

aa-3. 実データによるデータサイエンス・AI の演習（オープンデータ、政府統計データ、 クロス集計表、相関

(人工知能)

URL: <https://www.kkaneko.jp/ai/mi/index.html>

金子邦彦



アウトライン

1. オープンデータ
2. 政府統計データ (Excel を使用)
3. クロス集計表 (Excel を使用)
4. 相関, 相関係数 (Excel を使用)

今日の授業では, **Office 365** のインストールを済ませておくことが便利です.

3-1 オープンデータ

オープンデータ

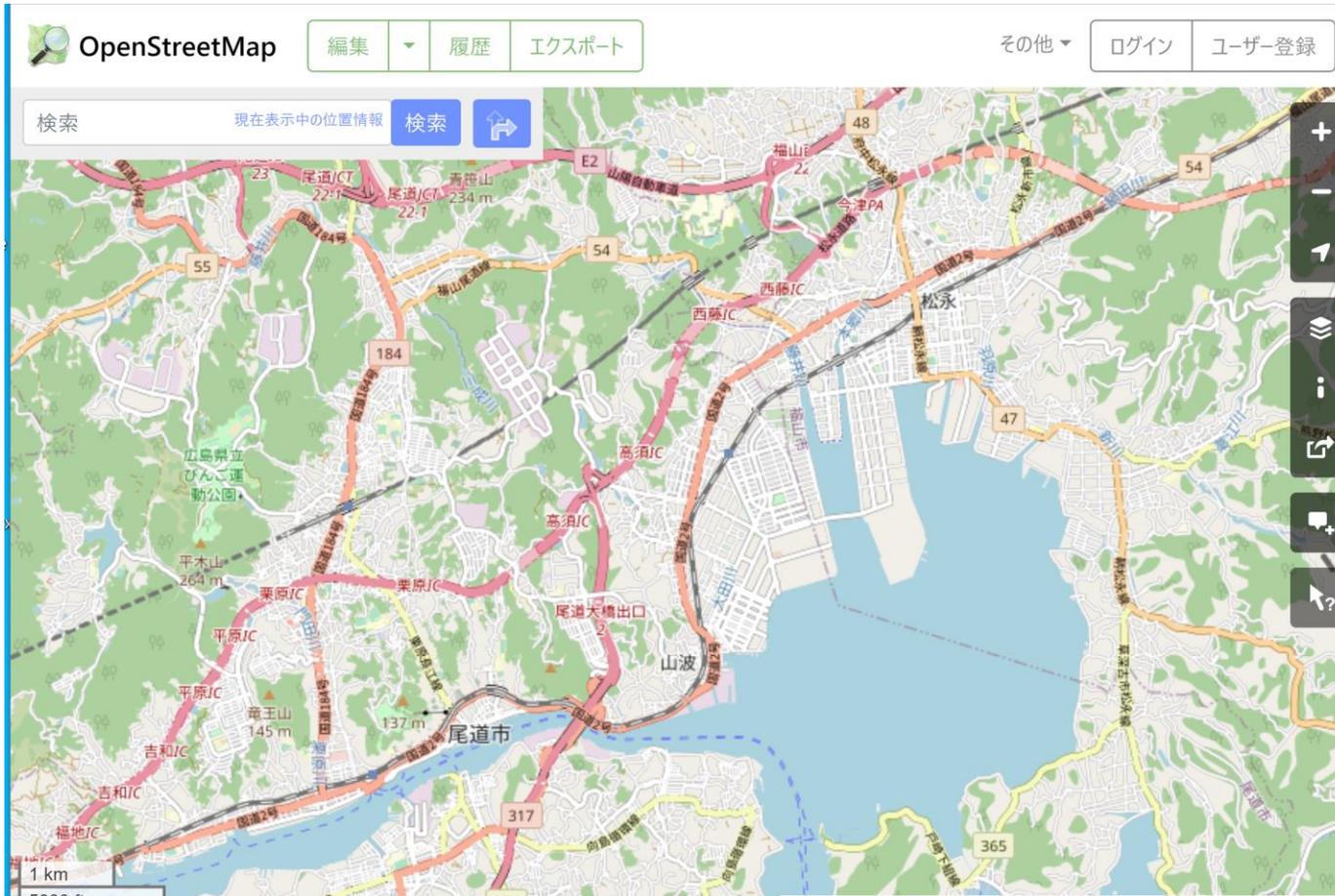


- **オープンデータ**は、誰でも自由にアクセスし、利用できるように公開されたデータ
- 政府, 企業, 個人のデータなどさまざま

オープンデータにより生まれたサービス



- **OpenStreetMap** <https://www.openstreetmap.org/>
オンラインの地図サービス



オープンデータのメリット



① 新しい価値の創造

新しいビジネスやサービスや研究

② コストの節約

データ収集や整理にかかる必要を節約

③ 透明性の向上

政府や企業の活動が，オープンデータにより透明になる

オープンデータの例



- 地震予知, 防災

気象庁強振観測データ

<https://www.data.jma.go.jp/eqev/data/kyoshin/jishin/index.html>

気象庁震度データ

<https://www.data.jma.go.jp/svd/eqev/data/bulletin/shindo.html>

- 環境調査

環境省大気汚染状況データ <https://www.env.go.jp/air/osen/>

環境省水環境総合情報サイト <https://water-pub.env.go.jp/water-pub/mizu-site/mizu/kousui/dataMap.asp>

- 医療情報

国立がん研究センターがん情報サービス <https://ganjoho.jp/public/index.html>

- 自然言語処理

Wikipedia <https://ja.wikipedia.org/wiki>

オープンデータ利用上の注意点



どのようなデータであれ，利用者側が注意深く利用する必要
がある

- データの品質や信頼性，プライバシーの侵害をしていないことを確認してから，利用
- 著作権を尊重，出典明示，公開者が定める利用条件を遵守

「公開されているデータだから信頼できる，何をしてもよい」というわけではない

オープンデータと人工知能の関係



- **機械学習**は、データを用いて、意思決定や問題解決を行うことができる人工知能の技術
- 機械学習に必要なデータとして、**オープンデータ**を利用可能
- 人工知能研究者向けに、研究論文、プログラム、データの情報をオープンにしているサイトもある

Paper with Code <https://paperswithcode.com>

3-2 政府統計データ

政府統計データ



- 日本政府は、統計データを公表している
人口，労働，経済，環境など
- これは、ビジネス戦略，政策立案，研究，学習に
役立つデータである

総務省統計局が公開する統計表



- 人口
- 労働力
- 国民掲載
- 物価
- 社会生活基本調査

など多数

ダウンロードページ

<https://www.stat.go.jp/data/guide/download/index.html>

総務省統計局が公開する統計表



・ダウンロードページ

<https://www.stat.go.jp/data/guide/download/index.html>

人口推計の結果の概要 令和5年4月報より（抜粋）

人口推計
令和5年4月20日
総務省統計局

人口推計 － 2023年（令和5年）4月報 －

【2023年（令和5年）4月1日現在（概算値）】	
<総人口> 1億2447万人で、前年同月に比べ減少	▲60万人（▲0.48%）
【2022年（令和4年）11月1日現在（確定値）】	
<総人口> 1億2491万3千人で、前年同月に比べ減少	▲53万1千人（▲0.42%）
・15歳未満人口は 1447万4千人で、前年同月に比べ減少	▲28万8千人（▲1.95%）
・15～64歳人口は 7421万4千人で、前年同月に比べ減少	▲24万4千人（▲0.33%）
・65歳以上人口は 3622万5千人で、前年同月に比べ増加	1千人（0.00%）
<日本人人口> 1億2196万9千人で、前年同月に比べ減少	▲75万9千人（▲0.62%）

