

## 第1章 統計学でできること

このパートではまず、統計学を使う一般的な意義を説明するとともに、サンプルデータに触りながら基礎的な分析手法を学んでいきます。

このパートでは統計学の役割をその名の通り「まとめてはかる学問」だと位置づけた上で「まとめる」ことのメリットとして現状を把握するための記述、現状を左右しうる要因の探索、限られた情報からの予測、特定の要因が本当に原因なのかの検証といった点を1つ1つ説明していきます。

統：まとめる

計：はかる

学：がくもん

「まとめて」「はかる」前の状態

「この間のお客さんたち超激アツだった」  
「へー...どれくらい？」

- 体温？
- 満足度？
- 購買金額？



「まとめる」ためには



購買金額  
2万円



購買金額  
3万円



購買金額  
7万円

要するに一言で言うと？

「足して2で割ったような～」という表現を応用

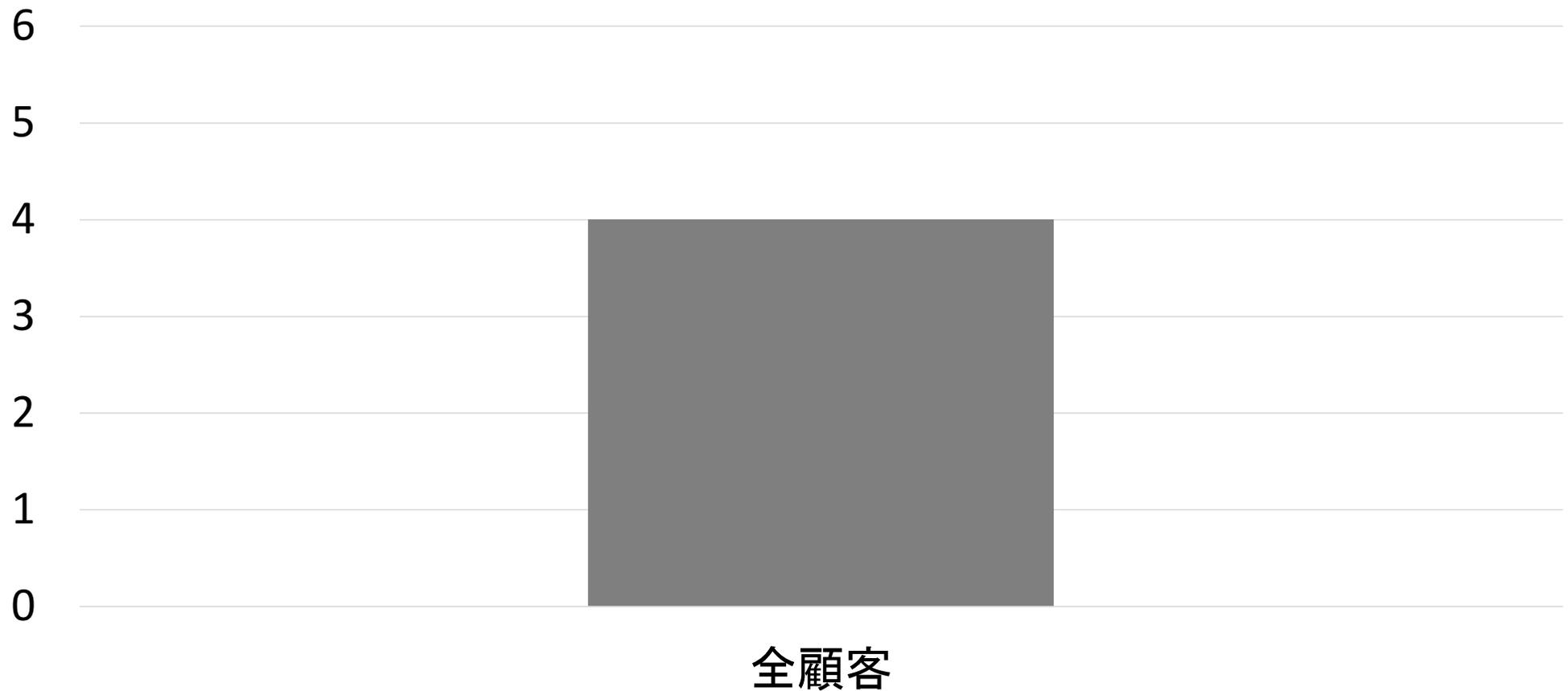
$$(2+3+7) \div 3 = 4$$

全員分足して 人数で割る

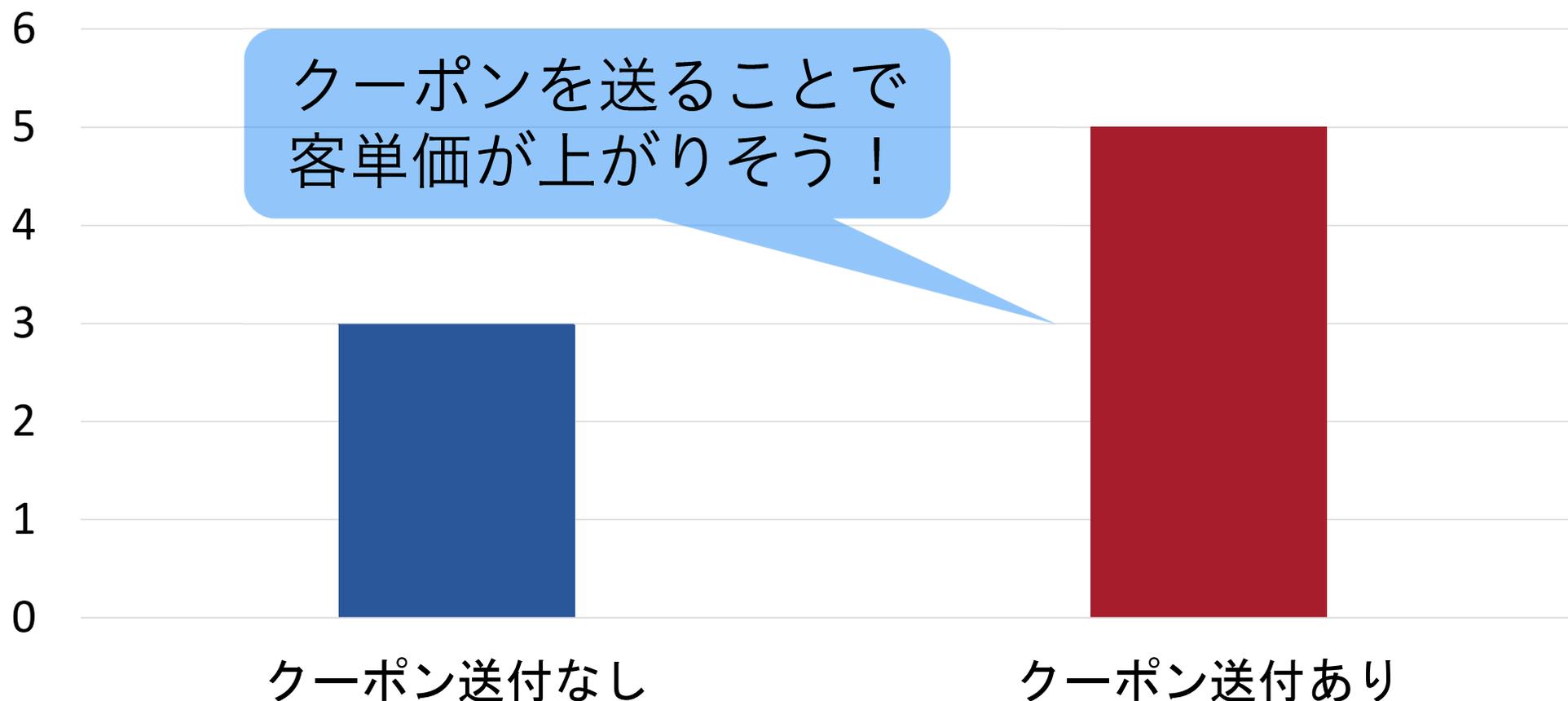
これがよく使う平均値

1. 現状を把握するための**記述**ができる
2. 現状を左右するような要因の**探索**ができる
3. 今後限られた情報だけで**予測**ができる
4. その要因が本当に原因なのか**検証**できる

## 顧客の平均購買金額(万円)



## 顧客の平均購買金額(万円)



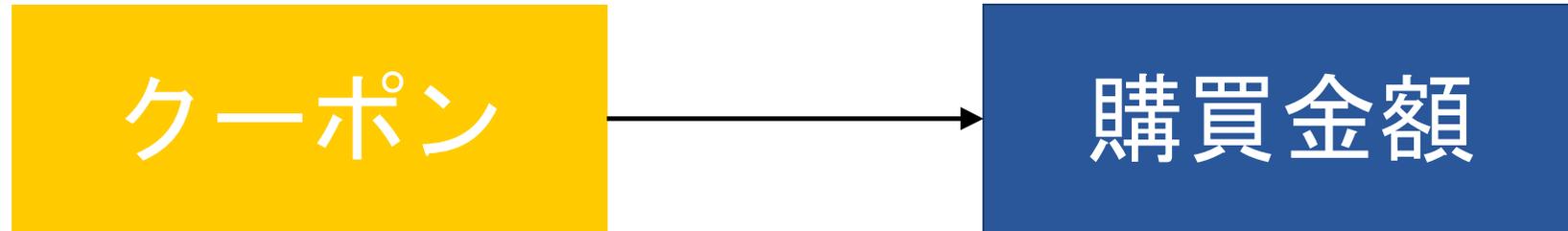
通常お店には毎日100人来店

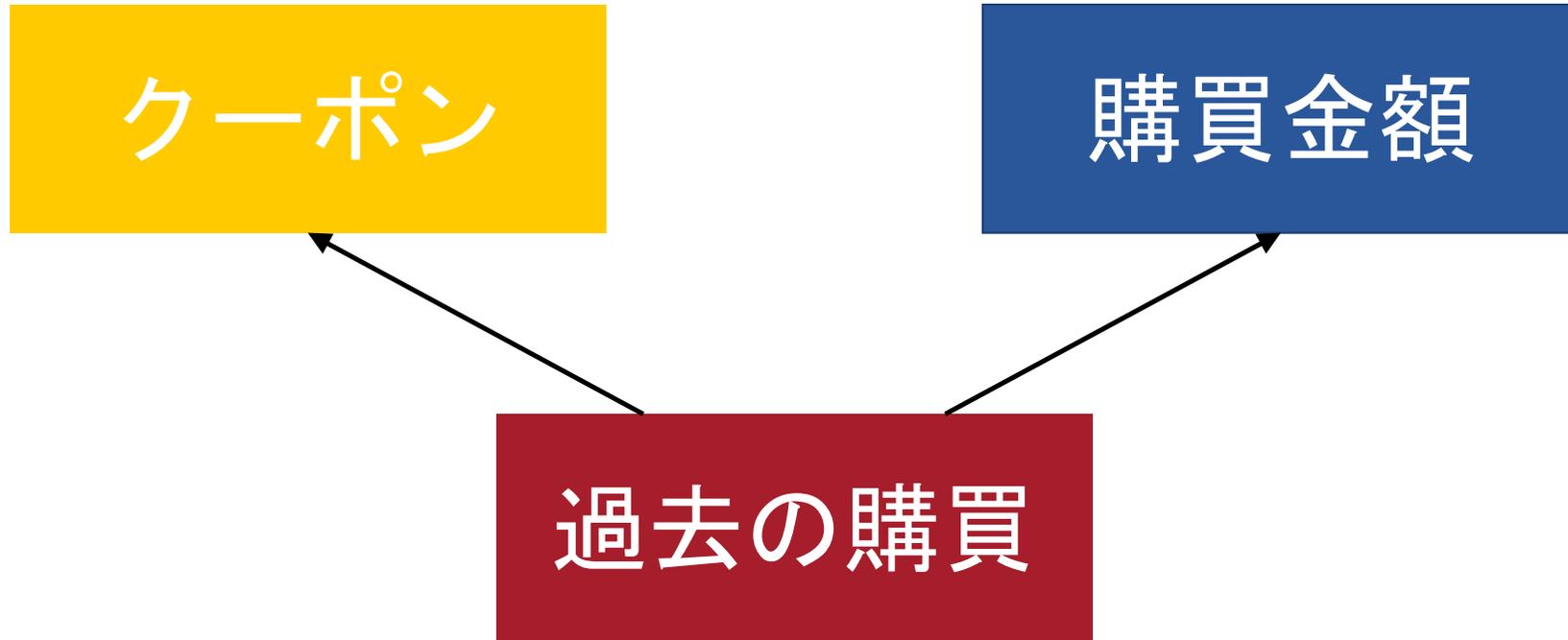
クーポンを配布していない50人は3万円利用

クーポンを配布した50人は5万円利用

結果平均4万円利用

では全員にクーポンを配布した時の売上増は？





先程のグラフで見えてるのはこの部分だけかもしれない

クーポン

購買金額

過去の購買



このパートでは具体的なビジネスの活用事例を紹介した上で、それに即したサンプルデータおよび演習課題の状況設定を提示します。

具体的には、今回取材した活用事例では店舗や事業所の出店に関するものが多いため、消費者視点で業態のわかりやすい飲食事業のケースを題材として、効率的な出店を考えるようにしました。

## <活動の概要>

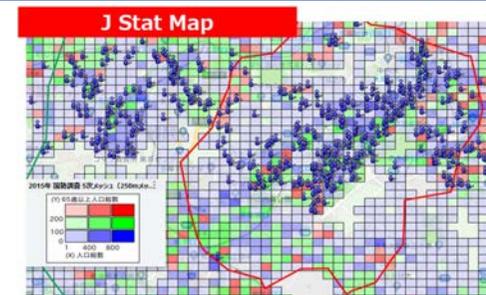
フランチャイズ（FC）研究会は日本唯一のFC専門研究実践団体である。2000年初頭から活動を開始し、現在はさまざまな分野を専門とする中小企業診断士約50名によって構成されている。FC本部や、FC加盟店、FC加盟希望者を対象として、事業支援や経営指導、各種セミナーの提供を通じて、FCビジネスの健全な発展とノウハウ開発を目的としている。クライアントの業態は飲食、物販、サービスなどFCビジネス全般にわたっている。

## <政府統計データ活用事例>

### jSTAT MAPを活用したフランチャイズビジネスにおけるドミナント戦略

#### 活用方法

店舗の営業エリア探索、顧客分布、競合調査、商圈分析、出店候補地探索、売上予測など。家計調査データを用いて新規店舗の売上を予測することや、BtoBビジネスの場合、経済センサスの事業所数データなどを参照し配送ルートの効率性判断を行っている。また、政府統計を活用した商圈分析のノウハウをセミナーで提供し、書籍「jSTAT MAP徹底活用 立地診断ガイドブック」で、jSTAT MAPの基本的な使い方から幅広い応用方法までを豊富な事例とともにわかりやすくまとめている。



#### 背景・課題

FC本部やFC加盟企業は、新規出店の際の立地診断、売上予測、ドミナント戦略など出店に関する分析を行う場合、有償のGISソフトやデータを購入する必要があった。しかし購入にはある程度の費用がかかるため、ソフトやデータの購入にまで至らず、勘と経験に頼った出店を行い、その結果、採算分岐点にのらなかったり、自社店舗同士の商圈重複などが起きるようなことも多かった。

#### 活用データ

国勢調査、経済センサス、家計調査、道路交通センサスなど

#### 効果

政府統計が整備されたことで、それまで有料だったデータやツールと同等機能が無料で利用できるようになり、データに基づいたマーケティングと縁が遠かった中小企業も、自社で商圈分析などを始められる環境が整った。また、コンサルティング活動では、これまでデータ提供が難しかった企業へ政府統計を活用した立地診断の分析結果データなどの提供が可能となった。

### <事業の概要>

『株式会社ゼンリンマーケティングソリューションズ』は、2020年4月に『株式会社ゼンリンジオインテリジェンス』『株式会社ゼンリンビジネス』『大東マーケティングソリューションズ株式会社』の3社が合併し発足。業務内容は、エリアマーケティングビジネス、DM発送代行ビジネス、セールスプロモーションビジネス、ダイレクトマーケティングビジネスである。従業員数103名（2020年4月1日現在）。このうち、エリアマーケティングビジネスでは、主にマーケティング諸活動支援及びコンサルティング業務を行っている。クライアントの業種は限定せず、地図の上でマーケティングの課題を解決したいという企業に対し、GISに関するあらゆるサービスを提供している。

### <政府統計データ活用事例>

## 地図と政府統計データをベースとしたGISサービスの提供

### 活用方法

クライアントに対し、地図データならびに政府統計データをベースとしてコンサルティングを行っている。また、政府統計データの第三者提供権を取得し、町丁目データやメッシュデータへ按分推計したり、時系列データ化するなどの加工をして販売している。例えば、厚生労働省の「介護保険事業状況報告調査」からエリアごとの要介護者数を把握するために要支援・要介護認定数を地域メッシュ単位に按分して推計するなどを行っている。



### 背景・課題

単に統計データを見ただけで理解し、施策の立案などにつなげられる企業はさほど多くはないため、景気が悪くなってくると会社のリソースをどこに注力していくかということデータを考えることが多くなり、ニーズが増えてきたと感じている。以前は多店舗展開が多かったが近年は統廃合の手伝いなども多い。

### 活用データ

国勢調査、経済センサス、商業統計、家計調査、住宅・土地統計調査など

### 効果

多店舗展開においては1,000店舗が壁になると考えられる。それ以上展開する場合は、データに基づいてきちんとした計画を立てて進めることが重要である。顧客の中にはそのようにして2,000店舗を達成したところがある。また、近年は既存店舗を維持するにはどうすればよいかということを考える企業が多く、例えばデータを分析することで配送ルートの効率化を図り、営業エリア内をトラック何台、運転手何人で回ればよいかを考えることができる。

### <事業の概要>

1953年創業。人工衛星や航空機などに搭載した各種センサーを駆使して収集した国内外の空間情報に加工・処理・解析を施し、クライアントの課題解決に向けたあらゆる空間情報サービスを提供する。約1,200の自治体に対し、固定資産税の評価業務や、道路・上下水道の維持管理などの支援を行う。また民間企業に対しては、物流効率化、マーケティング支援、企業防災支援などを提供している。連結売上高541億円、従業員2677名。東証一部上場。（2020年6月末現在）

### <政府統計データ活用事例>

## 地図と政府統計データを組合わせたエリアマーケティングツール

### 活用方法

エリアマーケティングツール「MarketPlanner GIS」に、標準で国勢調査や、商業統計、人口推計データなどを収録。オプションとしては、経済センサスや国勢調査をベースにした50mメッシュの推計人口データ、全国消費実態調査をもとにした月間消費支出額の推計データ、年収別推計世帯数などもある。クライアントは「MarketPlanner GIS」を用いて人口の集計、地図上への表示、レポートの作成、運転時間の計算、交通量把握などをすることが可能になった。



### 背景・課題

同社は、GISサービスに必要な計測技術、GISエンジン、地図製作技術、システム保守サービスを有していた。それに政府統計データを組み合わせることで、更なる差別化を図った。

また、これまで小売業や金融業などの企業は、出店場所の検討やチラシ配布などの業務に際し、GISソフトや政府統計データなしに該当地域の調査分析をしようとすると、非常に煩雑なデータ収集、集計、計算作業を行わなければならない、手作業によるミスも頻発していた。例えば、住民基本台帳のデータを各自治体のウェブからダウンロードして集計しようとすると、自治体ごとに集計の粒度やデータ形式が異なるためそれを整えるだけでも数日から数週間を要した。地図上にデータを展開する時も、手集計したものを、地図に手書きで表示していた。

### 活用データ

国勢調査、経済センサス、商業統計、住宅・土地統計調査、全国消費実態調査など

### 効果

GIS技術に加えて、政府統計データ、さらには自社統計コンテンツを組み合わせることで、お客様の業務に役立つサービス化を図ったことが多くの実績を有することにつながった。

クライアント企業は政府統計データと同社のGISツールを活用することで集計作業が大幅に効率化され、自社で所有するシステム（営業支援CRM、不動産業向け物件管理システムなど）内部の住所データを、地図上に展開し、政府統計データもあわせて分析することが可能となった。小売業の出店検討、チラシ配布エリアの選定、金融機関の支店統廃合、消費財メーカーが店舗近隣居住者の属性を分析し小売業に仕入の提案を行うなど、地図ソフトと政府統計データを掛け合わせたシステムの活用の幅は枚挙にいとまがない。

**<事業の概要>**

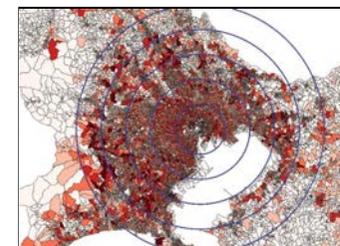
日本不動産研究所は1959年に設立認可された不動産の総合調査研究機関（2011年に一般財団法人化）。約550人の人員と全国的なネットワークを有し、不動産に関する調査研究、鑑定評価、コンサルティングを主な事業とする。調査研究部門は全国の地価・賃料情報公表に代表される統計資料を一般に公開。鑑定部門は豊富なデータをもとにした客観的な鑑定評価を自治体、一般事業者、リート・ファンド運営者などに対して提供する。コンサルティング部門では政府機関への土地・まちづくり政策に関する提言や、不動産事業者、建設会社、設計会社への不動産市場・開発効果・有効活用・将来予測分析業務などを行っている。

**<政府統計データ活用事例>**

**e-Statのオープンデータを活用した調査研究やコンサルティング**

**活用方法**

主にコンサルティング業務において、これまで研究・調査のベースとして政府統計データを活用し、高い頻度でe-Statを利用している。分析業務でよく使う項目は、町丁字単位の商業売上、事業所数といった経済指標や人口世帯数、年収世帯割合などの基礎データである。e-Statからダウンロードしたデータは、エクセルやアクセスなどで分析するほか、ArcGISやQGIS、MapInfoなどのGISソフトに取り込み、さらに自社独自データをも活用した分析を行っている。



**背景・課題**

長年、不動産業界では主に「実績と知見」に基づいて事業を行っていたが、統計情報のオープン化が進み、ビッグデータの活用の必要性がより重要視されるようになったこともあり、近年特にデータを元にした客観的分析と、それに基づく戦略（開発エリアやアセットタイプ、開発時期など）の策定に対するニーズが高まっている。

e-Statができる前は、調査研究の元となる政府統計のデータソースが紙であることが多く、まずは入力作業が必要だったことから人手と時間がかかっていた。当時の作業手順は地図に手で円を書き、地図上に町丁目をピックアップし、人口を記入し...という手作業の連続であった。

**活用データ**

国勢調査、経済センサス、商業統計、家計調査、労働力調査、地価公示・地価調査など

**効果**

e-Statに政府統計データが集約されてからは、各省庁のウェブサイトを巡回する手間がなくなり非常に便利になったと感じている。これらのオープンデータを活用することで、調査やコンサルティング業務などの事業につなげることができ、顧客からの相談に解決策を提案することができる。さらに今後、政府統計データが調査完了時から迅速にオープン化され、CSV形式に加えてGISにて利用可能な形式での提供を期待する。

- あなたは全国に「ちょっと高級な」焼肉店を展開する会社の幹部
- 新たな店舗を出店したい
- デモデータから「どのような店舗を」「どのあたりに出店するのが良さそうか」を考えましょう

まずは「記述」の観点から、データには量的なもの と 質的なものがあることを説明した上で、量的なものは平均値、質的なものは割合の形で集計を行うことを説明します。

その上で、何かのグループで分けて平均値や割合を集計することで「探索」ができることを示し、その差がたまたまの差と言えるようなのかどうか、それぞれ対応する統計的仮説検定のエクセルでの実行方法を示します。

数値型  
(量的)

大小自体が意味を持つもの  
(例：年齢・来店回数・金額・・・)

分類型  
(質的)

異なる属性に分類するもの  
(例：性別・居住地域・職業・・・)



購買金額  
2万円



購買金額  
3万円



購買金額  
7万円

要するに一言で言うと？

「足して2で割ったような～」という表現を応用

$$(2+3+7) \div 3 = 4$$

全員分足して 人数で割る

これがよく使う平均値

平均値だけじゃわからないこと



平均4万円

平均値だけじゃわからないこと



これも平均4万円

およそいくつからいくつまでの値なのか  
平均値より大きいものと小さいもののバランス  
最も多数派なのはおよそいくつくらいなのか



ヒストグラムが便利

ぱっと見把握しにくいデータでも…

27

19

30

7

3

22

17

23

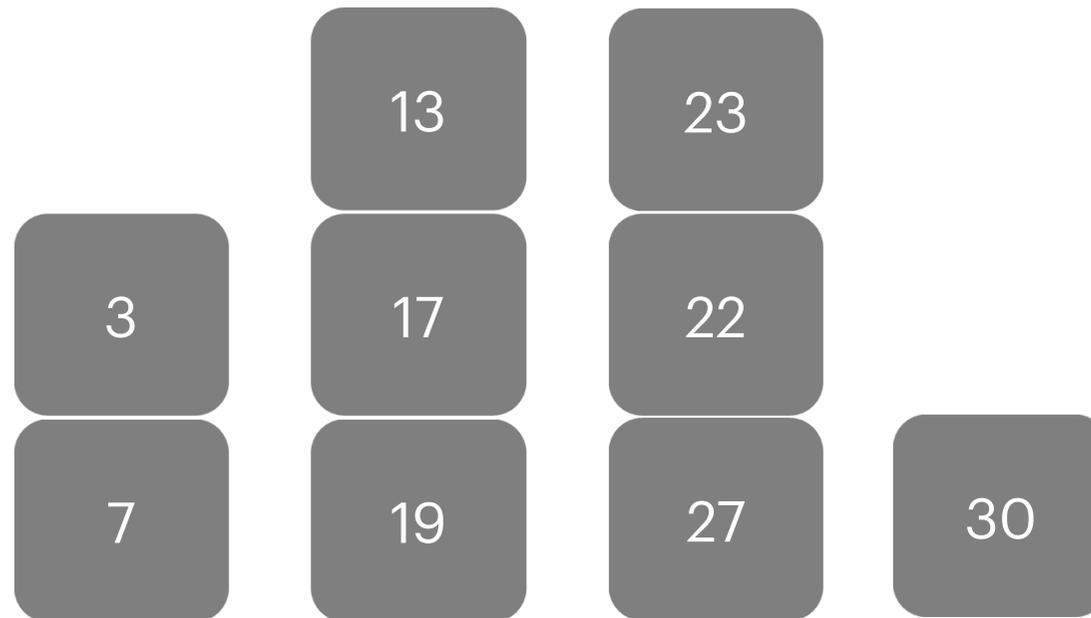
13

ぱっと見把握しにくいデータでも…



十の位だけ見てみると

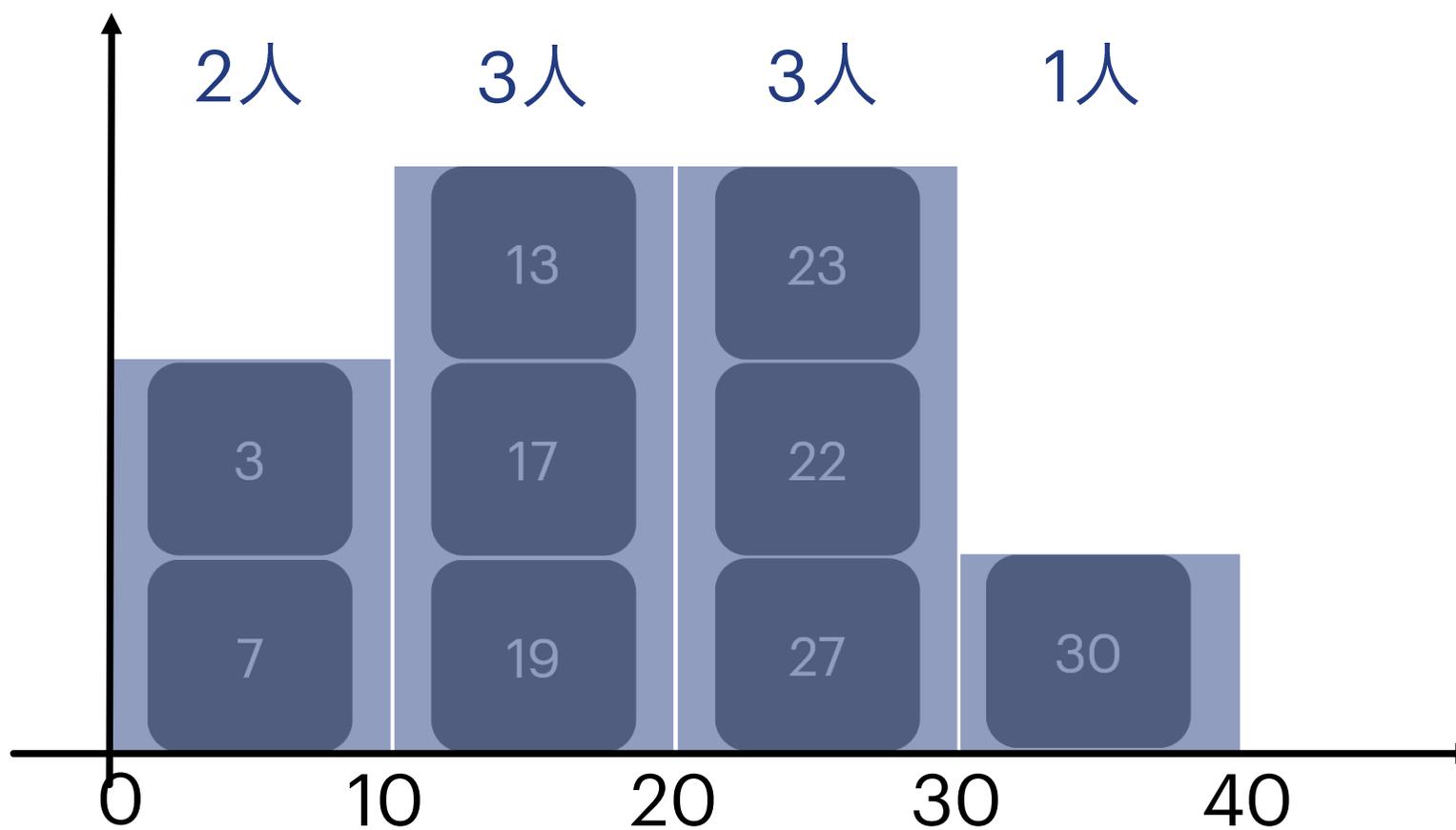
ざっくりカテゴリーに区切ると把握しやすい



ざっくりカテゴリーに区切ると把握しやすい

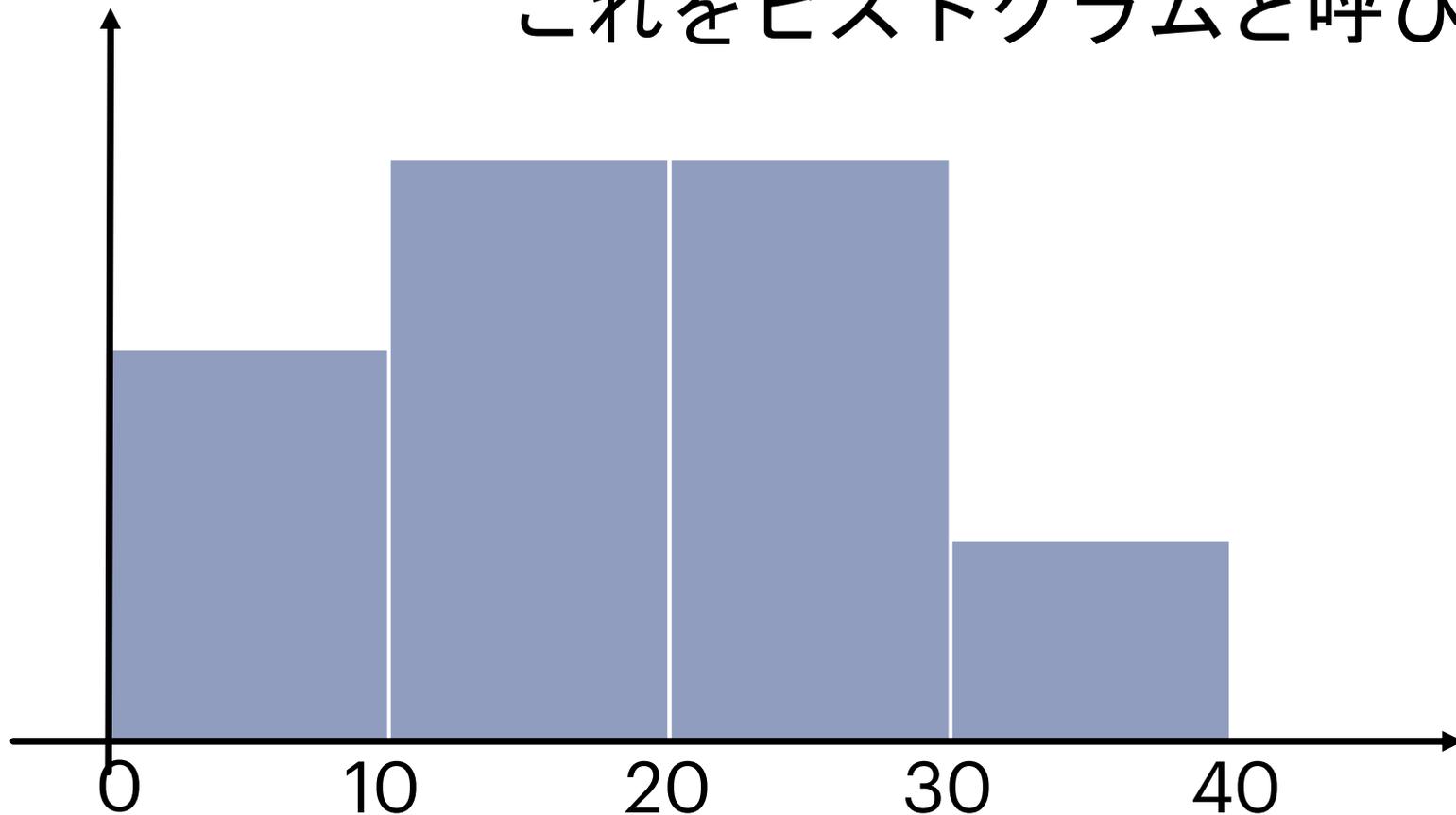


これを図にすると...



