
外国語教育研究ハンドブック

研究手法のより良い理解のために

【増補版】

考えてみよう！

解答例

第 1 章

- 1) 研究をはじめるとピックを選ぶ際にどのような方法があるでしょうか。

答：文献調査（入門書，オンライン検索，他の論文など）

- 2) あるプログラムが成功したかを調査するために，英語の総合能力を測定する事前テストと事後テストを実施しました。効果を調査するためにはその他に何をしなければならないでしょうか。

答：対照群を作る。

第 2 章

- 1) 妥当性と信頼性の違いは何でしょうか。

答：妥当性は「測定しようとしているものを正しく測定している程度」であり，内容や構成概念に関すること。信頼性は「その測定を何度やっても，同じ人には同じ結果が得られるであろう精度」のこと。

- 2) 信頼性係数が低い場合には，測定がうまくいっていない（誤差が多い）という原因以外にどのようなことが考えられるでしょうか。

答：項目数が少ない，受験者数が少ない，受験者の能力が等質でばらつきが少ない。

第 3 章

- 1) ある生徒の中間テストの英語の得点が 50 点，期末テストの英語の得点が 40 点だとします。この場合，当該生徒の学力は低下傾向にあるといえるでしょうか。

答：この段階ではいえない。各テストの平均点や標準偏差といった情報が追加され，この生徒の置かれている相対的な位置やテストの難易度がわかった時点で，判定が可能となる。

- 2) 平均点 45 点, 標準偏差 15 のテストで, ある生徒の得点が 80 点であったとすると, この生徒の偏差値はいくらになりますか。

答: $(80 - 45) \div 15 = 2.33$ が z スコア。これに 10 倍して 50 をたすと, 73.3 (73) という高い偏差値が得られる。

第 4 章

- 1) 文系クラスと理系クラスの実力テストにおける英語の点数を比較して, 文系クラスの点数が統計的に有意に高いか調べたいとします。この場合の帰無仮説は何で, 対立仮説は何でしょうか。また, なぜ推測統計をもちいる必要があるのでしょうか。

答: 帰無仮説は「2つのクラスの母平均に差はない」。対立仮説は「2つのクラスの母平均に差がある」。推測統計は, 今回の文系クラスと理系クラスだけではなく, 日本全国の同レベルの文系学習者と理系学習者まで結果を一般化させることを目指しているため使われる。

- 2) 統計的検定を行う前 (実験前) にすべきことは何でしょうか。また検定結果を解釈する際に, p 値以外に何に注意しなければならないでしょうか。

答: 検定を行う前には, どれぐらいのサンプルサイズがあれば, 検定力が一定以上 (0.8 以上) に保てるか検定力分析で調べてから必要サンプルサイズを集める。検定後の解釈では, p 値以外に効果量や検定力 (もしくは第 2 種の誤り) に注意しなければならない。

第 5 章

- 1) 英語の中間テストの結果をみて, 1 組 ($n = 40$) と 2 組 ($n = 38$) の間に平均点で差があるようにあなたは感じました。そこで t 検定をもちい仮説を検証することにしました。この場合, 「独立した t 検定」と「対応のある t 検定」のどちらの t 検定を使うべきでしょうか。また, 有意な差があると認める基準としては, p の値をいくりに設定すればよいで

しょうか。

答：「独立した t 検定」をもちいる。 p は 0.05 以下 ($p < .05$) にする。

- 2) 上のケースで、1組と2組の平均点の差を、リスニング、ライティング、リーディング、合計点の4分野にわけて確認しようと考えました。この際、どのようなことに気をつければよいのでしょうか。

答： t 検定を繰り返して利用すると、第1種の誤りが起りやすいため、ボンフェローニの補正を利用する。

第6章

- 1) 同じ学年の5つのクラスが受験した英語の模擬試験の結果（平均点）に統計的な差があるかどうか確認しようと考えました。この場合、「対応なし」と「対応あり」のどちらの一元配置分散分析をもちいればよいのでしょうか。

答： 同じ受験者が受けたテストではないため、「対応なし」をもちいる。

- 2) あるクラスの生徒が1年間に受けた5回の英語の模擬試験の結果（平均点）に差があるかどうか確認しようとしたとき、どのようなことに気をつけなければいけないのでしょうか。

答： 5回の模擬試験の内容が比較に値するものであるかに気をつける。たとえば、1回目の試験がリスニング（中心）の試験で、2回目の試験がリーディング（中心）の試験であった場合、その平均点の比較には意味がないかもしれない。

