

## 統計教育における評価と指導方法に関するガイドライン

### 大学レポート\*

#### Guidelines for Assessment and Instruction in Statistics Education (GAISE)

#### College Report

Martha Aliaga, George Cobb, Carolyn Cuff, Joan Garfield (chair), Rob Gould, Robin Lock, Tom Moore, Allan Rossman, Bob Stephenson, Jessica Utts, Paul Velleman, Jeff Witmer

### 概 略

アメリカ統計学会 (ASA) は、統計教育における評価と指導方法に関するガイドライン (GAISE) 事業に資金を提供した。この事業は2つのグループから成り立っており、1つは K-12 教育 (訳注 高等学校までの教育) に重点を置き、もう1つは大学の入門コースに重点を置いている。このレポートはその大学グループによって作成された提案を示している。

レポートでは、大学の入門コースの歴史を手短に述べ、ジョージ・コブ氏による 1992 年のレポート<sup>(1)</sup> について概説している。コブ氏のレポートは、これらのコースを教える際の、一般的に認められた提案事項であると考えられている。入門コースの現在の状況として、入門コースの教育についての調査結果をまとめている。その後、私たちは、統計的リテラシーを持つことを目指して学生の到達目標を提示した。コブ氏のレポートで提案された内容を基本として、統計学入門コースを教えるための 6 つの提案を表した。この 6 つの提案は次のようなものである：

1. 統計的リテラシーを重視し、統計的思考力を育成する
2. 本物のデータを使う
3. 単なる手順に関する知識よりも、概念の理解に重点を置く
4. クラス内での活動的な学習を育む
5. 概念の理解とデータ解析を向上させるために情報機器を活用する
6. 学生の学びを向上させたり、見極めたりするような評価を用いる

レポートは、どのようにこれらの変革を行っていくのかという提案と、提案を詳細に例示するため、付録に数多くの例を載せて締めくくっている。

---

\*英語版原典の翻訳・転載については、The American Statistical Association より許可を得ている。  
Copyright 2010 by The American Statistical Association. All right reserved.

## 序論

GAISE プロジェクトは、K-12 カリキュラムにおける統計学の指導と評価、そして大学の統計学入門コースのための ASA 推奨のガイドラインを発展させるため、2003 年 ASA から選ばれたメンバーによる新しい試みで始められた助成金によって、資金が提供された。

私たちの大学課程のガイドラインに対する取り組みには、e-メールによる討論や個人的な小グループの集まりでの討論も含まれていた。私たちの討論は、現行の基準やガイドライン、統計の指導法や学習法に関する調査結果、統計的推論の根底にある重要な概念についての指導や評価に焦点を置く必要性についての最新の議論や提案を調べることから始まった。

### 入門コースの歴史と発展

現代の統計入門コースをずっとさかのぼっていくと、統計的手法についての初期の本が起源となっている。1925 年にその初版が発行された R.A.フィッシャー、「研究者のための統計的方法」は実践的な科学者たちを教育するのが目的だった。その 12 年後、ジョージ・スネデカーの「統計的方法」は、同じ内容を発展させたものあったが、対象とする読者に少し違いが見られた。フィッシャーの本より、スネデカーの本は、まだ学位を取得中の未来の科学者たちのための課程における教科書であった；統計学はかなり実務的で職業を主眼としているにもかかわらず、学術的な科目として確立され始めていた。1961 年までにはフレッド・モステラー、ロバート・ルーク、ジョージ・トーマスらによる、「確率の統計的な応用」(Probability with Statistical Applications) の出版に伴い、統計学はより幅広い学術的カリキュラムへと発展したが、ここでまた問題点が生じた。この早い時期には、統計学はその妥当性について、確率に大きくよりかからなければならなかった。

1960 年代の終わりごろから 1970 年代の初期の間に、ジョン・チューキーの探索的データ解析の考えが、カリキュラムに革命的と言ってよい、ひと組の変革をもたらした。つまり、ある種類のデータ解析の手法を確率に基づいたモデルから切り離すことで、データ解析が独立した知的活動としての地位を獲得し始めたこと、そして大きくて扱いにくい計算機につきっきりで何時間も過ごす必要なく、学生がデータ解析ができるように、間に合わせのデータツールを集めたものを導入したことである。コンピューターが後に、初等的な統計カリキュラムにおける「データ革命」を完全なものとしたが、チューキーの「探索的データ解析」(EDA) の考えが最初の技術的な躍進と、作り物の例を避ける新しい精神をもたらした。

大きな影響を及ぼす 2 つの本が、1978 年に現れた：デヴィッド・フリードマン、ロバート・ピサーニ、ロジャー・パーヴズによる「統計学」、とデヴィッド S. ムーアによる「統計：概念と論争」。この 2 冊の本の出版は、私たちが現代統計学入門コースとみなしているものの誕生を記している。

内容の進化が他の傾向と同時進行してきた。このうちのひとつは在籍者数の著しく、そ

して安定した伸びを示している点である。この点について、ふたつの統計がここにある。

- ・数理科学会議によると、2年制の大学では、統計学の登録者は、1970年の微積分学の登録者数の27%であったものが2000年の微積分学の登録者数の74%へと拡大した。
- ・統計学のAP(Advanced Placement)テストは1997年に初めて実施された。その最初の年にそれを受けた生徒は、7,500人いた。これは、APテストにおける他のどの教科の実施初年の受講者数よりも多かった。翌年には、15,000人の生徒がそのテストを受け、その翌年は25,000人、そのまた次の年は35,000人であり、2004年には、65,000人以上の生徒が統計学のAPテストを受けた。

コースの内容変化と登録者数のめざましい増加の両方によって、受講者の背景、興味、意欲が多様化し、拡散するという三番目の変化をもたらした。統計学は、農学と生物学の未来の科学者たちという狭いグループを対象としたスネデカーの本のような教科書を用いて教えられるコースから離れて、中学校から学士以降まで、あらゆるレベルの生徒に教えられる、そして様々な関心と目的を持つ、コースとなっていった。1940年代の教師は、スネデカーの「統計的方法」を使用し、たいていの学生は定量的スキルを身につけており、就職設計によって十分動機づけられていると考えていた。今日の統計学入門コースの教師は、異なった学生のグループを対象としている。ほとんどが若いうちから統計学を取っていて、高校でもますます増えている；差し迫った必要性から統計学にひきつけられた学生はほとんどおらず；定量的スキルのレベルは実に様々である。その結果、今日の教師は半世紀前のものより相当多くの動機づけや説明をすることへの取り組みに直面している。

統計学入門の、「何を、なぜ、誰が、いつ」が変化してきただけでなく、「どうやって」も変化してきた。ここ2、30年で、学生はどうやって統計学を学ぶのか、そしてどうやって私たち教師は彼らの学習を効果的にサポートできるのかということに重点を置く、並はずれたレベルの活動が見られてきた。

### 1992年のコブ氏のレポート

1991年春、ジョージ コブ氏は、数学の分野の重要な問題点を強調するため、アメリカ数学学会カリキュラム実行プロジェクト(MAA)の一部として、統計教育の電子メールグループを編成した。そのレポートはMAAの本、変革を求める声に注意せよ<sup>(2)</sup>に掲載され出版された。それには次の提案が含まれていた：

#### 統計的思考力を力説せよ

どの入門コースにも、学生が統計的思考力の基本要素を学ぶ手助けとなる中心的な目標を掲げるべきである。上級コースの多くも、同じ基本要素をより明確に強調するような改善をすべきである。すなわち：

- ・データの必要性。個人的な決定を根拠（データ）に基づいて行う必要性と、根拠によらない仮定に基づいて行う推論に従って行動することに潜んでいる危険性を認識する

こと。

- ・データ作成の重要性。問題を明確に定式化することと、実際に適切な問題を扱っている良質なデータを手に入れることは、困難であり時間のかかることだということを認識すること。ほとんどの人は、自分でこの経験をするまで気がつかないようだ。
- ・変動のさまざまな原因。変動が至る所にあるということを認識すること。これは学問としての統計学の真髄であり、講義ではよく理解できない。経験する必要がある。
- ・変動の数量化と解釈。以下のことを考慮することで、変動は計測し説明できることを認識すること。(a)ランダム性と分布；(b)パターンと偏差（適合と残差）；(c)パターンに対する数学的モデル；(d)モデルとデータとの対話（診断）

### データや概念を増やし、理論と方法を減らす

統計学のほとんどのコースは、理論と方法を少し減らして、データと概念をさらに重視することで改善できる。実行可能な最大の範囲まで、計算とグラフが自動化されるべきである。

### 活動的な学習を育む

一般的に統計学の教師は、講義で教える部分をずっと少なくし、プロジェクトや実験作業、そしてグループによる問題解決と討論活動のような、他の手段をもっと使用すべきである。伝統的な講義の環境であっても、学生に、より活動的に関わらせることは可能である。

3つの提案がたいへん幅広く応用されようとしている。（例えば、コースに微積分が必須条件なのか、学生が特定の統計的方法を学ぶことを期待されているかを気にかける必要があるのか、など）このコブ氏のレポートを作成した電子メールグループの仕事はレポートを完成させることで終わったが、そのグループのメンバーの多くが、大学生向けの統計学に関する ASA と MAA の合同委員会のメンバーとして活動し、特に普及と実施の活動に力を入れている。

### 統計学入門コースの最新の状況

コブ氏のレポートの出版後 10 年間にわたって、統計教育において多くの変革が実行された。近年では、多くの統計学者が、統計学入門を教えることを目的とした統計学教育における改革運動にかかわるようになった。<sup>(3)</sup> そして国立科学財団は、この改革の実施面を企画する数多くのプロジェクトに資金を提供してきた。ムーア<sup>(4)</sup> は、その改革の特徴を、内容（データ解析を増やし確率を減らす）、教育（講義を少なく、より活動的な学習に）、情報機器（データ解析と模擬実験）の変革という意味で述べている。

1998 年と 1999 年に、ガーフィールド<sup>(5)</sup> は、入門コースがどのように教えられているのかわかりやすく、教育改革運動の影響を調査し始めるために、数学と統計学部出身の多くの統

計学指導者と心理学、社会学、商学、経済学部出身の少数の統計学指導者を調査した。この調査結果として、入門コースでは大きな変革が行われていて、変革の主な領域は情報機器の使用であり、そのコースの指導者は多くの時間を必要としていることがわかった。結果は意外にも学部間での違いは少なく、大きな違いは、グラフ電卓の使用の増加、2年制大学の数学科で教えられているコースにおいて、活動的な学習と新しい評価法、統計学部の指導者によるウェブ機器の使用の増加、また、なぜ変革を行う理由（より多くの数学指導者が統計教育からの提案に影響されている）に見られた。レポートでは、また、より情報機器の活用ができるように、さらに多くの変革がなされる予定であるということを一貫して述べている。

今日の統計入門コースは、実際多くの学問分野や学部にまたがって教えられているコースである。これらのコースを登録した学生たちは、様々な背景（例えば、数学、心理学など）や到達目標（例えば、研究プロジェクトで統計分析を行いたいと望んでいる学生もいれば、一般的な定量的推論で求められるスキルを身につけたい学生もいる。）を持っている。

昔のように、これらのコースの中には、大規模なクラスで教えられているものや、小規模なクラスで（あるいは新入生研修で）教えられているものもある。コンピューター実習室で教えられている学生もいれば、電卓しか使わないコースで教えられている学生もいる。あるいはまた、クラスメートや教員と対面しない遠隔地教育によるコースを取っている学生もいる。また、クラスの中には10週間の四半期で教えられるものもあれば、15週間の半期にわたって教えられるものもある。これらのクラスの1週間における時間数は3~6時間の範囲となっている。

今日の学生の到達目標は、概念の理解や統計的リテラシーと統計的思考力の育成に、より重点が置かれる傾向にあり、一連のツールと手続きを学ぶコースは少なくなっている。情報化の時代の中でデータを扱う必要性が伸び続けており、情報機器やソフトウェアの発展によって、ツールと手続きは、多くの人々にとって入手しやすく、使いやすくなっている。このように、手続きの仕組みを教える必要性の減少と、より多くの人々に基本的概念をしっかり理解させることの重要性の増加によって、これらのツールを賢く使用し、理解することが必要となってきた。次のセクションに述べられているが、これらの新しい到達目標は、学生にとっての重要な学習目標の達成を支援するような、多くの統計学入門コースの再検討と改訂の必要性を強化するものである。