これを図にすると・・・

これをヒストグラムと呼びます



数値型  
(量的)



平均値にまとめる  
グラフはヒストグラム

分類型  
(質的)

異なる属性に分類するもの  
(例：性別・居住地域・職業・・・)

例えばこういう状況だと

## 顧客満足度についてのアンケート項目

「0:わからない」が20人

「1:不満」が30人

「2:満足」が50人

平均すると1.3点？

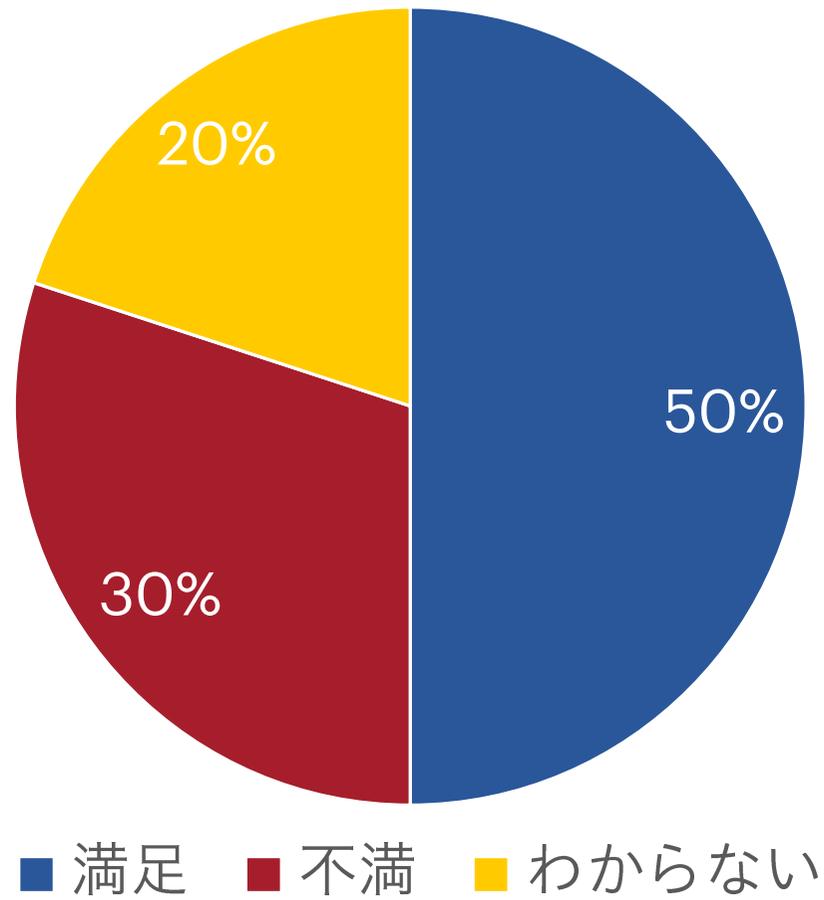
## 「質的に異なる」分類の場合

回答	人数
わからない	20
不満	30
満足	50
合計	100

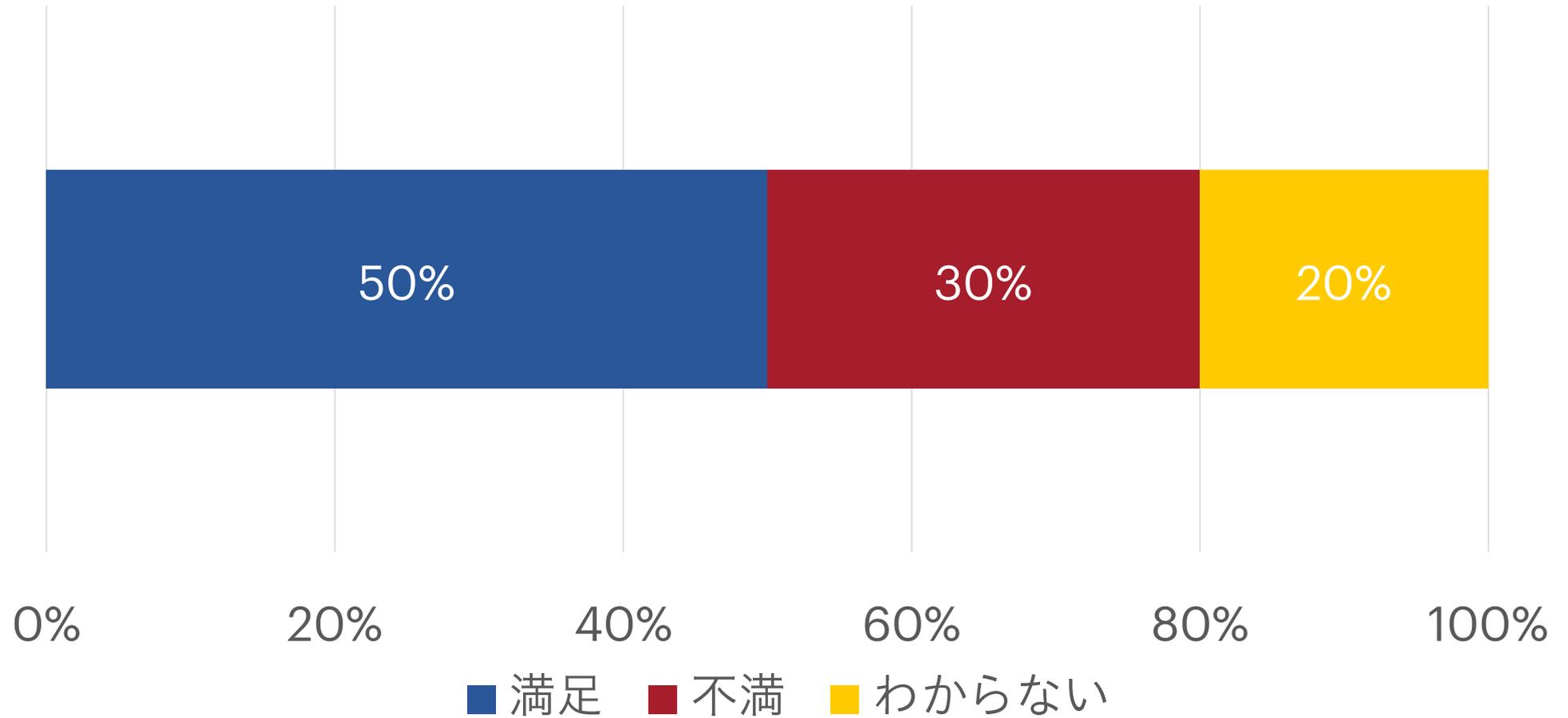
回答	人数	割合
わからない	20	20.0%
不満	30	30.0%
満足	50	50.0%
合計	100	100

回答	人数	割合
わからない	20	20.0%
不満	30	30.0%
満足	50	50.0%
「だいたいこのへん」もすぐにわかる		
合計	100	100

## グラフにするとしたら①



## グラフにするとしたら②



数値型  
(量的)



平均値にまとめる  
グラフはヒストグラム

分類型  
(質的)



分類ごとの割合にまとめる  
グラフは円グラフか帯グラフ

記述

機会やリスクの大きさを把握



探索

いったいどのような「違い」が  
「機会やリスク」を左右する？

これが何かを  
探索したい！

## 説明変数

アウトカムを  
「説明」するかも  
しれない要因

関連

## アウトカム

ビジネスで  
最大化/最小化  
したい項目

それぞれのデータを「探索する」ときは・・・

数値型  
(量的)

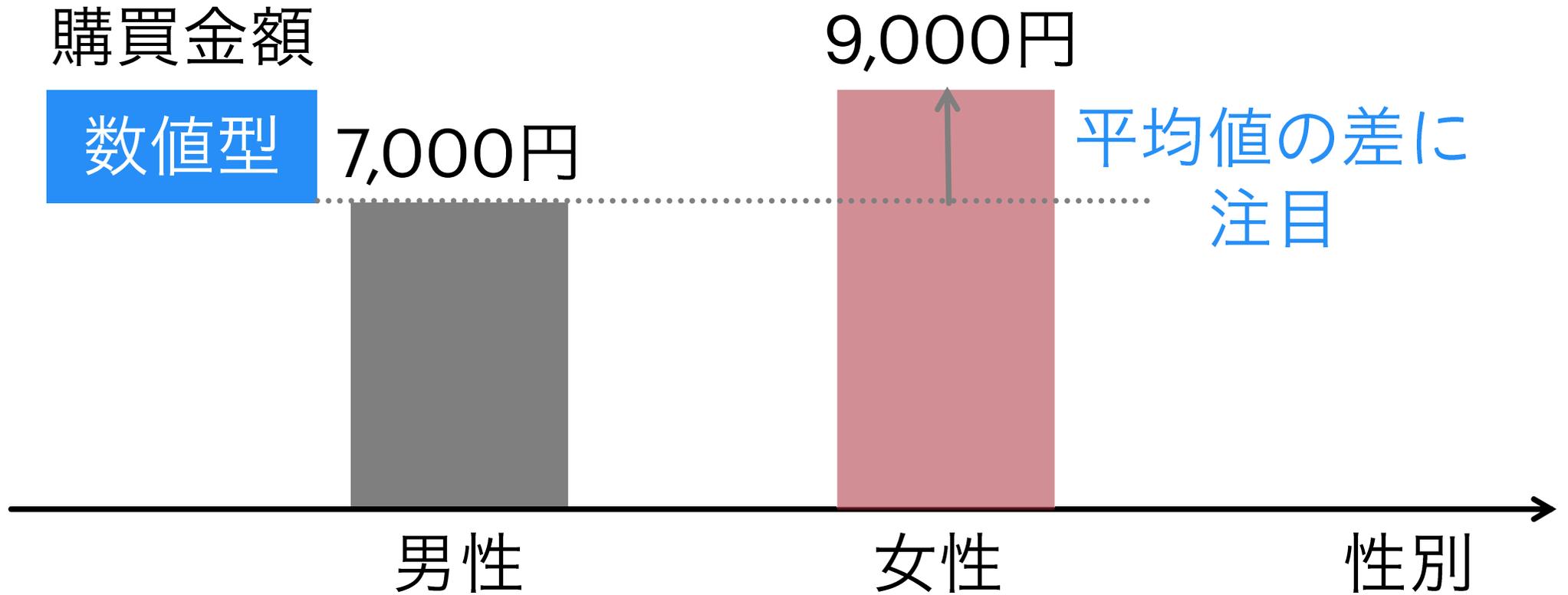


何かの要因で分けて  
平均値の差を比較

分類型  
(質的)



何かの要因で分けて  
割合の差を比較



満足したか

分類型

33.3%

75.0%

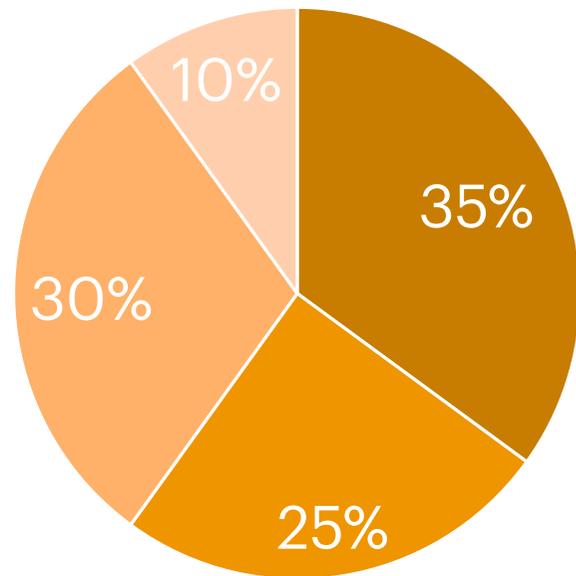
割合の差に  
注目

男性

女性

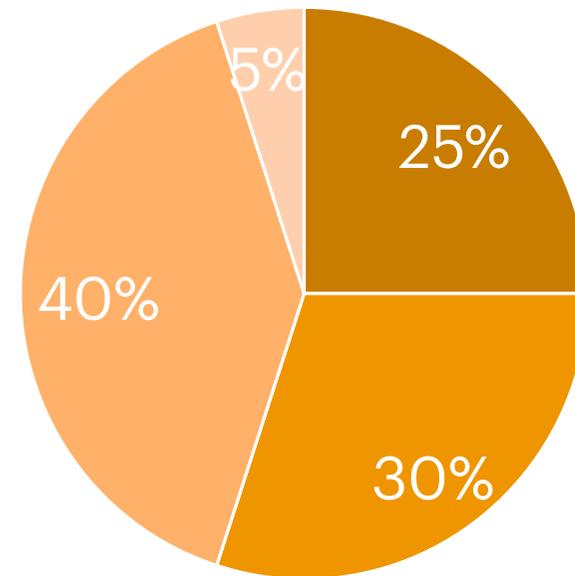
性別

男性



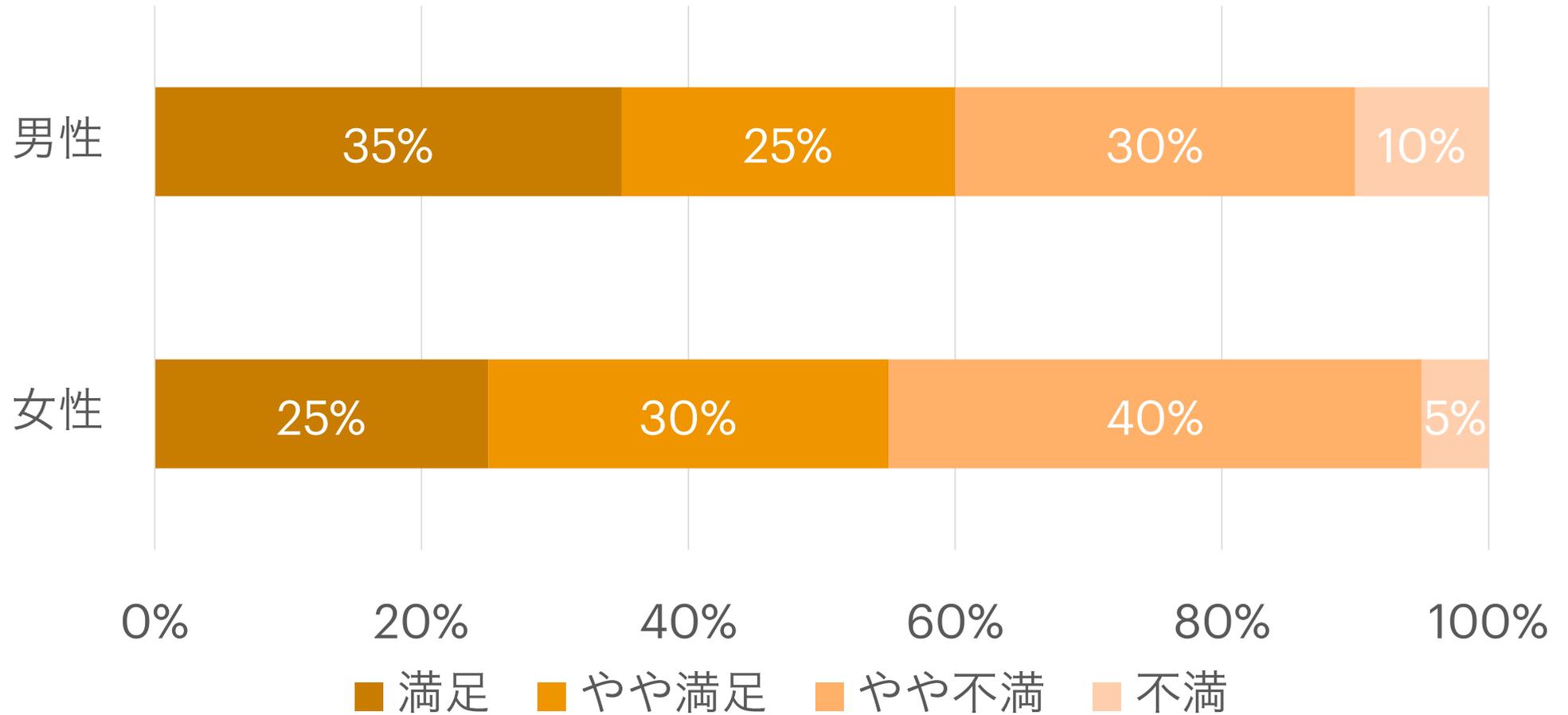
■ 満足      ■ やや満足  
■ やや不満   ■ 不満

女性



■ 満足      ■ やや満足  
■ やや不満   ■ 不満

# 帯グラフや棒グラフの方が比較しやすい



ピボットグラフで平均値を集計するには(この後デモします)

- データ全体を選択してピボットグラフを挿入
- グループ分けしたい項目を「列」に
- 平均値を求めたい項目を「 $\Sigma$ 値」に
- 値フィールドの設定を「平均値」に

# デモ①-1

# デモ①

自動保存 サンプルデータ配布版.xlsx

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 ヘルプ

ピボットテーブル 挿入 アドインを入手 個人用アドイン 推奨グラフ

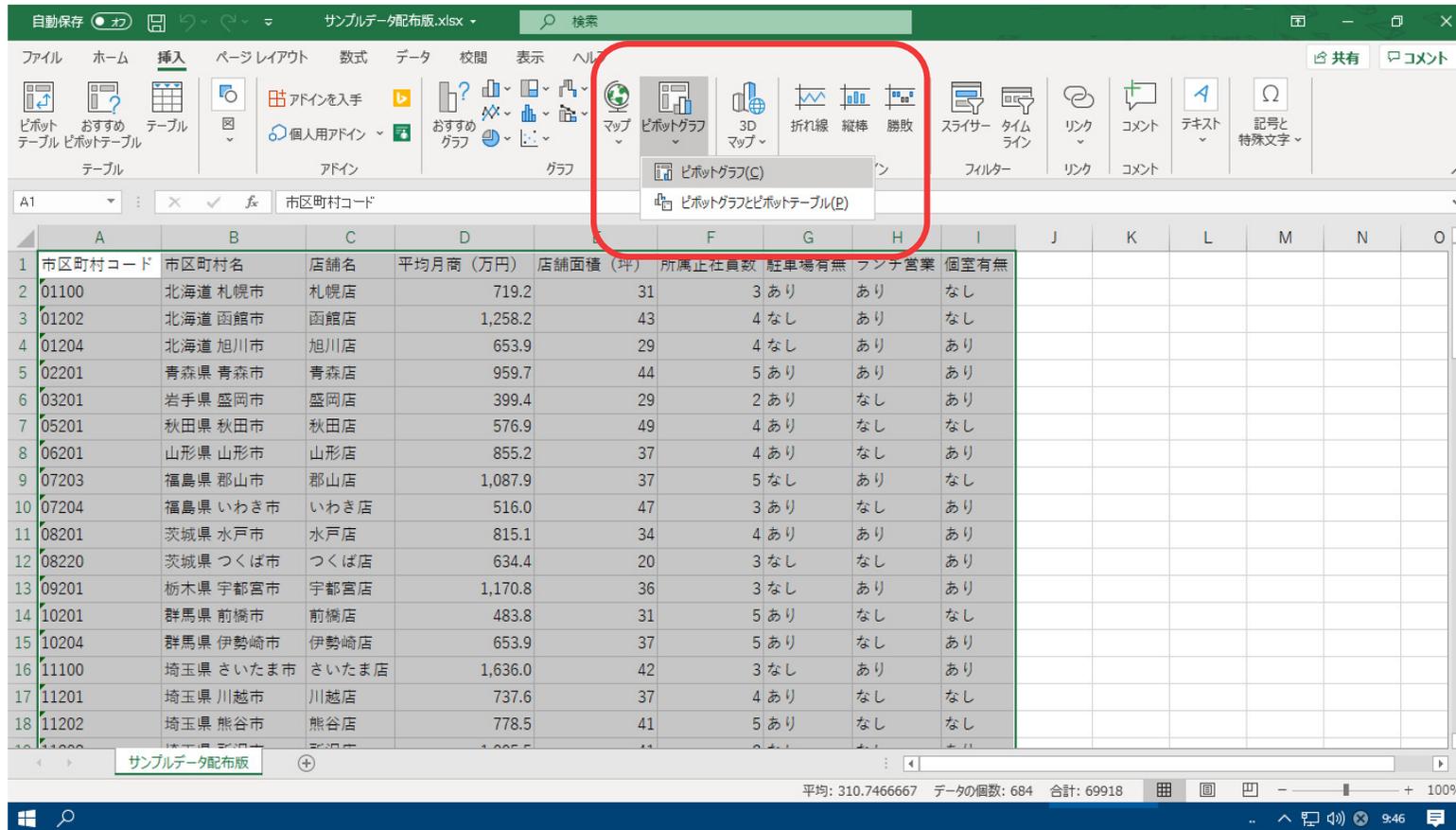
グラフ 3Dマップ 折れ線 縦棒 勝敗 スライサー タイムライン フィルター リンク コメント テキスト 記号と特殊文字

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
	市区町村コード	市区町村名	店舗名	平均月商 (万円)	店舗面積 (坪)	所属正社員数	駐車場有無	ランチ営業	個室有無						
1	01100	北海道 札幌市	札幌店	719.2	31	3	あり	あり	なし						
2	01202	北海道 函館市	函館店	1,258.2	43	4	なし	あり	なし						
3	01204	北海道 旭川市	旭川店	653.9	29	4	なし	あり	あり						
4	02201	青森県 青森市	青森店	959.7	44	5	あり	あり	あり						
5	03201	岩手県 盛岡市	盛岡店	399.4	29	2	あり	なし	あり						
6	05201	秋田県 秋田市	秋田店	576.9	49	4	あり	なし	なし						
7	06201	山形県 山形市	山形店	855.2	37	4	あり	なし	あり						
8	07203	福島県 郡山市	郡山店	1,087.9	37	5	なし	あり	なし						
9	07204	福島県 いわき市	いわき店	516.0	47	3	あり	なし	あり						
10	08201	茨城県 水戸市	水戸店	815.1	34	4	あり	あり	あり						
11	08220	茨城県 つくば市	つくば店	634.4	20	3	なし	なし	あり						
12	09201	栃木県 宇都宮市	宇都宮店	1,170.8	36	3	なし	あり	あり						
13	10201	群馬県 前橋市	前橋店	483.8	31	5	あり	なし	なし						
14	10204	群馬県 伊勢崎市	伊勢崎店	653.9	37	5	あり	なし	あり						
15	11100	埼玉県 さいたま市	さいたま店	1,636.0	42	3	なし	あり	あり						
16	11201	埼玉県 川越市	川越店	737.6	37	4	あり	なし	なし						
17	11202	埼玉県 熊谷市	熊谷店	778.5	41	5	あり	なし	なし						
18	11203	埼玉県 新沼市	新沼店	1,087.9	37	5	あり	なし	あり						

