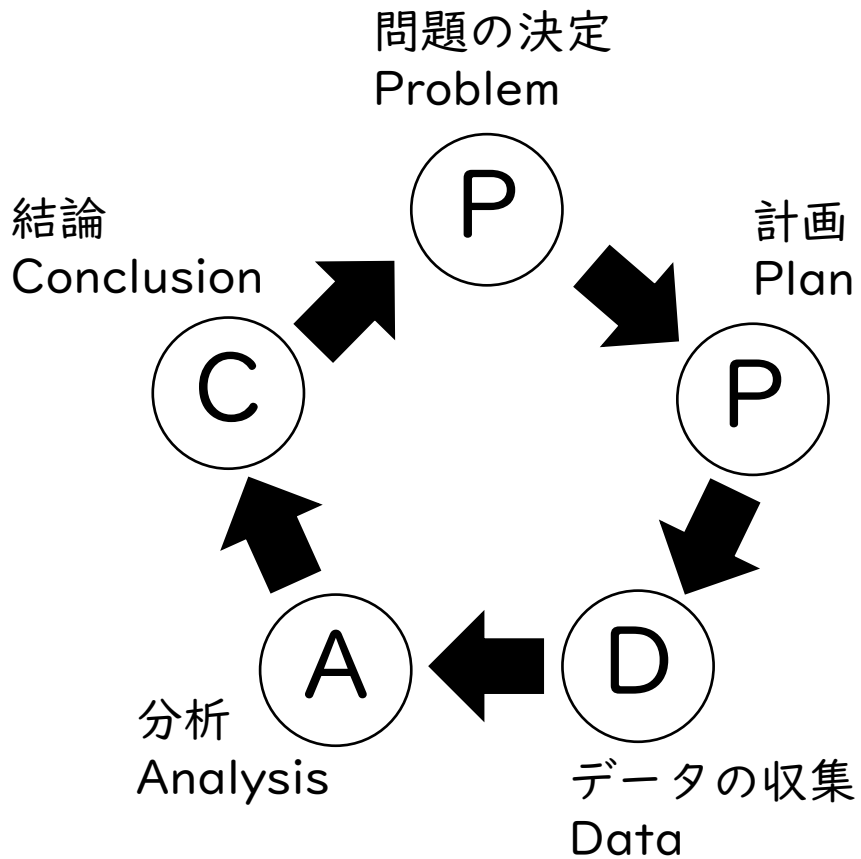


# データサイエンス7th

『アンケート調査』  
～データの収集と整理～

# アンケート調査はPPDACをまわすべし



- P : 分析の目的は何なのかを明確にする
- P : 調査対象を決定する
- D : 調査データを収集する
- A : 調査データを分析する
- C : 調査データの分析から解釈・活用へ

実践事例：「アンケートで身近な問題を解決しよう！」

<https://www.youtube.com/watch?v=VFzCUPp30jg>

# P：分析の目的は何なのかを明確にする

例えば、「高校生の好きな色を調査しなさい」と言われたら、どうしますか？

○アンケートを実施して、その結果を統計処理ソフトウェアで表やグラフにしたり、簡単な統計処理をすればいい

●どんな調査データをとればいいのか？

対象はどうするのか？**全数調査**？**サンプリング**(**標本調査**)でランダムにできる？

●調査データをもとにどのようなことを示すのか？

●指標をもとにどのような結論を述べるのか？ . . . .

