

チャンネル登録をお願いします！

第14講 データの活用とデータベース

教材のダウンロードは以下URLにアクセスし、印刷して受講すると効果的です。

赤字を書き込んでいくとより効果的です！！！！

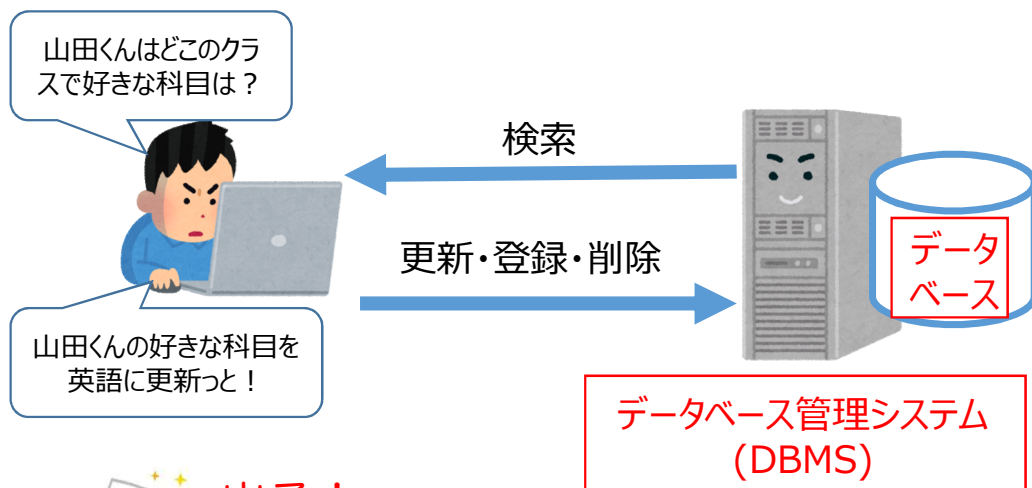
<https://joho-tarou.com/>



JOHO 太郎「情報 I」スーパー講義

14-1 データの活用とデータベース

情報システムでは、データベースを使ってデータを管理する！



学生テーブル

学生番号	氏名	クラス	好きな科目
1	山田 一郎	1-1	社会
2	鈴木 次郎	1-2	国語
3	田中 三郎	1-1	理科
4	山本 史郎	1-3	英語
5	上杉 五郎	1-2	数学
...

テスト結果テーブル

学生番号	テスト	科目	点数
1	中間テスト	国語	100
1	中間テスト	数学	90
1	中間テスト	英語	80
2	中間テスト	国語	30
2	中間テスト	数学	40
2	中間テスト	英語	50
...

関係データベース(RDB)

表を使ってデータを管理する方式



出る！

データベースといえば「関係データベース」を指すと言ってよく、ほぼこれしか問われない。



ビッグデータ

・・・ コンピュータの発達によって、大量なデータを蓄積し、分析できるようになった。

構造化データ

・・・ 構造が定義されている、表形式で定義されたデータを指す。

非構造化データ

・・・ 構造が定義されていない、内容に規則性のない、画像、動画、本の文章、センサの記録等を指す。

14-2-1 データの管理 (関係データベースの関係論理演算)



部門表と結合して、東京にいる、従業員コード・従業員名・場所の表を抽出してください。

結合演算

選択演算

射影演算

従業員コード	従業員名	場所
1000	山田 太郎	東京
2000	田中 花子	東京

SQL

= データを扱うための言語

```
SELECT 従業員コード, 従業員名, 場所
FROM 従業員表, 部門表
WHERE 従業員表.部門 = 部門表.部門コード
AND 従業員表.部門 = 10
```

POINT



- ✓ 関係データベースは表形式でデータを管理します
- ✓ 表の事を「表」もしくは「テーブル」と呼びます
- ✓ データベースの表に対し、抽出、削除、更新、作成を行う命令をSQL (エスキューエル、シークエル)と呼びます

14-2-2 データの管理（関係データベースにおける正規化）

正規化

… データの重複をなくし整合的にデータを取り扱えるように表(テーブル)を設計すること

「従業員表」を設計

従業員コード	従業員名	手当	部門コード	部門
1000	山田 太郎	1000	10	東京
2000	田中 花子	500	10	東京
3000	鈴木 一郎		20	大阪
4000	佐藤 二郎	3000	20	大阪