準備完了

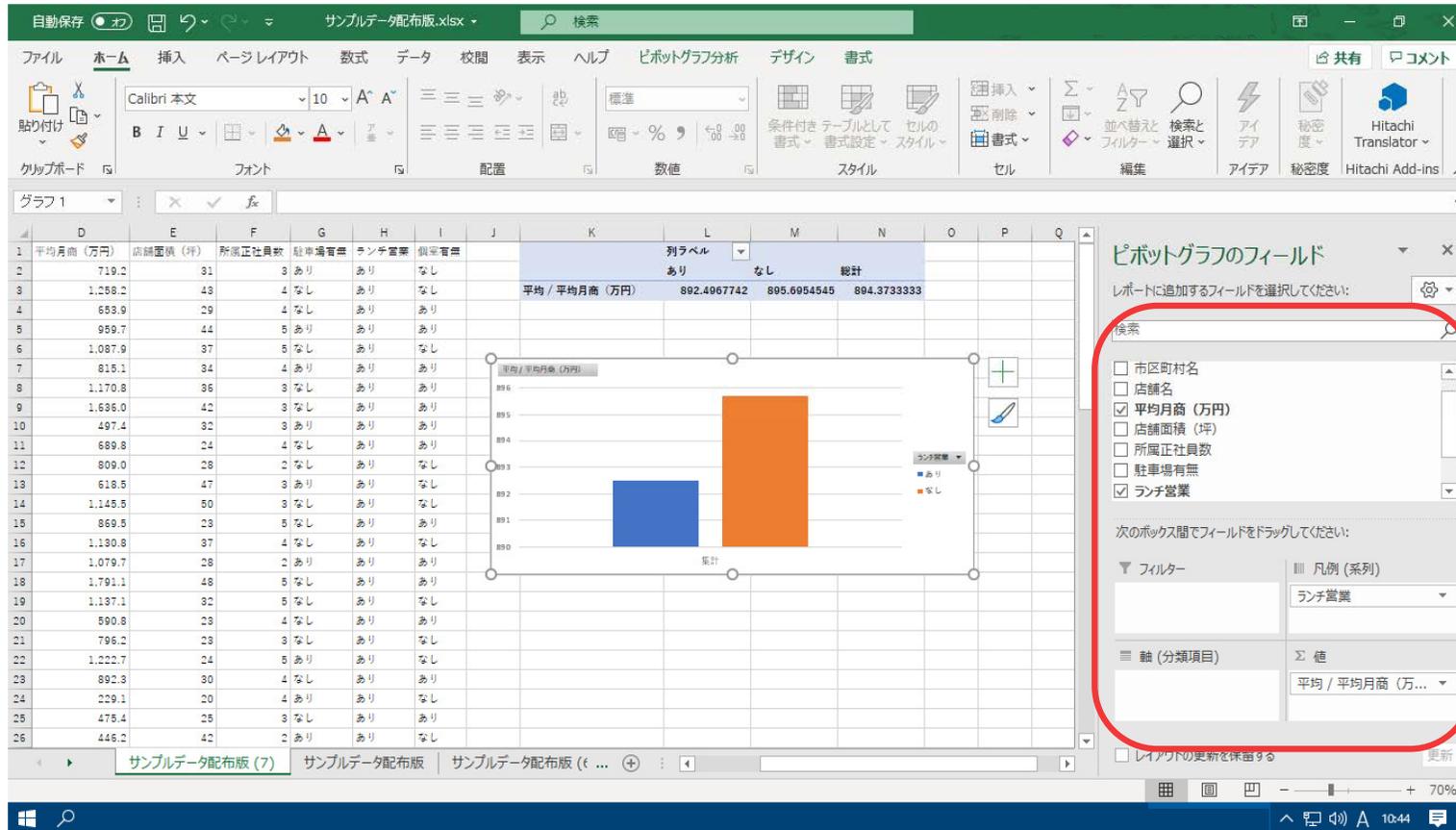
9:46

# デモ①



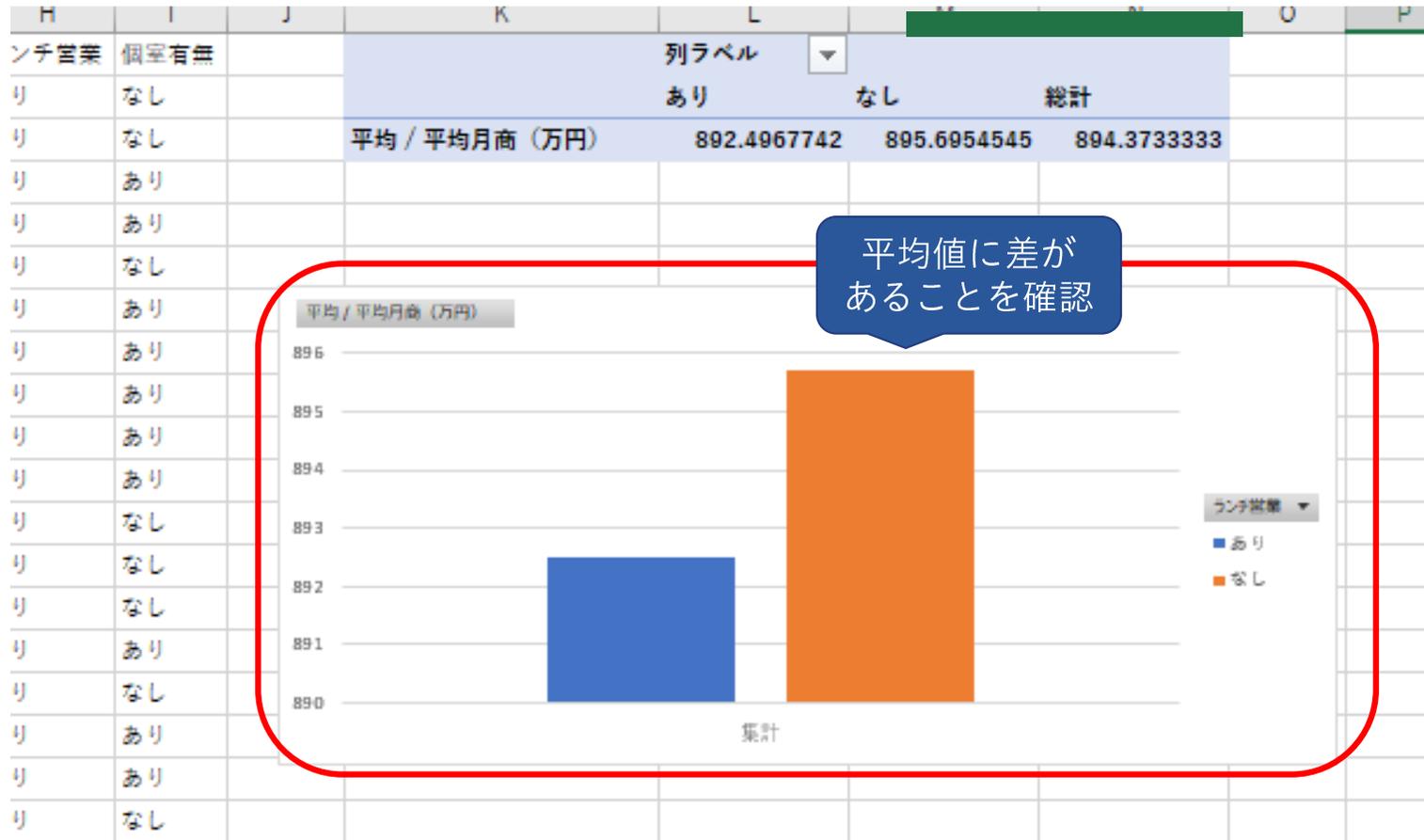
	A	B	C	D	F	G	H	I	J	K	L	M	N	O
1	市区町村コード	市区町村名	店舗名	平均月商 (万円)	店舗面積 (坪)	所属正社員数	駐車場有無	ランチ営業	個室有無					
2	01100	北海道 札幌市	札幌店	719.2	31	3	あり	あり	なし					
3	01202	北海道 函館市	函館店	1,258.2	43	4	なし	あり	なし					
4	01204	北海道 旭川市	旭川店	653.9	29	4	なし	あり	あり					
5	02201	青森県 青森市	青森店	959.7	44	5	あり	あり	あり					
6	03201	岩手県 盛岡市	盛岡店	399.4	29	2	あり	なし	あり					
7	05201	秋田県 秋田市	秋田店	576.9	49	4	あり	なし	なし					
8	06201	山形県 山形市	山形店	855.2	37	4	あり	なし	あり					
9	07203	福島県 郡山市	郡山店	1,087.9	37	5	なし	あり	なし					
10	07204	福島県 いわき市	いわき店	516.0	47	3	あり	なし	あり					
11	08201	茨城県 水戸市	水戸店	815.1	34	4	あり	あり	あり					
12	08220	茨城県 つくば市	つくば店	634.4	20	3	なし	なし	あり					
13	09201	栃木県 宇都宮市	宇都宮店	1,170.8	36	3	なし	あり	あり					
14	10201	群馬県 前橋市	前橋店	483.8	31	5	あり	なし	なし					
15	10204	群馬県 伊勢崎市	伊勢崎店	653.9	37	5	あり	なし	あり					
16	11100	埼玉県 さいたま市	さいたま店	1,636.0	42	3	なし	あり	あり					
17	11201	埼玉県 川越市	川越店	737.6	37	4	あり	なし	なし					
18	11202	埼玉県 熊谷市	熊谷店	778.5	41	5	あり	なし	なし					

②データ全体を選択して、挿入>ピボットグラフを選択します。



③割合を求めたい項目を「Σ値」に、「凡例」に比較したい項目を入れて、値フィールドの設定を「平均値」に設定します。

# デモ①



④完成。これでピボットグラフで平均値を集計することができました。

- データ全体を選択してピボットグラフを挿入
- グループ分けしたい項目を「軸（分類項目）」に
- 割合を求めたい項目を「 $\Sigma$ 値」に
- 「凡例」比較したい項目を
- 値フィールドの設定で計算の種類を「行集計に対する比率」に

# デモ①-2

# デモ①-2

自動保存 [保存] サンプルデータ配布版.xlsx 検索

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 ヘルプ

共有 コメント

ピボットテーブル おすすめのピボットテーブル テーブル

図

アドインを入手 個人用アドイン

おすすめグラフ

グラフ

マップ

ピボットグラフ

3Dマップ

ツアー

折れ線 縦棒 勝敗

スパークライン

スライサー タイムライン

フィルター

リンク

コメント

テキスト

記号と特殊文字

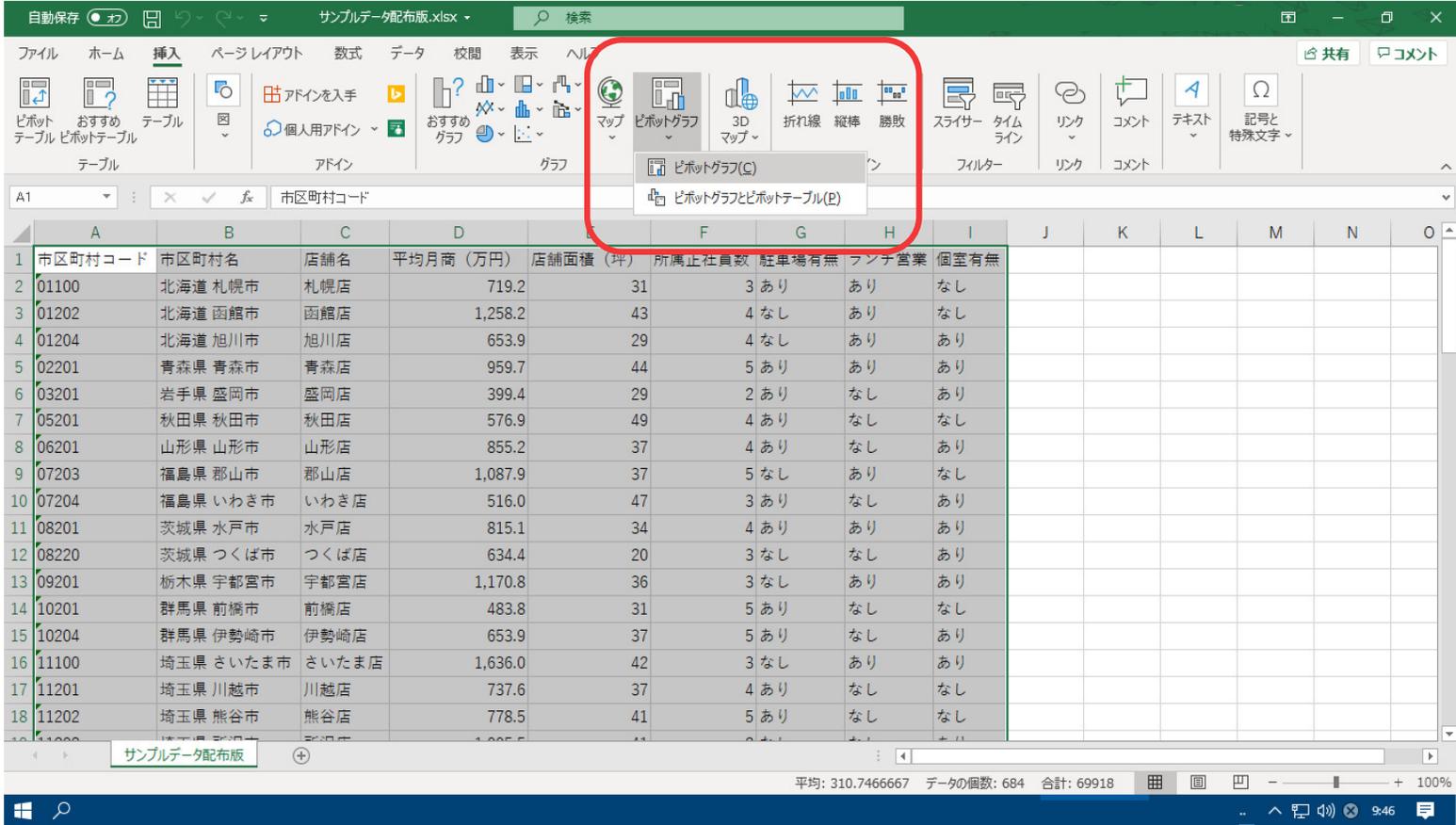
K9

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	市区町村コード	市区町村名	店舗名	平均月商 (万円)	店舗面積 (坪)	所属正社員数	駐車場有無	ランチ営業	個室有無						
2	01100	北海道 札幌市	札幌店	719.2	31	3	あり	あり	なし						
3	01202	北海道 函館市	函館店	1,258.2	43	4	なし	あり	なし						
4	01204	北海道 旭川市	旭川店	653.9	29	4	なし	あり	あり						
5	02201	青森県 青森市	青森店	959.7	44	5	あり	あり	あり						
6	03201	岩手県 盛岡市	盛岡店	399.4	29	2	あり	なし	あり						
7	05201	秋田県 秋田市	秋田店	576.9	49	4	あり	なし	なし						
8	06201	山形県 山形市	山形店	855.2	37	4	あり	なし	あり						
9	07203	福島県 郡山市	郡山店	1,087.9	37	5	なし	あり	なし						
10	07204	福島県 いわき市	いわき店	516.0	47	3	あり	なし	あり						
11	08201	茨城県 水戸市	水戸店	815.1	34	4	あり	あり	あり						
12	08220	茨城県 つくば市	つくば店	634.4	20	3	なし	なし	あり						
13	09201	栃木県 宇都宮市	宇都宮店	1,170.8	36	3	なし	あり	あり						
14	10201	群馬県 前橋市	前橋店	483.8	31	5	あり	なし	なし						
15	10204	群馬県 伊勢崎市	伊勢崎店	653.9	37	5	あり	なし	あり						
16	11100	埼玉県 さいたま市	さいたま店	1,636.0	42	3	なし	あり	あり						
17	11201	埼玉県 川越市	川越店	737.6	37	4	あり	なし	なし						
18	11202	埼玉県 熊谷市	熊谷店	778.5	41	5	あり	なし	なし						
19	11203	埼玉県 新沼市	新沼店	1,087.9	37	5	あり	なし	あり						

準備完了

9:46

## デモ①-2



自動保存 [保存] サンプルデータ配布版.xlsx 検索

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 ヘルプ

ピボット テーブル 挿入 アドインを入手 アドイン おすすめ グラフ マップ

ピボットグラフ(ピ) 3D マップ 折れ線 縦棒 勝敗

スライサー タイムライン リンク コメント テキスト 記号と特殊文字

A1 : X ✓ 市市区町村コード

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
	市市区町村コード	市市区町村名	店舗名	平均月商 (万円)	店舗面積 (坪)	所属正社員数	駐車場有無	ランチ営業	個室有無						
1	01100	北海道 札幌市	札幌店	719.2	31	3	あり	なし	なし						
2	01202	北海道 函館市	函館店	1,258.2	43	4	なし	あり	なし						
3	01204	北海道 旭川市	旭川店	653.9	29	4	なし	あり	あり						
4	02201	青森県 青森市	青森店	959.7	44	5	あり	あり	あり						
5	03201	岩手県 盛岡市	盛岡店	399.4	29	2	あり	なし	あり						
6	05201	秋田県 秋田市	秋田店	576.9	49	4	あり	なし	なし						
7	06201	山形県 山形市	山形店	855.2	37	4	あり	なし	あり						
8	07203	福島県 郡山市	郡山店	1,087.9	37	5	なし	あり	なし						
9	07204	福島県 いわき市	いわき店	516.0	47	3	あり	なし	あり						
10	08201	茨城県 水戸市	水戸店	815.1	34	4	あり	あり	あり						
11	08220	茨城県 つくば市	つくば店	634.4	20	3	なし	なし	あり						
12	09201	栃木県 宇都宮市	宇都宮店	1,170.8	36	3	なし	あり	あり						
13	10201	群馬県 前橋市	前橋店	483.8	31	5	あり	なし	なし						
14	10204	群馬県 伊勢崎市	伊勢崎店	653.9	37	5	あり	なし	あり						
15	11100	埼玉県 さいたま市	さいたま店	1,636.0	42	3	なし	あり	あり						
16	11201	埼玉県 川越市	川越店	737.6	37	4	あり	なし	なし						
17	11202	埼玉県 熊谷市	熊谷店	778.5	41	5	あり	なし	なし						

サンプルデータ配布版

平均: 310.7466667 データの個数: 684 合計: 69918 100%

②データ全体を選択して、挿入>ピボットグラフを選択します。

自動保存 ● オフ サンプルデータ配布版.xlsx 検索

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 ヘルプ ピボットグラフ分析 デザイン 書式

Calibri 本文 10 A A 標準

条件付き書式 テーブルとして セルの書式設定 スタイル

挿入 削除 書式

並べ替えとフィルター 検索と選択 アイデア 秘密度 Hitachi Translator

共有 コメント

グラフ 1

	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	平均月商 (万円)	店舗面積 (坪)	所属正社員数	駐車場有無	ランチ営業	個室有無		個数/駐車場有無	列ラベル								
2	719.2	31	3	あり	あり	なし		あり	なし	総計	38.71%	61.29%	100.00%				
3	1,288.2	48	4	なし	あり	なし		なし			59.09%	40.91%	100.00%				
4	653.9	29	4	なし	あり	あり											
5	959.7	44	5	あり	あり	あり											
6	1,087.9	37	5	なし	あり	なし											
7	815.1	34	4	あり	あり	あり											
8	1,170.8	36	3	なし	あり	あり											
9	1,636.0	42	3	なし	あり	あり											
10	497.4	32	3	あり	あり	あり											
11	689.8	24	4	なし	あり	あり											
12	809.0	28	2	なし	あり	なし											
13	618.5	47	3	あり	あり	なし											
14	1,145.5	50	3	なし	あり	なし											
15	869.5	23	5	なし	あり	あり											
16	1,130.8	37	4	なし	あり	なし											
17	1,079.7	28	2	あり	あり	あり											
18	1,791.1	48	5	なし	あり	あり											
19	1,137.1	32	5	なし	あり	なし											
20	590.8	23	4	なし	あり	あり											
21	796.2	23	3	なし	あり	なし											
22	1,222.7	24	5	あり	あり	なし											
23	892.3	30	4	なし	あり	あり											
24	229.1	20	4	あり	あり	なし											
25	475.4	25	3	なし	あり	あり											
26	446.2	42	2	あり	あり	なし											

ピボットグラフのフィールド

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

- 店舗面積 (坪)
- 所属正社員数
- 駐車場有無
- ランチャ営業
- 個室有無

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

凡例 (系列)

駐車場有無

軸 (分類項目)