アンケートする際には、具体化する

●高校生は、男女、年齢、地域で分けるべきなのか？**バイアス**がかからないように

●好きな色とは、何をさしているのか？服、靴、携帯、小物etc

※何のためにデータをとっているのか、目的を明確にすることが重要

# P : 調査対象を決定する

例えば、「高校生の好きな色を調査しなさい」と言われてアンケートを実施する。  
どのような工夫がいる？

○調査対象となる**母集団**を決定する

●**全数調査**を実施できる？

全国の高校生（全日制・定時制・通信制）の生徒数は、約319万人(R5年度)。  
具体的には、全日制約291万人、定時制約7万9千人、通信制約20万人

◎**サンプリング**(**標本調査**)を実施する

**バイアス**がかからないような調査データがのぞましい

考慮すること：いつ？どこで？誰に？（男女、年齢）どのように？ etc

※サンプリングの方法には工夫が必要

# D：調査データを収集する

調査データは、嘘の混入やノイズ(誤った情報)、データの偏りが生じる。

○どこでアンケートをとる？

学校、SNS上、街頭 etc

○アンケート項目をどのように工夫するか？

例えば「好きな色は何ですか？」と聞かれたときに、それぞれイメージする対象物が異なる

※調査対象が人間である場合、ノイズや偏りを完全になくすことは不可能

◎収集したデータを分析する前に整理することが必要

# 「データを整理する」 どのようなことをするのか？

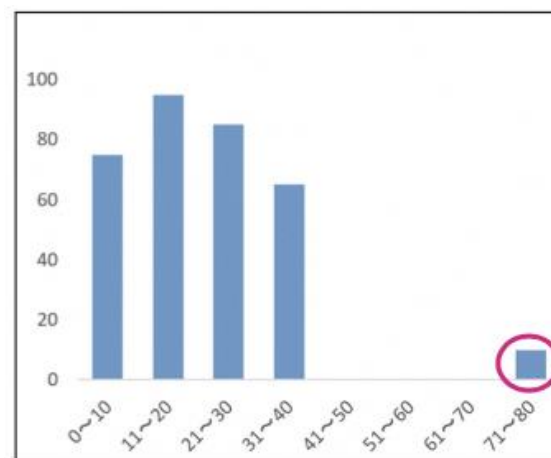
## (1) コーディング

- ・情報を記号化すること。

例：収集したデータが「はい」、「いいえ」の文字データであれば、「はい」を1、「いいえ」を0に対応させる など

## (2) 外れ値

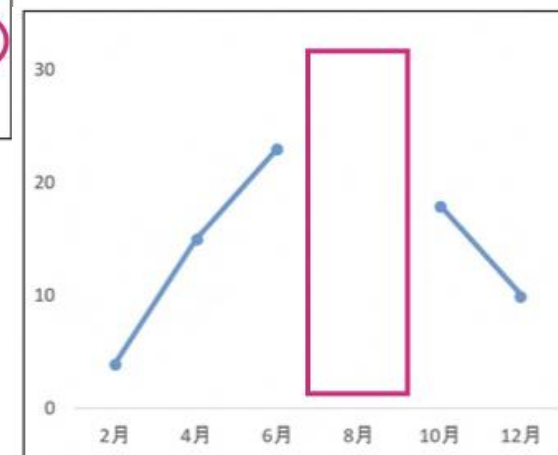
- ・ほかの多数のデータから大きく離れた値。



● 図1 外れ値

## (3) 欠損値

- ・存在するはずだった値。
- ・データの値が不明（空欄）であったり、収集されていなかったりする場合に生じる。



● 図2 欠損値

何故、整理することが必要か？

→

データの種類によって統計手法が異なるから

データにはどのような種類があるのか？

# 量的データ

量的データは数量的な意味があり計算可能

購入パン	価格	個数	合計金額
メロンパン	220	3	660

価格 × 個数 = 合計金額

$$220 \times 3 = 660$$

# 質的データ

---

質的データは分類としての意味をもつ  
計算はできない

購入パン	価格	個数	合計金額
メロンパン	220	3	660

# データにはどのような種類があるのか？

データ種類	尺度	性質	可能な計算	例
量的データ	比例尺度	大小関係、差、比率にも意味がある	+、-、×、÷	大きさ、価格、身長、体重、売上、速度、年齢
	間隔尺度	目盛が等間隔になっているもの	+、-	温度（℃）、西暦
質的データ	順序尺度	順序には意味があるが間隔には意味がないもの	大小関係の比較	満足度、成績の段階評価、ランキング
	名義尺度	他と区別し分類するためのもの		名前、国籍、血液型、電話番号、メーカー名、学籍番号

## <例題：尺度水準>

次の各項目は，どの尺度水準に該当するか考えなさい。

- ①アンケートの5段階評価
- ②郵便番号
- ③体温
- ④価格

## <例題：尺度水準>

次の各項目は，どの尺度水準に該当するか考えなさい。

### 【解答】

- ①アンケートの5段階評価 → 順序尺度
- ②郵便番号 → 名義尺度
- ③体温 → 間隔尺度
- ④価格 → 比例尺度



## 振り返り

1. データには大きく分けて、**質的データ**と**量的データ**の区別がある
2. **比例尺度**は大小関係、差、比率にも意味がある
3. **間隔尺度**は目盛が等間隔になっているもの
4. **順序尺度**は順序には意味があるが間隔には意味がないもの
5. **名義尺度**は他と区別し分類するためのもの

データの尺度ワークショップ

スーパーのデータを収集しよう！



Life is Tech!