正規化

従業員表

従業員コード	従業員名	手当	部門
1000	山田 太郎	1000	10
2000	田中 花子	500	10
3000	鈴木 一郎		20
4000	佐藤 二郎		20

部門表

部門コード	場所
10	東京
20	大阪



- ✓ データの追加・更新・削除などに伴うデータの不整合や喪失が起きるのを防ぐ
- ✓ メンテナンスの効率を高められる

14-3 データの収集と種類

データベースを作るためのデータ収集方法を整理しよう！

アンケート		尺度	データの種類
Q1 通学方法を教えてください 1 : 徒歩、2 : 電車、3 : バス、4 : その他	名義尺度	区別のために用いられる番号	質的データ 文字のデータ
Q2 次の科目に好きな順に順位を記入ください。 国語 ()位、数学 ()位、英語 ()位	順序尺度	数値の大小、順序は決まっている	
Q3 本日の体温を教えてください ___度	間隔尺度	数値の間隔に意味がある	量的データ 数値のデータ
Q4 通学時間を教えてください ___分	比例尺度	数値の差のみならず、 比にも意味がある (15分と30分では2倍の差がある)	

データ分析	意味	中学生小遣い	分析時の注意点	
代表値	平均値	データの合計÷データの数	2,536円	欠損値 無回答等で得られない
	中央値	大きさの順で並べて真ん中の値	2,000円	外れ値 他と比べて大きな(小さな)値
	最頻値	最も多く出てくる値	1,000円	POINT 代表値に影響を与えるため、 明らかな誤りであれば、取り除く等を考える