Σ 値

ランチ営業

個数 / 駐車場有無

レイアウトの更新を保留する

更新

③割合を求めたい項目を「Σ値」に、「凡例」に比較したい項目を入れて、値フィールドの設定で計算の種類を「行集計に対する比率」に設定します。

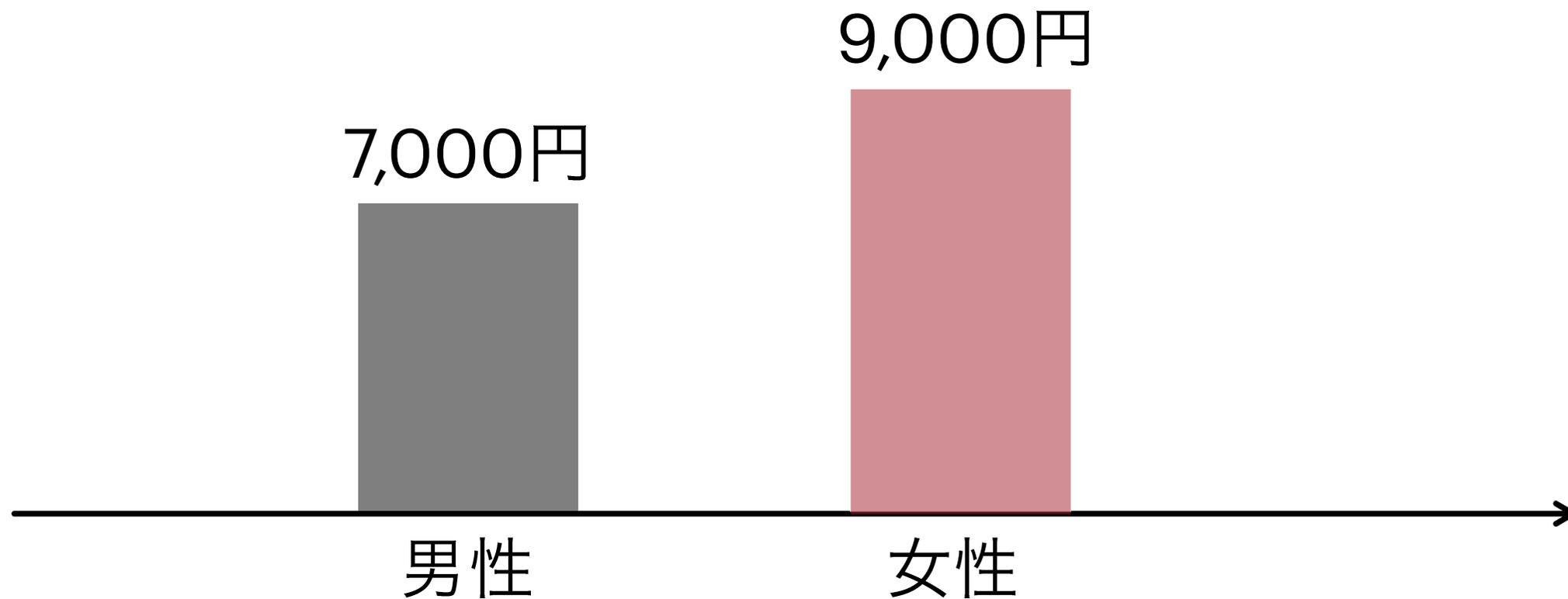


④完成。これでピボットグラフで割合を集計することができました。

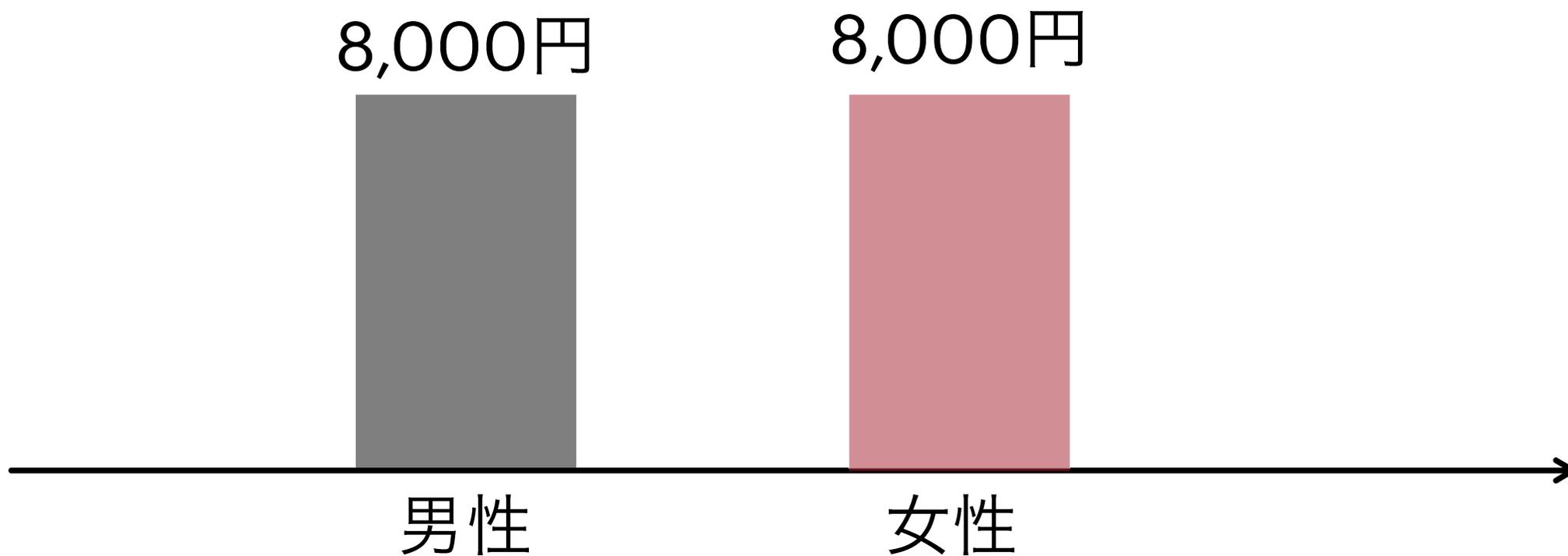
でもこんな反論も…

それぐらいの差、**たまたま**じゃないか？

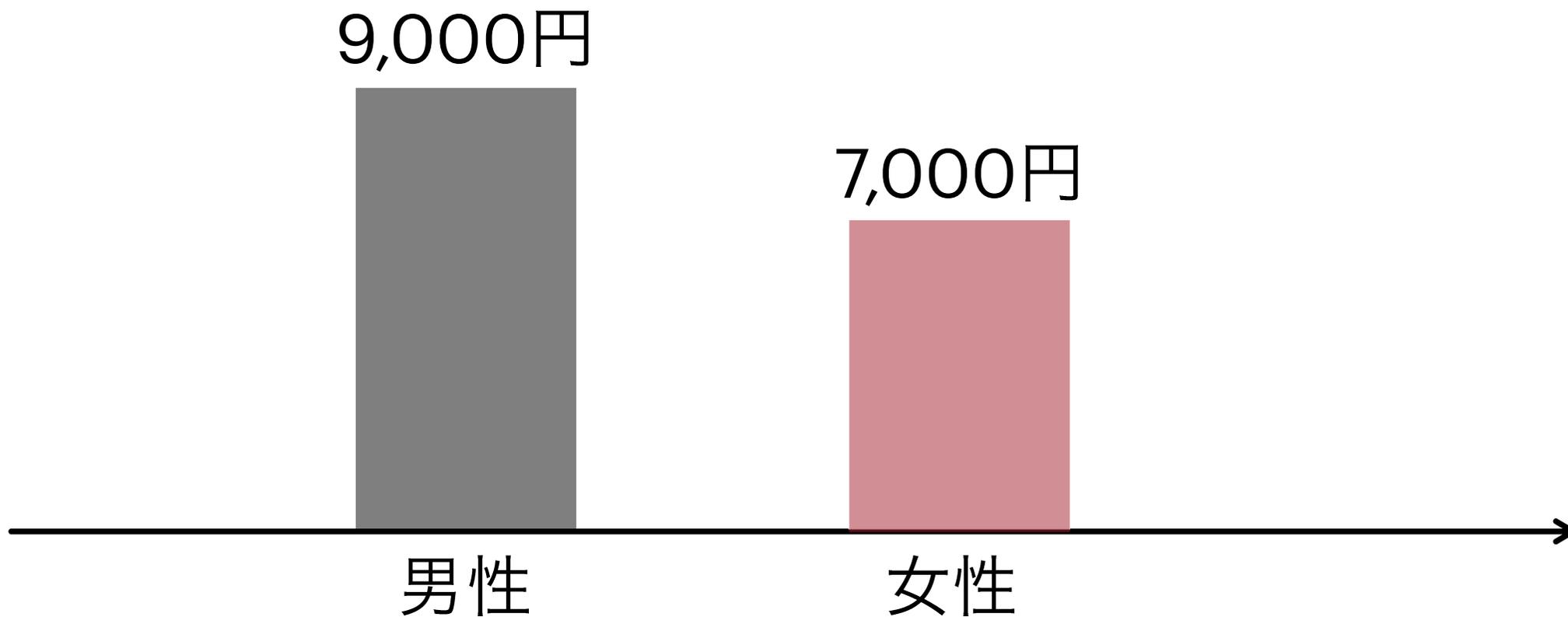
この結果がもしたまたまだとしたら次にデータを取ると…



平均値の差が消えてしまうことも…



逆転することも…



想定する状況：

男女各10人ずつ調査対象者

それぞれの人がみんな「無作為に0～10回来店」

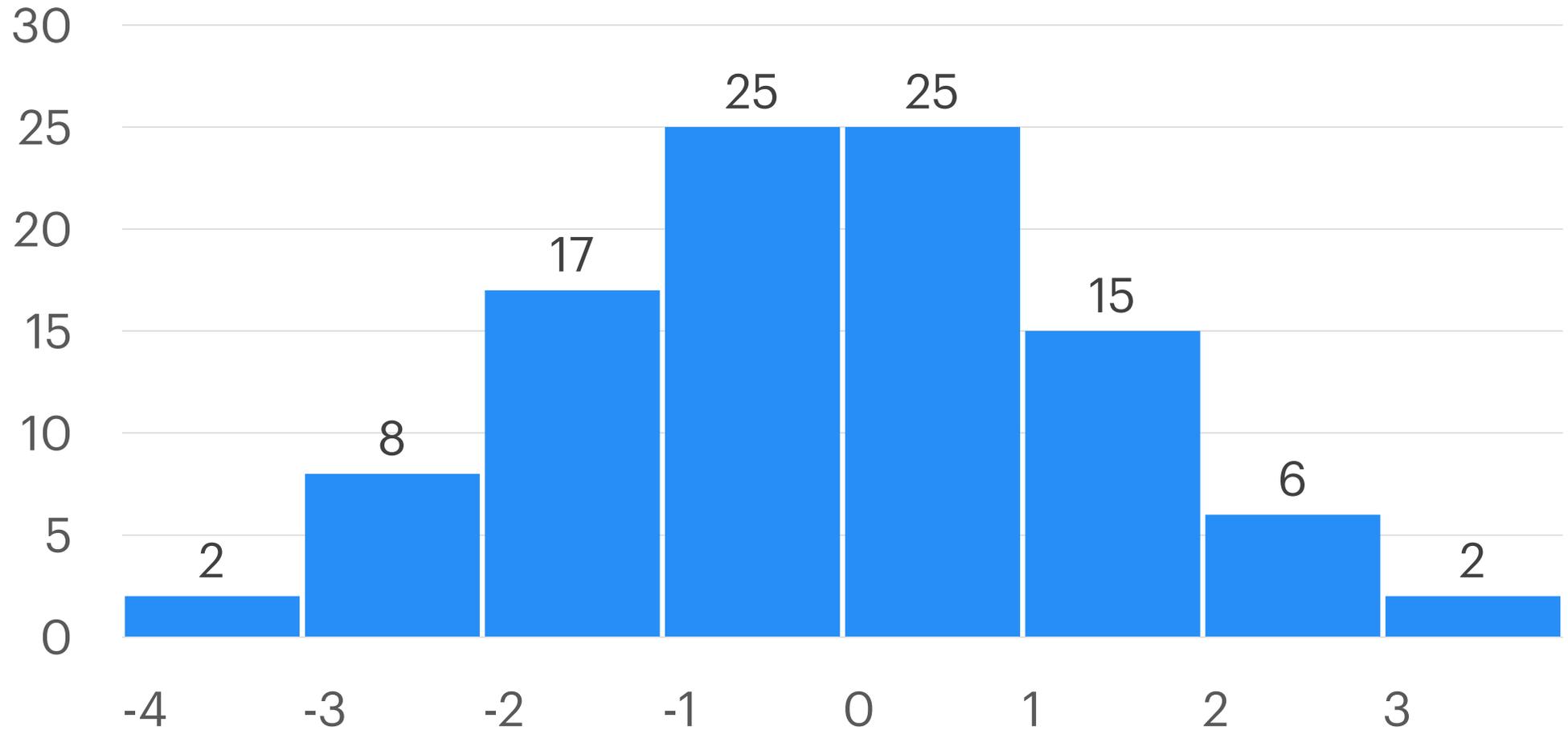
男女間の平均来店回数の差は？

この計算を100回やってみよう

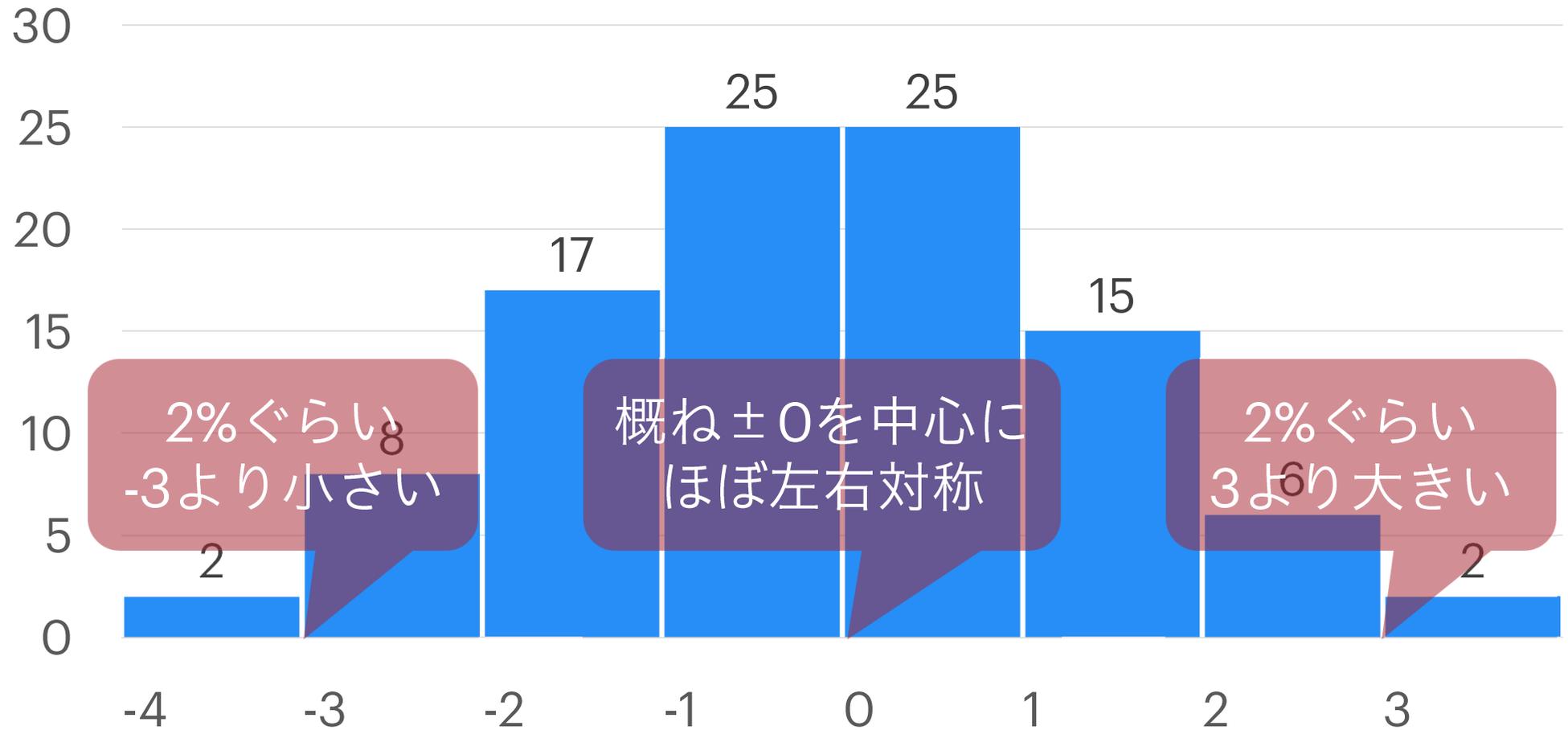
**=RANDBETWEEN(数字1, 数字2)**

**意味：数字1から数字2までの範囲で  
ランダムな整数を出力します**

## 実際にやってみた結果



# 実際にやってみた結果



おまけ：割合について「**たまたま**の差」を経験したい人は

想定する状況：

男女各10人ずつ調査対象者

それぞれがみんな「五分五分で満足と回答」

男女間の平均満足度の差は？

この計算を100回やってみよう

「たまたまの差じゃないの？」と言われてたら

仮に今後無制限にデータをとっていったら  
差がなくなる状況でたまたま  
これほど以上の差がつく確率(p値)を計算

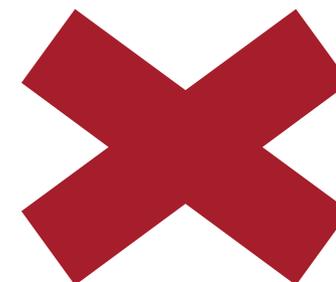
差異の大きさ

比較に使った  
データの数

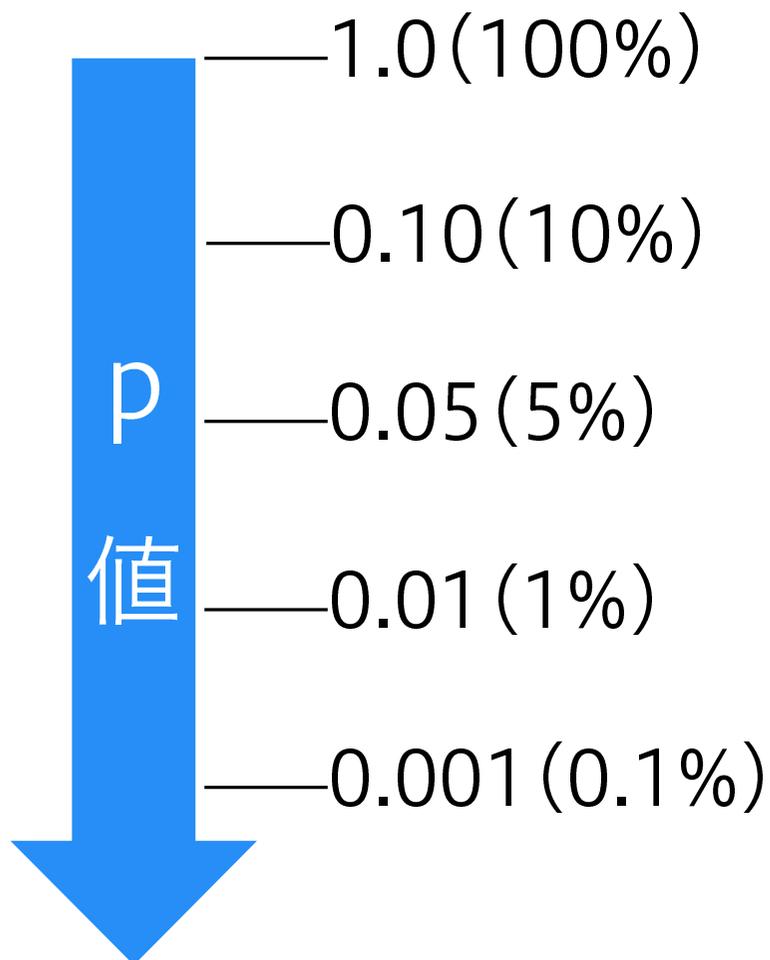
元の値の  
バラつき

判断

総合的に  
判断



p 値



たまたまであってもおかしくない

たまたまじゃないかも…

たまたまとは考えにくい

かなりたまたまと考えにくい

断言したい(けどダメ)

「平均値に全く差がなかったとして」と仮定

「どちらが大きいにせよ2.5%未満(計5%未満)でしか起こらない」  
変な結果になっていないかどうかを確認

変だという結果なら「仮定がおかしい」つまり  
「**たまたま**とは考えにくい」と判断

エクセルで「平均値の差」の仮説検定をするには(このあとデモします)

「平均値の差」の仮説検定であるt検定を行う関数：

=ttest(

1グループ目のデータ範囲,

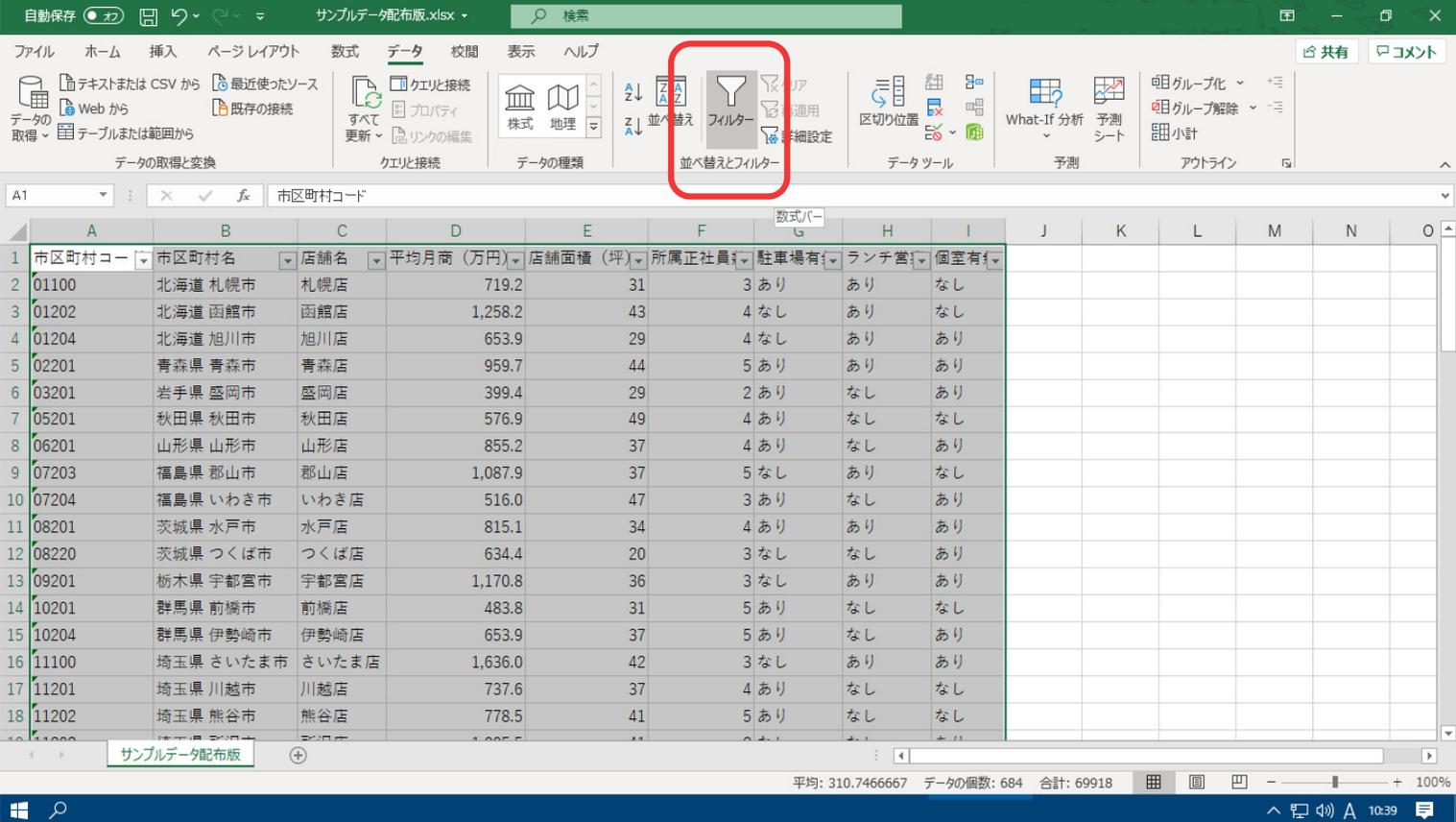
2グループ目のデータ範囲,

2, 3

)

# デモ②

## デモ②



The screenshot shows the Microsoft Excel interface with the 'Data' ribbon selected. The 'Filter' button, represented by a funnel icon, is highlighted with a red rectangular box. Below the ribbon, a spreadsheet is visible with columns labeled A through O. The data in the spreadsheet includes municipality codes (e.g., 01100, 01202), names of municipalities (e.g., 北海道 札幌市, 北海道 函館市), store names (e.g., 札幌店, 函館店), and various numerical and categorical data points. The status bar at the bottom indicates '平均: 310.7466667 データの個数: 684 合計: 69918'.

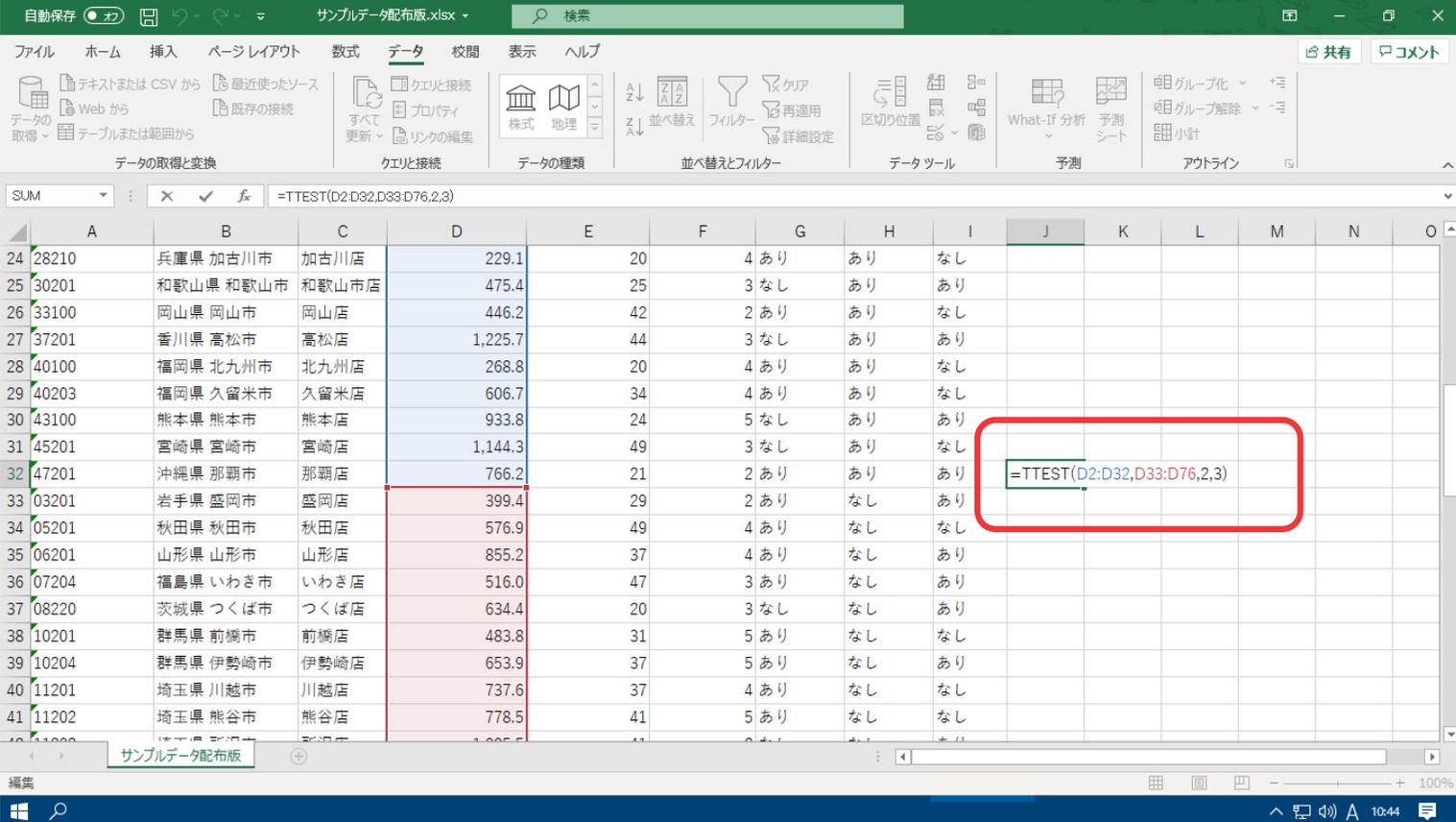
①データ全体を選択して、データ>フィルタを押下します。

## デモ②

市区町村コード	市区町村名	店舗名	平均月商 (万円)	店舗面積 (坪)	所属正社員数	駐車場有	ランチ営業	個室有
01100	北海道 札幌市	札幌店	719.2	31	なし		なし	
01202	北海道 函館市	函館店	1,258.2	43	なし		なし	
01204	北海道 旭川市	旭川店	653.9	29	あり		あり	
02201	青森県 青森市	青森店	959.7	44	あり		あり	
03201	岩手県 盛岡市	盛岡店	399.4	29	なし		なし	
05201	秋田県 秋田市	秋田店	576.9	49	あり		あり	
06201	山形県 山形市	山形店	855.2	37	なし		なし	
07203	福島県 郡山市	郡山店	1,087.9	37	なし		なし	
07204	福島県 いわき市	いわき店	516.0	47	あり		あり	
08201	茨城県 水戸市	水戸店	815.1	34	あり		あり	
08220	茨城県 つくば市	つくば店	634.4	20	あり		あり	
09201	栃木県 宇都宮市	宇都宮店	1,170.8	35	あり		あり	
10201	群馬県 前橋市	前橋店	483.8	31	なし		なし	
10204	群馬県 伊勢崎市	伊勢崎店	653.9	37	あり		あり	
11100	埼玉県 さいたま市	さいたま店	1,636.0	42	あり		あり	
11201	埼玉県 川越市	川越店	737.6	37	なし		なし	
11202	埼玉県 熊谷市	熊谷店	778.5	4	なし		なし	

② 「ランチ営業」列のフィルタメニュー▼から「昇順」を選択すると、あり→なしの順に並びが変わります。

## デモ②

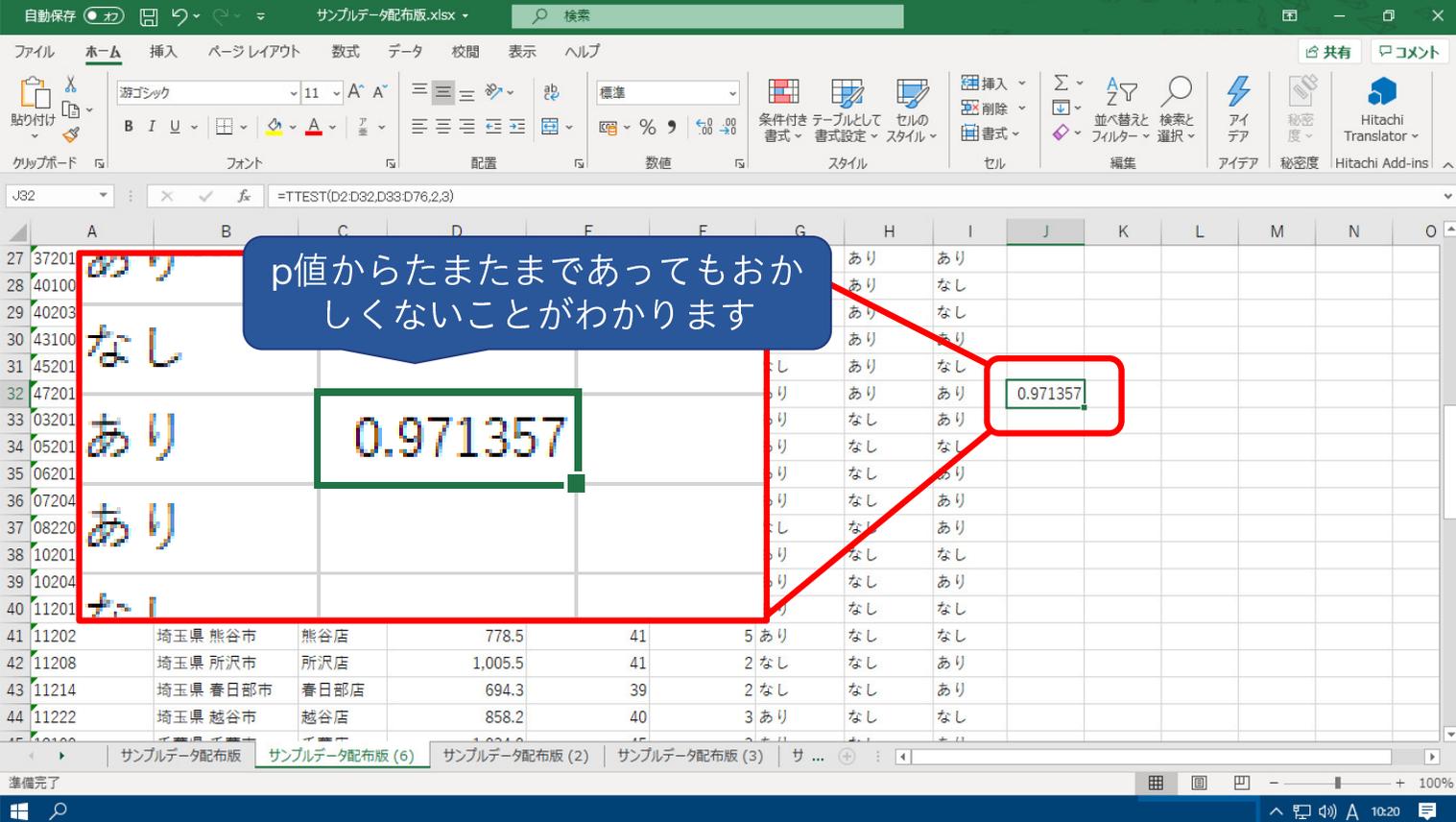


The screenshot shows the Microsoft Excel interface with the following data in columns A through I:

	A	B	C	D	E	F	G	H	I
24	28210	兵庫県 加吉川市	加吉川店	229.1	20	4	あり	あり	なし
25	30201	和歌山県 和歌山市	和歌山市店	475.4	25	3	なし	あり	あり
26	33100	岡山県 岡山市	岡山店	446.2	42	2	あり	あり	なし
27	37201	香川県 高松市	高松店	1,225.7	44	3	なし	あり	あり
28	40100	福岡県 北九州市	北九州店	268.8	20	4	あり	あり	なし
29	40203	福岡県 久留米市	久留米店	606.7	34	4	あり	あり	なし
30	43100	熊本県 熊本市	熊本店	933.8	24	5	なし	あり	あり
31	45201	宮崎県 宮崎市	宮崎店	1,144.3	49	3	なし	あり	なし
32	47201	沖縄県 那覇市	那覇店	766.2	21	2	あり	あり	あり
33	03201	岩手県 盛岡市	盛岡店	399.4	29	2	あり	なし	あり
34	05201	秋田県 秋田市	秋田店	576.9	49	4	あり	なし	なし
35	06201	山形県 山形市	山形店	855.2	37	4	あり	なし	あり
36	07204	福島県 いわき市	いわき店	516.0	47	3	あり	なし	あり
37	08220	茨城県 つくば市	つくば店	634.4	20	3	なし	なし	あり
38	10201	群馬県 前橋市	前橋店	483.8	31	5	あり	なし	なし
39	10204	群馬県 伊勢崎市	伊勢崎店	653.9	37	5	あり	なし	あり
40	11201	埼玉県 川越市	川越店	737.6	37	4	あり	なし	なし
41	11202	埼玉県 熊谷市	熊谷店	778.5	41	5	あり	なし	なし

The formula bar shows the formula: `=TTEST(D2:D32,D33:D76,2,3)`. A red box highlights the cell in row 32, column J, which contains the same formula: `=TTEST(D2:D32,D33:D76,2,3)`.

③ 任意のセルに、`=ttest(1グループ目のデータ範囲, 2グループ目のデータ範囲, 2, 3)`と入力します。



自動保存 検索 サンプルデータ配布版.xlsx

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 ヘルプ

共有 コメント

貼り付け 遊ゴシック 11 A A 標準

条件付き書式 テーブルとしてセルの書式設定 スタイル

挿入 削除 書式

Σ 並べ替えとフィルター 検索と選択 アイデア 秘密度 Hitachi Translator

J32 =TTEST(D2:D32,D33:D76,2,3)

27	37201	あり																	
28	40100	あり																	
29	40203	あり																	
30	43100	なし																	
31	45201	なし																	
32	47201	あり																	
33	03201	あり																	
34	05201	あり																	
35	06201	あり																	
36	07204	あり																	
37	08220	あり																	
38	10201	あり																	
39	10204	あり																	
40	11201	あり																	
41	11202	埼玉県 熊谷市	熊谷店	778.5	41	5	あり	なし	なし										
42	11208	埼玉県 所沢市	所沢店	1,005.5	41	2	なし	なし	あり										
43	11214	埼玉県 春日部市	春日部店	694.3	39	2	なし	なし	あり										
44	11222	埼玉県 越谷市	越谷店	858.2	40	3	あり	なし	なし										

準備完了

サンプルデータ配布版 サンプルデータ配布版 (6) サンプルデータ配布版 (2) サンプルデータ配布版 (3) サ...

10:20

④結果が表示されます。これで「平均値の差」の仮説検定をすることができました。

A1	B1	C1
A2	B2	C2
A3	B3	C3
A4	B4	C4
A5	B5	C5

一個のセルだけの範囲なら  
直接セル名を指定

A4

A1	B1	C1
A2	B2	C2
A3	B3	C3
A4	B4	C4
A5	B5	C5

「開始セル：終了セル」で範囲指定

A2:A4

A1	B1	C1
A2	B2	C2
A3	B3	C3
A4	B4	C4
A5	B5	C5

同じ行や列でなくても  
長方形に指定ができる

A2:C4

A1	B1	C1
A2	B2	C2
A3	B3	C3
A4	B4	C4
A5	B5	C5

開始位置と終了位置の順番は不問  
(お作法としては「左上：右下」)

A1	B1	C1
A2	B2	C2
A3	B3	C3
A4	B4	C4
A5	B5	C5

C5:B2

コンマで区切れば  
複雑な範囲指定もできる(非推奨)

A1	B1	C1
A2	B2	C2
A3	B3	C3
A4	B4	C4
A5	B5	C5

B1,A3:B4,C2:C5

片側

「男性側が多いか」など  
一方向の違いのみに興味

両側

「両者に差があるか」  
両方向の違いに興味

対

「同じ人の二時点の値」など  
データ間に対応がある場合

等分散

グループごとのバラつき方が  
同じであると仮定

非等分散

グループごとのバラつき方が  
同じであると仮定しない

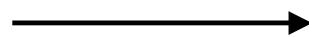
それぞれのデータを「探索する」ときは・・・

数値型  
(量的)



何かの要因で分けて  
平均値の差をt検定で比較

分類型  
(質的)



何かの要因で分けて  
割合の差を比較

満足したか

分類型

33.3%

75.0%

割合の差に  
注目

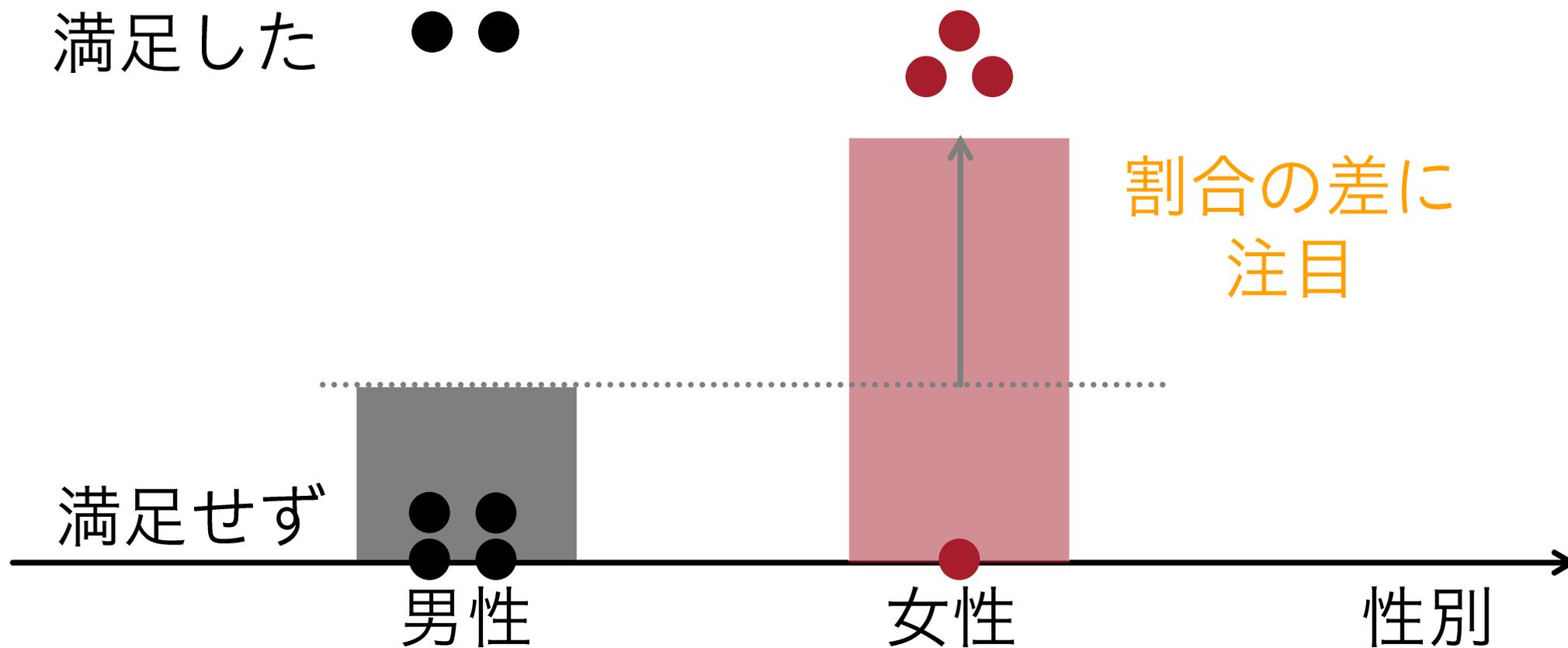
男性

女性

性別

割合に対するこんな反論は…

それぐらいの差、**たまたま**じゃないか？



表で考えてみると

もし差がつかないのなら...