第7章

- 1) 二元配置分散分析で、主効果、交互作用ともに有意でした。次にどのような下位検定を行えばよいのでしょうか。

答： 要因の水準ごとに単純主効果の検定を行い、その結果が有意であれば、研究目的に応じて多重比較を行う。

- 2) 3つのクラスで3種類の英語テキスト (A1, A2, A3) を使って授業を行いました。それぞれのクラスで、英語の習熟度と動機の高さを測定し、クラスごとに差があるかどうかを調べたいとします。この場合、どのような分散分析をもちいればよいでしょうか。またその理由も述べましょう。

答：多変量分散分析 (multivariate analysis of variance) を行う。従属変数が英語の習熟度、動機の高さの2つあるため。

第8章

- 1) 指導の前後で同じ生徒が回答した、「4. 強くそう思う」から「1. まったくそう思わない」までの4件法の質問紙で得られたデータの差を検定しようと思います。尺度が順序尺度であると考えたため、ノンパラメトリック検定をもちいることにしましたが、本章であげた4つの方法のうち、どれをもちいればよいでしょうか。

答：対応のある2群のデータ分析であるため、ウィルコクソンの符号付順位和検定をもちいる。

- 2) 1クラス20名の生徒が3回受けたテストの結果に統計的な差があるか検討しようと思いました。パラメトリック検定をもちいるべきか、ノンパラメトリック検定をもちいるべきか、どのように決定すればよいのでしょうか。

答：データの正規性の前提が満たされているかどうか、コルモゴロフ・スミルノフ検定やシャピロ・ウィルク検定で検討する

第9章

- 1) ある英語リーディングテストとある英語語彙テストのスコアの相関が $r = .55$ であったとします。2つのテストがスコアの散らばりを共有している割合はいくらになるのか計算してみましょう。

答： $r = .55$ を二乗して、 $r^2 = .30$ 、つまり30%の共有率であることがわかる。

2) ある会社で、社員の年齢構成と生まれ月をしらべて表9-4のようなデータを得ました。相関を計算すると $r = .51$ という相関係数でしたが、これを根拠に「年齢」と「誕生月」の間の関係は比較的強い、と主張してよいでしょうか。

年齢	誕生月	年齢	誕生月
29	1	51	11
26	2	60	9
33	3	30	4
33	7	31	8
45	6	45	6
47	12	23	7
32	5	41	6
22	9	25	6
56	10	27	2
55	5	30	3

N=20

▲表 9-4 ある会社での社員の年齢と誕生月

答： そのように主張するには無理がある。理由は、(1) 年齢が比例尺度、誕生月は名義尺度と考えられるため、ピアソン相関の適用条件に反している危険性が高いこと、および (2) 常識的にみても、また理論的にみても、両者には関係が期待できないため、サンプルサイズの少なさなどにより偶発的に高い相関係数が得られた危険性が高いこと、の2つ。

相関係数は、明らかに無関係なデータ・セット（任意の2つのデータ・セット）を対象にしても出すことができるため、高い値が出たからといって、「強い関係がある」とはすぐに主張しない方がよい。なお、通常は、先行研究や理論の枠組みなどを参照にして、関係が

ありそうな変数を選び出し、相関分析を行うため、今回のような例は回避することが可能。

第 10 章

- 1) 大学入学時の TOEIC スコアから大学卒業時の TOEIC スコアを予測する場合、単回帰分析と重回帰分析のどちらをもちいますか。また、予測変数と目的変数はそれぞれ何ですか。

答：単回帰分析。予測変数は大学入学時の TOEIC スコア、目的変数は大学卒業時の TOEIC スコア。

- 2) 大学入学時の TOEIC スコアから大学卒業時の TOEIC スコアを予測したいのですが、大学入学時の TOEIC スコアが正規分布していないことがわかりました。この場合でも、回帰分析を使用することは可能でしょうか。

答：可能。回帰分析において、予測変数は正規分布している必要はない(10.4 の適用条件 7 を参照)。

第 11 章

- 1) 3つのクラスで「授業の予習は行いますか」という質問について調査したところ、表 11-11 のような結果が得られました。カイ二乗検定を行う場合、事前に確認する前提はどのようなものと考えられるでしょうか。

	はい	いいえ	ときどき
A 組	19	6	15
B 組	9	19	11

▲表 11-11 質問への回答人数

答：1) データが名義尺度であるか、2) データが累積の頻度であるか(パーセンテージではないか)、3) 同じ人が2度カウントされていないか(データが独立しているか)、4) すべてのセルで期待値が5以上であるか、5) どのような分割表になるデータか、という前提

を確認する。

- 2) 大学の英語教科書 95 冊においてグループ活動が取り扱われている回数を、教科書の種類別（リーディング用、ライティング用、オーラル用）に調べたところ、表 11-12 のような結果になりました。しかし、このままではカイ二乗検定を適用することはできません。なぜでしょうか。また、どのように改善すれば、カイ二乗検定を行うことができるでしょうか。