### 入門コースの学生の到達目標：統計学の教育を受ける意義

学生が統計的リテラシーをもち、データをうまく使いこなす消費者となるような教育に傾倒したコースを受け持っている人もいる。これは、芸術鑑賞のコースにすこし似ている。学生が統計解析の実施者になるような教育にかなり偏ったコースを持っている人もいる。これは、スタジオアートコースと類似している。ほとんどのコースは消費者と実施者を混

合したものであるが、その混合のバランスが、私たちが示すそれぞれの提案の重要性を決定づけるものとなる。

すべての統計入門コースにとって望ましい結果は、統計学の教育を受けた学生を生み出すということである。そしてそれは、学生が統計的なりテラシーを育成し、統計的に考えることができる能力を向上させるべきだということを意味する。

以下の目標は、そのような学生が何を知り、何を理解すべきか、ということを表している。この知識を身につけるには、統計的な技能を多少学ぶ必要があるが、特定の技能を身につけることは、それらを学ぶ経過の中で生まれる知識と同じくらい重要であるわけではない。それ故、私たちは特別な内容をカバーすることを勧めているわけではない。

学生は以下のことについて信念を持ち、理解すべきである：

- データは逸話に勝つ
- 変動は自然なことであり、予想できるものであり、量として表すことが可能である
- 無作為に標本を選ぶことで、調査や実験の結果を、標本の抜き取られた母集団へと一般化することができる。
- 比較に基づいた実験における無作為割り付けによって、原因と結果の結論を引き出すことができる
- 連関は因果とは限らない。
- 統計的な有意性は、必ずしも実用的な有意性を意味するわけではない。特に標本サイズの大きな研究では顕著である。
- 統計的に有意ではない差や関係を見つけたとしても、母集団において違いがないとか、関係がないということを必ずしも意味するものではない。特に、標本サイズの小さな調査では顕著である。

学生は次のことに気づくべきである：

- 調査や実験における偏りの共通した出所
- データが集められた方法に基づいて 統計的推論の結果を一般化できる母集団を決定する方法。
- データの集められた方法に基づいて、連関から因果関係を引き出せる場合を決定する方法
- 「正規」、「無作為」、「相関」のような言葉は、普通の使い方とは異なり、統計学において特定の意味を持っているかもしれないこと。

統計学が、問題の解決に取り組んでいるプロセスの部分を、学生は理解すべきである：

- データを得たり生みだしたりする方法
- データ解析の第一段階として、データをどうグラフ化するのかということと、興味の

ある問題に答えるのに十分な場合を知る方法

- データのまとめとグラフ表示を解釈する方法—疑問に答えることと条件を調べることの両方（統計的手順を正しく使うため）
- 統計的な推論の適切な利用方法
- 統計解析の結果の伝え方

学生は、以下の点を含めて統計的推論の基本的な考え方を理解すべきである：

- 標本分布の概念と、それをデータのサンプルに基づいて、どのように統計的推論に応用するのか（標準誤差のアイデアも含めて）
- 有意水準や p 値も含めた統計的有意の概念
- 信頼度や誤差限界の概念も含めた信頼区間の概念、

最後に、学生が知っておくべきこと：

- 背景を考慮して統計的な結果をどう解釈すべきか
- 統計的情報を含んでいるニュースの話や学術雑誌の記事を批判する方法。その際には、発表に何が欠けているのか、その情報を生みだすのに使用された研究の流れや方法も含まれる。
- いつ統計学者に支援を求めるべきか

## 提案

私たちは、コブレポート<sup>(6)</sup>に見られる3つの独創的な目標の中の考えを支持し、今日の実況を考慮しながらそれらを拡張させてきた。これらの提案の意図は、前のセクションで述べた学習の目標のリストを、学生が達成する手助けをすることである。

### 提案1：統計リテラシーを重要視し、統計的思考力を育成する

私たちは、統計的リテラシーを、基本的統計用語（例えば、統計用語と記号が何を表しているのかを知り、統計グラフが読み取ること）と統計の基本的考えを理解することとして定義する。統計リテラシーについては、Gal<sup>(7)</sup>、Rumsey<sup>(8)</sup>、Utts<sup>(9)</sup>を参照のこと。

統計的思考力は、統計学者が統計問題に取り組んだり解決したりするときに使う思考の形であると、定義されてきた。統計的思考力にはデータの必要性の理解、データ作成の重要性、変動の原因、変動の数量化と説明<sup>(10)</sup>として述べられてきた。私たちは以下の例と類推によって、統計学思考力を示す。

### 漏斗の例

上が幅広くあらゆる場合に対応し、また、下の方は狭く、対応できる部分が少ない漏斗を考えてみなさい。統計学者は現実的な問題解決者である。依頼人が問題（例えば、治療効

果はあるのか)を提示した時、統計学者はその問題に有効な、現実的な答えを得ようとする。話を伝えるには単純なグラフで十分である場合も多い。また、より詳細な図表によって、その問題に答えられる場合もある。そうでなければ、ある種の詳細な計算が必要になるかもしれない。全体な状況を単純化して、簡単な統計的検定によって、治療の効果はあるということを確認するかもしれない。もし、状況の単純化が難しければ、手の内にあるモデルの状況をもっと調べて、より厳密な統計的検定が使われるかもしれない。異なる統計学者たちは、与えられたデータセットに対して、すこし異なった解析をおこなうかもしれないが、主要な結論部分では大体同意するだろう。そして小さな点についてのみ、依頼人がこれらの点を気にかけるかどうか気になるだろう。もし問題に答える手順や標準的な方法がないのなら、統計学者はその時だけ新しいツールを開発するため新しい原理の開発を行うだろう。学生にある一定のツールや手順を示したり、どんな与えられた状況でも使用するべき最高の手順があったりするか、その手順のみが認められるという印象を与えるよりむしろ、この統計家が行う思考をモデルとするべきだ。

## 大工仕事との類似

第1週の大工仕事(統計学)コースでは、私たちは様々な種類のかんな(統計指標)の使い方を学んだ。第2週では、様々な種類ののこぎり(グラフ)の使い方を学んだ。その後、金づち(信頼区間)の使い方について学んだ。それから、様々なタイプの木材(統計的検定)の特徴について学んだ。コースの終わりまでには、大工仕事(統計学)の多くの面をカバーした。しかし、私が望んでいたのはテーブルの作り方(問題に答えるためのデータの集め方や統計解析の方法)を学ぶことであり、そのやり方を私は一度も学んでいなかった。私たちは、統計学の実践的な行動とは問題に答えるためにデータを集めて解析することだ、ということを学生に教えるべきである。

## 教師への提案：

- ✓ 具体的な例に取り組み、発想から推論まで、統計的問題を解く際にかかわってくる問題やプロセスを明らかにしながら、学生のための統計的思考のモデルとせよ。
- ✓ 情報機器を活用せよ。データを管理し、調査し、推論し、推論の過程の背後にある状況をチェックするために、情報機器を効果的に用いる方法を学生に示せ。
- ✓ 学生に、統計的思考力を育成し、それを活用するような練習課題を与えよ。この中には、結論のない問題やプロジェクトも含めるべきである。
- ✓ 学生にどのツールを使用するのかを教えたり、またそれをただ実行させたりするのではなく、適切な問題や情報機器を選ぶ練習課題をたくさん学生に与えよ。
- ✓ 学生の統計的思考力を評価し、フィードバックせよ。

付録にプロジェクトの例、活動、統計的思考力を育成し、評価するのに使用できる評価問題を示している。



## 提案 2：本物のデータを使用せよ

統計教育において、どのように、そしてなぜそのデータが作成されたり、収集されたりしたのかという問題を考えたり、解析結果を問題の背景と関連づけるたりすることを実際に行うためには、本物のデータを使うということが重要である。学生にとって関心のある実際のデータセットを使うこともまた、そのデータや適切な統計的検定について、学生に考えさせるのに良い方法である。アーカイブされたデータ、教室で生成されたデータ、模擬実験のデータも含めて、本物のデータには多くの種類がある。時には、ある特定の点（例えば、アンスコムデータの4つのデータセットが、同じ程度の相関関係であるのに、どのようにして著しく異なった散布図となっているのかということについて例示している）について例示するため、あるいは、ある特定の概念を評価するために、仮想的なデータセットが使用されることがあるかもしれない。一般的なデータ解析や調査のためではなく、この特定の目的のために作成された、あるいは現実的なデータのみを使うことが重要である。本物のデータを扱うという重要な局面は、学生が良い問題を定式化し、データがどのように作成されたのかに基づいて、それらの問題に適切に答えるためにデータが使えるようになる手助けをしている。

### 教師への提案

- ✓ ウェブのデータアーカイブや、教科書、ソフトウェアのパッケージ、教室での調査あるいは活動から、良い生のデータを探しなさい。もし、機会があれば、実践的な科学者から直接（学術雑誌やその人の所属する機関を通して）本物のデータを探し出さなさい。そのようなデータを使うとクラスを活気づけることができ、新しく発見されたデータを他の人に伝えることで、他の教師のための良いデータセットの蓄積を増やすことができる。データ要約ウェブサイト、学術雑誌の記事、世論調査のウェブサイト、教科書から、本物のデータに基づいた概略を探したり使ったりしなさい。
- ✓ 状況に合わせて、問題に答えるためにデータを用いたり、問題を定式化したりしなさい。
- ✓ データセットに使われた問題が学生の興味のあるものなのか確かめなさい。もし、誰も気にかけないような問題なら、それは入門クラスのデータセットとして良いとは言えない。（例：誰も聞いたことのない種の身体測定値）注：すべての学生にとって関心のあるデータセットはほとんどない。それゆえ、様々な背景からのデータセットを使うべきである。
- ✓ アンケートを作成したり、データを収集したりする前に、教室で生成したデータを使って、統計的問題を練り上げたり、データの解析の計画を立てたりしなさい。（例：様々な形のヒストグラムができ、関係を調査するための興味深いカテゴリカル変数を使いそうな質問をすること）。授業で学生から集められたデータは、学生にとって恥

ずかしいということがありうる情報を含んでいないこと、また学生のプライバシーが保たれるような情報であることが重要である。

- ✓ 大量のデータセットを入力するのに時間を費やすよりも、むしろ少量のデータセットあるいはデータセットの一部を使って、生のデータ入力を、学生に実習させなさい。大量のデータセットはコンピューターで使えるようにしなさい。
- ✓ コースの様々な部分で、変数の異なる部分集合を使用しなさい。しかし、全体を通して、同じデータセットを繰り返し用いなさい。(例：ふたつのグループを比較するために、平行箱ひげ図(ボックスプロット)を用い、それから同じデータに対して2標本t検定を適用すること。特性を調べるためヒストグラム(柱状図)を使用し、それから仮説検定のための条件を確認すること。)

付録には、課題、プロジェクト、テストなどでデータに使用する適切な方法(並びにあまり良くない方法)の例が示してある。

### 提案 3：単なる手順の知識よりも概念の理解を重要視しなさい

入門コースでは、教育内容として多くのことを含んでいるコースが多くあり、学生は、ただ表面上でしか理解していない、うまく統合できていない、すぐに忘れてしまうような知識の集めるだけで終わってしまう場合もある。もし学生が概念の重要性を理解していなければ、一連の手順を知っていることに、ほとんど価値がない。もし学生が重要な概念をよく理解していれば、特定の手順を容易に学べるだろう。学生の心の中には、手順の段階には注意を必要とすることが多すぎて、有用な教師でさえ概念の方向に向かって指導できていない。

手順よりも概念の方により注目することは、学生にとっても学部の要求としても政治的に難しいかもしれないということを認識しなさい。しかしながら、入門コースで概念の基礎がよくできている学生は、次の段階のコースで学習する研究方法、回帰、実験計画、あるいは統計手法を学習するための準備がよくできているだろう。

### 教師に対する提案：

- ✓ 様々な手法をカバーすることではなく、概念を発見することを本来の目的と考えなさい。
- ✓ 背景にあるアイデアにはあまり注目せず、多くの技能を扱うよりも、少ない技能を用いて例示することで学生が主要な概念の理解ができることを重要視すること。
- ✓ 核となる概念にもっと深い意味で焦点を当てるには、入門コースの内容を削減しなさい。多くのトピックをカバーすることよりも概念を重要視したシラバスの例は、付録に載せている。

結果の解釈をさらに強調させるため、決まりきった計算は情報機器を使って行いなさい。

数学という言葉は主要なアイデアを簡略に表現することができるが、概念の理解を高めるために公式を使用し、理解から遠ざけるような計算は避けなさい。例えば、 $s = \sqrt{\frac{\sum (y - \bar{y})^2}{n-1}}$  は、学生が、広がりやを測る尺度としての偏差基準の役割を理解するのに役立つ。しかし、 $s = \sqrt{\frac{\sum y^2 - (\sum y)^2/n}{n-1}}$  は計算のむずかしさを上回るような教育的な価値はない。