	男性	女性	合計
満足	2	3	5
不満足	4	1	5
合計	6	4	10

表で考えてみると

こうなっているはず

	男性	女性	合計
満足	23	32	5
不満足	43	12	5
合計	6	4	10

「割合の差」の仮説検定である $\chi^2$ 乗検定を行う関数：

=chitest(

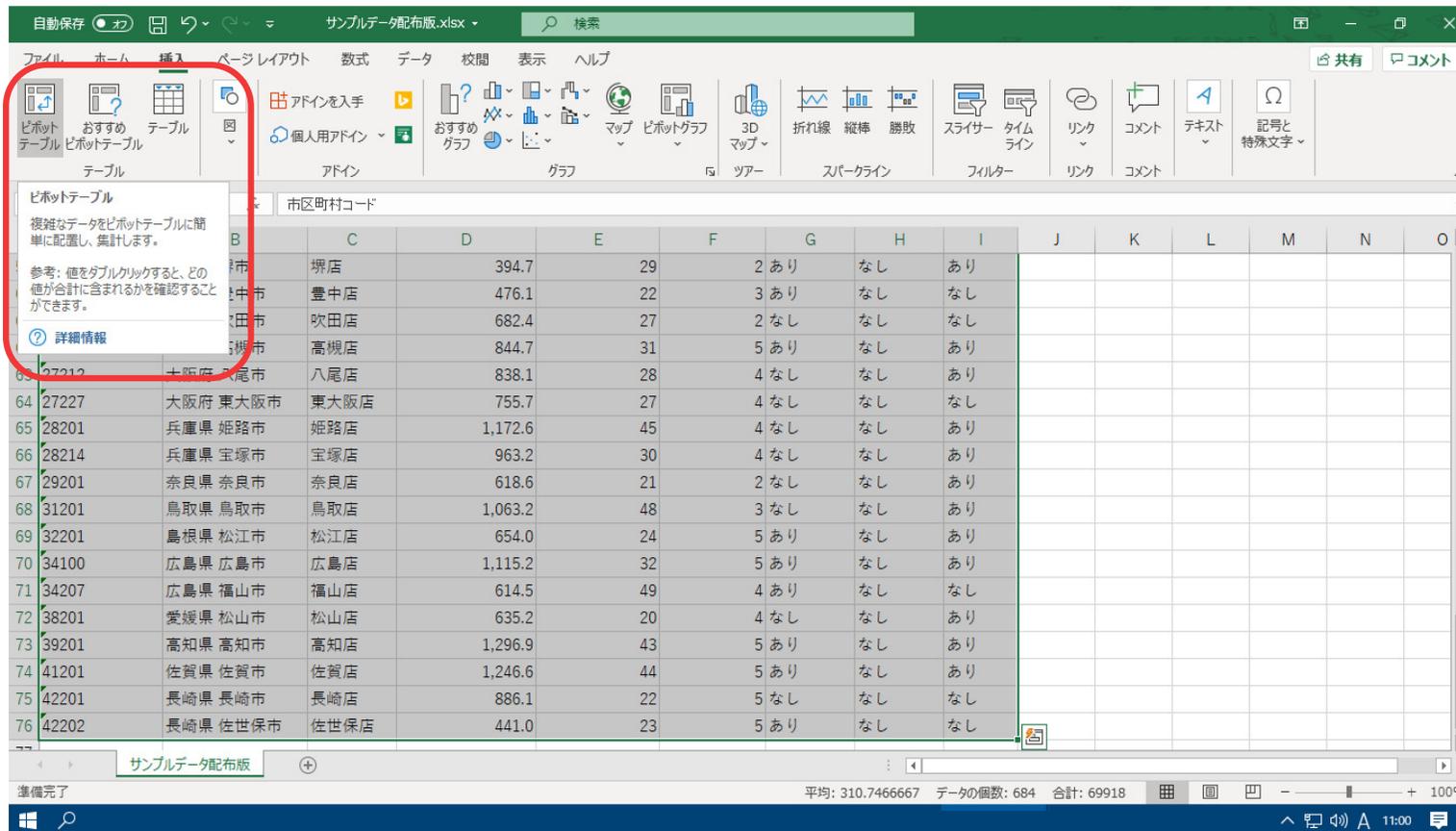
実際の集計表のデータ範囲,

「差がついてない」場合の集計表のデータ範囲

)

- 元データを選んでピボットテーブルを挿入
- 行・列に集計したい分類型の項目を挿入
- 値フィールドにどちらかの項目を挿入
- 次に「差がついてない場合の集計表」を作成
- 最後にCHITEST関数を実行

# デモ③



自動保存 ● オフ サンプルデータ配布版.xlsx 検索

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 ヘルプ

共有 コメント

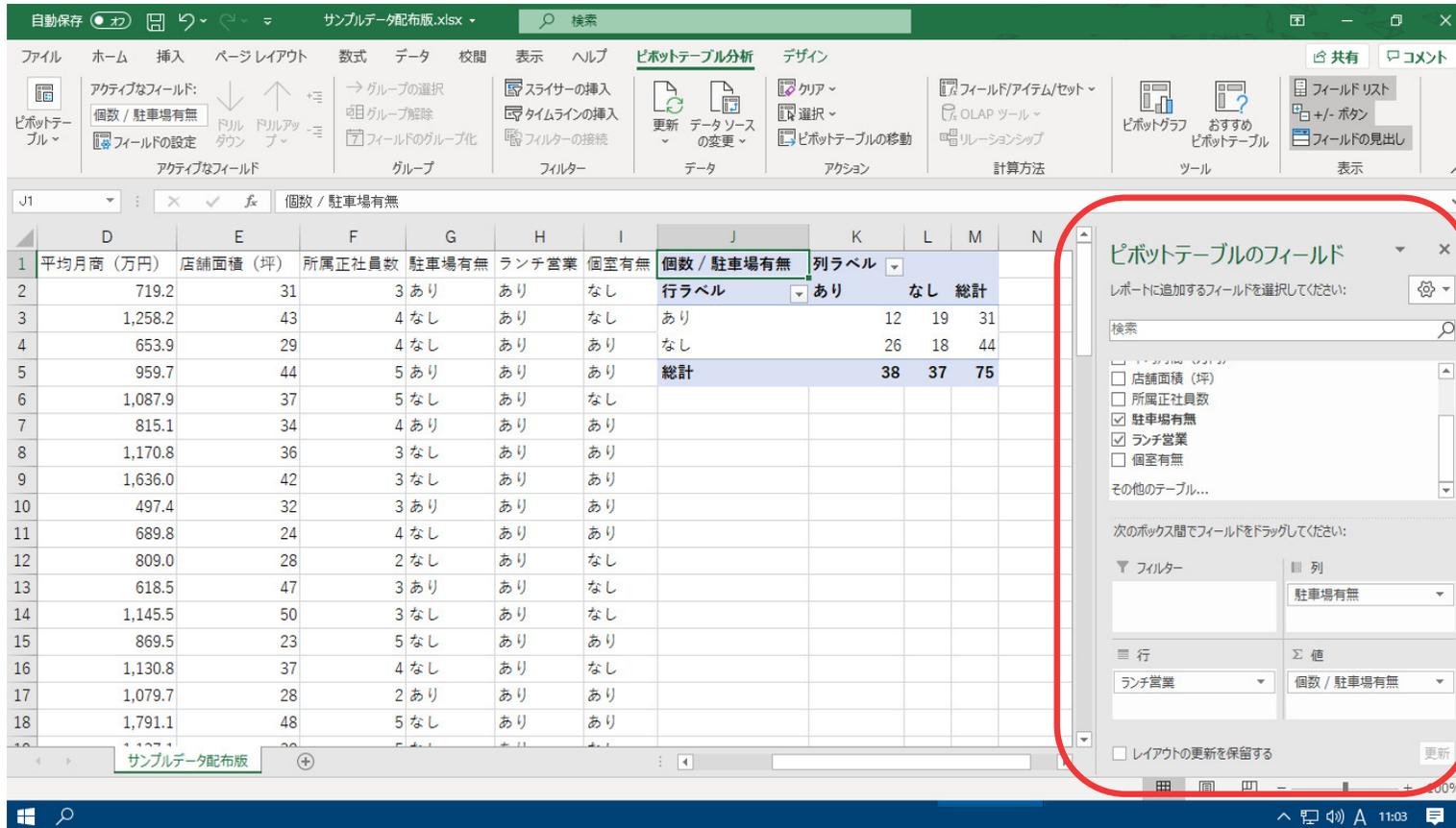
ピボットテーブル (highlighted)

ピボットテーブル  
複雑なデータをピボットテーブルに簡単に配置し、集計します。  
参考: 値をダブルクリックすると、どの値が合計に含まれるかを確認することができます。  
詳細情報

	B	C	D	E	F	G	H	I	J	K	L	M	N	O
	市区町村コード													
		堺市	堺店	394.7	29	2	あり	なし	あり					
		堺市中区	豊中店	476.1	22	3	あり	なし	なし					
		吹田市	吹田店	682.4	27	2	なし	なし	なし					
		高槻市	高槻店	844.7	31	5	あり	なし	あり					
63	27212	大阪府 八尾市	八尾店	838.1	28	4	なし	なし	あり					
64	27227	大阪府 東大阪市	東大阪店	755.7	27	4	なし	なし	なし					
65	28201	兵庫県 姫路市	姫路店	1,172.6	45	4	なし	なし	あり					
66	28214	兵庫県 宝塚市	宝塚店	963.2	30	4	なし	なし	あり					
67	29201	奈良県 奈良市	奈良店	618.6	21	2	なし	なし	あり					
68	31201	鳥取県 鳥取市	鳥取店	1,063.2	48	3	なし	なし	あり					
69	32201	鳥根県 松江市	松江店	654.0	24	5	あり	なし	あり					
70	34100	広島県 広島市	広島店	1,115.2	32	5	あり	なし	あり					
71	34207	広島県 福山市	福山店	614.5	49	4	あり	なし	なし					
72	38201	愛媛県 松山市	松山店	635.2	20	4	なし	なし	あり					
73	39201	高知県 高知市	高知店	1,296.9	43	5	あり	なし	あり					
74	41201	佐賀県 佐賀市	佐賀店	1,246.6	44	5	あり	なし	あり					
75	42201	長崎県 長崎市	長崎店	886.1	22	5	なし	なし	なし					
76	42202	長崎県 佐世保市	佐世保店	441.0	23	5	あり	なし	なし					

準備完了 平均: 310.7466667 データの個数: 684 合計: 69918 100%

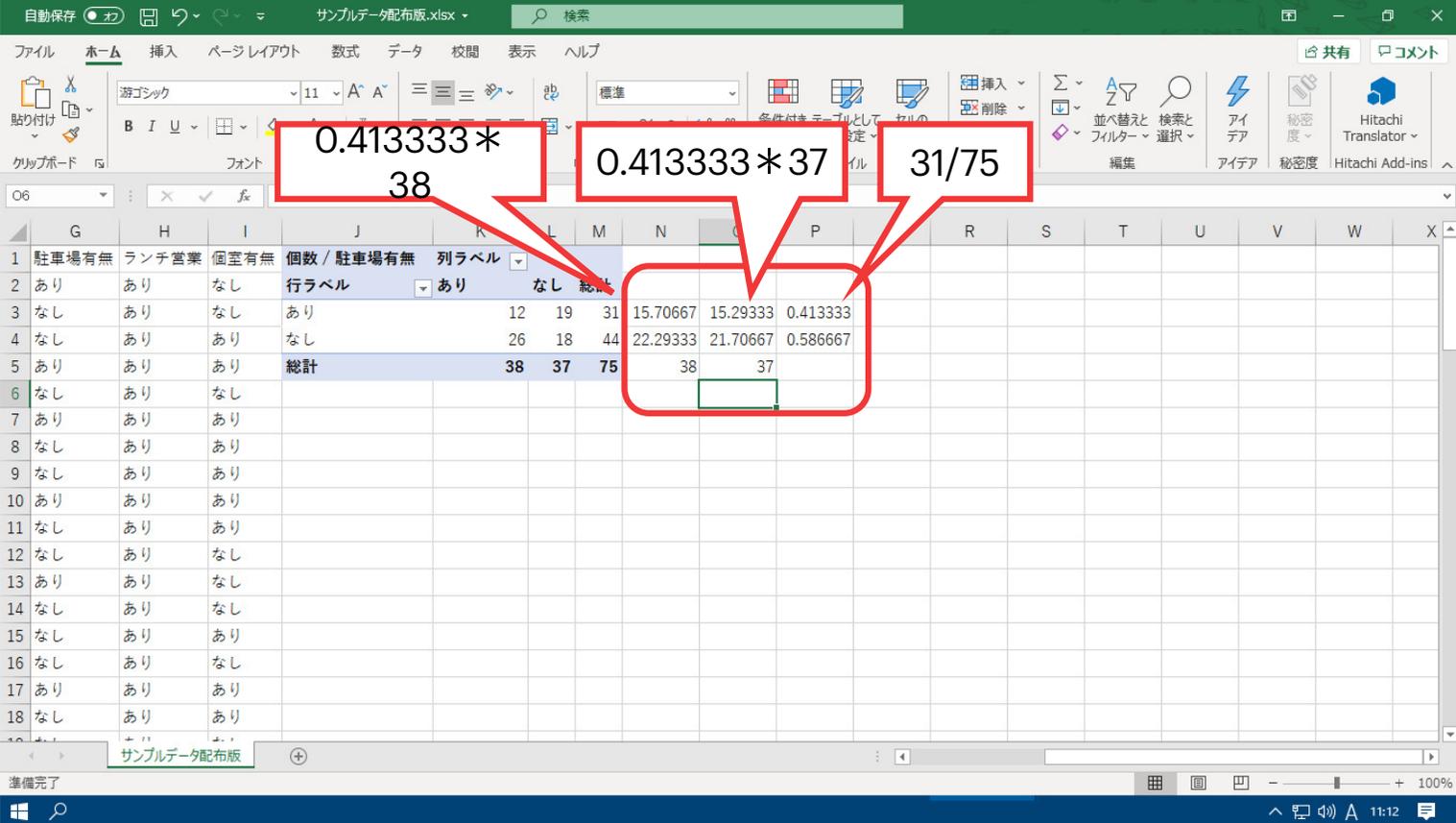
①データをすべて選択し、挿入>ピボットテーブルを選択します。



The screenshot shows an Excel spreadsheet with a PivotTable. The PivotTable is set to show counts for '駐車場有無' (Parking Available) across different categories. The task pane 'ピボットテーブルのフィールド' (PivotTable Fields) is open, showing the following configuration:

- レポートに追加するフィールドを選択してください: (Select fields to add to the report)
- 検索: (Search)
- フィールドリスト:
  - 店舗面積 (坪)
  - 所属正社員数
  - 駐車場有無
  - ランチ営業
  - 個室有無
- その他のテーブル... (Other tables...)
- 次のボックス間でフィールドをドラッグしてください: (Drag fields between the following boxes):
  - フィルター (Filter): (Empty)
  - 列 (Column Labels): 駐車場有無
  - 行 (Row Labels): ランチ営業
  - 値 (Values): 個数 / 駐車場有無
- レイアウトの更新を保留する (Save layout updates):
- 更新 (Refresh): (Button)

②ピボットテーブルのフィールドで、行・列に集計したい分類型の項目を、また挿入値フィールドにどちらかの項目を設定します。



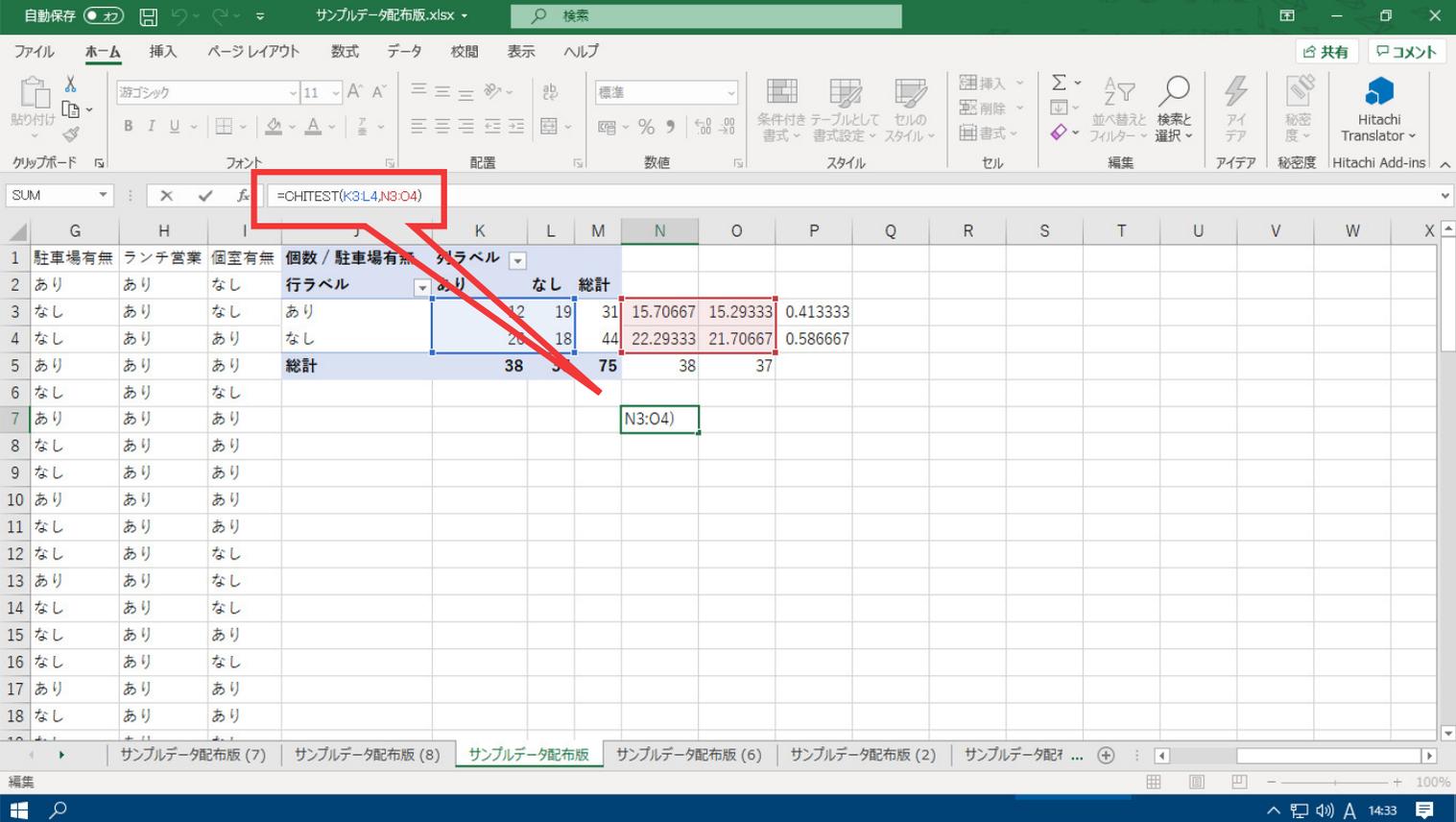
The screenshot shows an Excel spreadsheet with a pivot table. The pivot table is located in columns G through N and rows 1 through 18. The pivot table has three columns: '駐車場有無' (Parking lot presence), 'ランチ営業' (Lunch service), and '個室有無' (Private room presence). The pivot table is filtered by '行ラベル' (Row label) set to 'あり' (Yes). The pivot table shows the following data:

駐車場有無	ランチ営業	個室有無	個数	駐車場有無	列ラベル
あり	あり	なし	12	あり	なし
なし	あり	なし	19	なし	なし
なし	あり	あり	31	あり	なし
なし	あり	あり	26	なし	なし
あり	あり	あり	75	総計	なし
なし	あり	なし	38		
あり	あり	あり	37		
なし	あり	あり			
なし	あり	あり			
あり	あり	あり			
なし	あり	あり			
なし	あり	なし			
あり	あり	なし			
なし	あり	あり			
なし	あり	なし			
あり	あり	あり			
なし	あり	あり			

Callouts in the image point to the following values:

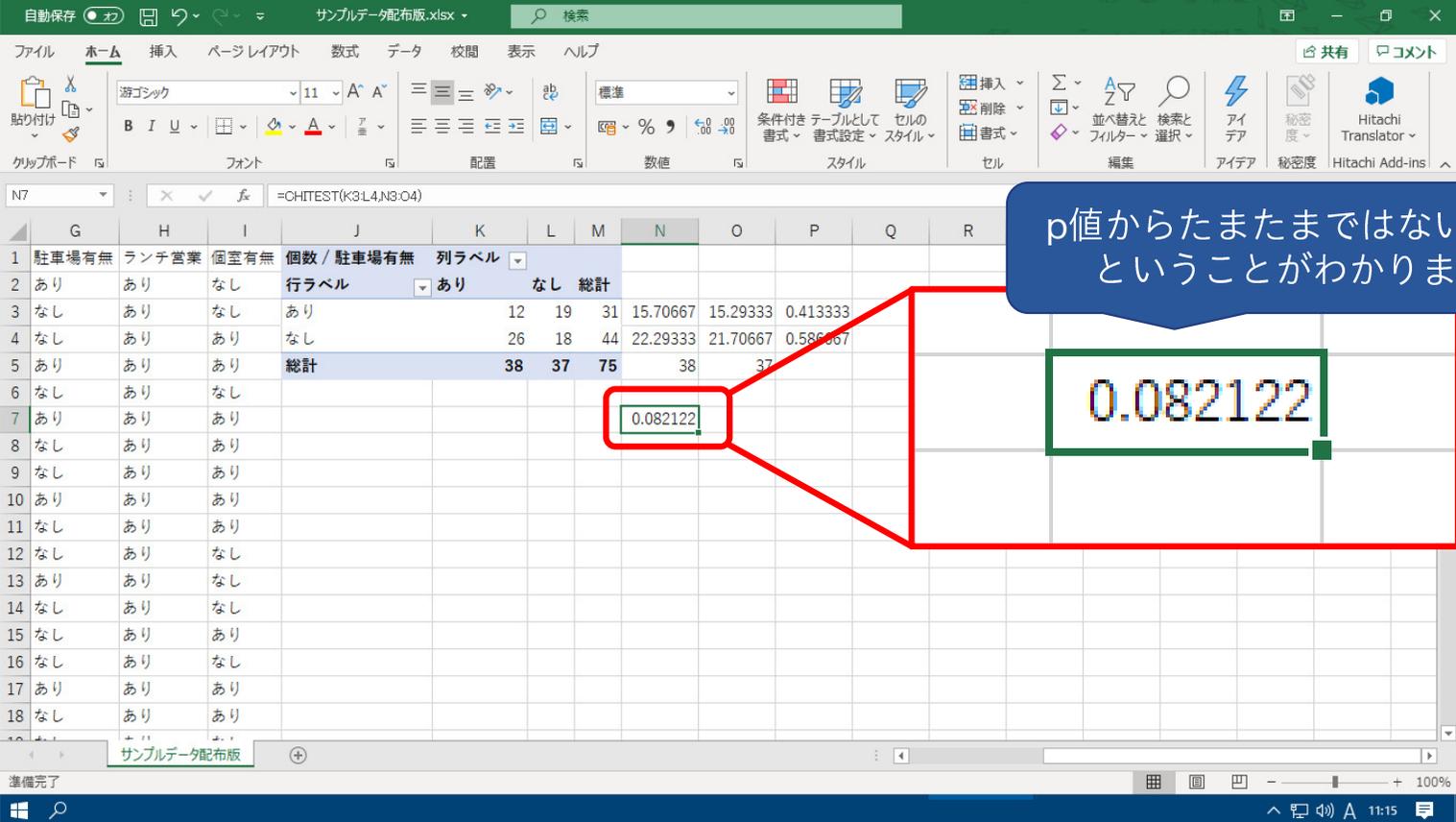
- $0.413333 * 38$  (pointing to the value 15.70667 in cell N3)
- $0.413333 * 37$  (pointing to the value 15.29333 in cell N4)
- $31/75$  (pointing to the value 0.413333 in cell N5)

③ 「差がついてない場合の集計表」を作成します。



駐車場有無	ランチ営業	個室有無	個数	駐車場有無 行ラベル	割合
あり	あり	なし	31	なし	0.413333
なし	あり	なし	12	あり	0.1570667
なし	あり	あり	19	なし	0.24375
なし	あり	あり	26	あり	0.333333
あり	あり	あり	44	なし	0.566667
なし	あり	なし	38	総計	0.4875
あり	あり	あり	75	総計	0.95

④任意の場所に =chitest(実際の集計表のデータ範囲,「差がついてない」場合の集計表のデータ範囲)を入力します。



自動保存 [ON] サンプルデータ配布版.xlsx 検索

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 ヘルプ

共有 コメント

貼り付け 遊ぶ シック 11 A A 標準

条件付き書式 テーブルとしてセルのスタイル

挿入 削除 書式

並べ替えとフィルター 検索と選択

アイデア 秘密度 Hitachi Translator

編集 アイデア 秘密度 Hitachi Add-ins

N7 : X ✓ fx =CHITEST(K3:L4,N3:O4)

	G	H	I	J	K	L	M	N	O	P	Q	R
1	駐車場有無	ランチ営業	個室有無	個数 / 駐車場有無	列ラベル							
2	あり	あり	なし	行ラベル	あり	なし	総計					
3	なし	あり	なし	あり		12	19	31	15.70667	15.29333	0.413333	
4	なし	あり	あり	なし		26	18	44	22.29333	21.70667	0.586667	
5	あり	あり	あり	総計		38	37	75		38	37	
6	なし	あり	なし									
7	あり	あり	あり									
8	なし	あり	あり									
9	なし	あり	あり									
10	あり	あり	あり									
11	なし	あり	あり									
12	なし	あり	なし									
13	あり	あり	なし									
14	なし	あり	なし									
15	なし	あり	あり									
16	なし	あり	なし									
17	あり	あり	あり									
18	なし	あり	あり									

準備完了

11:15

④結果が表示されました。これで「割合の差」の仮説検定ができました。

次に「グループ分けする」という質的な要因を探索するだけでなく、数値と数値の関係性をみるための単回帰分析の方法を学びます。

まずエクセルのグラフツールで散布図を描き、その中の回帰直線および数式を確認する方法を示すほか、ここでもやはり「たまたまのトレンド」と言えるのかどうか、分析ツールメニューから得られるp値の確認方法を示します。

それぞれのデータを「探索する」ときは・・・

数値型  
(量的)



何かの要因で分けて  
平均値の差をt検定で比較

分類型  
(質的)



何かの要因で分けて  
割合の差を $\chi^2$ 乗検定で比較

探索する要因(説明変数)

数値型  
(量的)

分類型  
(質的)

数値型  
(量的)

この  
場合は？

平均値と  
t検定

分類型  
(質的)

割合と  
 $\chi^2$ 乗検定



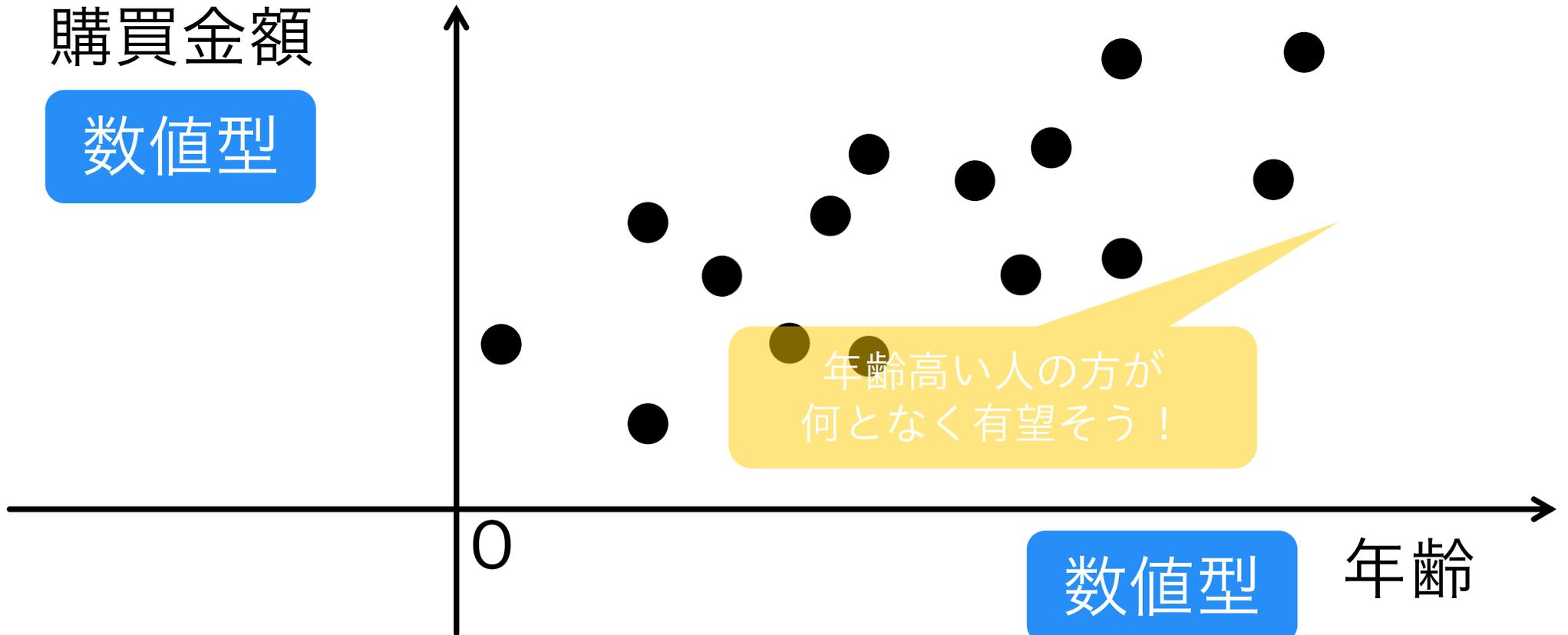
棒グラフで「平均」や「割合」に  
まとめるのではなく「1人1人の値」を表わしたい

