変換表による設定により設定した表 (alldata) の N2 が「対象外」だったら

動画 [enqNN_kingaku1.mp4](#): アンケート金額の計算式の設定 1

動画 [enqNN_kingaku2.mp4](#): アンケート金額の計算式の設定 2

図 2 は、分析用シートの完成例です。セル内での改行 (**alt**+**enter**) や列幅、行高を調整しました。

4 人数（件数）の分析 (1 次元集計)

表計算ソフトウェアには、ピボットテーブルと呼ばれる集計を行う機能があります。この機能を使うとクロス集計（2 次元集計）などを簡単に行うことができます。

まず、1次元の集計を行います。現在の住居について、「一人暮らし(A)」、「家族と一緒に(B)」の人数を計算してみます。

4.1 範囲名の定義

図 2 の表を使って分析を行います。範囲の一番上の行（1 行目）を系列名として使います。ピボットテーブルでは同じ系列名は使えませんので、もし、同じ名前の系列名があれば変更してください。

表が大きいので、ウインドウ枠の固定をし、この表をいろいろな分析で使うので、A1:O9871 (O9871 の 1 文字目はアルファベットの O(オー) です) に範囲名「alldata」を付けます。

ウインドウ枠の固定 B2 をクリック → 表示 → ウインドウ枠の固定 → ウインドウ枠の固定
 名前の定義 A1:O9871 を範囲指定 → 数式 → 名前の定義 → 名前の定義 → 名前に「alldata」を入力
 範囲指定方法：A1 をクリック → (Ctrl)+(Shift)+(↓) → (Ctrl)+(Shift)+(→)

動画 [enqnn.defname2.mp4](#): アンケートウインドウ枠の固定, 名前の定義

4.2 ピボットテーブルによる集計と円グラフの作成

■ピボットテーブルによる集計

- (1) 範囲名「alldata」を指定 (または A1:O9871 (アルファベット (列) の O の行 9871) を範囲指定)
- (2) 挿入 → ピボットテーブル
- (3) 新規ワークシートを選択 → OK
- (4) フィールドの設定 (図 4) 参照

ドロップするエリア フィールド 値フィールド設定

行 現在の住居

値 現在の住居 データの個数

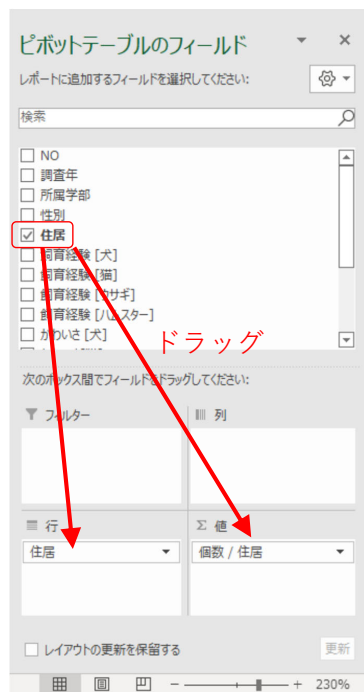


図4 ピボットテーブルの設定（1次元集計）

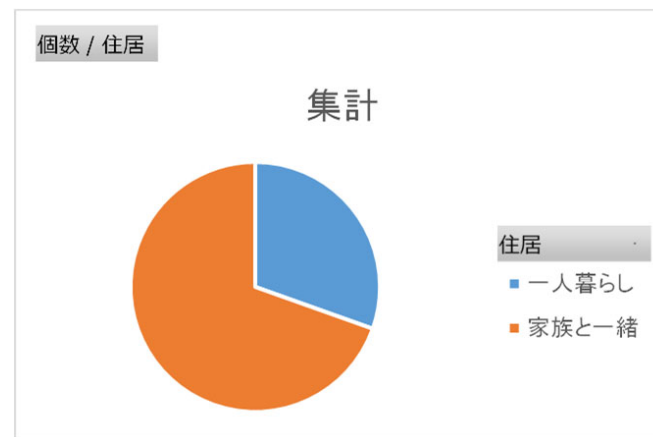


図5 住居別構成割合（1次元集計）

■ピボットグラフによる可視化 円グラフの作成は、ピボットグラフを利用します。

ピボットグラフの作成 (図5)

- (1) ピボットテーブルの部分をクリック (例 A3)
- (2) メニューの ピボットテーブル分析 → ピボットグラフ
- (3) 円グラフの中から一番左のものを選択し, OK

動画 [enqnn_pieC.mp4](#): アンケート：住居別構成割合 円グラフ

ピボットグラフは、グラフの分析の視点を変えて、自動的にグラフを変更するのに使います。たとえば、図 4 で、行や値のフィールドにドロップする項目を性別にすれば、性別の集計になります。論文やレポートなどためにレイアウトを整えるには、別途、通常のグラフを作成することあります。

5 クロス集計

クロス集計は、図 6 の右下のように複数の属性毎（学部、住居別）に個数を数えたり、平均値を求めることを言います。

5.1 学部別の住居

- (1) 範囲名「alldata」を指定 (または A1:O9871 を範囲指定)
- (2) 挿入 → ピボットテーブル
- (3) 新規ワークシート を選択 → OK
- (4) フィールドの設定 (6 の左下)

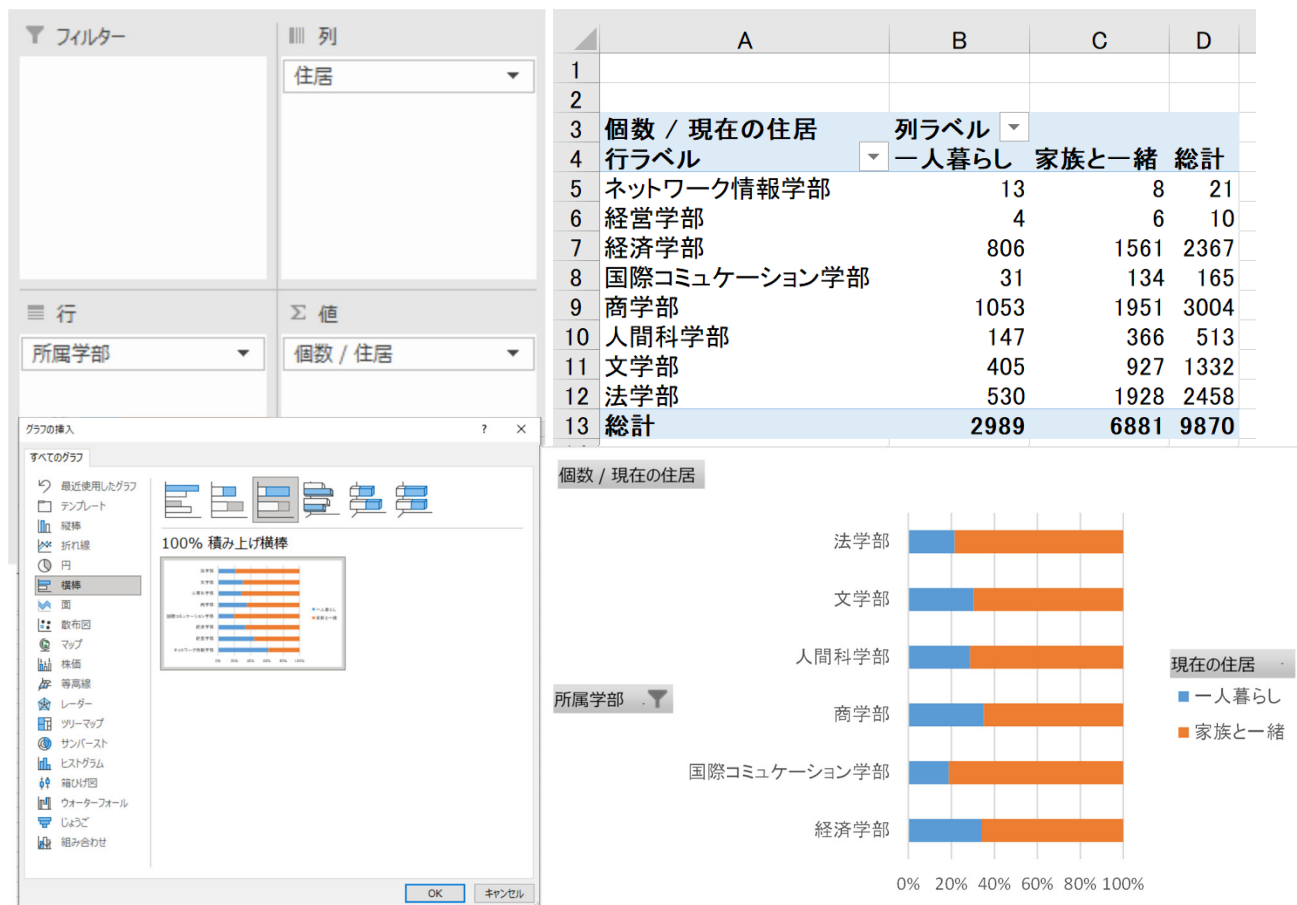


図 6 学部別住居（クロス集計）

ドロップするエリア	フィールド	値フィールド設定
列	現在の住居	
行	所属学部	
値	現在の住居	データの個数

(5) 図 6 の右上のような表が作られます。

(6) ピボットグラフの作成. 横棒の中の 100% 積み上げ横棒を選びます (図 6 の左下)