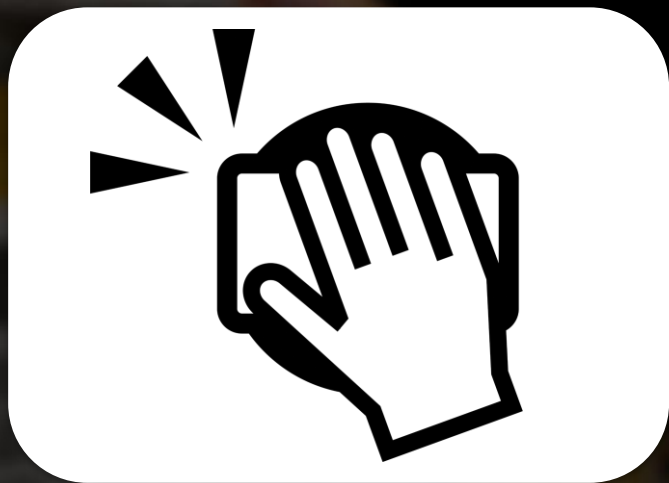
*Lesson*

# Amazon Go



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

# Amazon Go

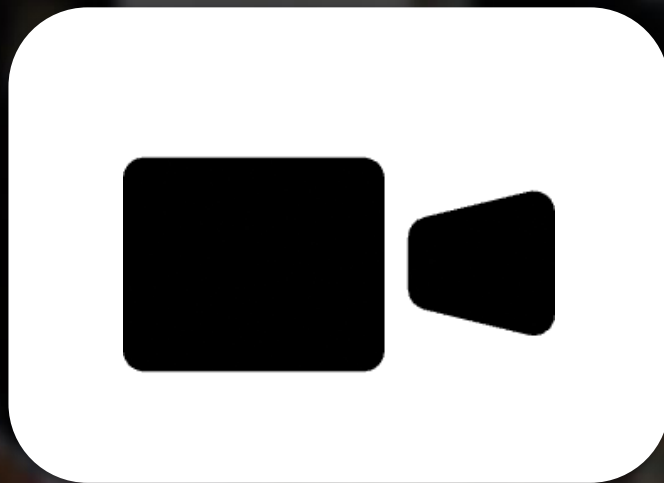


## コード読み取り機器

入店するには、まずスマホをかざして  
ゲートにある機器で自分のQRコードを  
スキャン



顧客データ (住所、年齢、等)

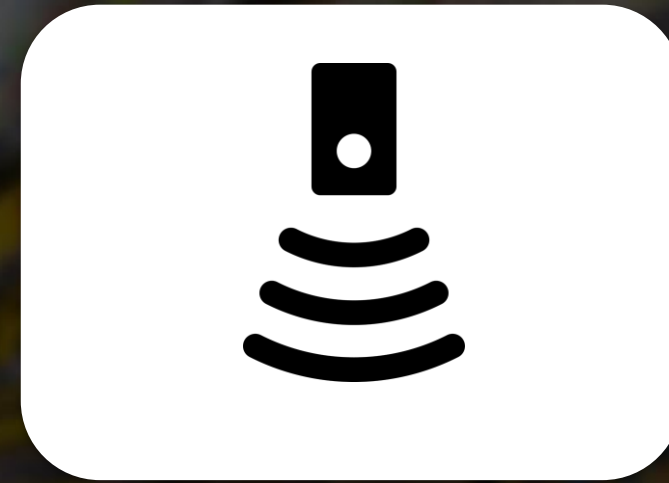


## カメラとマイク

人の動きを追跡



顧客の店内移動履歴データ



## 赤外線、圧力、重量センサー

商品の数や移動などを詳細に追跡



顧客が棚に戻した商品データ

The background of the image is a blurred, dark-toned photograph of a grocery store aisle. The shelves on both sides are filled with various products, likely canned goods or packaged items, with some colorful labels visible. The perspective is looking down the aisle, creating a sense of depth. The overall lighting is dim, and the image has a soft, out-of-focus quality.