年齢（5歳階級）、男女別人口 Population Estimates by Age (Five-Year Groups) and Sex

年齢階級 Age groups	2023年4月1日現在（概算値） （令和5年） April 1, 2023 (Provisional estimates)			2022年11月1日現在（確定値） （令和4年） November 1, 2022 (Final estimates)					
	総人口 Total population			総人口 Total population			日本人人口 Japanese population		
	男女計 Both sexes	男 Male	女 Female	男女計 Both sexes	男 Male	女 Female	男女計 Both sexes	男 Male	女 Female
総数 Total	12447	6051	6395	124,913	60,742	64,171	121,969	59,283	62,687
0～4歳 years old	418	214	204	4,236	2,168	2,068	4,152	2,125	2,027
5～9	490	251	239	4,937	2,529	2,408	4,854	2,487	2,368
10～14	528	270	257	5,301	2,716	2,585	5,232	2,681	2,552
15～19	550	282	267	5,514	2,831	2,683	5,417	2,781	2,636
20～24	621	319	302	6,266	3,218	3,049	5,834	2,986	2,848
25～29	641	329	312	6,416	3,296	3,119	5,935	3,027	2,909
30～34	640	328	312	6,441	3,298	3,143	6,065	3,094	2,971
35～39	712	362	350	7,199	3,663	3,536	6,920	3,522	3,397
40～44	786	398	387	7,929	4,019	3,910	7,694	3,913	3,781
45～49	928	470	458	9,439	4,781	4,659	9,252	4,700	4,552
50～54	954	481	473	9,457	4,767	4,690	9,278	4,696	4,582
55～59	814	407	407	8,094	4,046	4,049	7,954	3,989	3,965
60～64	747	370	377	7,458	3,690	3,767	7,360	3,650	3,710
65～69	740	360	380	7,506	3,651	3,855	7,434	3,620	3,814
70～74	904	426	477	9,288	4,382	4,906	9,235	4,358	4,877
75～79	728	327	401	7,076	3,173	3,903	7,040	3,158	3,883
80～84	576	240	336	5,743	2,392	3,351	5,719	2,383	3,336
85～89	399	145	254	3,962	1,434	2,528	3,949	1,429	2,520
90～94	204	58	146	1,998	564	1,434	1,993	563	1,431
95～99	59	12	47	564	112	452	563	112	451
100歳以上 and over	9	1	8	88	11	77	88	11	77

労働力調査（基本集計） 2023年2月分結果より（抜粋）

	年平均			月次（季節調整値）			
	2020年	2021年	2022年	2022年11月	12月	2023年1月	2月
完全失業率	2.8%	2.8%	2.6%	2.5%	2.5%	2.4%	2.6%

2020年基準 消費者物価指数 全国 2023 3月分より（抜粋）

「ポイント」

- 総合指数は2020年を100として104.4
前年同月比は3.2%の上昇
- 生鮮食品を除く総合指数は104.1
前年同月比は3.1%の上昇
- 生鮮食品及びエネルギーを除く総合指数は103.2
前年同月比は3.8%の上昇

e-Stat は、政府統計の総合窓口

<https://www.e-stat.go.jp/>



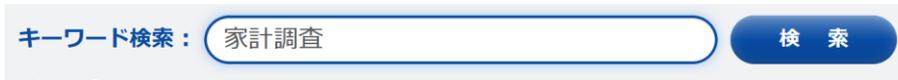
The screenshot shows the e-Stat website homepage. At the top left is the e-Stat logo with the tagline '政府統計の総合窓口'. To its right is the text '統計で見る日本' and a description: 'e-Statは、日本の統計が閲覧できる政府統計ポータルサイトです'. On the top right, there are links for 'お問い合わせ | ヘルプ | English' and a 'ログイン' button. Below the header is a navigation bar with links: '統計データを探す', '統計データの活用', '統計データの高度利用', '統計関連情報', and 'リンク集'. The main content area is divided into two columns. The left column features a '統計データを探す' section with a sub-header '(政府統計の調査結果を探します)' and a 'その他の絞り' button. It contains three large buttons: 'すべて' (政府統計一覧の中から探します), '分野' (17の統計分野から探します), and '組織' (統計を作成した府省等から探します). Below these is a search bar with the text 'キーワード検索: AED' and a '検索' button. The right column features a '統計データの高度利用' section with a '利用ガイド' button. It contains three buttons: 'マイクロデータの利用' (公的統計のマイクロデータの利用案内), '開発者向け' (API、LODで統計データを取得), and '統計関連情報' (統計分類・調査計画等). At the bottom of the left column, there is a '統計データを活用する' section with four buttons: 'グラフ' (主要指標をグラフで表示 (統計ダッシュボード)), '時系列表' (主要指標を時系列表で表示 (統計ダッシュボード)), '地図' (地図上に統計データを表示 (統計GIS)), and '地域' (都道府県、市区町村の主要データを表示).

① 統計データの検索

- ・ 分野による検索



- ・ キーワードによる検索, その他



② 必要な項目を選ぶ (表示項目選択)

③ レイアウトの設定

④ ダウンロード

手元のパソコンにファイルがダウンロードされる

家計調査



- 全国約9千世帯を対象として、**家計の収入・支出、貯蓄・負債**などを毎月調査

e Stat 家計調査のデータを用いた分析



「プリン」にお金を使うことが多い都市は？

	A	B	C	D	E	F	G	H	I	J	K
10											
11									/品目分類	1.1E+07	
12	表章項目	表章項目	世帯区分	世帯区分	時間軸 (月)	時間軸 (月)	地域区分	地域区分	/品目分類	548 プリン	【円】
13	1	金額	3	二人以上	0	2E+09	#####	0	全国		132
14	1	金額	3	二人以上	0	2E+09	#####	1003	0110	札幌市	107
15	1	金額	3	二人以上	0	2E+09	#####	2003	0220	青森市	146
16	1	金額	3	二人以上	0	2E+09	#####	3003	0320	盛岡市	109
17	1	金額	3	二人以上	0	2E+09	#####	4003	0410	仙台市	183
18	1	金額	3	二人以上	0	2E+09	#####	5003	0520	秋田市	126
19	1	金額	3	二人以上	0	2E+09	#####	6003	0620	山形市	122
20	1	金額	3	二人以上	0	2E+09	#####	7003	0720	福島市	217
21	1	金額	3	二人以上	0	2E+09	#####	8003	0820	水戸市	147
22	1	金額	3	二人以上	0	2E+09	#####	9003	0920	宇都宮市	182
23	1	金額	3	二人以上	0	2E+09	#####	10003	1020	前橋市	154
24	1	金額	3	二人以上	0	2E+09	#####	11003	1110	さいたま市	182

謝辞

<https://www.youtube.com/watch?v=0RSIrNTFnJw>

を参考にし、説明の追加等を行っている

演習

資料：20 ～ 32

【トピックス】

e-Stat 家計調査データの活用

**事前準備として Office 365 の
インストールが必要**

e Stat 家計調査データのダウンロード



<https://www.e-stat.go.jp/>

① 統計データの検索

次のどちらかの方法で

分野での検索：**企業・家計・経済** → **家計調査**

● 統計データを探す (政府統計の調査結果を探します) その他の絞込

すべて
政府統計一覧の中から探します

分野
17の統計分野から探します

組織
統計を作成した府省等から探します

キーワード検索： 検索

キーワード検索：**家計調査** → 一覧の中の「**家計調査**」をクリック

● 統計データを探す (政府統計の調査結果を探します) その他の絞込

すべて
政府統計一覧の中から探します

分野
17の統計分野から探します

組織
統計を作成した府省等から探します

キーワード検索： 検索

② 「家計調査」の画面で、スクロールして、少し下の方にあるデータベースをクリック

家計調査

詳細

一覧形式で表示

家計調査は、統計理論に基づき選定された全国約9千世帯を対象として、家計の収入・支出、貯蓄・負債などを毎月調査しています。

家計調査の結果は、我が国の景気動向の把握、生活保護基準の検討などの基礎資料として利用のほか、地方公共団体、民間の会社などでも利用されています。

二人以上の世帯の結果は、主に、地域・世帯属性ごとに1世帯当たり1か月間の収支金額にまとめ毎月公表、単身世帯及び総世帯の家計収支に関する結果並びに二人以上の世帯の貯蓄・負債に関する結果を四半期ごとに公表しています。

家計調査	データベース	件数 更新日	ファイル	件数 更新日	概要
家計調査		112件 2024-04-05		23,841件 2024-04-05	

検索結果をさらに絞込み

統計分野（大分類）で絞込み



③ 種類は「月次」を選ぶ

家計調査

詳細

一覧形式で表示

家計調査は、統計理論に基づき選定された全国約9千世帯を対象として、家計の収入・支出、貯蓄・負債などを毎月調査しています。

家計調査の結果は、我が国の景気動向の把握、生活保護基準の検討などの基礎資料として利用のほか、地方公共団体、民間の会社などでも利用されています。

二人以上の世帯の結果は、主に、地域・世帯属性ごとに1世帯当たり1か月間の収支金額にまとめ毎月公表、単身世帯及び総世帯の家計収支に関する結果並びに二人以上の世帯の貯蓄・負債に関する結果を四半期ごとに公表しています。