動画 [enqnn-gakubzyukyo.mp4:アンケート:学部別住居 \(クロス集計表\)](#)

図 6 では、「経営学部」や「ネットワーク情報学部」という回答がありますが、経営学部・ネットワーク情報学部対象の授業ではこのアンケート調査は行われませんでした。そこで、学部を使った集計には除外することにします。また、国際コミュニケーション学部は 2020 年に設置されたので、このアンケートでは、2020 と 2021 年の 2 年分しかなく、サンプル数も 162 と比較的小数ですが、分析対象とします。

(1) 図 6 の右下のグラフグラフの「所属学部」ボタンをクリック

(2) ネットワーク情報学部と経営学部のチェックを外し、「OK」をクリック

(3) 図 7 のようなグラフになります。

動画 [enqnn-gakubzyukyo2.mp4:アンケート:学部別住居 \(クロス集計表, 学部修正\)](#)

5.2 犬と猫のかわいさの関係

ペットのかわいさについて、5 段階で評価しています。このデータを 1~5 に置き換えています (図 3)。大きければ大きいほど、「かわいい」と評価した回答です。このように、順序に意味があるデータは、順序尺度で測られているといえます。

そこで、犬と猫のかわいさの関係がどのようになっているのか、ピボットテーブルを使って分析してみます。

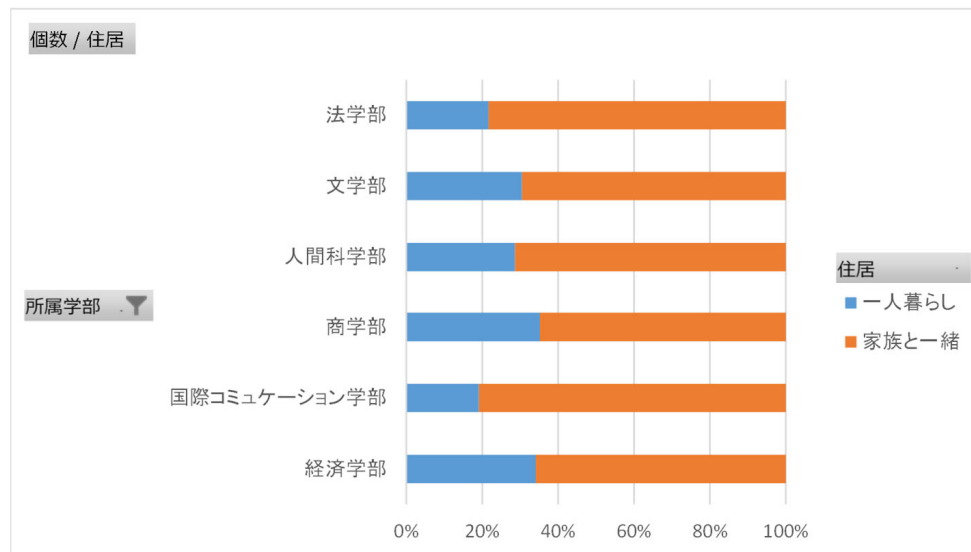


図7 学部別住居の帯グラフ

(1) 範囲名「alldata」を指定 (または A1:O9871 を範囲指定)

(2) 挿入 → ピボットテーブル

(3) フィールドの設定とその結果 (図 8)

ドロップするエリア	フィールド	値フィールド設定
列	かわいさ (犬)	
行	かわいさ (猫)	
値	かわいさ (犬)	数値の個数 (▼ をクリックして, 値フィールドの設定 → 数値の個数)

数値の個数 / かわいさ[犬]		列ラベル				
行ラベル		犬のかわいさ				
		1	2	3	4	5 総計
猫のかわいさ	1	362	15	16	45	107 545
	2	17	53	11	50	137 268
	3	15	16	258	104	296 689
	4	27	32	83	593	895 1630
	5	64	51	168	686	5769 6738
総計		485	167	536	1478	7204 9870

図 8 犬と猫のかわいさの分布（ピボットテーブル）

(4) 挿入 → ピボットグラフ → 縦棒 → 3-D 縦棒 (縦棒の一番右) (図 9)

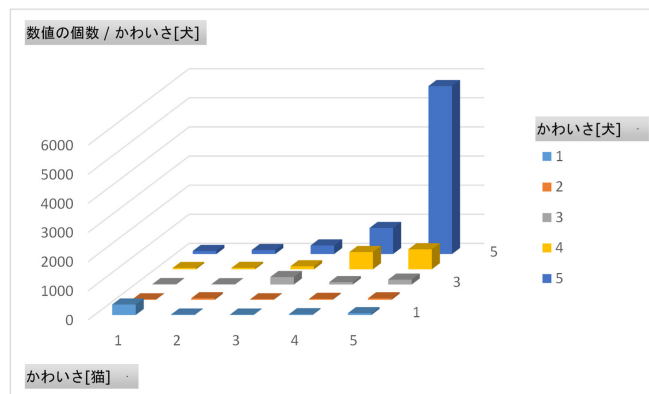


図 9 犬と猫のかわいさの分布（ピボットグラフ）

動画 enqnn_3ddogcat.mp4: アンケート: かわいさの比較

■ヒートマップの作成 3D 縦棒ですと、見る角度によって高さの直観的な印象が異なったり、一部縦棒が棒の陰になり見えなかったりします。そこで、図 10 のようなヒートマップを作成します。

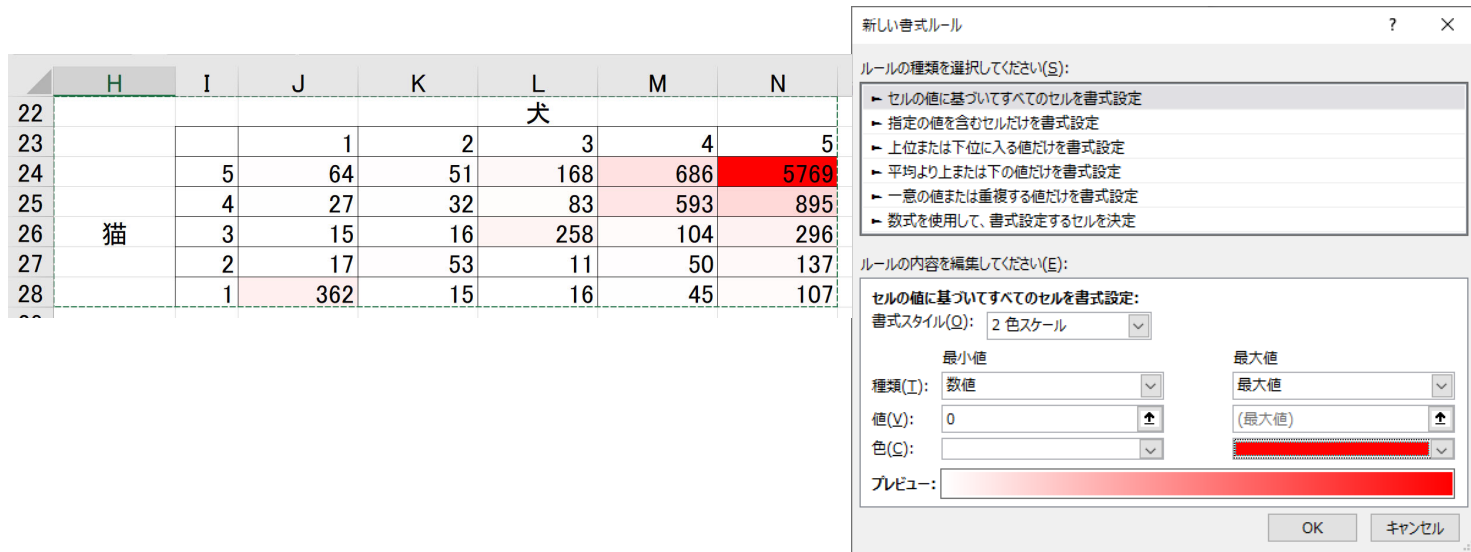


図 10 犬と猫のかわいさの分布（ヒートマップ，左）とその設定（右）

- (1) 表の右下に、表示する範囲 (H22:N28) を決めます。
- (2) 表頭 (22~23 行)、表側 (H,I 列) を記入します。ただし、右上をかわいさの大きな数値するため、I 列は上から、5, 4, ..., 1 とします（散布図と同じようにします）。
- (3) 各行についてピボットテーブルの値を転記します。ただし、フィールドの設定を変更してもうまく対応できるよう、=B9 のように、計算式を手入力します。

(a) 24 行目 (猫 5) の設定: J24: (B9 は, キーボード入力. クリックでの入力は不可)

:

(b) 28 行目 (猫 1) の設定: J28: (B5 は, キーボード入力. クリックでの入力は不可)

(c) J 列の計算式を K 列から N 列まで複写

複写元: , 複写先:

(4) ヒートマップの背景色の設定

(a) J24:N28 を範囲指定 (ヒートマップで背景色つける範囲)

(b) メニューの → → →

(c) 新しい書式ルールを定義します (図 10 右).