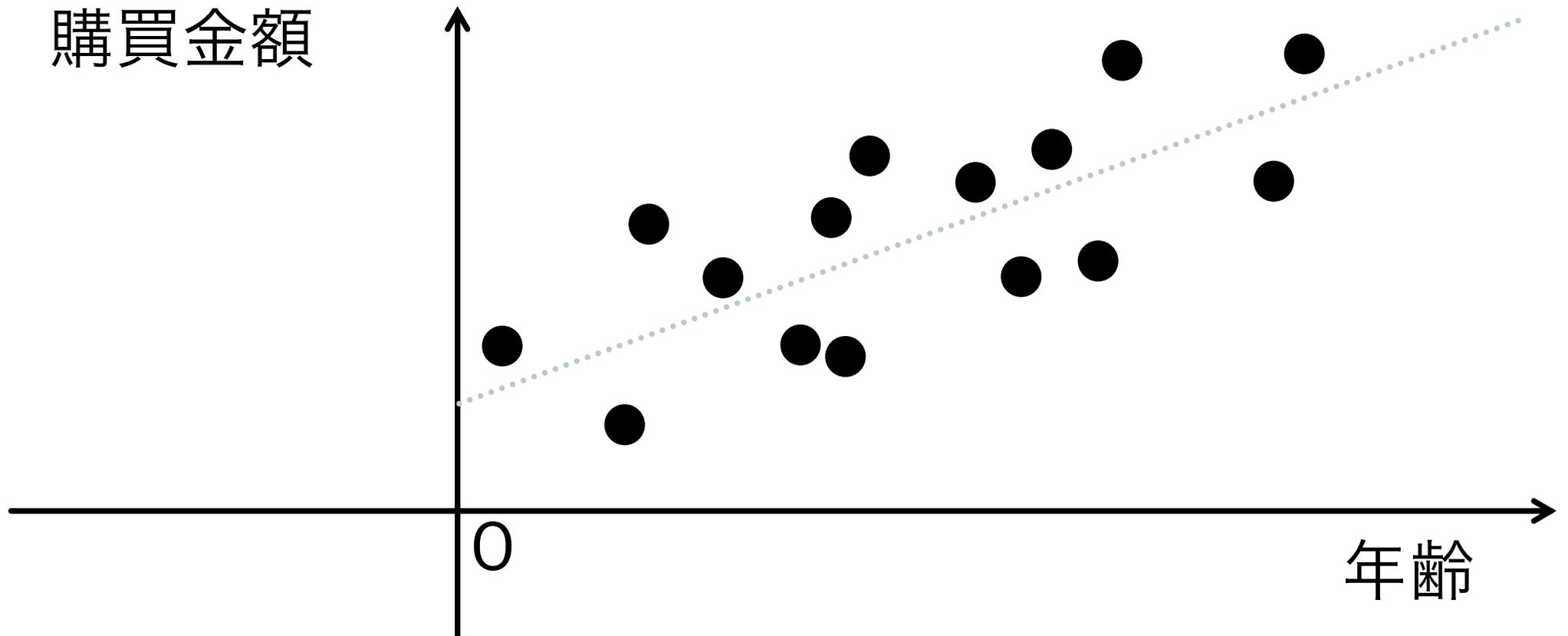


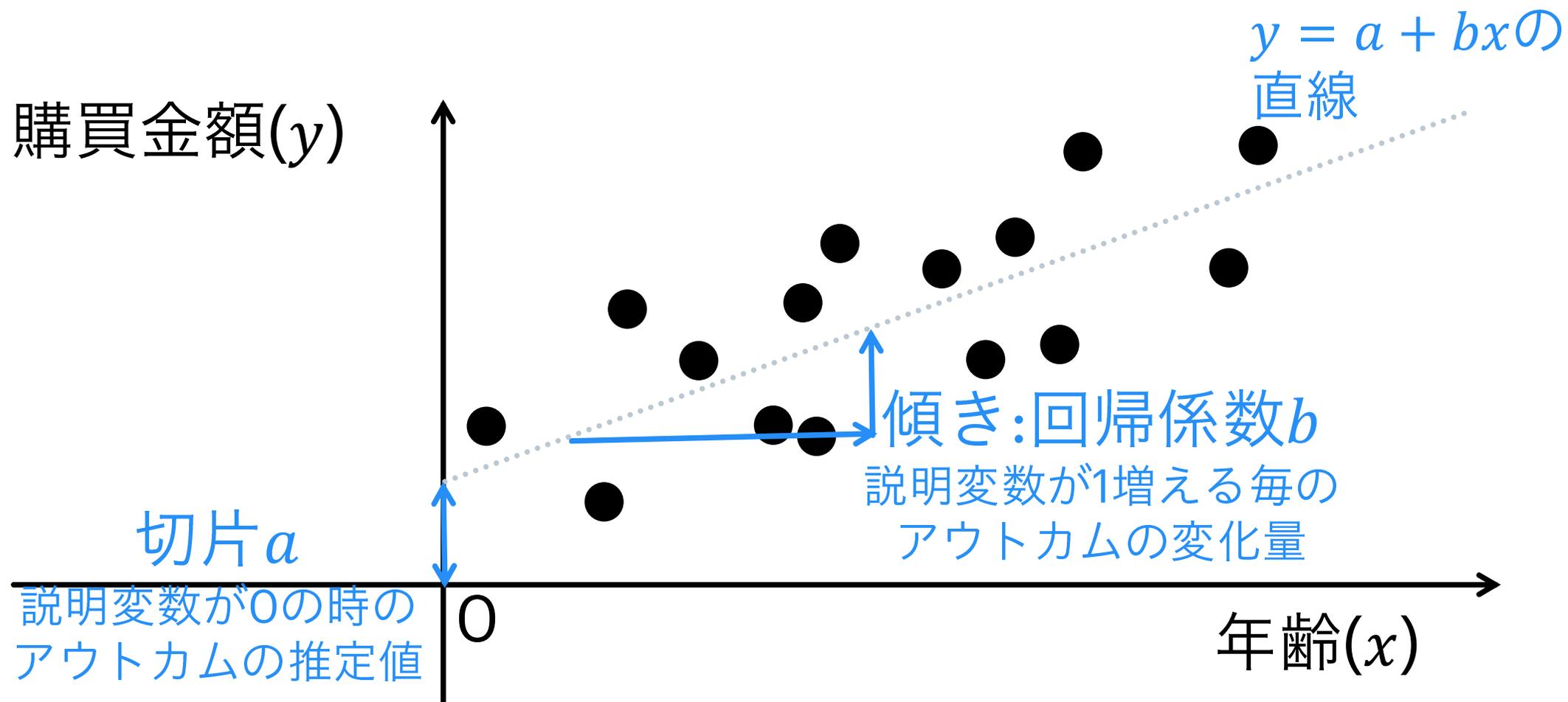
1人のデータを1つの点で表わして  
座標で表わしたものが散布図

購買金額

数値型







切片が300

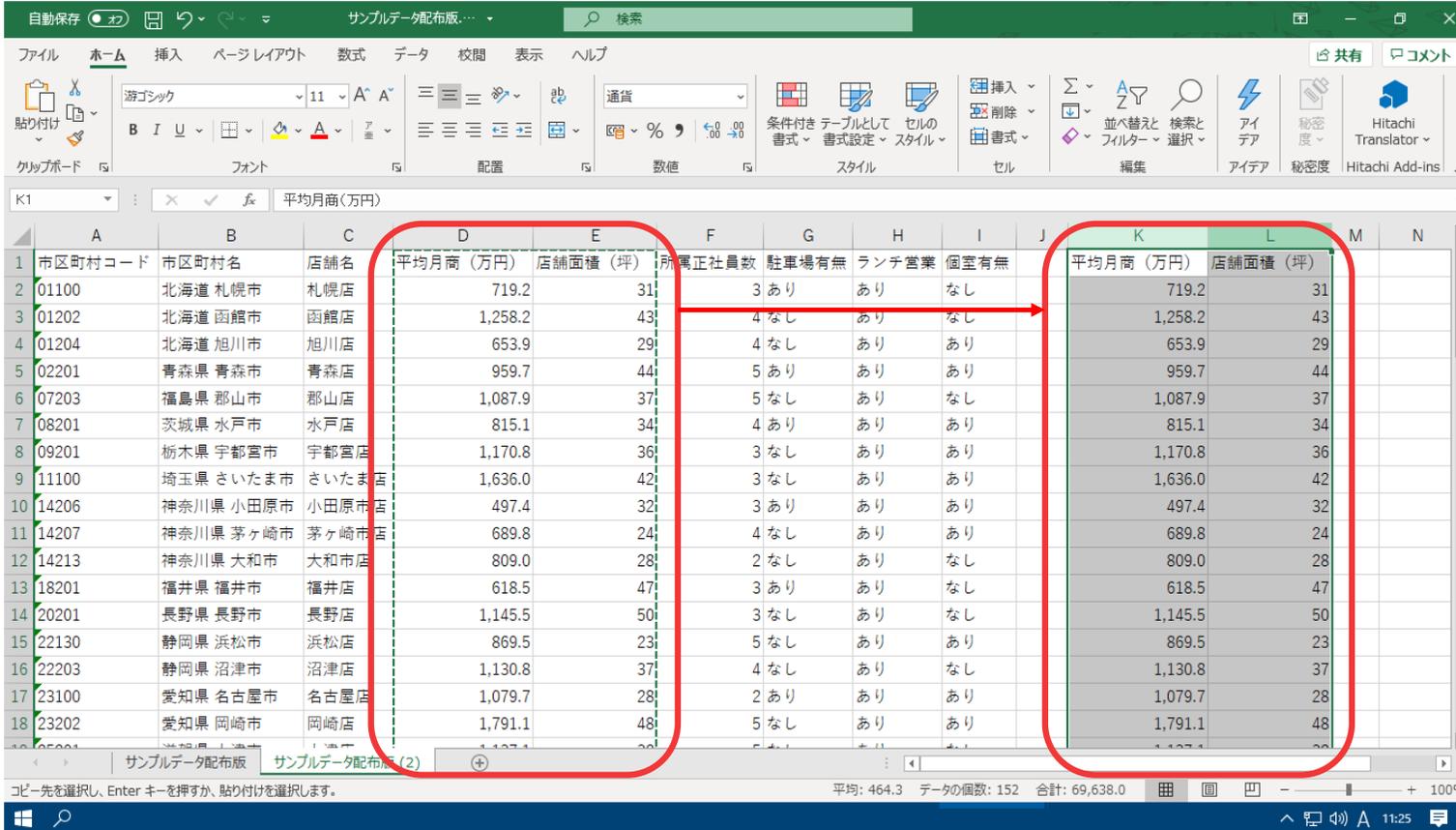
年齢に対する回帰係数(傾き)が200のとき

年齢から推定される購買金額=300+年齢×200

40才なら・・・ $300 + 40 \times 200 = 8300$ 円ぐらい買いそう

- グラフメニューから散布図を挿入
- 横軸と縦軸それぞれに数値型の項目を選択  
(横軸が説明変数・縦軸がアウトカム)
- 近似曲線の追加で「線形近似」を選ぶ
- 「グラフに数式を表示する」も選ぶ

# デモ④



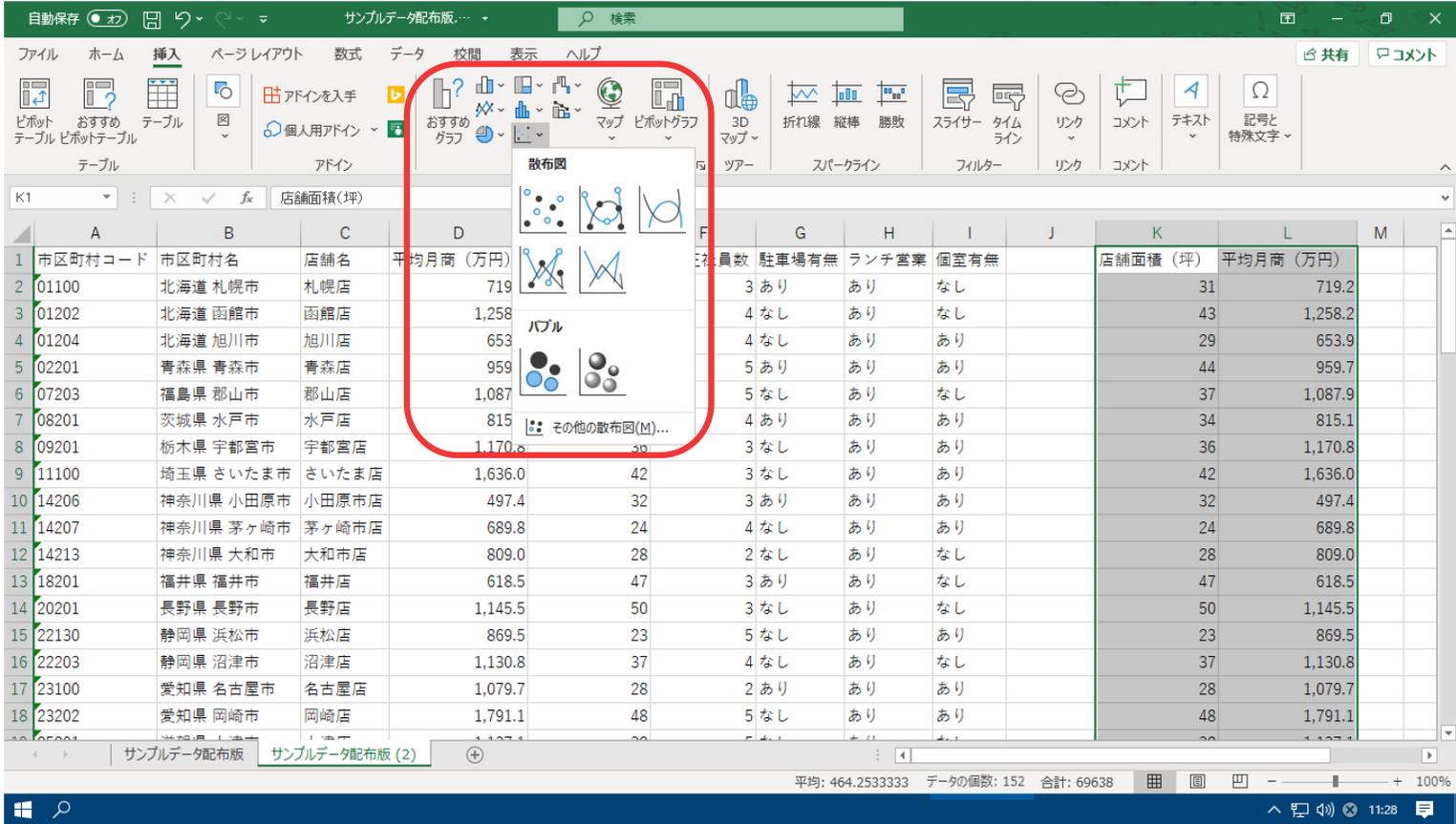
市区町村コード	市区町村名	店舗名	平均月商 (万円)	店舗面積 (坪)	所属正社員数	駐車場有無	ランチ営業	個室有無	平均月商 (万円)	店舗面積 (坪)
01100	北海道 札幌市	札幌店	719.2	31	3	あり	あり	なし	719.2	31
01202	北海道 函館市	函館店	1,258.2	43	4	なし	あり	なし	1,258.2	43
01204	北海道 旭川市	旭川店	653.9	29	4	なし	あり	あり	653.9	29
02201	青森県 青森市	青森店	959.7	44	5	あり	あり	あり	959.7	44
07203	福島県 郡山市	郡山店	1,087.9	37	5	なし	あり	なし	1,087.9	37
08201	茨城県 水戸市	水戸店	815.1	34	4	あり	あり	あり	815.1	34
09201	栃木県 宇都宮市	宇都宮店	1,170.8	36	3	なし	あり	あり	1,170.8	36
11100	埼玉県 さいたま市	さいたま店	1,636.0	42	3	なし	あり	あり	1,636.0	42
14206	神奈川県 小田原市	小田原店	497.4	32	3	あり	あり	あり	497.4	32
14207	神奈川県 茅ヶ崎市	茅ヶ崎店	689.8	24	4	なし	あり	あり	689.8	24
14213	神奈川県 大和市	大和市店	809.0	28	2	なし	あり	なし	809.0	28
18201	福井県 福井市	福井店	618.5	47	3	あり	あり	なし	618.5	47
20201	長野県 長野市	長野店	1,145.5	50	3	なし	あり	なし	1,145.5	50
22130	静岡県 浜松市	浜松店	869.5	23	5	なし	あり	あり	869.5	23
22203	静岡県 沼津市	沼津店	1,130.8	37	4	なし	あり	なし	1,130.8	37
23100	愛知県 名古屋市	名古屋店	1,079.7	28	2	あり	あり	あり	1,079.7	28
23202	愛知県 岡崎市	岡崎店	1,791.1	48	5	なし	あり	あり	1,791.1	48

①表中から「平均月商」「店舗面積」をコピーして抜き出します。

	A	B	C	D	E	F	G	H	I	J	K	L	M
	市区町村コード	市区町村名	店舗名	平均月商 (万円)	店舗面積 (坪)	所属正社員数	駐車場有無	ランチ営業	個室有無		店舗面積 (坪)	平均月商 (万円)	
2	01100	北海道 札幌市	札幌店	719.2	31	3	あり	あり	なし		31	719.2	
3	01202	北海道 函館市	函館店	1,258.2	43	4	なし	あり	なし		43	1,258.2	
4	01204	北海道 旭川市	旭川店	653.9	29	4	なし	あり	あり		29	653.9	
5	02201	青森県 青森市	青森店	959.7	44	5	あり	あり	あり		44	959.7	
6	07203	福島県 郡山市	郡山店	1,087.9	37	5	なし	あり	なし		37	1,087.9	
7	08201	茨城県 水戸市	水戸店	815.1	34	4	あり	あり	あり		34	815.1	
8	09201	栃木県 宇都宮市	宇都宮店	1,170.8	36	3	なし	あり	あり		36	1,170.8	
9	11100	埼玉県 さいたま市	さいたま店	1,636.0	42	3	なし	あり	あり		42	1,636.0	
10	14206	神奈川県 小田原市	小田原市店	497.4	32	3	あり	あり	あり		32	497.4	
11	14207	神奈川県 茅ヶ崎市	茅ヶ崎市店	689.8	24	4	なし	あり	あり		24	689.8	
12	14213	神奈川県 大和市	大和市店	809.0	28	2	なし	あり	なし		28	809.0	
13	18201	福井県 福井市	福井店	618.5	47	3	あり	あり	なし		47	618.5	
14	20201	長野県 長野市	長野店	1,145.5	50	3	なし	あり	なし		50	1,145.5	
15	22130	静岡県 浜松市	浜松店	869.5	23	5	なし	あり	あり		23	869.5	
16	22203	静岡県 沼津市	沼津店	1,130.8	37	4	なし	あり	なし		37	1,130.8	
17	23100	愛知県 名古屋市	名古屋店	1,079.7	28	2	あり	あり	あり		28	1,079.7	
18	23202	愛知県 岡崎市	岡崎店	1,791.1	48	5	なし	あり	あり		48	1,791.1	

②左側に「店舗面積」がくるように並び替えます

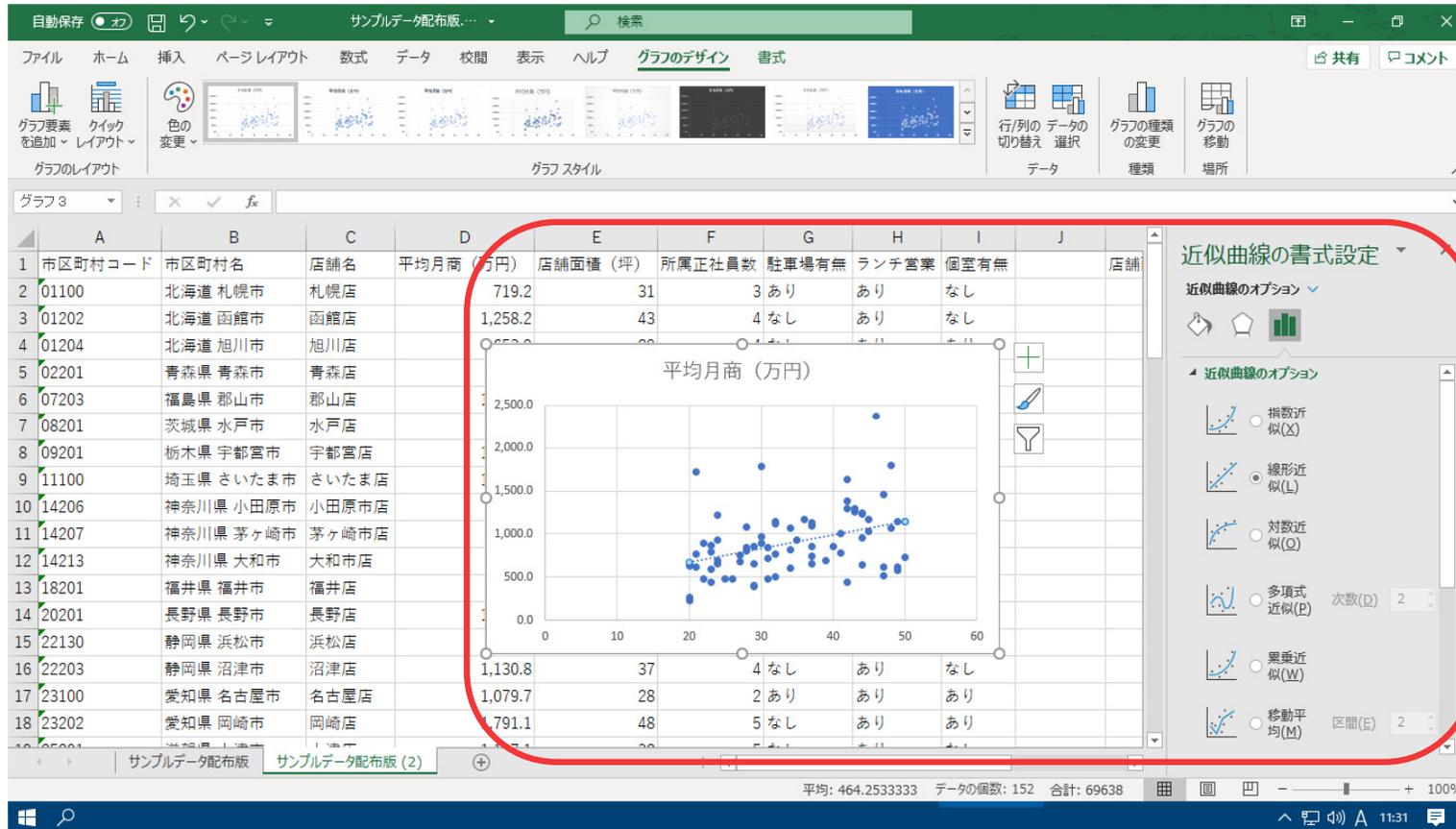
# デモ④



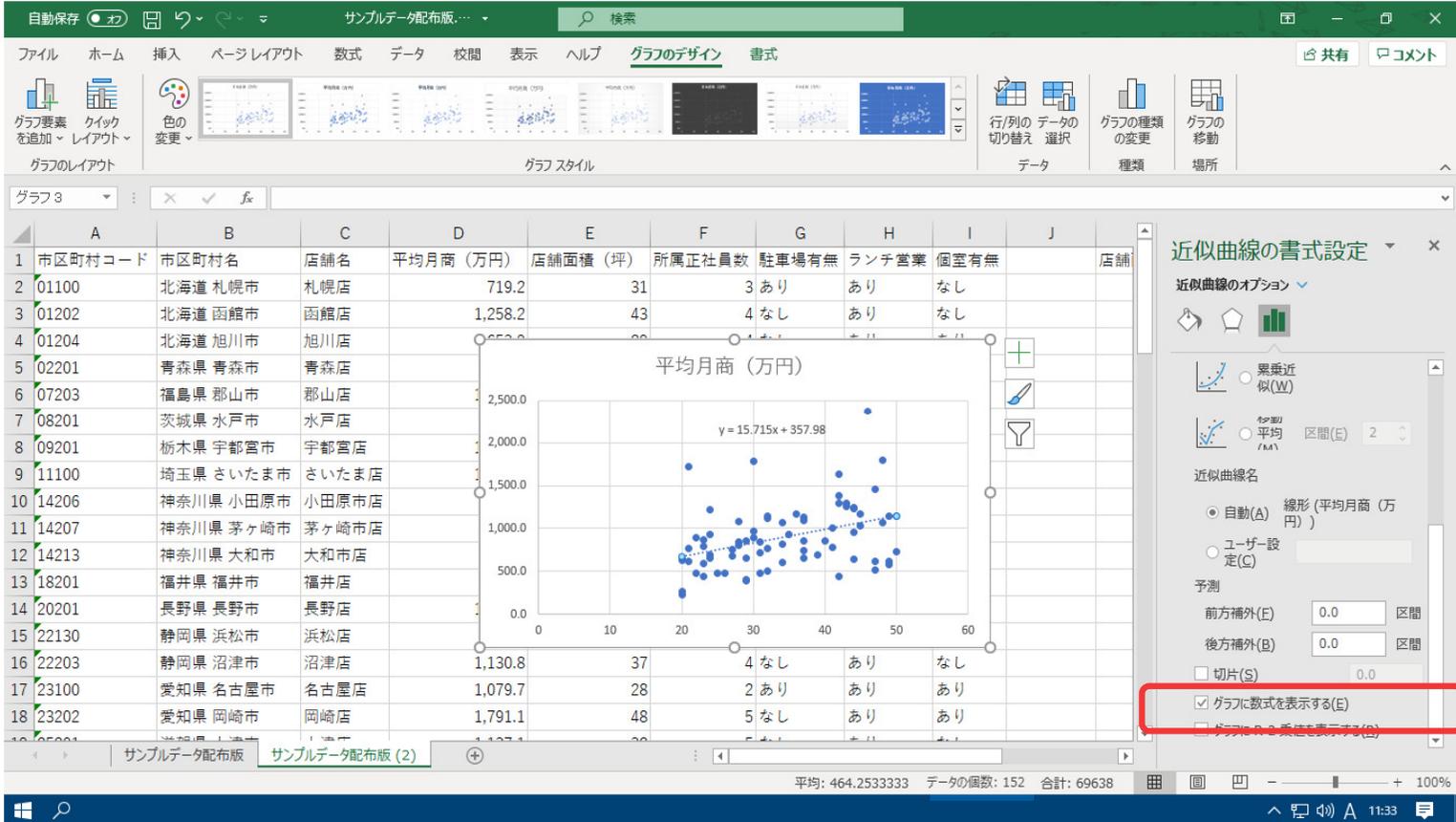
The screenshot shows the Microsoft Excel interface with the 'Insert' ribbon active. The 'Charts' group is expanded, and the 'Scatter' chart type is selected. The spreadsheet data is as follows:

市区町村コード	市区町村名	店舗名	平均月商 (万円)	店舗面積 (坪)
01100	北海道 札幌市	札幌店	719	31
01202	北海道 函館市	函館店	1,258	43
01204	北海道 旭川市	旭川店	653	29
02201	青森県 青森市	青森店	959	44
07203	福島県 郡山市	郡山店	1,087	37
08201	茨城県 水戸市	水戸店	815	34
09201	栃木県 宇都宮市	宇都宮店	1,170.8	36
11100	埼玉県 さいたま市	さいたま店	1,636.0	42
14206	神奈川県 小田原市	小田原店	497.4	32
14207	神奈川県 茅ヶ崎市	茅ヶ崎市店	689.8	24
14213	神奈川県 大和市	大和市店	809.0	28
18201	福井県 福井市	福井店	618.5	47
20201	長野県 長野市	長野店	1,145.5	50
22130	静岡県 浜松市	浜松店	869.5	23
22203	静岡県 沼津市	沼津店	1,130.8	37
23100	愛知県 名古屋市	名古屋店	1,079.7	28
23202	愛知県 岡崎市	岡崎店	1,791.1	48

③ 「店舗面積」列と「平均月商」列を選択し、挿入>散佈図を選択すると散佈図が表示されます。



④散布図内で、右クリック>近似値曲線を表示 を選択し、書式設定で「線形近似」を選択すると図内に近似値曲線が表示されます。

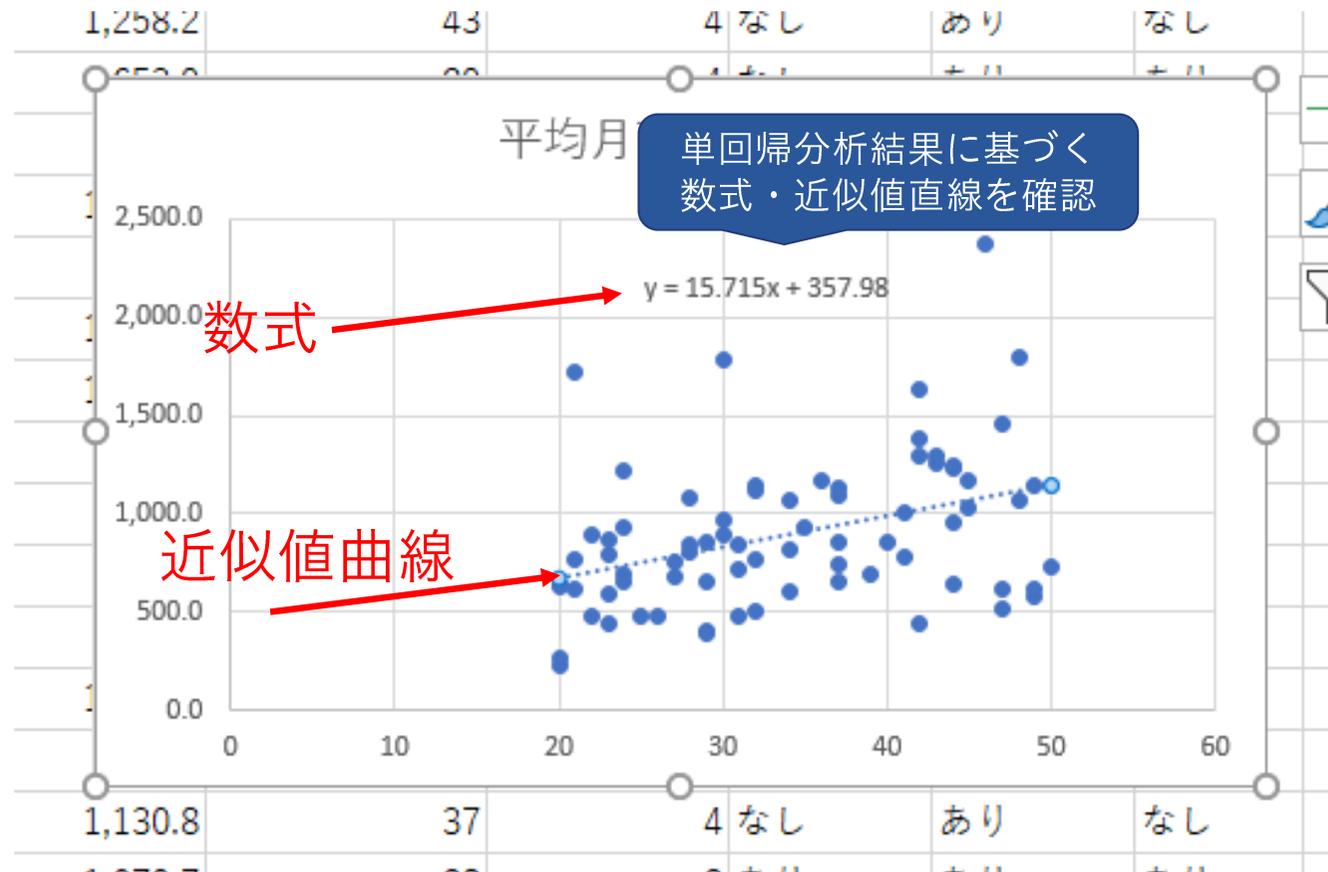


The screenshot shows an Excel spreadsheet with a scatter plot. The x-axis is labeled '店舗面積 (坪)' (Store Area) and the y-axis is '平均月商 (万円)' (Average Monthly Revenue). A linear trendline is shown with the equation  $y = 15.715x + 357.98$ . The 'Format Trendline' task pane is open on the right side of the chart area. In this pane, under the 'Options' section, the checkbox 'Show equation on chart' (グラフに数式を表示する(E)) is checked and highlighted with a red rectangular box.