家計調査	公開（更新）日
■ 家計収支編 [85件]	
└ 二人以上の世帯 [55件]	
└ 月次 [16件]	2024-04-05
└ 四半期 [9件]	2024-02-06
└ 年次 [30件]	2024-02-06

④ 少し下の方にある「品目分類（2020年改訂）（総数：金額）」の「DB」を選ぶ

提供周期		月次		
表番号	統計表	調査年月	公開（更新）日	表示・ダウンロード
用途分類				
001	用途分類（総数）	-	2024-04-05	→ DB API
002	用途分類（年間収入五分位階級別）	-	2024-04-05	→ DB API
003	用途分類（世帯主の定期収入五分位階級別）	-	2024-04-05	→ DB API
004	用途分類（世帯人員別）	-	2024-04-05	→ DB API
005	用途分類（世帯主の年齢階級別）	-	2024-04-05	→ DB API
006	用途分類（世帯主の職業別）	-	2024-04-05	→ DB API
008	用途分類（世帯主の勤め先企業規模別）	-	2024-04-05	→ DB API
009	用途分類（住居の所有関係別）	-	2024-04-05	→ DB API
品目分類				
010	品目分類（2020年改訂）（総数：金額）	-	2024-04-05	→ DB API
011	品目分類（2020年改訂）（総数：数量）	-	2024-04-05	→ DB API
010	品目分類（平成27年改訂）（総数：金額）	-	2020-02-07	→ DB API
011	品目分類（平成27年改訂）（総数：数量）	-	2020-02-07	→ DB API
010	品目分類（平成22年改訂）（総数：金額）	-	2015-02-06	→ DB API
011	品目分類（平成22年改訂）（総数：数量）	-	2015-01-30	→ DB API
010	品目分類（平成17年改訂）（総数：金額）	-	2010-01-29	→ DB API
011	品目分類（平成17年改訂）（総数：数量）	-	2010-01-29	→ DB API

⑤ 「表示項目選択」をクリック。



The screenshot shows the e-Stat website interface. At the top, there is a navigation bar with the e-Stat logo and the text '統計で見る日本' (Japan seen through statistics). Below this, there are several menu items: '統計データを探す' (Search for statistics data), '統計データの活用' (Use of statistics data), '統計データの高度利用' (Advanced use of statistics data), '統計関連情報' (Statistics related information), and 'リンク集' (Link collection). The main content area is titled '統計表・グラフ表示' (Statistics table and graph display). It contains a table with the following information:

統計名	家計調査 家計収支編 二人以上の世帯
表番号	010
表題	品目分類 (2020年改定) (総数: 金額)

Below the table, there is a message: '統計表を表示するには項目の絞り込みや表示位置の変更が必要です。『表示項目選択』から項目を絞り込むか、『レイアウト設定』から表示位置のページ上部 (欄外) への変更を行ってください。(画面表示可能セル数 1)'

The interface also features a navigation bar with '統計表表示' (Statistics table display) and 'グラフ表示' (Graph display) options. The '表示項目選択' (Select display items) button is highlighted with a red box. Below the navigation bar, there are several dropdown menus for '表章項目' (Table item), '金額' (Amount), '世帯区分' (Household type), '地域区分' (Region type), and '再表示' (Refresh). The '表示項目選択' button is located in the left sidebar.

⑥ 「品目分類」の「項目を選択」をクリック

2/5	品目分類 (2020年改定)	690/690	<input type="button" value="項目を選択"/>
-----	----------------	---------	--------------------------------------

「全解除」をクリック



「プリン」で検索

- 世帯人員
- 18歳未満人員
- 65歳以上人員
- 65歳以上無職者人員
- 有業人員

※複数項目を選ぶ場合は、Ctrlキーを押しながらクリックしてください。
連続した範囲であれば、マウスの左クリックを押した状態でマウスを移動するか、Shiftキーを押しながらクリックしてく

<input type="checkbox"/> 選択状態の反転	<input type="button" value="反転"/>
<input checked="" type="checkbox"/> 複数項目の選択/解除	<input type="button" value="選択"/> <input type="button" value="解除"/>
<input checked="" type="checkbox"/> 配下グループの選択/解除	<input type="button" value="選択"/> <input type="button" value="解除"/>
<input checked="" type="checkbox"/> 同一階層の選択/解除	<input type="button" value="選択"/> <input type="button" value="解除"/>
<input checked="" type="checkbox"/> 時間軸の範囲選択	<input type="button" value="選択"/> <input type="button" value="解除"/>

プリン

「プリン」にチェック。「確定」をクリック

表示切替 選択: 1 全項目: 690 未選択: 689

- 348 プリン
- 345 他の洋生菓子
- 350 せんべい

キャンセル

確定

⑦ 「時間軸（月次）」の「項目を選択」をクリック



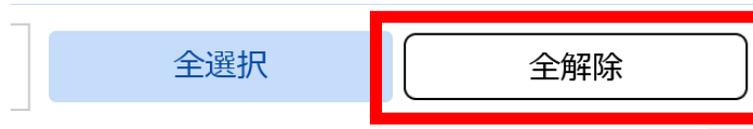
5/5

時間軸（月次）

458/458

項目を選択

「全解除」をクリック



スクロールし、「2024年2月」をチェック

2023年11月

2023年12月

2024年1月

2024年2月

「確定」をクリック

キャンセル

確定

⑧ これで、表示項目選択が終わった。「確定」をクリック。（間違いがあっても、あとで、やり直すことができる）

統計表・グラフ表示

統計名	家計調査 家計収支編 二人以上の世帯
表番号	010
表題	品目分類（2020年改定）（総数：金額）

統計表を表示するには項目の絞り込みや表示位置の変更が必要です。「表示項目選択」から項目を絞り込むか、「レイアウト設定」から表示位置のページ上部（欄外）への変更を行ってください。（画面右上参照）

統計表表示

表示項目選択

初期状態は全項目表示状態です。表示を変更するには、以下の手順で設定を変更してください。

1. **項目を選択** ボタンをクリックして表示項目を設定してください。
2. **確定** ボタンを押すと変更内容の表示が更新されます。

画面表示セル数： 1
画面総セル数： 212
(統計表は最大100,000セルまで表示可能、グラフは最大15,000セルまで表示可能)

項目	事項名	説明	選択/全項目 ?
1/5	表章項目		1/1 項目を選択
2/5	品目分類（2020年改定）		1/690 項目を選択
3/5	世帯区分		4/4 項目を選択
4/5	地域区分		53/53 項目を選択
5/5	時間軸（月次）		1/458 項目を選択

初期状態に戻す（全項目表示）

キャンセル **確定**

⑨ 「レイアウト設定」をクリック

表示項目選択 >

レイアウト設定 >

表示項目	金額
世帯区分	二人以上の世帯 (2000年~)
	348 プリン [円]
2024年2月	149

⑩ 「地域区分」は「行」のところにドラッグ&ドロップ。「設定して表示を更新」

レイアウト設定 ?

レイアウト変更は、下記の項目名の を押して「ページ上部」「列」「行」を選択して項目を移動してください。表示順の変更は を押して調整してください。変更を保存して表示に反映させるには、[設定して表示を更新](#) を押してください。

1 レイアウト設定

ページ上部 (欄外)	列
<ul style="list-style-type: none"> 表示項目 世帯区分 	<ul style="list-style-type: none"> 品目分類 (2020年改定)
<ul style="list-style-type: none"> 行 地域区分 	

画面表示セル数: 53
画面総セル数: 212
(最大 100,000 セルまで表示可能)

キャンセル [設定して表示を更新](#)

⑪ プリンについての結果を確認

表示項目選択 > レイアウト設定

表章項目		金額	世帯区分	二人以上の世帯 (2000年~)
		348 プリン 【円】		
2024年2月	全国			149
	01100 札幌市			149
	02201 青森市			166
	03201 盛岡市			113
	04100 仙台市			193
	05201 秋田市			221
	06201 山形市			147