# Mission



# 尺度を当てはめていこう！

データ種類	尺度	性質	可能な計算	例
量的データ	比例尺度	大小関係、差、比率にも意味がある	+、-、×、÷	大きさ、価格、身長、体重、売上、速度、年齢
	間隔尺度	目盛が等間隔になっているもの	+、-	温度（℃）、西暦
質的データ	順序尺度	順序には意味があるが間隔には意味がないもの	大小関係の比較	満足度、成績の段階評価、ランキング
	名義尺度	他と区別し分類するためのもの		名前、国籍、血液型、電話番号、メーカー名、学籍番号



共有タイム

# 共有タイム

データ種類	尺度	性質	可能な計算	例
量的データ	比例尺度	大小関係、差、比率にも意味がある	+、-、×、÷	大きさ、価格、身長、体重、売上、速度、年齢
	間隔尺度	目盛が等間隔になっているもの	+、-	温度（℃）、西暦
質的データ	順序尺度	順序には意味があるが間隔には意味がないもの	大小関係の比較	満足度、成績の段階評価、ランキング
	名義尺度	他と区別し分類するためのもの		名前、国籍、血液型、電話番号、メーカー名、学籍番号

# 共有タイム

データ種類	尺度	性質	可能な計算	例
量的データ	比例尺度	大小関係、差、比率にも意味がある	+、-、×、÷	大きさ、価格、身長、体重、売上、速度、年齢
	間隔尺度	目盛が等間隔になっているもの	+、-	温度（℃）、西暦
質的データ	順序尺度	順序には意味があるが間隔には意味がないもの	大小関係の比較	満足度、成績の段階評価、ランキング
	名義尺度	他と区別し分類するためのもの		名前、国籍、血液型、電話番号、メーカー名、学籍番号

# 共有タイム

データ種類	尺度	性質	可能な計算	例
量的データ	比例尺度	大小関係、差、比率にも意味がある	+、-、×、÷	大きさ、価格、身長、体重、売上、速度、年齢
	間隔尺度	目盛が等間隔になっているもの	+、-	温度（℃）、西暦
質的データ	順序尺度	順序には意味があるが間隔には意味がないもの	大小関係の比較	満足度、成績の段階評価、ランキング
	名義尺度	他と区別し分類するためのもの		名前、国籍、血液型、電話番号、メーカー名、学籍番号

# 共有タイム

データ種類	尺度	性質	可能な計算	例
量的データ	比例尺度	大小関係、差、比率にも意味がある	+、-、×、÷	大きさ、価格、身長、体重、売上、速度、年齢
	間隔尺度	目盛が等間隔になっているもの	+、-	温度（℃）、西暦
質的データ	順序尺度	順序には意味があるが間隔には意味がないもの	大小関係の比較	満足度、成績の段階評価、ランキング
	名義尺度	他と区別し分類するためのもの		名前、国籍、血液型、電話番号、メーカー名、学籍番号

# どの尺度に入るのかわからなかったデータ

データ種類	尺度	性質	可能な計算	例
量的データ	比例尺度	大小関係、差、比率にも意味がある	+、-、×、÷	大きさ、価格、身長、体重、売上、速度、年齢
	間隔尺度	目盛が等間隔になっているもの	+、-	温度（℃）、西暦
質的データ	順序尺度	順序には意味があるが間隔には意味がないもの	大小関係の比較	満足度、成績の段階評価、ランキング
	名義尺度	他と区別し分類するためのもの		名前、国籍、血液型、電話番号、メーカー名、学籍番号

# データ例

データ	獲得方法	データ種類	尺度
来客者の思い入れのある 都道府県ランキング	アンケート	質的 量的	順序 尺度
来客者の身につけているブランド	カメラでの画像認識	質的 量的	名義 尺度
いらっしゃいませと店員が 言った回数	防犯カメラのマイク	質的 量的	比例 尺度
クーポンが使われた数	POS	質的 量的	比例 尺度
セルフレジ使用率	POS	質的 量的	比例 尺度
カードを利用する人の割合	POS	質的 量的	比例 尺度
来客者の年齢	ポイントカード	質的 量的	比例 尺度
マイバック持参している人の割合	POS	質的 量的	比例 尺度
アルコール消毒をする人の割合	アルコールディスペンサー	質的 量的	比例 尺度