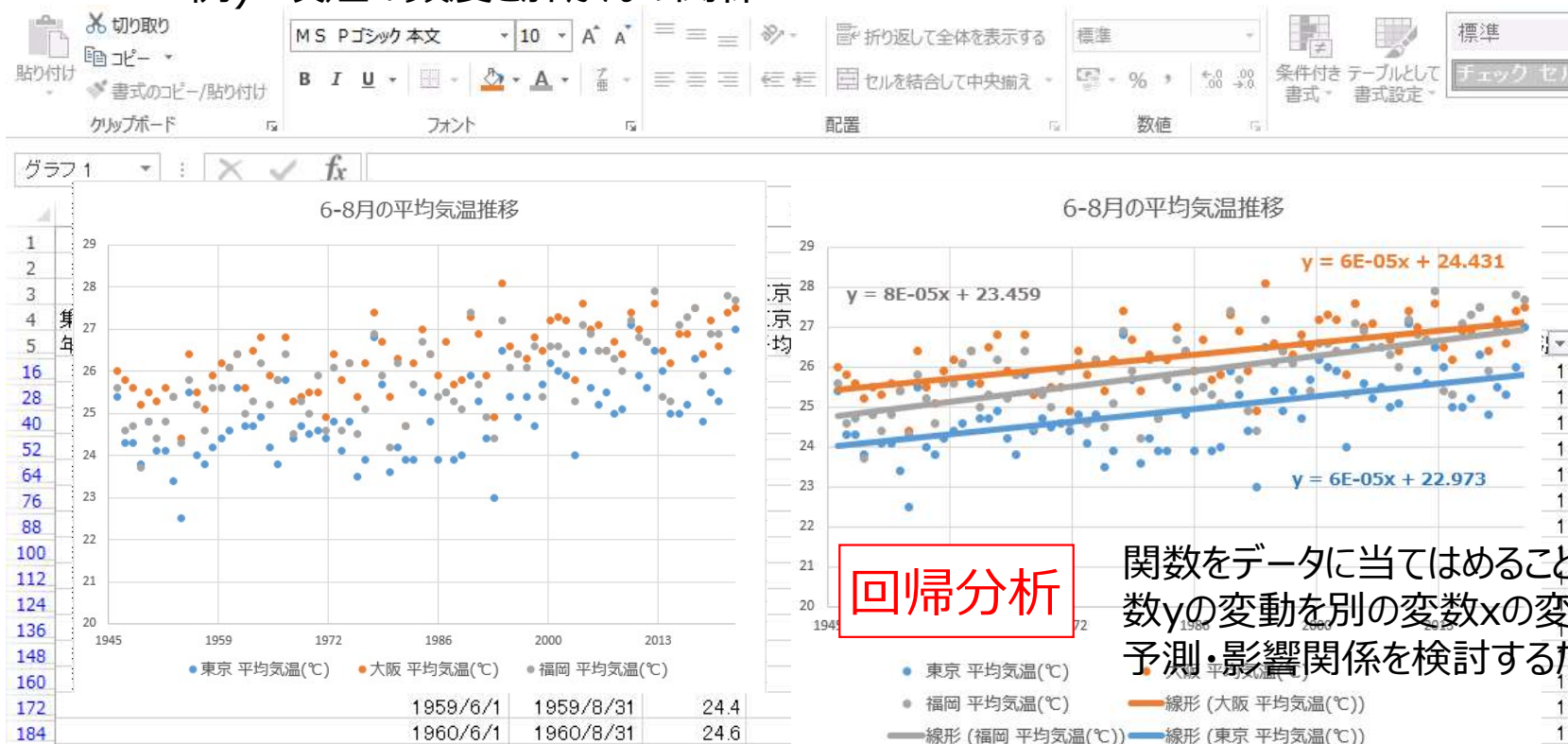
14-4 2つのデータの関係

相関関係

2つの事柄の間で、一方が増加すると他方が増加または減少する関係
例) アイスクリームの売上と溺死事故の件数

因果関係

一方が原因となり他方が結果となる関係
例) 喫煙の頻度と肺がんの関係



第14講 確認問題 問1

関係データベースにおける射影演算はどれか。

- ① 2つの表から列の値で関連付けした上で一つの表として抜き出す
- ② 表から特定の列を表として抜き出す
- ③ 表から特定の条件に当てはまるレコードを表として抜き出す
- ④ 同じ構成である2つの表について同じレコードがあれば、それを表として抜き出す

正解は②となります。

①：結合演算の説明です。

③：選択演算の説明です。

④：集合演算の共通部分の選択の説明です。

第14講 確認問題 問2

以下の発注データを発注に関する情報と商品に関する情報に分割して、正規化を行った場合に、結果の表の組み合わせで適切なものを選びなさい。ただし、単価は商品番号に対して一つで決まるものとする。

発注番号	発注者	商品番号	商品名	個数	単価
H001	山田 太郎	1000	りんご	3	100
H002	田中 一郎	2000	みかん	2	50
H003	鈴木 二郎	1000	りんご	4	100
H004	山本 花子	3000	トマト	1	110

①

発注番号	発注者		
商品番号	商品名	個数	単価

②

発注番号	発注者	商品番号	
商品番号	商品名	個数	単価

③

発注番号	発注者	商品番号	個数
商品番号	商品名	単価	

④

発注番号	発注者	商品番号	個数	単価
商品番号	商品名			

正解は③となります。商品番号を結合させる項目として分割し、商品に紐づく、商品名・単価が重複することなく、分割することを考えます。

①は発注と商品を「商品番号」で結合させられず、元の表を作ることができません。②は商品に「個数」が入っているため、商品データの重複が改善されません。④は発注側に「単価」が入ることで、「単価」が重複されたままとなります。

第14講 確認問題 問3

次のデータの平均値と中央値の組合せはどれか。

[データ]

10, 20, 30, 30, 40, 40, 70, 200, 200, 300

	平均値	中央値
①	95.7	40
②	95.7	35
③	94	35
④	94	40

正解は④となります。

平均値： $(10 + 20 + 30 + 30 + 40 + 40 + 70 + 200 + 200 + 300) / 10 = 94$

中央値：データを値の大小で並べた時に、中央に位置する値です。

※データ個数が偶数の時は中央の値が2つになる。この時は、2つの平均値を取る！

10, 20, 30, 30, 40, 40, 70, 200, 200, 300

→ $(40 + 40) / 2 = 40$

第14講が完了しました！

・ 修了 ・
おめでとう！

チャンネル登録をお願いします！