

統計学

4月13日（火）
第1回

兵庫県立大学
情報科学研究科 / 社会情報科学部

山本 岳洋

t.yamamoto@sis.u-hyogo.ac.jp

- 山本の自己紹介
- 講義「統計学」について
 - ー 成績評価, 進め方, スケジュールなど
- 簡単なアンケート
- 統計学をなぜ学ぶか・何を学ぶか
- 今週の課題の説明
- Teams使います.

temtbw

自己紹介

● 山本 岳洋（やまもと たけひろ）

- － 社会情報科学部 准教授
- － 1984年生まれ，広島県出身.



● 専門分野

- － 情報検索
- － ヒューマンコンピュータインタラクション

こんなことをやっています

新しい情報アクセス技術



モバイル協調検索技術

smoking cancer risk	
diseases caused by smoking	
cigarettes price increase	
smoking benefits	
smoking ruins your looks	

Much info is still unexplored.
I have to keep on searching.



「多様な情報閲覧」を支援するインタフェース



視線情報に基づく検索意図理解

こんなことやってます

企業との共同研究

YAHOO! ラボ (Rerank)

検索クエリ: 豚肉 ビーマン 件数: 200

「検索結果中のキーワードを選択し、または画像をクリックすることにより、検索結果を並び替えることができます。」

青梗肉餅を強調しました 豚肉 ビーマン で検索した結果 1 - 200 件目 / 約3,200,000件

1位 ~ 100位

1. もやしとビーマンの豚肉炒め(青梗肉餅風) - レシピ情報 - Yahoo!グルメ

もやしとビーマンの豚肉炒め(青梗肉餅風) 240人が登録... 豚肉 もやし 簡単 ビーマン おつまみ ヘルシー チンジャオ... 2. フライパンにゴマ油をひきビーマンを軽く炒め、フライパンの端に寄せて下味を付けた豚肉を入れほぐしながら炒める...

http://recipe.gourmet.yahoo.co.jp/16571231516497/ 元9位

200g cm幅 キャベツ グルメ コショウ ゴマ
サラダ タケノコ ニンニク フライパン



INTERNET Watch

記事検索

ヤフー、実験的サービスの公開サイト

ヤフーは10日、Yahoo! JAPANの実験的なプロダクトを公開する「Yaf

Yahoo!ラボは、Yahoo! JAPANのエンジニアが主導となって、新規性の紹介するサービス。各プロダクトには、ユーザーが5段階の評価やコメントバックをもとに正式サービスへの可能性を探っていく。

● 類似画像検索や、検索結果の再ランク機能など公開

Yahoo!ラボ開設時点で公開されているプロダクトは、選択した画像から「VisualSeeker」、検索キーワードを都道府県別・性別・年代別などの割りのなかみ」、キーワード検索結果の結果から気になる単語を選択すること「Rerank.jp」など。また、ヤフーの地図実験サイト「Lat Long Lab」のる。

「VisualSeeker」は、あらかじめ登録されている約100万点の画像からの画像を検索できる技術。ランダムに表示されている画像から1点を選ぶと、次々に表示されていくモードのほか、ユーザーが描いた絵やアップロード検索する機能も備える。