ルールの種類: セルの値に基づいてすべてのセルを書式設定

書式スタイル: 2 色スケール

最小値: 種類を数値, 値を 0, 色は 白にします

最大値: 種類を最大値, 色は原色に近い色 (例では 赤) にします

(d)

動画 [enqnn.Hmap1.mp4](#): アンケート: かわいさの比較 (ヒートマップ)

図 10 右や図 9 を見ると, 圧倒的に, 犬, 猫ともに「かわいい」もしくは「どちらかといえばかわいい」と回答していることがわかります.

■他の組み合わせの分析 例題では, 犬と猫の比較でしたが, 他の種類間の比較は次のようにします. 例としてハムスターとウサギに変更します.

※フィールドを追加する前に, 前のフィールドを削除します. 削除しないと, ピボットテーブルが拡大して, ヒートマップなどが消えることがあります.

- (1) ピボットテーブル (例えば A1) を選択し、右クリックで **フィールドリストを表示する** で、ピボットテーブルのウィンドウを表示させます。
- (2) 列のボックスから、**かわいさ [犬]** を削除します (**▼** をクリックして、**フィールドの削除**)
- (3) **かわいさ [ハムスター]** を列フィールドにドラッグ
- (4) 行ボックスから、**かわいさ [猫]** を削除し、**かわいさ [ウサギ]** を行フィールドにドラッグ
- (5) 値ボックスから、**数値の個数 かわいさ [犬]** を削除し、**かわいさ [ハムスター]** を値ボックスにドラッグ
- (6) 値ボックスの **値フィールドの設定** を **数値の個数** に変更
- (7) 自動的に、ピボットテーブル、3D グラフ、ヒートマップが書き換わったと思います。ただし、表頭、表側の「犬」や「猫」の文字の書き換えは手作業で行います。

動画 [enqnn.Hmap2.mp4: アンケート: 他の組み合わせのかわいさの比較 \(3D 棒, ヒートマップ\)](#)

練習：他の 2 種のペットの組み合わせで、かわいさの関係を分析してみましょう。

6 数値の分析

6.1 購入対象のペット毎の人数

購入対象のペット毎の人数をピボットテーブルで求めましょう。ただし、**対象ペット** は **行** にドロップします (グラフ作成のとき、1 つの系列として扱うようにするため)。

ドロップするエリア	フィールド	値フィールド設定
行	対象のペット	
値	対象のペット	データの個数

次に、図 11 のように、棒グラフと円グラフを表示させましょう。

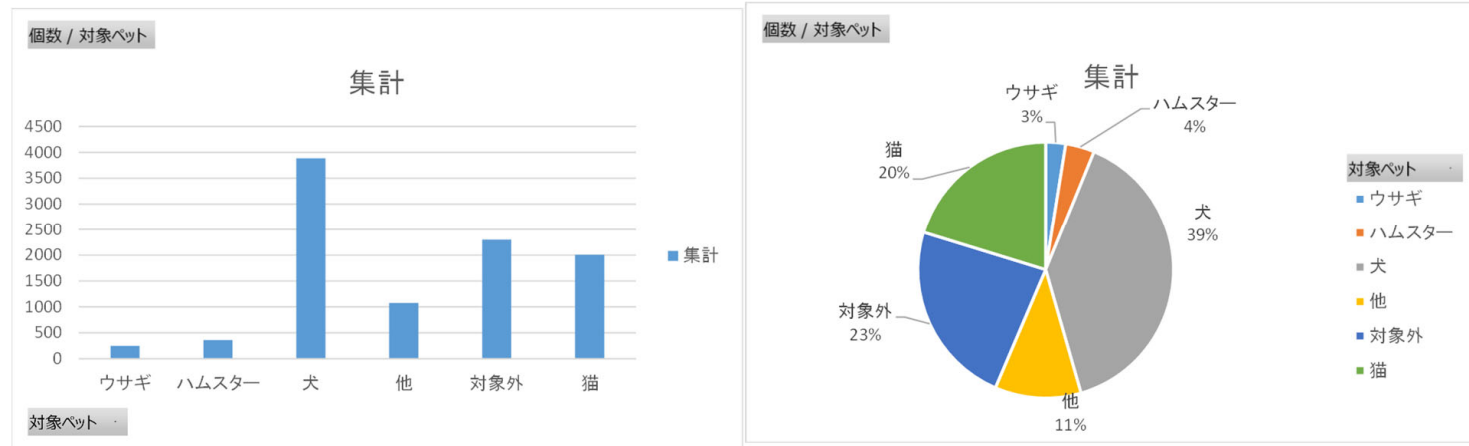


図 11 ペット毎の人数の棒グラフ (左) と円グラフ (右)

動画 [enqnn_petc.mp4](#): アンケート：購入対象ペットの構成

6.2 購入対象のペット毎の想定支出金額の平均値

金額の列を「想定支出金額」と呼ぶことにしましょう。ペット毎に、想定支出金額を回答した人の人数と平均値を求めます。この場合、購入対象のペットを選択しても、「想定支出金額」を回答しない（空白、空値）人もいます。また、購入対象のペットの選択で、「他」や「対象外」を選択した回答も集計対象から外します。次のように 2 つのピボットテーブルを作成します。

ドロップするエリア	フィールド	値フィールド設定
行	対象ペット	
値	金額	平均

ドロップするエリア	フィールド	値フィールド設定
行	対象ペット	
値	金額	数値の個数

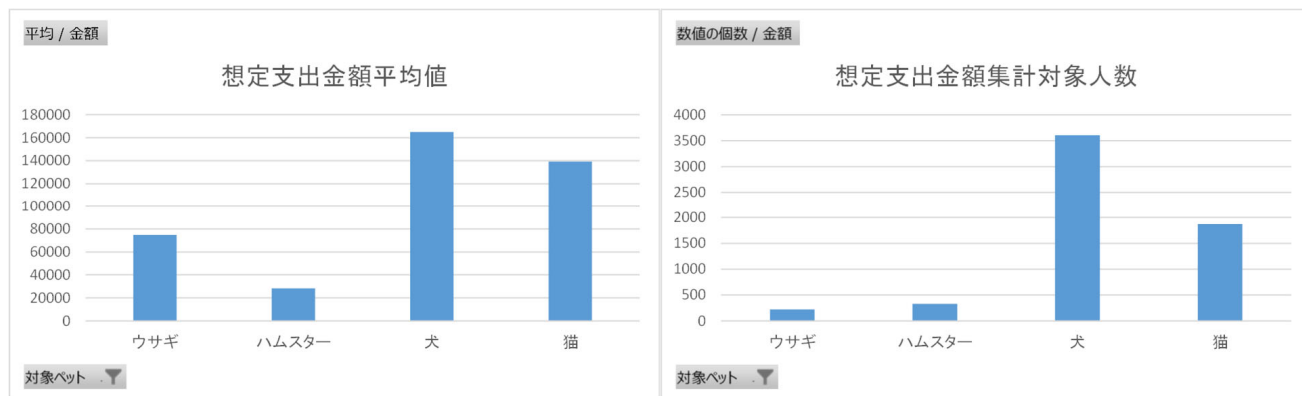


図 12 ペット毎の想定支出金額の平均値（左）とその対象件数（右）

動画 [enqnn.kingakusyukei.mp4](#): アンケート：ペット毎の想定支出金額の平均値, 対象件数

図 11 と図 12 は若干異なります。これは、前述の対象ペットを選択したのに、金額を記入しなかった回答の影響です。

7 時系列の分析

時間とともに変化するデータは時系列データと呼ばれています。このアンケートデータも「調査年」の系列があり、時間とともにペットへの関心がどう変化しているか分析できます。

