

データサイエンス講座
第3・4回
～データの捉え方～

これまでのデータサイエンス講座では・・・

▶ 第1回「データの見方」

- データはランダムに集めることが大切
⇒ データはばらつくものだから

▶ 第2回「データビジュアライゼーション（視覚化）」

- データのどこに着目しているか、どのような特徴があるかを読み手に伝わるように視覚化する

第3回・第4回では・・・

「データの捉え方」を考える。

探究を行っていく上で根拠（エビデンス）を集め、
それをもとに考察し結論を導いていく過程は欠かせないものである。

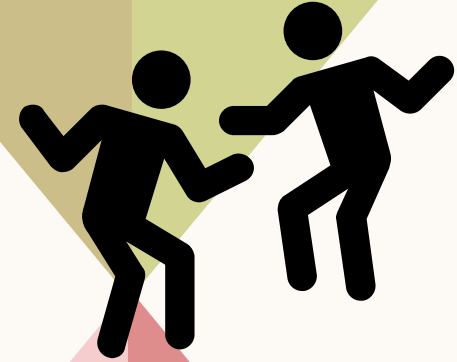
では、これから内容に入る前に・・・



【一人でじっくり考える時間】
⇒自分の考えをワークシートにしっかり書くこと。



【周りの人と共有する時間】
⇒自分と違う考えを聴いて、書き取っておくこと。
LearnWizに意見を書き込もう！また、書き込まれた意見に共感できたら「いいね」を押しておこう！



【ディスカッションする時間】
⇒積極的に参加して、意見交換を行うこと。

【アクティビティ①】（資料 1）

ISSチャレンジの研究から「データの捉え方」を考えてみよう！

調理法	生	蒸し	ゆで	ゆで汁	冷凍
タマネギ皮あり	59.9	86.4	71.7	67.4	
タマネギ皮なし	78.6	81.8	56.7	45.5	66.9
ナス皮あり	62.7	89.2	82.9	77.9	
ナス皮なし	91.6	85.3	73.0		54.4
レンコン皮あり	65.4	59.3	47.2	47.7	84.7
レンコン皮なし	17.4	61.6	28.4	78.9	
レモン皮あり	91.1	58.5	81.8	69.0	
レモン皮なし	73.8	74.7	79.5		
ショウガ皮あり	90.7	82.9	79.9	62.9	
ショウガ皮なし	85.0	73.8	83.4		



**「抗酸化力に変化があった」
と考察をしていいのだろうか？**

2020年度ISSチャレンジより「食品の調理法ごとに抗酸化力の測定」

「資料 1」の研究データから

周りの人と情報交換をして、自分では見つけられなかった内容をワークに記入しておきましょう。

右下のQRコードを読み取って、自分の意見を書き込もう！書き込まれた意見に共感できたら、「いいね」や返信コメントで反応しておこう！



【アクティビティ②】（資料２） 実験結果のデータから、自分の意見を決めよう！

A.

「関係している」
と**言える**

B.

「関係している」
とは**言えない**

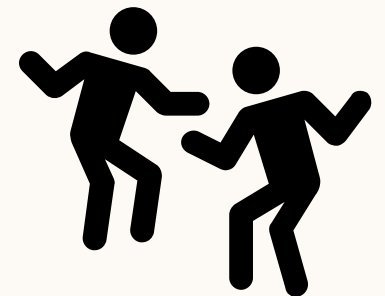
C.