⑫ 「ダウンロード」を3回クリック

↓ ダウンロード API ✕

キャンセル ダウンロード

No.	ファイル名	サイズ	ダウンロード
1	FEH_00200561_220505211256.csv	6.0 K	ダウンロード

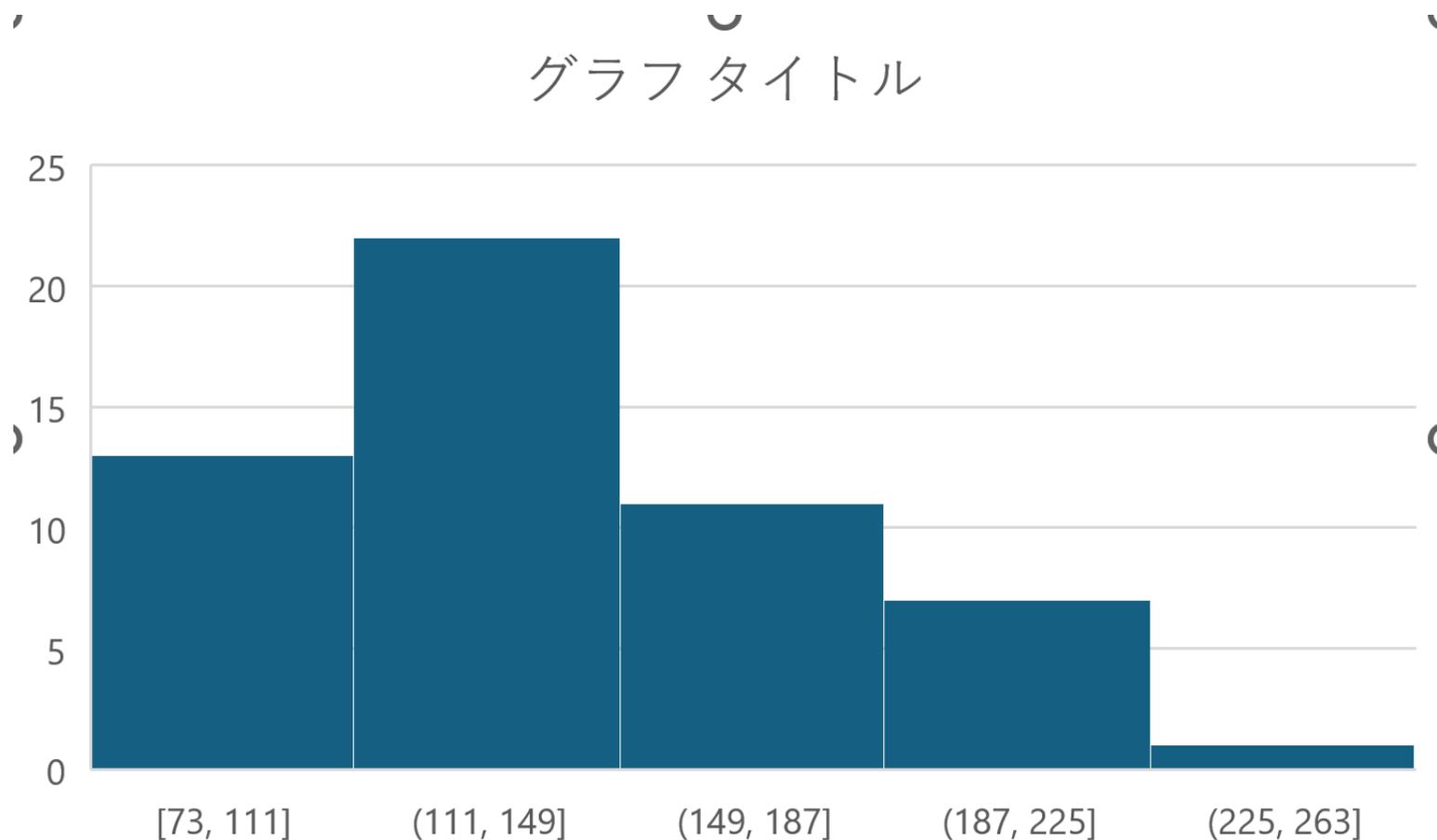
キャンセル

⑬ ダウンロードしたファイルを Excel で開くことができる

12	表章項目	表章項目	世帯区分	世帯区分	時間軸 (月)	時間軸 (月)	地域区分	地域区分	/品目分類	348 プリン 【円】
13	01	金額	03	二人以上	2.02E+09	#####	00000	全国		149
14	01	金額	03	二人以上	2.02E+09	#####	01003	01100 札幌市		149
15	01	金額	03	二人以上	2.02E+09	#####	02003	02201 青森市		166
16	01	金額	03	二人以上	2.02E+09	#####	03003	03201 盛岡市		113
17	01	金額	03	二人以上	2.02E+09	#####	04003	04100 仙台市		193
18	01	金額	03	二人以上	2.02E+09	#####	05003	05201 秋田市		221
19	01	金額	03	二人以上	2.02E+09	#####	06003	06201 山形市		147
20	01	金額	03	二人以上	2.02E+09	#####	07003	07201 福島市		231
21	01	金額	03	二人以上	2.02E+09	#####	08003	08201 水戸市		223

⑭ ヒストグラムの作成

(セルの範囲を選び, 「挿入」, ヒストグラムを選ぶ)



① 統計データの検索

分野での検索：**企業・家計・経済**→**家計調査**

キーワード検索：**家計調査**

② 必要な項目を選ぶ（表示項目選択）

品目分類：プリン

時間軸（月次）：最新月

③ レイアウトの設定

ページ, 行, 列の設定

④ ダウンロード

まとめ



- **政府統計データは、様々な分野を網羅**

例えば、家計調査では、プリンの支出状況などが調査されている。

- **政府統計データの活用により、さまざまな分析が可能**
どの都市がプリンが好まれているかを知ることができる

- このようなデータを個人で収集することはほぼ不可能

- **政府統計データを利用することで、広範なデータを取得し、重要な洞察を得ることができる。**

演習

資料：36

【トピックス】

**e-Stat 家計調査データの活用，
各自の自己研鑽**

- 各自で、**e Stat 家計調査データ**で、**プリン以外の品目**について調べてみる
- 余裕があれば、Excel を使い、**ヒストグラム**を作成してみる