7.1 購入対象ペットの調査年ごとの推移

図 13 のように、調査年ごとに 購入対象ペットがどのように変化するかどうかがグラフ化（折れ線グラフ、このグラフは散布図で作成）してみます。

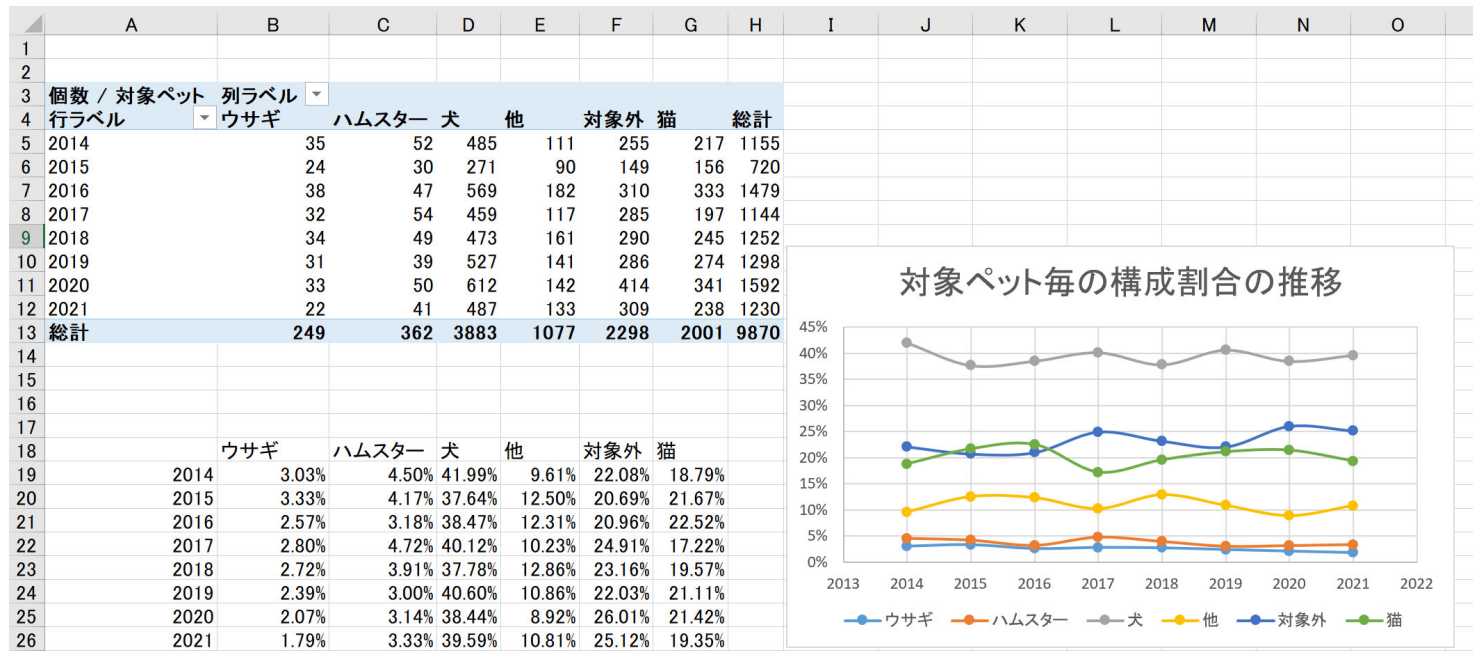


図 13 ペット毎の人数割合

図 13 の左上のように、ピボットテーブルを使い、各調査年、ペットの種類毎に人数を数えます。
ドロップするエリア フィールド 値フィールド設定

行	調査年
列	対象ペット
値	対象ペット 個数

動画 [enqnn.lineC.mp4: アンケート 調査年ごと対象ペット ピボットテーブルの作成](#)

この表から、図 13 の左下のように各調査年でそれぞれのペットを選択した割合を計算します。ただし、ピボットテーブル (A3:H13) はコピーしにくくなっています、そこで、「= 計算式」を使って、値を転記しながら求めます。

(1) 年度の複写 A19: , 複写元: A19, 複写先: A20:A26

(2) ペット名の複写 B18: , 複写元: B18, 複写先: C18:G18

(3) 割合の計算 各ペットの個数を総計で割ります。

(a) B19: (ウサギの個数を総計で割る。ただし、総計の列はいつも H 列なので、H の前に \$ を付け絶対参照にします。)

(b)

(スピルを使って計算することもできます。)

動画 [enqnn.lineC2.mp4: アンケート 調査年ごと対象ペット 割合の計算](#)

図 13 右のように、散布図で推移のグラフ（折れ線グラフ）を作成します。

(1) A18:G26 を範囲指定

(2) メニューの → → →

(3) X 軸と凡例が逆の場合：グラフを右クリックして、 →

(4) グラフを整える

8 ペット毎, 調査年ごとの分析 (想定支出金額の中央値など)

8.1 平均値での分析

ペットの種類を列, 調査年を行に, 値を金額の平均値とすれば, 図 14 のようなピボットテーブルと金額の平均値のグラフを描くことができます.

ドロップするエリア	フィールド	値フィールド設定
行	調査年	
列	対象ペット	
値	金額	平均

また, ピボットグラフの 折れ線の中の折れ線 (一番左) で, 図 14 のようなグラフを描くことができます. また, 「他」と「対象外」の金額は 0 になるように計算式で設定したので, 列ラベル をクリックして, 「他」と「対象外」のチェックを外し, 表示しないようにしましょう.

図 14 をみると, 想定支出金額の平均値が犬が 13 万円~21 万とかなり高額になっており, また, 年を追うごとに上昇しているように見えます. 想定支出金額の値を見ると, 所々に大きな金額 (外れ値) を回答しています. 人間が主観的に回答したデータなど外れ値を含むデータは, 平均値の他に外れ値の影響が少ない中央値なども求めた方がよいでしょう.

8.2 箱ひげ図

箱ひげ図は, サンプル (データ) の散らばり具合を見るときに使われます. ほぼすべてのサンプルがどの範囲に散らばっているのか, また, 約半分のサンプルがどの範囲に散らばっているのかや, 中央値や平均値がどこに位置しているのかを直

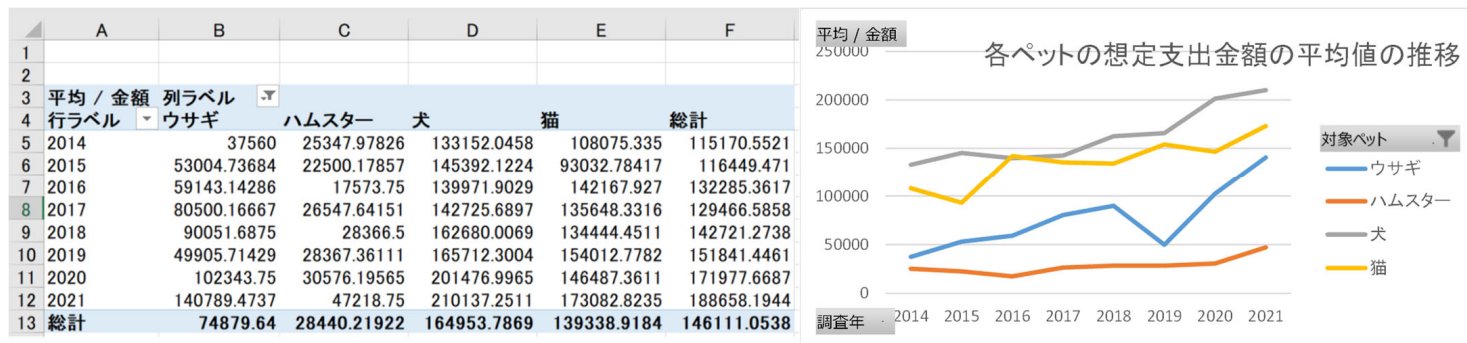


図 14 ペット毎の平均金額の推移

感的に把握できます。

図 15 は、箱ひげ図の大まかな見方です。黒の線や文字が箱ひげ図で、色の付いた線や文字は説明です。2つの系列 A,B のデータ（サンプル）があり、そのデータがどのように分布しているかを示しています。真ん中の長方形の部分とそのひげ（T 字形の線）が出ています。

上下のひげの端から端までがサンプルが分布している範囲を示しています。A のサンプルは 10 から 130 のあいだにほぼあること、B のサンプルは 10 から 140 のあいだにほぼあることを示しています。

長方形の部分は、真ん中の 50% のサンプルが分布している範囲を示しており、真ん中の線が中央値を表しています。A は 50 から 100 のあいだに半分のサンプルが、B は 27.5 から 90 のあいだに半分のサンプルがあることを示しています。「×」印は平均値を表しています。

「ほぼ 100%」と書いたのは、外れ値（極端に大きなサンプルや小さいサンプル）があるとき、ひげの外側に別途、点（●）を打つことになってます。B は、200 という外れ値があるので、ひげの外側に点を打っています。

A は、中央値の線が箱の上方であったり、下のひげが長く伸びています。中央値 88 付近に沢山のサンプルがあり、小さ

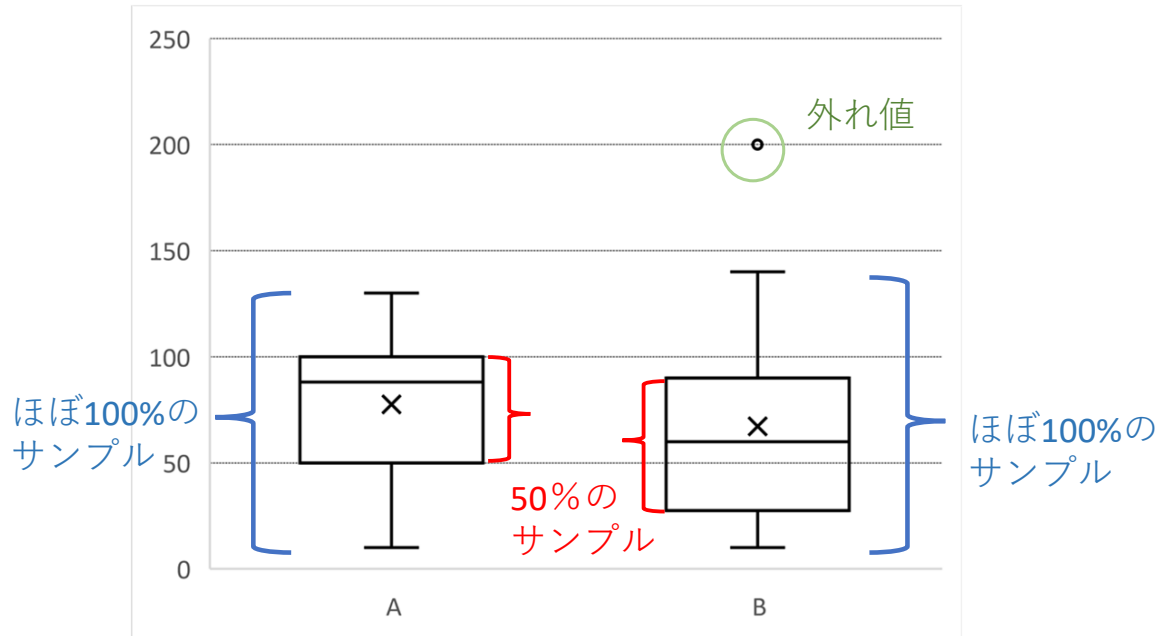


図 15 箱ひげ図 概略

な値のサンプルが広く分布していることを示しています。逆に、B は中央値 60 より上に、広い範囲にサンプルが分布していることがわかり、上方に外れ値があることもわかります。

