

統計学

7月28日（火）
第12回（最終回）

兵庫県立大学 社会情報科学部

山本 岳洋

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

2020年度前期・火曜3限 神戸商科キャンパス 全学共通科目

本講義の準備（毎回提示する予定）

2

- slackを見られるようにしておいてください
- 質問やコメントあれば随時slackの
#統計学 チャンネルやDMに書いておいてください
- マイク・カメラはオフにしておいてください
- Webexについては毎回録画して
後で掲載する予定です（失敗しなければ）
 - SNS等で共有しないようにお願いします

- 期末レポートについて
 - 締切: 8月18日（火） 23:59
- 前回の課題の解説・コメント返信
- 仮説検定の補足
- 振り返り・授業アンケート

期末レポートについて

期末レポート

5

- 定期試験は対面・オンラインともに
実施しません
- 締切: 8月18日（火） 23:59
- ユニパのクラスプロファイルより提出

レポートの作成方法

6

(中間レポートと同じです)

- Wordで作成しPDFを提出 (基本はこれ)
- Wordがない人はGoogleドキュメントで作成しPDFを提出
- 数式の記述が辛い人や上記が使えない人は手書きのレポートを撮影し提出
- 慣れている人や余裕がある人はMarkdownで作成しPDFを提出

提出方法と成績評価は関係ありませんが、
できる限りWord等での数式記述になれてほしいと思います

成績評価

(初回講義のとおり)

7

60%

- (ほぼ) 毎週出している課題
- 講義中のミニ演習

40%

- 中間レポート
- 期末レポート

を基準として総合的に評価する

先週の課題の解説・コメント返信

- 講義科目Aでは、定期試験の平均点が80点となるよう問題を作成している。いま、ためしに無作為に抽出した49名の受講生にその試験を受けてみてもらったところ、平均点77点、標準偏差10点であった。さて、この定期テストの平均点は80点といえるだろうか。帰無仮説を「母集団の平均点が80点である」として有意水準5%で両側検定を行う。なお、母集団のテストの点は正規分布に従うと仮定してよく、サンプルサイズ n が大きいため検定をする際に母標準偏差を標本標準偏差で代用して良い。

問1. 答え

10

対立仮説は **母集団の平均点は80点でない**（母平均 \neq 80 点）である．有意水準 5%で両側検定を行うため，帰無仮説を棄却するためには，帰無仮説が正しいと仮定したときの統計量 $Z =$

$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ が棄却域 $Z \leq -1.96$ または $Z \geq 1.96$ に含まれるかどうかを判定すれば良い．実際に統計量 Z を求めると， $Z =$

$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \frac{77-80}{10/\sqrt{49}} = -2.10$ となり， Z は棄却域に**含まれる**．よって，

有意水準5%で帰無仮説は**棄却される**．すなわち，テストの平均点は80点ではないといえる．

- ある工場の機械は、牛肉を平均300 g、標準偏差10 gの正規分布に従うようにブロックに切り分けるように調整されている。いま、この機械が切り分けた牛肉ブロック10個を無作為抽出し重さを測ったところ、平均が295 gであった。この機械は平均300 gに正しく切り分けていると言えるだろうか？帰無仮説を母平均=300 gとして、有意水準5%で両側検定を行うことを考える。
 - － 第1種の過誤とはどのような誤りを言葉で説明せよ
 - － 同様に、第2種の過誤について説明せよ。

帰無仮説と対立仮説

- 帰無仮説
 - 母平均 = 300 g
- 対立仮説
 - 母平均 \neq 300 g

仮説検定における2種の誤り

13

		真実の世界 (本当に成り立っている事実)	
		帰無仮説が正しい	帰無仮説は正しくない
仮説検定の結果	帰無仮説が正しいと判断 (帰無仮説を採択)	正しい判断	誤り (第2種の過誤)
	帰無仮説が正しくないと判断 (帰無仮説を棄却)	誤り (第1種の過誤)	正しい判断

第1種の誤り (第1種の過誤)

14

- 帰無仮説が正しいにもかかわらず、
帰無仮説を棄却してしまうこと
- 問2では . . .
 - 母平均が本当は 300gであるのに
母平均は 300 g でないと結論づけること