#### 提案 4：教室での活動的な学習を育成する

教室において活動的な学習方法を用いることは、協同学習を促進する方法として価値のある方法であり、学生がお互いから学ぶことができる。活動的な学習によって、学生は重要な統計のアイデアを発見し、構成し、理解するようになり、統計的思考力を形作るようになる。活動によって学生が学習に従事することが多くなり、学習のプロセスを楽しくさせるという点で、さらに利点がある。活動的な学習方法のその他の利点は、学生に統計用語を使ってコミュニケーションを取らせ、チームで仕事することを学ぶことができる点にある。活動は教師に、生徒の学習に対する形式的ではない評価方法を提供し、生徒がどの位順調に学習を続けているかについて、教える側がフィードバックすることができる。教師は教材を教えるための活動の能力を過小評価してはいけないし、講義を過大評価してもいけないということが重要である。そういうわけで、大きな講義クラスでも活動を組み入れることについて、提案がされた。

活動的な学習に含まれるタイプは：

- ・グループや個別の問題解決の活動と討論
- ・実習室での活動（身体的な活動とコンピューターを使用した活動）
- ・現場で学生が生み出したデータに基づくデモンストレーション

#### 教師に対する提案：

- ✓ 実際の問題の文脈に活動の基礎を置くこと。それゆえ、データは、「データを集めるために集める」のではなく、問題に答えるために集められるべきである（無論当然のことである）。
- ✓ 講義と活動、討論、実習を混合させること。
- ✓ 物理的な説明ができるコンピューターシミュレーションから始めなさい。（例、サイコロを振る、カードをきる）
- ✓ 学生からデータを集めること。（匿名で）
- ✓ データを分析する前に、活動としてデータを集める研究の結果について、学生が予測するよう奨励すること。このことは統計的手法の必要性を与える。（もし、すべての

結果が予想可能なら、データも統計も必要ないことになる)

- ✓ 学生にひとつひとつ手順の段階を踏ませるような活動をさせないで、データや問題について考えさせたり、討論させたりすること。
- ✓ 問題を明らかにし、学生に問題を通して学習させ、同じ授業時間内に活動を要約する時間が十分にあるか確認できるよう、前もって計画すること。次の授業時間に活動を完結させるのは困難である。次の授業時間の始めであっても、要点を繰り返したり、報告したりする時間があるか確認すること。
- ✓ 学生の実行したことや学習したことに対して、彼らに多くの意見・感想を与えること。
- ✓ 活動の重要な構成要素として評価も含めること。

#### 大規模クラスの活動的な学習実施に対する提言：

- ✓ 大規模クラスでは、学生が作成するデータサイズが大きくなることを利用しなさい。
- ✓ 大規模クラスでは、学生にペアで活動させることが、大きなグループでさせるより簡単である場合がある。
- ✓ できれば、活動を行う際には、実習と討論は分けて実施しなさい。

#### 提案 5：概念を発達させデータ解析に情報機器を利用しなさい

情報機器は統計学者の研究の方法を変えてしまった。そして、何をどうやって教えるのかも変えるべきである。例えば、正規分布の確率表のような統計表は、もはやp値を見つける際には必要なくなってきた。コンピューターを利用して計算できるようになっている。私たちは、コンピューター操作を教えるのではなく、むしろ、学生に結果の解釈や条件のチェックを重要視させながら、データ解析を行うために情報機器は使われるべきだと考えている。情報機器の利用は、また、学生が概念を可視化し、模擬実験によって抽象的な考えの理解を発展させる手助けとなるように利用すべきである。

ツールの中には、この2つのタイプの利用ができるものも多くある。その一方で、統計ソフトウェアパッケージとウェブ上のアプレットで補完的に用いる必要がある場合がある。使用されるツールは言うまでもなく、数値を計算するためのものとしてだけ見るのではなく、概念的なアイデアを調べ、学生の学習を高めさせる方法としても見るのが大切である。私たちは、情報機器の利用を、単なる計算（例：グラフ計算機に100個の数値を入力し、統計的指標を計算する）や誤った正確さ（小数点以下の桁数が多いところまで結果を出す）のために使うことに対して警告する。情報機器がすべて望まれた特徴を持っているわけではない。今後も新しいものが常に出現してくるであろう。

使える情報ツール

- ・ グラフ計算機

- ・統計パッケージ
- ・教育ソフト
- ・タブレット
- ・表計算ソフト（スプレッドシート）
- ・データアーカイブ、オンラインテキスト、データ解析ソフトを含めた WEB 上の情報源
- ・教室での反応システム

### 情報機器の使用法に対する教師への提言

- ✓ 大量の本物のデータセットを入手する
- ✓ 計算を自動化する
- ✓ 統計グラフを作ったり、適切に修正したりする
- ✓ 抽象的な概念を例証するために模擬実験をする
- ✓ 「もし...だったら、どうなる」といった問題を調べる
- ✓ レポートを作成する

### 情報ツールを選ぶ際に教師が考えるべきこと

- ・ データ入力の容易さ、多様なファイル形式からデータを取り込む能力
- ・ 対話能力
- ・ データ、グラフ、数値計算の間の強力なつながり
- ・ 特定の参加者にとって使用が簡単であること
- ・ 学生にとって入手可能で携帯できること

### 提 案 6： 学生の学びを調べ、向上させるために評価を用いる

教師が評価したものを学生は価値あるものとして受け入れる。そのため、到達目標に合わせた評価を行う必要がある。評価では、単に技能や手続き、計算した結果を重要視するのではなく、主要なアイデアを理解することを重要視すべきである。これは、コースの中で使われた評価（例：小テスト、中間考査、小規模なプロジェクト）や、総括的評価（コースの成績）で行われるべきである。有用で、タイミングよく行ったフィードバックが、学習を促すのに必要不可欠である。異なったタイプのコースの中では、評価の種類によって実用的なものもあれば、そうでないものもあるだろう。しかし、大規模クラスにおいても適切な評価を実行することは可能である。

### 評価のタイプ

- ・ 宿題
- ・ 小テストと定期試験
- ・ プロジェクト

- ・活動
- ・口頭での発表
- ・記述式のレポート
- ・詳細なレポート
- ・記事の批判

### 教師への提言

- ✓ コースの必須の構成要素として、評価を組み込みなさい。教師がクラスでしていることとうまく合った評価を行うことは、2週間前にクラスでやったことに焦点を置くことより、もっと効果的である。
- ✓ 学生の学びをもっと完全に評価をするために、さまざまな評価方法を使用する。
- ✓ メディアのニュース記事やグラフについて解釈したり、批判したりするような評価を用いることで、統計的リテラシーを評価する。
- ✓ 学生のプロジェクトや、決まりきった答えはない調査タスクのような評価を用いて統計的思考力を評価する。

### 大規模クラスにおける学生の評価に関する提案：

- ✓ 個別のプロジェクトの代わりに小グループによるプロジェクトを使用しなさい。
- ✓ 成績を出す前に、フィードバックを行い、プロジェクトを完全にするために仲間同士の評価を利用しなさい。
- ✓ グラフの適切な解釈を選んだり、妥当な統計手順を選んだりすることを重視している項目を使用しなさい。
- ✓ 学生の口頭発表のために討論の時間を活用しなさい。

### 実際に行うために

統計教育はフィッシャーとスネデガー以来、大きく発展してきた。さらに、世界中の統計学の教師たちは、概して現代的な取り組みや教授法を採用することに熱心である。それにもかかわらず、私たちが教え方を変えるのは必ずしも容易ではない。ある意味で、私たちはみんな教師であるとともに、学習者でもあり、少しヤドカリに似ている。成長するためには、慣れ親しんだ守りの殻をまず脱ぎ捨て、新しい、より大きなひとそろいの習慣と可能性に落ち着くまで、傷つきやすい時期を耐えなければならない。

私たちは、このレポートで多くのアイデアを提示してきた。最初はまず小さなステップを踏み、提案された方向に動くよう読者に忠告する。

小さなステップの例として以下のようなものがある：

- ・コースに活動を加える
- ・学生に小さなプロジェクトをさせる

- ・アプレット（小規模プログラム）を講義に組み入れする
- ・ソフトウェアの使用を学生に実演してみせる
- ・本物のデータセットの使用を増やす
- ・概念の理解により重点を置くために、最近扱おうとしている内容のリストからいくつかのトピックを削除する。

教育に関する哲学は教科書の選定に影響するだろうが、このレポートでの提案は教科書の選び方の提案ではない。それらは教え方に関するものである。入手できる情報源はたくさんある。MAA ノートの中の統計教育に関する巻、発展的な学部教育コンソーシアム (CAUSE) ([causeweb.org](http://causeweb.org))、統計関係討論リスト ([www.lawrence.edu/fac/jordanj/isostat.html](http://www.lawrence.edu/fac/jordanj/isostat.html))、SIGMAA – MAA 内の統計教育グループ ([www.pasles.org/sigmaastat](http://www.pasles.org/sigmaastat))、そして ASA ウェブサイト、特に統計学教育センター ([www.amstat.org/education](http://www.amstat.org/education))、そして統計教育セクション ([www.amstat.org/sections/educ](http://www.amstat.org/sections/educ)) などたくさんある。

### 未来に向けて GAISE する

これまで多くの改革がなされてきたが、統計学入門コースを改善する余地はまだたくさんある。さらに、小中高等学校で統計学の見方を学んで大学に入ってくる学生が増えているので、このコースは柔軟性を持たせ、変えやすいように調整しなければならない。AP のコースは統計教育の視野を変え続けている。ここでは、一般的な入門コースについて述べてきたが、ビジネス統計、数理統計のような他のコースについても考える必要があり、入門コースの次に行われるコースについてもその内容や到達目標として、入門コースで育成したことをベースとして、しっかりとした概念理解が築かれるものにしていくよう心を配らなければならない。

(脚注)

- (1) ジョージ・コブ 変革を求める声に注意せよ：カリキュラム実行に対する提案 (MAA 文書 No.22), 統計学教育の章, 3-43 ページ. アメリカ数学学会 1992 年 ワシントン DC
- (2) ジョージ コブ 変革を求める声に注意せよ：カリキュラム実行に対する提案 (MAA 文書 No.22), 統計学教育の章, 3-43 ページアメリカ数学学会. 1992 年 ワシントン DC
- (3) ジョージ コブ 統計教育再考；国立科学基金会議、統計教育ジャーナル、1 (1)、1993 年 URL <http://www.amstat.org/publications/jse/v1n1/cobb.html>
- (4) デヴィッド ムーア 新教育学と新しい内容：統計事例 国際統計学評論、65:123-165,1997
- (5) ジョアン ガーフィールド 統計学の教育改革の査定 国立科学基金への最終報告 2000 URL <http://education.umn.edu/EdPsych/Projects/Impact.html>
- (6) ジョージ コブ 変革への要望に耳を傾ける：カリキュラムの実行への提案 (MAA 記録

- No.22) 統計学教育の章 3-43 ページ 1992 年 ワシントン DC アメリカ数学学会
- (7) イッド ガル 大人の統計応用：意味・構成要素・責任 国際統計学批評 70:1-51 2002
- (8) D J ラムゼイ 統計学入門コースの目的としての統計応用力 統計教育ジャーナル 10(3) 2002 URL <http://www.amster.org/publications/jse/v10n3/rumsey2.html>
- (9) ジェシカ アッツ 教育を受けた国民は統計学と確率について何を知るべきか？ アメリカの統計学者 57 (2) : 74-79 2003
- (10) ジョージ コブ 変革への要望に耳を傾ける：カリキュラムの実行への提案 (MAA 記録 No.22) 統計学教育の章 3-43 ページ 1992 年 ワシントン DC アメリカ数学学会