図 16 は、箱ひげ図の詳細です。箱ひげ図は四分位数（最小値、第 1 四分位数、中央値 (第 2 四分位数)、第 3 四分位数、最大値) と平均値を基に描かれます。サンプルを小さい順に並べ替えて、四分位は、小さいサンプルから $1/4$ を第 1 四分位、 $2/4$ を第 2 四分位 (中央値)、 $3/4$ を第 3 四分位と呼びます。第 1 四分位から第 3 四分位のあいだに、おおよそ 50% のサンプルが含まれます。

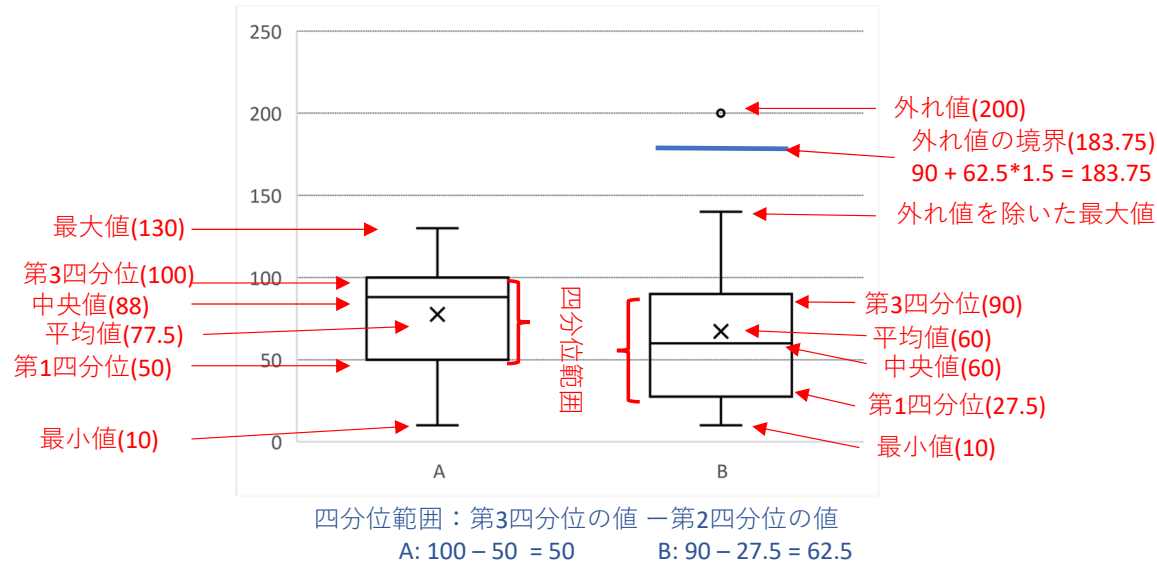


図 16 箱ひげ図 2

あまりに大きな値，小さな値は，外れ値として，箱ひげ図の外に出し，点を打ちます．外れ値かどうかの基準（境界）は，第3四分位 + [四分位範囲 × 1.5] より大きな値を上の方のひげより上に点を打ち，第1四分位 - [四分位範囲 × 1.5] より小さな値を下のひげより下に点を打ちます．外れ値以外のサンプルでの最大値を上方向のひげの端，最小値を下方向のひげの端として箱ひげ図を描きます．

8.3 箱ひげ図の描き方（ペット毎の箱ひげ図の作成）

Excelでは、図 17 のように分類項目別（例題ではペット別）に、箱ひげ図を作成する機能があります。しかし、現行の Excel では、空値を含んだデータの箱ひげ図は作成できません。そこで、金額が空値（対象ペットが「対象外」で金額を空値に設定したサンプルも含む）の行を削除した表を作成し、箱ひげ図を作成します。ただし、元のデータ（シート「alldata」など）は残すため、元のデータの必要な部分をコピーして作成し、ソート機能を使って金額が空値以外のサンプルを集めて箱ひげ図を作成します。

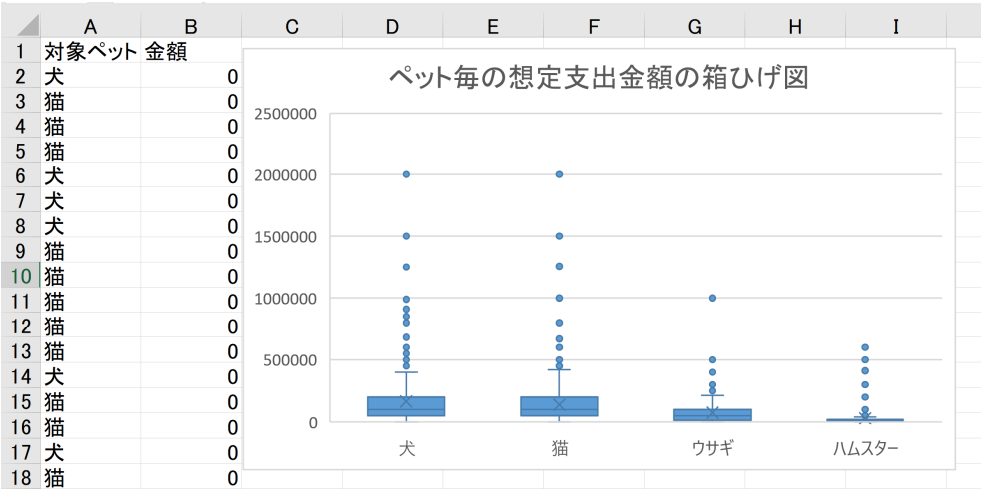


図 17 ペット別箱ひげ図 1

箱ひげ図のためのデータを作成

- (1) 新しいワークシートを作成します (例：箱ひげ図)
- (2) alldata の必要な部分対象ペットと金額の列を 箱ひげ図 にコピー（値貼り付けを利用）します (図 17 のような表を作成).
 - (a) alldata の N1 から N 列の最下行を範囲指定 (N1:O9871) を範囲指定し、コピー.
1 行目から最下行の範囲指定方法 (1 行目のセルをクリック → (ctrl) + (Shift) + (↓))
 - (b) 箱ひげ図 の A1～に 値 で貼り付け
 - (c) 同様に、alldata の金額の列を 箱ひげ図 の B 列に 値 で貼り付け
- (3) B 列を基準に、A 列と B 列の値を B 列の値の小さい順に並べ替えます.
 - (a) A1 から B 列の最下行 (A1:B9871) を範囲指定
 - (b) メニュー データ → 並べ替え
 - (c) 先頭行をデータの見出しとして使用する にチェックを入れる
 - (d) 最優先されるキー： 金額
 - (e) 並べ替えのキー： セルの値
 - (f) 順序： 小さい順
 - (g) OK
- (4) 6031 行目まで数値が記入されたサンプル、6032 行以降が空白のセルになっています。A1:B6031 が箱ひげ図のデータになります (図 17 の左)。金額の欄を空値にする条件の違いにより、6031 は若干異なることがあります。この数値が記入されたサンプルが何行目目までか (例では 6031) の数値はメモしておきましょう。

箱ひげ図を描画

- (1) A1:B6031 を範囲指定 (名前ボックスに、A1:B6031 と入力することでも可能)

(2) メニューの 挿入 → グラフ → 統計グラフ → 箱ひげ図

図 17 の右側のグラフが表示されます。

各ペットについて、大きな金額の方向に多数の外れ値があり、データの分布が見えにくくなっています。そこで、縦軸の範囲を 0～500000 の範囲に絞って表示します (図 18)。

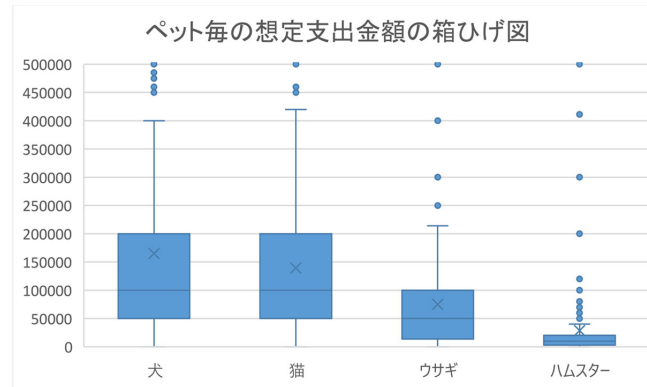


図 18 ペット別箱ひげ図 2 (縦軸の範囲を変更)