市区町村コード	市区町村名	店舗名	平均月商 (万円)	店舗面積 (坪)	所属正社員数	駐車場有無	ランチ営業	個室有無	店舗
01100	北海道 札幌市	札幌店	719.2	31	3	あり	あり	なし	
01202	北海道 函館市	函館店	1,258.2	43	4	なし	あり	なし	
01204	北海道 旭川市	旭川店	652.0	20	2	あり	あり	あり	
02201	青森県 青森市	青森店	1,000.0	30	3	あり	あり	あり	
07203	福島県 郡山市	郡山店	1,000.0	30	3	あり	あり	あり	
08201	茨城県 水戸市	水戸店	1,000.0	30	3	あり	あり	あり	
09201	栃木県 宇都宮市	宇都宮店	1,000.0	30	3	あり	あり	あり	
11100	埼玉県 さいたま市	さいたま店	1,000.0	30	3	あり	あり	あり	
14206	神奈川県 小田原市	小田原市店	1,000.0	30	3	あり	あり	あり	
14207	神奈川県 茅ヶ崎市	茅ヶ崎市店	1,000.0	30	3	あり	あり	あり	
14213	神奈川県 大和市	大和市店	1,000.0	30	3	あり	あり	あり	
18201	福井県 福井市	福井店	1,000.0	30	3	あり	あり	あり	
20201	長野県 長野市	長野店	1,000.0	30	3	あり	あり	あり	
22130	静岡県 浜松市	浜松店	1,000.0	30	3	あり	あり	あり	
22203	静岡県 沼津市	沼津店	1,130.8	37	4	なし	あり	なし	
23100	愛知県 名古屋市	名古屋店	1,079.7	28	2	あり	あり	あり	
23202	愛知県 岡崎市	岡崎店	1,791.1	48	5	なし	あり	あり	

④書式設定下部にある「グラフに数式を表示する」にチェックを入れます。

# デモ④



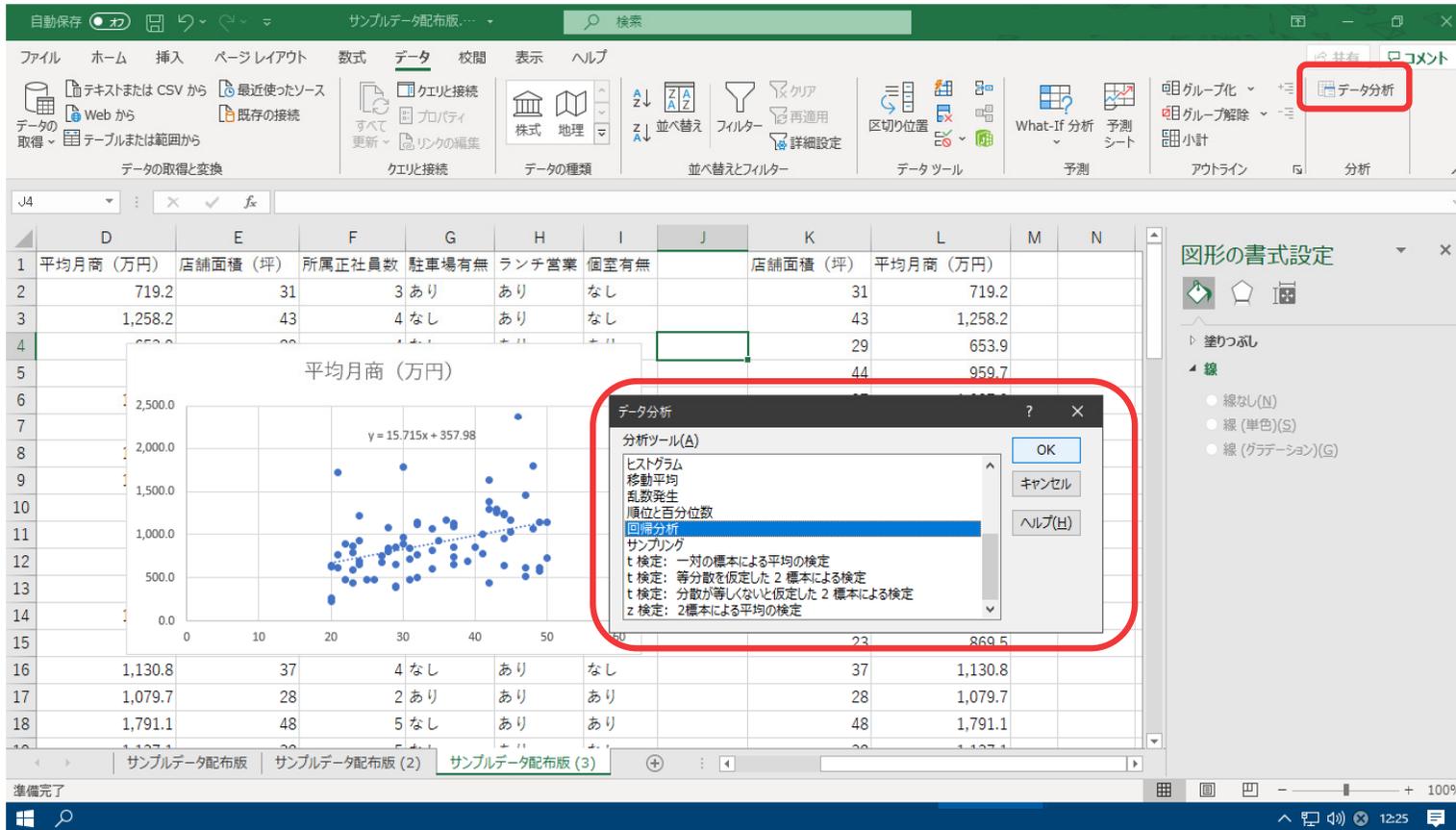
⑤図内に近似値曲線と数式が表示されたら完成です。

もちろん回帰直線のトレンドに対してもこんな反論が…

それぐらいのトレンド、**たまたま**じゃないか？

- データメニューから「データ分析」（アドイン）
- 「回帰分析」を選択
- Yの方にアウトカム・Xの方に説明変数
- 項目名を含む選択なら「ラベル」に✓
- どこか空白の場所を「一覧の出力先」に
- 最後の表の「係数」の欄と「P-値」の欄に注目

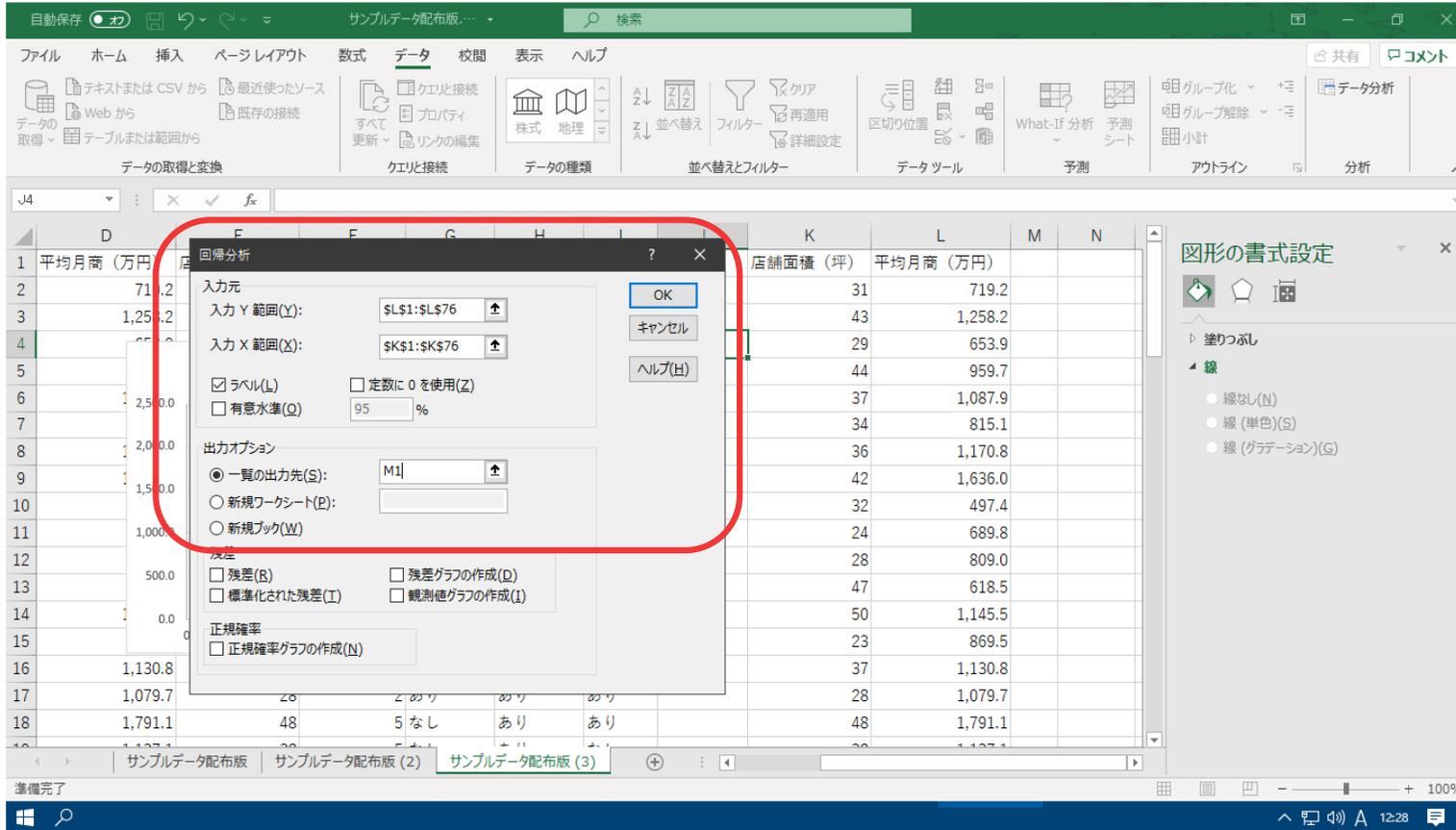
# デモ⑤



The screenshot shows the Microsoft Excel interface. The 'Data' ribbon is selected, and the 'Data Analysis' button is highlighted with a red box. A scatter plot titled '平均月商 (万円)' is displayed, showing a positive correlation between two variables. The regression equation is  $y = 15.715x + 357.98$ . The 'Data Analysis' task pane is open, and 'Regression Analysis' is selected, also highlighted with a red box.

平均月商 (万円)	店舗面積 (坪)	所属正社員数	駐車場有無	ランチ営業	個室有無	店舗面積 (坪)	平均月商 (万円)
719.2	31	3	あり	あり	なし	31	719.2
1,258.2	43	4	なし	あり	なし	43	1,258.2
653.9	29	4	あり	あり	あり	29	653.9
959.7	44	4	あり	あり	あり	44	959.7
1,130.8	37	4	なし	あり	なし	37	1,130.8
1,079.7	28	2	あり	あり	あり	28	1,079.7
1,791.1	48	5	なし	あり	あり	48	1,791.1

- ①データ > データ分析 > 回帰分析を選択します。  
(データ分析メニューはアドインから追加できます)

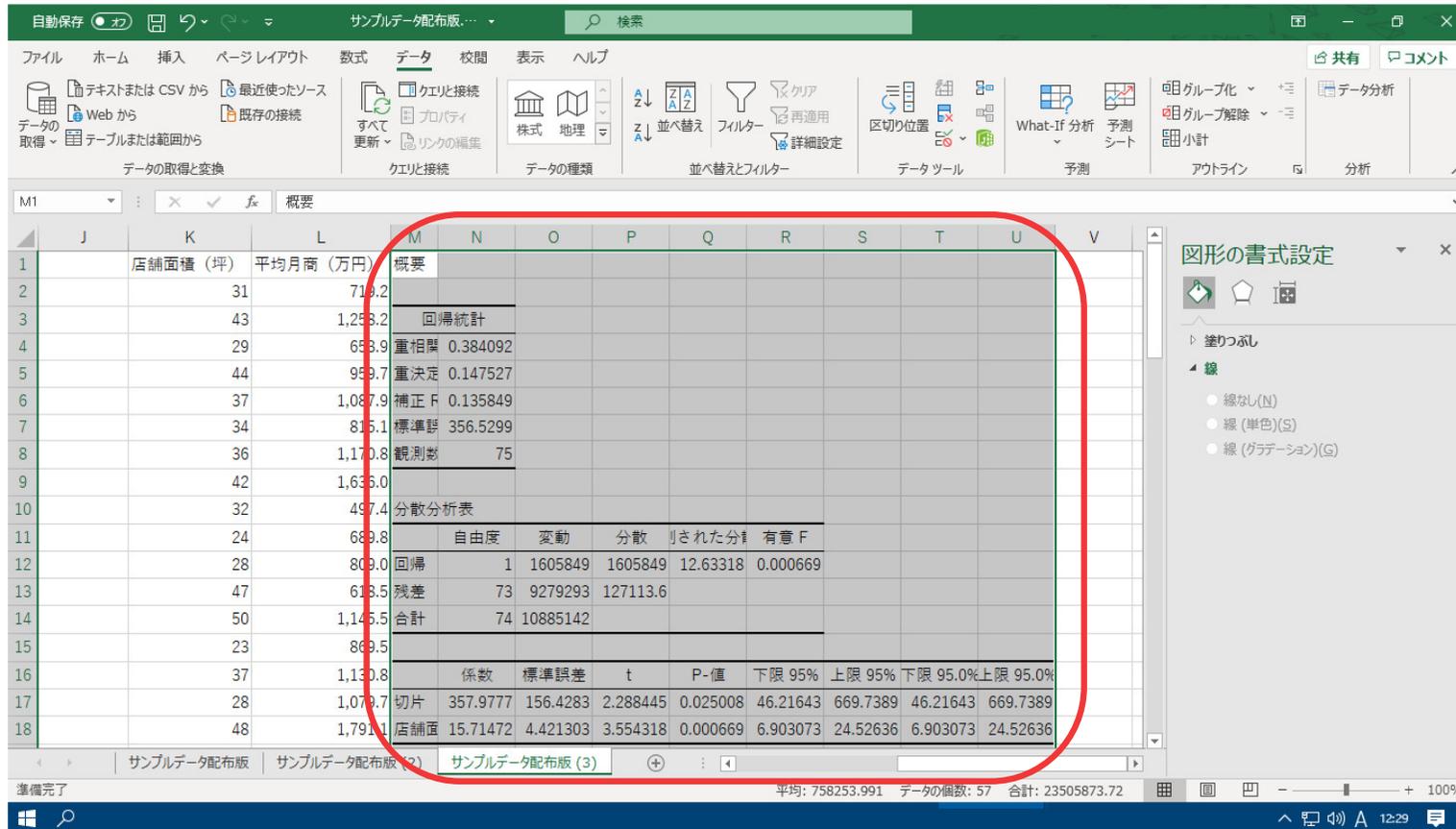


The screenshot shows the '回帰分析' (Regression Analysis) dialog box in Excel. The dialog is open over a spreadsheet with columns for '平均月商 (万円)' and '店舗面積 (坪)'. The '入力 Y 範囲(Y):' is set to '\$L\$1:\$L\$76' and '入力 X 範囲(X):' is set to '\$K\$1:\$K\$76'. The 'ラベル(L)' checkbox is checked. The '出力先(S):' is set to 'M1'. The '正規確率' section has '正規確率グラフの作成(N)' checked.

店舗面積 (坪)	平均月商 (万円)
31	719.2
43	1,258.2
29	653.9
44	959.7
37	1,087.9
34	815.1
36	1,170.8
42	1,636.0
32	497.4
24	689.8
28	809.0
47	618.5
50	1,145.5
23	869.5
37	1,130.8
28	1,079.7
48	1,791.1

②Yには「平均月商」列、Xには「店舗面積」列を指定します。  
 選択した「平均月商」「店舗面積」は一行目がラベルなので「ラベル」にチェックを入れます。出力オプションは任意の位置を指定します。

# デモ⑤



概要	
回帰統計	
重相関	0.384092
重決定	0.147527
補正 F	0.135849
標準誤	356.5299
観測数	75

分散分析表					
	自由度	変動	分散	割された分	有意 F
回帰	1	1605849	1605849	12.63318	0.000669
残差	73	9279293	127113.6		
合計	74	10885142			

	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	357.9777	156.4283	2.288445	0.025008	46.21643	669.7389	46.21643	669.7389
店舗面	15.71472	4.421303	3.554318	0.000669	6.903073	24.52636	6.903073	24.52636

③回帰分析の結果が表示されます。

# デモ⑤

1)月商 (万円)	概要								
719.2									
1,258.2	回帰統計								
653.9	重相関	0.384092							
959.7	重決定	0.147527							
1,087.9	補正 F	0.135849							
815.1	標準誤	356.5299							
1,170.8	観測数	75							
1,636.0									
497.4	分散分析表								
689.8		自由度	変動	分散	割された分	有意 F			
			5849	16058					
			293	12711					
			142						
869.5									
1,130.8		係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
1,079.7	切片	357.9777	156.4283	2.288445	0.025008	46.21643	669.7389	46.21643	669.7389
1,791.1	店舗面	15.71472	4.421303	3.554318	0.000669	6.903073	24.52636	6.903073	24.52636

数式で利用する切片・係数となります

p値からたまたまとは考えにくいことが分かります

④係数と p 値はここに表示されています。

ここではまず、「ダミー変数」という考え方を通して単回帰分析とt検定が本質的には同じようなことをやっているということを示した上で、単回帰分析の限界と、それを重回帰分析で解決できることを説明します。

次に先程と同じ分析ツールメニューを使って重回帰分析を実行し、その結果の解釈方法について学びます。

探索する要因(説明変数)

数値型  
(量的)

分類型  
(質的)

数値型  
(量的)

単回帰  
分析

平均値と  
t検定

分類型  
(質的)

割合と  
 $\chi^2$ 乗検定

探索する要因(説明変数)

数値型  
(量的)

分類型  
(質的)

数値型  
(量的)

単回帰分析でも実は  
平均値とt検定

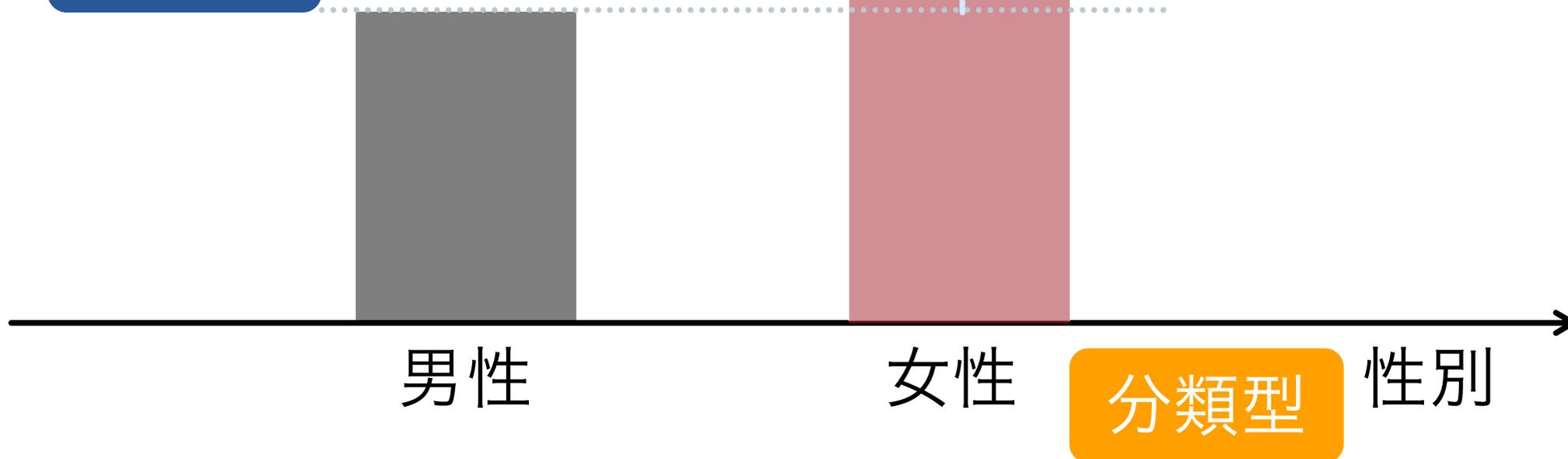
分類型  
(質的)

割合と  
 $\chi^2$ 乗検定

購買金額

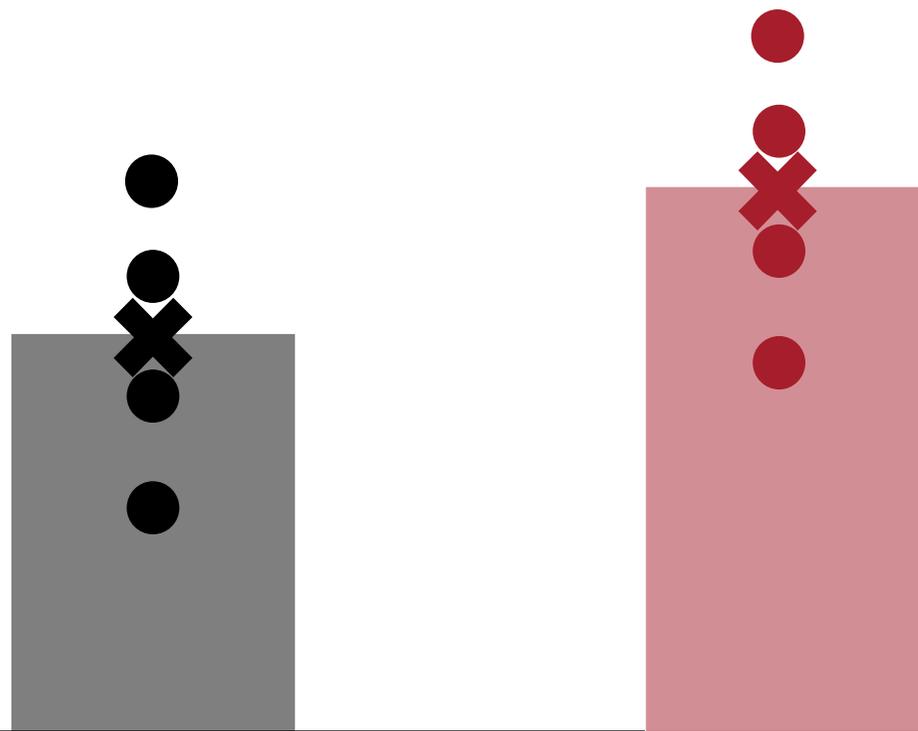
数値型

平均値の差を  
検証



購買金額

数値型



男性

女性

分類型

性別

数値型

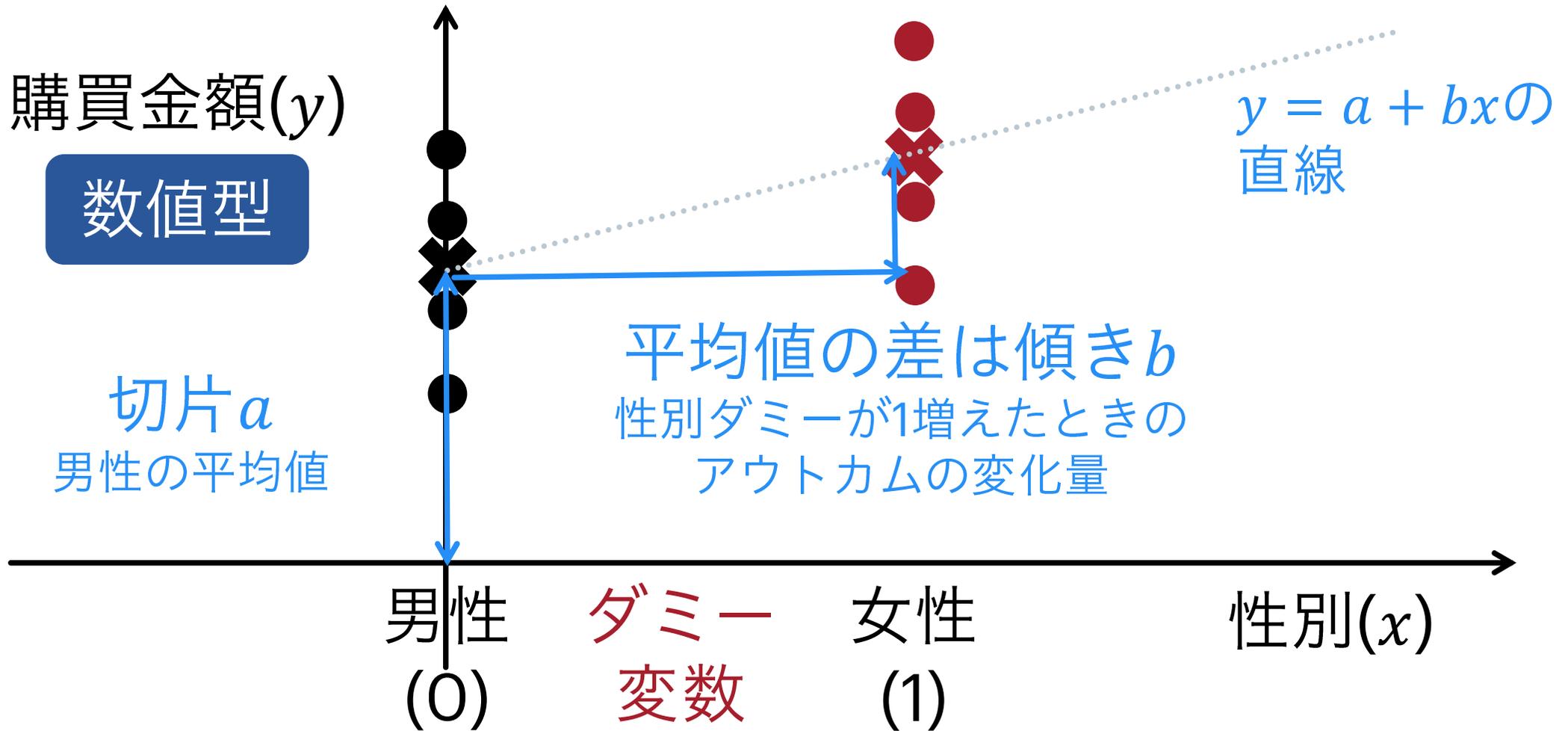


男性  
(0)

ダミー  
変数

女性  
(1)

性別



質的に異なる分類を数値として扱いたい



ある分類に該当するなら「1」

そうでなければ「0」としたダミー変数を用いる

=IF(条件式,

条件式が当てはまるときの値,

そうでないときの値)

例： =IF(C2="特定の 카테고리",1,0)

探索する要因(説明変数)

数値型  
(量的)

分類型  
(質的)

数値型  
(量的)