仮説検定における2種の誤り

15

		真実の世界 (本当に成り立っている事実)	
		帰無仮説が正しい	帰無仮説は正しくない
仮説検定の結果	帰無仮説が正しいと判断 (帰無仮説を採択)	正しい判断	誤り (第2種の過誤)
	帰無仮説が正しくないと判断 (帰無仮説を棄却)	誤り (第1種の過誤)	正しい判断

第2種の誤り (第2種の過誤)

16

- 帰無仮説は本当は正しくないのに、
帰無仮説を正しいと結論づけてしまうこと
- 問2の例では . . .
 - 母平均が本当は 300g ではないのに
母平均は 300 g であると結論づけること

問2. 仮説検定の結果

17

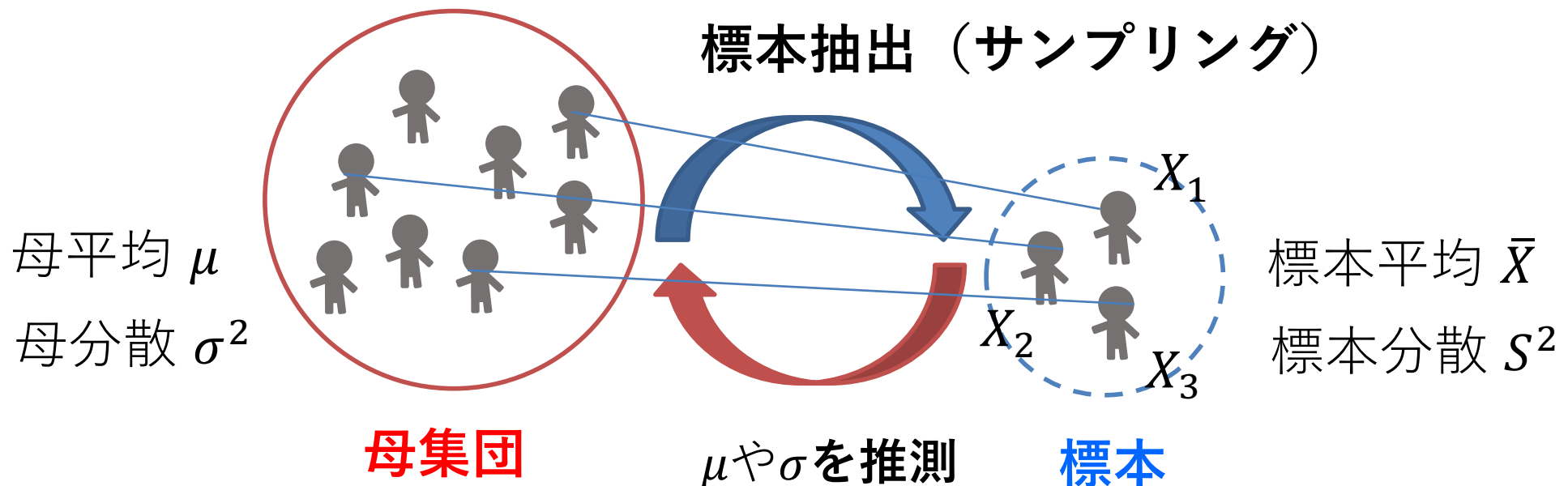
対立仮説は **母平均 \neq 300 g** である. 有意水準 5%で両側検定を行うため, 帰無仮説を棄却するためには, 帰無仮説が正しいと仮定したときの統計量 $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ が棄却域 **$Z \leq -1.96$** または **$Z \geq 1.96$** に含まれるかどうかを判定すれば良い. 実際に統計量 Z を求めると, $Z = \frac{295 - 300}{10/\sqrt{10}} = \textbf{-1.58}$ となり, Z は棄却域に**含まれない**. よって, 有意水準5%で帰無仮説は**棄却されない**. すなわち, この機械は牛肉を**平均 300 g に切り分けているとも平均300 g に切り分けていないともいえない**.