動画 [enqnn.boxplot.mp4](#): アンケート 箱ひげ図の作成

図 18 を見ると、犬と猫の箱ひげ図はほぼ同じで半分の回答が 5 万から 20 万であることが分かります。犬と猫ではほぼ同様の金額を考えていることが分かります。図 17 では、犬のほうが上方への外れ値が多く、その影響などで、平均値 (×印) は若干、犬のほうが高くなっています。これは、図 14 で平均値が少し犬の方が大きいことに表れています。

ウサギは明らかに犬と猫よりも低い範囲に、ハムスターはさらに低い範囲に分布していることがわかります。

8.4 調査年・ペット別の分析

箱ひげ図では、四分位の値や平均値を図示することができました。本節では、関数等によって、調査年・ペット別の平均値・四分位点の値・調整平均値の値を求めてみます。

8.4.1 調査年・ペット別の平均値・四分位・調整平均値の計算

■調査年・ペット別のデータの抜き出し ピボットテーブルには、中央値や第 1 四分位、第 2 四分位、調整平均値を求める機能がありません。そこで、調査年、ペットの両方 または 片方を指定して、その指定されたサンプルのみを対象にして、関数を使い、各値を求めます。

図 19 の 1 行目の N 列から X 列、T 列の 2 行目から 13 行目の表頭、表側を入力してみましょう。また、V2,X2 に仮に対象ペット名 W3,X3 に仮に対象調査年を入力しておきましょう。

図 19 は、計算表で、P 列は V2 で指定したペットのみ値を表示し他は空白としたもの、Q 列は W3 で指定した調査年のみ値を表示し他は空白としたもの、R 列は X2, X3 で指定したペット・調査年のみ値を表示し他は空白としたものです。計算式は次のようになります。

セル	計算式
P2	=IF(N2=\$V\$2,O2,"")
Q2	=IF(B2=\$W\$3,O2,"")
R2	=IF(AND(N2=\$X\$2,B2=\$X\$3),O2,"")

この P2～R2 の計算式を下方へ複写します。

動画 [enqnn-QUARTILE1.mp4](#): アンケート 指定ペット、調査年の抜き出し

	A	O	P	Q	R	S	T	U	V	W	X
1	NO	金額	ペット指定 金額	調査年指 定金額	調査年・ ペット指定 金額			全体	ペット指定 金額	調査年指定 金額	調査年・ ペット指定 金額
2	1	10000	10000	10000			指定ペット		犬		猫
3	2						指定調査年			2014	2014
4	3	100000	100000	100000			件数	6030	3599	709	196
5	4						平均値	143958.94	162404.40	112512.12	98422.66
6	5	50000		50000	50000		最小値	0	0	0	0
7	6						第1四分位	50000	50000	30000	30000
8	7	100000	100000	100000			中央値	100000	100000	100000	70000
9	8						第3四分位	200000	200000	150000	150000
10	9	0	0	0			最大値	1500000	1500000	1500000	1000000
11	10	200000	200000	200000			10%調整平均値	118004.40	136601.32	88594.22	81756.42
12	11						標準偏差	160260.49	166352.46	144605.13	110557.00
13	12						変動係数	1.11	1.02	1.29	1.12

図 19 調査年ペット別計算表 (2024 年度版の当初のテキストに誤りがあったので訂正 (2025/1/7))

■調査年・ペット別の集計値の計算 ペット毎の想定支出金額の集計値として、回答者数、最小値、第 1 四分位数、中央値 (第 2 四分位数)、第 3 四分位数、最大値、平均値、10% 調整平均値、標準偏差、変動係数を計算します。

サンプルを小さい順に並べ替えて、四分位は、小さいサンプルから 1/4 を第 1 四分位、2/4 を第 2 四分位 (中央値)、3/4 を第 3 四分位と呼びます。第 1 四分位から第 3 四分位のあいだに、おおよそ 50% のサンプルが含まれます。

10% 調整平均値は、小さいサンプルから 10% 大きいサンプルから 10% のサンプルを除外して、残った 80% のサンプルの (単純) 平均値を求めたものです。

標準偏差は、データのチラバリを表す指標で、平均値からおおよそどれくらい離れているかの指標です。変動係数は異な

る系列間でチラバリを比較するための指標で、標準偏差を平均値で割った値です。

9871 は O 列の最下行の行番号です。

セル	計算式	
U4	=COUNT(O\$2:O\$9871)	COUNT は、数値の個数を数える関数
U5	=AVERAGE(O\$2:O\$9871)	
U6	=QUARTILE(O\$2:O\$9871,0)	=MIN(O\$2:Q\$9871) に同じ (小さい方から 0/4)
U7	=QUARTILE(O\$2:O\$9871,1)	第 1 四分位 (小さい方から 1/4)
U8	=QUARTILE(O\$2:O\$9871,2)	第 2 四分位, 中央値, Median 関数に同じ (小さい方から 2/4)
U9	=QUARTILE(O\$2:O\$9871,3)	第 3 四分位 (小さい方から 3/4)
U10	=QUARTILE(O\$2:O\$9871,4)	最大値, MAX 関数に同じ (小さい方から 4/4)
U11	=TRIMMEAN(O\$2:O\$9871,0.2)	10% 調整平均値, 0.2 は 上下の各 10%, 合わせて 20% のサンプルを除いた平均値
U12	=STDEV(O\$2:O\$9871)	標準偏差
U13	=U12/U5	変動係数 = 標準偏差 ÷ 平均値

QUARTILE 関数は、四分位を求める関数で、最初の引数に範囲、第 2 の引数に小さい方から $0/4$ のいくつかの位置の値かを指定します。0 で $0/4 = 0\%$ 小さい方から 0% 最小値, 1 で $1/4 = 25\%$ 小さい方から 25% 第 1 四分位, 2 で $2/4 = 50\%$ 小さい方から 50% 第 2 四分位 = 中央値, 3 で $3/4 = 75\%$ 小さい方から 75% 第 3 四分位 4 で $4/4 = 100\%$ 小さい方から 100% 最大値を表します。

TRIMMEAN 関数は、調整平均値を求める関数で最初の引数に範囲、第 2 の引数で、合わせて何パーセントの値を削除するかを指定します。

STDEV 関数は、標準偏差を求める関数で最初の引数で範囲を指定し、標準偏差を平均で割って変動係数を求めています。

U4:U13 の計算式を U4:X13 に複写します.

動画 [enqnn_QUARTILE2.mp4](#): アンケート 四分位点などの計算

V2,W3,X2,X3 の値を変更することにより, それぞれの集計値を求めてることができます. ペット別の集計値を比較する表を作成するには次のようにします.

- (1) V2 にペット種 (例えば 犬) を入力
- (2) 集計値が計算されます.
- (3) V2:V13 をコピーします.
- (4) 集計値記載用のワークシートに 値貼り付け をします.
- (5) (1) に戻り別のペット種を入力しくり返します.

動画 [enqnn_QUARTILE3.mp4](#): アンケート 対象ペット・調査年の変更

8.4.2 ヒストグラム (度数分布表) の作成

■データの抜き出しと空白のセルの削除 箱ひげ図の作成と同様に Excel では, ヒストグラムの作成では空白のセルは受け付けません. そこで, データを抜き出し, 別のワークシートに貼り付け, ソートで数値だけのセルにし, ヒストグラムを描きます.

犬の金額のヒストグラムを作成してみます.

- (1) V2 に犬を設定します,
- (2) P 列に犬の金額が表示されます.
- (3) P2 から P 列の最下行 (P2:P9871) を範囲指定 ((**Ctrl**) + (**Shift**) + (**↓**)) して, コピー
- (4) 犬のヒストグラム作成用のワークシートを作成し, A1 から値貼り付け (A1 をクリックし, 形式を選択して貼り付け (値))

- (5) 犬のヒストグラム作成用のワークシートで、A1:P9871 を範囲指定
- (6) データ → 並べ替え
- (7) 先頭行を見出しとして使用する のチェックされないことを確認
- (8) 最優先されるキー： A 列
- (9) 並べ替えのキー： セルの値
- (10) 順序： 小さい順
- (11) OK

動画 [enqnn_histram1.mp4](#): アンケート ヒストグラム用サンプルを抜き出すソートされた一番下の数値のセルが何行目か確認し、メモしておきます。例題の場合、3599 行目になります。

図 20 のような階級幅が 50000 円のヒストグラムを作成します。

■ヒストグラム（グラフ）の作成

- (1) ヒストグラム化する範囲（数値のみ、例では、A1:A3604）を範囲指定
- (2) メニューの 挿入 → グラフ → 統計グラフの挿入 → ヒストグラム
- (3) ヒストグラムが作成される。適当に横幅を長くする。
- (4) 階級幅を 50000 にするため、ビン（ヒストグラムの棒の区間幅）を 50000 にする。
X 軸で右クリックし、軸の書式設定
ビンの幅を 50000 に変更