## 付録 A： 活動と計画例

### クラス活動の望ましい特徴

- 活動は現実社会をまねたものであるべきだ。それは「忙しい仕事」のように見えてはいけない。例えば、2項式の実験を行うためにコインやカードを使うのなら、それらが表す現実社会の2項式の実験を説明しなさい。z
- クラスは、活動をどのように行うのかについての決定に幾分関わるべきである。ステップの詳細な作り方に従うだけでは多くは学べない。
- クラスで決定したことは、授業で学んだ知識を必要とするものであるべきだ。例えば、学生が実験を計画しているなら、どうやって実験するのか直感的に決めるのではなく、授業で学んだ適切な実験計画の法則を考慮するべきだ。
- もし可能なら、学生が活動中に全体のプロセスがわかるように、活動に企画、データ収集と分析を含めるべきである。
- どのように活動を計画するのか討論するために学生をチームで取り組ませ、どのように行うのか討論するためにもう一度集合させるほうがよい場合もあるが、クラスで一緒に最初の計画やその他の決定に取り組ませた方がいい場合もある。それは、話し合う問題がどの程度難しいか、また、それぞれのチームが全く同じ方法で行う必要があるかどうかにもよる。
- 活動は、何が、なぜ行われているのかという概観で始まり、終わるべきである。
- 活動は楽しいものでなければいけない！

### 改善できる活動

#### ペプシ対コカコーラ活動

今日、私たちはペプシとコークのどちらの方がおいしいのかテストする。グループを4つに分けなさい。実験者として、グループからひとり選びなさい。注：もし自分が実験者でないなら、教室の前でできることは控えなさい。

1. 教室の前のテーブルの上に、ふたつの大きな炭酸の入ったびんがある。ひとつはペプシでもうひとつはコークのびんである。A, B とラベルのついたカップもある。実験者はテーブルの方に行って、コインを投げなさい。もし表が出たら、カップ A にペプシを注ぎ、コークはカップ B に注ぎなさい。どちらがどちらか覚えておきなさい。カップをチームに持って帰りなさい。
2. チームのメンバーに両方味見をさせなさい。どちらの方が好きか記録しなさい—カップ A なのかカップ B なのか。
3. 実験者はチームのメンバーに、より好まれたのはコークなのかペプシなのか、もう教えなさい。
4. 実験者はチームのメンバーひとりひとりに、このプロセスを一度ずつ繰り返しなさい。それから今度は、全部の生徒が一度はそれを行えるように、他のチームのひとり

が実験者に味見テストをこなさい。

5. クラス全体で集合する。教師が生徒のうち何人がコークの方を好きなのか尋ねる。
6. 教科書で公式を調べ、比率の信頼区間を求めなさい。クラスの中でコークの方が好きな生徒の比率を信頼区間で作図しなさい。
7. どちらの飲み物方がクラスでは好まれたか、仮説検定をこなさい。

## 中心極限定理活動

この演習の目的は、中心極限定理を証明することである。この定理が私たちに伝えているもの何なのか覚えておきなさい。それは、大きな被調査者母集団の意味は：

- ・ ベル型に近い
- ・ 母集団と同じ意味を持つ
- ・ 母集団基準の偏差  $\sigma/\sqrt{n}$  に等しい標準偏差を持つ。

批評：テストは二重盲検法ではない。実験者がどの飲み物がどれなのかわからないようにする理由はない。最初に実験を設置する人は飲み物容器にラベルをつけたりはずしたりでき、それらを飲み物1、飲み物2と呼ぶことができる。それから飲み物はA,Bとラベルを貼られたカップに前もって用意される。提示する順番はそれぞれの味見をする人により、無作為に行うべきである。

批評：これは、少なくとも2つの理由から良い活動とは言えない。まず、それは全く現実社会の動機を持っていないし、統計は退屈だという神話を増強してしまう。次に、指示が完全すぎる。

学生側の調査の余地がない；彼らは単に従うべき「方法」を与えられているだけだ。

中心極限定理が当てはまることを証明するために、これらの指示に従ってください。

1. ペアに分かれなさい。各ペアでひとつサイコロを持つ。
2. 交替でサイコロを25回ずつ振るので50回振ることになる。上になった数字を毎回把握しておきなさい。
3. 結果のヒストグラムを書きなさい。サイコロの表面は等しく似かよっている。だから、ヒストグラムも均一の形になるはずである。そうなることを確かめなさい。
4. 50回振った中間値と標準偏差を求めなさい。
5. ひとつのサイコロを振った時の中間値と標準偏差は、それぞれ3.5と1.708である。自分たちの50回振った中間値は3.5に近いか？標準偏差は1.708に近いか？
6. クラス全体で集合しなさい。50回の中間値を持つはずである理論上の曲線を描きなさい。それが釣り鐘型で母集団の中間値と等しい中間値を持つということ覚えておきなさい。だから、この場合は3.5であり、標準偏差は  $1.708/\sqrt{50} \approx .24$  になるはずである。
7. 各組のペアに、自分たちの50回の中間値を曲線上に印をつけさせなさい。中心極限定理を使って予想されることを考慮に入れると、それらが道理にかなっているかどうか注目しなさい。

## この活動をどう改善すべきか？

統計学を基にした活動（シェーファーおよびその他）からの出典である「セントと中心極限定理」活動は、より指針に沿った中心極限定理を例証するための例を示している。他にも活動を基本とした統計学からの良い例がある：

- ・ 仮説検定活動（ひと組のトランプから無作為にカードを引き、常に同じ色を引く）への導入はうまく機能する。
- ・ グラフを変数に合わせるのは多くの議論と学習を生む。
- ・ ふぞろいの四角形が、当然のことながら標準になった。
- ・ 入門コースでは、無作為に抽出された返答は主要ではないが、それは統計学的な思考と幾分かかわりがある。

## 活動とプロジェクトの追加例

### データ収集と分析： クラスのプロジェクト

ここで述べられているプロジェクトのようなプロジェクトの考えは、ロバート ワードロップの統計学：流動性の存在における学習（アイオワ デュビューク：ウィリアム C. ブラウン 1995）から由来している。これらのプロジェクトは、同様に認知心理学者、ダニエル カーネマンとエイモス トバルスキーによる研究に基づいている。

「司令官のジレンマ」のふたつの解釈について考えなさい。

第一版：優勢な敵の軍に脅され、司令官はジレンマに陥る。彼の部下である諜報部の将校たちが、ふたつ使える経路のうちひとつを使って、兵士たちを安全なところに導かない限り、彼らは待ち伏せされて捕えられ、彼らのうちの600人は死んでしまうだろうと言った。ひとつ目のルートを通れば200人は救える。ふたつ目のルートを通れば三分の一の確率で600人の兵士たちが助かり、三分の二の確率で誰も助からない。彼はどちらのルートを取るべきか？

第二版：優勢な敵の軍に脅され、司令官はジレンマに陥る。彼の部下である諜報部の将校たちが、ふたつ使える経路のうちひとつを使って、兵士たちを安全なところに導かない限り、彼らは待ち伏せされて捕えられ、彼らのうちの600人は死んでしまうだろうと言った。ひとつ目のルートを通れば400人は死ぬだろう。ふたつ目のルートを通れば三分の一の確率で兵士たちは誰も死なない、三分の二の確率で600人が死ぬ。彼はどちらのルートを取るべきか？

質問のふたつのバージョンは、ふたつの同じ答を持っている；両方とも同じ状況を述べている。ふたつの質問は言葉の使い方が違うだけである：ひとつは亡くなる命について述べ、もうひとつは助かる命について述べている。

この形式のひとつの質問は、ふたつの質問を「問題の論じ方」で比較する単純な無作為抽出の実験へと容易に導く。ひとつのテーマを形成し、無作為の数字の表を使ってそれらをふたつのグループに分類し、質問のひとつのバージョンをそれぞれのグループに割り当てなさい。結果は、 $2 \times 2$ の数表にまとめられる。

質問	答え	
	A	B
バージョン1		
バージョン2		

データはふたつの比率を比較することによって分析できる。例えば、フィッシャーの直接確率検定か、連続性の補正とカイ二乗検定を使う。

ワードロップの本の中にある、集合1、2の練習問題にはこの構造について数多くの種類が載っており、それらの多くは学生によって行われる。ここに4つだけ簡略されたものがある。

- 歴史博物館の人に、歴史の本のある特定の主張の中で説得力のあるものを見つけたか尋ねてみなさい；主張は支持するデータの表がある／ない。
- 学生会の女性に、ある見知らぬ男性が近づいてきてお酒を飲みましょうと誘われたら行くかどうか尋ねなさい；男性は魅力的／魅力的ではないと表現される。
- ソフトクリームを注文している客に、普通のがよいか、ワッフルコーンがいいか尋ねなさい；ワッフルコーンは「自家製」である／「自家製」ではないと述べられる。
- 大学生に、(1)落ち込んでいる友達にカウンセリングを受けることを勧めるか？それとも、(2)自分が落ち込んだらカウンセリングを受けに行くか？

ふたつの答があるふたつのバージョンに基づくプロジェクトは、利点を多く持っている：

- ・ データ収集が、適切な時間内に完成できる。
- ・ 無作為抽出は、結果が正式な推論にふさわしいということを保証する。
- ・ 無作為抽出は、データ収集の可能性と分析の確率モデル使用とのつながりを明白にする。
- ・ 分析の方法は、比較的単純でわかりやすい。
- ・ 構成（2×2 数表）は、たいへん幅広く応用が利くものである。
- ・ 最後に、書式は無制限で、それによって学生は幅広いアプリケーションの範囲を得ることができ、テーマやペアの質問を選ぶ際に、そこから構想や独創性を活かす十分な機会が得られる。

### チーム構成の関係についての質問

これらの助言は、教師向けのものである。学生向けの助言はプロジェクト4 チーム形成にある。

目標：学生に研究用の疑問を作らせ、それに答える手助けとなるように、データを収集し説明する経験を与えなさい。

与えるもの：(N=学生数； T=チーム数)

- ・ N枚の見出しカード、あるいはそれぞれのT色のカラー（またはボードのスペースを

使用；下を見なさい)

- ・ T枚 あるいは 2T枚の透明シートとペン(2T枚の理由となるステップ3を見なさい)
- ・ T個の 計算機

学生は4~6人のチームで取り組むべきである。プロジェクト例4 チーム形成を見なさい。

**ステップ 1：**各チームは、関連があるかどうか知りたいふたつの分類別の変数を明確に述べる。例えば、誰かが長男（長女）（あるいはひとりっこ）で、屋内の活動を好むか屋外の活動を好むか（最近の調査では、第一子は屋内の活動の方を好み、下の子供は屋外の活動の方を好むということを示している）；男性／女性と何かについての意見；クラス（上級生／下級生等）と、車を持っているかどうか等。時間内に容易に終わらせるため、ひとつの変数につき、ふたつの分類に制限したいと思うかもしれない。

データ収集には、ふたつの可能な方法がある一見出しカード（または紙片）あるいはボードを使用する。次の数段階はそれぞれ両方の方法で述べられるだろう。

**ステップ 2：**各チームは、チームの数の色の中から、ひとつの色の見出しカードを割り当てられる。例えば、チーム1は青かもしれないし、チーム2はピンクかもしれない、等々。

ボード：各チームに、疑問を書くためのチョークボードのスペースを割り当てなさい。

**ステップ 3：**各チームはクラス全体に自分たちのふたつの疑問を尋ねなさい。カード：チームは質問を透明シートに書いて皆に見せる。各チームが交替で教室の前に行く。学生はチームカラーに応じて答を見出しカードに書き、それらを集める。例えば、クラスの学生全員が、自分たちの答えをチーム1の質問にはブルーの見出しカードに、チーム2の質問にはピンクの見出しカードに、等々。ボード：チームのメンバーは、各学生が適切なセルに#の記号を書けるようになっている2方向の表とボードに、質問を書く。

プロジェクト2.2から採用、指導者の教材マニュアル、統計学に注意せよ

アツとヘッカード

注：これはひとつの分類的、そして定量的変数とふたつのサンプルの推測をする  
時のために保持されたデータで行うこともできる。

**ステップ 4：**カード：各チームが質問をして学生たちが自分たちの答を書いた後、カードが集められ、それぞれのチームに渡される。例えば、チーム1は青のカードをすべて受け取るというように。ボード：クラス全員がボードの自分たちの割り当てのところに行き、#の記号を自分たちに合った表のセルのところに書く。

**ステップ 5：**各チームが集計し、要約し、自分たちの質問に対するデータのグラフを画面に表示する準備をする。結果は透明シートに書かれる。

**ステップ 6：**各チームが結果をクラスで発表する。

**ステップ 7**：そのデータが大きな母集団から無作為に抜きとったものだということにした  
いのだったら、カイ二乗検定で独立したものとして扱う時のために、結果を取っておくこ  
とができる。

## プロジェクト 4： チーム形式

チームメンバー：

- |          |          |
|----------|----------|
| 1. _____ | 4. _____ |
| 2. _____ | 5. _____ |
| 3. _____ | 6. _____ |