3-3 クロス集計表

クロス集計表



クロス集計表は、**2つ以上のデータを分析**するために使用される表

(例) **最終卒業学校卒業時の居住地と性別**

- データ全体の把握, 比較, 分析が容易になる

クロス集計表の例

名前	性別	申し込み
A	男性	済
B	男性	未
C	女性	済
D	女性	未
E	男性	済
F	男性	未

元データ



	男性	女性
済	2	1
未	2	1

クロス集計表

クロス集計表の例

最終卒業学校卒業のあと、引っ越しをしていないのは全体の何パーセントか？

最終卒業学校卒業時の居住地（男女別）

	まだ卒業していない	現在と同じ居住地	その他
男性	19.2	17.8	63.0
女性	17.3	10.6	72.1

パーセント

ここから導かれる結論は？
性別と居住地についての関連

その他：
居住地が違う，不詳

- **元データからクロス集計表の作成は、Excel を用いて簡単にできる**

Excel でピボットテーブル（クロス集計用）を作成（1 / 5）



次のファイルを使用

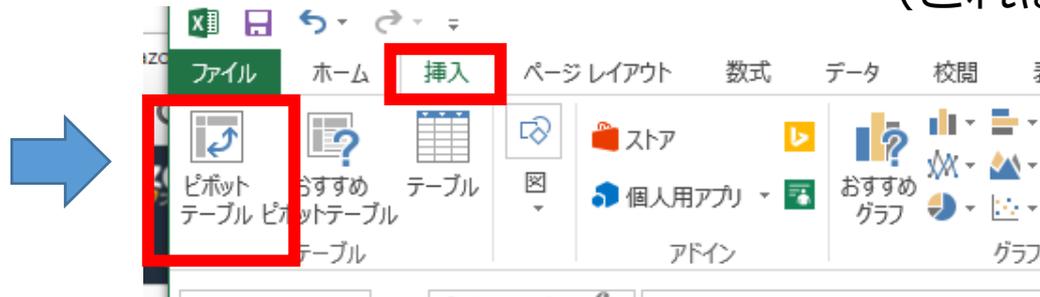
<https://www.kkaneko.jp/ai/mi/32.xlsx>

	A	B	C
1	名前	性別	申し込み
2	A	男性	済
3	B	男性	未
4	C	女性	済
5	D	女性	未
6	E	男性	済
7	F	男性	未
8			

元データ

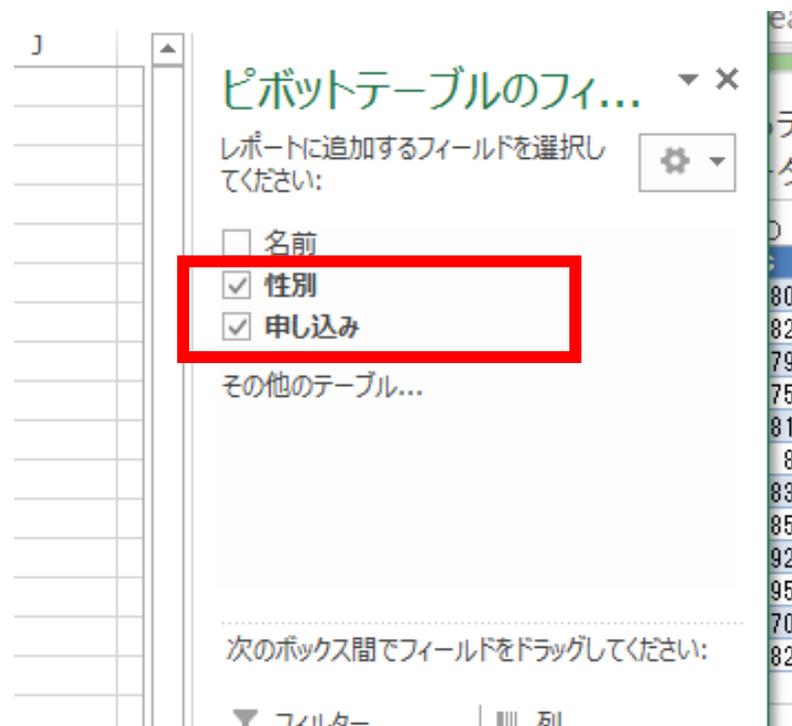
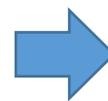
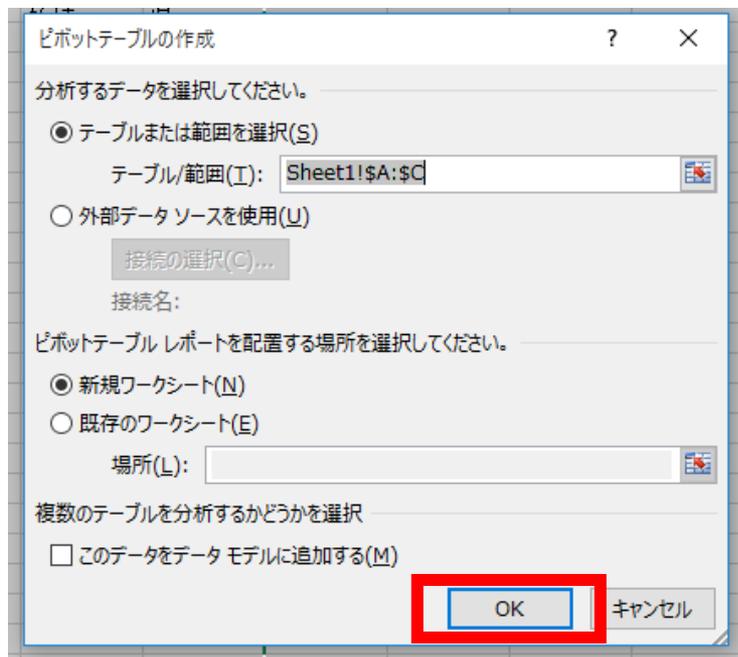
	A	B	C
1	名前	性別	申し込み
2	A	男性	済
3	B	男性	未
4	C	女性	済
5	D	女性	未
6	E	男性	済
7	F	男性	未
8			

- ① 左上をクリック. 全セルが選択される
(これは**範囲の選択**)



- ② リボンで「挿入」→「ピボットテーブル」

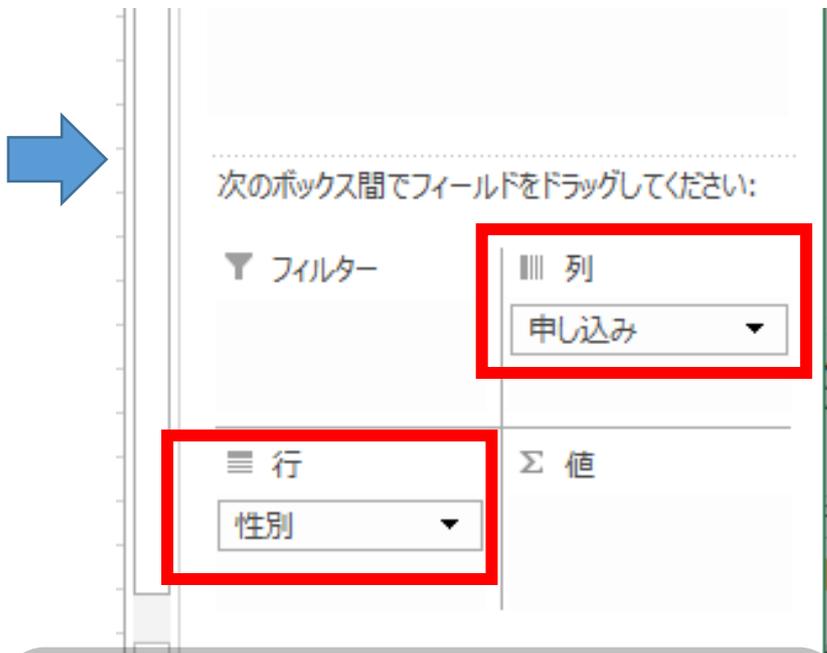
Excel でピボットテーブル（クロス集計用）を作成（2 / 5）



③ 「OK」をクリック

④ 「性別」と「申し込み」をチェック

Excel でピボットテーブル (クロス集計用) を作成 (3 / 5)



⑤ 列のところに「申し込み」が、
行のところに「性別」が来るよう
に、**ドラッグして調整**



Excel でピボットテーブル（クロス集計用） を作成（4 / 5）



ピボットテーブルのフィ...

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

- 名前
- 性別
- 申し込み

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
	申し込み

■ 行	Σ 値
性別	
名前	

⑥ 「名前」をチェック

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
	申し込み

■ 行	Σ 値
性別	
名前	

⑦ 行の下の「名前」をドラッグして、

▼ フィルター	■ 列
	申し込み

■ 行	Σ 値
性別	個数 / 名前

⑧ 値の下にドロップ

Excel でピボットテーブル（クロス集計用） を作成（5 / 5）



	A	B	C	D	E
1					
2					
3	個数 / 名前 列ラベル ▼				
4	行ラベル ▼	未	済 (空白)	総計	
5	女性	1	1		2
6	男性	2	2		4
7	(空白)				
8	総計	3	3		6
9					

クロス集計表ができた

3-4 相関

- **相関**は、2つの変数の間に関連性があるかを示す
- **相関がある**場合、一方が変化すると、もう一方も変化
する傾向にある

【相関ありの場合】

- Xが増えると、Yが増える傾向がある（**正の相関**）
勉強時間が増えると、得点上がる
- Xが増えると、Yが減る傾向がある（**負の相関**）
ガソリン代が上がると、車の利用が減る

【相関なしの場合】

XとYに関係がない

足のサイズと勉強時間に関係がない

相関係数



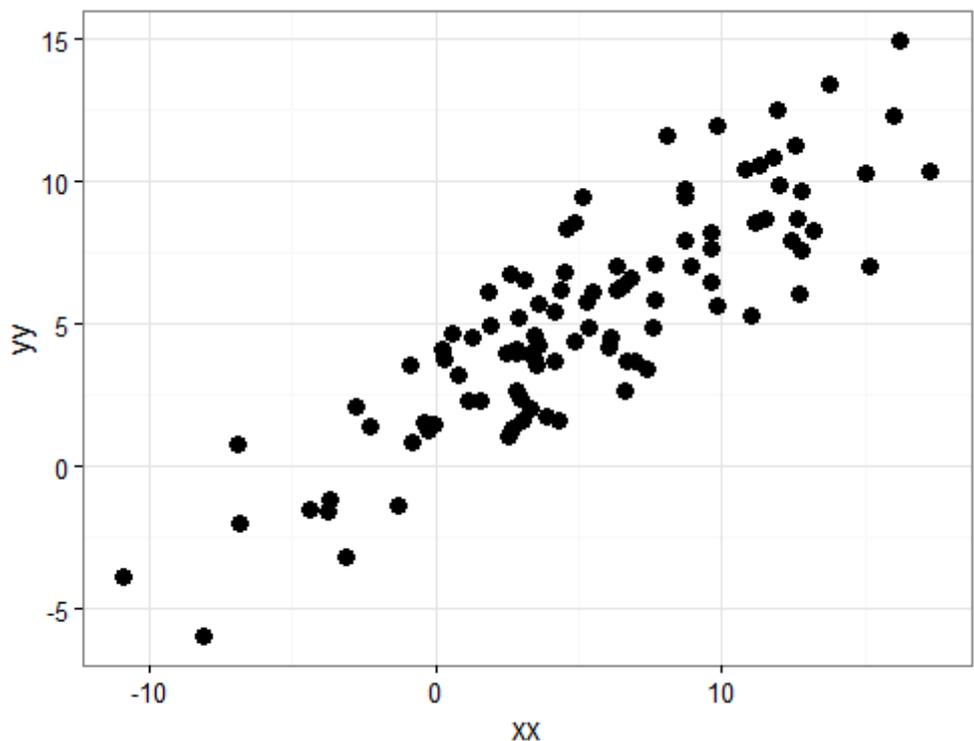
- **相関係数**は、**相関**を算出した**数値**。範囲は**-1から1まで**
- 相関係数を算出することで、変数間の関係の分析ができる