200 万を越えるサンプルは、異常値として取り除いているので、最後は 2,000,000 までです。

動画 [enqnn_histram2.mp4](#): アンケート ヒストグラムの作成

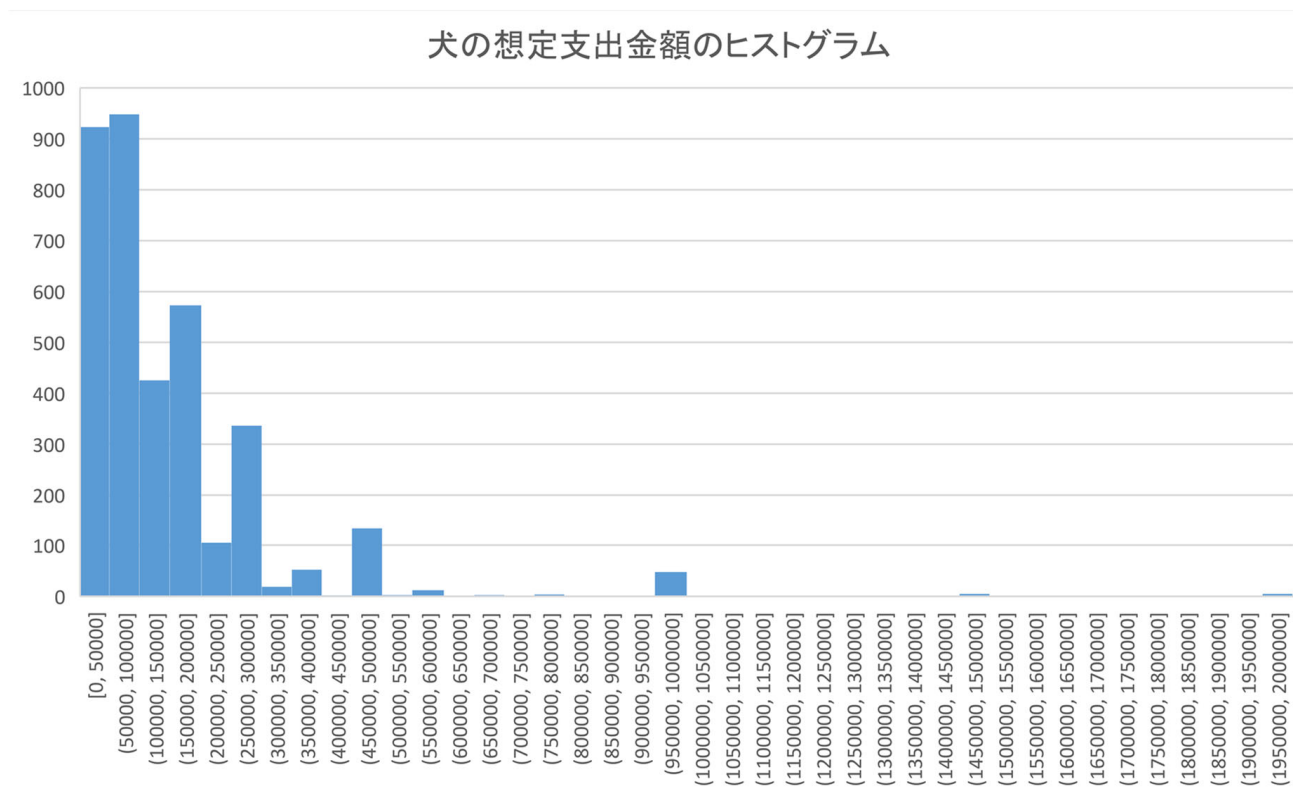


図 20 犬の金額のヒストグラム

図 20 は、度数のピークが金額が少ない階級に集中し、右に歪んだグラフになります。他の形状については、[なるほど統計学園：ヒストグラムのヒストグラムの形状を参照ください](#)。

9 自由記述欄の分析

2023 年度から実施した自由記述欄をテキスト分析を行います。

- 1) 共起ネットワーク図や階層的クラスター分析により、全体の傾向を把握することと、
 - 2) 出現する抽出語により回答を分類し、自由記述欄にどのような分類の回答が多いかなどを分析をしてみましょう。
- アンケートでは、「ペットに関する意見や思いを記述してください。」と表示し、自由記述で回答してもらいました。
- 分析は、著者により KH Coder を使って、語を抽出し、各回答者の回答を集計単位として

- 1) の共起ネットワークと階層的クラスター分析を行い、
- 2) 各回答、抽出語の一覧表を作成し、分類するための語（複数）を指定して、その語を含む回答を分類していきます。

9.1 準備

9.1.1 Excel ファイルのダウンロード

2023 年度の自由記述を含むデータファイルは、[テキストのページ](#) からダウンロードできます。

このファイルは、アンケート番号と調査年が付加しています。P 列 (自由記述の回答の列) に、1 人の回答者の自由記述が 1 つのセルに入力されています。回答内容を見てみてください。

9.1.2 表記のゆれの修正例 (下記ファイルは修正済み)

自由記述欄を眺めていくと、表記の揺れが見られます。例えば、「かわいい」と「可愛い」の 2 つの表記があります。KH Coder では、ひらがなだけの形容詞の語は「形容詞 B」として一般的な語として分類されるので、すべて、漢字の「可愛い」に置換して分析することにしました (参考 URL:https://www.jstage.jst.go.jp/article/mathling/31/1/31_36/_pdf/-char/ja)。

他にも、重要そうな（文章の特徴に影響を与える）語で表記の揺れがあれば、置換して分析できます。

次のようにして、置換をしました。

- (1) P 列のみを置換の対象にするので、**P**列をクリック
- (2) **検索する文字列:** かわいい **置換後の文字列:** 可愛い
- (3) **次を検索**をクリックし、正しければ、**すべて置換**をクリックします。

動画:アンケート：テキスト分析：表記の揺れ修正

表記の揺れ 修正済みファイルは、**テキストのページ**からダウンロードできます (enq_freeWtext2023_replace.xlsx)。

9.1.3 KH Coder による分析（分析結果）

KH Coder というソフトウェアを利用して分析を行いました。2024 年度は、2023 年度アンケートの分析結果を提示して、それを考察、利用することにしました。2024 年度のアンケート結果は、別途、**テキストのページ**からダウンロードできます。

前述の「かわいい」を「可愛い」に置換したファイルを利用して、分析を行い、その結果を提示します。

- (1) 形態素解析で、形態素（単語）を抽出する。

形態素解析は、自然言語で書かれた文（日本語の文）意味を持つ単位（形態素、単語）に分割し、分析を行うことです。

- 形態素解析エンジンは、「Mecab」を利用
- 抽出語リストは、「抽出語リスト_2023.xlsx」というファイルで配布 (**テキストのページ**からダウンロード)

- (2) 共起ネットワーク

- 対象とする抽出語は、「最小出現回数:10」、「最大出現回数:200」を設定
- **強い共起関係ほど濃い線に** と **最小スパニング・ツリーだけを描画**、**係数を表示**にチェックを入れて分析を行いました。

- 共起ネットワーク図は、「共起ネットワーク_2023.pdf」というファイルで配布 ([テキストのページ](#) からダウンロード)

(3) 階層的クラスター分析

- 最小出現回数:10 最大出現回数:200
- 階層的クラスター図は、「階層的クラスター分析_2023.pdf」というファイルで配布 ([テキストのページ](#) からダウンロード)

9.2 共起ネットワーク図の解釈

- ○の大きさが語の出現頻度を表しています.
- 線で結ばれていると、語と語が共起（ともに出現）していることを示し、線が太いほど、その共起が強いことを示しています. 数値は下記の共起の強さ (Jaccard 係数) を表しています.
- サブグラフは、強く結びついている語のグループをまとめて、同じ色で表示しています.
- 注意：○の近さに意味はありません.

9.2.1 共起の強さ (類似係数) の計測方法 – Jaccard 係数 –

語と語が共起する強さ (類似係数) は、1つ (1人) の回答 (この例) を単位とした Jaccard 係数を使っています (デフォルト). $|A \cap B|$ を語 A,B がともに出現する回答数, $|A \cup B|$ を語 A,B が少なくともどちらかが出現する回答数とします. 語 A と B の Jaccard 係数は、次の式で計算されます.

$$\text{語 } A \text{ と } B \text{ の Jaccard 係数} = \frac{|A \cap B|}{|A \cup B|} = \frac{\text{ともに出現数}}{\text{どちらかの語の出現数}}$$

語 A が 20 個の段落に出現し、語 B が 60 個の段落に出現し、語 A と B がともに出現する段落数を 15 とします。(語 A と B の少なくともどちらかが出現する段落は、 $20 + 60 - 15 = 65$ になります)。このときの Jaccard 係数は、

$$\text{語 } A \text{ と } B \text{ の Jaccard 係数} = \frac{15}{20 + 60 - 15} = \frac{15}{65} = 0.25$$

となります。類似度係数としては、Jaccard 係数の他、Cosine 係数が使われることがあります(どちらを使うかは担当教員の指示に従ってください)。

9.3 回答の分類

抽出語リスト、共起ネットワーク図、階層的クラスター分析の図 およびアンケートの自由記述欄を見ながら、分析者が指定した語の出現によって、自由記述の各回答がどのようなことに述べているのかを分類(重複あり、どれにも分類されないこともあり)してみます。