指 示：

1. クラスのメンバーに面白い関係があるかもしれないと思う分類の変数をふたつ考えな  
さい。もしそうしたければ、例えば、GPA-高低（ $GPA \geq 3.0$  のような切り捨てを使って）  
のように、定量的変数を分類の変数に変えてもよい。それぞれの変数は、割り当てられた  
時間内に終わらせられるように、ふたつの分類を持つべきである。

2. どちらが説明変数なのか、どちらが反応変数なのか組み立てながら、自分の状況で意  
味を通るなら、下のふたつの変数をリストに入れなさい。

- ・ 説明変数：
- ・ 反応変数：

3. 各チームがチョークボードに場所を割り当てられる。ひとりのチームメンバーがボー  
ドのところに行って、自分たちのふたつの質問を書かなければならない。また、皆がそれ  
に答えられるよう、#の記号を書きこむ四角のある「2方向」の表も書きなさい。

4. 全員がボードのところに行って、各チームの質問に対して適切な四角に#の記号を書  
く。

5. 皆がボードのところに行ってデータを記入したら、自分のチームの質問に対して下の  
表に合計を入力しなさい。また、それぞれの変数に対して何の分類なのか入力する。

反応変数

説 明 変 数	分類 1	分類 2	合 計
分 類 1			
分 類 2			
合 計			

6. 自分のチームの透明シートに表示できるように、適切な数とグラフの要約を作りなさい。必要があれば、発見したことの手短な要点を下と後ろに書く。

7. チームの中からひとりが、透明シートを使ってチームの結果をクラスで発表する。

8. このシートと透明シートを提出する。

### ふたつの条件のもと、手先の器用さを比較する

これらの指示は、教師向けである。学生向けは「プロジェクト5 チーム形成」にある。

目標：学生に計画、遂行、分析、実験という経験をさせなさい。

与えるもの：(N=学生の数、 T=チームの数)

- ・ ふたつの色の違う乾燥豆がそれぞれ 30 粒くらい入ったボウル T 個
- ・ 2T 個の空の紙コップまたはボウル
- ・ T 個の ストップウォッチ あるいは秒針のある時計

話：ある会社にはふたつのタイプの小さな部品を仕分けするのを仕事としている従業員が大勢いた。従業員たちはけがを繰り返しがちであった。それで会社は、従業員たちが入れ替わったら、能力の高い人を使ったり、能力の低い人を使ったりして生産性で大きな損失が出るだろうかと考えた。(あるいは、もしゴム手袋を使っているのなら、話はそうなりうるだろう。健康上の理由で彼らは手袋を必要とするだろう) それゆえ、この比較をする実験を計画し、遂行し、ぶんせきするのである。学生たちは、2色の豆をボウルから色分けしてふたつの紙コップに移すのにどの位時間がかかるのか見るためにタイムを測る。比較は利き手とそうでない手を使った後に行われる。代替案は、例えば 30 秒というように時間を決めて、いくつの豆がその時間内に移せたか見るというものである。

**ステップ 1**：クラスで、どのように実験を行うのか話し合うこと。これは、まずチームで行う。下の提案をみること。

- ・ 処置は何か？実験の単位は何か？
- ・ 考慮すべき実験計画の方針は以下のとおりである。実験を計画し実行する際に、できるだけそれらを多く使用しなさい。どうしてそれらが使われるのか話し合いなさい。
  - つり合ったペアの作成と区画
  - 実験単位に対する処置法の無作為抽出、あるいは処置法の順序の無作為化
  - 目隠し試験、または二重目隠し試験
  - グループの制御
  - 偽薬

プロジェクト 12.2 から採用、指導者の教材マニュアル、統計学に注意せよ  
アツツとヘッカード

注：変分は、熟練の人とそうでない人の代わりに、ゴム手袋をつけて仕事をさせるのと、つけないでさせること。その場合、ゴム手袋のN組のペアが必要になる。

➤ 学習の影響あるいは疲労

- ・ 興味の媒介変数は何か
- ・ どのような分析のタイプが適切か—仮説検定、信頼区間、または両方

各生徒が自分の手で一度タスクを完成させるように、クラスで決定すべきである。なぜこれが、クラスを半分に無作為に分けて、クラスの半分は利き手を使い、もう半分が利き手ではない方を使うより望ましいのか？ 順番はどのようにして決められるのか？ すべての生徒にそれは同じであるべきなのか？ 練習は許されるのか？ 目隠し試験や二重の目隠し試験は可能なのか？

**ステップ 2：チームに分かれて実験を実行すること。**

プロジェクト 5 チーム形成は、チームメンバーにタスクを割り当てるひとつの方法を示している。

**ステップ 3：記述統計学と推論の準備**

クラス全体で集合し、違いの幹葉図を描きなさい。この分析の必要条件が見合っているかどうか話し合いなさい。異常値はあったか？ もしそうなら、それらは説明がつくか？ その違いに対する中間値と標準偏差を、誰かに計算させなさい。

**ステップ 4：推論**

チームで集合させなさい。各チームは、中間値の相違に対して信頼区間を求め、仮想検定を実行させること。

**ステップ 5：クラスで再集合し、結論について話し合うこと。**

**被調査者母体で実験を計画し分析する方法についての提案**

**プロジェクト 5**

**計画の問題：**

- つり合ったペアの作成と区画：各学生はペアで、自らの手でタスクを一度行うように、使用されなければならない。
- 実験のユニットの処置の無作為抽出、あるいは処置の順序の無作為抽出：それぞれの学生に対してどちらの手を使うのかについて、無作為に抽出する。
- 目隠し実験か二重目隠し実験か：明らかに、学生はどちらの手が使われているのか知っているが、タイムキーパーは知る必要はない。
- グループ制御：この実験には関係がない。



- 偽薬： この実験には関係がない。
- 学習

効果あるいは疲労：学習効果はありそうなので、練習問題を数回分作りたくなるかもしれない。また、各生徒の両手の順番を無作為に抽出することが、このことに役立つだろう。

### 可能な計画の例：

学生にコインを投げさせる。表が出たら利き手から。裏が出たら利き手ではない方から。豆を分離させるのにどの位時間がかかるのか見るために、時間を測る。時間を計測する学生は、見ないことで、状況に対して目隠しをしていることにできる。

### 分析：

- 興味の要素は何か？

答え：それぞれの人の関心の確率変数を、以下の「手先の器用さの違い」であると定義する。

$d$  = 利き手ではない方の手で、余計に必要とした秒数

$d$  = 聞き手ではない方の手でかかった時間－利き手でかかった時間

Define  $\mu_d$  = 母集団中間値 手先の器用さの違い

- 帰無仮説と対立仮説は何か？●

$$H_0: \mu_d = 0$$

$$H_A: \mu_d > 0$$

(利き手でより速い)

- 信頼区間は適切か？●

そう、それは、従業員が利き手だと、どの位早く仕事を達成できるかについての情報を

$$\bar{d} \pm t^* \left( \frac{s_d}{\sqrt{n}} \right)$$

与える。信頼区間の公式は、 $\bar{d} \pm t^* \left( \frac{s_d}{\sqrt{n}} \right)$  となり、そこでは  $t^*$  が  $df = n - 1$  で、臨

界  $t$  値であり、 $s_d$  が点差の標準偏差となる。検定を行うために、 $t = \frac{\bar{d} - 0}{s_d / \sqrt{n}}$  を計算し、確率値を求めるために、臨界  $t$  値と比較しなさい。

## プロジェクト 5： チーム形式

チームメンバー：

- |          |          |
|----------|----------|
| 1. _____ | 4. _____ |
| 2. _____ | 5. _____ |
| 3. _____ | 6. _____ |

## 指 示：

チームで取り組む。各チームは、豆の入ったボウルと2つの空のカップを取る。豆をボウルから空のカップに、色分けして移す。どの位かかるか見るために、タイムを測る。これを、それぞれの手で各2回行う。それぞれの手で一度、順序は無作為に決める。

- これらの作業を計画する。望むなら、それぞれの回で作業を入れ替えてもよい。
  - 調整係 — 実演する
  - 無作為抽出係 — どちらの手から始めるか、人が代わるたびにコイントスで決める。
  - タイムキーパー — タスクを行うそれぞれの人のタイムを測るため、秒針のある時計を持っていなければならない。
  - 記録係 — 下の表に結果を記録する。
- 誰から始めるのか決める。無作為抽出係は、どちらの手で先に行うのか告げる。最初の人々が2番目の手に変える前に、それぞれの人が一度タスクを完成しておくべきである。そうすると、手を変える前に皆が休む時間ができる。
- タイムキーパーは、ボウルからカップに色分けしてひとつずつ豆を移す間、その人の時間を測る。
- 記録係は時間をメモし、表に記録する。
- これを各チームメンバーについて繰り返す。
- その後、それぞれの人が、1回目に使わなかった方の手で2回目を行う。
- それぞれの違いを計算する。

## クラスの結果

ここにデータを記録する：

名 前	利き手ではない方の手のタイム	利き手のタイム	d = 違い 利き手ではない方 - 利き手

テストされ、予想される母集団特性値は：

信頼区間：

仮説検定 — 推測と結果

## 付録 B： 評価項目の例

難易度の、特質に関する問題と解説を伴った、評価項目の例をいくつか挙げる。

まず、難易度の本質に関する問題と解説を伴う、評価項目の例をいくつか取り上げる。

検定では使用を避ける評価項目：正/誤、背景や説明なしで計算のみ、計算や分析を入力するにはデータが多すぎる項目、定義や公式の記憶を試すだけの項目。

### 項目 1

ある教師は、前期で初歩的な統計の 2 セクションを教えた。午前 8 時のクラスの期末考査の中間値と標準偏差は、78 と 8 で、午後 4 時のクラスは 75 と 10 であった。これらの数字をよく調べると、より出来の良い学生は、恐らく午前 8 時の方に届け出していると教師は考える。それで彼女は、自分の学生の 2 つのグループの期末考査の点数の中間値は同じかどうか、テストすることにした。

### 項目 2

ある経済学者は、男性と女性の重役の月給の中間値を比較したい。彼は、それぞれから 10 の被調査者母体を無作為に取り、t 検定を行う。結果の確率値は 0.45 である。

1. 帰無仮説と対立仮説を述べなさい。
2. 統計的結論を出しなさい。
3. 統計学の学習をしていない人にもわかるような言葉で、自分の結論を述べなさい。

批評：教師はすべての母集団のデータを持っているので、統計的な推量は必要ない。

批評：質問は t 検定の必要条件を述べておらず、被調査母体の規模が小さい。条件はここでは、ほとんど守られていない。月給はほぼ確実に非対称である。

### 項目 3

次のうちどれが、確率値の限定をするのか？

- A. 無効仮説が当てはまるなら、それは無効仮説を拒否する可能性である。
- B. 無効仮説が当てはまるなら、それは無効仮説を拒否しない可能性である。
- C. 観察されたものと同じくらい、それは極端な観察データの可能性である。
- D. それは無効仮説が当てはまる可能性である。

## 評価項目を改善する方法を示す例

正/誤項目は、うまく書かれても、その話題について何も知識はなく、項目を正しくする可能性は常に 50% あるので、学生についての情報をあまり多く与えない。ある最新の取り組みは、項目を 3 つ、あるいはそれ以上の選択肢による、強制選択質問に変えることである。

#### 項目 4

集合データの標準偏差の大きさは、中心がどこにあるのかによる。正しいか、正しくないか？

こう変える：

集合データの標準偏差の大きさは、どこに中心があるのかによるか？

- A. そうである。中間値が高ければ高いほど、標準偏差は高くなる。
- B. そうである。標準偏差を計算するために中間値を知る必要があるから。
- C. そうではない。標準偏差の大きさは、分布の位置に影響されない。
- D. そうではない。標準偏差の大きさは、中央値とどの位違うのかではなく、値がどの位異なっているのかしか測れないから。

批評：これらのうちどれも、完全に正しいわけではない。BとDの答えは明らかに間違っている。Aの答えは重要レベルである。Cの答えは継続すれば正しい（が、可能性は低い）だろう。「... あるいはさらに極端になるだろう、無効仮説が当てはまると仮定すれば。」

#### 項目 5

+1の相関関係は、-1の相関関係よりも強い。正しいか、正しくないか。

このように書き変える：

教育研究ジャーナルに載っている最近の記事に、数学の成績と総合的数学の適性の間の相関関係は、+0.8であると報告している。また、数学の成績と数学の不安テストの間では-0.8の相関関係であると報告している。次のどの解釈が最も正しいか？