判断できない



それぞれの立場から、ディスカッションを行います。

- ▶ **まず始めに、同じ考えを持つ人たちとグループを作ります。**
 - 「関連している」「関連していない」「判断できない」 のいずれかに自分を考えを決める。
 - 教室担当の先生の指示に従って、まとまって座るようにしてください。
- ▶ **グループの中である程度意見をまとめましょう。**
 - 同じ意見もあるだろうし、違う見方もあると思うので、まずはグループ内で情報共有を。
 - 発言をする代表者を決めてもいいOK。もちろん、発言する人は何人いても大丈夫。
- ▶ **教室担当の先生は、このディスカッションの「ファシリテーター」です。**
- ▶ **ディスカッションできる時間は、約 10 分です。**



ディスカッションを通して・・・

自分の考えがどう変化したか、ワークシートに記入をしましょう。

- ★考えが変わった人は、「**どうして**」変わったのか。
- ★自分の意見に確信を持てた人は、「**どこに**」確信を持つことができたのか。
- ★迷い込んだしまった人は、「**なぜ**」迷い込んでしまったのか

必ず、自分の意見に対して根拠（エビデンス）を入れて文章にしておきましょう。



【ディスカッションの内容を共有しよう！】



右下のQRコードを読み取って、自分の意見を書き込もう！
書き込まれた意見に共感できたら、「いいね」や返信コメントで反応しておこう！

【本日のまとめ】

資料 1 , 資料 2 とともに...

実験データ（アンケート調査結果）

調査したサンプルの結果でしかなく、偶然得られた結果の可能性もある。
調査の**狙いが達成できたかどうかはわからない。**

結局

狙いが対象としている**すべてを調べる**

毎回、対象するすべて（全員）を調べるのは不可能に近い！！

**実験の結果や、アンケートの狙いが
達成できたかどうかを定量的に評価
する方法**

あります！！



・・・ということで、次回、7実験の結果やアンケートの狙いが達成できたかどうかを検証する方法とは何か??