「統計学」でやったことと これからの話題

本講義でやったこと: 統計的推測の入り口

19

- 母集団から標本抽出された標本の平均（**標本平均**）や分散（**標本分散**）などの情報をつかって、母集団の平均（**母平均**）や分散（**母分散**）などについて予測を立てたい



本講義でやったこと: 統計的推測の入り口

20

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

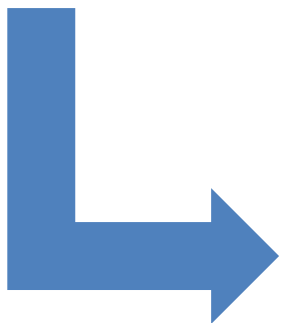
母集団の推定

仮説検定

確率の基礎

集合・確率

確率分布



本講義に関連するこれからの講義

21

(山本独自の見方)

大量のデータに基づき
予測をするための
理論・手法

機械学習

人工知能

データマイニング

時系列データや多変数
データに関する
詳細な理論・手法

統計的モデリング

多変量解析

標本を正しく収集し
分析するための
方法論

社会調査法

確率・統計

代表的・実用的な
検定・分析手法

統計学

確率や統計の基礎

- 本講義で扱った，正規分布を用いた検定は実問題ではあまり用いることはない
 - － 実際に扱う標本の統計量は正規分布ではない分布に従うことがほとんど
- ただし，正規分布を用いた検定の理論が分かっているならば，分布が変わっても考え方は理解できる

いろいろな仮説検定

学部ごとのPC保有率の標本調査結果

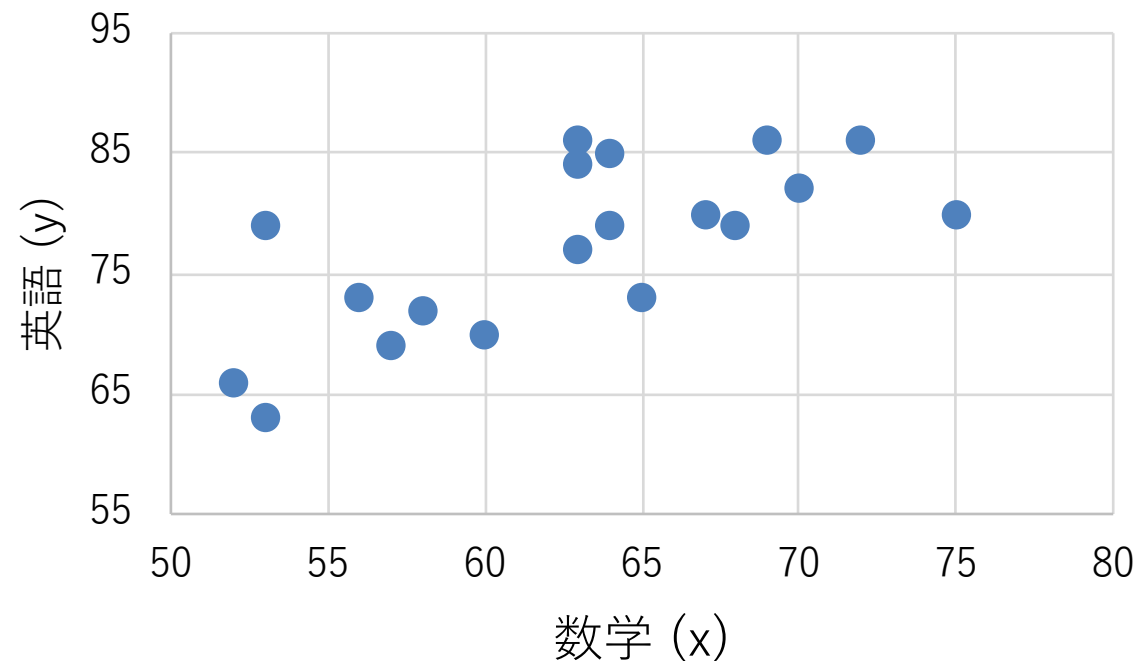
		PC保有 (y)		合計
		持っている	持っていない	
学部 (x)	社会情報科学部	20	10	30
	国際商経学部	40	30	70
合計		60	40	100