1に近い値： **相関**あり。正の相関

0に近い値： **相関なし**

-1に近い値： **相関**あり。負の相関

正の相関



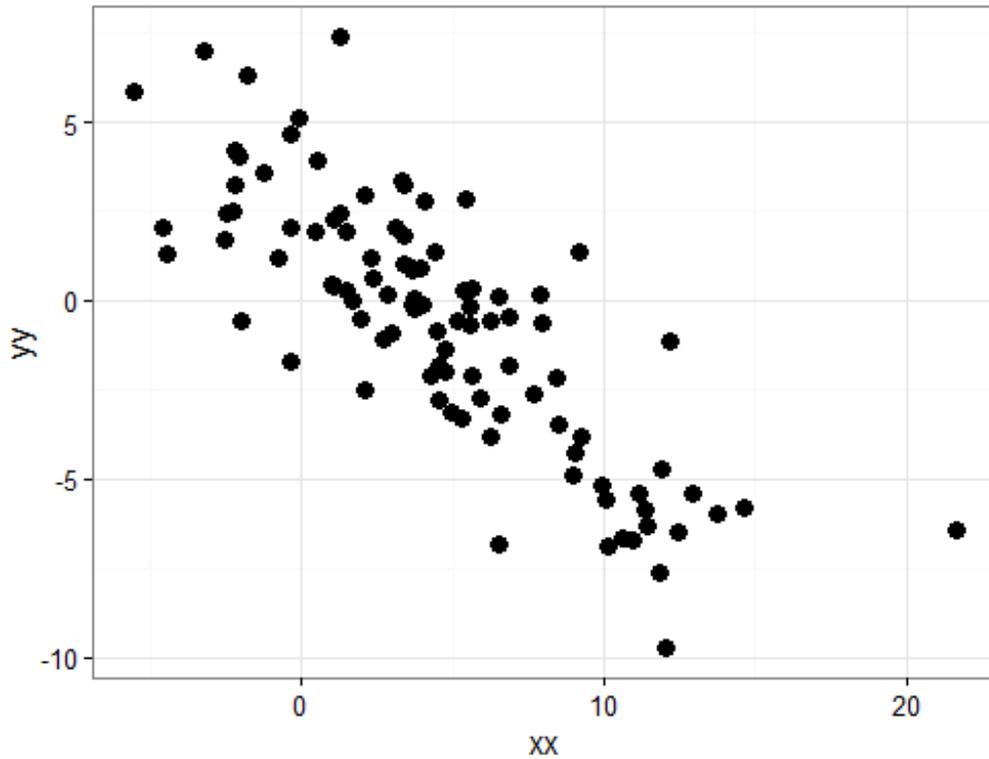
2つの変数 xx , yy に相関がある

xx の値が増えると
 yy の値が増える傾向がある
(正の相関)

相関係数の算出結果

0.8620027 (1 に近い値)

負の相関



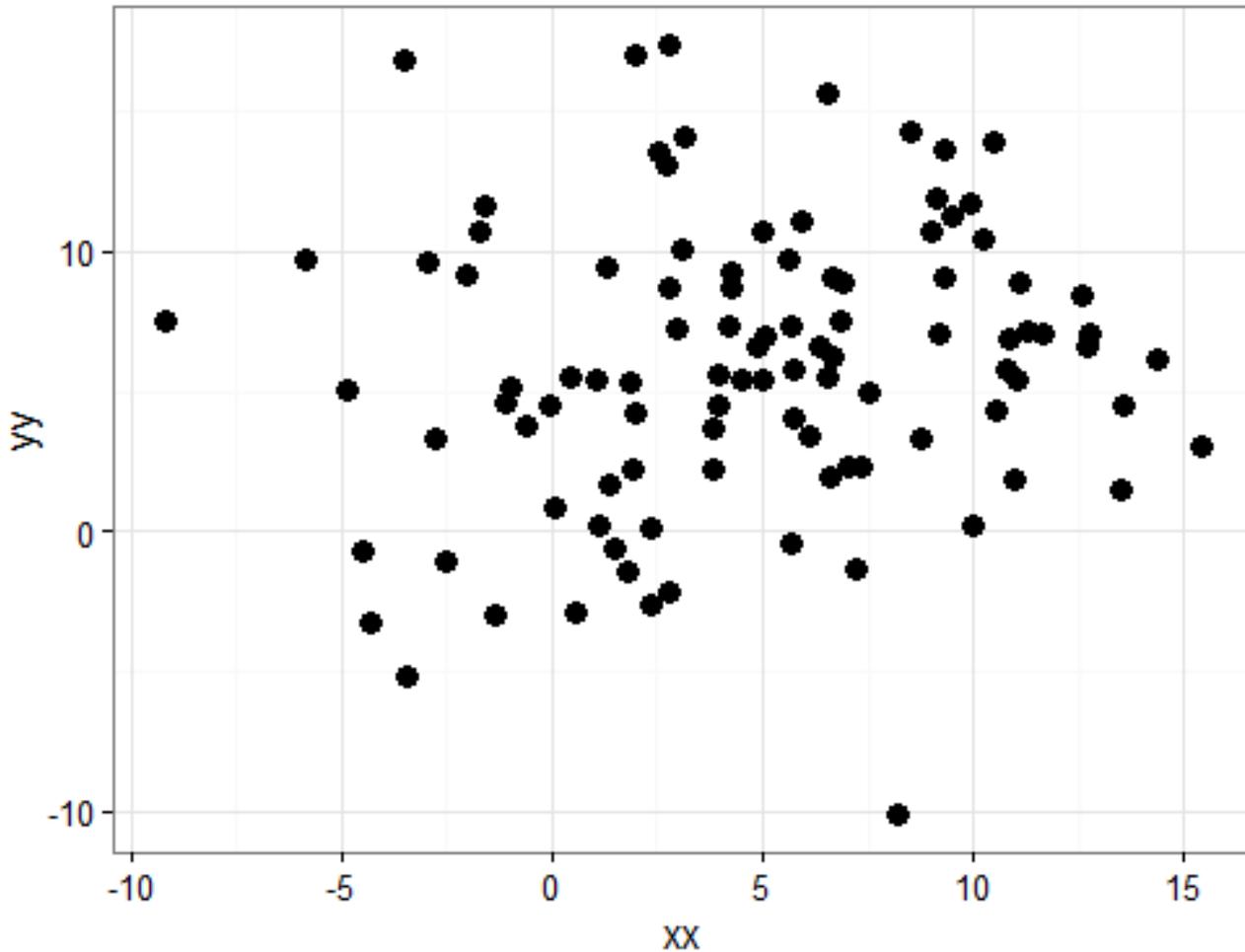
2つの変数 xx , yy に
相関がある

xx の値が増えると
 yy の値が減る傾向がある
(負の相関)

相関係数の算出結果

-0.8502535 (-1 に近い値)