- A 相関関係+0.8は、相関関係-0.8より強い関係を示している。
- B 相関関係+0.8は相関関係-0.8とちょうど同じくらい強い。
- C どちらの方が相関関係が強いとは言えない。

**背景は、現実の状況で、学生が統計学的考えを知ったり扱ったりする手助けとするのに重要である。**

#### 項目 6

XとYが深く関連づけられていることが、いったん証明されると、Xの変化がYに変化を引き起こすことを立証するためには、どんな形の研究をすることが必要になってくるのか？

背景が付け加えられる：

ある研究者は実験的薬と T<sub>4</sub>リンパ球細胞レベルの関係を、HIV/AIDS 患者について研究している。T<sub>4</sub>リンパ球は免疫システムの一部で、HIV 感染者にそれが低下したときに見られる。その 2 つの変数、薬の服用量、T<sub>4</sub>細胞レベルが高く関連しているといったん証明されると、服用量の変化が T<sub>4</sub>細胞レベルでの変化をもたらすということを証明するために、どんなタイプの研究が必要となるのか？

- A. 相関関係研究
- B. 制御実験
- C. 予想研究
- D. 調査

**概念や解釈を損ない、学生にとって問題の重点となってしまうかもしれない、繰り返し、またはうんざりする検査の計算は避けるよう努める。**

#### 項目 7

一年目の教科課程のコースでは、メジャーリーグスポーツの放送権における、展開の経過を分析するために、ダーウィンの法則を適用するよう学生に求める、20 点のレポート問題を含んだ最終試験を使用した。コース内の 4 人の教授の採点に一貫性を調べるため、6 段階のレポートのサンプルが無作為にそれぞれの教授から選ばれた。点数は下の表にまとめられている。4 人の指導者の中間値における違いをテストするために、分散分析を行いなさい。

指導者	点数					
アフィンジャー	18	11	10	12	15	12
ビューリユー	14	14	11	14	11	14
クリアリー	19	20	15	19	19	16
ディーン	17	14	17	15	18	15

一年目の教科課程コース… (上と同じ)…。点数は、全体のサンプルに対する記述的統計学と一方向の分散分析の出力と共に、下の表にまとめられている。

#### 記述的統計学

変数	N	Mean	Median	TrMean	StDev	SEMean
点数	24.00	15.00	15.00	15.00	2.92	0.60

#### 一元配置分散分析

\*\*\*分散分析表 除外\*\*\*

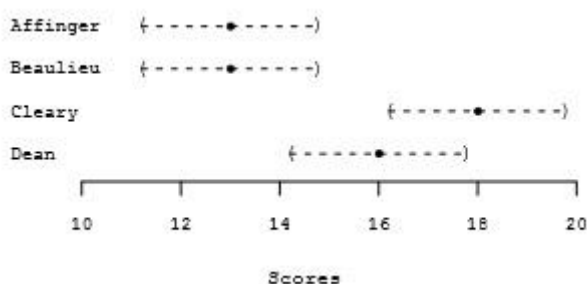
レベル	N	Mean	StDev
-----	---	------	-------

アフィンジャー	6	13.00	2.97
ビューリユー	6	13.00	1.55
クリアリー	6	18.00	2.00
ディーン	6	16.00	1.55

統合標準偏差 = 2.098

批評：上の質問形式には、結果を得るためにかなり計算機を使う必要があるし、解釈を問うということもあり得ない。下の改訂版はそれでもまだいくらか計算を必要とする（それは与えられる計算の出力量により調整できる）が、計算は比較的効率よくできる。与えられる計算の出力が何であるのかよくわかっている学生なら特に。

統合標準偏差に基づく中間値の各 95%信頼区間



1. 残念ながら、出力からの分散分析表がないので、それを構成するために上記の情報を使い、4人の指導者たちに割り当てられた平均点における大きな違いをテスト（5%レベルで）する。仮定と結論を必ず含むこと。残りの部分（または次の質問）を完成するのに必要である表の一部を作るのにてこずったら、筋の通った推量をするか援助（要点だけの料金で）を求めなさい。
2. 分散分析表ができたなら、アフィンジャー博士から与えられた平均点の、95%信頼区間を構成しなさい。注：解答はグラフが示すものと一致していること。

## 良い評価項目の追加例

### 項目 8

ひとりの学生が1学期中に教科書に費やす金額を、Yで表しなさい。ナンシーが統計学を心得ていて、秋の前期1と春の後期2がどの位なのか知りたいと仮定して、それを比較しなさい。特に、彼女は？1の平均と？2の平均に関心があると仮定しなさい。ナンシーはいくつか統計学コースを取っていて、信頼区間や仮定検定をどう解釈するかも含め、統計学をよく知っていると想定されるかもしれない。それぞれの学期から無作為にサンプルを抽出し、データを分析してレポートを書きなさい。次の4人の人たちのアドバイスを聞きな

さい：

1. ラッドはこう言う、「 $H_0 : \mu_1 = \mu_2$  vs.  $H_A : \mu_1 \neq \mu_2$  の  $\alpha = .05$  テストを行い、ナンシーに  $H_0$  を認めるべきではないか伝えなさい。」
2. リンダはこう言う、「 $\bar{y}_1 - \bar{y}_2$  の 95% 信頼区間を出しなさい。」
3. スティーブは言う、「 $H_0 : \mu_1 = \mu_2$  vs.  $H_A : \mu_1 \neq \mu_2$  のテストを行い、ナンシーにそのテストから出た確率値を報告しなさい。」
4. グロリアは言う、「 $\bar{y}_1$  と  $\bar{y}_2$  を比較しなさい。もし、 $\bar{y}_1 > \bar{y}_2$  なら、それから  $\alpha = .05$  を使って  $H_0 : \mu_1 = \mu_2$  vs.  $H_A : \mu_1 > \mu_2$  をテストし、ナンシーに、 $H_0$  を認めないかどうか言いなさい。もし、 $\bar{y}_1 < \bar{y}_2$  なら、それから  $\alpha = .05$  を使って  $H_0 : \mu_1 = \mu_2$  vs.  $H_A : \mu_1 < \mu_2$  をテストし、ナンシーに  $H_0$  を認めるかどうか告げなさい。」

4つのアドバイスを最も良いものから悪いものまでランクをつけ、なぜそのようにランクをつけたのか説明しなさい。つまり、なぜそれが他のものより良いのか説明しなさい。

### 項目 9

研究者たちは、ふたつの母集団から無作為に、被験者のサンプルを抽出し、そのデータにウィルコクソン・マン・ウィトニーテストを応用した；そのテストで、無指向性代替を使った確率値は.06. だった。以下に述べているものは、それぞれ正しいか正しくないか、またそれはなぜなのか答えなさい。

1. ふたつの母集団の分布が同じだという確率は6%である。
2. もしもふたつの母集団の分布が本当に同じなら、この研究者たちが観察した違いが同じくらい極端なこのふたつのサンプルの違いは、6%の確率でしか起こらない。
3. もし、このふたつの母集団を比較する新しい研究がされたら、 $H_0$  が再び拒否される確率は6%ある。
4. もしも  $\alpha = .05$  で指向性代替が使われ、データが代替仮説に特定された方向で分けられたら、 $H_0$  は拒否されるだろう。

### 項目 10

月曜のCNNのウェブページ ([www.cnn.com/HEALTH/9612/16/faith.healing/index.html](http://www.cnn.com/HEALTH/9612/16/faith.healing/index.html)) の記事は、このような文で始まる、「月曜に発表された調査によると、宗教的信仰は患者の治療を促すと、家庭医たちは圧倒的に信じている」と。後からその記事はこう述べる、「医学研究者たちは、宗教の利点はストレスを減らすことで免疫システムを助けるというとても簡単なことかもしれない」と。そしてハロルド・コーニング博士は、伝えられるところによると、こう述べている、「定期的に教会に通っている人は、うつ病にかかる割合が、あまり教会に行かない人の半分である」と。

コーニング博士が述べていることと、記事の示している、宗教的信仰と習慣がうつ病との闘いを助けるという主張を批判するために、統計学の用語を用いなさい。批評には次の単語のうち、いくつかを選ぶ必要があるだろう：

観察に基づく研究、目隠しの、二重目隠しの、精度、偏り、被調査者母体、擬似の、誤りを証明、因果関係、結合、無作為、有効な、信頼できる

### 項目 11

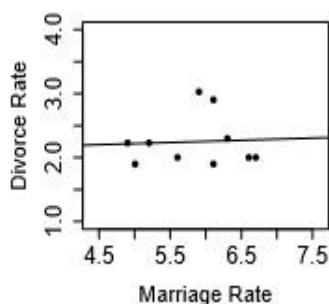
フランシスコ フランコ (98 年度クラス) は、100 個のハーシーキッス (アーモンド入り) の重さを測った。被調査母体の平均は 4.80 グラムであり、SD は.28 グラムであることを彼は発見した。この背景の状況で、平均の標本分布によって、中間値は何であるのか説明しなさい。

### 項目 12

ある園芸家は 3 つのタイプの豆の種、A タイプ、B タイプと C タイプの産出を比較したい。彼女は、A タイプの種を無作為に 3 つに分け、それぞれを庭の東の部分、中央部分、そして西の部分にいくらか植えた。その後、B タイプと C タイプも同様に行った。

1. この園芸家はどんな種類の実験計画を使用しているのか？
2. なぜ、この種の計画がこの状況で使用されるのか？ (背景の状況で説明せよ)

### 項目 13



この散布図は離婚率  $y$  と婚姻率  $x$  がどの程度、データ採集した 10 カ国の中で関連しているのかを示している。回帰線が図表に書きくわえられた。

1. アメリカはもともと集められた国々の中で 10 点の国ではない。たまたまアメリカの婚姻率が他の 10 カ国より高いというだけである。さらに、他の国々のパターンを考慮に入れると、アメリカの離婚率は予想より高い。アメリカを集合データに加えると、回帰線にどう影響するのか？ それはなぜだろう？
2. アメリカが集合データに加えられた後の散布図と回帰線について考えなさい。残余の図表の略図を書きなさい。三角形の図表に軸線を分類し、アメリカを特定しなさい。



#### 項目 14

研究者たちはふたつの薬、フォルモテロールとサルブタモールを、運動から引き起こされる喘息の患者の治療法として、噴霧溶液で擬似薬とで比較したいと思った。患者たちは薬か擬似薬を飲み、運動をし、「強制呼気量」を測ることになっていた。30人の被験者が有効だった。(A.N. トソイとその仲間たち、*ヨーロッパ呼吸器ジャーナル* 3 (1990) : 235 ; ベリーを通じて、*統計学：ベイジアン*の相関関係に基づく)

1. これは実験なのか、観察的研究とすべきなのか？ その理由は？
2. この背景の状況では、擬似薬の効果は何か？
3. ここでの乱塊法計画 (RBD) の始め方を手短かに説明しなさい。
4. 乱塊法計画はどのように役立つのか？つまり、このような状況で、RBD を使用する主な利点は何か？

#### 項目 15

私は、114 クラスから 8 人の生徒が、第 2 回目の試験 (4 月) の前の復習授業に出席していることに気がついた。これら 8 人の生徒たちの平均点は、復習授業に出席しなかった 21 人の生徒たちの平均点より低かった。この情報を復習授業の有効性における研究に、私が使いたいと考えていると想定しなさい。

1. これはどんな種類の研究か、観察的、あるいは実験的なのか？ それはなぜか？
2. ここで懸念されるのは、どんな種類のサンプリング誤差や標本抽出バイアスなのか？
3. (仮説) 私は友人のジョージに、出席した 8 人と出席しなかった 21 人のデータを渡した。ジョージは仮説検定を行うためにそのデータを使用した。仮説検定は意味があるのだろうか？もしそうなら、 $H_0$ は何か？そうでなければ、なぜなのか？

#### 項目 16

次の 3 つの状況について、それぞれ自分が行う分析のタイプを述べなさい (例 : 1 標本 t 検定、回帰、独立性カイ二乗検定、カイ二乗適合度検定等)、もしあなたが未補正のデータをすべて持っているなら。そして自分が分析を行う変数の役割を明細に述べなさい、しかし、分析を実際に行ってはいけない。

1. エリザベス ラーンツ ('02 のクラス) は、13 人の学生の脈について運動の影響を測った。彼女は運動 (30 回のジャンピングジャックをする) の前後に脈を計測し、変化の平均は 55.1 で変化の SD は 18.4 だということを発見した。このデータをどう分析するか？
2. 3 つの HIV 治療法について、子供の HIV の進行を防ぐ効果がテストされた。薬 A を与えられた 276 人の子供たちのうち、259 人が生き、17 人が亡くなった。B の薬