	リーディング用	ライティング用	オーラル用	合計
回数	30	25	50	105

▲表 11-11 質問への回答人数

答：同じ 1 冊の教科書での取り扱い回数の多さが影響している危険性があり、データが独立しているといえないため、カイ二乗検定を適用することができない。改善策としては、グループ活動の「回数」別にデータまとめるのではなく、1 冊の教科書での取扱回数を 0～5 回、6～10 回、11 回以上などのカテゴリーにわけ、それぞれに入る教科書の「冊数」をまとめる（以下の表を参照）。

	リーディング用	ライティング用	オーラル用	合計
0～5 回	16	16	6	38
6～10 回	8	9	8	25
11 回以上	6	6	20	32
合計	30	31	34	95

第 12 章

- 1) 学習方略についての質問紙調査を行いました。因子の間にはある程度の相関が予想されます。そのときに使用する因子軸はどのような回転が適切でしょうか。

答：斜交回転（プロマックス回転）

- 2) 因子分析を行った結果、3 因子解が得られました。しかし項目 1 が第 1 因子と第 2 因子の両方にほぼ同じ負荷量を持っています。項目内容を検討しても、どちらの因子に入っても構わないものでした。この場合はどうしますか。

答：その項目を除外して再分析を行う。

第 13 章

- 1) コレスポンデンス分析を使うと何ができるでしょうか。

答：クロス集計表の行と列の関係を少数の次元にまとめ視覚的に提示することができる。

- 2) 「英語総合力」を示す点数を、合計点ではなく、たくさんの変数から効率よく合成するには多変量解析のうち、どの手法が使えるでしょうか。

答：主成分分析。

第 14 章

- 1) 生徒から集めたテストのデータを SEM で分析します。SEM を適用する際に満たすべき前提は何でしょうか。

答：受験者数は十分か、観測変数の正規性があるか、欠損値はないか。

- 2) 論文を読んでいると、SEM を使った分析が行われていました。とても興味があり、著者が調べていないモデルも検証してみたいと思います。論文の Appendix に観測変数の平均値・標準偏差が掲載されています。

これらの情報を使い、モデルを検証することはできるでしょうか。

答： 相関行列（または分散共分散行列）がないため、検証はできない。

第 15 章

1) 古典的テスト理論と項目応答理論の違いは何でしょうか。

答： 古典的テスト理論は素点を使い、順序尺度であり、受験者能力はテストに依存し、項目困難度は受験者集団に依存する。項目応答理論はロジットが値の単位となり、間隔尺度であり、受験者能力と項目困難度は独立して推定できる（本文表 15-1 の内容）。

2) ラッシュ・モデルで受験者能力や項目難度のミスフィットは、何を意味しており、その原因はどのようなことが考えられるでしょうか。

答： ラッシュ・モデルにデータがあっていない。まぐれ当たりや、測定しようとしている構成概念以外の内容が含まれている可能性がある。

第 16 章

1) 通常の研究では、学習者 1 人ひとりデータをデータと見なします。メタ分析を行う際に、このデータに相当するのは何でしょうか。

答： 先行研究（内の効果量）

2) メタ分析を行った結果、効果量の異質性の検定が有意になりました。その後、どのような分析を行うべきでしょうか。

答： 仲介変数ごとに先行研究を区分・統合する。

第 17 章

1) 2 つの異なる科学的立場とは何ですか。

答： 1 つが、私たちの認識活動の外に客観的な世界の存在を認める立場。
もう 1 つが、世界の有り様は、私たちの言葉を基盤とする認識活

動に規定されており、私たちが、ありのままに客観的世界を認識することはできないとする立場。

- 2) 質的研究に求められる3つの要件とは何ですか。また、なぜ3つの要件は必要なのでしょう。

答：厚い記述，調査者自らのリフレクション，データと分析プロセスの開示または明示。この3つの要件を満たすことによって，反証可能性を保証することができるから。

第 18 章

- 1) 刺激提示による再生面接法においては，どのようなものが刺激として使えるのでしょうか。

答：観察記録，フィールドノートの記述，ビデオや音声の記録，日記・ジャーナル，脳波等の生理学的データ，レポート・テスト結果など。

- 2) (面接者と)異なる文化圏出身の被面接者を対象にするとき，どのようなことに注意をすればよいのでしょうか。

答：質問の等価性（ある質問が相手の文化でも同じ意味を持つのか），被面接者の代表性（被面接者をその文化の代表者としてとらえてよいのか），疑似類似性（たとえば volunteer とボランティアのように，一見同じようで異なるニュアンスを持つ言葉），異なる学校教育制度のような背景知識の違い，日本文化への馴染み具合（留学生の場合）。