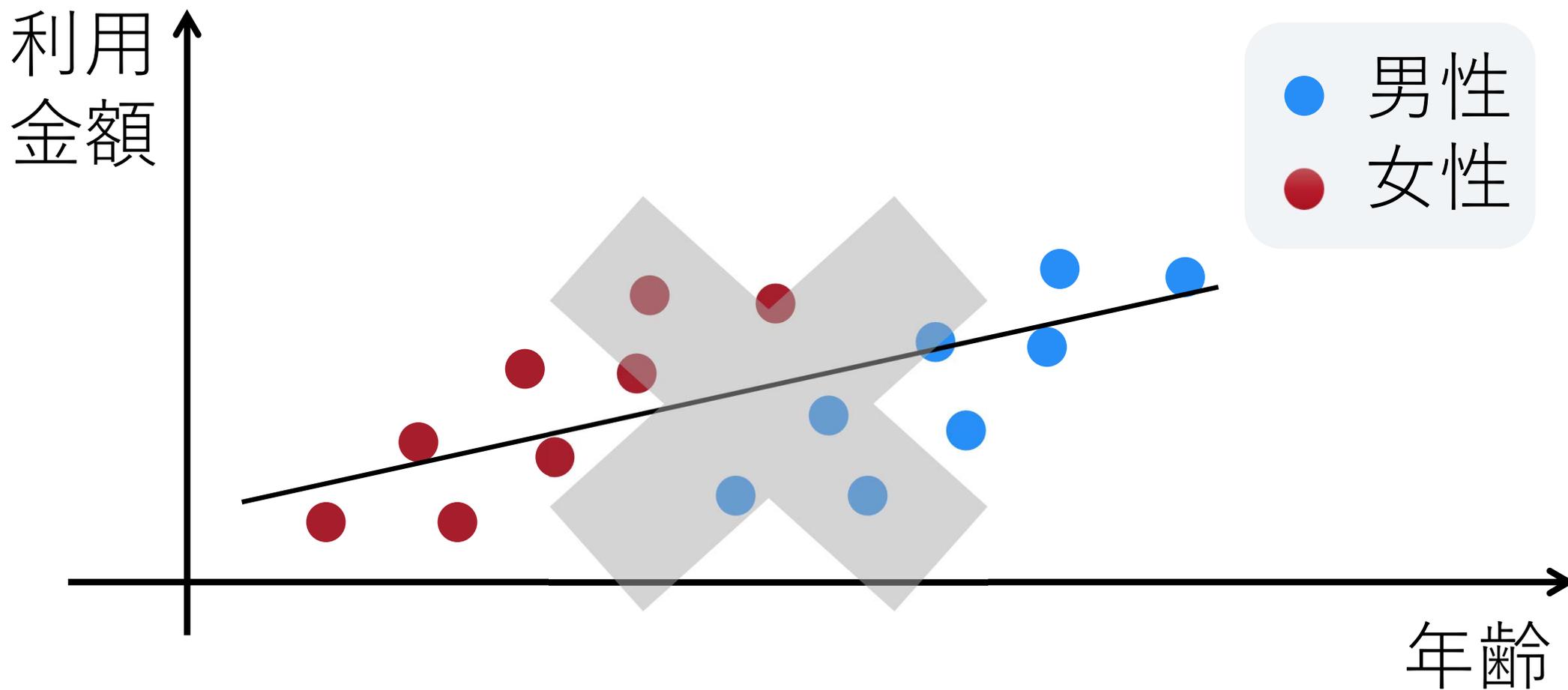
単回帰  
分析

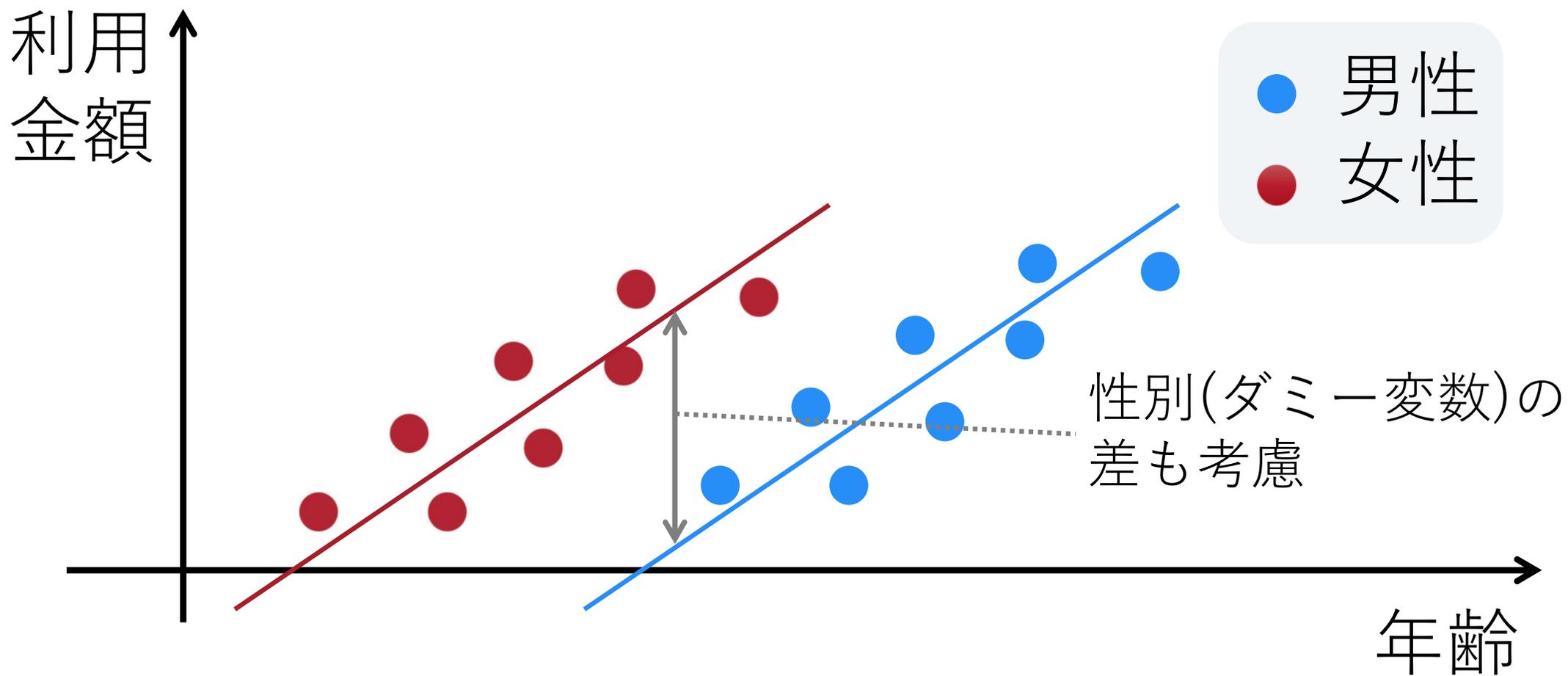
平均値と  
t検定

単回帰分析でOK

分類型  
(質的)

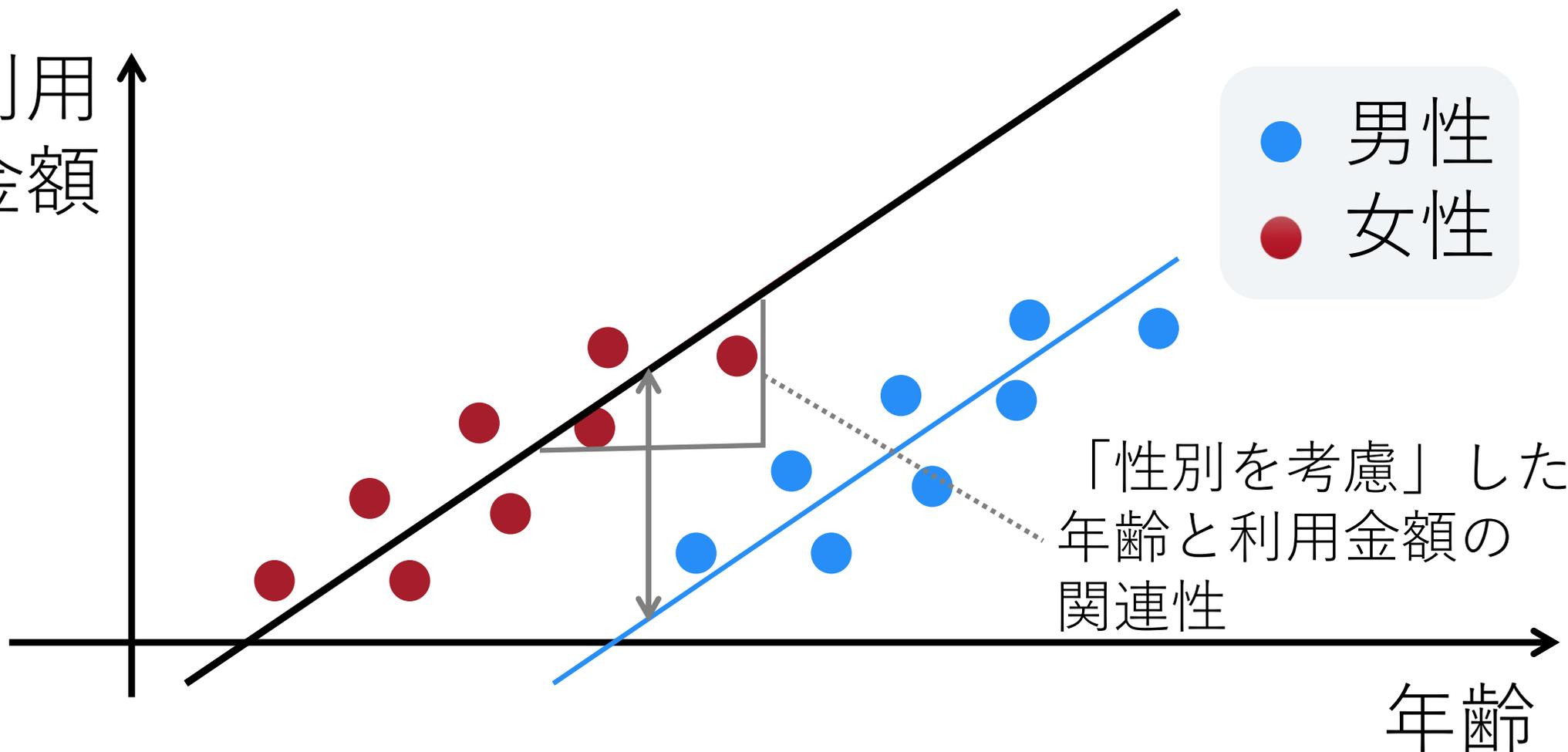
割合と  
 $\chi^2$ 乗検定





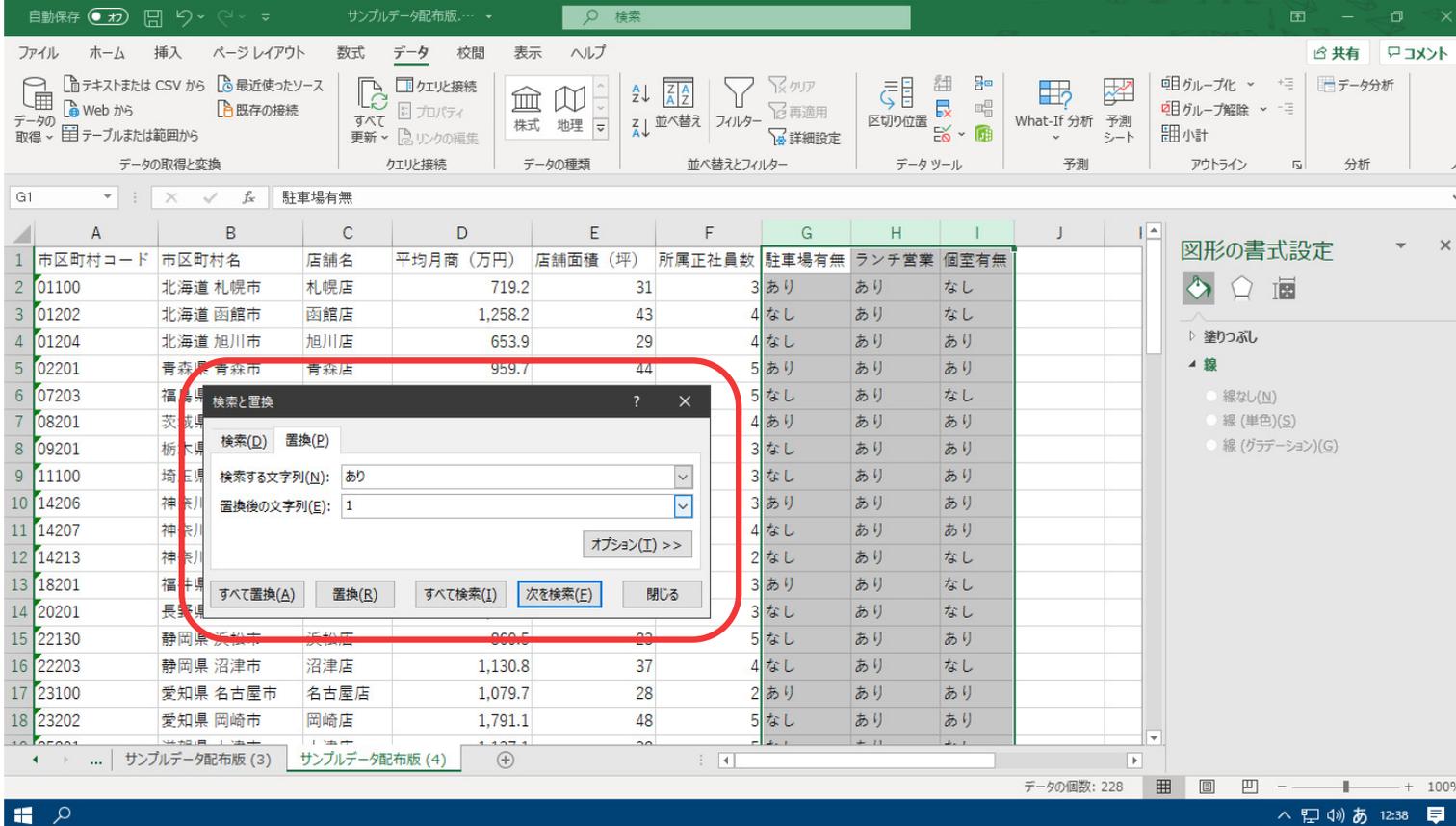
利用  
金額

● 男性  
● 女性



- 予め用いたい説明変数列を横並びに整形
- データメニューから「データ分析」（アドイン）
- 「回帰分析」を選択
- Yの方にアウトカム・Xの方に説明変数の範囲
- 項目名を含む選択なら「ラベル」に✓
- どこか空白の場所を「一覧の出力先」に
- 最後の表の「係数」の欄と「P-値」の欄に注目

# デモ⑥



市区町村コード	市区町村名	店舗名	平均月賃 (万円)	店舗面積 (坪)	所属正社員数	駐車場有無	ランチ営業	個室有無
01100	北海道 札幌市	札幌店	719.2	31	3	あり	あり	なし
01202	北海道 函館市	函館店	1,258.2	43	4	なし	あり	なし
01204	北海道 旭川市	旭川店	653.9	29	4	なし	あり	あり
02201	青森県 青森市	青森店	959.7	44	5	あり	あり	あり
07203	福島県				5	なし	あり	なし
08201	茨城県				4	あり	あり	あり
09201	栃木県				3	なし	あり	あり
11100	埼玉県				3	なし	あり	あり
14206	神奈川県				3	あり	あり	あり
14207	神奈川県				4	なし	あり	あり
14213	神奈川県				2	なし	あり	なし
18201	福井県				3	あり	あり	なし
20201	長野県				3	なし	あり	なし
22130	静岡県 浜松市	浜松店	868.6	23	5	なし	あり	あり
22203	静岡県 沼津市	沼津店	1,130.8	37	4	なし	あり	なし
23100	愛知県 名古屋市	名古屋店	1,079.7	28	2	あり	あり	あり
23202	愛知県 岡崎市	岡崎店	1,791.1	48	5	なし	あり	あり

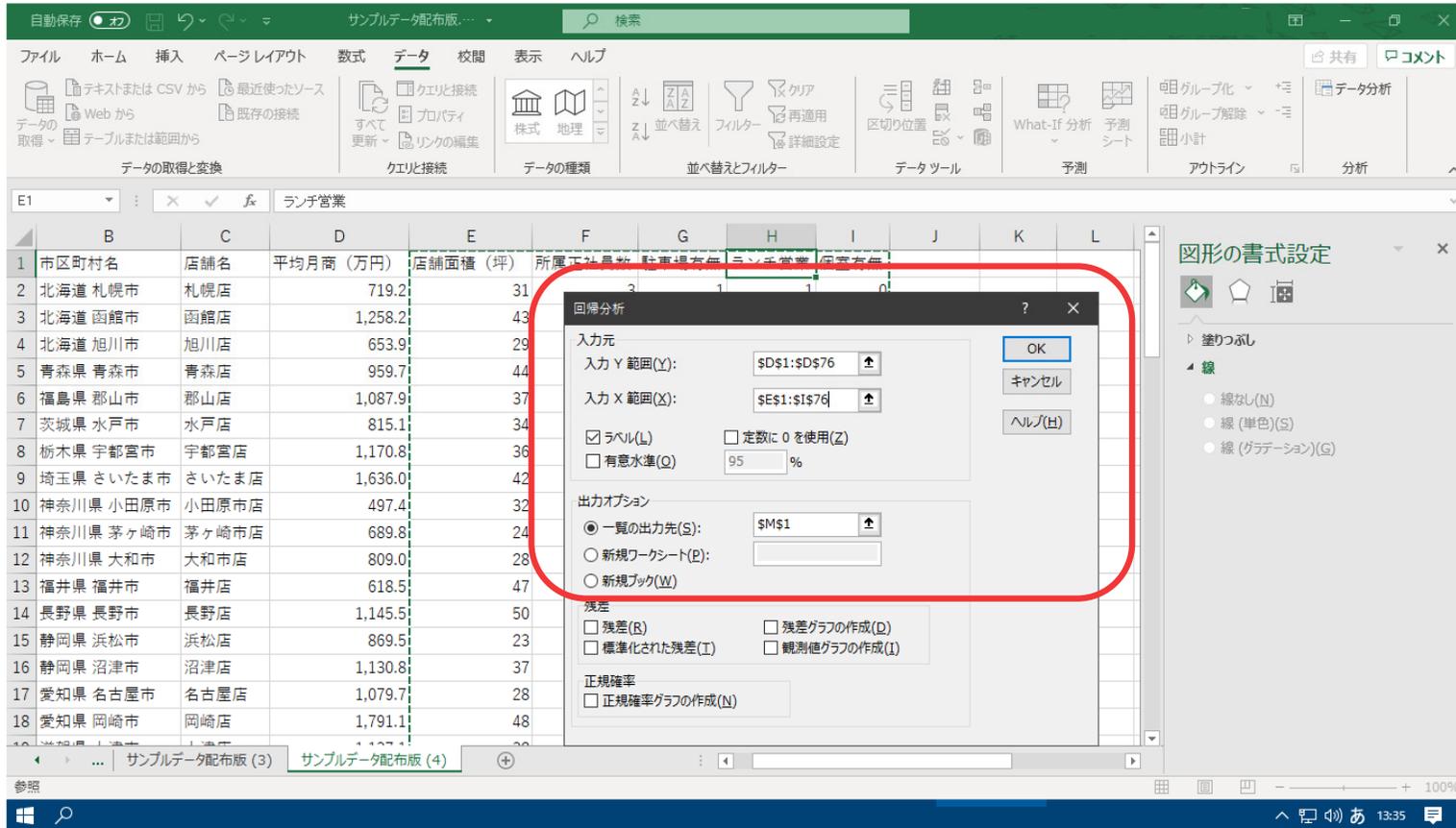
①検索と置換機能を使用して、「駐車場有無」列、「ランチ営業」列、「個室有無」列をダミー変数（あり = 1 / なし = 0）に置き換えます。（検索と置換機能を使用）

The screenshot shows the Microsoft Excel interface. The 'Data' ribbon is selected, and the 'Data Analysis' button is highlighted with a red box. The 'Data Analysis' dialog box is open, showing a list of analysis tools. The '回帰分析' (Regression Analysis) option is selected. The spreadsheet data is visible in the background, with columns for '市区町村コード', '市区町村名', '店舗名', '平均月商 (万円)', '店舗面積 (坪)', '所属正社員数', '駐車場有無', 'ランチ営業', and '個室有無'.

市区町村コード	市区町村名	店舗名	平均月商 (万円)	店舗面積 (坪)	所属正社員数	駐車場有無	ランチ営業	個室有無
01100	北海道 札幌市	札幌店	719.2	31	3	1	1	0
01202	北海道 函館市	函館店	1,258.2	43	4	0	1	0
01204	北海道 旭川市	旭川店	653.9	29	4	0	1	1
02201	青森県				5	1	1	1
07203	福島県				5	0	1	0
08201	茨城県				4	1	1	1
09201	栃木県				3	0	1	1
11100	埼玉県				3	0	1	1
14206	神奈川県				3	1	1	1
14207	神奈川県				4	0	1	1
14213	神奈川県				2	0	1	0
18201	福島県				3	1	1	0
20201	長野県 長野市	長野店	1,145.5	50	3	0	1	0
22130	静岡県 浜松市	浜松店	869.5	23	5	0	1	1
22203	静岡県 沼津市	沼津店	1,130.8	37	4	0	1	0
23100	愛知県 名古屋市	名古屋店	1,079.7	28	2	1	1	1
23202	愛知県 岡崎市	岡崎店	1,791.1	48	5	0	1	1

②データ > データ分析 > 回帰分析 を選択します。

# デモ⑥



自動保存 [オフ] サンプルデータ配布版... 検索

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 ヘルプ

データの取得と変換 クエリと接続 データの種類 並べ替えとフィルター データ ツール 予測

ランチ営業

	B	C	D	E	F	G	H	I	J	K	L
1	市区町村名	店舗名	平均月商 (万円)	店舗面積 (坪)	所属正社員数	駐車場有無	ランチ営業	個室有無			
2	北海道 札幌市	札幌店	719.2	31	3	1	1	0			
3	北海道 函館市	函館店	1,258.2	43							
4	北海道 旭川市	旭川店	653.9	29							
5	青森県 青森市	青森店	959.7	44							
6	福島県 郡山市	郡山店	1,087.9	37							
7	茨城県 水戸市	水戸店	815.1	34							
8	栃木県 宇都宮市	宇都宮店	1,170.8	36							
9	埼玉県 さいたま市	さいたま店	1,636.0	42							
10	神奈川県 小田原市	小田原市店	497.4	32							
11	神奈川県 茅ヶ崎市	茅ヶ崎市店	689.8	24							
12	神奈川県 大和市	大和市店	809.0	28							
13	福井県 福井市	福井店	618.5	47							
14	長野県 長野市	長野店	1,145.5	50							
15	静岡県 浜松市	浜松店	869.5	23							
16	静岡県 沼津市	沼津店	1,130.8	37							
17	愛知県 名古屋市	名古屋店	1,079.7	28							
18	愛知県 岡崎市	岡崎店	1,791.1	48							

図形の書式設定

塗りつぶし

線

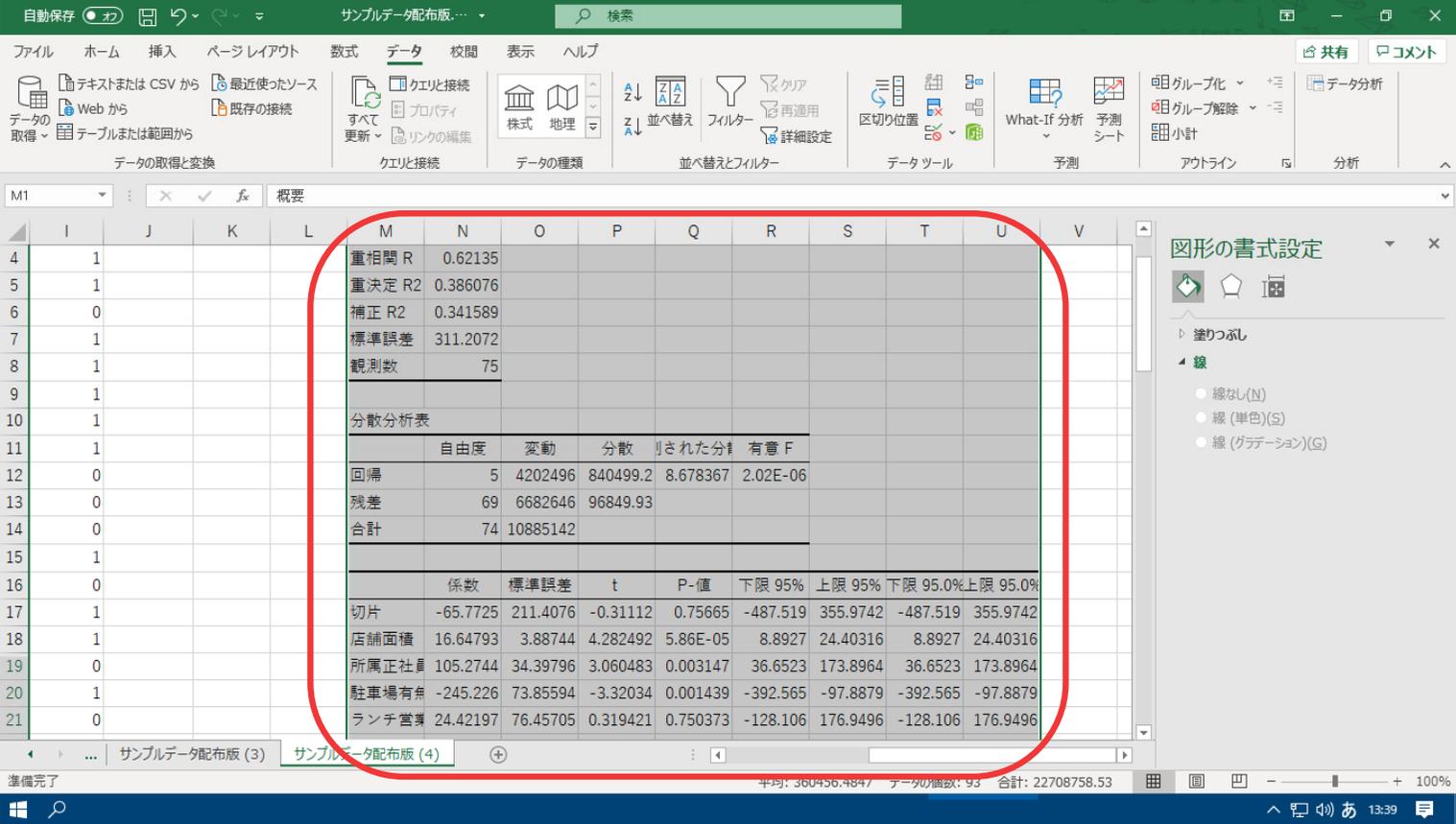
- 線なし(N)
- 線 (単色)(S)
- 線 (グラデーション)(G)

参照 100%

13:35

③Yには「平均月商」列、Xには「店舗面積」列、「所属正社員数」列、「駐車場有無」列、「ランチ営業」列「個室有無」列を指定します。出力オプションは任意の位置を指定します。

# デモ⑥



The screenshot shows the Excel interface with the following data tables highlighted by a red circle:

	M	N	O	P	Q	R	S	T	U	V
4	1	重相関 R	0.62135							
5	1	重決定 R2	0.386076							
6	0	補正 R2	0.341589							
7	1	標準誤差	311.2072							
8	1	観測数	75							
9	1									
10	1	分散分析表								
11	1		自由度	変動	分散	割された分散	有意 F			
12	0	回帰	5	4202496	840499.2	8.678367	2.02E-06			
13	0	残差	69	6682646	96849.93					
14	0	合計	74	10885142						
15	1									
16	0		係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
17	1	切片	-65.7725	211.4076	-0.31112	0.75665	-487.519	355.9742	-487.519	355.9742
18	1	店舗面積	16.64793	3.88744	4.282492	5.86E-05	8.8927	24.40316	8.8927	24.40316
19	0	所属正社員	105.2744	34.39796	3.060483	0.003147	36.6523	173.8964	36.6523	173.8964
20	1	駐車場有無	-245.226	73.85594	-3.32034	0.001439	-392.565	-97.8879	-392.565	-97.8879
21	0	ランチ営業	24.42197	76.45705	0.319421	0.750373	-128.106	176.9496	-128.106	176.9496

④回帰分析の結果が表示されます。

重相関 R	0.62135								
重決定 R <sup>2</sup>	0.386076								
補正 R <sup>2</sup>	0.341589								
標準誤差	311.2072								
観測数	75								
分散分析表									
	自由度	変動	分散	割された分散	有意 F				
回帰	5	4202496	840499.2	8.678367	2.02E-06				
		82646	96						
		85142							
	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%	
切片	-65.7725	211.4076	-0.31112	0.75665	-487.519	355.9742	-487.519	355.9742	
店舗面積	16.64793	3.88744	4.282492	5.86E-05	8.8927	24.40316	8.8927	24.40316	
所属正社員	105.2744	34.39796	3.060483	0.003147	36.6523	173.8964	36.6523	173.8964	
駐車場有無	-245.226	73.85594	-3.32034	0.001439	-392.565	-97.8879	-392.565	-97.8879	
ランチ営業	24.42197	76.45705	0.31942	0.750373	-128.106	176.9496	-128.106	176.9496	

数式で利用する切片・  
係数となります

それぞれのp値を確認し  
たまたまかの判定を行います

⑤係数と p 値はここに表示されています。

探索する要因(説明変数)

数値型  
(量的)

分類型  
(質的)

数値型  
(量的)

単回帰分析  
平均値とt検定  
単/重回帰分析でOK

分類型  
(質的)

割合と  
 $\chi^2$ 乗検定