相関なし



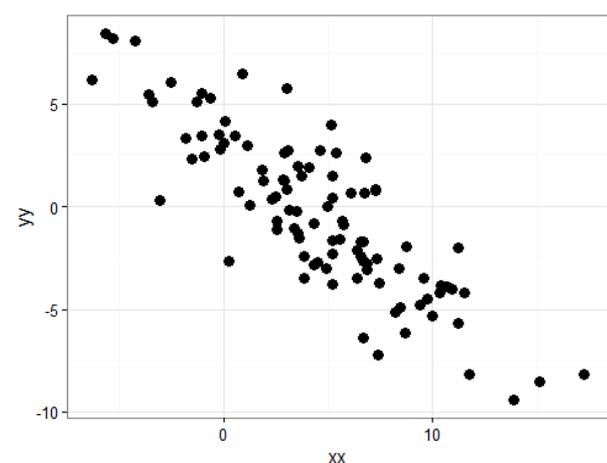
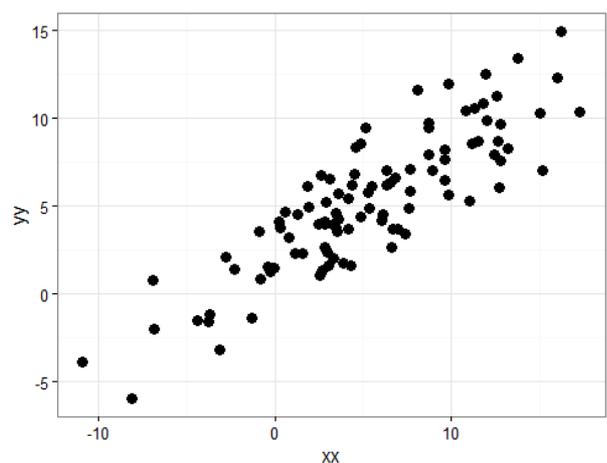
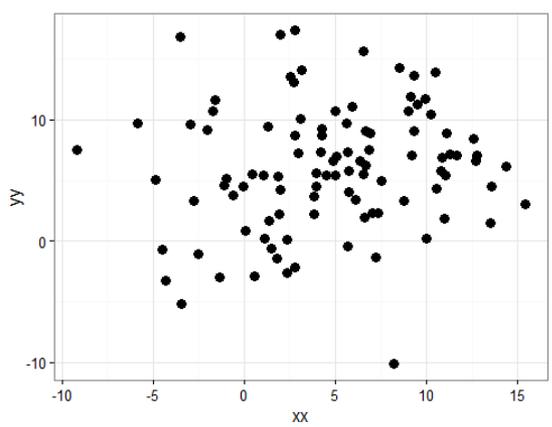
相関係数の算出結果

0.1252164 (0 に近い値)

相関係数のまとめ



- 1に近い値： 相関あり。 正の相関
- 0に近い値： 相関なし
- -1に近い値： 相関あり。 負の相関



0.1252164 (0に近い値)

相関なし

0.8620027 (1に近い値)

正の相関

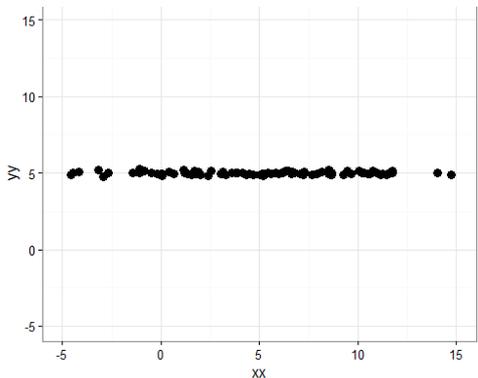
-0.8502535 (-1に近い値)

負の相関

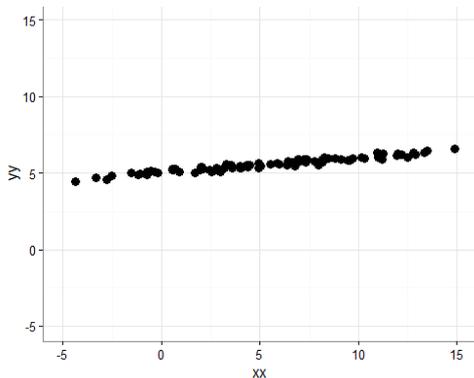
相関係数の性質



「相関の強弱」の尺度である。「傾き」ではない

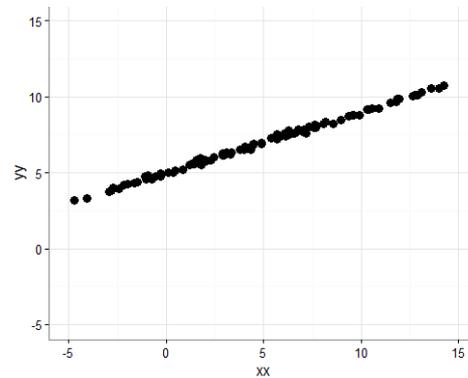


```
> cor(d9$xx, d9$yy)
[1] -0.06027409
```



```
> cor(d10$xx, d10$yy)
[1] 0.9743955
```

1 に近い値



```
> cor(d11$xx, d11$yy)
[1] 0.998944
```

1 に近い値

2つの量の間の関係性の分析

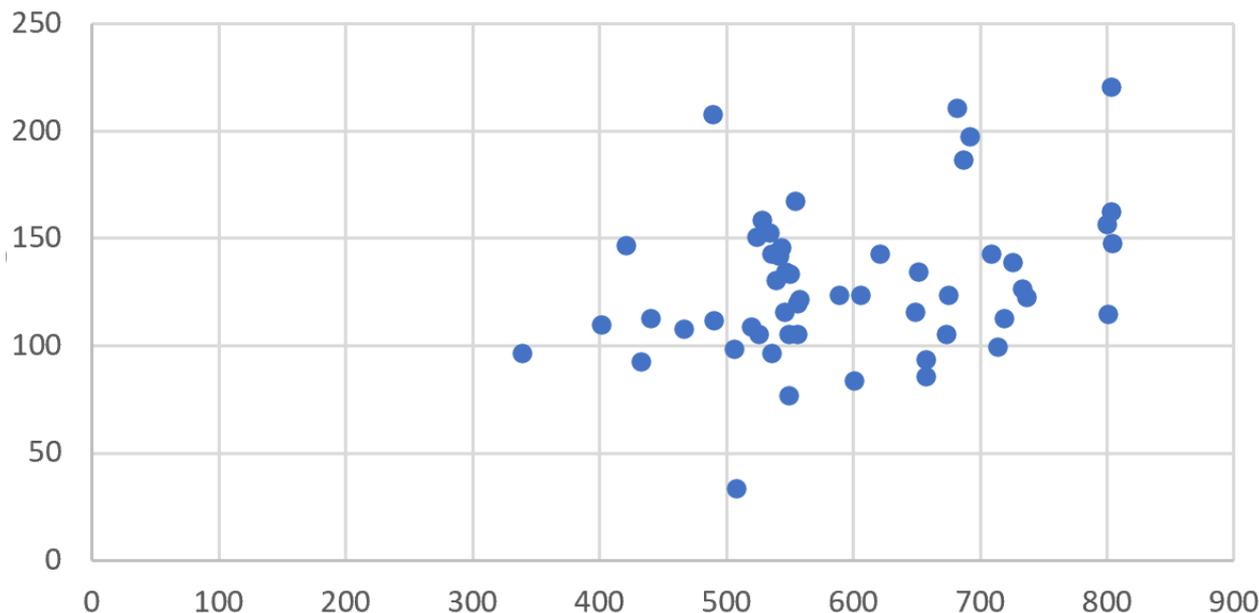
- 広告を増やすと，売上高が増えそうか
- 相関が高い複数の金融商品を扱うと，リスクが高いか
- 遺伝子と疾患に関係がありそうか

- プリンとケーキの関係を分析を説明
- 政府統計データを使用

e Stat 家計調査のデータを用いた分析



「プリン」にお金を使うことと、「ケーキ」にお金を使うことに**相関**があるか？



横軸：ケーキ，縦軸：プリン

e Stat 家計調査のデータの活用



- e Stat で、表示項目として、**ケーキとプリン**を選択

項目の表示・非表示を設定してください

2/5 品目分類 (2020年改定) ?