そこで、分析者が、近い語(共起する語、クラスター分析距離が近い語)の組を指定し、その後のいずれが出現したら、その語の組で表す意見を述べていると判断します。

9.3.1 分類のキーワード群の選択

階層クラスター分析の図や共起ネットワーク図を見ながら分析を行います。階層クラスター分析の図で、近い位置に保健所、引き取る、..., 保護といった語が表示されます。これらは、ペットショップから購入するのではなく、保護施設などから引き取ることにについて述べた記述です。そこで、これらの語(キーワード)を含む回答を、たとえば、「保護施設からの引き取り」という分類にすることができるかどうか検討してみます。使われている回答を見ながら、「保護施設からの引き取り」に分類するキーワードを選んでいきます(これは、著者の主観による部分もあるので、自分自身で判断してみてください)。

保健所 「保健所」は「引き取り」、「譲りうけ」などの語が近くにあり、「保護施設からの引き取り」に関連した回答である

ことが分かりますので、この分類のキーワードに加えてみます。

引き取る これも保健所と同様、「保護施設からの引き取り」に関連した回答であることが分かります。

捨てる これは、「保護施設からの引き取り」に関連した意見もありますが、捨て猫、捨て犬関連の記述も多いです。この分類のキーワードに加えないことにします。（「捨て」などの活用形も「捨てる」に分類されています）

子 これも、同様にこの分類のキーワードに加えないことにします。

施設 「保護施設からの引き取り」に関連した意見が多いですが、宿泊施設など別の意見もあります。「保護施設からの引き取り」に関連した意見の多くは、「保護」という語も含んでいるのでそちらでカバーできると考え、これも、この分類のキーワードに加えないことにします。

知る この分類のキーワードに加えないことにします。

購入 「保護施設からの引き取り」だけではなく多様な意見です。この分類のキーワードに加えないことにします。

シヨップ 「保護施設からの引き取り」だけではなく多様な意見です。この分類のキーワードに加えないことにします。

保護 「保護施設からの引き取り」に関連した語であることが分かります。この分類のキーワードに加えてみます。

そこで、著者の案として、次の語のいずれかを含む回答を「保護施設からの引き取り」に関連した意見を述べた回答とします。ただし、「保護施設からの引き取り」に関して、肯定的、否定的、両方の意見があるので、すべて肯定的として分析するのではなく、関連した意見として分類します。

保健所、引き取る、保護

記述内容から分類するには、人間による判断が必要なこともあります。ただし、人間による判断は、判断する人の主観が入り、人により異なることが発生します。ここでおこなった語を含むかどうかでの分類はある程度客観的に分類できますが、語の選択自体に主観が入ります。

9.3.2 Excel:文書×抽出語表による分類

KH Coder で、「文書×抽出語表」で、各回答が抽出語を何個含むのかの一覧表を見ることができます。そのファイルが、「文書抽出語表_2023.xlsx」で配布しています。

	A	B	C	D	E	F	CV	CW	CX	CY	CZ	DA
1	h1	h2	h3	h4	h5	id	心	子		保護施設からの引き取り		
2	0	0	0	0	1	1	0	0		0		
3	0	0	0	0	2	2	0	0		0		
4	0	0	0	0	3	3	0	0		0		
5	0	0	0	0	4	4	0	0		0		
6	0	0	0	0	5	5	0	0		0		

図 21 自由記述分析：回答別抽出語表（CY 列は計算式で設定）

図 21 から、各回答（各行）のなかで、「保健所」、「引き取る」、「保護」の列のどれかが 1 以上のとき、「保護施設からの引き取り」に関連した意見があるとします。「保健所」は AD 列、「保護」は AL 列「引き取る」は BZ 列、に出現回数が書かれているので、3 つの合計が 1 以上であれば、保護施設からの引き取りの列を 1 に、そうで無い場合 0 にします。CY 列にその計算式を設定します。

CY1:	保護施設からの引き取り
CY2:	=IF(AD2+AL2+BZ2>0,1,0)
複写元 CY2	複写先 CY3:CY1085

動画:アンケート：テキスト分析：「保護施設からの引き取り」の記述の有無を計算

9.3.3 Excel:テキスト分析用の Excel ファイルに追加

enq_freeWtext2023.xlsx に、enq_freeWtext2023_抽出語.xlsx の「保護施設からの引き取り」をの情報を追記します。
enq_freeWtext2023.xlsx を起動してください。

■自由記述の有無 自由記述欄に何らかの入力した回答と、無回答を別々に集計するため、Q 列に自由記述欄が記入したか (1)、記入なし (0) を表示します。

Q1:	自由記述有無
Q2:	=IF(P2<>"" ,1,0)
複写元 Q2	複写先 Q3:Q1085

■保護施設からの引き取りの転記 enq_freeWtext2023.xlsx の R 列に、「文書×抽出語」表（文書抽出語表_2023.xlsx）の CY 列（保護施設からの引き取り）の値を値貼り付けで、コピーアンドペーストします。

複写元 文書抽出語表_2023.xlsx の CY1:CY1085 複写先 (値貼り付け) enq_freeWtext2023.xlsx の R 列

動画:アンケート：テキスト分析：「保護施設からの引き取り」のを転記

9.4 自由記述欄の分類による分析 (ピボットテーブル)

「保護施設からの引き取り」に分類された回答についてピボットテーブルを使って分析をします。例として、男女別にどれくらいの割合で「保護施設からの引き取り」に分類された回答をしているかを求めます。

図 22 自由記述分析：ピボットテーブルの設定

図 22 は、Excel のピボットテーブル設定画面のスクリーンショットです。左側のワークシートには、性別（女性、男性、空白）ごとの「保護施設からの引き取り」の件数（個数）が示されています。右側の「ピボットテーブルの...」タスクパネルには、フィールド「保護施設からの引き取り」が選択されており、集約方法として「個数」が選ばれています。赤い枠で「個数」が囲まれており、青い吹き出しで「個数に変更」という注釈が付けられています。

行ラベル	女性	男性 (空白)	総計
0	479	523	23
1	35	23	1
総計	514	546	24

図 22 自由記述分析：ピボットテーブルの設定

(1) A1:R1085 を範囲指定します。

(2) メニューの **挿入** → **ピボットテーブル** → **テーブルまたは範囲から**

(3) 図 22 のように設定します。

列 **性別** をドラッグアンドドロップします。男女無回答別に分析します。

行 **保護施設からの引き取り** をドラッグアンドドロップします。1 が「保護施設からの引き取り」に関する回答を

した数、0 がそれ以外の回答数になります。

値 保護施設からの引き取り で、値フィールドを 個数 にします。

(4) 図 22 の左の表が作成されます。

動画:アンケート：テキスト分析：「保護施設からの引き取り」のピボットテーブルでの分析

人数	女性	男性	記入無し	全体
保護施設記述無し	479	523	23	1025
保護施設記述有り	35	23	1	59
合計	514	546	24	1084
割合	女性	男性	記入無し	全体
保護施設記述無し	93.19%	95.79%	95.83%	94.56%
保護施設記述有り	6.81%	4.21%	4.17%	5.44%

図 23 自由記述分析：「保護施設からの引き取り」に関する得る回答の人数と割合

図 23 は、図 22 を見やすいようにまとめたものです。5.2 節で書いたように、=B5 のように、(クリックを使わずに) 手入力で計算式を設定しましょう。

図 24 のように設定するすれば、自由記述欄に記入した人のみを対象にした「保護施設からの引き取り」を記載した人の割合を求めることができ、図 25 のようにまとめることができます。

図 24 自由記述分析：自由回答の有無別ピボットテーブルの設定

	女性	男性	回答無し	全体
回答者数	514	546	24	1084
自由記述あり	388	419	15	822
保護施設記述あり	35	23	1	59
割合				
保護施設記述あり（分母：全回答）	6.81%	4.21%	4.17%	5.44%
保護施設記述あり（分母：自由記述ありの回答）	9.02%	5.49%	6.67%	7.18%

図 25 自由記述分析：自由回答の有無別「保護施設からの引き取り」に関する回答の人数と割合

9.5 練習問題

- 階層的クラスター分析の図などで、「責任」と「命」の分析をしてみましょう。
 - (1) 「責任」と「命」の両方の語を分析対象にするか、片方のみを分析対象の語にするかのも検討してください。
 - (2) 文書抽出語表_2023.xlsx の CZ 列に、この単語（もしくは、単語群）が含まれているかどうかを示しましょう。
 - (3) 文書抽出語表_2023.xlsx の S 列に (3) の CZ 列を転記（値貼り付け）してください。
 - (4) ピボットテーブル等を利用して、分析をしてみましょう。
- 階層クラスター分析の図、共起ネットワーク図、抽出語リストなどを見て、何らかの共通した意見を分析しましょう。
 - (1) 階層クラスター分析の図の 1 つまたは複数のクラスター、共起ネットワーク図の Subgraph などを見て、何らかの意見や感想で使われている語を検討します。
 - (2) enq_freeWtext2023_抽出語.xlsx の DA 列以降に、このキーワードの語（もしくは、語群）が含まれているかどうかを示しましょう。
 - (3) enq_freeWtext2023.xlsx の T 列以降に (2) の DA 列 (以降) を転記（値貼り付け）してください。
 - (4) ピボットテーブル等を利用して、分析をしてみましょう。