- 社会情報科学部生の方が
PCの保有率が高そう

－ 独立性検定（カイ二乗分布）

いろいろな仮説検定

標本として集めた学生20名の
数学と英語成績の散布図

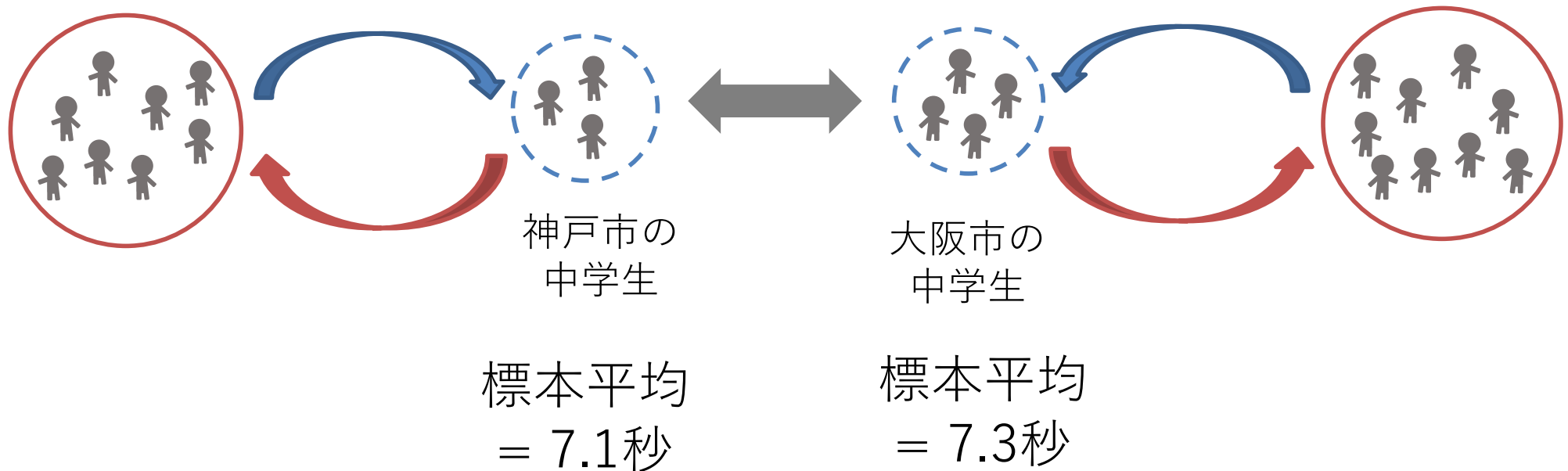


- 英語と数学に相関関係がありそう
 - 無相関の検定 (t分布)

いろいろな仮説検定

25

2つの都市の中学生の50m走タイムを比較してみたい



- 神戸市の学生の方が大阪市の学生よりも早い！

－ 2標本の平均の差の検定 (t分布)

いろいろな仮説検定

(重回帰分析)

26

ある商品の売り上げを顧客の情報から予測したい

商品Aの売り上げ

$$= \underline{1.3} \cdot \text{年齢} + \underline{2.5} \cdot \text{所得} - \underline{0.2} \cdot \text{家族構成} + \underline{4.3} \cdot \text{趣味}$$

- 本当にその係数には意味があるの？

－ 偏回帰係数の検定 (t分布)

振り返り・授業アンケート

- 教科書「確率・統計」よりも
少し詳細に説明しました
 - － 参考書 東大出版「統計学入門」から入門として抑えるべき箇所をピックアップした感じです
- 数式や定義に従ってものごとを考える
基礎力がついたと自信をもってください

- **Excelを積極的に使いました**

- コンピュータを使ってやりたいことを実現する，ということの1例として捉えてもらえればと思います

- **後期では R や Python といったプログラミング言語を学びます**

- だんだん，実現したいものを作れるようになります

- 前半は資料を見ながら課題を解いて、
講義で解説・補足という流れ

- 個人的には来年も前半はこの形でいけそうと感じましたが…

- 後半は まず講義，その後資料を
見ながら課題， という流れ

- 実は後半も前半と同じ形式でもいける？

山本の反省点

- 講義中の休憩があまりとれなかった
- Excelは説明動画に丸投げで、みなさん個々人がどれだけExcelを使えたのか確認不足
- 証明系の課題が分量不足

授業アンケートに回答してください 32

- ユニパにログイン
 - メニューの授業評価
 - 授業評価回答



- 質問項目 10

- 教科書・参考書で「使った」と思うものを全てあげてください
 - 教科書「確率・統計」
 - 小波先生「統計学入門」
 - 東大出版「統計学入門」

- 質問項目 11

- 本講義の全体的な難易度感について教えてください
 - 特に問題なく一人で取り組めた
 - Webや友だちに聞きながらできた
 - ついていくのがしんどかった