を与えられた 281 人のうち、274 人の子供たちが生き、7 人が亡くなった。C の薬を与えられた 274 人の子供たちのうち、264 人が生き、10 人が亡くなった。このデータをどう分析するのか？

3. ある研究者は前腕の長さとうりとの関係に興味を持った。彼は、16 人の女性に前腕の長さとうりを測った。これらのデータをどう分析するか？

### 項目 17

4 つの標本から、私はデータデスクに、データを平行な構成で作図させた。それから、 $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$  の検定を行い、 $\alpha = .05$  レベルで  $H_0$  を認めなかった。また、 $H_0 : \mu_1 = \mu_2 = \mu_3$  も検定し、 $\alpha = .05$  レベルで  $H_0$  を認めなかった。しかしながら、 $\alpha = .05$  を用いて  $H_0 : \mu_2 = \mu_3$  を検定した時、 $H_0$  を認めた。同様に、 $\alpha = .05$  を用いて  $H_0 : \mu_1 = \mu_4$  を検定したとき、 $H_0$  を認めた。

1. あなたの仕事は、データを平行な構成でグラフを描くことである。つまり、私がテストについて話したことに基づいており、データがどのように見えるのかという考えを持つべきである。グラフを描くのに、その考えを使いなさい。構想に加える三角形で標本の間値を示しなさい。
2. パート 1 でグラフ化した同じ標本の間値で、データを得ることは可能であるが、 $\alpha = .05$  レベルで  $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$  は認められるという仮定に対しては得られない。この背景のグラフを描きなさい。つまり、パート 1 から得た同じ標本間値（三角形）を保ちなさい。しかし、 $H_0$  が認められるなら、データがどのように異なっているのか示しなさい。

### 項目 18

アトレイ チャーク（'02 クラス）は、12 の朝食シリアルは無作為標本についてデータを集めた。彼は  $X$ =繊維（グラム/オンス）、 $Y$ =価格（セント/オンス）と記録した。データの散布図は、線形の関係を示している。適合回帰モデルは、

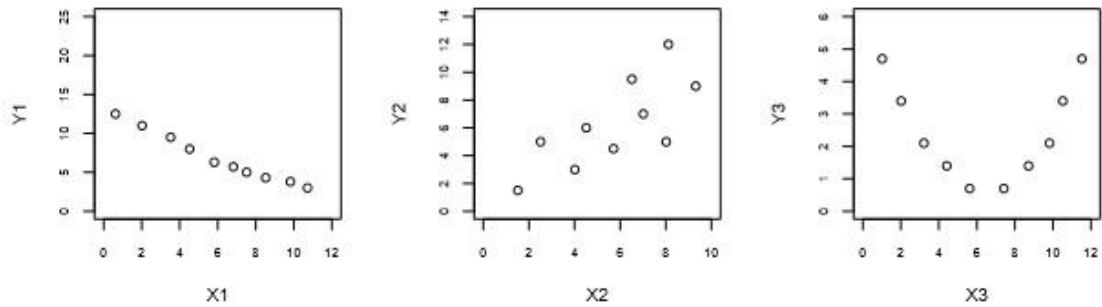
$$\hat{y} = 17.42 + 0.62x$$

サンプルの相関係数  $r$  は、0.23 である。 $b_1$  の SE は、.81 で、また、 $s_{y|x} = 3.1$  である。

1.  $r^2$  を求め、この問題の状況での  $r^2$  を説明しなさい。
2. あるシリアルには 2.63 グラムの繊維（オンスあたり）があり、17.3 セント（オンスあたり）の価格であると、仮定しなさい。このシリアルの残りはどうか？
3. この問題における状況で、 $s_{y|x} = 3.1$  の値について説明しなさい。つまり、 $s_{y|x} = 3.1$  であるというのは、どういう意味なのか？
4. この問題の状況では、「回帰の影響」によって何を意味しているのか説明しなさい。

### 項目 19

下のそれぞれの散布図におけるデータのサンプルの、相関関係を概算しなさい。



### 項目 20

整合実験では、ジョーズピザとドミノズピザの通常のチーズピザの味を比較する。それぞれの被験者は印のない 2 切れのピザ、各タイプから 1 切れ、無作為の順序で味見をし、どれを好むか述べる。この研究に参加している 50 人の被験者のうち、21 人がジョーズピザの方を好んでいる。

1. ドミノズピザよりジョーズピザの方を好む母集団の比率に対する 96%の信頼区間を求めなさい。
2. 望ましい信頼区間の誤差が 4%の場合、どの位の規模の母集団が必要か？

### 項目 21

心臓病学者 5 人に 1 人が、動脈硬化を防ぐため、一日に一錠アスピリンを飲んでいるという主張があった。この主張が事実だと仮定しなさい。もし 1,500 人の心臓病学者が無作為に選ばれたら、1,500 人のうち少なくとも 275 人が一日に一錠アスピリンを飲む確率はどのくらいか？

### 項目 22

散布図が、変数間の関係の概要を、視覚的に適切に示すのかどうか特定しなさい。それぞれの場合で、自分の推論を説明しなさい。

1. 血圧と年齢
2. 国の地域とより強い銃規制法についての意見
3. SAT (大学進学適性テスト) の言語と数学の点数
4. ハンドスパン (親指と他の指を開いた手幅) と性 (男性か女性か)

### 項目 23

次の文章は、それぞれいくつか統計学の分析を求める状況を述べている。それぞれに対して指示に従いなさい：

1. (下のリストより) 適用するのにふさわしい統計学の手順の名前をあげなさい。同

じ手順を 2 回以上使ってもよい。また、ひとつの質問に対して正しい答が複数あるものもあるかもしれない。

2. 問題の中では確率値も与えられる。その特定の問題に対して結論を導くのに、それを使いなさい。ただ  $H_0$  を認めないだとか、あるいは  $H_0$  を否認できないと言うだけではいけない。それ以上のことを述べなさい。(5%の重要度を想定すること)

---

信頼区間 (中間値 $p$ の…)	通常の分布
母集団規模の決定	相関関係
中間値テスト	単一線状回帰
比率テスト	複合回帰
中間値の差 (ペアデータ)	二方式図表 (カイ二乗検定)
中間値の差 (ふたつの独立標本)	中間値の差に対する分散分析表
比率の差	中間値に対する二方式分散分析表

---

あなたが選ぶ統計学的手順：

A. 人類学者たちは同じ地域にふたつの埋葬地を発見した。彼らは、その地域にはいくつかの部族が暮らしていたこと、そしてそれらの部族は頭蓋骨の長さの違いによって分類されてきたことも知っている。彼らは、それぞれの埋葬地から発見され無作為抽出された頭蓋骨の標本を測り、そのふたつの埋葬地が異なる部族によって作られたのか決定したいと考えた。(確率値 = 0.0082)

B. ハワイの農園主協会では糖分をより多く含む果肉を生産するため、3つの新種のパイナップル (A、B、C と呼ぶ) を開発中である。それぞれの種類で 20 本 (全部で 60 本) が無作為に 2 エーカーの畑に分散された。収穫後、パイナップルの結果として糖分の含有量が測られた。そして生産高が各種類に対して記録された。3つの種類で、糖分の平均含有量に大きな差があるか? (確率値 = 0.987)

C. 研究者たちは、子供向けテレビ番組における暴力シーン調査モニター (VICTIM) に、土曜の朝のテレビ番組における暴力的演技描写の頻度について研究するよう依頼された。彼らは、12 週間にわたってこの時間帯に放送された 40 番組の無作為標本を抽出した。標本 40 番組のうち 28 番組が暴力描写シーンが過剰であると判定された。土曜朝のテレビ番組すべての母集団について述べるのに、彼らはどのようにこれらの情報を使うべきか?

D. 就職指導室では、最上級生の計画と彼らがどのように専門を関連づけていくのか、関心

を持っている。多数の学生が調査され、彼らの専攻（自然科学、社会科学、人文）や将来の計画（大学院、就職、未定）によって分類されている。専攻のタイプと将来の計画は関連しているか？（確率値 = 0.047）

E. 2年生マガジンでは15歳の無作為サンプルに、性交経験があるかどうか（はい、いいえ）尋ねた。男子と女子では回答に差があるかどうか見るものであった。（確率値 = 0.029）

F. ベトナム戦争中、死体の（殺された敵の数）が各軍の部隊から毎週報告された。これらの数の最後の桁はかなりでたらめだったと思われる。しかし、その数がでっちあげだったかもしれないという疑惑が持ち上がった。これをテストするために、死体のカウント数の大きな無作為標本が調べられ、最後の桁が0あるいは5だった頻度が記録された。でたらめな数をでっちあげている人たちがこれらの桁を使う頻度は、でたらめな可能性が示す頻度よりも低いと、心理学者たちは述べている。（すなわち、103の方が100とカウントするよりも「本当」らしく聞こえる。）もしデータが本物のカウント数なら、0か5で終わる数の比率は、約0.20となるだろう。（確率値 = 0.002）

G, シャーロック ホームズの冒険のひとつで、彼は犯行現場で犯人が残した足跡を発見し、それらの距離を測った。多くの人の見本をとり身長と歩幅の距離を測り、彼は容疑者の身長を予想できると自信を持って発表した。どうやって予想するのか？

#### 項目 24

自宅保有者に売られるタイプのラドン検知器は、どの位正確なのか？この疑問に答えるため、大学の研究者たちは、ラドン1リットル当たり105ピコキュリーまでさらす特別な部屋に検知器を12個置いた。表示された探知機の数値は以下のものであった。ミニタブからの記述統計学のプリントアウトがそれに続く。

91.9 97.8 111.4 122.3 105.4 95.0  
103.8 99.6 96.6 119.3 104.8 101.7

変数	N	Mean	Median	TrMean	StDev	SE Mean	Minimum	Maximum	Q1	Q3
数値	12	104.13	102.75	103.54	9.40	2.71	91.90	122.30	96.90	109.90

1. このタイプすべての探知機の間接値20という数値は、105という真値とは異なっているというもっともらしい証拠はあるのか？ $\alpha = .05$ . で適切な仮説検定を行いなさい。
2. この問題と関連しているタイプ I エラーは何か？

3. この問題と関連しているタイプIIエラーは何か？
4. もし、探知機の数値が5ピコキュリー低い(105であるべきところを実際は100を示す)なら、タイプIIのエラーの確率は何か？

### 項目 25

アメリカ食品医薬品局 (FDA) の研究によると、コーヒー一杯には平均 115mg のカフェインが含まれており、カップごとの量は淹れ方によって 60 から 180mg の幅がある。自分の好きな淹れ方を使って、95%で 5mg 以内のカフェイン含有の中間値を概算するために、FDA の実験をやってみたいと仮定しなさい。このような問題では、母集団の標準偏差を範囲の 4 分の 1 まで概算することができる。何杯のコーヒーを淹れなければならないか？

### 項目 26

ある特定の薬を適用することで、禿げた男性に髪を修復できるとある広告では主張している。この主張を試すために使用する実験計画の概要を作りなさい。この実験で 20 人の禿げた男性に使用するお金があると想定しなさい。

### 項目 27

幼児の鉄分不足に関する研究では、様々な食事養生計画に従っている幼児のサンプルを比較した。ひとつのグループには母乳で育った幼児を含んでおり、一方他のグループの子供たちは鉄分の補強を含まない粉ミルクで育っている。ここに、12 カ月の時点での血中のヘモグロビンレベルに対する結果の概要がある。

グループ	N	$\bar{X}$	s
母乳	23	13.3	1.7
粉ミルク	19	12.4	1.8

子供たち(母乳で育ったものと粉ミルクで育ったもの両方)の血中のヘモグロビンレベルは普通に分散していると想定しなさい。観察された違いの統計学的重要性を決定するために、有意性検定を行いなさい。

### 項目 28

+0.4 の相関関係と-0.6 の相関関係では、どちらの方が強い直線的関係を示しているか？自分の選択について簡潔に説明しなさい。

### 項目 29

内科医のグループが、医療診断テストに使われたテストと同じ入念なテストに対して、嘘発見器を受けさせた。もし 1,000 人の人たちが嘘発見器を受けさせられ、500 人は本当のこ