本日の講座はこれで以上です。

Teamsに投稿されたFormsにて、本日の講座についてのアンケートをお願いします。



データサイエンス講座
第3・4回
～データの捉え方～

【前回の続き・・・】

「C：判断できない」を選択した人たちの理由

学力の差が考慮されていない

睡眠時間以外の外的要因
がある可能性

10点満点のテストでは
問題は少ない

集中力は数値化できるのか？

これらの内容に関しては、

研究デザインを考える

ときに、よく考えて欲しい内容で、とってもとっても大事なこと。

今回は・・・

【全体を調査することが不可能で、サンプル数が少ないときに、得られたデータをどのように評価するか】

【アクティビティ①】

17

「影響があった」・・・と判断したいとき、何を基準にしますか？

40人というサンプル数は、十分であると判断したとき・・・
得られたデータから「影響があった」という判断をしたい。

周りの人たちと共有すると同時に、LearnWizにも書き込み
をして色々な人たちの意見も読んで見よう！



⇒ 人によって基準は異なるのは当たり前・・・

【アクティビティ②】

18

「睡眠時間と集中力に**関係がない**」
としたら？？



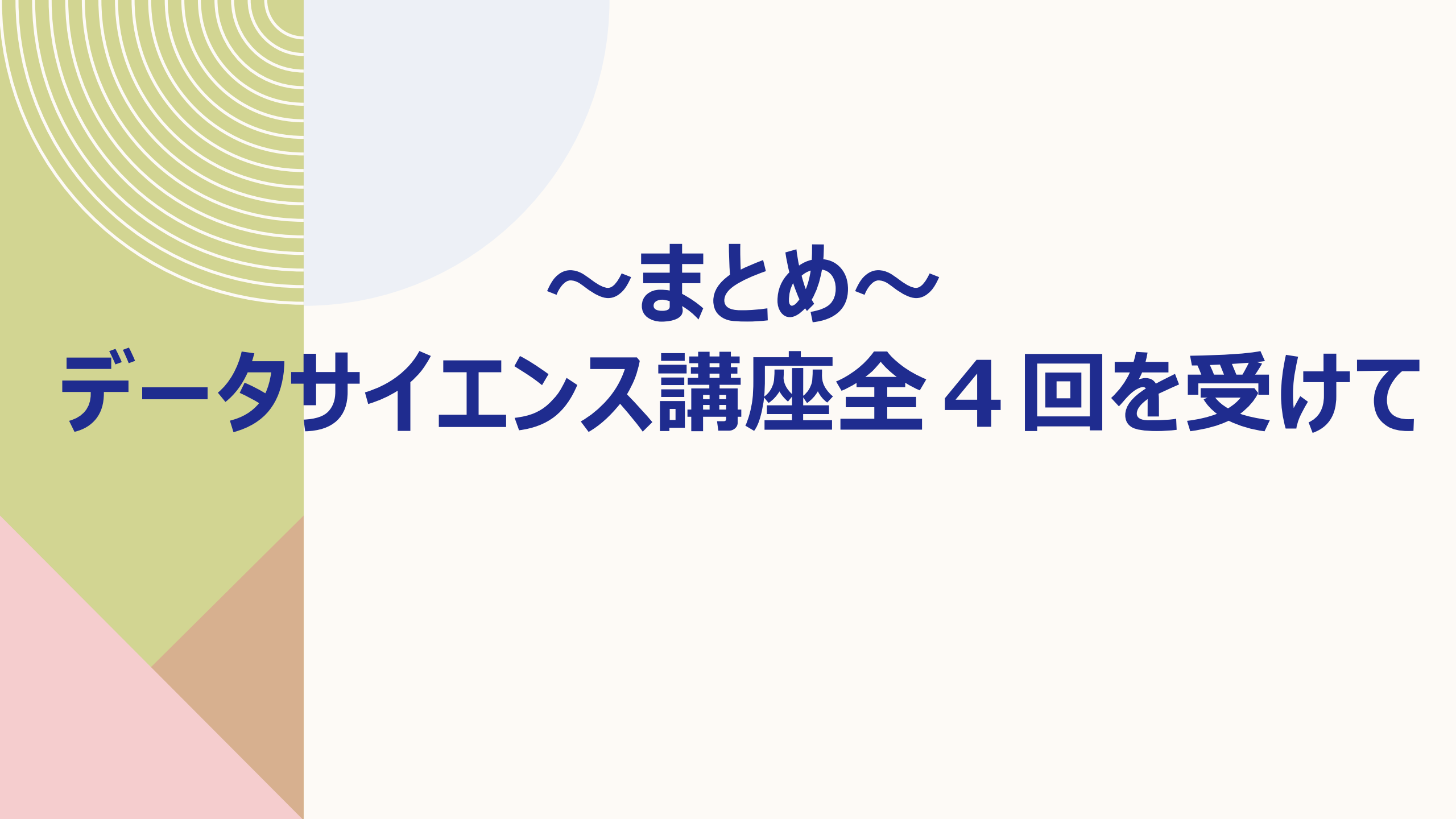
どんな結果が得られることが「関係ない」という
判断に繋がるのか考えてみよう！

まとめると…

少ないデータで評価するとき

『検定』という。

- ▶ 先に基準を決めて（どこまでが偶然起きたこととするか）
 - ▶ 決めた基準に対して『起こらない』ことを考える
⇒『起こらない』ことが**起こりにくい**と判断されたとき、
差があった（影響があった）という判断がされる。
- ➡ 「起こる起こらない」は、『確率』などを用いて考えていく



～まとめ～

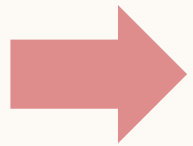
データサイエンス講座全 4 回を受けて

【今回の架空の実験データ】にあてはめると・・・

「睡眠時間は集中力には**影響がなかった**」と仮定して、検定を行うことにした。

その判断基準を作り、得られたデータを当てはめると・・・

『**かなり起こりにくい**』という**判断結果**であることがわかった。



この実験条件のもとでは

「睡眠時間は集中力に影響があった」

独立 2 標本 t 検定

なので...

実験データを見る前に



ここなら偶然ではないという基準

ということを先に考えるが必要になってくる。