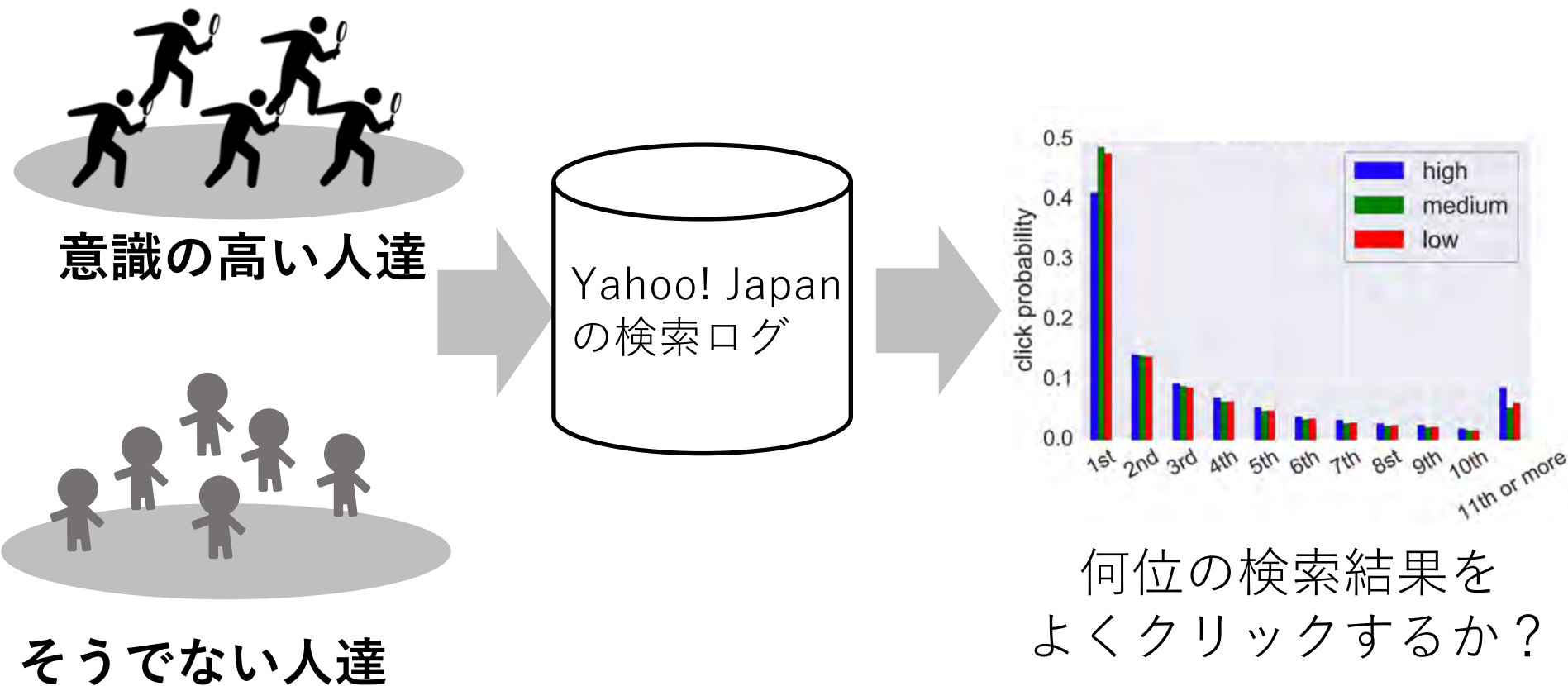
また、色の要素を無視して形状だけで類似画像を検索する機能や、複数候補から検索機能も備える。

ファミ通
EYE・COMJ
Hot-Dog
RESS)などの懐しの創刊号を電化、NTTドコモdマガジン」で限定配信
015/07/01
witter公式アプリ
ニュースのキューション機能、本限定でiOS版が提供開始
015/07/01

Q: 普段のWeb検索で、
正しい情報を取得するよう
どの程度心がけていますか？

こんなことやってます

意識の低い人は高い人に比べて、
上位の検索結果だけをみて検索を終えることが多い
(ポジションバイアスの影響が強いことを示唆)



神戸新聞2020年1月17日

9



神戸新聞
@kobeshinbun



【語りかける25年のことば】神戸新聞社は兵庫県立大学社会情報科学部と共同で、1995年1月17日以降、神戸新聞に掲載された約135万本の記事を分析、可視化しました。25年間で「復興」は37105回、「支援」は25501回使われました。*写真は2020年1月17日朝刊よりkobe-np.co.jp/rentoku/sinsai/#阪神淡路大震災



♡ 91 11:08 - 2020年1月17日



引用: <https://twitter.com/kobeshinbun/status/1217992167722930178>

本講義「統計学」

- <https://tyamamot.github.io/statistics-2021/>
 - 配布した資料や講義スケジュールはここで一覧できるようにします。ブックマークしておいてください。



stats

本講義関係でなにかパスワードを求められたら
このパスワードを入力してください

教科書・参考書

どれか1冊は用意してください

- 教科書

- 岡本和夫 「**新版 確率統計**」，実教出版

- 参考書

- 小波秀雄 「**統計学入門**」 （無料）

<http://konamih.sakura.ne.jp/Stats/Text/>

- 参考書 「**統計学入門**」 東京大学出版会

変更があればそのたびに通知します

対面での試験が行える前提での案です

- **30%:**（ほぼ）毎週出す課題
 - － 今のところ**金曜日21時締切**予定
- **10%:** 講義中のミニ演習
 - ・ その日にやればOKにする予定（今日はなし）
- **60%:** 中間テストおよび定期試験