とを言い、500人は嘘をついたということになったら、嘘発見器は、本当のことを言っている人のうち約185人は嘘つきで、嘘つきのうち120人は本当のことを言っていたということを示すだろうということ、彼らは発見した。嘘発見器テストの応用では、その人が嘘つきであると示されるまで、各個人は本当のことを言う人だと推定される。この問題の背景でタイプIエラーは何か？この問題の背景でタイプIエラーの確率は何か？この問題の背景でタイプIIエラーは何か？この問題の背景でタイプIIエラーの確率は何か？

### 項目 30

聴覚学者たちは、高齢者のためのカナダのプログラムで、聴覚障害のある患者用のリハビリテーションプログラムを最近開発した。ある特定の高齢者ホームの住人から単純無作為に抽出された30人のサンプルと高齢者たちは、以下のようにコード化された感音性難聴の程度とタイプで診断された：1 = 正常の限度内で聞こえる 2 = 高周波難聴 3 = 軽度難聴 4 = 軽～中程度難聴 5 = 中程度難聴 6 = 中程度～重度難聴 7 = 重度～完全難聴。データは以下のとおりである：

6	7	1	1	2	6	4	6	4	2	5	2	5	1	5
4	6	6	5	5	5	2	5	3	6	4	6	6	4	2

1. データの箱ひげ図を描きなさい。
2. データについて適切な説明をきなさい。
3. このカナダのプログラムにおける、高齢者の難聴の中間値に対して95%の信頼区間を求めなさい。上記のデータの中間値と標準偏差は、それぞれ4.2と1.808である。その区間を説明きなさい。

### 項目 31

公益事業の会社では、農業関連の顧客は、ピークの時間の料金が異なるなら、この時間帯に電気を使わないようにするのかどうか知りたいと思った。顧客は標準料金のままいくのか、時間帯によって違う構造でいくのか、無作為に割り当てられた。ピーク時とそうでない時間帯の使用を記録する特殊なメーターが設置された；メーターを読み取った技術者は、各顧客がどちらの構造の料金を割り当てられているのか知らなかった。

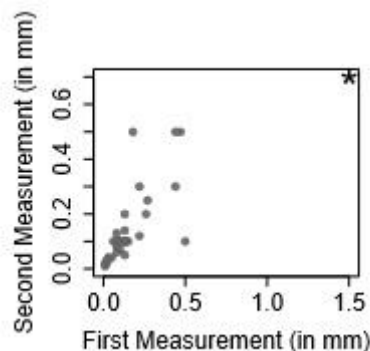
1. これは観察研究なのか、実験なのか？自分の意見を擁護きなさい。
2. 説明変数、反応変数は何か？
3. この取り組みにおける潜在的交絡変数をリストきなさい。
4. これは整合計画なのか？自分の答を弁護きなさい。

### 項目 32

学期の始めに、私たちは統計学の教科書のページ幅を測った。下は最初に測ったもの 対 二

番目に測ったものの分散図である。

1. この分散について述べなさい。
2. アステリスクがついている点とついていない点の相関関係を概算しなさい。



### 項目 33

レジャー調査ジャーナルにおける研究は、学業成績とレジャー活動との関係を調査した。159人の高校生の被調査母体のうち、それぞれの人が毎週いくつぐらいのレジャー活動に参加しているのか答えるよう依頼された。そのリストの中で、読書、作文、算数を含む活動は、「学習的レジャー活動」と分類された。結果のいくつかは下の表にある：

	中間値	標準偏差
GPA	2.96	0.71
レジャー活動の数	12.38	5.07
学習的レジャー活動	2.77	1.97

これらの数字に基づき（そして、GPAは0と4の間の値であり活動の数はマイナスにはなりえないということを理解して）、上記の変数それぞれに対して起こりうる歪みについて討論しなさい。

### 項目 34

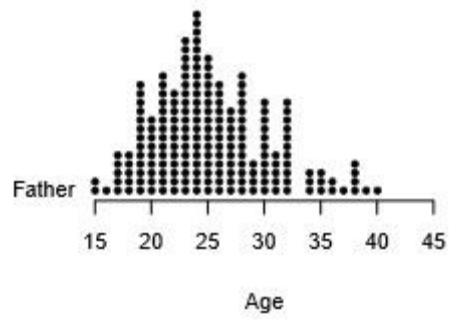
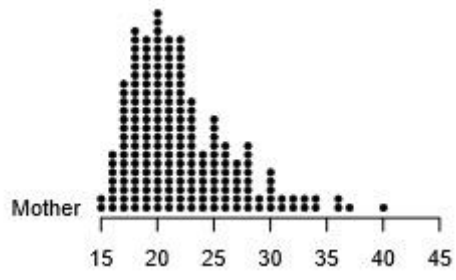
イベントAとBは互いに素である。AとBが独立したものになりえるかどうか討論しなさい。

### 項目 35

200人の母親の被調査者母体と、200人の父親の被調査者母体が抽出された。第一子をもうけた時の母親の年齢と、第一子をもうけた時の父親の年齢が記録された。

1. 母親の年齢に対するデータについて説明しなさい。
2. 父親の年齢に対するデータについて説明しなさい。
3. 分布を比較しなさい。
4. ふたつの母集団を比較したい時、年齢の相関関係を調べるためにある提案がされる。これは良い提案か？それはなぜそうなのか、そうでないのか？





## 付録 C : 技術の使用例

この例は、現実社会の背景から始まり、トランプを使って学生に物理的に模擬実験させ、模擬実験を自動化するためにコンピューター技術を持ちこんでいる。

### コカイン中毒の治療効果を調べる技術を基盤とした擬似体験

コカイン中毒の治療研究は、中毒患者がコカインから離れているのに効果があるふたつの薬を、比較する実験の結果を発表した (DM バーンズ、「コカイン中毒のサイクルを破れ」、サイエンス 241 号、1988、1029-1030 ページ) 治療を求めている 48 人のコカイン中毒者のグループは、無作為に 24 人ずつのふたつのグループに分けられた。ひとつのグループは、デシプラミンという新しい薬で治療され、一方他のグループはリチウムを与えられた。結果は下の表にまとめられ、その表では、元に戻らなかった患者をうまく治療されたとみなした。

	元に戻らず	元に戻る
デシプラミン	14	10
リチウム	6	18

私たちは、この特殊な実験において、デシプラミンのほうがリチウムよりうまくいったということを観察したが、この改善は統計学的に重要だと結論づけてよいのだろうか。(つまり、薬が同じ位効果的で、それはデシプラミンのグループでもっとうまくいった事例がたまたまその位多く出る無作為の分け方だったとしたら、そのような大きな違いがでると予想しただろうか。) わたしたちは、模擬実験を通してこの疑問に答える。まずトランプを切ることを基本とした実演を実際に行う。それから治療グループの中毒患者の無作為な分け方で違いが見られるコンピューターの模擬実験を使って述べる。

### 物理的模擬実験

54 枚ひと組のトランプ (2 枚のジョーカーを含む) を取りあげ、黒いカード (スペードかクラブ) 6 枚を取り除く。残りのトランプはコカインの実験における被験者にマッチさせる。赤いカード全部とジョーカーは元に戻ってしまった患者を表し、20 枚の黒いカードはうまく治療できた患者を表している。このひと組のトランプをシャッフルし、それぞれ 24 枚ずつの山になるようにふたつに配ると、中毒患者の分け方を、成功がどちらの薬を飲んだかに依らないふたつの治療グループに分ける模擬実験ができる。これを行い、それぞれのグループを「成功」(黒いカード) と「元に戻る」(赤/ジョーカー) でカウントし、二方向の表に記入しなさい。

	成功	元に戻る
デシプラミン		
リチウム		

いったん表にひとつの数を記入すると、残りも埋められるということに気づきなさい。わかっているように、各治療グループに 24 人いて、20 人は後戻りしない、28 人は戻る（そういう理由から  $2 \times 2$  の表にはたった 1 自由度しかない時々いうのである）。それでは内容を単純化するために、デシプラミングループの「元に戻らない」の数のように、ひとつのカウントを見てみよう。

トランプを再び、すべてシャッフルし、デシプラミングループに 24 枚配る。そして、黒いカードの枚数を数える。

デシプラミングループの「元に戻らない」の数 = \_\_\_\_\_

クラスで結果を構想の中にためておきなさい（デシプラミングループに分けられた 24 枚のカードの各無作為グループ黒いカードの数を数える）。実際の実験で観察された 14 の事例と同じくらい（またはもっと）大きな数は黒いカードではどの位頻繁にでてきたか？

もともとのデータの確率値は、どちらの薬も同じ位効果的だと想定して、14 以上の「元に戻らない」の事例がデシプラミングループに行くという無作為の割り当ての比率である。クラスの構想の中にあるデータを使って比率を概算しなさい。

### コンピューター模擬実験

14 以上の元に戻らないケースをデシプラミングループに入れる無作為割り当ての比率をより正確に概算するために、コンピューター模擬実験に移る。

ふたつの列と 48 行でできている集合データ（オンラインで提供）から始める。最初の列（治療法）には「デシプラミン」値が最初の 24 行にあり、「リチウム」が残り 24 行にある。2 番目の列（結果）には「元に戻らない」値と「元に戻る」値が、コカイン実験よりもともとあった  $2 \times 2$  の表にあるデータに合わせるために、書かれている。

結果がどの薬を摂取したのかによらない、治療グループの被験者の新しい無作為割り当てを表すために、コンピューターで「結果」の列の値を並べかえなさい。デシプラミンの治療グループの「元に戻らない」事例の数を数えなさい。そしてその結果をどこかに残しておきなさい。このプロセスを 1,000 回<sup>11</sup>自動的に繰り返しなさい。

1,000 回の模擬実験の数の分布のヒストグラムかドットプロットを見なさい。デシプラミングループの 14 もの「元に戻らない」事例は、普通とは思えないか？

デシプラミングループに 14 以上の成功がある模擬実験の数を数えなさい（グラフから、あるいは実行可能なら、模擬実験カウント列を区分して）、そしてもともとのデータの確率値のもうひとつの近似値を得るため、1,000 ごとに分けなさい。成功した事例のより大きな数（14）が、偶然デシプラミングループに現れたというのは道理にかなっていると思えるか？あるいは、デシプラミンはコカイン中毒の治療にリチウムよりよく効くかもしれないと、結論づける方がより適切なのだろうか？

- <sup>11</sup> 技術の二者択一：ここで最も難しい段階は、多くの無作為割り当てをすることである。例えばファソムのように、パッケージの中には全くそういった目的のために設計された使用が簡単なツールもある。他にもミニタブのように、前もって作っておき、ループ（繰り返し）で繰り返せるマクロスを使ってちょっとしたプログラミングができるものがある。いくぶん啓発的に欠ける模擬実験が、超幾何分布から無作為なデータの生成もできる統計学のパッケージで、行える。そうすると学生は物理的な無作為化とのつながりを見失ってしまうだろうが。最後に、やる気のある指導者は、必要な模擬実験を行ったり、結果を収集するのにアプレット構成する（ウェブ上で見つかるかもしれない）ことができる。

## 付録 D：裏付けのない実際的な現実のデータ例

### 未補正のデータ（勧められない）

下記のデータの最小 2 乗直線を求めなさい。X = 5 のときの Y を予測するのにそれを使いなさい。

X	1	2	3	4	6	8
Y	3	4	6	7	14	20

### （より良いがまだ最善ではない）実際のデータ

下のデータはレストランの 6 つのテーブルの客の数と、食事後テーブルに残されたチップの多さを示している。テーブルで食事した人の数から、チップの額を予測するために、データを使って最小 2 乗直線を求めなさい。結果を使い 5 人で食事したテーブルのチップの額を予測しなさい。

食事した人	1	2	3	4	6	8
チップ	\$ 3	\$ 4	\$ 6	\$ 7	\$ 14	\$ 20

批評：作られたデータには何の背景もない（勧められない）。問題は、意義のある解釈の可能性がなく、ただ単に計算するだけということにある。

批評：問題を、より魅力的にする背景が加えられ、学生に実用的な統計学の使用を示している。

### 本物のデータ（勧められる）

下のデータは、小テストの点数（20 点中）と中間考査の成績（100 点中）を、前期でこのコースを取った 8 人の学生のサンプルについて示している。これらのデータを使い、小テストの点数から中間考査の点数を予想する最小 2 乗直線を求めなさい。小テストと中間考査は今期は同じ位の難易度であり、同じ直線関係を今年適用すると想定し、小テストで 17 点取った学生の中間考査で取る成績は何点が予想されるか？

小テスト	20	15	13	18	18	20	14	16
中間考査	92	72	72	95	88	98	65	77

批評：データは、そのコースを取っている学生にとって、興味がある実際の背景から取りあげている。