表示切替 選択: 2 全項目: 690 未選択: 688 全選択

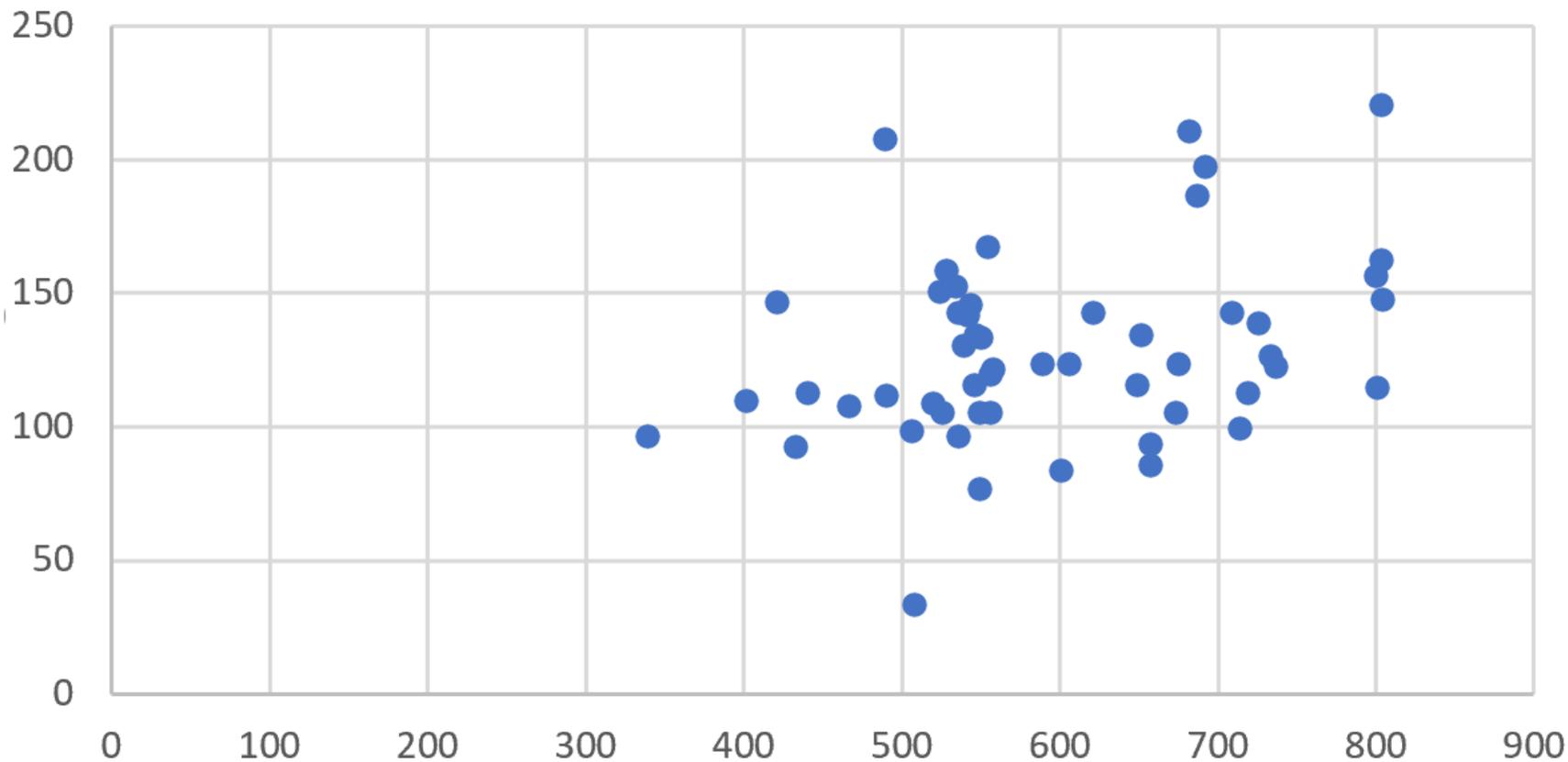
1.8 果子类

- 340 ようかん
- 341 まんじゅう
- 342 他の和生菓子
- 343 カステラ
- 344 ケーキ
- 347 ゼリー
- 348 プリン

- データをダウンロード, Excel で開く

A	B	C	D	E	F	G	H	I	J	K
表章項目	表章項目	世帯区分	時間軸	時間軸 (月)	地域区分	地域区分	品目分類	344 ケー	348 プリン	
2	1 金額	3 二人以上	2E+09	#####	0 全国			551	133	
3	1 金額	3 二人以上	2E+09	#####	1003 01100 札幌市			529	158	
4	1 金額	3 二人以上	2E+09	#####	2003 02201 青森市			340	96	
5	1 金額	3 二人以上	2E+09	#####	3003 03201 盛岡市			804	162	
6	1 金額	3 二人以上	2E+09	#####	4003 04100 仙台市			658	85	
7	1 金額	3 二人以上	2E+09	#####	5003 05201 秋田市			715	99	
8	1 金額	3 二人以上	2E+09	#####	6003 06201 山形市			441	112	
9	1 金額	3 二人以上	2E+09	#####	7003 07201 福島市			527	105	
0	1 金額	3 二人以上	2E+09	#####	8003 08201 水戸市			537	142	
1	1 金額	3 二人以上	2E+09	#####	9003 09201 宇都宮市			490	207	
2	1 金額	3 二人以上	2E+09	#####	10003 10201 前橋市			607	123	
3	1 金額	3 二人以上	2E+09	#####	11003 11100 さいたま市			683	210	
4	1 金額	3 二人以上	2E+09	#####	12003 12100 千葉市			805	147	
5	1 金額	3 二人以上	2E+09	#####	13003 13100 東京都区部			550	105	
6	1 金額	3 二人以上	2E+09	#####	14003 14100 横浜市			535	152	
7	1 金額	3 二人以上	2E+09	#####	15003 15100 新潟市			540	130	
8	1 金額	3 二人以上	2E+09	#####	16003 16201 富山市			544	145	
9	1 金額	3 二人以上	2E+09	#####	17003 17201 金沢市			804	220	
0	1 金額	3 二人以上	2E+09	#####	18003 18201 福井市			557	119	
1	1 金額	3 二人以上	2E+09	#####	19003 19201 甲府市			650	115	
2	1 金額	3 二人以上	2E+09	#####	20003 20201 長野市			738	122	
3	1 金額	3 二人以上	2E+09	#####						

グラフタイトル

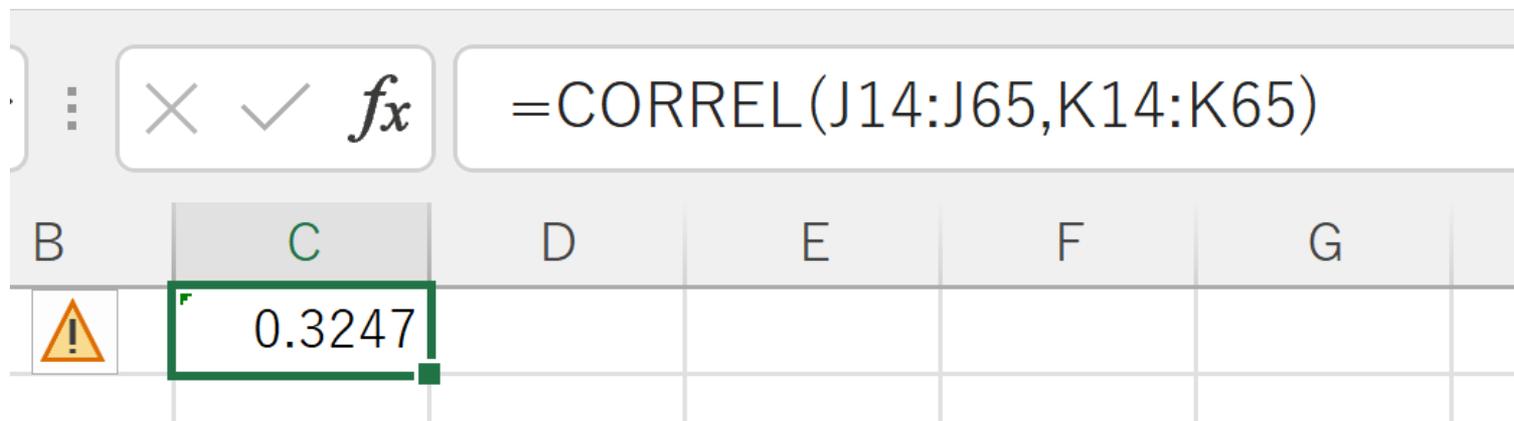


横軸：ケーキ，縦軸：プリン

Excel で相関係数を求める



- ① 相関係数を求めたいセルを選んでおく
- ② メニューの「数式」
- ③ 「その他の関数」, 「統計」, 「CORREL」
- ④ 配列 1 と配列 2 の範囲を指定



相関係数として 0.3247 が求まった
→ 「相関がある」とは言い切れない

まとめ



- **政府統計データ**は、人口、世帯、企業活動、産業構造、物価・地価、文化などのさまざまなデータがある。
- **クロス集計表**は、2つ以上の要因に関するデータを分析するために使用される表で、**データの確認や分析に役に立つ**。
- **相関係数**は、2つの変数間の関係性を示す指標。**0に近い値の場合には関係が低い**。
- **オープンデータ**は、**誰でも自由にアクセスし、利用できる**ように**公開されたデータ**。社会課題解決、効率化、透明化の向上に有用。