を基準として総合的に評価する

スケジュール (4月13日版)

15

1.	4月13日	1変数データの記述と要約
2.	4月20日	2変数データの記述と要約
3.	4月27日	確率の基礎
4.	5月11日	確率変数・確率分布
5.	5月18日	離散型の確率分布・二項分布
6.	5月25日	連続型の確率分布・正規分布
7.	6月1日	中間テスト
8.	6月8日	母集団と標本の基礎的概念
9.	6月15日	推定量の性質
10.	6月22日	大数の法則と中心極限定理
11.	6月29日	母平均・母比率の推定
12.	7月6日	仮説検定の基礎的概念
13.	7月13日	仮説検定の手順
14.	7月20日	仮説検定の演習・まとめ (期末レポート予定)
15.	7月23日	まとめと発展的な話題
		定期試験

最新のスケジュールは
講義ページを参照
してください

「統計学」の基本的な進め方 (あくまで予定ですが)

16

1. 前の週に出した課題の解説, 資料の補足
2. ミニ演習 (その日中にやればOKの課題)
3. 今週の資料と課題のポイント説明
4. 余った時間で質問対応など

前の週に出した課題 の解説・補足+ コメントへの返答	ミニ演習	今週の資料の ポイント解説	質疑 など
----------------------------------	------	------------------	----------



本講義で扱う内容の概観

(ざっくりと)

17

記述統計

データの整理

- 1変数データ
- 2変数データ

推測統計

母集団と標本

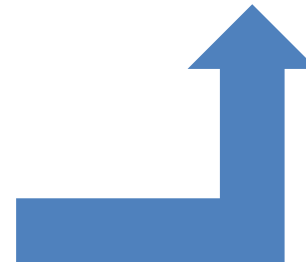
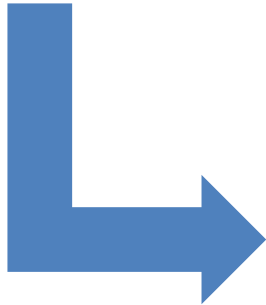
母集団の推定

仮説検定

確率の基礎

集合・確率

確率分布



ありそうな質問

- **文系出身ですが授業についていきますか**
 - － 過去2年実施していますが、文系・理系で成績の差に違いはありませんでした
- **教科書は買った方がよいですか？**
 - － 小波先生の参考書でもいけそう，という人はそれでもOKです． といっても例年ほとんどの人が教科書を購入しているようです
- **他にも講義の進め方， 評価に関して疑問や要望があれば遠慮なく連絡してください**

統計学をなぜ学ぶか
統計学で何を学ぶか

アンケート

(成績評価と一切関係ありません)

TeamsにURLを張ります

<https://forms.gle/Tvz5MPsWD7iVP4XHA>



- 世界の人口のうち、極度の貧困にある人の割合は、過去20年でどう変わったでしょう？
 - a. 約2倍になった
 - b. あまり変わっていない
 - c. 半分になった

- 世界中の1歳児の中で、なんらかの病気に対して予防接種を受けている子どもはどのくらいいるでしょう？
 - a. 20%
 - b. 50%
 - c. 80%

出題の引用元



正しいデータに
基づいて判断すること

人間は分かりやすい
ストーリーを信じやすい

ハンス・ロスリング他, FACTFULNESS (ファクトフルネス) 10の思い込みを乗り越え、データを基に世界を正しく見る習慣, 日経BP, 2019.