第 19 章

- 1) SC-GTA には，理論的飽和に達するために，どのような「比較」の方法がありましたか。

答：1つが，収集されたデータにもとづく比較で，もう1つが，収集されたデータと架空の状況との比較。後者を理論的比較と呼ぶ。

- 2) 言語データの切片化の取り扱いについて、ストラウス・コービン版と修正ストラウス・グレイザー版の違いを説明しましょう。

答：ストラウス・コービン版は、言語データの切片化を行う。一方、修正ストラウス・グレイザー版は、切片化を行わない。そのかわりに分析ワークシートを使ってオープン・コーディングを行う。

第 20 章

- 1) SCQRM で質的研究法を修正する場合の注意点は何か。

答：修正の意思決定は研究目的と相関的に行い、修正理由と内容を研究目的にあわせながら明記する。

- 2) SCQRM では一般化可能性をどのように考えていますか。

答：一般化可能性は「A と B は類似性が高いため A の知見は B にも当てはまるかもしれない」という類推によって支えられている。

第 22 章

- 1) 次の文中の間違いを訂正し、APA スタイルに従って正しい形式にしてみましょう。

The researchers attempted to determine the relation between the age of the mother at the child's birth and (1) the child's intellectual development, (2) the child's social development: and (3) the mother's personal adjustment.

答： The researchers attempted to determine the relation between the age of the mother at the child's birth and (a) the child's intellectual development, (b) the child's social development, and (c) the mother's personal adjustment.

- 2) 面接者と次の文献情報を、APA スタイルの参考文献情報に変更してみましょう。

著者：Mills, N / Pajares, F/ Herron, C

発行年：2006年

ジャーナル名：Foreign Language Annals, 39巻, pp. 273-292

論文タイトル：A Reevaluation of the Role Anxiety.

サブタイトル：Self-efficacy, Anxiety, and their Relation to Reading and Listening Proficiency.

DOI：10.1111/j.1944-9720.2006.tb02266.x

答： Mills, N., Pajares, F., & Herron, C. (2006). A reevaluation of the role of anxiety: Self-efficacy, anxiety, and their relation to reading and listening proficiency. *Foreign Language Annals*, 39(2), 276–295. <https://doi.org/10.1111/j.1944-9720.2006.tb02266.x>

第24章

1) 頻度論の統計とベイズ統計の違いを書き出してみよう。

答： 頻度論の統計

- ・客観確率
- ・母数が定数、データが変数
- ・ p 値を使う
- ・信頼区間を使う
- ・Significant

ベイズ統計

- ・主観確率
- ・母数が確率変数、データが定数
- ・ベイズファクター (BF) を利用する。
- ・信用 (確信) 区間を使う。
- ・Compatible

2) ベイズ統計の問題点は何か考えてみよう。

答： ・事前分布の設定により結果が異なる

- ・結果が断定的ではないので、解釈に困る場合もある
- ・現時点では認知度が低い。

第 25 章

- 1) 「効果量が大きかったが有意差はなかった。サンプルサイズを増やせば、有意な差が得られるだろう」という主張は何が誤っているでしょうか。

答：効果量が大きかったとしても、小さいサンプルサイズから計算される効果量の信頼区間の幅が広いため、誤差が大きい。つまり、効果量の推定精度自体が低く、サンプリング・エラー（誤差）の影響がないとは言い切れない。また、区間推定での慎重な解釈が必要となる。

- 2) 線形モデル (LM), 一般化線形モデル (GLM), 一般化線形混合モデル (GLMM) のそれぞれの違いは何か説明してください。

答：線形モデル (LM) は t 検定, 分散分析, 回帰分析などの従来からよくもちいられる正規分布のデータを対象とした手法。一般化線形モデル (GLM) は, 正規分布以外の確率分布を持ったデータを扱えるようにした手法。一般化線形混合モデル (GLMM) は, さまざまな分布に対応できることに加えて, GLM を拡張して, モデルの中にランダム (変量) 効果 (random effect) を組み込んだ手法。

外国語教育研究ハンドブック【増補版】

研究手法のより良い理解のために

考えてみよう！ 解答例

2023年5月30日 作成

編著者———竹内 理／水本 篤

発行所———株式会社 松柏社

〒102-0072 東京都千代田区飯田橋 1-6-1

Tel. 03 (3230) 4813

Fax. 03 (3230) 4857

<https://www.shohakusha.com/>

本書を無断でコピー・スキャン・デジタル化等の複製をすることは、著作権上の例外を除いて禁じられています。本書を代行業者の第三者に依頼しスキャン・デジタル化することも、個人や家庭内の利用であっても著作権法上認められません。

Copyright © 2023 by Osamu Takeuchi, Atsushi Mizumoto, and other authors