なぜ統計学を学ぶか

● ビッグデータ時代

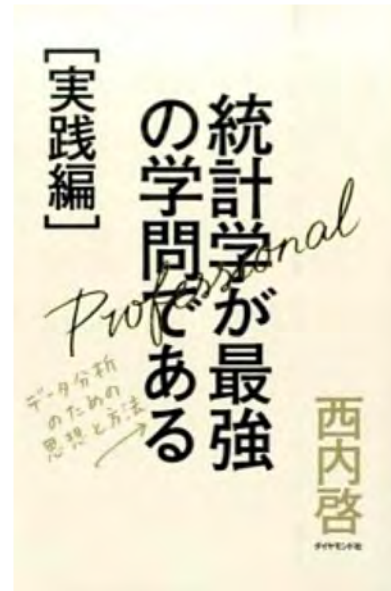
- 人々の購買データ
- ウェブ・ソーシャルメディアデータ

● データに基づいた意思決定

- EBM（エビデンスに基づく医療）
- EBPM（エビデンスに基づく政策）
- 統計学の知識が不可欠
- 「経験と勘」からの脱却

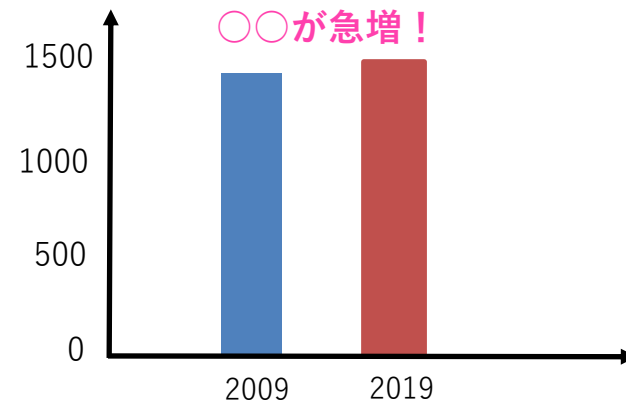
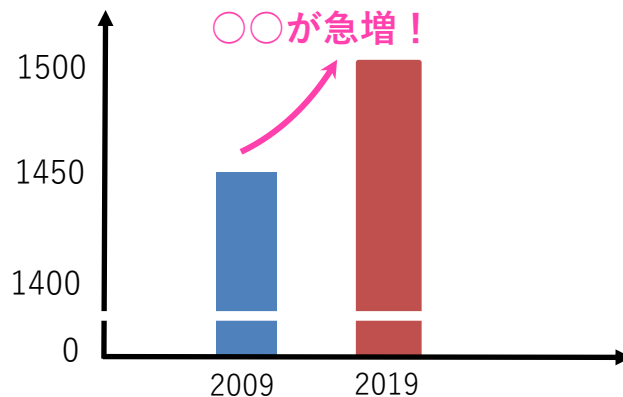
なぜ統計学を学ぶか

- 統計家は今後の最もセクシーな職業だ
 - Hal Varian (Googleチーフエコノミスト)



統計学を学ぶ必要のある人

- 将来，研究や業務でデータ分析に関わる人
 - － 適切にデータを収集し，適切に判断をする
 - － ほぼすべての人？
- 教養としての統計リテラシー
 - － 統計の「嘘」にだまされない・「嘘」をつかない



「新型コロナウイルス以外の理由で病院を受診した患者について、大阪市立大学が（新型コロナウイルスに）感染した経験があるかどうか血液中の抗体を調べたところ、300人余りのうちのおよそ1%の患者から抗体が検出されたことが分かりました。」

- 引用: NHKニュース: <https://www3.nhk.or.jp/news/html/>

20200501/k10012414101000.html (2020年5月2日閲覧)

さて、このことから、「大阪府では、人口880万人の1%にあたる8.8万人が新型コロナウイルスの抗体を持っているかもしれない」と推測することは妥当だろうか？

実際にその病気にかかっている人に対して検査をしたときに、その人を正しく陽性であると判定できる確率を検査の感度という。また、病気にかかっていない人を正しく陰性と判定できる確率を特異度という。感度90%の検査であれば、病気を持っている人がその検査を受ければ90%の確率で陽性と判定され、10%の確率で陰性と判定される、ということである。

さて、いま全国民の1%がかかっていることが分かっている病気がある。その中から国民を1人選び、感度99%特異度99%の検査をしたところ、陽性と判定された。このとき、この陽性と判定された人が本当は病気にかかっていない確率はいくつだろうか？ 1%, 50%, 99% から最も近いものを選び。

統計学で学んでほしいこと

- データから何が言えて、何が言えないか？
を普段から意識できるようになる
- 確率や統計はそこに答えを
与える方法論の1つだと思います

- 講義資料「1変数データの記述と要約」を学習し課題を解くこと

ー ほぼ数学1「データの分析」の復習です

- 課題: 以下のGoogleフォームより提出

<https://forms.gle/wCZtH8K8jdHdcVrU6>

- 締切: 4月16日 (金) 21:00

- 今週の課題の正答率は成績評価に含めません

- Teamsの練習？
- 「1変数データの記述と要約」の
ポイント解説